Insurgency and Small Wars: Estimation of Unobserved Coalition Structures

Francesco Trebbi
Eric Weese

October 2017

Discussion Paper No.1628

GRADUATE SCHOOL OF ECONOMICS

KOBE UNIVERSITY

ROKKO, KOBE, JAPAN

Insurgency and Small Wars: Estimation of Unobserved Coalition Structures

Francesco Trebbi and Eric Weese*

October 2017

Abstract

Insurgency and guerrilla warfare impose enormous socio-economic costs and often persist for decades. The opacity of such forms of conflict is an obstacle to effective international humanitarian intervention and development programs. To shed light on the internal organization of otherwise unknown insurgent groups, this paper proposes two methodologies for the detection of unobserved coalitions of militant factions in conflict areas. Our approach is based on daily geocoded incident-level data on insurgent attacks. We provide applications to the Afghan conflict during the 2004-2009 period and to Pakistan during the 2008-2011 period, identifying systematically different coalition structures. Further applications to global terrorism data are discussed.

*University of British Columbia, Vancouver School of Economics, CIFAR and NBER, francesco.trebbi@ubc.ca; Kobe University, weese@econ.kobe-u.ac.jp respectively. The authors would like to thank Eli Berman, Ethan Bueno de Mesquita, James Fearon, Patrick Francois, Camilo Garcia-Jimeno, Jason Lyall, Carlos Sanchez-Martinez, Jake Shapiro, Drew Shaver, Austin Wright and seminar participants at UCSD, Penn, Chicago Harris, Columbia, Stanford, Berkeley, Rochester, UQAM, Kobe, Tokyo, Toronto, CIFAR, and Princeton for useful comments and discussion and the researchers at the Princeton University Empirical Studies of Conflict Project for generously sharing their incident data online. Nathan Canen provided excellent research assistance. We are grateful to the Social Science and Humanities Research Council for financial support.

1 Introduction

Among many political sources of welfare loss, few compare in magnitude to military conflict and, in the post World War II period in particular, to the losses ascribed to civil war and violent insurgency [O'Neill, 1990]. Insurgency, defined as armed rebellion against a central authority[1], is also one of the most opaque forms of conflict. Intertwining connections with the population blur the lines between combatants and civilians [Kilcullen, 2009]. The relative strength and even the identity of potential negotiating counterparties are often unclear, and in the words of Fearon [2008] "*there are no clear front lines.*" Such forms of conflict have disproportionately affected poor countries and are gaining central status in the literature on the political economy of development [Blattman and Miguel, 2010; Berman and Matanock, 2015]. Our paper offers a contribution to the economic analysis of these irregular wars that is both methodological and empirical.

The paper proposes two methodological approaches for the estimation of the latent internal organization of multiple insurgent groups. Motivating our interest in this particular problem is uncontrived. Take the case of Syria after 2011. Lack of knowledge about the organization of the anti-Assad insurgency and the structure of cross-alliances among militant groups emerged as a crucial obstacle to targeted Western military action and a deterrent to humanitarian relief intervention in the Syrian civil war [Jenkins, 2014]. The political risks of supporting anti-regime rebels (possibly aligned with radical Islamist movements) were deemed too high. In this backdrop, to April 2017, the Syrian conflict alone displaced 11.3 million civilians and caused the death of over $250,000$ individuals[2] according to the United Nations Office for the Coordination of Humanitarian Affairs [United Nations, 2016].

Empirically, we focus on the costly insurgencies of Afghanistan and Pakistan[3]. On the U.S. side alone, Afghan operations cost the lives of more than $1,800$ troops between 2001 and 2011, and more than \$444 billion in military expenses.[4] Statistics for

[1] According to O'Neill [1990] "*Insurgency may be defined as a struggle between a nonruling group and the ruling authorities in which the nonruling group consciously uses political resources (e.g., organizational expertise, propaganda, and demonstrations) and violence to destroy, reformulate, or sustain the basis of one or more aspects of politics.*"

[2] This number was last estimated in 2016. After that the UN stopped keeping track of death tolls due to inaccessibility of several conflict areas and lack of verification of conflicting claims by government and rebel forces.

[3] Unfortunately, we are not aware of suitable data for Syria. In Section 5 we further discuss the case cases of Iraq, Syria, and Libya, all instances where our methodologies could be potentially of use.

[4] Soon into the operation, the U.S. military acknowledged through a drastic adjustment in tactics

Afghan civilians appear less certain, but the adverse effects are painfully obvious even to the casual observer. In Pakistan, a perduring effort in placating sectarian insurgencies has been accompanied by a heavy toll of $18,583$ people killed and $19,356$ injured between 2012 and 2015 alone [Pak Institute for Peace Studies, 2016, p.8]. This effort has diverted valuable resources from aid and development assistance programs and public goods provision.

As a more concrete application, consider the case of Afghanistan specifically. The insurgency during 2004-2009 has been approximately described as originating from an alliance between the Afghan Taliban insurgents and al Qa'ida foreign fighters against the Afghan government and the supporting international security assistance force (ISAF). While this was surely a facet of this early phase of the Afghan war, front lines were uncertain and the unity of the insurgents doubtful. Importantly, policymakers disagree about whether the Taliban were a unified fighting organization, or rather an umbrella coalition of heterogeneous forces. Some were skeptical of the degree of control that Taliban leader Mullah Mohammed Omar exerted over the powerful Haqqani faction and the Dadullah network.[5] Similarly, the Hizb-i Islami faction was considered by many a separate entity from the Taliban proper.[6]

Other observers promoted an opposite view. In an insightful qualitative essay Dorronsoro [2009] states: "*The Taliban are often described as an umbrella movement comprising loosely connected groups that are essentially local and unorganized. On the contrary, this report's analysis of the structure and strategy of the insurgency reveals a resilient adversary, engaged in strategic planning and coordinated action*".Evidence in support of this position includes the existence of the *Layha* (a centralized code of conduct for Mujahidin), as well as the strong centralizing tendencies of the *Obedience to the*

that the Afghan conflict differed substantially from previous large-scale military operations.

[5] For example, the UN report [2013] stated that *"Despite what passes for a zonal command structure across Afghanistan, the Taliban have shown themselves unwilling or unable to monopolize anti-State violence. The persistent presence and autonomy of the Haqqani Network and the manner in which other, non-Taliban, groupings like the Lashkar-e-Tayyiba are operating in Afghanistan raises questions about the true extent of the influence exerted by the Taliban leadership."* Brahimi [2010] reports a statement by Ashraf Ghani, current Afghan president, in a lecture for the Miliband Programme at LSE indicating *"The Taliban are not a unified force - they are not the SPLA in Sudan or the Maoists in Nepal"* while Giustozzi [2009] states that "*The Taliban themselves are not fully united and the insurgency is not limited to the Taliban."*

[6] Smith [2005] believes that the Taliban are "not a single monolithic movement, but a series of parallel groupings." Fotini and Semple [2009] state explicitly that "*the Taliban is not a unified or monolithic movement*" and Thruelsen [2010] that "*the movement should not be seen as a unified hierarchical actor that can be dealt with as part of a generic approach covering the whole of Afghanistan.*"

Amir (a manual endorsed by Mullah Omar).[7]

This disagreement among experts is unsettling. Understanding the extent of territorial control and population support of insurgent groups is essential for military operations. Furthermore, knowledge of the internal organization and cohesion of rebel groups can be used to prevent selective violence by insurgents, and ultimately help with the relief and reconstruction of affected areas.[8]

This paper proposes a model of the attack organization of insurgency and shows how the covariance structure of data on violent events can be used to estimate the number of different active insurgent groups and their territories of influence, features that are typically unobservable to the econometrician. In this particular application we make use of the specific covariance structure emerging from the fact that insurgent groups with the ability to launch simultaneous and geographically separated attacks appear to do so. This relies on the conclusions of the existing literature regarding the signalling value for organized violent groups to launch coordinated attack in multiple locations, a mechanism that we also explore theoretically. Deloughery [2013] provides a recent review of this literature and presents systematic evidence of the advantages of simultaneous attacks for terrorist organizations in terms of media coverage and appeal in the recruitment of new fighters, incentives that operate within insurgencies as well.[9]

The covariance of the attacks that is observed in excess to random co-occurrence can be employed to analyze the underlying structure of the insurgent groups organization. Attacks across different locations on the same day are assumed either the result of random chance or signalling an insurgent group with a presence in those locations. A transparent model of attacks is used to distinguish between random chance and organized group behavior. After estimating the number of different guerrilla groups and their territorial extent, we assess the extent of coordination in attacks and present applications of our analysis to Global Terrorism Database (GTD) data.

The paper addresses several questions. First, when faced with multiple violent incidents in multiple regions, how can one decide whether the simultaneous incidents observed are isolated idiosyncratic events, as opposed to organized attacks by coalitions of assailants? Second, how can one identify from incident data alone how many distinct in-

[7] These are available in English translation as Munir [2011] and Ludhianvi [2015], respectively.

[8] One example of the importance of understanding insurgent group structures for post-conflict negotiations comes from Colombia. The recent appearance of the Bandas Criminales Emergentes (BACRIM) in lieu of the AUC paramilitary combatants has been a central issue in the work of Colombia's Reconciliation Commission in deploying resources and rebuilding state institutions and control at the local level.

[9] Additional citations for the qualitative literature are provided in Appendix A.

surgent groups (if any) are attacking? Third, can we quantify the extent of coordination in insurgent attacks?

A country experiencing an insurgency is described as a set of points at which violent incidents can occur in each period. Each point in this set represents the centroid of an administrative district and each period is one day. Attacks on the same day in two different districts will occur with greater-than-random frequency if the same insurgent group is operating in both areas and chooses to coordinate. Using a variety of assumptions regarding what the reference cross-district covariance in attacks would be in the case where there were no organized groups, we calculate which sets of districts are more correlated than would be expected by chance alone. We then use this information to estimate the cluster of districts in which each guerrilla group operates.

We present estimators that allow for a single district to be contested by multiple guerrilla groups or by a single group, or by none. The estimators provide the number of guerrilla groups operating, the geographic area of each of these, and the intensity of each group's activity in each district. Let us emphasize that the methods we present can accommodate slow-moving trends in violence over time and are robust to aggregate shocks (e.g. weather, seasonality, or U.S. troop movements) that might affect insurgent activity in many areas simultaneously.

The main empirical results of the paper are as follows. For the early period 2004-2009 insurgent activity in Afghanistan is best represented by a single organized group, rather than several independent groups, and that the extent of this group is largely determined by ethnic boundaries. This result is probed by partialling out slow-moving trends and spurious correlation induced by large-scale U.S. military activity, to constraining the analysis to districts with a number of incidents above specific thresholds, and to limiting the analysis to incidents explicitly claimed by the Taliban.We also conduct an analysis of the Pakistani insurgency (which includes the Pakistani Taliban, known as Tehrik-i-Taliban Pakistan or TTP), using multiple data sets for the period 2008-2011. In the case of Pakistan, our methodology detects multiple insurgent groups (four, in fact) and is completely consistent with the extant qualitative literature on insurgency in the country.

An increasing amount of attention has been devoted within the fields of development economics and political economy to the study of armed conflict within countries, in particular civil wars and insurgency. Both Political Science and Economics have provided some of the most recent and novel insights in the study of insurgency.[10] As underlined by

[10] These include Berman [2009], Berman et al. [2011], Condra and Shapiro [2012], and Bueno de Mesquita [2013]. Economists have been interested in the analysis of violence and conflict at least as far

Blattman and Miguel [2010], a remarkable characteristic of this recent wave of research has been a strong empirical bend and an increasing attention to micro-level (typically incident-level) information. The use of geocoded micro data in this area is a departure from more established "macro" empirical approaches, which were based on country-level information or aggregate conflict information. Some of this recent wave of research has also taken the direction, which we share, of focusing on the network structure of conflict. One such recent example is the work on the complex alliances in the post-Mubutu Congo wars studied by König et al. [2017].

This paper is one in the new "micro" style, with a specific emphasis on the analysis of insurgency and small wars. Economic and statistical evidence on the role of anti-government guerrilla activities is still sparse, even though such activities cause substantial damage worldwide and appear from a quantitative perspective to be the predominant form conflict in civil wars since 1945 [Fearon, 2008; Ghobarah et al., 2003]. Insurgents' strategies are generally not well understood, and neither are the subtleties of their interactions with the noncombatant population [Gutierrez-Sanin, 2008; Kilcullen, 2009]. A particular incentive for further study is that insurgent activity is also often linked to terrorist activities, and thus there is a connection with the growing literature on the economics of terrorism [Bueno de Mesquita and Dickson, 2007; Benmelech, Berrebi, Klor, 2012].

This work is also related to the conflict studies literature focused on the internal organization of insurgent and terrorist groups. Berman [2009] offers an analysis of the internal management of defection risk within terrorist groups. Shapiro [2013] reports varied accounts (including about al Qa'ida) of the organizational structures and discusses agency problems related to the trade-off between cohesion/effectiveness versus operational secrecy (a security-control dilemma). In our theory we discuss the complementary problem of signalling group strength to noncombatants in order to broadcast to potential recruits or induce support from the civilian population. With specific emphasis on Afghanistan, Staniland [2014, ch.5] discusses the Taliban organization in the periods 1994-2001 and post 9/11. While the author discusses the Taliban as an "integrated organization" with a defined centralized structure and unique leadership (p. 136-137), the presence of the Haqqani Network and of Hezb-i Islami is also considered. What seems clear from the conflict literature is that studying group organization may offer a useful perspective in understanding insurgency. This is the main goal of the paper.

back as Schelling [1960] and Tullock [1974].

2 A Simple Model of Insurgent Attacks

This section presents a model of insurgent behavior and choice of organization of attacks useful for the interpretation of our main empirical exercise.

Let $\mathcal{N}$ be the set of districts in a country, indexed by $i \in \mathcal{N}$, with a total of $N = |\mathcal{N}| > 2$ districts. Violent occurrences in i at time t, x_{it} , can be of two types: (i) unorganized; (ii) organized by an insurgent group.[11] Let $\ell_i \geq 0$ be the number of unorganized militants in i.

Each organized group j consists of: (i) a leadership; (ii) a number $\alpha_j > 0$ of rank-and-file members. Let $\mathcal{J}$ be the set of organized insurgent groups active anywhere in the country, indexed by $j \in \mathcal{J}$, and $J = |\mathcal{J}|$ be the total number of groups. Let $\alpha_{ij} \geq 0$ be the number of j rank-and-file members present in district i, and $N^j \geq 1$ the number of districts where group j operates. Note that $\alpha_j = \sum_i \alpha_{ij}$. $\mathcal{J}^i \subseteq \mathcal{J}$ is the subset of organized groups present in i (i.e. the groups for which $\alpha_{ij} > 0$ in i). Time is discrete, indexed by t.[12]

Leaders of organized insurgents are assumed to be rational, expected utility maximizers. We assume the leader's utility to depend on two separate dimensions of the attacks that the group launches: (i) their *signalling value*; (ii) their *military value*.

Noncombatants (like the econometrician) observe x_{it} but do not know if the attacks they witness are the result of organized versus random violence, nor the location and strength of j (i.e. $\mathcal{J}^i$, α_{ij}, and ℓ_i are unknowns). By the *signalling value* of the attacks that they carry out, we mean that j leaders place a preference weight $\zeta > 0$ on signalling to the noncombatant population j's span of control, N^j, and its strength α_j. The qualitative literature supports the idea that simultaneous attacks have a signalling motivation. For example, Barno [2006] gives a specific example of a simultaneous attack on three border checkpoints where the media appears to have been deliberately alerted to the attack. In this case, publicity appears to have been the main objective of the attack. According to Kilcullen [2009], "the insurgents treat propaganda as their main effort, coordinating physical attacks in support of a sophisticated propaganda campaign" (p. 58).

Based on the realized attacks, noncombatants solve a statistical identification prob-

[11]We make a distinction between attacks initiated by local unorganized militants and those initiated by members of organized groups, to allow for the possibility that there are no organized insurgent groups present in a district, even though violence may be observed there.

[12]In our analysis the time periods used will be days. This high-frequency attack data is useful in our application because it reduces the number of attacks that are simultaneous simply by random chance, a feature whose importance will be clear in what follows.

lem and infer N^j, α_j. Based on their estimates $\hat{N}^j, \hat{\alpha}_j$, they can rationally update their beliefs. The j leadership values this process, for instance because noncombatants' beliefs about j may start from priors that are unfavorable to the group's recruitment efforts, so updating to estimates closer to the true N^j, α_j may further the group's goals.[13] Let us indicate j's leaders' utility gain from signalling by $W(\hat{N}^j, \hat{\alpha}_j)$, increasing in both arguments and normalized to $W = 0$ if the noncombatants' problem is not statistically identified (i.e. noncombatants cannot draw any inference from the attack data given assumptions on beliefs and primitives).[14] We do not spell out the details of the noncombatants' updating process about the distribution of N^j, α_j, rather we focus the analysis on whether N^j, α_j are statistically identified or not.

By *military value* of an attack, we mean the immediate tactical value of carrying out a specific act of violence (e.g. disrupting a road block, killing an enemy combatant, gaining control of an hill post, etc.). We assume that each district i provides the leadership a time-varying, military value (benefit) $b_{ijt} \in \{b^l; b^h\}$ for each attack that may be carried out, where $0 < b^l < b^h$. Military value is either high or low with probability $p(b^h) = 1 - p(b^l) = \theta$, drawn identically and independently across groups, districts and time, and constant within groups, districts and time (i.e. at t, all attacks in i have identical military value to group j). At each period t the central leadership of j privately observes the distribution of military values across all districts (i.e. which districts have high value and which low value). Noncombatants do not observe the realization of military values.

2.1 Rebel Organization

The organizational design problem we propose follows a general group of problems commonly discussed in the political economy literature[15].

One rank-and-file member can execute one attack per period. To induce attacks, leaders provide costly incentives to their foot soldiers. There are numerous ways in which incentive schemes can be designed in insurgencies (Shapiro [2013]), but to discipline the

[13] The case in which noncombatants' beliefs may be favorable to the group can also be studied, but it is not considered here, as it trivializes the choice of attack organization below. In essence, we focus on insurgent groups that wish to project more strength than what it is currently attributed to them by noncombatants.

[14] This signalling aspect is a well known feature of insurgent behavior and a substantial qualitative narrative highlights the wider recruiting appeal of larger and stronger insurgent groups and the more effective terror campaigns associated with a broader group's reach (for references see Arce and Sandler [2007]). We use the function W as a reduced-form representation of these potential gains.

[15] See Besley and Persson [2017] and references therein.

complexity of this problem, we assume that group j's leadership can offer (monetary or nonmonetary) rewards r following two different *attack plans*:

Definition 1. *Plan 1 (Centralized): Let $\mathcal{N}^j \subseteq \mathcal{N} : j \in \mathcal{J}$ and $\alpha_{ij} > 0$, the central leadership of j each period t sets a unique reward r_{jt} for every rank-and-file member across all districts $i \in \mathcal{N}^j$ in the country.*

Definition 2. *Plan 2 (Decentralized): Let $\mathcal{N}^j \subseteq \mathcal{N} : j \in \mathcal{J}$ and $\alpha_{ij} > 0$, for every $i \in \mathcal{N}^j$ the central leadership of j delegates to a local leadership setting independently each period t a unique reward r_{ijt} for every rank-and-file member within district i.*

Plan 1 implies coordination of j rank and file within all districts in the country where j operates, irrespective of military value. This generates correlation in attack patterns, which will prove useful for signalling.[16] Plan 2 implies coordination of j rank and file only within district i. This may be preferable whenever district-specific military benefits are expected to differentially arise over time, for instance if a military base in i becomes suddenly vulnerable or state dignitaries are in official visit.[17] The simple dichotomy between plan 1 and 2 is sufficient to lay out a trade off between attack plans that privilege costly signalling of j's strength through coordinated efforts[18] versus flexibility to district-specific windows of opportunity for attack. The choice of coordination of insurgents' attacks arises from the trade off between value of signalling strength to civilian noncombatants relative to the value of capturing district-idiosyncratic military opportunities that may randomly arise (but jam the signalling value of attacks). While this is by no means the sole trade off governing insurgent strategies, it offers a coherent perspective on the degree of coordination observed in actual data.[19]

[16]The choice of having coordination through rewards is also supported by evidence from the literature, as that from Bahney et al. [2013]. While studying al Qa'ida in Iraq, they show cross-sectionally constant rewards: "*mean and median salaries are relatively constant across governorates, ranging between* 93,000 *and* 98,000 *[IQD]*" (p.520). The literature on insurgent compensation is more limited that that of insurgent coordination, but this evidence is consistent with plan 1. Plan 2 instead stylizes a more geographically targeted reward structure. Notice that the second part of plan 2 implies that it is costly to change organizationally from a decentralized to a centralized plan.

[17]It could also be thought of as if rewarding j members to attack becomes suddenly cheaper in district i than i'. For example, a local holy shrine is desecrated, pictures of counterinsurgency abuse are leaked, a drone strikes a peaceful gathering or a funeral.

[18]Insurgent coordination is achieved through mobile communication and ICT, as discussed in Shapiro and Siegel [2015].

[19]In Appendix B.1 we discuss how it is possible to nest the choice of attack plan within a full game of strategic offensive allocation against a rationally defending counterinsurgency force (e.g. for the Afghan case think of ISAF), as presented and characterized in Bier, Oliveros, and Samuelson [2007].

2.2 Timing of the Game and Strategies

The timing of the game is as follows. The attack plan $m = 1, 2$ are decided by all groups J at $t = 0$, for the duration of $T \geq 1$ periods. The reward amounts are decided every subsequent period $t \geq 1$.[20] At any such t, attack actions by the rank and file take place, and then per-period payoffs from military values realize.

Indicate with V_1^j j's leaders' present discounted expected total military value of carrying out attacks under plan 1 net of the reward payments to the rank and file for T periods. With V_2^j the same under plan 2.

The optimal plan will be decided at time 0 based on:

$$\max_{m=1,2} \left\{ V_m^j + \zeta W(\hat{N}^j, \hat{\alpha}_j | m) \right\}$$

subject to all other insurgent groups optimally selecting their attack plans and rewards and where the notation $W(.|m)$ indicates the differential utility gain from inference by noncombatants based on attack plan m. The only additional condition that we will impose to this optimization is for the reward menus to be binary. That is, r_{jt} or r_{ijt} will take two values $\{r^l; r^h\}$, high or low rewards for group attacks decided at $t = 0$, that possibly differ across groups and plans.

More specifically, an equilibrium strategy for group j will include a plan of attack m, a reward menu $\{r_j^l; r_j^h\}$, and a decision rule for how rewards are to be disbursed. This decision rule will depend on local or aggregate military opportunities which will generate an endogenous frequency of high rewards p_{ij} over T periods. Such value p_{ij} could be also interpreted as the "attack intensity" level. Each group strategies will also condition on the equilibrium strategy of all other groups (of which they are aware) and the solution of the identification problem faced by noncombatants. Finally, we exclude the possibility by groups of ex ante publicly communicating p_{ij} or any components of their attack strategy (other than the reward menu) to noncombatants at $t = 0$, as this would likely induce sizeable losses from counterinsurgency preemptive response, which we do not model here.

Putting aside strategic considerations across groups, the arms of the trade-off in attack plan choice can be intuitively illustrated. The decentralized plan 2 allows the leadership to tailor rewards to conditions on the ground. When the military value is

[20] Note that the assumption $T > 1$ is only necessary to justify the statistical analysis on a time series of attacks across districts over T consecutive periods. This is also for realism, as changing mode of organization is likely costly.

high in i, j's leadership can induce more attacks by increasing rewards r_{ijt}. Due to the idiosyncratic nature of the military benefits, however, this is detrimental to inducing attack correlation across districts pairs (i, i') and we will show this correlation will be useful for identification (exactly as in our empirical application). Under plan 2, as rewards are set independently, the (i, i') correlation will be 0 after T periods. Under the centralized plan 1, the leadership may instead suffer a mismatch between incentives and military values. That is, the group leadership may have to set high rewards even in districts where the military value of an attack is temporarily low in order to have all of its members attacking across districts in sync, and vice versa.[21] Anticipating our results, we will prove below that $W(.|m = 1) \geq W(.|m = 2)$.

There are however strategic considerations to take into account. Even if each group's payoffs depend only on the military value and signalling of own strength to noncombatants, it is possible that other groups's choices of attack plans or reward affect a group decision. The first strategic consideration is whether under plan 2 it will be possible to signal j_1's presence to noncombatants by increasing rewards r, thus differentially inducing more attacks (something akin costly signalling by "burning money") relative to a competing insurgent group j_2. While under plan 2 this will signal to noncombatants the presence of a group in the district, under the assumptions of our model, we can show that this will not solve the noncombatants identification problem. The second consideration is whether, under the assumption of, say, $J = 2$, it will be possible for group j_2 to gain identification/signalling value, by decentralizing when instead group j_1 decides to centralize. In other words, can j_2 gain identification simply because all other known groups choose a plan of attack that allows them to be identified?

Before tackling these issues, it is necessary to posit a microfoundation for rank-and-file decisions. This will define the properties of the attack data generating process at the basis of the statistical problem faced by noncombatants.

[21] In fact, it is easy to show that the net expected military value is higher when one can choose rewards in a decentralized than a centralized manner (i.e. $V_1^j \leq V_2^j$). Under plan 1 at best the leader will be able to select the reward level that maximizes the country-wide military value at t (i.e. will pick a high reward level when the majority of j's rank and file members are in high military value districts, and vice versa). We show this more formally in Appendix B.3. Under plan 1, coordinated rewards will induce coordinated attacks, however, which will signal presence in all districts in which j operates. Correlations of attacks will not be 0.

2.2.1 Rank and File Microfoundation

The microfoundation of the behavior of unorganized local militants and rank-and-file members is described by a static discrete choice problem.

Random Violence. At t, any unorganized local militant ι has the opportunity of committing an act of random violence, yielding a deterministic payoff $\bar{u}$ and an outside option from not engaging in violence, yielding deterministic utility $\underline{u}$. For each violent attack, ι also receives an i.i.d. type-I extreme value random shock $\bar{\varepsilon}_\iota$. This captures the possibility of ι's having idiosyncratic motives for attack. If ι decides not to attack s/he receives $\underline{\varepsilon}_\iota$, again i.i.d. type-I extreme value. We assume that if launched, an attack is successful. The probability that ι launches an attack is then given by:

$$\Pr\left(\bar{u} + \bar{\varepsilon}_\iota \geq \underline{u} + \underline{\varepsilon}_\iota\right) = \frac{e^{\bar{u}}}{e^{\bar{u}} + e^{\underline{u}}} = \eta.$$

η does not change over time (this assumption is relaxed in Section 3.4). As the decision by unorganized militants to attack is independent of the decision of anyone else (unorganized militant or group member), the number of random attacks per day in district i is distributed binomial $B\left(\ell_i, \eta\right)$, which for small η is approximated by $Poisson\left(\eta \ell_i\right)$. The covariance in unorganized attacks between two districts i and i' will be zero, as the attack decisions are independent and the probability of an attack is constant.

Organized Violence. Suppose individual ι is a rank and file member of group j. Under plan 1, the probability that ι launches an attack, given rewards r_{jt}, preferences $u\left(.\right)$, and outside option $\underline{u}$,[22] is:

$$\Pr\left(u\left(r_{jt}\right) + \bar{\varepsilon}_\iota \geq \underline{u} + \underline{\varepsilon}_\iota\right) = \frac{e^{u(r_{jt})}}{e^{u(r_{jt})} + e^{\underline{u}}} = \epsilon_{jt}. \tag{1}$$

ϵ_{jt} is the probability that a j member will attack at time t and is time varying. The higher r_{jt}, the higher the probability of an attack. Depending on the value of $r_{jt} \in \mathbb{R}$, j members will be more likely to attack on some days than others. Probability (1) is the same for all members of group j, and whether any given member attacks is independent of other attack decisions, conditioning on the attack probability ϵ_{jt}. The number of organized attacks by j in district i at t is therefore distributed $B\left(\alpha_{ij}, \epsilon_{jt}\right) \simeq Poisson\left(\alpha_{ij}\epsilon_{jt}\right)$. For the entire country this implies a time-varying distribution of attacks due to j's organized

[22] Note that the fact that the outside option is common across different types of militants is irrelevant for the analysis below, it can be relaxed, and $\underline{u}$ is not a parameter which the model will be able to empirically identify.

violence $Poisson(\alpha_j \epsilon_{jt})$. Under plan 1, the covariance of attacks across time between two j members is $\text{Var}(\epsilon_{jt})$ and depends on the sequence of rewards offered over time.

Under plan 2, the number of organized attacks by j at t in district i is distributed $B(\alpha_{ij}, \epsilon_{ijt}) \simeq Poisson(\alpha_{ij}\epsilon_{ijt})$, where $\epsilon_{ijt} = \frac{e^{u(r_{ijt})}}{e^{u(r_{ijt})}+e^{\underline{u}}}$. Hence, for group j, ϵ_{ijt} differs by both t and i.

Given the framework above and the independence between random violence and organized group activity, no matter the attack strategy of each group operating in i, total attacks x_{it} will follow Poisson mixture distributions, if organized groups are present, and Poisson distributions, if not.

2.3 Informational Value of the Plan of Attack

We now focus on the informational dimensions of the attack plan choices.

A final set of assumptions are needed to define the signalling problem. Let $J \leq 2$ (the case of higher J is discussed in Appendix B). We will further assume that:

Assumption 1: If $J > 0$, then there exists $j \in \mathcal{J}$ and $i \in \mathcal{N}^j$ such that (i) noncombatants have the prior $\Pr(i \in \mathcal{N}^j) = 1$; (ii) $N^j > 1$; and (iii) $J^i = 1$.

In other words, there exists an organized group j and a district i where noncombatants know that j is present in i, and j is the only group therein (e.g. the stronghold of the group), but that particular group also operates in more than one district.

Noncombatants are assumed to know η. Recall that the noncombatant population (like the econometrician) simply observe total attacks x_{it} over time. We can now prove that "money burning" cannot help in signalling for a group:

Lemma 1. *Consider two groups j_1 and j_2 only one of which is present in district i and both operating under a decentralized attack plan 2 with rewards set to induce $\epsilon^h_{j_1} \geq \epsilon^l_{j_1}$ and $\epsilon^h_{j_2} \geq \epsilon^l_{j_2}$, with $\epsilon^h_{j_1} \neq \epsilon^h_{j_2}$ and $\epsilon^l_{j_1} \neq \epsilon^l_{j_2}$. Assume noncombatants know both η and the respective $\epsilon^h_{j_1}$, $\epsilon^l_{j_1}$ and $\epsilon^h_{j_2}$, $\epsilon^l_{j_2}$. It is not possible for the noncombatants to identify α_{ij_1} nor α_{ij_2}.*

Proof. All proofs are in Appendix B.

Lemma 1 is useful in showing that the reward amounts per se cannot help groups in signalling their strength and in solving noncombatants' identification problem. This naturally limits strategic interaction on this particular design margin. Further, it also shows how even in presence of a single organized group, credit claiming by an otherwise powerless second group j_2 with $\alpha_{j_2} = 0$ cannot be easily countered just based on rewards. We will focus the analysis below on the case:

Assumption 2: ϵ^h and ϵ^l are common across all groups and known to the noncombatants.

A final assumptions is needed to make the statistical identification problem for noncombatants non-trivial (implying a sufficiently spread out presence of groups across districts):

Assumption 3: If $J > 0$, (i) there are at least 3 districts with an organized group presence; and (ii) at least 2 districts have $|\mathcal{J}^i| < J$.

We discuss in detail the sensitivity of our main results to Assumptions 1-3 in Appendix B.2. We can state:

Proposition 1. *Let Assumptions 1-3 hold. Then, for any finite large T, the informational gain for noncombatants from attack plan 1 is larger than that from plan 2. Specifically:*
(i) If all groups adopt a centralized plan 1, noncombatants will be able to identify parameters N^j and α_j, for all j.
(ii) If all groups adopt a decentralized plan 2, noncombatants will not be able to identify N^j and they will not be able to identify α_j, as long as the noncombatants' prior places non-zero weight on the possibility that there are two groups.
(iii) If $J = 2$ and j_1 operates under plan 1 and j_2 under plan 2, noncombatants will be able to identify N^{j_1} and α_{j_1}, but not N^{j_2} and α_{j_2}, as long as the noncombatants' prior places non-zero weight on the possibility that there are more than two groups.

The proposition is useful to frame the choice of attack behavior. Based on the result of proposition 1 and our discussion of flexibility of attack, we can posit that if the value placed by the j group leadership on the signalling gain from plan 1, $W(\hat{N}^j, \hat{\alpha}_j | m = 1)$, is sufficiently valuable relative to the flexibility gain from plan 2, as dictated by the preference parameter ζ, then plan 1 will be selected. Otherwise plan 2 will be selected.[23]

Proposition 1 operates under the assumption of military value benefits b_{ijt} being idiosyncratic and private to each organized group j. This implies groups cannot ride

[23] In making these considerations, it is worth noticing that the expected difference $V_2^j - V_1^j$ may be an upper bound of the flexibility gains from plan 2, as j may decide to endogenously change some district-specific rewards to induce correlation of attacks for a subset of districts (see Appendix). Full characterization of this problem will be function of the specific assumptions on the distribution of members across districts, the random violence parameters, and the extent of the gains from flexibility of attack. These are all parameters unobserved in the empirical analysis that follows and not essential to our estimation methods, therefore such exercise is not pursued further here.

each other's coattails in order to project strength beyond what they actually hold. If we relax such assumption and allow groups to coordinate, we can show:

Proposition 2. *Let Assumptions 1-3 hold and* $J = 2$*. Then, for any finite large* T*, no group* j_1 *can project* $\hat{N}^{j_1} > N^{j_1}, \hat{\alpha}_{j_1} > \alpha_{j_1}$ *by imperfectly coordinating attacks with another group* j_2 *(i.e. Corr*$(\epsilon_{j_1,t}, \epsilon_{j_2,t}) < 1$*). For any finite large* T*, two groups* j_1 *and* j_2 *cannot both project* $\hat{N}^{j_1} > N^{j_1}, \hat{\alpha}_{j_1} > \alpha_{j_1}$ *and* $\hat{N}^{j_2} > N^{j_2}, \hat{\alpha}_{j_2} > \alpha_{j_2}$ *by perfectly coordinating attacks (i.e. Corr*$(\epsilon_{j_1,t}, \epsilon_{j_2,t}) = 1$*).*

The proposition shows how imperfect coordination cannot be a solution for signalling power beyond what actually held by the group. It achieves the same signalling value of centralized attacks uncoordinated across groups. If imperfect coordination across groups is minimally costly, then it will likely be avoided. If perfect coordination across groups is achievable, projecting both higher N and α for both groups will not be feasible. The intuitive reason is that noncombatants will still be able to assess the joint strength of the union of the two groups across districts as if it were a single group (per proposition 1). Their sum is given, so both groups cannot project higher α.

2.4 Additional Discussion

We maintain the working assumption that for any pair of organized groups under plan 1, j, $j' \in \mathcal{J}$, $\text{Corr}(\epsilon_{j,t}, \epsilon_{j',t}) = 0$. Daily frequency of attacks can help ruling out the possibility of certain groups riding the coattails of others, placing us most likely within the first part of Proposition 2 (where groups would not gain from $0 < \text{Corr}(\epsilon_{j,t}, \epsilon_{j',t}) < 1$).[24] But this is also to say that, if two groups j and j' can coordinate in their use of time-specific rewards (and on the size of such rewards) and do so *systematically at daily frequency*, we will consider them de facto the same organized group (see proof of Proposition 2 for a discussion).[25] This limitation of our approach applies to what is arguably a very small subset of groups.

Consider the members of group j under plan 1. If there are α_{ij} members in district i and $\alpha_{i'j}$ members in district i', then the covariance in attacks over time between these two districts, due to the presence of members of group j, is $\alpha_{ij}\alpha_{i'j}\sigma^2$, where σ^2

[24]Under this assumption, it follows that the covariance of attacks between two members of different groups is zero. This orthogonality condition parallels what is standard (but less theoretically argumented) procedure in the factor analysis literature.

[25]Besides the organizational complexity of matching day after day the attack behavior of another group, there is also a potential loss of signalling value of the group identity, as noncombatants will be uncertain about whether it is j or j' launching the attacks, as shown in the proof of Proposition 2.

is the variance $\mathrm{Var}(\epsilon_{jt})$ and below will be assumed constant across groups, a necessary normalization because, differently from the theoretical analysis above, we do not have access to information on ϵ. Summing over members of all groups following the same reward strategy, the covariance over time in total attacks x between districts i and i' is:

$$\mathrm{Cov}(x_{it}, x_{i't}) = \sigma^2 \sum_{j \in \mathcal{J}} \alpha_{ij} \alpha_{i'j}. \tag{2}$$

It is not necessary to assume knowledge of the optimal plan of attack of insurgents in order to perform the analysis that follows, but it will be evident from the data that plan 1, with simultaneous attacks, is playing a role in our empirical setting. Simultaneous insurgent and terror attacks across districts are documented and frequent in both Afghanistan and Pakistan.[26]

There are features of the data that could potentially reject plan 1. To provide an example here, consider that the structure introduced under plan 1 imposes the very specific covariance structure (2). We have assumed that the members of an insurgent group do not move between districts (a given group j has a certain membership α_{ij} in district i) and j members will either be encouraged to attack in a given period (a high ϵ_{jt}) or not (low ϵ_{jt}). It follows that (2) has to be strictly positive if the same insurgent group j has members in both i and i', as attacks in both i and i' will be higher in periods when ϵ_{jt} is high and lower in periods when ϵ_{jt} is low. Consistently with these assumptions, in the data the observed covariances $\mathrm{Cov}(x_{it}, x_{i't})$ are systematically positive, in accord with what is implied by plan 1.[27] The qualitative research of Deloughery [2013] and others, as discussed in Section 1, also buttresses this finding. Within our setting the obvious alternative of employing plan 2 without coordination (or with limited partial coordination) would imply for a vast majority of districts $\mathrm{Cov}(x_{it}, x_{i't}) = 0$. This appears counterfactual.[28]

[26]See our discussion of the Global Terrorism Database evidence in Section 5. 15% of all events recorded in the Global Terrorism Database are coordinated attacks. We will proceed under the assumption that the selected plan is 1, with the obvious caveat that the detection of groups employing plan 2 would require much longer district-specific time series of attacks necessary for the estimation of the district specific Poisson mixture components discussed in the proof of Proposition 1.

[27]Permutation tests of the sort discussed later indicate that the mean covariance is positive at any reasonable confidence level. Results available upon request.

[28]Outside our framework, a very different mechanism could be one in which members of an insurgent group are mobile and in any given period have the choice of moving and attacking in one of many districts. This alternative model would imply that organized groups movement should lead to negative covariances between districts i and i', as insurgent group members who attack in district i could not

3 Estimation

This section presents our estimation methods for parameters describing the number of insurgent groups and their presence across locations.

3.1 Covariance Decomposition

Let Γ be the covariance matrix for attacks discussed in Section 2, where the entry in row i and column i' gives the covariance in attacks across time for these two districts. Analysis will be based on this matrix, and others created from it. The covariance matrix Γ can be decomposed as $\Gamma = \Gamma_D + \Gamma_L$, where Γ_D is a diagonal matrix and Γ_L is a low rank matrix of the form:

$$\Gamma_L = \sigma^2 \begin{bmatrix} \sum_j \alpha_{1j}\alpha_{1j} & \sum_j \alpha_{1j}\alpha_{2j} & \dots & \\ \sum_j \alpha_{2j}\alpha_{1j} & \sum_j \alpha_{2j}\alpha_{2j} & & \\ \dots & & \sum_j \alpha_{ij}\alpha_{ij} & \\ & & & \dots \end{bmatrix} \quad (3)$$

This decomposition is considered because the diagonal entries of the covariance matrix are a sum of variance from unorganized militants and variance from organized groups, and only the latter is of interest.[29]

As a normalization, we set $\sigma^2 = 1$. Let $\hat{\Gamma}_L$ be the resulting estimate of the low rank matrix: for details, see Appendix C.[30]

3.2 Non-overlapping Insurgent Groups

We desire both an estimate $\hat{J}$, the total number of organized insurgent groups, as well as an estimate $\hat{\alpha}_{ij}$ for each district i and group j, giving the number of insurgent

also be attacking in district i' in the same period. This is also counterfactual.

[29]The diagonal entries of Γ do not in general have a useful form. The situation even with mixture Poisson distributions does not appear to be simple: see Ashford and Hunt [1973] for the Poisson-Gamma distribution, and Karlis and Xekalaki [2005] for mixture Poisson distributions in general.

[30]Because of this decomposition, we effectively do not use pure random uncoordinated violence as part of our procedure for the detection of groups: random violence acts only as noise and is not used as part of the estimator. Thus, areas where some group adherents may be present, but that group has no organizational capacity will not be detected by our method. For example, ISIS may have successfully radicalized some individuals in North America and convinced them to conduct attacks, but these were all lone wolf style attacks, because ISIS does not have the infrastructure to safely coordinate attacks in North America.

members of the group operating in that district. The set of estimates $\{\hat{\alpha}_{ij}\}$ will have a total of $N \times \hat{J}$ elements. It turns out to be easiest to first produce the $\{\hat{\alpha}_{ij}\}$ estimates for various values of $J \in \{1, ..., J_{\max}\}$, and then choose a $\hat{J}$ based on examining this set of estimates.[31] We will thus begin by assuming that J is known, and consider how to compute estimates $\{\hat{\alpha}_{ij}\}$ given J. After this, we will then consider how to choose $\hat{J}$.

An approach based on standard clustering techniques will be presented in this subsection, as well as one based on eigenvalues. Matrix factorization will be considered in Section 3.3.[32]

Estimation via standard clustering techniques requires an additional assumption different from those that will be needed for Section 3.3. Specifically, it is necessary to assume that the various insurgent groups present do not have overlapping territories. That is, there is one organized group j present in any given district i.[33] Based on this assumption, reordering the districts i allows Γ_L to be written as a block-diagonal matrix:

$$\Gamma_L = \begin{bmatrix} \Gamma_L^1 & ... & 0 \\ ... & \Gamma_L^j & ... \\ 0 & ... & \Gamma_L^J \end{bmatrix} \tag{4}$$

where there are a total of J organized groups, and each block Γ_L^j has the form given in Equation 3. To produce estimates $\{\hat{\alpha}_{ij}\}$ we will first determine which organized group is present in each district, and then we determine the strength of this group in the district.

To determine which organized group is present in each district, we will follow a modified k-means type approach[34]. Begin by constructing a scaled version of Γ_L:

$$\Gamma_L^{\text{cor}} = D(\sum_j \alpha_{\cdot j}\alpha_{\cdot j})^{-1/2}\Gamma_L D(\sum_j \alpha_{\cdot j}\alpha_{\cdot j})^{-1/2},$$

where $D(.)$ indicates a diagonal matrix with the specified vector on the diagonal. This

[31] The exact choice of $J_{\max}$ is not important.

[32] The approach based on clustering and that based on matrix factorization are largely complementary in that they rely on different assumptions. This provides a form of cross-validation for our results, which is important given the novelty of these methodologies within the field of political economy.

[33] This is a direct consequence of the standard assumption that factors should be orthogonal, combined with the fact that insurgent prevalence α must be non-negative.

[34] In many clustering approaches transformations or scaling of the original data is common. An example is spectral clustering (see Luxburg [2007]), a methodology we also employed in the working paper version of the paper. We discuss this approach, and reasons why our current approach may be preferable, in Appendix E.

process is occasionally referred to as "sphering" and it often improves the quality of the clustering. By assumption, for each district i, $\alpha_{ij} = 0$ for all but one group j, and thus Γ_L^{cor} is constructed by dividing row i and column i of Γ_L by the value of α_{ij} for the single group j that is present in i.

The "cor" superscript is used because Γ_L^{cor} is positive semi-definite with all diagonal entries equal to one, and thus has the form of a correlation matrix. However, observe that each off-diagonal entry $\gamma_{ii'}$ has now been divided by $\alpha_{ij}\alpha_{i'j}$ if the same insurgent group is present in districts i and i'. Thus, in exactly the same way as (4), after suitable rearrangement Γ_L^{cor} is a block diagonal matrix with entries consisting only of zeros and ones:

$$\Gamma_L^{\text{cor}} = \begin{bmatrix} 1_{N_1} & \ldots & 0 \\ \ldots & 1_{N_j} & \ldots \\ 0 & \ldots & 1_{N_J} \end{bmatrix} \tag{5}$$

where 1_{N_j} is an N_j by N_j matrix consisting entirely of ones, and corresponding to the N_j districts that have group j present in them.

Running a k-means type clustering algorithm on Γ_L^{cor} would be trivial, but only the finite sample version is available. Let $\hat{\Gamma}_L^{\text{cor}}$ be the correlation matrix associated with the finite sample covariance matrix $\hat{\Gamma}_L$.[35] This $\hat{\Gamma}_L^{\text{cor}}$ will have off-diagonal entries that are neither zero nor one. For many k-means algorithms, a distance matrix rather than a correlation matrix is needed. Such a distance matrix can easily be constructed using cosine distances: $1 - \gamma_{ii'}^{\text{cor}}$ is the cosine distance between i and i', where $\gamma_{ii'}^{\text{cor}}$ is the off-diagonal entry of Γ_L^{cor} corresponding to districts i and i'.[36] The cosine distance between two districts with the same group present will be zero asymptotically, while it will be one when the districts have different groups present. Given J, any reasonable clustering algorithm should thus be able to recover which insurgent group is present in which district, given enough data.[37] Once districts have been clustered into groups, estimates

[35] The correlation matrix $\hat{\Gamma}_L^{\text{cor}}$ is readily obtained by imposing diagonal elements equal to 1 and appropriately rescaling rows and columns of the covariance matrix $\hat{\Gamma}_L$ by the square root of the corresponding diagonal entry of $\hat{\Gamma}_L$.

[36] The construction of a distance matrix is trivial because any correlation matrix is also an interpoint angle matrix, and these angles can be used directly to construct a cosine distance matrix.

[37] A weighted clustering approach appears to be called for, because a district with very low α_{ij} for the group j that is present will have very noisy off-diagonal entries. We do not explore optimal weights, instead using ad-hoc weights corresponding to the square root of the diagonal entries of $\hat{\Gamma}_L$. Krishna and Narasimha [1999] provide a weighted k-means algorithm, based on genetic optimization: we use the Hornik, Feinerer, Kober, and Buchta [2012] implementation of this algorithm. Using unweighted clustering instead does not change any of the results discussed below substantially.

for $\{\alpha_{ij}\}$ can be obtained. Appendix D provides the relevant details.

We now consider how to produce an estimate $\hat{J}$ of the number of insurgent groups present. A major difficulty we face in determining the number of groups is that it is not obvious how to construct a null distribution for potential test statistics. For example, suppose that we wished to test for $J > 2$ versus the null hypothesis that $J = 2$. The distribution of plausible test statistics under the null would in general depend on whether one of these two groups is very large compared to the other, or whether they are of roughly equal size. The only case where this difficulty is avoided is in the test of $J > 1$ versus the null hypothesis that $J = 1$, because under the null there are no nuisance parameters, as there is only one group present and thus every district must be assigned to that group. In this special case where the null hypothesis is $J = 1$, we will show that a permutation test can be constructed due to the simplicity of the group structure under the null. Our approach will thus be based on the repeated splitting of groups, as the only test we have available is one that asks whether a group should be split in two. The technique we use turns out to match that of Bruzzese and Vistocco [2015], except that in their case they assume that their data has a hierarchical form, whereas in our case we are dealing with a block diagonal correlation matrix that does not have any hierarchy.

We begin by looking at the set of all districts $\mathcal{N}$. Run a standard clustering procedure using distances based on Γ_L^{cor} to split these districts into two clusters, $\mathcal{N}_1$ and $\mathcal{N}_2$. Let $Q(\mathcal{N}_1, \mathcal{N}_2)$ be some test statistic that takes a high value when the division of $\mathcal{N}$ into $\mathcal{N}_1$ and $\mathcal{N}_2$ looks like a "good" division. To determine whether we actually want to split the $\mathcal{N}$ districts into the two groups $\mathcal{N}_1$ and $\mathcal{N}_2$, we will use a permutation test to calculate a cutoff value for Q.

If we do not split $\mathcal{N}$ into two groups, we are done and our estimate of the number of groups is $J = 1$. If we do split $\mathcal{N}$ into two groups, we apply our method recursively. That is, let our new set of districts be $\mathcal{D} = \mathcal{N}_1$, compute a clustering of these districts into clusters $\mathcal{D}_1$ and $\mathcal{D}_2$ using the relevant portion of Γ_L^{cor}, and then test whether we should actually split $\mathcal{D}$ into $\mathcal{D}_1$ and $\mathcal{D}_2$. If we do split, we continue the recursion downwards. If not, we move to considering $\mathcal{D} = \mathcal{N}_2$. At the end of this procedure, we will have a partition of $\mathcal{N}$ into groups, where each of these groups should not be split further according to a permutation test.

The standard choice of permutation test would be to follow Bruzzese and Vistocco [2015], and consider permutations of attacks that generate different Γ^{cor} matrices. The test statistic Q would be based on how much of the covariance matrix can be explained by splitting the districts being considered into two groups, rather than leaving them as a single group. This standard approach runs into problems with the calculation of the null distribution of Q, because a reference distribution for $\hat{\Gamma}_L^{\text{cor}}$ needs to be calculated.[38] This calculation appears to be extremely complicated, because the answer depends on the finite sample behavior of $\hat{\Gamma}_L^{\text{cor}}$, which is not well understood. We avoid this problem by modifying the Bruzzese and Vistocco [2015] approach, and use a Q defined with respect to a set of auxiliary covariates Z, rather than Γ^{cor}.[39]

To see why this simplifies the problem, note that in the model the only source of correlation in insurgent attacks across districts is through ϵ. In particular, our model assumes that if the same insurgent group is present in both districts i and i', the correlation in attacks between districts will not depend on the relationship between any other covariates of i and i'. For example, it does not matter whether i is geographically adjacent to i', or geographically distant. We will now add one additional assumption. Suppose that the districts where a given insurgent group is present are less dispersed in terms of these auxiliary covariates Z than a set of randomly chosen districts. For simplicity, we will focus specifically on geography, with Z_i being a vector indicating which other districts are geographically adjacent to i, but our approach is potentially more general.

Let $Z_{ii'} = 1$ if districts i and i' are geographically adjacent. Let $Q(\mathcal{D}_1, \mathcal{D}_2)$ describe the geographic dispersion of the insurgent group territories when the set $\mathcal{D}$ of districts are being split into two groups, according to the following formula:

$$Q(\mathcal{D}_1, \mathcal{D}_2) = \sum_{i \in \mathcal{D}_1} \sum_{i' \in \mathcal{D}_1} Z_{ii'} + \sum_{i \in \mathcal{D}_2} \sum_{i' \in \mathcal{D}_2} Z_{ii'} \tag{6}$$

That is, the test statistic Q is a very simple calculation regarding whether $\mathcal{D}_1$ and $\mathcal{D}_2$ represent distinct geographic regions (in which case Q should be high, as there are a great many adjacencies), or whether the districts in $\mathcal{D}_1$ and $\mathcal{D}_2$ are randomly interspersed (in which case Q should be low). We do not worry about normalizing Q with respect to the total number of adjacencies, because the threshold value of Q will be computed using exactly the same sets of districts $\mathcal{D}_1$ and $\mathcal{D}_2$.

The intuition for this test statistic is that, if all the districts in $\mathcal{D}$ are really part of the same group, then a split of these districts into two groups $\mathcal{D}_1$ and $\mathcal{D}_2$ will be based on finite sample noise, which is by assumption uncorrelated with geography. Thus, the

[38] Specifically, we would need to know how much adding a second group should improve model fit, if there is only actually one group in the data.

[39] The standard approach is presented in Appendix Table H.1. The results are the same as in the main text.

additional groups should not be correlated with geography, and thus values of Q should be quite low, as the group labels are randomly assigned. A threshold value is easy to generate using Montecarlo permutations of the group structure: randomly permute the identities of all the districts, thereby forcing group membership to be unrelated to geographic location.

One does not have to exclusively rely on the Q statistic to estimate the number of groups J. In fact, this parameter is also recoverable using an entirely different approach, one based on the spectral properties of Γ_L. Notice that each Γ_L^j in (4) has rank 1, implying the rank of Γ_L is J.[40] The rank of Γ_L can be then consistently estimated by applying the intuition of Ahn and Horenstein [2013], using the eigenvalues of Γ_L.

Ahn and Horenstein's "eigenratio" approach proceeds as follows. Suppose that we were interested in estimating the rank of Γ_L. Let $\hat{\lambda}_k$ be the $k-$th largest eigenvalue of $\hat{\Gamma}_L$. Asymptotically, the first J of these eigenvalues will be positive and bounded away from zero, while the remaining $N - J$ will go to zero. Ahn and Horenstein consider the "eigenratio"

$$\text{ER}_k = \hat{\lambda}_k / \hat{\lambda}_{k+1}. \tag{7}$$

Asymptotically, ER_k will converge to some positive value c_k for $k < J$. However, it will diverge to infinity for $k = J$, as the denominator becomes increasingly close to zero while the numerator remains bounded away from zero. A simple estimate for $\hat{J}$ can then be obtained by choosing the $\hat{J}$ that gives the highest value for $\text{ER}_{\hat{J}}$.[41]

3.3 Potentially Overlapping Insurgent Groups

We now relax the assumption that insurgent groups do not overlap. As in the approach discussed above, we will begin by assuming that J is known, and estimate $\{\alpha_{ij}\}$. We then produce an estimate $\hat{J}$ based on a comparison of these estimates for different values of J. The estimates for $\{\alpha_{ij}\}$ will be based on non-negative matrix factorization, and the $\hat{J}$ estimate will be based on a modification of the ER_k discussed above.

We first construct an estimator for the $\{\alpha_{ij}\}$, given an assumed number of groups J.

[40] This is because the vectors $\alpha_{\cdot j}$ and $\alpha_{\cdot j'}$ describing insurgent group presence are orthogonal for $j \neq j'$.

[41] Ahn and Horenstein [2013] require that there be some exogenous maximum number of possible factors, $J_{\max}$. We follow this, and use $J_{\max} = 80$ for this paper. Simulations are provided in Appendix G.

Consider choosing $\hat{\alpha}_{ij}$ for each district i and group j to satisfy the set of restrictions

$$\hat{\gamma}_{ii'} = \sum_j \hat{\alpha}_{ij}\hat{\alpha}_{i'j}.$$

where $\hat{\gamma}_{ii'}$ is the relevant entry in $\hat{\Gamma}_L$, estimated in (15) in Appendix C. If there are N districts, there are $N(N+1)/2$ restrictions: one for each off-diagonal element in one half of the symmetric covariance matrix, plus the diagonal elements. If there are J groups, there are $N \times J$ parameters to be estimated: one $\hat{\alpha}_{ij}$ for each district i and group j. A necessary condition for identification is thus that $(N+1)/2 \geq J$.

In the data the number of districts is large relative to plausible numbers of groups, and thus this inequality holds strictly and a penalty function is required. An obvious estimator for $\{\alpha_{ij}\}$ would then be the squared Frobenius norm

$$\underset{\hat{\alpha}_{ij} \geq 0}{\operatorname{argmin}} \sum_i \sum_{i'} \left(\hat{\gamma}_{ii'} - \sum_j \hat{\alpha}_{ij}\hat{\alpha}_{i'j} \right)^2. \tag{8}$$

Unfortunately, solving this optimization problem directly by searching the space of $\{\alpha_{ij}\}$ is challenging because the problem as stated is non-convex in $\{\alpha_{ij}\}$. A variety of algorithms have been proposed for solving this problem. We will use the "Procrustes rotation" algorithm of Huang, Sidiropoulos, and Swami [2014]. This algorithm does not attempt to minimize (8), but instead solves a related optimization problem based on a spectral decomposition of $\hat{\Gamma}_L$. Huang and Sidiropoulos [2014] show that this algorithm is effective at solving (8), despite the fact that this objective is not used as part of the algorithm.[42]

We now consider how to produce an estimate $\hat{J}$. In Section 3.2, the rank of Γ_L was J, because the vectors $\alpha_{\cdot j}$ and $\alpha_{\cdot j'}$ describing insurgent group presence would be orthogonal for $j \neq j'$. This is no longer true if groups have the potential to overlap. Instead of the rank of Γ_L, we thus base our estimate $\hat{J}$ on the completely positive rank of Γ_L: that is, the rank of A, where $\Gamma_L = AA^{\mathrm{T}}$, and all entries of A are non-negative. Without further

[42]See Appendix F for more details. A previous version of this paper optimized (8) directly, using the algorithm of Birgin, Martinez, and Raydan [2000]. The (qualitatively identical) results from this direct approach are available upon request. Huang, Sidiropoulos, and Swami [2014] is orders of magnitude faster, converging in seconds or minutes rather than hours or days.

assumptions this decomposition is not identified. For example,

$$\Gamma_L = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{9}$$

could be decomposed either into $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ or $\tilde{A} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Huang and Sidiropoulos [2014] summarize assumptions under which the non-negative factorization of Γ_L becomes unique for practical purposes:[43] each factor must have at least $J - 1$ non-zero entries, and the non-zero entries of one factor must not be a subset of the non-zero entries of any other factor. In the example above, the second assumption is violated by matrix $\tilde{A}$. We will assume that the Huang and Sidiropoulos [2014] assumptions are satisfied. Thus, faced with the covariance matrix in (9), we would conclude that $\hat{J} = 1$ based on the factorization employing matrix A.

If Γ_L were known, the number of organized groups could thus be calculated immediately by producing a non-negative factorization of Γ_L. However, only the finite sample $\hat{\Gamma}_L$ is actually available, and in general this matrix will not have a non-negative factorization due to finite sample variation.

To address this problem, we will use a modification of the eigenratio approach. The intuition behind the Ahn and Horenstein [2013] approach appears very general. Consider a rank k approximation to an $N \times N$ matrix. The first k eigenvectors can be used to create such an approximation. How much better is a rank $k + 1$ approximation? If the $k + 1$th eigenvalue is very small relative to the kth eigenvalue, then considering a rank $k + 1$ matrix instead of a rank k matrix does not improve the approximation very much, and ER_k will thus be very high. We can apply this intuition to the case of the group structure of $\hat{\Gamma}_L$. Let A_k be an approximate non-negative factorization of $\hat{\Gamma}_L$ with k factors. How much better would A_{k+1} be as an approximation to $\hat{\Gamma}_L$? Asymptotically, if Γ_L was produced by k groups, the improvement will be zero.

A ratio equivalent to Ahn and Horenstein's "eigenratio" can then be expressed as

$$\text{NNR}_k = \frac{||\hat{\Gamma}_L - A_k A_k^{\text{T}}||_F^2 - ||\hat{\Gamma}_L - A_{k-1} A_{k-1}^{\text{T}}||_F^2}{||\hat{\Gamma}_L - A_{k+1} A_{k+1}^{\text{T}}||_F^2 - ||\hat{\Gamma}_L - A_k A_k^{\text{T}}||_F^2} \tag{10}$$

where $||.||_F$ is the Frobenius norm. The intuition for NNR_k is exactly that of the ER_k:

[43] Conditions theoretically guaranteeing the uniqueness of the factorization are more complicated: see the references in Huang and Sidiropoulos [2014].

if Γ_L has a completely positive rank of k, then the $k+1$th factor should not help explain Γ_L, and thus NNR_k should diverge to infinity. In contrast, values of NNR_k for $k < J$ will converge to finite values.[44]

The ER estimator has a finite sample tendency to estimate $\hat{J} = 1$, because the eigenvalues of random matrices are generally distributed so that the first few eigenvalues are spaced further apart than most of the remaining eigenvalues.[45] This effect has been noted previously by Ferson and Kim [2012], and Guo-Fitoussi and Darne [2014] perform an extensive simulation-based analysis.[46] The attack datasets that we consider in this paper, however, are noisier than the data generally used by researchers studying factor models in macro or finance. We are thus particularly interested in the finite sample properties of the estimator when the signal-to-noise ratio is very low. In Appendix G, we provide figures illustrating the behavior of the eigenratio estimator as the signal vanishes. The simulations that we perform are effectively identical to those conducted in Guo-Fitoussi and Darne [2014], as well as the original Montecarlo exercises of Ahn and Horenstein [2013]. To the best of our knowledge, however, the figures that we produce have not previously appeared in the literature. This includes Appendix Figure G.3d, showing the distribution of the eigenratio estimator under the null hypothesis that there is no group structure.

The finite sample behavior of the eigenratio estimator is shared by estimators using NNR. It will thus be important to check whether the values of NNR obtained might have arisen by random chance from data with no actual group structure. Consider the value of $\max_{k<J_{\max}} \text{NNR}_k$. We wish to compare this test statistic to its distribution under the assumption that there is no actual group structure, obtaining appropriate p-values.[47]

To do so, consider a "reference distribution" where there are no organized groups. Randomly generate attack data based on this distribution, calculate an equivalent to $\hat{\Gamma}_L$ based on this randomly generated data, calculate a value for NNR_k based on this matrix, and then repeat this process 100 times. We consider three different reference distributions. Specifics are provided in Appendix H.

[44] Ding, He, and Simon [2005] describe from a more general perspective the similarities between NNMF and spectral clustering.

[45] Classic references here include the Wigner [1955] semi-circular distribution and the Marchenko-Pastur [1967] distribution.

[46] As Mirza and Storjohann [2014] point out, the effect is visible in the original Ahn and Horenstein [2013] simulations.

[47] Other hypothesis tests are difficult to perform: the distribution of eigenvalues resulting from random variation in finite samples is not obvious. We thus do not report confidence intervals for $\hat{J}$. For similar reasons, we also do not report confidence intervals for $\{\hat{\alpha}_{ij}\}$ below.

3.4 Robustness: Potentially changing district environments

Both the non-overlapping and overlapping approaches just described assume that the covariance in attacks by group members across districts remains the same even across long periods of time. In the observed data, however, it may be the case that in earlier years certain districts are the focus of many attacks, while in later years activity shifts to other districts. These sorts of long-term changes can be accounted for by considering only the covariance in attacks across districts within shorter time windows.

Let Γ_m be calculated the same as Γ from Equation (3), but using only daily attack data from month m. As the number of days of data used to calculate estimates of Γ_m does not increase asymptotically for any given month m, estimation based on a single Γ_m would be inconsistent. Aggregating across months, however, results in a consistent estimator that is robust to changes in attack probabilities between districts at monthly frequency.

Specifically, assume that the probability of an attack in district i in month m, either from unorganized militants or an organized group, now changes with a parameter ζ_{im}. That is, the probability of an attack from a unorganized militant is now $\zeta_{im}\eta$, and the probability of an attack from member of organized group j is now $\zeta_{im}\epsilon_{jt}$. Let $D(\cdot)$ indicate a diagonal matrix with the given entries on the diagonal. If ζ were known, the standardized matrix $\tilde{\Gamma}_m = D(\frac{1}{\zeta_m})\Gamma_m D(\frac{1}{\zeta_m})$ could be summed to create $\tilde{\Gamma} = D(\sum_m \zeta_m)\tilde{\Gamma}_m D(\sum_m \zeta_m)$. $\tilde{\Gamma}$ could then be used to estimate $\{\alpha_{ij}\}$. In reality, ζ is unobserved; however, dividing by the observed number of attacks creates a feasible estimator, with α identified up to scale.

This approach can be employed with both estimation based on clustering and that based on non-negative matrix factorization. Appendix I provides further details.

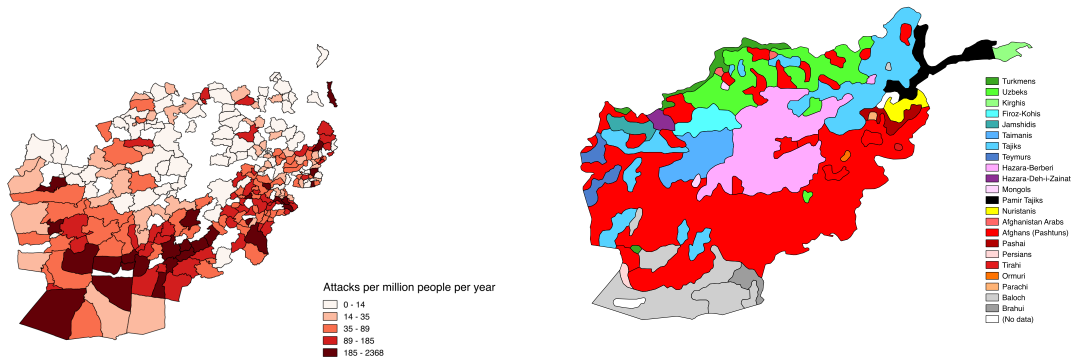
4 Data

Incident-level information is the main input to our empirical analysis. Both Afghanistan and Pakistan were covered by the Worldwide Incidents Tracking System, a discontinued U.S. government database [Wigle 2010].[48] Data is available for location, date, and type of violent incidents from the beginning of 2003 to the end of 2009.[49] The violent incidents cataloged in the WITS data are episodes of violence initiated by insurgents,

[48] The data remains accessible online courtesy of the Empirical Studies Of Conflict (ESOC) project at Princeton University.

[49] The following two examples illustrate the typical form of incident descriptions:

Figure 1: Afghanistan data



(a) Total attacks per capita (b) Ethnic Groups

or acts of random violence. The data does not include violence directly connected to military counterinsurgency operations, such as for instance a U.S. military attack on a Taliban safe house or the bombing of a fortified compound: see Appendix N for details.

The location reported for an attack in WITS is given as latitude and longitude coordinates. This would seem to suggest that attacks could be analyzed as some sort of spatial point process. Closer inspection, however, reveals that the latitude and longitude coordinates reported are not those of the actual location of the attack, but rather the coordinates of a prominent nearby geographic feature. Sometimes this is a city or village, but for the vast majority of incidents the location given is that of the centroid of the district in which the incident occurred. In Afghanistan, the "district" is the lowest-level political unit and the unit of geographic location in our model. We also note that a few districts have been split in recent years: this paper uses 2005 administrative boundaries, which specify 398 districts. The WITS data effectively provides panel data at the district-day level, with $N = 398$ and $T = 2082$. District-level geographic locations are also used for the Pakistan WITS data.

According to the data, there are some days where as many as 64 different districts in Afghanistan are affected by simultaneous insurgent attacks. However, there are also 123 districts with no reported incidents over the entire 2004-2009 time period. It is apparent

"*On 27 March 2005, in Laghman, Afghanistan, assailants fired rockets at the Governor House, killing four Afghan soldiers and causing minor damage. The Taliban claimed responsibility for the attack.*" And "*On 19 February 2006, in Nangarhar, Afghanistan, a suicide bomber detonated an improvised explosive device (IED) prematurely near a road used by government and military personnel, causing no injuries or damage. No group claimed responsibility.*"

Figure 2: Pakistan data

(a) Total attacks per capita

(b) Ethnic Groups

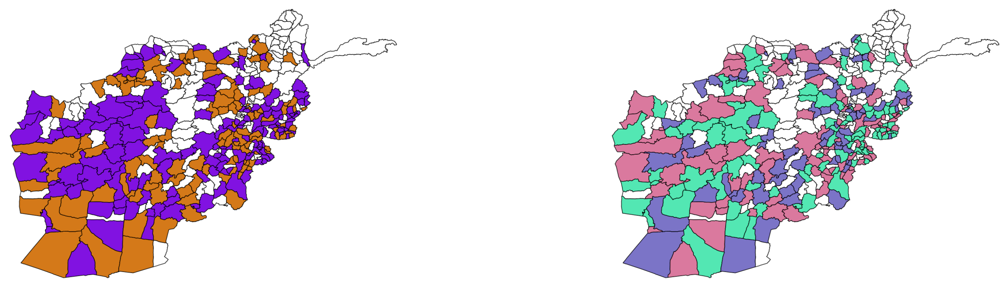
to even the most casual observer that attacks are concentrated in certain areas of the country.

For Pakistan, the BFRS dataset [Mesquita et al. 2015] is also available. This is similar to WITS, in that it provides daily data on violent incidents, including geographic information. BFRS data is available until 2011, and over the WITS time frame of 2004-2009, BFRS contains approximately twice as many incidents as WITS. Because of the greater number of attacks recorded, we prefer the BFRS data to the WITS data. We make use only of BFRS data from mid-2008 until the end of the sample in 2011, because qualitative evidence suggests that the structure of insurgent groups during this 3.5 year period was relatively stable. In early 2008 national elections took place in Pakistan, producing a new executive after the resignation of General Pervez Musharraf, and indicate an important structural breakpoint politically. The Taliban also strengthened considerably around 2008 [Iqbal and De Silva 2013].During the period in analysis, there is also qualitative as well as case study evidence of an active organizations in Sindh (e.g. Sindhudesh Liberation Army) and in Balochistan (e.g. Baloch Republican Army).

For geographic data on ethnicities, we use the Soviet Atlas Narodov Mira data.[50] In Figure 1 we show the pattern of attacks by district in Afghanistan and the distribution of ethnicities. The concentration of attacks in the ethnic Pashtun areas is evident. In Figure 2 we report the same information for Pakistan. This data forms the basis for most of the analysis that will be performed in the following section.

[50]The version used is the “Geo-referencing of ethnic groups” data set of Weidmann et al. [2010].

Figure 3: Afghanistan groups via spherical k-means



(a) 2 Groups (b) 3 Groups

5 Results

We first analyze attack data from Afghanistan, and then consider the case of Pakistan. In both cases, we begin with the spherical k-means clustering and splitting approach outlined in Section 3.2, and then proceed to the non-negative matrix factorization approach of Section 3.3. For all these analyses, we will use attack covariance matrices calculated using only within-month variation, as described in Section 3.4, unless noted otherwise. This is because it is important to avoid contamination by long-term trends, as well as seasonal variation in conflict.

5.1 Afghanistan

To illustrate attack data from Afghanistan, Figure 3 shows clustering based on spherical k-means, as outlined in Section 3.2. Qualitatively, the clusters shown in the figure appear indistinguishable from random noise.[51] We test this hypothesis using the statistic proposed in (6). Column I of Table 1 shows the results of this analysis. We will use this column to provide some further detailed exposition of this approach.

We begin by considering the null hypothesis that all districts are associated with the same insurgent group, and ask whether we should instead split the districts into two groups. In Figure 5, this first potential split is indicated by (1). We calculate how geographically distinct these split groups would be, and also calculate how geographically

[51] These figures are calculated based on a "within month" covariance matrix, as described in Section 3.4. Results do not change with other approaches.

Figure 4: Eigenratios, Afghanistan

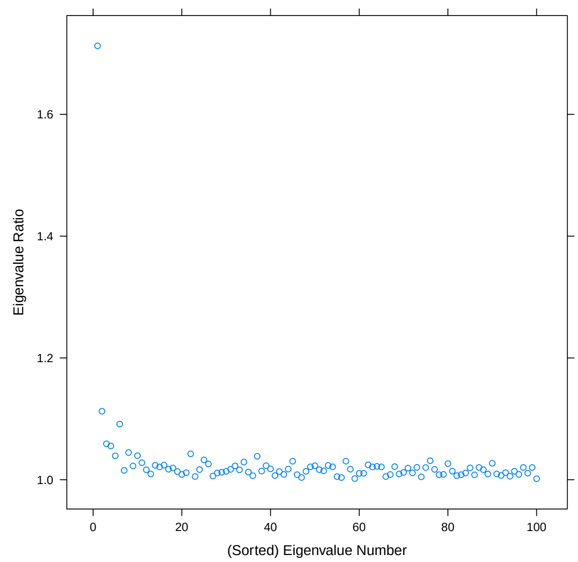


Figure 5: Hierarchical Splits

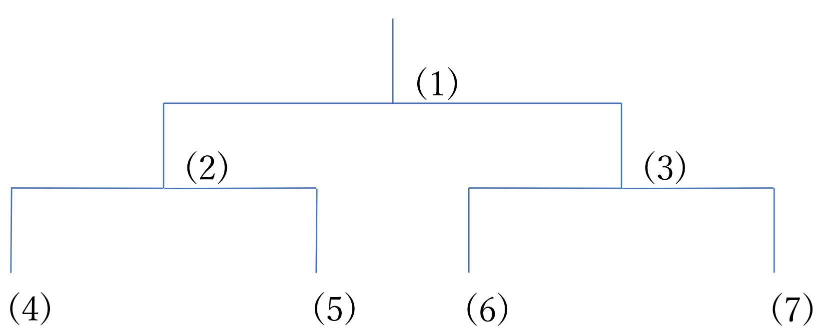


Table 1: Estimation of $\hat{J}$ based on hierarchical splits

		Afghanistan I	Pakistan II
Split at (1)?	Randomly shuffled data (mean)	284.99	137.67
	Std. dev.	11.09	8.52
	Actual data	289.00	159.00
	p-value	0.38	0.01
Split at (2)?	Randomly shuffled data (mean)		44.42
	Std. dev.		4.30
	Actual data		64.00
	p-value		0.00
Split at (3)?	Randomly shuffled data (mean)		34.30
	Std. dev.		4.61
	Actual data		47.00
	p-value		0.01
Split at (4)?	Randomly shuffled data (mean)		18.22
	Std. dev.		3.01
	Actual data		16.00
	p-value		0.71
Split at (5)?	Randomly shuffled data (mean)		14.01
	Std. dev.		2.90
	Actual data		19.00
	p-value		0.08
Split at (6)?	Randomly shuffled data (mean)		12.63
	Std. dev.		2.32
	Actual data		15.00
	p-value		0.23
Split at (7)?	Randomly shuffled data (mean)		9.21
	Std. dev.		1.87
	Actual data		10.00
	p-value		0.44

Each column computes a test statistic Q as described in Section 3.2, based on a within-month covariance matrix as described in Section 3.4. Figure 5 shows the order of the potential splits. Columns differ in the underlying attack data used:

Column I uses the full Afghanistan WITS dataset.

Column II uses the Pakistan BFRS dataset for May 2008 - October 2011.

distinct we would expect them to be under the null hypothesis that there is actually only one group. These calculations are shown in Table 1 on the rows following "Split at (1)?" The results displayed in column I of Table 1 show that for Afghanistan the actual data leads to groups that are no more geographically distinct than would be expected by random chance. Thus, for the Afghanistan data we stop at one group. The eigenratio approach of Ahn and Horenstein [2013] produces an identical result, as illustrated in Figure 4, where one large eigenratio at 1 is strikingly evident. Thus, both approaches suggest that the Taliban do not appear divided into multiple organized groups.

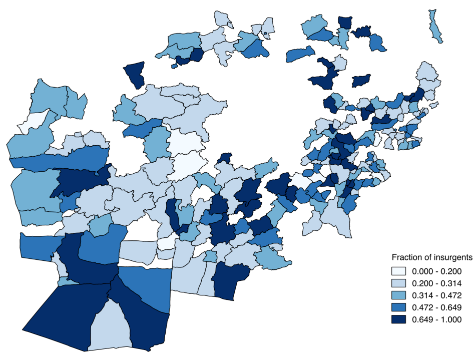
The group membership shown in Figure 3 involves a discrete partition of districts into insurgent groups. Some districts, however, might have many insurgents, while other districts might have few. Furthermore, there might be districts where more than one insurgent group is active. The model presented in Section 3.3 allows for these possibilities. An additional advantage of this model is that it provides a test against that null hypothesis that $J = 0$, and all attacks are the result of disorganized local actors. In contrast, the model used in Section 3.2 assumes that there is exactly one organized group present in each district, and thus this model cannot be used to test the hypothesis that there are actually no groups.

We thus move on to the alternate model of Section 3.3. A non-negative factorization of the attack covariance matrix, following Equation (8), turns out to give very similar results to those just discussed above. In contrast to the previous approach, multiple insurgent groups may now be active in any single district, and thus it is no longer possible to easily display the estimated insurgent group structure on a single map. Instead, we produce one map for each group. Figure 6 provides a visualization of this factorization in the case where $J = 2$. Again, there is no discernible pattern to the estimated insurgent groups.

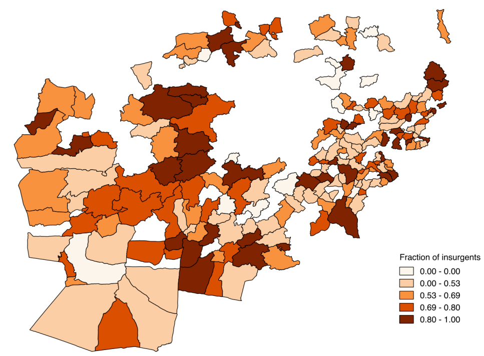
We are particularly interested now in whether we can reject the null hypothesis that $J = 0$, i.e. that there are no organized insurgent groups present at all. We begin by calculating the ratio described in Equation (10), and choose $\hat{J}$ so as to maximize this ratio. We then consider the distribution that this ratio would have if there were actually no organized groups. To do this, we use the permutation approach described in Section 3.3 and Appendix H.

Table 2 shows the results of this analysis for the Afghanistan data. There are four estimates of J provided. Beginning with the first two columns of the first row, $\hat{J} = 4$ in the case where districts are weighted proportionally to the number of attacks in the district. Continuing to the next three columns of the first row, $\hat{J} = 1$ if, in addition to this

Figure 6: Afghanistan, 2 groups via NNMF



(a) Group 1



(b) Group 2

Table 2: Estimated number of groups via NNR, Afghanistan

	Not by Month		By Month		
	I	II	III	IV	V
Afghanistan (WITS, Jan 2004 - Sept 2009, weighted)	4	4	1	1	1
(p value, vs. no group structure)	0.57	0.41	0.01	0.01	0.03
Afghanistan (WITS, Jan 2004 - Sept 2009, unweighted)	1	1	1	1	1
	0.02	0.02	0.02	0.03	0.06

Each row presents two estimates of $\hat{J}$, the number of groups present. Columns I and II show the first estimate, described in Section 3.3.

Columns III through V show the second estimate, based on the within-month covariance matrix as described in Section 3.4.

In each column, the p values presented are a test of the null hypothesis that there is no group structure. Other tests (e.g. $J = 1$ vs. $J = 2$) appear difficult to construct.

Columns I and III compute p values by comparing to a reference distribution where the time of the attacks within each district has been permuted. See Appendix H for a description of this and other reference distributions.

Column IV is the same as Column III, but the time of attacks is permuted only within each month.

Columns II and V consider only permutations that keep constant the total number of attacks in each district and on each day.

weighting, the covariance matrix is calculated considering only within-month variation in attacks using the approach described in Section 3.4. The second row provides estimates without weighting districts, and results in $\hat{J} = 1$ regardless of whether the approach in Section 3.4 is used or not.

Below each $\hat{J}$ estimate a p value is shown, corresponding to a test of the null hypothesis that in fact there is no group structure, with $J = 0$. We see that in general the null is rejected at the 95% level. The exception is the case where our estimate was $\hat{J} = 4$. With this specification, the model appears to have low power. This analysis supports the results obtained in Table 1, in that there appears to be one organized group of insurgents, rather than more than one. Furthermore, the observed NNR_1 values, calculated according to Equation 10, appear to be more extreme than would be the case if there were no organized groups at all. Table 2 shows that the observed data appears to be inconsistent with $J = 0$, a conclusion that we were not above to draw from the Q statistic results shown in Table 1.

Overall, in the Afghan case all our methodologies conclusively point to a single, organized Taliban insurgent group.

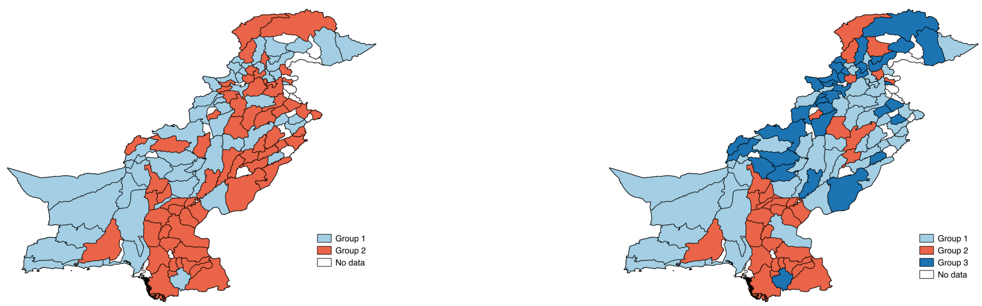
To conclude the discussion of Afghanistan, we refer to the working paper version of this article for an analysis of the ethnic and economic correlates of the spatial diffusion of the Taliban over the Afghan territory beyond its ethnic Pashtun areas. In the same setting we also performed a dynamic analysis of the spatial diffusion over time of the Taliban. We discussed there the specific pattern of "oil spot" evolution between 2004-2007 and 2008-2009, predicting expansion among contiguous districts. Due to space constraints, we do not explore such applications further here.

5.2 Pakistan

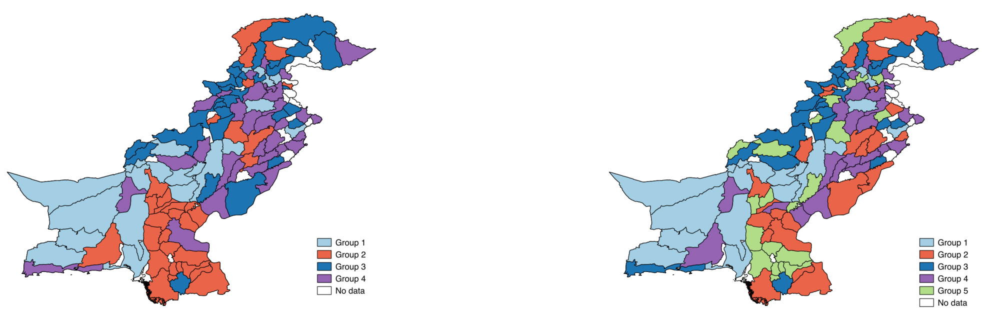
Clusterings based on the Pakistan attack data are shown in Figure 7. Unlike the results for Afghanistan shown in Figure 3, our clusterings for Pakistan, computed on the basis of the attack covariance matrix, result in groups that appear to be clustered geographically. For a more formal analysis, we consider the Q statistic results in column II of Table 1.

Unlike the case with Afghanistan in Column I, we do not stop immediately with an estimate of $\hat{J} = 1$. Instead, for Pakistan, the first set of rows in Table 1 shows that if the set of all districts is split into two groups, these groups are substantially more geographically distinct than would be expected if there were no actual group structure. We

Figure 7: Pakistan groups via spherical k-means



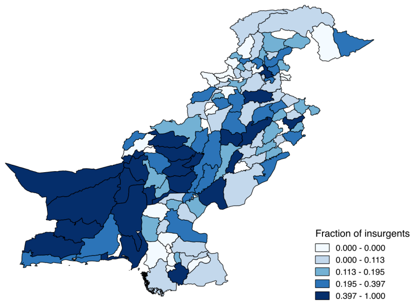
(a) 2 Groups (b) 3 Groups



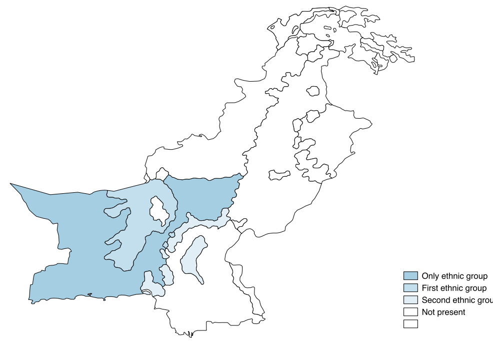
(c) 4 Groups (d) 5 Groups

thus split the set of districts into these two groups, and continue recursively, asking for each of these two groups whether the group should be further divided. These questions are indicated by (2) and (3) in Figure 5, and the next two sets of rows in Table 1. In each of these cases the potential splits appear to be more geographically distinct than would be expected by random chance, and so in each case the group is split, leading to a total of four groups. Continuing recursively, we consider whether any of the four groups we now have should be further split. These questions correspond to the final four sets of rows in Table 1. We see that none of these splits generate groups that are more geographically distinct than would be expected by random chance, and thus we do not split any of these groups. At this point there is no further recursion, with an estimate of $\hat{J} = 4$. The unified insurgent structure ($\hat{J} = 1$) that we recover for the Afghan case appears not to be present in Pakistan. This accords with the qualitative analysis, e.g.

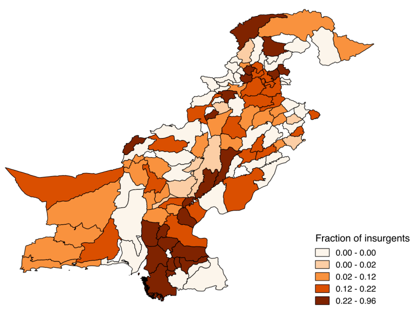
Figure 8: Pakistan, 4 groups via NNMF (left column) and ethnicities (right column)



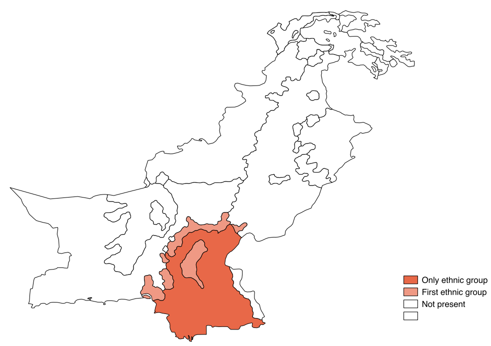
(a) Group 1



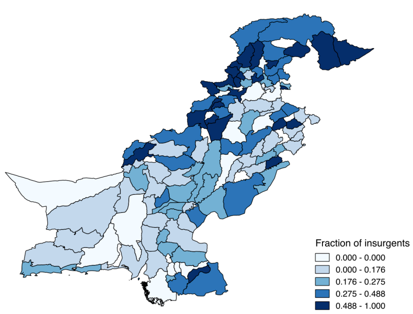
(b) Balochis



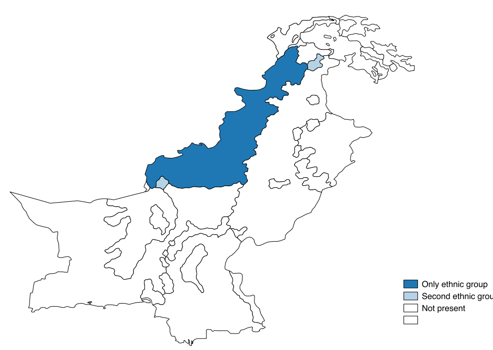
(c) Group 2



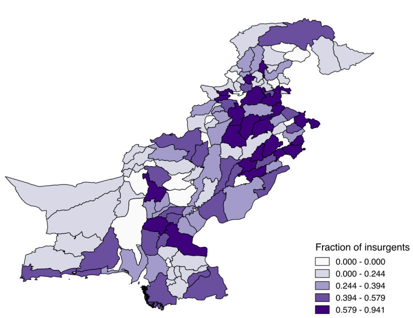
(d) Sindhis



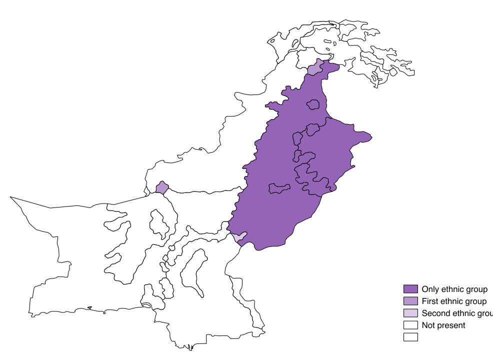
(e) Group 3



(f) Afghans



(g) Group 4



(h) Panjabis, Jhats, Awans

Table 3: Ethnic composition of groups shown in Figure 7c

	Group 1	Group 2	Group 3	Group 4
Baloch	0.62	0.25	0.00	0.12
	(0.09)	(0.09)	(0.10)	(0.10)
Sindhs	0.04	0.87	0.04	0.04
	(0.07)	(0.07)	(0.08)	(0.09)
Afghans	0.12	0.08	0.58	0.23
	(0.07)	(0.07)	(0.08)	(0.08)
Panjabis, Jhats, Awans	0.16	0.12	0.23	0.49
	(0.05)	(0.05)	(0.06)	(0.06)
Other	0.00	0.29	0.57	0.14
	(0.13)	(0.13)	(0.15)	(0.16)
N	115	115	115	115

Each column corresponds to a single regression without intercept.

The dependent variable is a dummy variable indicating whether a given district was clustered into the specified group number in the clustering shown in Figure 7c. Districts shown as white in the figure ("no data") are dropped: the remaining 115 districts are used in the regression.

The independent variables are a set of dummy variables, indicating whether the specified ethnicity was listed as the first ethnicity at the centroid of a given district.

Each row should sum to 1 because each coefficient in the table is a conditional mean giving the fraction of districts of the specified ethnicity that were clustered into the specified group, and the clustering in Figure 7c assigns each district to one group. Rows may not sum exactly to 1 because of rounding.

Standard errors in parentheses.

Dorronsoro [2009].

For completeness we now analyze the Pakistan data using the non-negative matrix factorization approach of Section 3.3. As with the Afghan data, results are generally in line with that obtained using the Q statistic approach. The left-hand column of Figure 8 shows a non-negative matrix factorization of the attack covariance matrix for Pakistan, using four factors. The result here is very close to that shown in Figure 7c. Furthermore, both of these figures show what appears to be a close relationship between the estimated group structure and the arrangement of ethnic groups in Pakistan. The relevant breakdown of these ethnic groups is shown in the right-hand column of Figure 8.

A qualitative comparison of the left and right columns of Figure 8 shows that there is one insurgent group present in Balochistan, another in the area populated by Sindhs, a third in the area populated by "Afghans" (i.e. Pashtuns), and a fourth in the Punjabi areas of Pakistan. The northernmost areas of Pakistan, with numerous smaller eth-

Table 4: Ethnic composition of groups shown in Figure 8

	Group 1	Group 2	Group 3	Group 4
Baloch	0.58	0.07	0.12	0.23
	(0.05)	(0.04)	(0.06)	(0.06)
Sindhs	0.14	0.35	0.19	0.32
	(0.04)	(0.03)	(0.05)	(0.05)
Afghans	0.15	0.07	0.48	0.30
	(0.04)	(0.03)	(0.04)	(0.05)
Panjabis, Jhats, Awans	0.22	0.11	0.23	0.44
	(0.03)	(0.03)	(0.03)	(0.04)
Other	0.05	0.11	0.61	0.24
	(0.07)	(0.06)	(0.08)	(0.09)
N	115	115	115	115

Each column corresponds to a single regression without intercept, concerning group $j \in \{1, 2, 3, 4\}$.

The dependent variable is $\hat{\alpha}_{ij} / \sum_{j' \in \{1,2,3,4\}} \hat{\alpha}_{ij'}$. This is the fraction of organized insurgents present in a district that are from group j.

This data is displayed in the left column of Figure 8, and it is available for the same 115 districts that were analyzed in Table 3.

The independent variables are a set of dummy variables, indicating whether the specified ethnicity was listed as the first ethnicity at the centroid of a given district.

Each row should sum to 1 (up to rounding) by the same argument as in Table 3: each coefficient in the table is a conditional mean for the ethnicity in question, every district is coded as one ethnicity, and the group shares must sum to one.

Standard errors in parentheses.

nicities, appear to be associated most closely with the "Afghans".[52] The major ethnic divisions of Pakistan can thus be successfully reproduced employing only the covariance matrix of data on insurgent attacks.

Tables 3 and 4 show the relationship between estimated insurgent groups and ethnic groups in a quantitative fashion. These tables are constructed to describe the distribution of ethnicities across the estimated groups. Each row corresponds to an ethnicity, and each row sums to 100% (+/- rounding error). The rows have been ordered so that the diagonal entries correspond to the qualitative relationship between groups and ethnicities just discussed. This is the same ordering of rows used in Figure 8.

We now consider an eigenratio type analysis of the Pakistan attack data. In the case of Afghanistan, analysis based on the Q statistic approach in Table 1 resulted in an estimate of $\hat{J} = 1$, but it was then necessary to use the results shown in Table 2 to show that the null hypothesis of $J = 0$ could be rejected. In contrast, with the Pakistan data,

[52]Adding a fifth group does not result in these "other" ethnicities being clustered into their own separate group: see Figure 7d. This may be because these areas consist of many small ethnic groups, and there is not a sufficient number of attacks for these smaller groups to be estimated correctly.

Table 5: Estimated number of groups via NNR, Pakistan

	Not by Month		By Month		
	I	II	III	IV	V
Pakistan (BFRS, May 2008 - Oct 2011, weighted)	1	1	1	1	1
	0.63	0.63	0.16	0.28	0.55
Pakistan (BFRS, May 2008 - Oct 2011, unweighted)	2	2	16	16	16
	0.73	0.68	0.00	0.01	0.03

Notes: same as Table 2, except with Pakistan data.

Table 1 gives $\hat{J} = 4$. This result would be extreme under a null hypothesis of $J = 0$, and thus it is not as important to seek alternate confirmation that there is indeed a group structure in the data.[53] This turns out to be fortunate, as the "eigenratio" type analysis shown in Table 5 is inconclusive in the case of the Pakistan data.[54]

6 Prevalence of Coordination

In Section 5 we employed the presence of simultaneous attacks to study the organization of insurgency in Afghanistan and Pakistan. A remaining question, however, is how prevalent these sort of attacks are. Were there only a tiny number of coordinated insurgent attacks, studying simultaneous incidents might appear simply a curiosity and studying them a distraction from more important issues. A more substantial number of these attacks, on the other hand, would make their relevance more obvious.

Suppose that our data actually contained no planned simultaneous attacks at all. According to our model in Section 3, the fact that more than one attack occurred on a given day would then be due purely to random chance. If we assume that the number of potential disorganized insurgents that could launch attacks is large, and the probability

[53]In order for Table 1 to lead to an estimate of $\hat{J} = 4$ it must be that for each of two groups, three groups, and four groups, the improvement in geographic clustering is at least one standard deviation better than would be expected if there were no group structure. A result of $\hat{J} = 4$ is thus already very extreme under the null that $J = 0$.

[54]Very few entries are statistically significant at the 95% level, and those that are appear to be computational artifacts of some sort, giving very high estimates for $\hat{J}$. In Appendix we confirm this by rerunning the analysis using the original eigenratio from Equation 7. We join a number of other researchers in discarding a low $\hat{J}$ estimate based on eigenratios in favour of other evidence: both Henzel and Rengel [2014] and Alquist and Coibion [2014] discard $\hat{J} = 1$ in favour of two factors, and Bleaney et al. [2012] discards $\hat{J} = 1$ or $\hat{J} = 3$ in favour of four or more factors, while Rezitis [2015] discards $\hat{J} = 2$ in favour of five factors.

η of any one of them launching an attack is low, then the distribution of attacks that we observe in any district should be approximately Poisson $(\eta \ell_i)$. The total number of attacks observed in any day, across all districts should thus also be Poisson $(\eta \sum_i \ell_i)$. We can thus check for the existence of simultaneous attacks by looking for overdispersion in the observed pattern of attacks relative to a Poisson, using the technique from Cameron and Trivedi [1990]. Within our model, overdispersion in presence of a single insurgent group j would occur in the mixture form:

$$x_{it} \sim p_{ij} Poisson\left(\epsilon^h \alpha_{i,j} + \eta \ell_i\right) + (1 - p_{ij}) Poisson\left(\epsilon^l \alpha_{i,j} + \eta \ell_i\right).$$

where the two cases correspond to the two potential values of ϵ_{jt}. According to this test, the distribution of attacks is overdispersed relative to a Poisson distribution (with $p < 0.01$) in both Afghanistan and Pakistan.[55] We can reject with high statistical confidence the absence of coordinated attacks in both Afghanistan and Pakistan.

We are further interested in assessing the quantitative importance of simultaneous organized attacks. The traditional definition of overdispersion refers to additional variance above and beyond what expected from a Poisson distribution, but this does not have an easy interpretation as a specific quantity of coordinated attacks. We will thus use a non-traditional definition of overdispersion, one that provides an estimate of how many simultaneous attacks there are in our dataset.

An overdispersed Poisson distribution will have more high realizations (days with a large number of attacks) compared to a Poisson distribution with the same mean. These additional high realizations correspond to the simultaneous attacks that are of interest.[56] We can then take the actually observed distribution of attacks across days and compare it to the theoretical Poisson distribution with the same mean. Let $g(x) = (\bar{f}(x) - f(x))x$, where f is the theoretical probability of observing x attacks in a day, given a Poisson with the mean equal to the actual mean number of attacks per day. Let $\bar{f}$ be fraction of days when x attacks occurred in the empirical distribution. Thus, $g(x)$ is the excess number of attacks in the empirical distribution, considering only days where there were exactly x attacks. Let $G(x) = max_x \sum_{i=x}^{\infty} g(x)$, where the object being maximized is the excess number of attacks in the situation where there were x or more attacks in a

[55] Including month fixed effects gives a p-value of approximately 0.002.

[56] That is, if there is a group that has an ϵ that is constant across all days, then it will lead to no overdispersion and thus, according to our definition, no simultaneous attacks of interest. There are attacks that occur on the same day by random chance: it is just that the fraction of days where there are two attacks will be that of a poisson distribution, which is our baseline, and thus there are no *additional* attacks due to coordination within the group.

day. We will define the number of excess attacks of interest as $G(x)$.

Using this definition, in our Afghanistan data 4% of all attacks are simultaneous attacks of interest, and in our Pakistan data, 9%. These figures point to the use of coordinated attacks in between one in twenty and one in ten episodes of violence for our samples. Our analysis in this paper has thus focused on a quantitatively relevant component of violence within these countries, comprising several hundred recorded incidents.

This evaluation of coordination relies on the Poisson approximation to our particular model of attacks. One might be concerned that the actual attack structure is, for some unknown reason, not an overdispersed Poisson distribution. To provide direct verification of the percentages reported above, we use attacks recorded in the Global Terrorism Database (GTD), "*currently the most comprehensive unclassified database on terrorist events in the world*".[57]

For our periods of interest, there are 5,464 attacks recorded in Afghanistan and 2,925 in Pakistan. For each attack, the GTD records whether there are other "related" attacks: these are generally attacks that occurred the same day but in a different location.[58] Importantly, this variable is not coded mechanically. Attacks are only coded by country experts as "related" if there actually appears to be intelligence evidence of intentional relationship between the attacks. For our period, 4% of the attacks in Afghanistan and 10% of the attacks in Pakistan have "related" attacks in the GTD, 511 attacks in total. This is extremely close to our 4% and 9% estimates above based on overdispersion of WITS and BFRS, strongly validating our approach.

One might further wonder whether this tight relationship between overdispersion and related attacks holds for countries other than Afghanistan and Pakistan. We consider all countries listed in the GTD, and calculate the percentage excess of attacks in each country using the GTD data. Figure K.6 shows a clear relationship between our measure of overdispersion and the fraction of attacks that are coded as related. Table K.8 considers regressions using our measure of overdispersion coded at the country-year level, which allows us to include country fixed effects. We see that in years in which there are more overdispersed attacks in a given country, that country is more likely to have a greater fraction of their attacks coded as related.

The GTD also proves extremely useful in performing a number of validity checks

[57]Available at https://www.start.umd.edu/gtd/

[58]The GTD codebook technically allows for "related" attacks that do not occur on the same day, but we confirm in Appendix K that here are no related attacks of this type in our sample.

of our main approach. These include: (i) identification of the insurgent groups based on GTD additional expert assessments; (ii) assessment of the insurgent structures we uncover for time periods beyond the ones considered in the main analysis; (iii) ruling out that coordinated attacks take place over time periods longer than the single day; (iv) evidence that a single group and not multiple coordinated groups is typically behind a set of simultaneous attacks; (v) evidence of the extent of credit claiming for coordinated attacks; (vi) support for the assumption of a trade-off between military value opportunity and signalling value of attacks; (vii) panel data evidence of the relationship between group strength and extent of coordination. This additional analysis is available in Appendix K, where all details on specifications and interpretation are provided.

7 Conclusions

This paper focuses on the empirical analysis of insurgency, with applications to Afghanistan and Pakistan. Often the only type of information available about level and geographical diffusion of insurgent activity comes from incident-level attack data. However limited that might be, thanks to it recent advances in the analysis of the economics of conflict and post-war reconstruction have been possible.[59] Adding to this literature, we show how incident-level data contains information useful to estimate latent structures and geographic span of influence of insurgent groups. The paper develops methodologies to detect unobserved rebel coalition structures based on co-occurrences of violent incidents across districts over time under different assumptions.

Progress in understanding insurgency seems key in furthering knowledge of the determinants and consequences of political violence in developing countries. Although much of the analysis in the paper is necessarily context-dependent, it is informative nonetheless for regional stabilization and local development goals. From a methodological perspective, our contributions have a more general appeal and show promise in other data environments.

REFERENCES

[1] Ahn, S.; Horenstein, A. (2013) "Eigenvalue Ratio Test for the Number of Factors. "Econometrica. 81(3): 1203-1227.

[59] Berman, Shapiro, and Felter [2011], Trebbi, Weese, Wright and Shaver [2017] are recent examples.

[2] Alquist, R.; Coibion, O. (2014) "Commodity-Price Comovement and Global Economic Activity ". NBER Working Paper 20003. March 2014.

[3] Ashford, J.R. and R.G. Hunt (1973) "The Distribution of Doctor-Patient Contacts in the National Health Service" Journal of the Royal Statistical Society Series A 137 (3), 347-383.

[4] David Barno (2006) Challenges in Fighting a Global Insurgency Parameters. Summer 2006. 15-29.

[5] Bleaney, M.; Mizen, P.; Veleanu, V.2012). "Bond Spreads as Predictors of Economic Activity in Eight European Economies ." University of Nottingham, Centre for Finance, Credit and Macroeconomics (CFCM) Discussion Paper. December 2011.

[6] Benmelech, Efraim, Claude Berrebi, and Esteban F. Klor. (2012). "Economic Conditions and the Quality of Suicide Terrorism." The Journal of Politics 74 (1): 113–128.

[7] Berman, Eli (2009). Radical, Religious and Violent: The New Economics of Terrorism. MIT Press.

[8] Berman, Eli, Joseph H. Felter, Jacob N. Shapiro, (2011) Can Hearts and Minds Be Bought? The Economics of Counterinsurgency in Iraq. Journal of Political Economy Vol. 119, No. 4: 766-819

[9] Berman, Eli, Aila Matanock. (2015). "The Empiricists' Insurgency." Annual Review of Political Science, Vol 18: 443-464.

[10] Besley, Tim, Torsten Persson. (2017). "The Joint Dynamics of Organizational Culture, Design, and Performance" mimeo LSE.

[11] Birgin, E.; Martinez, J.M.; Raydan, M. (2000). "Nonmonotone Spectral Projected Gradient Methods on Convex Sets." SIAM J. Optim.. 10(4): 1196–1211.

[12] Blattman, Christopher and Edward Miguel (2010) "Civil War" Journal of Economic Literature 2010, 48:1, 3–57

[13] Brahimi, A. (2010). "The Taliban's Evolving Ideology." Working Paper. LSE Global Governance. WP 02/2010.

[14] Bruzzese, D., Vistocco, D. (2015). "DESPOTA: DEndrogram Slicing through a PemutatiOn Test Approach". Journal of Classification. 32(2):285-304.

[15] Bueno de Mesquita, Ethan. (2013). "Rebel Tactics." Journal of Political Economy 121 (2): 323–357

[16] Bueno de Mesquita, Ethan, and Eric S. Dickson. (2007). "The Propaganda of the Deed: Terrorism, Counterterrorism, and Mobilization." American Journal of Political Science 51 (2): 364–381.

[17] Cameron, A.C. and Trivedi, P.K. (1990). Regression-based Tests for Overdispersion in the Poisson Model. Journal of Econometrics, 46, 347-364.

[18] Condra, Luke N., Jacob N. Shapiro, (2012) Who Takes the Blame? The Strategic Effects of Collateral Damage. American Journal of Political Science Vol. 56, No. 1: 167-187.

[19] Deloughery Kathleen (2013) Simultaneous Attacks by Terrorist Organisations. Perspectives on Terrorism, 7(6): 79-90.

[20] Ding, C., He, X., and Simon, H. (2005) On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering Proceedings of the Fifth SIAM International Conference on Data Mining, 606-610.

[21] Dorronsoro, Gilles (2009) The Taliban's Winning Strategy in Afghanistan. Carnegie Endowment for International Peace Paper.

[22] Fearon, James (2008) "Economic development, insurgency, and civil war" in Institutions and Economic Performance, ed. Elhanan Helpman, Harvard University Press

[23] Ferson, W.; Kim, M. (2012) "The factor structure of mutual fund flows. " International Journal of Portfolio Analysis and Management, 1(2), 112-143.

[24] Fotini, Christia, Semple, Michael (2009) "Flipping the Taliban- How to Win in Afghanistan" Foreign Affairs, 88, 34-45

[25] Ghobarah, Hazem Adam, Paul Huth and Bruce Russett. (2003) Civil Wars Kill and Maim People Long After the Shooting Stops." American Political Science Review 97(2):189-202.

[26] Giustozzi, Antonio (2009). "The Pygmy who turned into a Giant: The Afghan Taliban in 2009", LSE mimeo.

[27] Guo-Fitoussi, L.; Darne, O.. (2014) "A Comparison of the Finite Sample Properties of Selection Rules of Factor Numbers in Large Datasets". HAL Working Paper hal-00962247. March 2014.

[28] Gutierrez-Sanin, Francisco. (2008) Telling the Difference: Guerrillas and Paramilitaries in the Colombian War. Politics and Society 36(1):3-34.

[29] Henzel, S.; Rengel, M. (2014) Dimensions of Macroeconomic Uncertainty: A Common Factor Analysis. . SSRN Scholarly Paper 2507743.

[30] Hornik, K.; Feinerer, I.; Kober, M.; Buchta, C. (2012). Spherical k-Means Clustering. Journal of Statistical Software. 50(10).

[31] Huang, K., Sidiropoulos, N. (2014) "Putting nonnegative matrix factorization to the test: a tutorial derivation of pertinent cramer-rao bounds and performance benchmarking." IEEE Signal Processing Magazine. 31(3):76–86.

[32] Huang, K., Sidiropoulos, N., and Swami, A. (2014) Non-Negative Matrix Factorization Revisited: Uniqueness and Algorithm for Symmetric Decomposition. IEEE Transactions on Signal Processing 62(1):211-224.

[33] Khuram Iqbal and Sara De Silva (2013) Terrorist lifecycles: a case study of Tehrik-e-Taliban Pakistan. Journal of Policing, Intelligence and Counter Terrorism. Vol. 8, No. 1, 72-86.

[34] Jenkins, Brian Michael (2014) The Dynamics of Syria's Civil War, RAND Perspectives no. 115.

[35] Karlis, D., Xekalaki, E. (2005) "Mixed Poisson Distributions". International Statistical Review. 73(1):35-58.

[36] Kilcullen, David (2009) The accidental guerrilla: Fighting small wars in the midst of a big one. Oxford University Press

[37] König, Michael D., Dominic Rohner, Mathias Thoenig and Fabrizio Zilibotti, 2017 "Networks in Conflict: Theory and Evidence from the Great War of Africa." Econometrica.

[38] Krishna, K.; Narasimha, M. (1999). "Clustering High Dimensional Data: A Survey on Subspace Clustering, Pattern-based Clustering, and Correlation Clustering". Trans. Sys. Man Cyber. Part B. 29 (3): 433–439.

[39] Ludhianvi, M.R. (2015). Obedience to the Amir: An early text on the Afghan Taliban Movement. Trans. Y. Mitha and M. Semple. Berlin: First Draft Publishing.

[40] Luxburg, Ulrike von (2007) "A tutorial on spectral clustering" Statistics and Computing Volume 17, Issue 4, pp 395-416

[41] Marchenko, V.A.; Pastur, L.A. (1967). "Distribution of Eigenvalues for Some Sets of Random Matrices". Matematicheskii Sbornik. 72 (114): 507–536

[42] Mirza, H.; Storjohann, L. (2014). "Making Weak Instrument Sets Stronger: Factor-Based Estimation of Inflation Dynamics and a Monetary Policy Rule." Journal of Money, Credit and Banking. 46(4): 643-664.

[43] Munir, M. (2011). "The Layha for the Mujahideen: an analysis of the code of conduct for the Taliban fighters under Islamic law ". International Review of the Red Cross. Vol. 93 No. 881.

[44] O'Neill, Bard (1990) Insurgency and Terrorism, Inside Modern Revolutionary Warfare, Dulles, VA.: Brassey's Inc.

[45] Pak Institute for Peace Studies (2016) Pakistan Security Report 2015.

[46] Rezitis, A.N. (2015). "Empirical Analysis of Agricultural Commodity Prices, Crude Oil Prices and US Dollar Exchange Rates Using Panel Data Econometric Methods." SSRN Scholarly Paper 2631534. July 2015.

[47] Schelling, Thomas C. (1960) The Strategy of Conflict. Cambridge: Harvard University Press.

[48] Ben Smith (2005) Afghanistan: Where Are We? Conflict Studies Research Center, Central Asia Series. Report 05/30. June 2005.

[49] Thruelsen, Peter Dahl (2010) "The Taliban in southern Afghanistan: a localised insurgency with a local objective" Small Wars & Insurgencies, Volume 21, Issue 2, pp.259-276

[50] Trebbi, Francesco, Eric Weese, Austin Wright, Andrew Shaver. 2017. "Insurgent Learning" NBER WP 23475

[51] Tullock, Gordon (1974) The Social Dilemma, Blacksburg: Center for the Study of Public Choice, VPISU Press.

[52] United Nations (2013) Third report of the Analytical Support and Sanctions Monitoring Team, submitted pursuant to resolution 2082 (2012) concerning the Taliban and other associated individuals and entities constituting a threat to the peace, stability and security of Afghanistan. S/2013/656

[53] United Nations (2016) "Humanitarian Response Plan - Syrian Arab Republic" United Nations Office for the Coordination of Humanitarian Affairs.

[54] Weidmann, N.; Rod, J.K.; Cederman, L.E. (2010) "Representing ethnic groups in space: A new datset." Journal of Peace Research, 47(4), 491–499.

[55] Wigner, E.P. (1955) "Characteristic Vectors of Bordered Matrices With Infinite Dimensions". The Annals of Mathematics, 62(3), 548–564.

Online Appendices – Not For Publication

A Simultaneous Attacks: Qualitative sources

From a Western perspective, the 9/11 attacks in the United States are the most obvious example of the salience of such simultaneous violence, but the phenomenon is widespread. For example, in southern Thailand insurgent movements have adopted similar tactics: "On April 28, 2004 groups of militants gathered at mosques in Yala, Pattani, and Songkhla provinces before conducting simultaneous attacks on security checkpoints, police stations and army bases" [Fernandes, 2008]. The Indian Mujahideen, responsible for the 2008 Mumbai attacks, typically carry out simultaneous attacks [Subrahmanian et al., 2013]. Kurdish nationalists and the Tamil Tigers are known to have adopted simultaneous attacks as a strategy. In Africa, Boko Haram in northern Nigeria has carried out coordinated attacks on multiple targets such as churches, and Anderson [1974] describes coordinated attacks in Portuguese colonies. Simultaneous attacks and suicides have been a trademark of international jihadist organizations and of al-Qaeda in particular, making our approach well-suited to the Afghan insurgency case. Because the empirical covariance matrix of attacks is observed, these assumptions implying positive covariances driven by co-occurring incidents are readily verifiable and they are in fact supported by the data. See discussion at the end of Section 2.

B Proof of Propositions

Proof. Lemma 1. Denote p_{ij} the frequency with which group $j = j_1, j_2$ provides high rewards in i. This frequency is endogenous to our problem, but its derivation turns out not to be necessary for the noncombatants, who will be able to estimate p_{ij} from the data. Noncombatants are assumed to know $\epsilon^h_{j_1}$, $\epsilon^l_{j_1}$, $\epsilon^h_{j_2}$, $\epsilon^l_{j_2}$, η . Based on the data generating process of total attacks and the properties of Poisson distributions the number of attacks in i follows a either one of these Poisson mixtures:

$$x_{i,t} \sim p_{ij_1} Poisson\left(\epsilon^h_{j_1}\alpha_{i,j_1} + \eta\ell_i\right) + (1 - p_{ij_1}) Poisson\left(\epsilon^l_{j_1}\alpha_{i,j_1} + \eta\ell_i\right)$$

or

$$x_{i,t} \sim p_{ij_2} Poisson\left(\epsilon^h_{j_2}\alpha_{i,j_2} + \eta\ell_i\right) + (1 - p_{ij_2}) Poisson\left(\epsilon^l_{j_2}\alpha_{i,j_2} + \eta\ell_i\right).$$

Noncombatants will be able to detect the presence of a group based on the overdispersion. As the number of attacks follows a Poisson mixture, its conditional moments:

$$\begin{aligned}\lambda^1 &= \epsilon^h_{j_1}\alpha_{i,j_1} + \eta\ell_i \\ \lambda^2 &= \epsilon^l_{j_1}\alpha_{i,j_1} + \eta\ell_i\end{aligned}$$

or

$$\begin{aligned}\lambda^1 &= \epsilon^h_{j_2}\alpha_{i,j_2} + \eta\ell_i \\ \lambda^2 &= \epsilon^l_{j_2}\alpha_{i,j_2} + \eta\ell_i\end{aligned}$$

are identified following Feller [1943], Teicher [1961], and Teicher [1963] and so is the mixing weight p_{ij}. Identification of overdispersed Poisson distributions is well established and we will not reiterate it here. However notice that there exist

$$\alpha_{i,j_1} = \frac{\lambda^1 - \lambda^2}{\epsilon^h_{j_1} - \epsilon^l_{j_1}} \text{ and } \alpha_{i,j_2} = 0$$

and

$$\alpha_{i,j_2} = \frac{\lambda^1 - \lambda^2}{\epsilon^h_{j_2} - \epsilon^l_{j_2}} \text{ and } \alpha_{i,j_1} = 0$$

both consistent with the data. This proves the Lemma.■

Proof. Proposition 1. We begin with the first part of the proposition, under Assumptions

1-2. Indicate with $x_{i,t}$ the observed total attacks in district i at time t. If $J = 0$, $\text{Cov}(x_{i,t}, x_{i',t}) = 0$ for all $i, i' \in \mathcal{N}$. Under plan 1, this can only happen if all organized groups are only present in one district. This does not hold due to Assumption 1(ii). Hence, $J = 0$ is identified by the noncombatants.[60] This implies $N^j = 0$ is identified, implying that there are no organized combatants, $\alpha_j = 0$.

We now proceed with the more interesting case of $J > 0$. By Assumption 1, there exists $j_1 \in \mathcal{J}, i_1 \in \mathcal{N}^{j_1}$ such that noncombatants know $i_1 \in \mathcal{N}^{j_1}$ and in district i_1 there is only j_1 as an organized group. It follows that with a large T, for all $i \in \mathcal{N}$, $\text{Cov}(x_{i_1,t}, x_{i,t}) \neq 0$ if and only if $i \in \mathcal{N}^{j_1}$. This holds because the only organized group in i_1 is j_1, presenting no other source of correlation. This identifies $\mathcal{N}^{j_1} = \{i \in \mathcal{N} : \text{Cov}(x_{i_1,t}, x_{i,t}) > 0\}$ and $N^{j_1} = |\mathcal{N}^{j_1}|$.[61]

We now propose that α_{i_1,j_1} is identified. Denote p_j the frequency with which group j provides high rewards. This frequency is endogenous to our problem, but its derivation turns out not to be necessary for the noncombatants, who will be able to estimate p_j from the data. In fact, since j_1 is the only organized group in district i_1, the distribution of (realized) attacks is given by the two component Poisson mixture:

$$x_{i_1,t} \sim p_{j_1} Poisson\left(\epsilon^h \alpha_{i_1,j_1} + \eta \ell_{i_1}\right) + (1 - p_{j_1}) Poisson\left(\epsilon^l \alpha_{i_1,j_1} + \eta \ell_{i_1}\right). \tag{11}$$

As the number of attacks follows a Poisson mixture, its conditional moments $\lambda^h = \epsilon^h \alpha_{i_1,j_1} + \eta \ell_{i_1}$ and $\lambda^l = \epsilon^l \alpha_{i_1,j_1} + \eta \ell_{i_1}$ are identified following Feller [1943], Teicher [1961], and Teicher [1963] and so is the mixing weight p_{j_1}. Then, since η, ϵ^h and ϵ^l are known under Assumption 2, it follows that ℓ_{i_1} and α_{i_1,j_1} are also identified, as the two unknowns in a system of two linear equations $\lambda^h = \epsilon^h \alpha_{i_1,j_1} + \eta \ell_{i_1}$ and $\lambda^l = \epsilon^l \alpha_{i_1,j_1} + \eta \ell_{i_1}$. We can now use the identified variance of $\epsilon_{j_1,t}$:

$$\text{Var}(\epsilon_{j_1,t}) = (\epsilon^h - \epsilon^l)^2 p_{j_1}(1 - p_{j_1}),$$

due to the distribution of ϵ. This is identified by Assumption 2 combined with the identification of p_{j_1}. Summing over $i_2 \in N^{j_1}$ pairwise attack covariances divided by

[60] Another argument can be used for this part. If $\alpha_{ij} = 0, \forall i \in \mathcal{N}^j$, since $J = 0$, then district i's attacks follow a Poisson distribution with mean $\eta \ell_i$. By the identification of Poisson mixtures (Feller [1943], Teicher [1961], [1963]), noncombatants identify the number of components of the mixture (which equals to one) in each district. Hence, it is identified that there are zero organized groups in each district.

[61] We will not impose other assumptions on the structure of presence of groups. One other way where assumptions and identification could be obtained would be to impose structure on the adjacency matrix of group presence.

$\text{Var}(\epsilon_{j_1,t})$ yields:

$$\begin{aligned}\sum_{i_2 \in \mathcal{N}^{j_1}} \frac{\text{Cov}(x_{i_1,t}, x_{i_2,t})}{\text{Var}(\epsilon_{j_1,t})} &= \sum_{i_2 \in \mathcal{N}^{j_1}} \alpha_{i_1,j_1}\alpha_{i_2,j_1} \\ &= \alpha_{i_1,j_1} \sum_{i_2 \in \mathcal{N}^{j_1}} \alpha_{i_2,j_1} \\ &= \alpha_{i_1,j_1}\alpha_{j_1}\end{aligned}$$

and, hence, α_{j_1} is identified, as α_{i_1,j_1} is identified.

Note, however, that the noncombatants still have not identified whether there is another organized group or not. Since noncombatants know $\mathcal{N}^{j_1}$, they can identify whether $J = 1$ or $J > 1$. This is done by comparing the number of components of the Poisson mixture in each district to the districts in which j_1 is located. If $J = 1$ the number of attacks follows either a Poisson distribution with parameter $\eta\ell_i$ or a Poisson mixture

$$x_{i,t} \sim p_{j_1} Poisson\left(\epsilon^h \alpha_{i,j_1} + \eta\ell_i\right) + (1 - p_{j_1}) Poisson\left(\epsilon^l \alpha_{i,j_1} + \eta\ell_i\right)$$

which is identified. If $J = 1$, the proof of this part is finished.

Let $J = 2$. Denote the second organized group as j_2. There are two cases to consider, overlapping territories, $\mathcal{N}^{j_1} \cap \mathcal{N}^{j_2} \neq \varnothing$, or non overlapping territories, $\mathcal{N}^{j_1} \cap \mathcal{N}^{j_2} = \varnothing$. Only the former case is worthy of attention, as the latter case simply repeats the above argument utilized for j_1 (including for the case $N^{j_2} = 1$, which is always identifiable through the overdispersed Poisson argument discussed above).

Focusing on overlapping territories, it must be the case that there exists at least one district $i_3 \in \mathcal{N}$ such that $\text{Var}(x_{i_3,t}) > \text{Var}(x'_{i_3,t})$, where

$$x'_{i_3,t} \sim p_{j_1} Poisson\left(\epsilon^h \alpha_{i_3,j_1} + \eta\ell_{i_3}\right) + (1 - p_{j_1}) Poisson\left(\epsilon^l \alpha_{i_3,j_1} + \eta\ell_{i_3}\right)$$

and

$$\begin{aligned}x_{i_3,t} \sim\; & p_{j_1}p_{j_2} Poisson\left(\epsilon^h\left(\alpha_{i_3,j_1} + \alpha_{i_3,j_2}\right) + \eta\ell_{i_3}\right) \\ & +(1 - p_{j_1})(1 - p_{j_2}) Poisson\left(\epsilon^l\left(\alpha_{i_3,j_1} + \alpha_{i_3,j_2}\right) + \eta\ell_{i_3}\right) \\ & +p_{j_1}(1 - p_{j_2}) Poisson\left(\epsilon^h \alpha_{i_3,j_1} + \epsilon^l \alpha_{i_3,j_2} + \eta\ell_{i_3}\right) \\ & +(1 - p_{j_1})p_{j_2} Poisson\left(\epsilon^l \alpha_{i_3,j_1} + \epsilon^h \alpha_{i_3,j_2} + \eta\ell_{i_3}\right),\end{aligned}$$

for which all parameters are identified following Feller [1943], Teicher [1961], and Teicher [1963]. The intuition for this step is that in i_3 there must be excess variance coming from the presence of group j_2, as well as j_1 and unorganized combatants. These different forms of overdispersion can be detected by noncombatants. It follows that $\mathcal{N}^{j_2} = \{i \in \mathcal{N} : \text{Cov}(x_{i_3,t}, x_{i,t}) > 0\}$ is identified, and so is N^{j_2}.

Finally, one can identify α_{j_2} by first identifying α_{i,j_2} for any $i \in \mathcal{N}^{j_2}$, which again follows from our argument on overdispersion of Poisson distributions and the fact that p_{j_1} is known (without p_{j_1} labeling in a district with two groups which organized group is which would not be otherwise possible). This completes the first part of the proposition.

For the second part of the proposition, we note that under plan 2 the covariance of attacks across any two districts is 0. This is because ϵ is drawn i.i.d. across all districts, organized groups, and over time. We show that we can construct two observationally identical $\mathcal{N}^j$ for $j = j_1, j_2$ for the noncombatants, implying that N^{j_1} and N^{j_2} cannot be separately identified as long as the noncombatants' prior places non-zero weight on the possibility that $J > 1$.

Let the district i_1 be the one valid under Assumption 1. By Assumption 3(i), there are at least two other districts outside of i_1 with organized groups present. Denote them as i' and i''. By Assumption 3(ii) there exists at least one of i' and i'' which will only have one group present. Without loss of generality, we consider the case in which there is one organized group in i', and districts i_1, i', i'' are the only three districts with any organized group present.[62] There are two cases to consider: when there is one organized group in i'', and when there are two groups in i''.

Noncombatants can identify the presence of organized groups in all districts, since they know the number of organized groups in each district, due to the identification of the number of components in the Poisson mixtures.[63] However, if there is one organized group in i'', since the noncombatants' prior places non-zero weight on the possibility that $J > 1$, i' can be characterized by the presence of a second group j_2. Hence, the distribution $\mathcal{N}^{j_1} = \{i_1; i''\}$, $\mathcal{N}^{j_2} = \{i'\}$ is observationally identical to the noncombatants as $\mathcal{N}^{j_1} = \{i_1; i'\}$, $\mathcal{N}^{j_2} = \{i''\}$, satisfying the priors and all Assumptions.

If there are two organized groups in i'', then the noncombatants know both j_1 and j_2 are present in i''. However, they do not know which group is present in i', since Assump-

[62] The same rationale will follow in case where organized groups are present in more than three districts. Here we focus on the lack of identification of N^j within these three districts and keep the others constant.

[63] If $\Pr(J > 1) > 0$, they cannot know whether j_1 is present in i' or i'' (or both), or one of them has j_1 and the other j_2. Note that by Assumption 1, j_1 will be present in at least two districts.

tion 1 is satisfied with j_1 present in i_1 and i''. Hence, the distribution of organized group presence $\mathcal{N}^{j_1} = \{i_1; i''\}$, $\mathcal{N}^{j_2} = \{i'; i''\}$ is observationally identical to $\mathcal{N}^{j_1} = \{i_1; i'; i''\}$, $\mathcal{N}^{j_2} = \{i''\}$. Therefore, $\mathcal{N}^{j_1}$ is not identified. In this case, when the noncombatants' prior places non-zero weight on the possibility that $J > 1$, it follows that α_j cannot be identified (for any j) by the noncombatants, due to its definition of summing across elements of $\mathcal{N}^j$.

From the arguments of equation (11), even if one were able to identify an individual $\alpha_{i,j}$ for some i, j , noncombatants would not be able to aggregate and identify α_j, as $\mathcal{N}^j$ is not identified. This completes the second part of the proposition.

For the third part of the proposition, again noncombatants can identify the presence of organized groups in all districts, since they know the number of organized groups in each district, due to the identification of the number of components in the Poisson mixtures they can estimate from x_{it}.

We note that if j_1 is the group that operates under plan 1 then α_{i,j_1}will be identified per the argument in part (i) thanks to the nonzero covariances among districts where $\alpha_{i,j_1} > 0$. On the other hand, because j_2 operates under plan 2 in districts where both groups are present or only j_2 is present then an argument analogous to part (ii) applies. Particularly, if noncombatants' prior places non-zero weight (potentially infinitesimal) on the possibility that $J > 2$ there can be a group j_3 operating under plan 2 that can replace j_3 in any district in $\mathcal{N}^{j_2}$. This, again, can produce multiple observationally equivalent and belief-consistent $\mathcal{N}^{j_1}$, $\mathcal{N}^{j_2}$, $\mathcal{N}^{j_3}$ distributions of group presences even if in reality $\mathcal{N}^{j_3} = \varnothing$. The same argument as part (ii) for the lack of identification α_{j_2} follows. ■

Proof. Proposition 2. First let us formalize the concept of identity of the two groups. By Assumption 1, there exists $j_1 \in \mathcal{J}, i_1 \in \mathcal{N}^{j_1}$ such that noncombatants know $i_1 \in \mathcal{N}^{j_1}$ and in district i_1 there is only j_1 as an organized group. Also by Assumption 3 (ii) there exists a $j_2 \in \mathcal{J}, i_2 \in \mathcal{N}^{j_2}$ such that noncombatants know $i_2 \in \mathcal{N}^{j_2}$ and in district i_2 there is only j_2. By the identification of Poisson mixtures describing the attack processes in i_1 and i_2 (Feller [1943], Teicher [1961], Teicher [1963]), α_{i_1,j_1}, α_{i_2,j_2}, p_{j_1}, and $p_{j_2}$are known, and so are $\text{Var}(\epsilon_{j_1,t})$ and $\text{Var}(\epsilon_{j_2,t})$.

Indicate with $x_{i,t}$ the observed total attacks in district i at time t.

For the first part of the proposition, define the set where either group is present $\tilde{\mathcal{N}} = \{i \in \mathcal{N} : \text{Cov}(x_{i_1,t}, x_{i,t}) > 0 \wedge \text{Cov}(x_{i_2,t}, x_{i,t}) > 0 \}$. Consider that by assumption of the proposition, imperfect coordination implies:

$$\frac{\mathrm{Cov}(\epsilon_{j_1,t},\epsilon_{j_2,t})}{\mathrm{Var}(\epsilon_{j_1,t})\mathrm{Var}(\epsilon_{j_2,t})}=\rho<1.$$

For any i in $\tilde{\mathcal{N}}$, notice that the number of attacks follows either a Poisson distribution with parameter $\eta\ell_i$ or a Poisson mixture

$$x_{i,t}\sim p_{j_1}Poisson\left(\epsilon^h\alpha_{i,j_1}+\eta\ell_i\right)+(1-p_{j_1})Poisson\left(\epsilon^l\alpha_{i,j_1}+\eta\ell_i\right) \tag{12}$$

or a Poisson mixture

$$x_{i,t}\sim p_{j_2}Poisson\left(\epsilon^h\alpha_{i,j_2}+\eta\ell_i\right)+(1-p_{j_2})Poisson\left(\epsilon^l\alpha_{i,j_2}+\eta\ell_i\right) \tag{13}$$

or

$$\begin{aligned} x_{i,t} \sim\; & p_{j_1}p_{j_2}Poisson\left(\epsilon^h\left(\alpha_{i,j_1}+\alpha_{i,j_2}\right)+\eta\ell_i\right) \\ & +(1-p_{j_1})(1-p_{j_2})Poisson\left(\epsilon^l\left(\alpha_{i,j_1}+\alpha_{i,j_2}\right)+\eta\ell_i\right) \\ & +p_{j_1}(1-p_{j_2})Poisson\left(\epsilon^h\alpha_{i,j_1}+\epsilon^l\alpha_{i,j_2}+\eta\ell_i\right) \\ & +(1-p_{j_1})p_{j_2}Poisson\left(\epsilon^l\alpha_{i,j_1}+\epsilon^h\alpha_{i,j_2}+\eta\ell_i\right). \end{aligned} \tag{14}$$

Notice that in case (14) $\rho<1$ matters in the relative infrequency of $\left(\epsilon^l,\epsilon^h\right)$ or $\left(\epsilon^h,\epsilon^l\right)$ draws, but that all the Bernoulli draws within and across groups remain independent. All (12), (13) and (14) are finite Poisson mixtures whose parameters are identified. To distinguish between (12) and (13), which are two components mixtures, it is sufficient to check whether

$$\frac{\mathrm{Cov}(x_{i_1,t},x_{i,t})}{\alpha_{i_1,j_1}\mathrm{Var}(\epsilon_{j_1,t})}>\frac{\mathrm{Cov}(x_{i_2,t},x_{i,t})}{\alpha_{i_2,j_2}\mathrm{Var}(\epsilon_{j_2,t})},$$

in which case $i\in\mathcal{N}^{j_1}$ $i\notin\mathcal{N}^{j_2}$ and we set

$$\alpha_{i,j_1}=\frac{\mathrm{Cov}(x_{i_1,t},x_{i,t})}{\alpha_{i_1,j_1}\mathrm{Var}(\epsilon_{j_1,t})},$$

or vice versa (where the converse applies for j_2).

For case (14), we have a four components finite Poisson mixture, which is overidentified in α_{i,j_1} and α_{i,j_2}. Hence, these parameters are recoverable from its moments. These imply $\hat{N}^j=N^j,\hat{\alpha}_j=\alpha_j$ for all j, which proves the first part.

For the second part of the proposition, suppose group j_1 may perfectly communicate its sequence of military value opportunities and rewards employed to j_2, so they can *exactly* mimic each other. Under plan 1, any i_3 where both groups are present will conduce noncombatants to estimate $\alpha_{i_3} = (\alpha_{i_3,j_1} + \alpha_{i_3,j_2})$ from

$$x'_{i_3,t} \sim p_{j_1} Poisson\left(\epsilon^h \left(\alpha_{i_3,j_1} + \alpha_{i_3,j_2}\right) + \eta \ell_{i_3}\right) + (1-p_{j_1}) Poisson\left(\epsilon^l \left(\alpha_{i_3,j_1} + \alpha_{i_3,j_2}\right) + \eta \ell_{i_3}\right).$$

Here α_{i_3,j_1} cannot be separately identified from α_{i_3,j_2}, only their sum is identified. This also means that there could exists a set of beliefs for noncombatants consistent with the data and such that $\hat{\alpha}_{i_3,j_2} + \hat{\alpha}_{i_3,j_1} = \alpha_{i_3}$ with $\hat{\alpha}_{i_3,j_2} > \alpha_{i_3,j_2}$. But also it would must be that $\hat{\alpha}_{i_3,j_1} < \alpha_{i_3,j_1}$. In general, $\hat{\alpha}_{j_2} > \alpha_{j_2}$implies $\hat{\alpha}_{j_1} < \alpha_{j_1}$.

Notice however that for the same argument the two groups can now also claim presence in any district in the set $\mathcal{N}^{j_1} \cup \mathcal{N}^{j_2}$ (bar i_1 for j_2 and bar i_2 for j_1). This implies that a set of beliefs such that $W(\hat{N}^{j_1} > N^{j_1}, \hat{\alpha}_{j_1} < \alpha_{j_1} | m = 1)$ and $W(\hat{N}^{j_2} > N^{j_2}, \hat{\alpha}_{j_2} > \alpha_{j_2} | m = 1)$ would not be inconsistent with the data.

For completeness, notice that depending on the specific functional form assumed for signalling value $W(.)$ it may yet possible that $W(N^{j_1}, \alpha_{j_1} | m = 1) < W(\hat{N}^{j_1} > N^{j_1}, \hat{\alpha}_{j_1} < \alpha_{j_1} | m = 1)$. If this is the case and costs of perfect coordination are lower than such difference, both groups could gain from coordination in this special instance.■

B.1 *Bier, Oliveros, and Samuelson [2007]*

The choice of the mode of insurgent organization can be potentially integrated in models of strategic interaction between insurgents and counterinsurgency forces, with the latter organizing a defensive effort against a specific mode of attack. Consider plan 1 versus plan 2 in our analysis, for example. Such mode of attack choice is equivalent to a location choice in the environment presented in Bier, Oliveros, and Samuelson [2007, p.564]. There the authors show that a counterinsurgency strategic defensive efforts against a specific mode of attack will define the insurgents' incentives in choosing a specific mode of attack. It is plausible in our case that defensive efforts by counterinsurgency forces specifically against opportunistic attacks (for example, providing contingent protection to valuable targets moving across districts) may affect the choice of offense method, possibly shifting it towards employing simultaneous attacks across districts (i.e. a form of negative externality of defensive effort across attack modes).

Bier, Oliveros, and Samuelson [2007] prove existence and fully characterize the equilibrium behavior in pure strategies of such game of strategic defense. They also show that in equilibrium it is possible that the defender may optimally choose not to defend against a specific mode of attack, even if a lower risk could be costlessly achieved, if the vulnerability on that attack type margin is low relative to damage potentially inflicted with the alternative modes of attack.

B.2 *Sensitivity of proposition 1 to Assumptions*

Some comments on the role of Assumptions 1-3 are now presented.

- **Assumption 1. Role of Part (ii): The presence of a group in more than one district.** For the identification results to hold in proposition 1, we must have at least one organized group operating in more than one district. If that was not the case, plan 1 would no longer be a centralized case: all organized groups would be deciding independently and within districts, which is a decentralized (plan 2 scenario). There would be no correlation of attacks across districts. Hence, we would fall back into the second part of the proposition, with no identification. **Role of Parts (i) and (iii):** If we do not assume that there is a district in which only one known group operates (i.e. that is known to noncombatants), then groups are identifiable only up to a "relabeling" (i.e. a permutation of the group identities) within each district. To see this, let us modify Assumption 1 to Assumption 1':

  Assumption 1': If $J > 0$, then there exists $j \in \mathcal{J}$ and $i \in \mathcal{N}^j$ such that (i) noncombatants have the prior $\Pr(i \in \mathcal{N}^j) = 1$; (ii) $N^j > 1$.

  Suppose there are two groups: j and j_2 operating in district i and district i_2. Noncombatants will identify a positive covariance between attacks in districts i and i_2. Noncombatants however cannot know which attacks are from group j or j_2, as they do not know to whom attribute which mixture weights. They may equivalently attribute a mixture weight to group j or group j_2. As the noncombatants cannot know which portions of the attacks originates from which group, α_j and α_{j_2} cannot be identified.

- **Assumption 2. Unidentifiability of ϵ^h, ϵ^l separately from α.** Poisson mixtures can be identified up to the values $\epsilon_{j,t}\alpha_{i,j}$ within a district. As discussed, the scale $\epsilon^h - \epsilon^l$ cannot be identified and it will show up in the variance of $\epsilon_{j,t}$. Another way to see this is if one divides both values of ϵ by a constant factor, and multiply $\alpha_{i,j}$ by that factor in every district (and hence, α_j), one cannot separately identify them.

- **Assumption 3. Role of Part (i): Why we need organized groups present in three districts**. This is needed only in the second part of the proposition 1. If this is not the case, it is possible to be in the case of only two districts with organized group presence. Assumption 1 would require group j to be in both and the other organized group to be in the one of the two that is not i. Hence, N^j will be trivially identified even under plan 2. **Role of Part (ii): Needing at least two districts with only one group.** In case there are not two districts with only one group, then N^j will be identified even under plan 2. This is because in the district with the lone group, Assumption 1 guarantees that this group is known. Since the other districts have both groups, the noncombatants trivially know N^j for any j. Yet, α_j still will not be identified, as noncombatants do not know which attacks are due to which groups in the districts with both groups.

- **Case of $J > 2$.** This case does not change substantially the conclusions of proposition 1 other than in one aspect: $J-1$ district strongholds - i.e. districts for which the organized group identity is known and presence is unique - are necessary to anchor the labelling of the groups. This is the same role of Assumption 1 for proposition 1 and this is only necessary to assign a specific identity "label" to each group in the general case. This aspect may be relevant for the signalling purposes of each specific organized insurgent group, but it is not necessary for our empirical analysis of unobserved insurgent structures in Section 4 where no "labeling" of the J groups is necessary.

B.3 *Leaders' Military Valuation under Different Plans*

To see how the different choice of organization affects j's valuation of expected returns, V_m^j, simply consider a one period commitment $T = 1$ and exclude any considerations for signalling. Indicate with $I_{i,t}^h = 1$ if district i is a high military value district in period t, and with $I_{i,t}^h = 0$ otherwise. Assume that if $I_{i,t}^h = 1$ then rewards are set to r^h and a rank-and-file attacks with probability ϵ^h. If $I_{i,t}^h = 0$ then rewards are set to $r^l = 0$ and a rank-and-file attacks with probability ϵ^l.

Finally, assume that $\epsilon^l b^l < r^l$ and $r^h < \epsilon^h b^h$, so that attacks are worthwhile for the j leadership only in districts where returns are high.

Under plan 1, in expectation at $t = 0$, the leadership of group j obtains expected returns:

$$\begin{aligned} V_1^j &= \Pr(\sum_{i\in\mathcal{N}^j}\alpha_{ij}I_{i,1}^h \geq \alpha_j/2)\times\sum_{i\in\mathcal{N}^j}\alpha_{ij}\left(b^h\epsilon^h - r^h\right) \\ +\Pr(\sum_{i\in\mathcal{N}^j}\alpha_{ij}I_{i,1}^h &< \alpha_j/2)\times\sum_{i\in\mathcal{N}^j}\alpha_{ij}\left(b^l\epsilon^l\right), \end{aligned}$$

as it will set high rewards when a majority of its combatants are in high return districts, and low rewards otherwise.

Under mode 2, the expected value, at $t = 0$, for the j group leadership is:

$$V_2^j = \sum_{i\in\mathcal{N}^j}\alpha_{ij}\left[\theta\times\left(b^h\epsilon^h - r^h\right) + (1-\theta)\times\left(b^l\epsilon^l\right)\right].$$

Note that per-period net payoffs are unambiguously higher under (the more flexible) plan 2. So, if the expected utility of j's leadership only depended on V_m^j, the problem of attack plan choice would be trivial, as $V_2^j \geq V_1^j$. Given a discount factor δ, notice that these valuations still respect this inequality no matter the value of T, as the problem is stationary. Specifically, if one indicates as $V_m^j(\tau)$ a τ periods valuation, then $V_m^j(T) = \frac{V_m^j(1)\left(1-\delta^{T+1}\right)}{1-\delta}$, with the expected difference $\Delta V^j(T) = V_2^j(T) - V_1^j(T)$.

C Decomposition of Covariance Matrix

Let $\gamma_{ii'} = \sum_j \alpha_{ij}\alpha_{i'j}$ denote the off-diagonal entry on row i and column i' of Γ_L. Let $\bar{\gamma}_{ii'}$ be the corresponding entry of the covariance matrix in the observed sample. Unfortunately, no empirical counterpart to Γ_L is observed, and thus one will have to be created by modifying the diagonal of the observed covariance matrix $\bar{\Gamma}$.

To create a $\hat{\Gamma}_L$ from $\bar{\Gamma}$, a diagonal matrix $\hat{\Gamma}_D$ will be subtracted from the latter to produce the former. An intuitive method for doing this is "trace minimization", discussed at least as early as Ledermann [1940]. First, note that $\bar{\Gamma}$ is a (sample) covariance matrix, and is thus positive semi-definite. $\hat{\Gamma}_L$ should also correspond to a covariance matrix, and thus should also be positive semi-definite. Consider the optimization problem

$$\begin{aligned} &\min_{\hat{\Gamma}_D} \mathrm{Tr}(\hat{\Gamma}_L) \\ &\text{s.t. } \hat{\Gamma}_L = \bar{\Gamma} - \hat{\Gamma}_D, \quad \hat{\Gamma}_D \text{ diagonal}, \\ &\qquad \hat{\Gamma}_D \succ 0, \hat{\Gamma}_L \succ 0 \end{aligned} \tag{15}$$

Here Tr() denotes the sum of diagonal entries of a matrix, and $\succ 0$ indicates positive semi-definiteness. The intuition for trace minimization is that the "extra" variance present in the diagonal entries of Γ has the form of a full rank matrix, and thus in order to recover a low rank matrix such as Γ_L, as much of this as possible needs to be removed.[64]

If $N = 200$, the the semi-definite program corresponding to (15) involves $200 \times 199 = 39,800$ constraints: each off-diagonal entry $\bar{\gamma}_{ii'}$ in the positive semi-definite matrix $\bar{\Gamma}$ must be equal to the corresponding entry in $\hat{\Gamma}_L$. Problems of this size are feasible using modern semidefinite programming algorithms. We thus compute $\hat{\Gamma}_L$ using (15), and will use it as the basis for producing an estimate of insurgent group presence in the next two subsections.

[64] Saunderson et al. [2012] show that the intuition of Ledermann and others was correct in general. Specifically, the positive semi-definite matrix Γ_L can be recovered given Γ so long as it is sufficiently "incoherent", and this property is satisfied by most low rank matrices. Details are provided in Appendix D.1.

D Estimate for α

Begin by supposing that each organized group that is present has members in a large number of districts, and that no single district has a particularly large α_{ij}. Let I_j be the set of districts that have members of organized group j. Then, since an assumption of the model was that the organized groups do not overlap, an estimate of α_{ij} for $i \in I_j$ can be produced via the following approximation, using $\bar{\Gamma}^j$, the relevant block of the original $\bar{\Gamma}$.[65]

Specifically, note that a sum across the off-diagonal entries of a row of $\bar{\Gamma}$ corresponding to district i is $\sum_{i' \neq i} \alpha_{ij}\alpha_{i'j}$. If there are a large number of districts with members of j, then it is reasonable to use the approximation

$$\begin{aligned} \sum_{i' \neq i} \alpha_{ij}\alpha_{i'j} &\simeq \sum_{i'} \alpha_{ij}\alpha_{i'j} \\ &= \alpha_{ij} \sum_{i'} \alpha_{i'j} \\ &= \alpha_{ij} a_j \end{aligned} \tag{16}$$

where $a_j = \sum_{i'} \alpha_{i'j}$ is the same for any choice of district i within I_j. The row sums of the off-diagonal entries of each block of $\bar{\Gamma}^j$ thus give the relative prevalence of organized group members in each district in I_j.[66]

D.1 Recoverability of Low-Rank Matrix

We are interested in the conditions under which the $\hat{\Gamma}_L$ resulting from (15) will be a consistent estimator for Γ_L. It is clear that there are some matrices Γ_L for which the proposed method will be inconsistent:

[65] A potential alternative approach to the one presented here would be to use the diagonal entries of $\hat{\Gamma}_L$ to produce estimates of $\{\alpha_{ij}\}$. However, this matrix is itself the output of a semi-definite program based on $\bar{\Gamma}$. The approach presented below has the advantage of using the off-diagonal entries of $\bar{\Gamma}$ directly.

[66] While it would be possible to use non-linear programming or other techniques to develop an estimator with more desirable properties, the approximate estimator has at least two advantages. First, the estimator has an intuitive interpretation: $\bar{\Gamma}$ is a covariance matrix, and the sum across the off-diagonal entries of a row of $\bar{\Gamma}$ thus gives an indication (in a heuristic sense) of how closely linked attacks in a given district are with attacks in other districts. Second, if in the data a given district i experiences only a small number of attacks, then the off-diagonal entries $\bar{\gamma}_{ii'}$ will be relatively small for that district, and thus i will not introduce substantial noise into estimates $\hat{\alpha}_{i'j}$ for other districts i'. Developing an unbiased estimator that also possesses such properties appears to be a non-trivial undertaking.

Example 3. *Suppose that there are three districts, and two groups. Group memberships are* $\alpha_{\cdot 1} = (1, 0, \delta)$ *and* $\alpha_{\cdot 2} = (0, 1, \delta)$*, and thus*

$$\Gamma_L = \begin{bmatrix} 1 & 0 & \delta \\ 0 & 1 & \delta \\ \delta & \delta & 2\delta^2 \\ \end{bmatrix}$$

for some small value δ*. Suppose that there are disorganized insurgents such that* $\Gamma_D = I_3$*. The minimum trace heuristic of (15), will then give an estimate*

$$\hat{\Gamma}_L = \begin{bmatrix} \delta & 0 & \delta \\ 0 & \delta & \delta \\ \delta & \delta & 2\delta \\ \end{bmatrix}$$

which has lower trace than the true Γ_L *so long as* δ *is small.*

It is thus important to provide conditions for the matrix Γ_L such that the proposed method gives a consistent estimator. Saunderson et al. [2012] give such a characterization. First, Saunderson et al. [2012] define a subspace $\mathcal{U}$ as realizable if, for any Γ_L having column space $\mathcal{U}$, and any Γ_D, the minimum trace factorization algorithm of (15) applied to $\Gamma = \Gamma_D + \Gamma_L$ returns $\hat{\Gamma}_L = \Gamma_L$. Next, they define the "coherence" $\mu(\mathcal{U})$ of a subspace $\mathcal{U}$ of $\mathbb{R}^n$ as

$$\mu(\mathcal{U}) = \max_{i \in \{1,2,\ldots n\}} ||P_{\mathcal{U}} e_i|| \tag{17}$$

where e_i are the standard basis vectors, and $P_{\mathcal{U}}$ is the orthogonal projection matrix onto $\mathcal{U}$. They then provide the following sufficient condition:

Theorem 4 (Saunderson et al. 2012)**.** *If* $\mathcal{U}$ *is a subspace of* $\mathbb{R}^n$ *and* $\mu(\mathcal{U}) < 1/2$*, then* $\mathcal{U}$ *is realizable.*

From an intuitive perspective, this restriction on coherence is equivalent to nothing in the column space of Γ_L being too close to the standard basis vectors. In the context of estimating insurgent groups, the standard basis vectors represent groups that are only present in one district. It makes sense that groups of this sort will result in the procedure in (15) being inconsistent: a group that is only present in one district is indistinguishable from disorganized insurgents, as they both only appear in the diagonal entries of the covariance matrix.

Saunderson et al. [2012] also provide a further result, regarding the "realizability of random subspaces". They argue that "most" subspaces of dimension less than $n/2$ are realizable. The intuition here appears to be that a random subspace of low dimension is unlikely to include anything close to a standard basis vector. In general, then, if the number of groups is small relative to the number of districts, the heuristic given in (15) will provide a consistent estimator for the group structure. Cases where the estimator will not be consistent are those where one of the groups is overwhelmingly located in a single district.

E Spectral Clustering Estimator

Spectral clustering is based on the "graph Laplacian" matrix

$$L = D - \Gamma_L \tag{18}$$

where D is a diagonal matrix with entries equal to the row sums of Γ_L. The graph Laplacian thus has off-diagonal entries equal to the negative of those of the adjacency matrix, and diagonal entries such that all rows and columns sum to zero. The graph Laplacian L has a rank of $N - J$, and thus has J zero eigenvalues.[67] Spectral clustering focusses on the number of zero eigenvalues for the associated graph Laplacian matrix L, whereas the method used in the main text produces an estimate $\hat{J}$ of the number of insurgent groups by examining (in a very broad sense) the rank of Γ_L.

If Γ_L were known, the number of organized groups could be calculated immediately, and it would equal both the rank of Γ_L and the number of zero eigenvalues of L. However, the data available gives the sample covariances $\bar{\gamma}_{ii'}$ rather than the true $\gamma_{ii'}$, and thus a noisy $\hat{\Gamma}_L$ must be used instead of the true Γ_L. The simplest option for actually implementing a spectral clustering approach is to use a modification of Shi and Malik [2000]: use $\bar{\Gamma}$ to construct $\bar{L}$, and then count the "zero" eigenvalues of $\bar{L}$.

In a finite sample, however, these eigenvalues calculated from $\bar{L}$ are subject to finite sample variation. In particular, random variation will result in positive $\bar{\gamma}_{ii'}$ entries in some cases where the true $\gamma_{ii'}$ is zero, and negative $\bar{\gamma}_{ii'}$ entries in some cases where the true $\gamma_{ii'}$ is positive. This random variation will tend to increase the rank of the $\bar{L}$ relative to L. This problem is particularly severe for districts i for which there are few attacks: the data provides little information on the group structure in these districts, and if one object of interest is J, the total number of groups, the inclusion of these particularly noisy districts could result in a substantial amount of additional noise in the estimate $\hat{J}$.

[67] The number of zero eigenvalues of the graph Laplacian matrix corresponds to the number of connected components of the weighted undirected graph described by the adjacency matrix Γ_L. This is J, the number of blocks of Γ_L.

The intuition for this result is relatively straightforward. Each Γ_L^j block has rank one. The corresponding block of the diagonal matrix D has full rank. Setting the entries in this diagonal matrix so that rows and columns of the graph Laplacian L sum to zero ensures that the rows (and columns) of L corresponding to each Γ_L^j block are linearly dependent. The Γ_L^j block that was subtracted, however, is only rank one, and thus the null space of the resulting block of L must be rank one. This is true for every block in L, and thus the null space of L has dimension J. This will also be the number of zero eigenvalues of L.

A similar problem affects the approach presented in the main text, which is based on using the the largest eigenvalues (or other components) of $\hat{\Gamma}_L$. Finite sample variation will also affect these eigenvalues. The question thus arises whether it is better to use $\hat{\Gamma}_L$ directly, or instead use the corresponding graph Laplacian matrix L. Direct use of $\hat{\Gamma}_L$ requires confidence that the trace minimization algorithm in (15) will work well in finite samples, while use of L avoids this issue because the diagonal entries in question are subtracted away and thus are irrelevant. On the other hand, using L requires labelling some eigenvalues as "zero" eigenvalues, despite the fact that due to random noise all eigenvalues will probably be non-zero.[68] A particular concern here is that the eigenvalues in question are the smallest out of N eigenvalues. Monte Carlo exercises (available upon request) suggest that the approach based on using $\hat{\Gamma}_L$ directly has better finite sample performance. We thus use this approach in our analysis, as described in the main text. Below, we briefly discuss how the alternative approach (based on the smallest eigenvalues of the graph Laplacian) might be applied.

A heuristic method is available based on "eigengaps" similar to those used by Ng, Jordan, and Weiss [2002]. Sort the eigenvalues λ of L in increasing order, such that λ_1 is the smallest and λ_N the largest.[69] The difference $\lambda_{k+1} - \lambda_k$ is defined the kth eigengap. Ng, Jordan, and Weiss [2002] argue that a large eigengap indicates that perturbation of the eigenvectors of L would not change the clusters produced by spectral clustering. Luxburg [2007] thus suggests that the right choice for $\hat{J}$ is a number such that λ_k is "small" for $k \leq \hat{J}$, and the $\hat{J}$th eigengap is large.[70] The intuition here is that if there truly are $\hat{J}$ eigenvalues that are zero, then these appear to be non-zero in the finite sample only due to random variation. In contrast, the $\hat{J} + 1$th and larger eigenvalues would be strictly positive even if the true L were used. An examination of the $\hat{J}$th eigengap thus provides a heuristic test of whether the choice of $\hat{J}$ was reliable, or whether small changes due to random variation might result in a different number of zero eigenvalues.

[68] Eigenvalues that would be zero asymptotically will not be zero in a finite sample, because some of the entries that are zero in Γ_L will be positive in the calculated $\hat{\Gamma}_L$. When using a covariance matrix that includes this finite sample variation, it is thus necessary to account for the fact that eigenvalues that are zero in the population may not be zero in the sample.

[69] A first step to dealing with the problem of finite sample is to exclude districts with very few attacks from estimation: for the analysis of the Afghan data, we used data only for those districts in which there were 3 or more attacks (other cutoffs yielded similar results). This approach does not fully solve the underlying issue, however. For simplicity the notation here assumes that no districts are excluded on this basis and thus there are still N districts, and N eigenvalues.

[70] The underlying difficulty here is determining what exactly constitutes a "zero" eigenvalue, when there is finite sample variation. The presence of a large eigengap would thus provide some confirmation that an appropriate definition of "zero" has been chosen.

Using this approach, the estimated $\hat{J}$ corresponds to an eigenvalue such that λ_k is "small" for all $k \leq \hat{J}$. The presence of high eigengaps for very high values of k is not relevant for the eigengap procedure, so long as $J_{\max}$ is lower than these values. Luxburg [2007] suggests that the cutoff between "small" and "large" should not be larger than the minimum degree in the graph. This is trivially met by $\hat{J} = 1$, but would be violated by any much larger estimate. Although the "eigengap" approach is intended to be heuristic rather than formal, it is possible to compare the first eigengap to simulated data where there is no group structure. Compared to data where the attacks in each district have been reassigned to a random date, the first eigengap in the actual Afghanistan attack data is larger, and this difference is statistically significant at the 95% level.

More formal tests could also be constructed. Each off-diagonal $\bar{\gamma}_{ii'}$ entry will converges to $\gamma_{ii'}$ as the number of time periods grows, and the $\bar{\Gamma}_L$ matrix will converge to Γ_L. Thus, $\bar{L}$ will converge to L. Asymptotically, the correct number of the sample eigenvalues of $\bar{L}$ will approach zero. Thus, from a theoretical perspective, a test statistic similar to that given in Yao, Zheng, and Bai [2015] could be used to determine the number of zero eigenvalues. This test statistic appears to have originated from Anderson [1963], and a simplified version appears to be appropriate in this case: the eigenvalues that are converging to zero are doing so at a $\sqrt{T}$ rate, and thus for the K smallest eigenvalues, the test statistic $\sqrt{T}\sum_{k=1}^{K}\lambda_k$ or $T\sum_{k=1}^{K}\lambda_k^2$ could be used.[71]

Unfortunately, the asymptotic distribution of these test statistics is not clear, and it is also not obvious that a subsampling bootstrap approach would yield the correct distribution either. Simulations suggest that there are certain cases where the correct number of groups will only be obtained with high probability when a very large number of time periods are observed. Specifically, consider the case where α_{ij} is positive but very close to zero for some i and j. That is, there are members of group j in district i, but there are very few of them. In this case $\gamma_{ii'}$ will be very close to zero for all the other i' that contain members of group j. It is thus difficult to distinguish between i containing its own separate group, and i being a part of group j. This suggests that a formal test following this approach might be difficult to implement.

[71] The asymptotic argument is made with a fixed number of districts, N, and a growing number of time periods, T.

F NNMF Consistency

Conditions under which $\hat{\Gamma}_L$ will converge to Γ_L have been discussed in Appendix D.1. We now consider conditions under which a non-negative matrix factorization of Γ_L will recover the $\{\alpha_{ij}\}$ group structure. It is clear that the index numbering of the groups cannot be recovered, because Γ_L is invariant to relabelling of groups. The index numbering of groups is irrelevant throughout our analysis, however, and thus we are only concerned with whether the group structure can be recovered up to a reindexing.

Huang, Sidiropoulos, and Swami [2014] discuss uniqueness of symmetric non-negative factorizations at some length. They conclude that while there are no obvious necessary conditions to check for uniqueness, simulations reveal that multiplicity of solutions does not appear to be a problem unless the correct factorization is extremely dense: factorizations with 80% non-zero entries are still reconstructed successfully. The Γ_L matrices considered in this paper would generally be expected to have a relatively sparse factorization.

G Eigenratio type estimators: Simulations

To better understand the finite sample properties of eigenratio type estimators, we conduct a series of simulations. For simplicity, we do not use a model with discrete attacks, as presented in Section 2, but instead use a more standard model with normally distributed random variables. Let there be $J = 4$ groups, $N = 100$ districts, and $T = 2000$ days. Let there be exactly one group in each district, with $\alpha_{i1} \sim \text{Uniform}(0, 1)$ i.i.d. for $i \in \{1, ..., 25\}$, and no other group present in those districts. In the same fashion, only Group 2 is present in districts 26-50, only Group 3 in districts 51-75, and only Group 4 in 76-100.

Our simplified model of attacks is that in each period t for each group j, an i.i.d. draw $\epsilon_{tj} \sim N(0, \sigma^2)$ is made. The number of attacks is then given by

$$x_{it} = \sum_j \alpha_{ij}\epsilon_{tj} + u_{it} \tag{19}$$

where $u_{it} \sim N(0, 1)$, i.i.d.

We then consider eigenvalues associated with the (N by N) covariance matrix of attacks. We perform 100000 simulations for each of $\sigma^2 = 1$, $\sigma^2 = 0.1$, $\sigma^2 = 0.05$, and $\sigma^2 = 0$, generating a total of 400000 simulated sample covariance matrices.[72]

Figures G.1 - G.3 graphically display the results of these simulations. Figure G.1 shows the eigenvalues of the covariance matrix. We see that the group structure is immediately apparent at $\sigma^2 = 1$, still clear at $\sigma^2 = 0.1$, but somewhat unclear at $\sigma^2 = 0.05$. There is no group structure with $\sigma^2 = 0$, and thus Figure G.1d shows the distribution of eigenvalues under $J = 0$.

Figure G.2 shows eigenratios, with the leftmost eigenratio being the ratio between the largest (i.e. leftmost) and second-largest eigenvalues, and so forth. Here, on average the largest eigenratio clearly corresponds to $J = 4$ when σ^2 is large, but this is no longer the case with $\sigma^2 = 0.05$. Figure G.2d shows that the distribution of eigenvalues when $J = 0$ leads to a somewhat peculiar distribution of eigenratios: the first few and last few eigenratios are much larger than the others. Figure G.2d thus illustrates why it is important to have some maximum number number of possible groups, $J_{\max}$. The eigenratios associated with the very smallest eigenvalues (towards the right hand side of

[72]Note that in the main text, the choice of $\sigma^2 = 1$ is a normalization, because the $\{\alpha_{ij}\}$ are unknown, and a decrease in the choice of σ^2 would simply result in higher $\hat{\alpha}$ estimates. In contrast, in the simulations in this appendix, the distribution of the $\{\alpha_{ij}\}$ are given, and thus choosing a different value σ^2 changes the signal to noise ratio for the attack covariance matrix.

Figure G.1d) become quite large. With $N = 100$, and no $J_{\max}$, choosing $\hat{J}$ based on the largest of all the eigenratios would lead to many $\hat{J}$ estimates of 99 groups. However, as noted in Ahn and Horenstein [2013], any intermediate choice of $J_{\max}$ is unlikely to affect the results.

Figure G.3 shows the distribution of estimates $\hat{J}$ with $J_{\max} = 50$. Figures G.3a and G.3b show that the eigenratio approach works very well when the signal to noise ratio in the covariance matrix is relatively high. Figure G.3c, however, shows that with a noisier covariance matrix, the estimated values for $\hat{J}$ tend to be too low. Figure G.3d shows the distribution of estimates of $\hat{J}$ when there is no group structure.

In both of Figures G.3c and G.3d, $\hat{J} = 1$ is the modal estimate. Figure G.3d shows that the median estimated $\hat{J}$ is below the true value $J = 4$ (the mean is above, but this is less apparent from the figure). However, Figure G.3d shows the case with no group structure at all, and thus would not change regardless of the true value of J. The bias of the estimator thus cannot be signed: this is a natural result of J and $\hat{J}$ both being integers bounded between 0 and 50. Bias correction appears to be non-trivial.

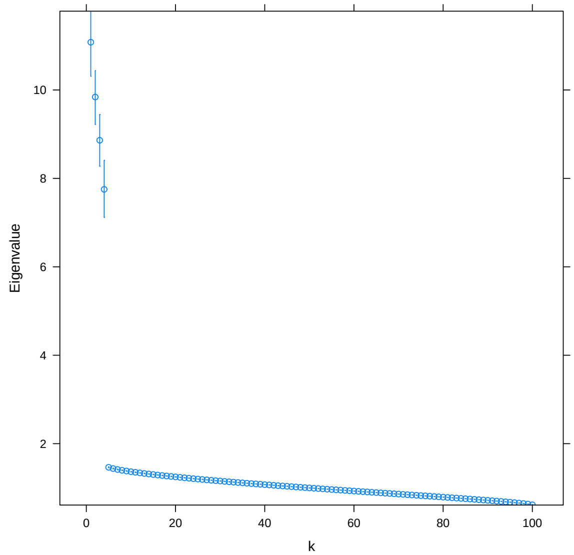
Figure G.3c provides a possible explanation for why estimates of $\hat{J} = 1$ appear so frequently in Table 5. The finite sample properties of eigenratio type estimators are such that there is a tendency to estimate low values of $\hat{J}$ in cases where the covariance matrix is noisy. This is due to the distribution of eigenvalues resulting from the noise, as shown in Figure G.1d. The evidence provided in Table 5 should thus mainly be taken as an indication that the null hypothesis of no group structure should be rejected. Figure G.3c shows how estimates $\hat{J} = 1$ occur frequently when there is actually a group structure with $J > 1$.[73]

G.1 Comparison with Hierarchical Splits

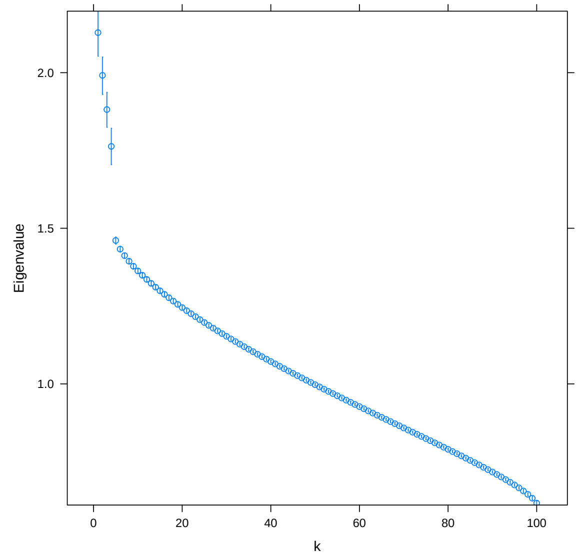
To compare our eigenratio type estimator with the estimator based on hierarchical splits, we need to simulate data with discrete attacks, as the permutation test used require integer numbers of attacks to permute. Let the number of attacks by group j in district i at time t be drawn from a Poisson(λ_{ijt}) distribution, where $\lambda_{ijt} = 0$ with probability 0.9, and $\lambda_{ijt} = \alpha_{ij}$ with probability 0.1 (this is equivalent to ϵ_{it} having a bernoulli distribution with probabilities 0.9 and 0.1). We consider $J = 2$ with the non-zero α_{ij} entries drawn from a Uniform$(0, 0.25)$ distribution, as well as $J = 4$ with the

[73] In the empirical literature, "low" estimates for the number of factors (compared to other methods) are obtained by Choi et al. [2014] and Wu et al. [2011]. Figures G.3b and G.3c appear in line with results reported (using actual data) in the supplement to Baurle [2013].

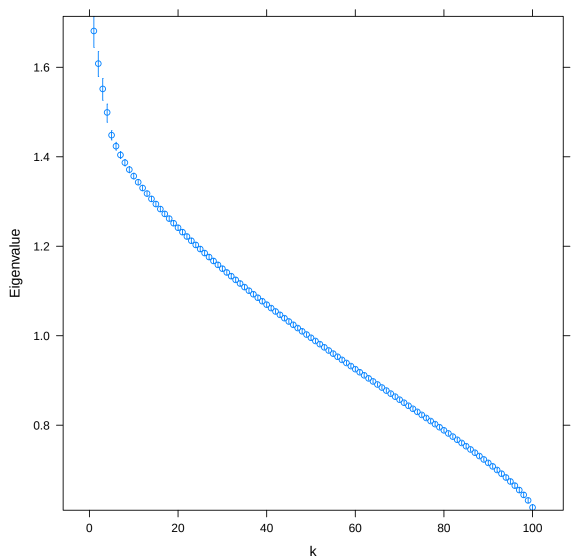
Appendix Figure G.1: Eigenvalues



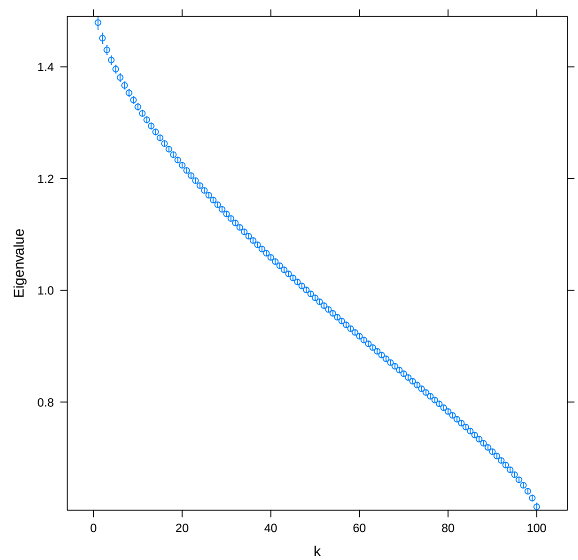
(a) $\sigma^2 = 1$



(b) $\sigma^2 = 0.1$



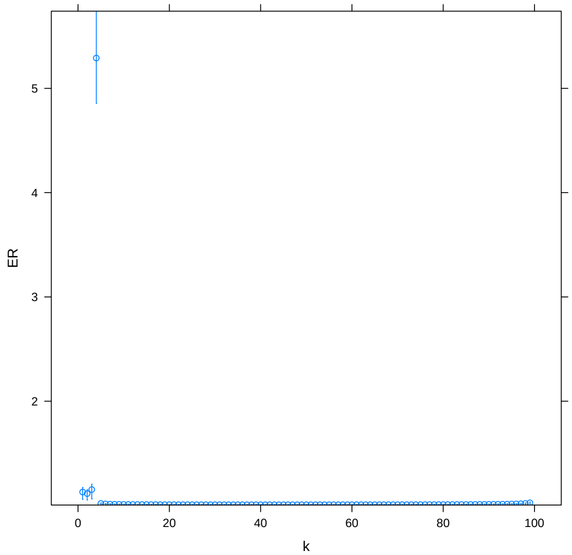
(c) $\sigma^2 = 0.05$



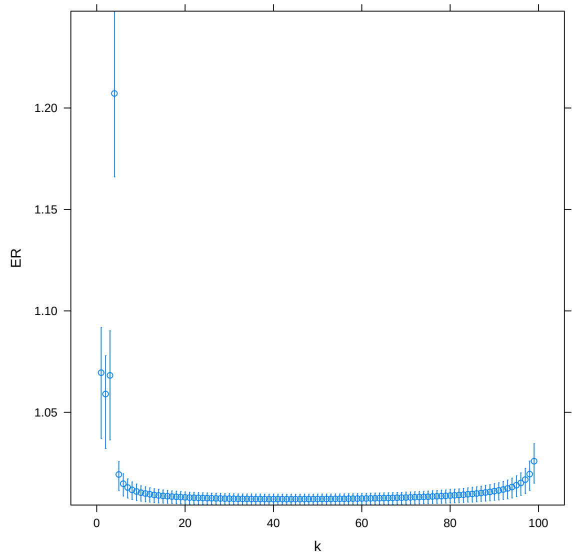
(d) $\sigma^2 = 0$

Points indicate means over 100000 simulations. Bars show interquartile range.

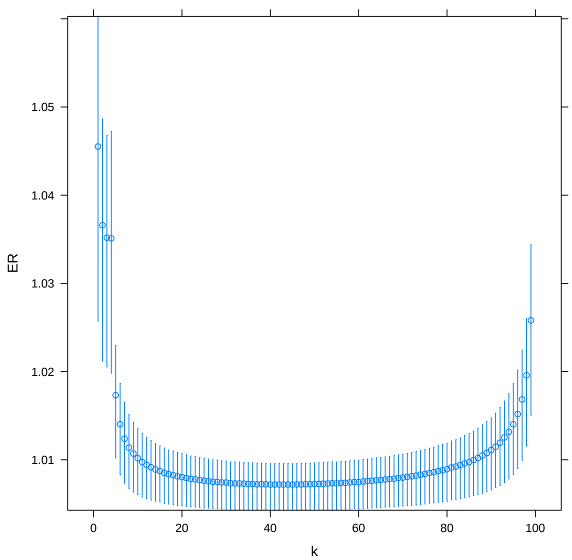
Appendix Figure G.2: Eigenratios



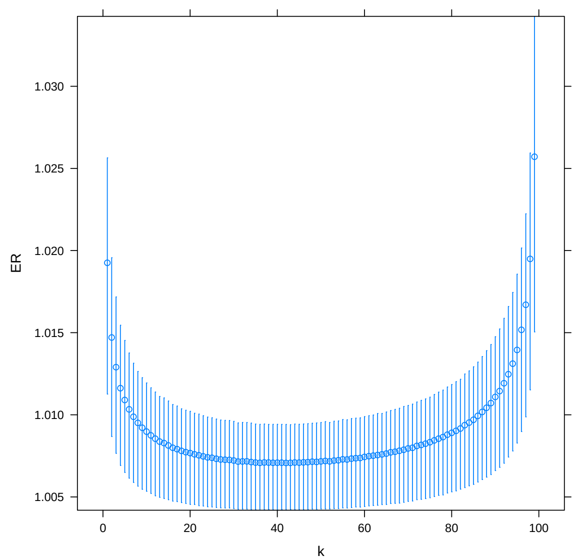
(a) $\sigma^2 = 1$



(b) $\sigma^2 = 0.1$



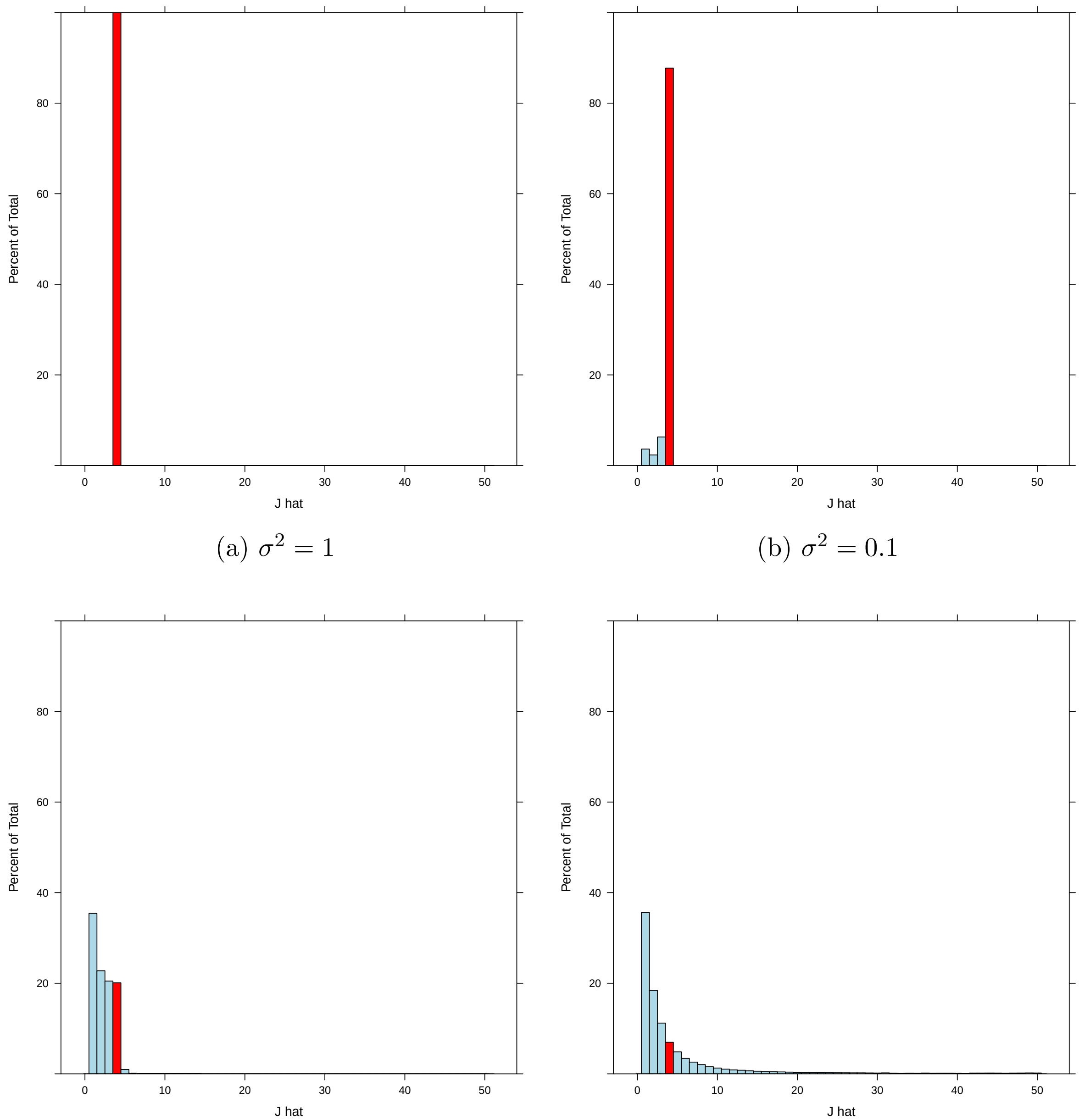
(c) $\sigma^2 = 0.05$



(d) $\sigma^2 = 0$

Points indicate means over 100000 simulations. Bars show interquartile range.

Appendix Figure G.3: Estimated number of groups ($\hat{J}$)



(a) $\sigma^2 = 1$

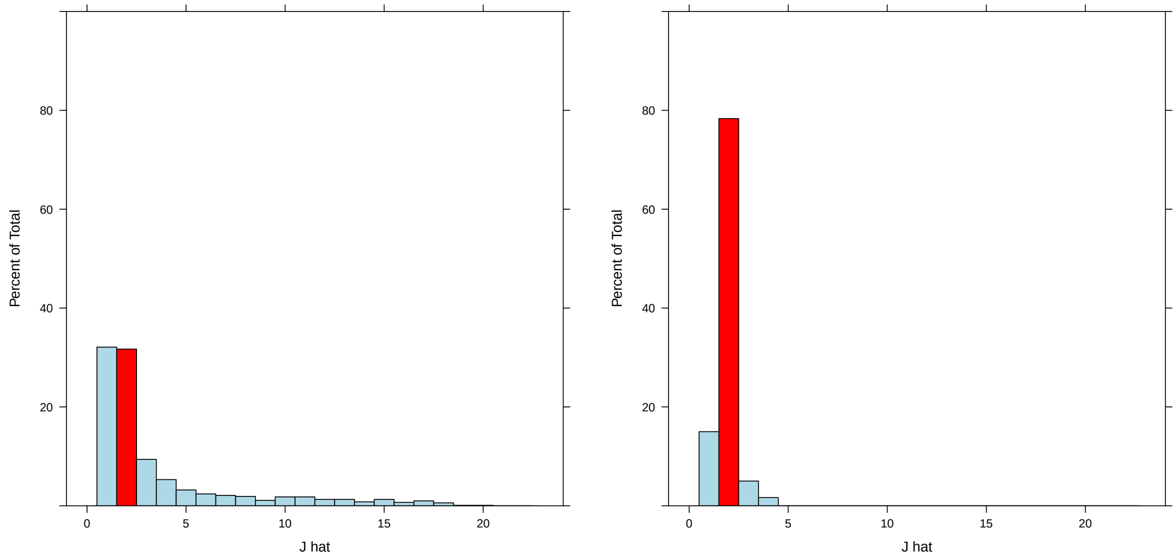
(b) $\sigma^2 = 0.1$

(c) $\sigma^2 = 0.05$

(d) $\sigma^2 = 0$

Histograms of estimated number of groups, over 100000 simulations. True value $J = 4$ shown in red.

Appendix Figure G.4: Estimated number of groups ($\hat{J}$)



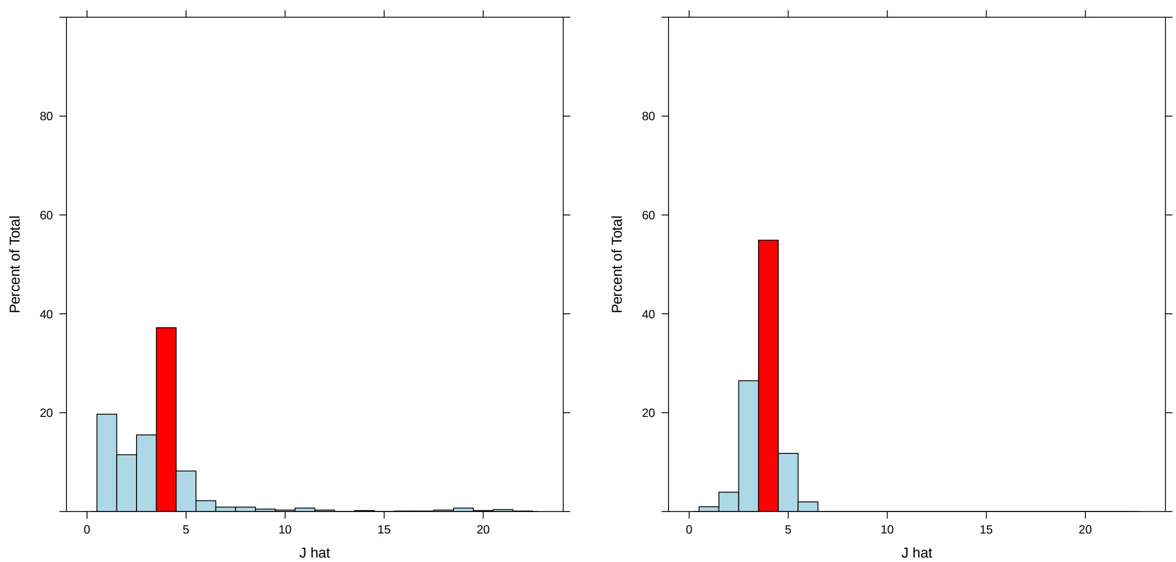
(a) Eigenratio (b) Hierarchical Splits

Histograms of estimated number of groups, over 100 simulations. True value $J = 2$ shown in red.

non-zero α_{ij} entries drawn from a Uniform$(0, 0.5)$ distribution.

Results are shown in Figures G.4 and G.5. In both cases, the method based on heirarchical splits substantially outperforms that based on eigenratios. A particular advantage of the hierarchical splits is that there are no estimates with very large values of $\hat{J}$, whereas with the eigenratio type approach a small number of simulations yield extremely large values for $\hat{J}$. The hierarchical split based method is also less likely to stop at $\hat{J} = 1$, and thus estimates in both tails appear to be less likely with this method.

Appendix Figure G.5: Estimated number of groups ($\hat{J}$)



(a) Eigenratio (b) Hierarchical Splits

Histograms of estimated number of groups, over 100 simulations. True value $J = 4$ shown in red.

H Reference Distributions

We consider three different "reference distributions". First, suppose that the structural model presented in Section 2 is correct. In this case, the distribution of the number of attacks by disorganized militants in district i is the same for all periods, with expected value $\eta \ell_i$. Thus, under the null hypothesis that there is no group structure, the observed attack data is weakly exchangeable: within a given district, permuting the time indices does not change the joint distribution of the attacks.[74] The total number of such permutations is huge, and thus rather than perform calculations using the entire set we consider only a random subset of these permutations. By construction, the permuted data exhibits no group structure: all the off-diagonal entries of the sample covariance matrix will be zero asymptotically. To construct the desired reference distribution, we treat each of these permutations as if it were the observed data.

Now, suppose instead that the structural model assumed is not exactly correct, and there is some cross-time variation in the expected number of attacks by disorganized militants within a district. Specifically, suppose that the probability that a disorganized militant launches an attack is not a constant η, but rather varies across months. The expected number of attacks on a given day in month m is then $\eta_{im} \ell_i$, and will differ by month. In this case, the observed attack data is still weakly exchangeable, but only within a given district *and* a given month. We can thus still construct a reference distribution, provided that observations are permuted only within each month for each district. In this case, the covariance matrices may not have all off-diagonal entries zero asymptotically: it could be that η_{im} and $\eta_{i'm}$ are positively correlated, for example.

Finally, suppose that the expected number of attacks by disorganized militants varies at the daily level, rather than the monthly level. The general case, with $\eta_{it} \ell_i$ attacks expected in district i at time t, is so general that it does not appear to allow for any permutations. However, suppose that the number of expected attacks is instead $\eta_t \ell_i$, where η_t now does not differ across districts.[75] This might be the case, for example, if there were particular days that, for whatever reason, generated large amounts of random

[74] The intuition here can be provided by an example. Suppose there are three periods. If there is no group structure, then the probability of observing $\{x_1, x_2, x_3\}$ in a given district must be equal to the probability of observing $\{x_1, x_3, x_2\}$, because the number of attacks is i.i.d. across time within a given district.

[75] This gives the disorganized militants the same structure an additional organized group. The test against the null hypothesis in this case is thus related to whether there is an organized group present that is active in some districts but not others. Under the null hypothesis, the off-diagonal entries of the sample covariance matrix should be directly proportional to the total number of attacks in the districts in question.

violence. In this case, observations are "approximately" weakly exchangeable via the following sort of permutation, inspired by Good [2002]. Find a pair of districts i and i', and a pair of times t and t', such that the following two conditions hold: there were the same number of attacks x in district i at time t and in district i' at time t', and there were the same number of attacks x' in district i at time t' and in district i' at time t. Permute the data by swapping x and x' in these four entries.[76] These permutations are attractive from an intuitive perspective, as they retain not only the same number of total attacks in each district, but also the same number of total attacks on each day. In the Afghan data, there are relatively few attacks on any given day and thus an enormous number of possible permutations of this sort.

H.1 Additional reference distribution

The purpose of generating permutations is to compute distributions of test statistics, and one of the most obvious test statistics is the fraction of covariance explained by the group structure. Covariance matrices are positive semi-definite, and thus have a spatial interpretation as points in Euclidean space that can be used in order to consider the "between sum of squares" and "within sum of squares" produced by any given clustering. With the permutations just proposed, however, the contribution of different districts to the total sum of squares will generally be different between different permutations, and thus some permutations may be more amenable to clustering than others. In addition, the permutations may in general be more amenable to clustering than the actually observed data, which complicates the interpretation of the permutation test. A way to avoid this would be to use only those permutations where each district makes the same contribution to the total sum of squares as in the actually observed data. While this would be an improvement, the correlation matrix Γ^{cor} is what is block diagonal, and thus is the most appropriate object to analyze using a sum of squares decomposition. To keep

[76] To see why this weak exchangeability holds "approximately", note that the distribution of attacks is binomial. Approximate the binomial with a Poisson distribution with expectation $\eta_t \ell_i$. Then for observations of the type just described

$$\begin{aligned}\Pr(x|\eta_t\ell_i)\Pr(x'|\eta_{t'}\ell_i)\Pr(x'|\eta_t\ell_{i'})\Pr(x|\eta_{t'}\ell_{i'}) &= \frac{(\eta_t\ell_i)^x}{x!}e^{-\eta_t\ell_i}\frac{(\eta_{t'}\ell_i)^{x'}}{x'!}e^{-\eta_{t'}\ell_i}\frac{(\eta_t\ell_{i'})^{x'}}{x'!}e^{-\eta_t\ell_{i'}}\frac{(\eta_{t'}\ell_{i'})^{x}}{x!}e^{-\eta_{t'}\ell_{i'}} \\ &= \Pr(x'|\eta_t\ell_i)\Pr(x|\eta_{t'}\ell_i)\Pr(x|\eta_t\ell_{i'})\Pr(x'|\eta_{t'}\ell_{i'})\end{aligned}$$

by rearranging terms. The canonical reference for multivariate permutations appears to be Pesarin [2001], although this specific type of permutation is not described. Good [2005] provides an accessible introduction to permutation tests.

the contribution of each district to the total sum of squares the same when considering this correlation matrix, we can add an additional requirement that the diagonal entries of the covariance matrix remain the same as those in the actually observed data. This ensures that the transformation to the correlation matrix will involve division by the same quantities as in the actual data, and thus the contribution of each district to the total sum of squares in the correlation matrix will remain the same in the permutation as in the actual data. The permutations that satisfy these additional criteria are a subset of the "swap" permutations discussed above; however, there does not appear to be a way to generate a permutation of the desired type by randomly choosing swaps. It is possible, however, to create valid permutations through the use of an integer program. Let the variables for this program be binary variables x_{ti}^r, which is equal to one if there were r attacks on day t in district i, and equal to zero otherwise. A valid permutation will satisfy the constraints

$$\sum_r x_{ti}^r = 1 \qquad \forall t, i \tag{20}$$

$$\sum_{t=1}^{T} x_{ti}^r = \sum_{t=1}^{T} x_{ti}^{r,\text{actual}}, \qquad \forall i, r \tag{21}$$

$$\sum_{i=1}^{N} \sum_r r x_{ti}^r = \sum_{i=1}^{N} \sum_r r x_{ti}^{r,\text{actual}} \qquad \forall t \tag{22}$$

$$\sum_{t=1}^{T} (\sum_r r x_{ti}^r)(\sum_r \sum_{i=1}^{N} r x_{ti}^{r,\text{actual}}) = \sum_{t=1}^{T} (\sum_r r x_{ti}^{r,\text{actual}})(\sum_r \sum_{i=1}^{N} r x_{ti}^{r,\text{actual}}) \qquad \forall i \tag{23}$$

where $x_{ti}^{r,\text{actual}}$ is a constant corresponding to the actually observed data. The first constraint simply ensures that there is a number of attacks on each day in each district. The second constraint ensures that distribution of attacks within each district is the same as in the actually observed data: this also ensures that the diagonal entries of the convariance matrix are the same as in the actually observed data. The third constraint ensures that the number of attacks on each day is the same as in the actually observed data.[77] The fourth constraint ensures that the sum of each row (and column) of the covariance matrix is the same as in the actually observed data.

[77]This is slightly weaker than the "swap" permutations described above, which preserve the distribution of attacks within each day. There does not appear to be a need for this stronger constraint, and so we relax it here.

A solution to this binary integer program always exists, because the actually observed data will always satisfy the constraints. To randomly generate a solution to the program, we choose a random objective function, and stop at the first integer solution obtained. Running the program repeatedly generates a random sample of permutations with the desired characteristics.

Table H.1 performs the same analysis as 1, except using the above reference distribution instead of using auxiliary geographic information.

Appendix Table H.1: Hierarchical model without geographic information

		Afghanistan I	Pakistan II
Split at (1)?	Randomly shuffled data (mean)	234.06	101.68
	Std. dev.	0.15	0.11
	Actual data	234.02	101.01
	p-value	0.40	0.00
Split at (2)?	Randomly shuffled data (mean)		47.02
	Std. dev.		0.09
	Actual data		46.78
	p-value		0.01
Split at (3)?	Randomly shuffled data (mean)		49.32
	Std. dev.		0.08
	Actual data		49.17
	p-value		0.04
Split at (4)?	Randomly shuffled data (mean)		17.01
	Std. dev.		0.03
	Actual data		17.01
	p-value		0.48
Split at (5)?	Randomly shuffled data (mean)		24.92
	Std. dev.		0.06
	Actual data		24.82
	p-value		0.08
Split at (6)?	Randomly shuffled data (mean)		20.08
	Std. dev.		0.06
	Actual data		19.98
	p-value		0.07
Split at (7)?	Randomly shuffled data (mean)		21.52
	Std. dev.		0.06
	Actual data		21.45
	p-value		0.14

I Estimation using monthly covariance matrices

Suppose that attack probabilities are relatively small. Then the number of attacks by unorganized militants can be approximated using a Poisson($\zeta_{im}\eta\ell_i$) distribution instead of using the actual Binomial($\zeta_{im}\eta, \ell_i$) distribution. Similarly, the distribution of attacks by members of an organized group can be approximated with Poisson($\zeta_{im}\epsilon_{tj}\alpha_{ij}$) in place of Binomial($\zeta_{im}\epsilon_{tj}, \alpha_{ij}$).

Now, suppose that there are a total of x_{im} attacks in district i. Conditional on there being a total of x_{im} attacks, the distribution of these attacks across days is given by a Multinomial(x_{im}, p_i) distribution, where p_i is a probability vector with elements of the form

$$p_{it} = \frac{\eta\ell_i + \sum_j \epsilon_{tj}\alpha_{ij}}{\sum_{t'}\left(\eta\ell_i + \sum_j \epsilon_{t'j}\alpha_{ij}\right)}$$

If in some other district i' there were $x_{i'm}$ attacks, then the covariance of daily attacks has the useful form

$$\begin{aligned}
\text{Cov}(x_{im\cdot}, x_{i'm\cdot}) &= x_{im}x_{i'm}\sum_t p_{it}p_{i't} - \frac{x_{im}}{T}\cdot\frac{x_{i'm}}{T} \\
&= x_{im}x_{i'm}\left(\sum_t p_{it}p_{i't} - \frac{1}{T}\cdot\frac{1}{T}\right) \\
\frac{\text{Cov}(x_{im\cdot}, x_{i'm\cdot})}{x_{im}x_{i'm}} &= \text{SCov}(p_{it}, p_{i't})
\end{aligned}$$

where $\text{SCov}(p_{it}, p_{i't})$ gives the sample covariance for a given draw of ϵ. The first line of the above holds because each attack decision is independent given both the total number of attacks and the realization of ϵ. If the ϵ are constructed such that $\sum_{t'}\epsilon_{t'j} = 1$, then the denominator in the expression above for p_{it} will simplify such that

$$\text{SCov}(p_{it}, p_{i't}) = \frac{\sum_j \alpha_{ij}\alpha_{i'j}\sigma_j^2}{(T\eta\ell_i + \sum_j \alpha_{ij})(T\eta\ell_{i'} + \sum_j \alpha_{i'j})}$$

The $T\eta\ell_i + \sum_j \alpha_{ij}$ term can be taken to be the "average" number of attacks, which implies that $\tilde{\alpha}_{ij} = \frac{\alpha_{ij}}{T\eta\ell_i + \sum_j \alpha_{ij}}$ is the fraction of attacks in district i that group j will be responsible for. Then

$$\text{Cov}(p_{it}, p_{i't}) = \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{i'j}\sigma_j^2$$

Here $\tilde{\alpha}$ and σ^2 are not separately identified. If the normalization $\sigma_j^2 = 1$ is used, then the estimated $\tilde{\alpha}$ describe relative degrees to which groups are more or less responsible for attacks, across districts.

J Gap Statistic

An approach used in a previous version of this paper is based on a modification of the "gap statistic" of Tibshirani, Walther, and Hastie [2001].[78] The intuition behind this technique is that adding an additional cluster increases the number of degrees of freedom in the model, and thus an appropriate estimate $\hat{J}$ must somehow counterbalance this in order to avoid overfitting the data. This is done by comparing the performance of the model with the actual data to a case with randomly generated data that is known not to have any group structure. In general, computational difficulties loom large in the clustering literature and asymptotic behaviour is given only limited consideration. Below, we present a simple computationally feasible estimator for $\hat{J}$.

In general, the gap statistic approach makes use of the value

$$\text{Gap}(k) = E^*[\log(W_k)] - \log(W_k). \tag{24}$$

Here W_k is some measure of variation left unexplained by the k clusters, and E^* is the expectation taken with respect to a "reference distribution" chosen to correspond to no cluster structure. In Equation 24, $\text{Gap}(k)$ quantifies an intuitive definition of the quality of a k group clustering of the observed data: the clustering is "good" only to the extent that it is *better* than what would occur with randomly generated data that had no group structure by construction. Following Tibshirani, Walther, and Hastie [2001], the estimated number of clusters $\hat{J}$ is then selected to be the smallest k such that

$$\text{Gap}(k) \geq \text{Gap}(k+1) - s_{k+1}, \tag{25}$$

where s_{k+1} is an estimated standard error of $\log(W_k)$ in the case where there is no cluster structure. $\text{Gap}(k)$ describes how much better the estimated cluster structure of k groups looks, compared to how it would be expected to look if there were in reality no cluster structure. If $\text{Gap}(k) \geq \text{Gap}(k+1)$, then this means that adding the $k+1$th cluster did not improve the clustering any more than would be expected with random unclustered data.

In the standard Tibshirani, Walther, and Hastie [2001] setup, if clustering were performed using Γ^{cor}, W_k would be the residual sum of squares or "within sum of squares":

[78] An alternative approach would be to compare $\hat{\Gamma}_L^{\text{cor}}$ to the correlation matrix that would be predicted from the k means results, and choose the number of groups that minimized the distance between these two matrices. We do not pursue this approach because it is not clear what sort of distance function might be appropriate.

the variation in Γ^{cor} that is not explained by the group structure. However, this standard approach runs into problems with the calculation of $E^*[\log(W_k)]$ because a reference distribution for $\hat{\Gamma}^{\text{cor}}_L$ needs to be calculated.[79] This calculation appears to be extremely complicated, because the answer depends on the finite sample behavior of $\hat{\Gamma}^{\text{cor}}_L$, which is not well understood. We avoid this problem by modifying the standard Tibshirani, Walther, and Hastie [2001] approach, and use a W_k defined with respect to a set of auxiliary covariates Z, rather than Γ^{cor}.[80]

To see why this simplifies the problem, note that in the model the only source of correlation in insurgent attacks across districts is through ϵ. In particular, our model assumes that if the same insurgent group is present in both districts i and i', the correlation in attacks between districts will not depend on the relationship between any other covariates of i and i': for example, it does not matter whether i is geographically close to i', or geographically distant. We will now add one additional assumption. Suppose that the districts where a given insurgent group is present are less dispersed in terms of these auxiliary covariates Z than a set of randomly chosen districts. For simplicity, we will focus specifically on geography, that is $Z_i = (\text{lat}_i, \text{long}_i)$, but our approach is potentially more general.

Let W_J describe the geographic dispersion of the insurgent group territories when there are J groups, according to the following formula:

$$W_J = \sum_{j=1}^{J} \frac{1}{2N_j} \sum_{i,i' \in I_j} d^2_{i,i'} \tag{26}$$

where $d_{i,i'}$ is the geographic (Haversine) distance between districts i and i'. As before, I_j is the set of districts where insurgent group J is present, and N_j is the cardinality of this set. W_J follows the "within sum of squares" formula from the analysis of variance literature. The intuition here is that, at numbers of groups beyond the true number of groups J, the additional groups will be based on finite sample noise, which is by assumption uncorrelated with geography. Thus, the additional groups should not be correlated with geography, and thus values of W_J should not decrease any faster than

[79] Specifically, we would need to know how much adding a $J+1$th group should improve model fit, if there are only actually J groups in the data.

[80] We employed the standard Tibshirani, Walther, and Hastie approach described here in a previous version of this paper. The results for Afghanistan were identical as those presented below. However, the results for Pakistan were not as easy to interpret. All results are available on request.

would be expected in the case where groups were randomly assigned.[81] The "random assignment" case, needed for determining $E^*[\log(W_k)]$, can be generated via Monte Carlo permutations of the group structure: randomly reassign geographic coordinates to each of the districts, thereby forcing group membership to be unrelated to geographic location.

Under the hypothesis of one organized groups of insurgents, the model is without degrees of freedom (all districts must be associated with that group) and hence the geographic variation left unexplained in both the actual data and the permuted (reference distribution) data is the same (all of it), leaving Gap(0) $= 0$ in the third row (marked "A"). Allowing for two groups of insurgents in the data leads to free parameters (which districts are associated with each group). However, it turns out that there is actually slightly less geographic variation left unexplained in the randomly permuted data compared to the actual data (16.760 vs. 16.765). This produces Gap(2) $= -0.005$ (marked "B"). The gap statistic (B-A) is thus -0.005, which, since it is negative, is definitely lower than 0.006, the estimated standard deviation for unexplained geographic variation in the case where districts are randomly assigned to insurgent groups. Thus, $\hat{J} = 1$ satisfies Inequality 25, and we conclude that there is only one insurgent group present in Afghanistan. The eigenratio approach of Ahn and Horenstein [2013] produces an identical result, as illustrated in Figure 4.

Unlike the case with Afghanistan in Column I, for Pakistan, a comparison of $J = 1$ with $J = 2$ shows that a clustering with two groups based on the attack covariance matrix is substantially more clustered geographically (with a log within sum of squares of 16.897) than would be expected by random chance (an average log within sum of squares of 16.921). This difference (0.024) is large relative to the standard deviation of clustering based on random noise (0.011), and thus Inequality 25 is not satisfied for $J = 1$.

We thus continue down the rows of Table J.2, and proceed to consider the possibility of three organized groups of insurgents. Here again, a clustering based on the actual attack covariance matrix is more geographically clustered than one would expect by random chance. This produces Gap(3) $= 0.176$ (marked "C"). The gap statistic (C-B) is thus 0.152, which is again higher than 0.015, the estimated standard deviation for $\log(W_3)$. The same situation arises with four groups (where $0.105 > 0.019$). In contrast, adding a fifth group results in a clustering that is not as geographically clustered as

[81] There may still be a decrease, but it should not be rapid. Eventually W must fall, because $W_N = 0$, with each group containing only one district.

when there were only four groups. We thus conclude that $J = 4$ in the case of Pakistan, at least in the 2008-2011 period. The unified insurgent structure ($J = 1$) that we recover for the Afghan case thus appears not to be present in Pakistan. This accords with the qualitative analysis in Dorronsoro [2009].

Appendix Table J.2: Gap Statistic

			Afghanistan I	Pakistan II
1 group	Randomly shuffled data (mean)		16.765	16.929
	Actual data	-	16.765	16.929
	Gap	A	0.000	0.000
2 groups	Randomly shuffled data (mean)		16.760	16.921
	Actual data	-	16.765	16.897
	Gap	B	-0.005	0.024
	Gap statistic (B minus A)		-0.005	0.024
	Randomly shuffled data (std. dev.)		0.006	0.011
3 groups	Randomly shuffled data (mean)		16.755	16.912
	Actual data	-	16.757	16.736
	Gap	C	-0.002	0.176
	Gap statistic (C minus B)		0.003	0.152
	Randomly shuffled data (std. dev.)		0.008	0.015
4 groups	Randomly shuffled data (mean)		16.749	16.903
	Actual data	-	16.755	16.622
	Gap	D	-0.005	0.281
	Gap statistic (D minus C)		-0.004	0.105
	Randomly shuffled data (std. dev.)		0.011	0.019
5 groups	Randomly shuffled data (mean)		16.745	16.894
	Actual data	-	16.738	16.758
	Gap	E	0.007	0.136
	Gap statistic (E minus D)		0.012	-0.145
	Randomly shuffled data (std. dev.)		0.011	0.021

Each column computes the gap statistic as described in Section 3.2, based on a within-month covariance matrix as described in Section 3.4. Columns differ in the underlying attack data used:
Column I uses the full Afghanistan WITS dataset.
Column II uses the Pakistan BFRS dataset for May 2008 - October 2011.

K Additional Analysis: Global Terrorism Database

Check 1. Our results indicate that the group structure we estimate for Pakistan corresponds to ethnic homelands. We might thus be concerned that in fact our method is not picking up individual insurgent groups, but rather some broader aspect of coordination within the same ethnic group. The GTD includes some information on the identities of attackers, and we can use this to cross-check our estimated group structure.[82] We examine this data for simultaneous attacks in Pakistan during the period that we study. In Balochistan, the GTD lists 38 attacks. Of these, 32% are ascribed to the Baloch Republican Army, and the remainder are listed as unknown. In the Federally Administered Tribal Areas and North-West Frontier Province, there were 167 attacks. Of these, 31% were ascribed to the (Pakistani) Taliban, 4% to Lashkar-e-Islam (which later joined the Taliban), and the remainder were unknown. Thus, in these cases, our results match what evidence is available: our method finds one group in Balochistan and one more in the area near the Afghan border. The GTD records very few attacks in Punjab, and most of these are Taliban attacks in the part of Punjab nearest to the Afghan border. A comparison for Punjab is thus unfortunately not available.

In Sindh, the GTD reports 24 attacks, but 20 of these involve an unknown group. In the next year, however, there are 54 attacks reported, with 61% of these ascribed to the Sindhu Desh Liberation Army. Our method thus appears to have picked up an organized group operating across Sindh a year or more before this would have been visible by examining the group identification in the best available datasets. Overall, we see that the GTD reports a single group corresponding to our estimated groups for Balochistan, Sindh, and the area near the Afghan border.

Check 2. As an additional verification of our model, we can consider whether our estimated group structure in Pakistan can predict the geographic structure of attacks in a later period. BFRS and WITS data is not available for more recent years, so we use data from the GTD for this analysis. We use data from the Nov. 2011 -Dec. 2016 period, and run a clustering of this data into four groups.[83] The resulting group structure is

[82]Neither the BFRS nor WITS record the group identity of the assailants in a systematic way.

[83]Another possibility would have been to examine data from a point *earlier* than our main period. One of the data requirements for our method to be effective, however, is that there must be simultaneous attacks in the districts that we wish to cluster. Although Pakistan has a long history of terrorism, much of this violence is concentrated in major cities. For example, in 1995 the Global Terrorism Database lists 666 attacks in Pakistan: of those, 614 of them occur in Karachi. The BFRS data similarly has 79% of all attacks occurring in Karachi. It is thus unsurprising that attempting to cluster other districts does not yield meaningful results. For comparison, only 9% of attacks occur in Karachi in 2009, and this is the most attacks in any district during that year. Because of this feature of the earlier data, it

shown in Figure K.9. There are obvious similarities here to the clustering on the original data shown in Figure 7. To quantify these similarities, we run regressions predicting the new group membership using the original group membership: these are shown in Table K.9. In both figures, we see a group that matches the Sindh ethnic homeland, another in Balochistan, and a third in the area near the Afghan border. The GTD data includes fewer attacks than the BFRS data, and has very few incidents in Punjab. We thus do not see any group corresponding to the Punjabi ethnicity, which is main difference between Figures 7 and K.9. Overall, however, the data shows a high degree of persistence in the structure of simultaneous attacks, which suggests that the methods we describe can be used to predict patterns of insurgent coordination in the future.

Check 3. In our estimation strategy, we calculate a covariance matrix based on daily attack data. One might be concerned that in fact we are discarding useful information by considering only coordination within a single day. For example, perhaps one of the ways an insurgency coordinates is to arrange sequential attacks over consecutive days. We can use the GTD to verify that this does not appear to be the case.

The GTD allows for the component attacks of a multiple attack to occur on different days. 96% of all multiple attacks, however, take place only on one day. In addition, most of the attacks that are spread across multiple days actually take place on two consecutive days, and in some cases the notes for the attacks indicate that the attack took place during a single night, with some components occurring before midnight and others after midnight. We thus see that almost all multiple attacks are indeed same-day simultaneous attacks, rather than spread out across time.[84]

Check 4. A concern is that the "groups" that result from our method do not match what a qualitative researcher would consider a group to be. For example, they might be too narrow, classifying as different groups what are in reality simply different branches of the same organization. Alternatively, the groups we estimate might be too broad, lumping together different insurgent organizations that merely cooperate occasionally on campaigns. As our definition of a group is based on same-day simultaneous attacks, we can address this concern by examining how these attacks are attributed to insurgent

---

is unfortunately not possible to track changes in the group structure in Pakistan across time.

[84]One of the major advantages that counterinsurgency forces have is that they are generally more numerous and better equipped than the insurgency that they are fighting. The insurgents, on the other hand, have the advantage of surprise, in terms of both timing and location of attacks. If an insurgent group were to advertise that they would attack a week later, the government would be able to place their forces on high alert, change their deployment strategy, cancel leave, and so forth. The insurgents thus face a higher cost in terms of casualties if they attack with advance warning. Horn [2013] cites a Taliban commander describing how simultaneous attacks prevent a concentration of security forces.

groups by qualitative analysts.

The Global Terrorism Database is again useful here, because it reports group identities where available. There are 6718 sets of multiple attacks in this database, with an average of 3.4 attacks in each set. Identities of the groups responsible are recorded for at least one attack in 67% of these sets. Only 170 sets of attacks (2.5% of the total) have multiple different groups recorded as being responsible for component attacks within the same set of attacks. Of these, the majority are cases where it is unclear whether the attacks were actually coordinated, and one of the groups is listed as unknown (the notes for these attacks often report this uncertainty). There are only 50 cases (0.7% of the total) where there are actually two distinct group names listed, and about half of these are cases where an identified group is clearly responsible for one of the attacks, but it is unclear whether it also committed the other one, and thus the second attack is listed with a more general group description (e.g. Revolutionary United Front vs. Rebels). There are only a few dozen cases where two different identified groups engage in a simultaneous attack. This happens, for example, in Colombia (ELN and FARC) and Chile (FPMR and MIR). It is thus true that sometimes multiple different groups will engage in simultaneous attacks, but these incidents comprise only a fraction of a percent of all simultaneous attacks.

We thus see that our simultaneous-attack based definition of a group is not too wide compared to the definition used by qualitative , because the GTD shows very little coordination of attacks between groups as they define them. A remaining danger is that our definition is too narrow, in that different cells in a group that has a cohesive objective may choose not to coordinate for some reason, and thus we detect too many groups using our method. However, we only detect one group in Afghanistan, and 4 in Pakistan. In Pakistan, separatists in Sindh and Balochistan have their own independent objectives, which are clearly not in alignment with Punjabi interests and also differ from those of the Taliban. It thus seems unlikely that we have detected too many groups in Pakistan, although there does not appear to be a more formal way of testing this using the data sources that we currently have available.

Check 5. Our model suggests that part of the value of the simultaneous attack relies on citizens knowing which group launched the attack, because the attack serves as a signal of this groups strength. It is thus more important that a simultaneous attack actually be attributed to a group, relative to a non-simultaneous attack. In particular, we should expect that groups will claim credit for these attacks at rates that are higher than for non-simultaneous attacks. Table K.5 shows that this appears to indeed be the case, even after controlling for variables that describe the total size and damage that the attacks cause.

Check 6. Another implication of our model is that types of attacks where decentralization is particularly important should be less likely to be simultaneous. For example, consider the difference between bomb attacks against a railroad, versus the assassination of senior government officials. The railroad is close to equally vulnerable every day, although there may be slight variations in the effect of a bombing due to differences in traffic. On the other hand, a given senior government official may be vulnerable to assassination only on certain days, and the probability of an attempt succeeding could vary greatly depending on when the attempt is made. Thus, there would be substantial costs to attempting to coordinate two assassinations: even if the coordinator had perfect information regarding when the targets were vulnerable, the time selected for the attack would still be a compromise that would not have either target at its most vulnerable. We should thus expect that assassinations are much less likely to be part of a simultaneous attack than bombings. Table K.4 shows that this appears to indeed be the case.

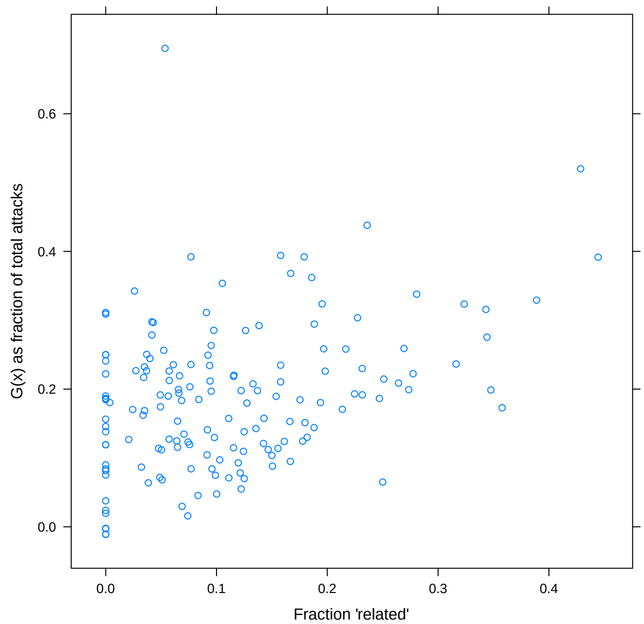
Check 7. Our theory of simultaneous attacks sketched in Section 1 is based on a cost-benefit trade-off of launching a simultaneous attack. In the case where a terrorist group is very weak and disorganized, it may be too difficult to attempt a simultaneous attack. This may, in fact, provide some of the attractiveness of launching such an attack, as it provides a credible signal that the terrorist group is strong and well organized.[85] Is there evidence that weaker groups are less likely to launch simultaneous attacks? The use of country-year data to answer this question is potentially problematic but provides a valuable starting point.

Table K.6 shows that a greater fraction of simultaneous attacks of interest is associated with a higher number of total attacks, even after controlling for country and year fixed effects. However, there is an obvious confounding effect here, because a group that is so weak that it can set off only a single bomb will not be able to launch any simultaneous attacks. One way to attempt to deal with this is by using lagged simultaneous attacks as an instrument for the fraction of simultaneous attacks this year: Columns III and IV of Table K.7 show that results do not change when this approach is used. So, at the very least, evidence from these conditional correlations seems not to counter our intuition.

We further address this point by considering districts of Afghanistan. The advantage

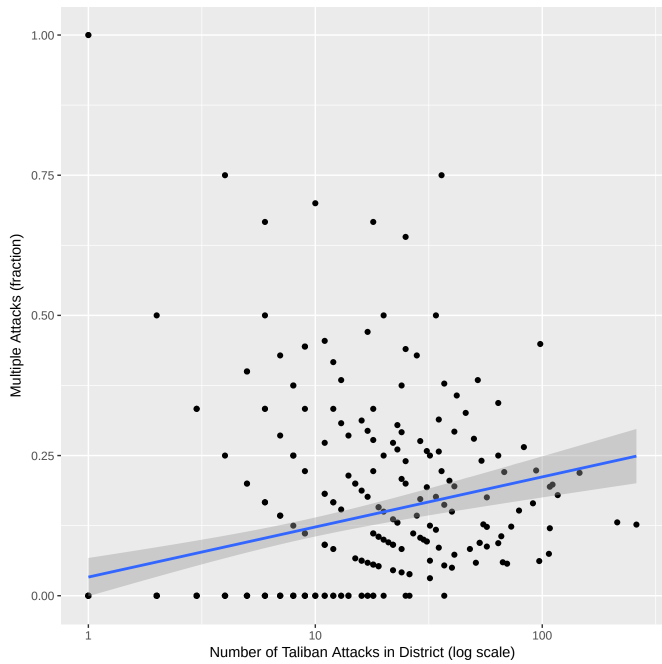
[85] Credibility of the signal follows from the same argument as in the educational signalling literature, where education is more costly for low ability individuals.

Appendix Figure K.6: Overdispersion and "Related" Attacks

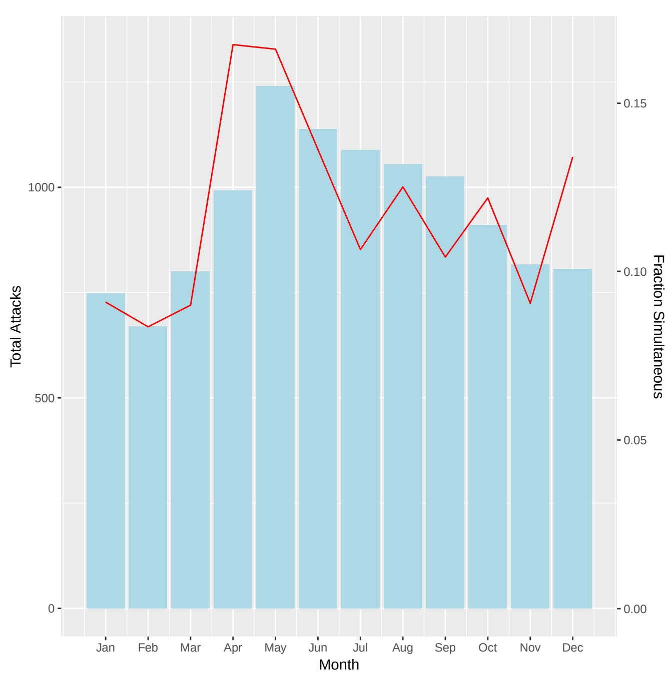


here is that even if there is only one attack in a district, it can still be a simultaneous attack because it is coordinated with an attack in another district. Our hypothesis is that districts where the Taliban are weak are districts where it would be very costly for them to coordinate, and thus are districts where they will not engage in simultaneous attacks. Figure K.7 and Table K.3 show that this indeed appears to be the case.

Appendix Figure K.7: Fraction of Taliban Multiple Attacks



Appendix Figure K.8: Seasonality in Multiple Attacks in Afghanistan



Appendix Table K.3: Fraction of Taliban Multiple Attacks

	OLS	OLS	Logistic	Logistic
	I	II	III	IV
(Intercept)	0.235	0.254	-1.253^{*}	-4.630^{***}
	(0.169)	(0.242)	(0.756)	(1.420)
log(Num Attacks)	0.037^{***}	0.039^{***}	0.308^{***}	0.477^{***}
	(0.009)	(0.010)	(0.051)	(0.077)
log(Population)	-0.022^{*}	-0.022	-0.247^{***}	-0.073
	(0.013)	(0.016)	(0.060)	(0.102)
log(Area)	-0.002	-0.002	-0.088^{**}	-0.217^{***}
	(0.008)	(0.012)	(0.037)	(0.063)
Night Lights 92, 00, 12		Yes		Yes
Province FE		Yes		Yes
Observations	341	341	341	341

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

Observations are districts in Afghanistan. Dependent variable is the fraction of Taliban attacks that are multiple attacks. Attacks by unidentified attackers and non-Taliban attackers are omitted.

Data source: Global Terrorism Database, 1998-2016

Appendix Table K.4: Is the attack part of multiple attacks?

	OLS	OLS	Logit	Logit
	(1)	(2)	(3)	(4)
Armed Assault	0.118***	−0.016	−2.011***	−3.514***
	(0.002)	(0.010)	(0.021)	(0.120)
Assassination	0.032***	−0.036***	−3.397***	−4.352***
	(0.005)	(0.012)	(0.077)	(0.201)
Bombing/Explosion	0.172***	0.070***	−1.568***	−2.584***
	(0.002)	(0.009)	(0.012)	(0.101)
Facility/Infrastructure	0.288***	0.042***	−0.903***	−2.904***
	(0.005)	(0.015)	(0.034)	(0.145)
Hijacking	0.109***	0.027	−2.100***	−3.035***
	(0.024)	(0.039)	(0.216)	(0.418)
Hostage-Barricade	0.141***	−0.043	−1.808***	−3.609***
	(0.024)	(0.034)	(0.197)	(0.372)
Hostage-Kidnapping	0.100***	−0.048***	−2.200***	−3.692***
	(0.005)	(0.015)	(0.043)	(0.171)
Unarmed Assault	0.173***	0.005	−1.562***	−3.211***
	(0.016)	(0.027)	(0.121)	(0.295)
Unknown	0.183***	0.032	−1.498***	−2.995***
	(0.007)	(0.022)	(0.048)	(0.212)
log(Num Perpetrators)		0.050***		0.435***
		(0.002)		(0.023)
log(Num Killed + 1)		−0.001		−0.006
		(0.004)		(0.041)
log(Num Wounded + 1)		0.015***		0.141***
		(0.003)		(0.031)
Country FE		Yes		Yes
Observations	89,338	14,156	89,338	14,156

Note: *p<0.1; **p<0.05; ***p<0.01
Observations are individual terrorist attacks in all countries. Dependent variable is binary: whether or the attack is part of a set of simultaneous (same day) attacks. Attack types are an exhaustive set of dummy variables.

Appendix Table K.5: Was the attack claimed by a terrorist group?

	OLS	OLS	OLS	Logit	Logit	Logit
	I	II	III	IV	V	VI
(Intercept)	0.131***	0.109***		−1.892***	−1.940***	
	(0.001)	(0.006)		(0.011)	(0.038)	
Multiple Attack	0.105***	0.112***	0.068***	0.720***	0.555***	0.531***
	(0.003)	(0.013)	(0.012)	(0.023)	(0.076)	(0.102)
log(Total Num Perpetrators)		0.004	−0.012***		0.020	−0.111***
		(0.003)	(0.003)		(0.017)	(0.025)
log(Total Num Killed + 1)		0.054***	0.023***		0.298***	0.169***
		(0.005)	(0.004)		(0.029)	(0.043)
log(Total Num Wounded + 1)		0.033***	0.022***		0.174***	0.184***
		(0.004)	(0.003)		(0.023)	(0.032)
Country FE			Yes			
Terrorist Group FE			Yes			Yes
Observations	87,901	13,441	13,441	87,901	13,441	13,441

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

Observations are individual terrorist attacks in all countries. Dependent variable is binary: whether or not a terrorist group claimed responsibility for the attack. In the case of multiple attacks, "Total Num" refers to the total number of perpetrators (etc.) in all of the attacks combined.

Column VI omits country fixed effects due to convergence issues (very few terrorist groups span multiple countries).

Appendix Table K.6: Dependent variable is Herf. fragmentation of terrorist groups

	I	II	III	IV
(Intercept)	0.53^*	0.06		
	(0.01)	(0.08)		
Overdispersion	-0.28^*	-0.32^*	-0.28^*	-0.28^*
	(0.04)	(0.03)	(0.03)	(0.03)
Max Possible Overdispersion		0.50^*	0.51^*	0.45^*
		(0.05)	(0.06)	(0.06)
FKMS Controls	No	Yes	Yes	Yes
Country FE	No	No	Yes	Yes
Year FE	No	No	No	Yes
N	1144	1143	1143	1143
R^2	0.05	0.22	0.85	0.86

Robust standard errors in parentheses

* indicates significance at $p < 0.05$

Observations are an unbalanced panel in country and year. Dependent variable is the Herfindahl fragmentation of terrorist attacks by terrorist group within a given country-year. The range of the dependent variable depends on the number of terrorist attacks that occurred: for example, with only one terrorist attack, the only possible fragmentation is 0, while with two terrorist attacks the possible levels are 0 and 0.5. The control variable "Max Possible Overdispersion" is the maximum possible fragmentation given the number of attacks that occurred. A more sophisticated adjustment appears not to exist: see Gotelli and Chao [2013] for discussion.

"FKMS Controls" are the covariates used in Table 1 of Freytag et al. [2011].[86]

Appendix Table K.7: Relationship between number of attacks and overdispersion

	OLS	OLS	IV	IV
	1	2	3	4
(Intercept)	1.517^{***}		0.543^{*}	
	(0.041)		(0.297)	
Overdispersion	2.681^{***}	1.567^{***}	9.447^{***}	10.719^{**}
	(0.198)	(0.115)	(1.574)	(4.413)
FKMS Controls	No	Yes	No	Yes
Country FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
Observations	1,940	1,939	1,568	1,568

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

Observations are an unbalanced panel in country and year. Dependent variable is the log number of terrorist attacks in a given country-year. "Overdispersion" is $G(x)$, as defined in the text. "FKMS Controls" are the covariates used in Table 1 of Freytag et al. [2011]. Columns 3 and 4 use the previous year's overdispersion as an instrument for current overdispersion.

Appendix Table K.8: "Related" attacks and overdispersion

	I	II	III	IV
(Intercept)	0.01	0.02		
	(0.00)	(0.02)		
Overdispersion	0.34^{*}	0.34^{*}	0.34^{*}	0.35^{*}
	(0.03)	(0.03)	(0.03)	(0.03)
FKMS Controls	No	Yes	Yes	Yes
Country FE	No	No	Yes	Yes
Year FE	No	No	No	Yes
N	2006	2005	2005	2005

Robust standard errors in parentheses

* indicates significance at $p < 0.05$

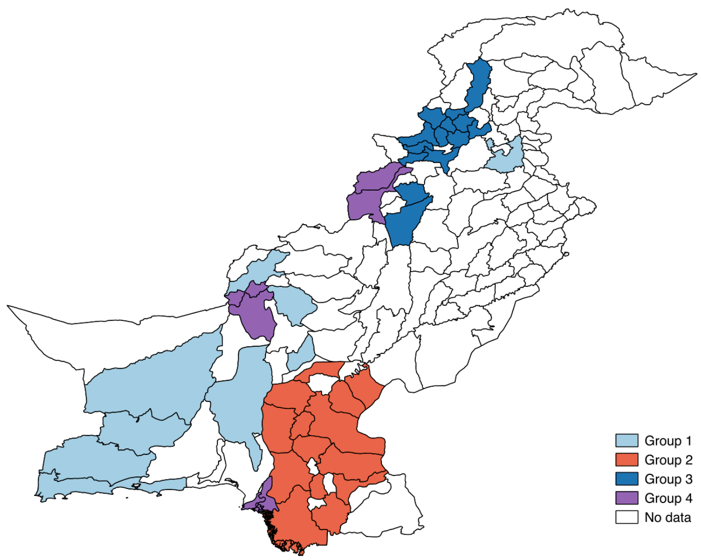
Observations are an unbalanced panel in country and year. Dependent variable is the fraction of terrorist attacks in a given country-year that had "related" attacks. "Overdispersion" is $G(x)$, as defined in the text. "FKMS Controls" are the covariates used in Table 1 of Freytag et al. [2011].

Appendix Table K.9: Pakistan Comparison using post- Nov 2011 GTDB data

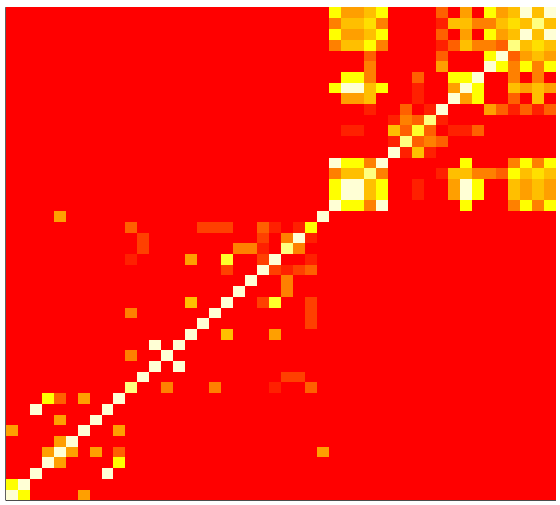
	Group 1	Group 2	Group 3	Group 4
Group 1 (mostly Baloch)	0.63	0.00	0.13	0.25
	(0.13)	(0.12)	(0.13)	(0.12)
Group 2 (mostly Sindhs)	0.07	0.80	0.07	0.07
	(0.09)	(0.08)	(0.10)	(0.09)
Group 3 (mostly Afghans)	0.08	0.08	0.77	0.08
	(0.10)	(0.09)	(0.11)	(0.09)
Group 4 (mostly Panjabis)	0.33	0.17	0.33	0.17
	(0.15)	(0.14)	(0.15)	(0.14)
N	42	42	42	42

Each column corresponds to a single regression without intercept.
The dependent variable is a dummy variable indicating whether a given district was clustered into the specified group number in the clustering shown in Figure K.9.
The independent variables are a set of dummy variables, indicating whether a given district was clustered into the specified group number in the clustering shown in Figure 7c. Districts shown as white ("no data") in either Figure 7c or K.9 are dropped: the remaining 42 districts are used in the regression.
Each row should sum to 1 because each coefficient in the table is a conditional mean giving the fraction of districts of the specified ethnicity that were clustered into the specified group, and the clustering in Figure K.9 assigns each district to one group. Rows may not sum exactly to 1 because of rounding.
Standard errors in parentheses.

Appendix Figure K.9: Pakistan Groups with post- Nov 2011 GTDB Data



Appendix Figure K.10: Covariance Matrix for post- Nov 2011 GTDB Data



Cells of cross-district covariance matrix, coloured from low covariance (red) to high covariance (white). Ordering of rows and columns is the default order for GIS maps of Pakistan, which places districts in the same province together. Three groups are clearly visible. The GTDB data contains very few attacks in Punjab: no group corresponding to Punjab is visible. This data is clustered to produce Figure K.9.

L Insurgency Organization & Economic Recovery

This section briefly discusses case studies chosen to highlight the economic importance of understanding insurgent organization in conflict and post-conflict environments. We focus on three different episodes: Iraq, Syria, and Libya.

Insurgent groups owe their success to their deep ties with noncombatant populations. By impeding reconstruction efforts, they can fuel popular dissatisfaction with central authorities, thereby maintaining a steady flow of recruits and ensuring logistic assistance for their agents. Insurgencies thus have a particular incentive to delay aggregate economic recovery.

In Iraq, insurgents disrupted the electricity grid and seized control of oil resources. Henderson [2005] describes the loop that linked insecurity and economic stagnation:

> *Inability to provide security had a profound impact on Iraq's economic recovery. In turn, inability to provide recovery had a profound impact on Iraq's security. Reconstruction delays fed into Iraqi feelings of resentment and despair, which fueled insurgency and crime, thereby worsening the security climate.*

The connection of the study of insurgency with economic development comes from this tight link between insurgent strategies and the failure of reconstruction efforts. Understanding the exact nature of the Iraqi insurgency early on in the conflict could have proven crucial in breaking the vicious cycle that Henderson [2005] observes.[87]

Uncertainty about the organization of the insurgency in post-2003 Iraq took several forms. First, there was disagreement regarding the extent to which attacks represented an insurgency at all.[88] There was also confusion regarding its magnitude: as late as the fall of 2004, the U.S. military still attributed 80 percent of attacks to random and not to political violence. Finally, there was heated debate about the organization of the insurgency, once it was clear that one existed.[89] Further complexity in the Iraqi case stemmed from signs of evolution over time, as the New York Times reported: *"the*

[87]Henderson is critical of the strategy actually used: *"as violence worsened, the response of coalition officials in charge of reconstruction was not to find a way to fight it more effectively. Instead, their response was to withdraw into the heavily protected world of the Green Zone."*

[88]Eisenstadt and White [2005] write that *"In the summer of 2003, Secretary of Defense Donald Rumsfeld and General John Abizaid (head of U.S. Central Command) publicly disagreed about whether the violence in the Sunni Triangle was the final act of former regime "dead-enders" or an incipient insurgency against the emerging political order"*. There was a similar disagreement in 2005 between Vice President Richard Cheney and General Abizaid.

[89]The New York Times quotes senior U.S. intelligence sources stating that "*It's not just one group*

insurgency was now organized regionally, and that evidence pointed to some planning across regional boundaries".[90]

The difficulty, and the importance, of understanding the structure of insurgencies is not limited to Iraq. Consider recent Western efforts in Syria: *"Sixteen months into the uprising in Syria, the United States is struggling to develop a clear understanding of opposition forces inside the country, according to U.S. officials who said that intelligence gaps have impeded efforts to support the ouster of Syrian President Bashar al-Assad."*[91]

Beginning with a series of pro-democracy protests in 2011, the situation in Syria quickly escalated into a full-blown civil war that has cost $250,000$ lives and displaced almost 11 million Syrian citizens to the beginning of 2016. In the backdrop of a ethnically and religiously divided population, this conflict quickly displayed a high degree of complexity in the heterogeneity of parties involved [Smith, 2012], including the Syrian state army loyal to Bashar al-Assad, Sunni Syrian rebels, the Islamic State, Jabhat al-Nusra, Kurdish forces, and Hezbollah. Lack of understanding of the structure of the insurgency in Syria has been one of the strongest deterrents to military and humanitarian involvement of Western powers in this conflict [Jenkins, 2014] and slowed down relief efforts.

Western countries were willing to lend support and provide prompt international aid to moderate Sunni organizations, but the difficulty laid in identifying these rebels and their true organizational linkages. The impossibility of separating the secular moderates from the religious extremists among the Sunni opponents of the Alawite-led government resulted in international paralysis. This led to further economic and social deterioration, radicalization, and escalation of the conflict. Syria is now a nearly failed state, fought over by Assad loyalists, the Islamic State, and the al-Qaeda affiliated Nusra front. Numerous attempts at a political solution by the Arab League and the United Nations have failed.

Another relevant case is Libya post-Colonel Gaddafi. This event would require in itself a fully accurate discussion, but as above for Iraq and Syria, we try to provide a basic picture from the perspective of the analysis of multi-group conflicts. After 2011 and the violent overthrowing of the Gaddafi regime, Libya gradually descended into full-blown

of insurgents rallying under one cause. It's multiple groups with different causes loosely tied together by the threads of anti-U.S. sentiment, some sort of Iraqi nationalism, Muslim-Arab unity or greed". The lack of familiarity with this type of enemy appeared evident: *"What makes it more difficult is that you're dealing with an insurgency without a single face"*.

[90] http://www.nytimes.com/2004/10/22/international/middleeast/22insurgents.html?pagewanted=2&_r=0

[91] http://www.washingtonpost.com/world/national-security/in-syria-conflict-us-struggles-to-fill-intelligence-gaps/2012/07/23/gJQAW8DG5W_story.html

factional violence with Islamic State factions jockeying for control of oil rich areas together with two main armed groups: the Tobruk government (elected democratically but in a deeply unstable political environment) and the Muslim Brotherhood-supported General National Congress. To further complicate the picture, other ethnic-based groups, like the Touareg, have also laid claim to certain parts of the former Libyan state. Repeated failures to achieve stable Unity governments and substantial ambiguity in the set of alliances struck among the various groups have severely hindered the pacification response led by the United Nations in the region. While the United Nations and the European Union have been holding off decisive intervention, the east/west divide in the country has been increasingly exacerbating.

M Covariance Matrix with differing values of σ^2

In the main text we assume that σ is constant for all districts, and we then normalize it to $\sigma^2 = 1$. Now suppose instead that some districts are easier to coordinate than others. Continue to assume that $\text{Var}(\epsilon_j) = 1$ for all groups j, but suppose that the signal to group j in district i is $\tilde{\epsilon}_{ij} = \tilde{\sigma}_i \epsilon_j$, where $\tilde{\sigma}_i$ is a district specific indicator of how much coordination will be occurring in this district. In this case we will have

$$(27) \quad \Gamma_L = \begin{bmatrix} \tilde{\sigma}_1\tilde{\sigma}_1 \sum_j \alpha_{1j}\alpha_{1j} & \tilde{\sigma}_1\tilde{\sigma}_2 \sum_j \alpha_{1j}\alpha_{2j} & & & \\ \tilde{\sigma}_2\tilde{\sigma}_1 \sum_j \alpha_{2j}\alpha_{1j} & \tilde{\sigma}_2\tilde{\sigma}_2 \sum_j \alpha_{2j}\alpha_{2j} & & & \\ ... & & \tilde{\sigma}_i\tilde{\sigma}_i \sum_j \alpha_{ij}\alpha_{ij} & & \\ \tilde{\sigma}_1\tilde{\sigma}_i \sum_j \alpha_{ij}\alpha_{1j} & & & ... & \tilde{\sigma}_i\tilde{\sigma}_{i'} \sum_j \alpha_{ij}\alpha_{i'j} \\ ... & & & & \end{bmatrix}$$

The transformation to a correlation matrix in this case will be

$$\Gamma_L^{\text{cor}} = D(\tilde{\sigma}_\cdot^2 \sum_j \alpha_{\cdot j}\alpha_{\cdot j})^{-1/2} \Gamma_L D(\tilde{\sigma}_\cdot^2 \sum_j \alpha_{\cdot j}\alpha_{\cdot j})^{-1/2},$$

where $D()$ indicates a diagonal matrix with the specified entries on the diagonal. The resulting Γ_L does not contain any $\tilde{\sigma}$ terms, and is thus identical to the Γ_L used in the main text. We thus see that district specific differences in coordination do not affect the analysis.

Now consider the case where σ differs across groups instead of across districts. That is, $\text{Var}(\epsilon_j) = \sigma_j$. In the case where groups do not overlap there is only one group per district, and thus the situation is identical to the above where $\tilde{\sigma}$ varied by district. In the case where groups do overlap, however, the transformation to Γ_L^{cor} would no longer eliminate the σ terms. Thus, if we assume that $\sigma^2 = 1$ for all groups when this is not in fact the case, our estimator for $\{\alpha_{ij}\}$ will be inconsistent. To see what will happen here, let $\tilde{\alpha}_{ij} = \sigma_j \alpha_{ij}$. The covariance matrix will have the form

$$(28) \qquad \Gamma_L = \sigma^2 \begin{bmatrix} \sum_j \tilde{\alpha}_{1j}\tilde{\alpha}_{1j} & \sum_j \tilde{\alpha}_{1j}\tilde{\alpha}_{2j} & & & \\ \sum_j \tilde{\alpha}_{2j}\tilde{\alpha}_{1j} & \sum_j \tilde{\alpha}_{2j}\tilde{\alpha}_{2j} & & & \\ ... & & \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{ij} & & \\ \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{1j} & & & ... & \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{i'j} \\ ... & & & & \end{bmatrix}$$

which is exactly the same as 3, except with $\tilde{\alpha}_{ij}$ replacing α_{ij}. Thus, our estimates $\hat{\alpha}_{ij}$ will be consistent for $\tilde{\alpha}_{ij}$. This would affect the estimates shown in Figures 6 and 8. If there is a group with low σ_j that thus launches almost no simultaneous attacks, this group would show up only in very light colours in these maps. This would not necessarily present a problem, since it would still be obvious where in the country such a group was operating. The only issue that would arise is that specific districts where there was overlap with other groups would seem to be dominated by those other groups, when the reality is that those other groups are simply engaging in more coordinated attacks.

If in reality σ differs based on pairs of districts, and so is actually $\sigma_{ii'}$, then the situation becomes more difficult. In the extreme case, insurgents in each district would coordinate with those in all adjacent districts but never with those that are further away. In this case, there is no plausible clustering of districts into groups, because each district exhibits the same similarity with all of its neighbours. The idea of clustering is that the underlying structure can be simplified into cluster memberships. In the extreme case this is ineffective, and thus our model is inappropriate.

A less extreme version of this would be that there is a group structure, but insurgents in the same group are more likely to coordinate with districts that are geographically close to them rather than districts that are further away. In this case, clustering the data could return meaningful results. The clustering algorithm would have to be carefully selected, however, to not incorrectly split a group just because there was some internal variation regarding which districts were coordinating with which other districts.

For example, suppose that districts are evenly spaced along a one dimensional line, and within the same insurgent group there will only be coordination between districts that are within a distance d of each other. In this case the covariance matrix does not consist of blocks as in Equation 4. Instead replacing each block will be a band, where the entries outside of the band are 0 because these district pairs, while in the same group, are too far away to coordinate. We thus have what we might call a diagonal matrix instead of a block diagonal matrix.

This situation would not be handled correctly by the approach we use in this paper, because we would incorrectly split a group based on the fact that it has this internal structure. It seems as though some sort of improved method should be able to cluster districts correctly here, because there is no correlation between districts in different groups but some positive correlation between at least some districts in the same group, and there are enough of these positive correlations to connect the entire group. One method that could potentially resolve this problem would be to use variant of correlation clustering [Bansal, Blum & Shuchi 2004]. We leave this for future work.

N Coding of attack vs. defence

A possible concern with the attack data we use is that, while classified as insurgent attacks, these incidents are actually in response to government actions. Thus, any correlation we discover between districts would not be indicative of the structure of insurgent groups, but rather the organization of the counter-insurgency.

There are two situations that are of particular concern. First, there is the danger that a police attack on an insurgent stronghold might be included in our data as an attack simply because the insurgents shoot back. Second, even if our data only includes incidents initiated by the insurgents in a tactical sense, these incidents may be initiated by the government in a strategic sense. For example, suppose that a mountainous area is known to be insurgent controlled, and the government wants to change this. It might send several patrols deep into the mountains. Insurgents that happen to be present in the area might then attack these patrols as targets of opportunity. These attacks could then show up in our dataset as simultaneous attacks, but this would be evidence of coordination by the government, rather than by the insurgents.

The easiest dataset to use to consider these issues is the Global Terrorism Database (GTD), which has much more detailed coding of events than either WITS or BFRS. The GTD has a smaller number of incidents overall, which is why we do not use it as our main datasource, but as shown in Section , this dataset gives effectively the same results, albeit with a smaller number of districts. Thus, if we can show that the above problems do not occur with the GTD, this suggests that they are not responsible for the results we report in the paper.

The GTD only includes incidents where non-state actors are the attackers. Thus, it specifically excludes incidents such as police raids. This can be seen in the dataset because a small number of incidents (about 0.1%) are coded as doubtful because the attack could been by a state actor. In a few of these, the additional notes explicitly give as the reason that the police may have fired first. The other 99.9% of attacks are not believed to be initiated by government forces, and thus simultaneous government attacks are not contaminating the data.

The second possibility is that a strategic decision by the government leads naturally to simultaneous attacks by the insurgents without any insurgent planning can also be checked using notes that accompany the GTD entries. Attacks on government forces could occur when these forces are on patrol, or they could occur when the government forces are stationary. If the forces are on patrol, it could be that they have entered an insurgent held area, and it is obvious that if many patrols simultaneously enter then they will be simultaneously attacked. On the other hand, if the forces are stationary, then there is no particular reason for the insurgents to naturally attack these forces simultaneously, unless there is coordination on the part of the insurgents. A police checkpoint, for example, could be attacked today, but could equally well be attacked tomorrow, and thus, beyond random chance, the simultaneous attacks that do occur would be due to insurgent coordination.

The question thus becomes whether insurgents strike mainly when government forces are on patrol, or when they appear to be stationary. In the GTD data, there are a total of 124 sets of simultaneous attacks listed for Afghanistan. In the summary description of these attacks, "patrol" occurs in descriptions in 4 sets of attacks, "checkpoint" appears in descriptions in 25 sets of attacks, and "post" or "checkpost" appears in descriptions in 31 sets of attacks. A qualitative examination of the descriptions suggests that many of the remaining attacks are aimed at targets that would best be described as "stationary" (e.g. police chiefs, embassies, towns). It thus appears that insurgents mainly attack government forces when they are stationary. This strongly suggests that government strategic decisions do not determine the precise day when the insurgents will attack, and thus the observed simultaneity really is due to insurgent coordination, rather than being a mechanical product of the strategy of government forces.

REFERENCES

[1] Anderson, Carl A. (1974) "Portuguese Africa: A Brief History of United Nations Involvement" Denver Journal of International Law & Policy 133

[2] Anderson, T.W. (1963) "Asymptotic Theory for Principal Component Analysis" Annals of Mathematical Statistics 122-148.

[3] Baurle, G. (2013) "Structural Dynamic Factor Analysis Using Prior Information From Macroeconomic Theory " Journal of Business & Economic Statistics. 31(2):136-150.

[4] Choi, W.G.; Kang, T.; Kim, G.Y.; Lee, B. (2014) "Global Liquidity Transmission to Emerging Market Economies, and Their Policy Responses" . SSRN Scholarly Paper 2580627. December 2014.

[5] Eisenstadt, Michael and Jeffrey White (2005) "Assessing Iraq's Sunni Arab Insurgency" The Washington Institute for Near East Policy Policy Focus No.50.

[6] Fernandes, Clinton (2008) Hot Spot: Asia and Oceania. ABC-CLIO

[7] Freytag, Andres, Jens J. Kruer, Daniel Meierrieks, Friedrich Schneider. (2011). The origins of terrorism: Cross-country estimates of socio-economic determinants of terrorism. European Journal of Political Economy 27 S516

[8] Good, Phillip. (2002) Extensions of the Concept of Exchangeability and their Applications. Journal of Modern Applied Statistical Methods. 1(2) 243-247.

[9] Good, Phillip. (2005) Permutation, Parametric, and Bootstrap Tests of Hypotheses. New York: Springer.

[10] Gotelli Nicholas J., and Chao Anne (2013) Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin S.A. (ed.) Encyclopedia of Biodiversity, second edition, Volume 5, pp. 195-211. Waltham, MA: Academic Press..

[11] Hastie, T., Tibshirani, R., and Friedman, J. (2001). The elements of statistical learning. New York: Springer.

[12] Henderson, Anne. (2005) The Coalition Provisional Authority's Experience: with Economic Reconstruction in Iraq: Lessons Identified. USIP Special Report No. 138. http://www.usip.org/files/resources/sr138.pdf

[13] Ledermann, W. (1940). "On a Problem concerning Matrices with Variable Diagonal Elements." Proceedings of the Royal Society of Edinburgh. 60(1). 1–17.

[14] Leites, Nathan and Charles Wolf. (1970). Rebellion and Authority. Chicago, IL: Markham.

[15] Ng, A. Y., Jordan, M., &Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), Advances in neural information processing systems, 14. Cambridge, MA: MIT Press.

[16] Pesarin, Fortunato (2001) Multivariate Permutation Tests, New York: Wiley.

[17] Saunderson, J.; Chandrasekaran, V.; Parrilo, P.; Willsky, A. (2012). "Diagonal and Low-Rank Matrix Decompositions, Correlation Matrices, and Ellipsoid Fitting." SIAM Journal on Matrix Analysis and Applications. 33(4): 1395–1416.

[18] Shi, J. and Malik, J. (2000). "Normalized cuts and image segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 888 – 905.

[19] Smith, B. (2012). "Syria: No End in Sight?" House of Commons Library. Research Paper 12/48.

[20] Subrahmanian, V. S., Aaron Mannes, Animesh Roul, R. K. Raghavan (2013) Indian Mujahideen: Computational Analysis and Public Policy, Springer.

[21] Tibshirani, R., Walther, G., and Hastie, T. (2001) "Estimating the number of clusters in a data set via the gap statistic". J. R. Statist. Soc. B, 63, Part 2, 411-423.

[22] Wu, Y.; Moon, H.R.; Deng, Y. (2011) "Factor Analysis on US Housing Price Indexes." USC Lusk Center Working Paper.

[23] Yao, J., Zheng, S., and Bai, Z. (2015) Large Sample Covariance Matrices and High-Dimensional Data Analysis, Cambridge University Press..