Econometrica Supplementary Material

SUPPLEMENT TO "INSURGENCY AND SMALL WARS: ESTIMATION OF UNOBSERVED COALITION STRUCTURES"

(*Econometrica*, Vol. 87, No. 2, March 2019, 463–496)

FRANCESCO TREBBI
Vancouver School of Economics, University of British Columbia, CIFAR, and NBER

ERIC WEESE
Faculty of Economics, Kobe University

APPENDIX A: INSURGENCY ORGANIZATION AND 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) described 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) observed.[1]

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.[2] 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

Francesco Trebbi: francesco.trebbi@ubc.ca
Eric Weese: eric.weese@aya.yale.edu

[1]Henderson was 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.

[2]Eisenstadt and White (2005) wrote 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.

 https://doi.org/10.3982/ECTA14436

it was clear that one existed.[3] Further complexity in the Iraqi case stemmed from signs of evolution over time, as the New York Times reported: "the insurgency was now organized regionally, and that evidence pointed to some planning across regional boundaries" (Schmitt and Shanker (2004)).

The difficulty, and the importance, of understanding the structure of insurgencies is not limited to Iraq. Consider newspaper reports on 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. (Miller and Warrick (2012))

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 an 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 and Michael (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 lay 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 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.

[3]The New York Times quoted senior U.S. intelligence sources stating that

> It's not just one group 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. (Schmitt and Shanker (2004))

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" (Schmitt and Shanker (2004)).

APPENDIX B: SIMULTANEOUS ATTACKS: THEORETICAL FRAMEWORK AND QUALITATIVE SOURCES

An insurgent group typically operates from an asymmetric position and does not usually aim for military victory over its adversary (Kilcullen (2009)). Baloch separatists need only convince the Pakistani government to allow an independent Balochistan, not necessarily topple the government. A group that appears strong, however, will have greater negotiating power vis-à-vis its opponent. It will also have more success in recruitment and fundraising, as the noncombatant population is more likely to side with a strong group. Launching simultaneous attacks is a signal of strength, because such an attack requires coordination.

The basic idea for our theoretical framework is provided by Shapiro (2013): insurgent groups face a trade-off between their degree of internal control and the safety of their members, because the mere act of communicating makes members more vulnerable to detection by government forces. Suppose that some particularly effective insurgent groups have managed to develop and maintain secure communication channels, while other groups are plagued by government moles and eavesdroppers. Insurgents benefit from the support of the civilian population for recruitment and fundraising, and civilians are more interested in supporting well-organized and effective groups than failing ones. In cases such as Afghanistan, insurgents also benefit from convincing foreign civilians of their strength, as these foreign civilians then pressure their governments to withdraw troops from an "unwinnable" conflict. Civilians do not know exactly how strong the insurgents are, and insurgents thus wish to somehow signal that their organization is strong and uncompromised, both to win local support and to force foreign withdrawal.

A simultaneous attack necessarily involves communication in order to coordinate the attack.[4] If a weak insurgent group is vulnerable to government surveillance when it attempts to communicate, while a strong group has successfully developed communication methods that escape detection, then a simultaneous attack is costlier for the weak group due to the exposure of its members. Simultaneous attacks thus fit into the standard Spence (1973) signaling framework: such an attack is a credible signal of strength because launching it is less costly for the strong group than the weak group.[5]

The qualitative literature supports the idea that simultaneous attacks have a signaling motivation. For example, Barno (2006) gave a specific example of a simultaneous attack on three border checkpoints where the media were deliberately alerted to the attack and publicity appears to have been the main objective. Deloughery (2013) provided a recent review of the literature and presented systematic evidence of the advantages of simultaneous attacks for terrorist organizations in terms of psychological warfare, media coverage, and appeal in the recruitment of new fighters, incentives that operate within insurgencies as well.[6]

In reality, insurgent groups launch a mix of simultaneous and individual stand-alone attacks. We posit that this is because there is a trade-off between the signaling value

[4]Shapiro and Siegel (2015) discussed how insurgent coordination is achieved through mobile phone communication and ICT.

[5]The cost of carrying out coordinated attacks may be lower for stronger groups through reasonable mechanisms other than a security rationale. Carrying out attacks often requires local knowledge and stronger groups may have broader recruitment networks to find operatives with the right skill set to carry out a coordinated attack, etc.

[6]According to Kilcullen (2009), "the insurgents treat propaganda as their main effort, coordinating physical attacks in support of a sophisticated propaganda campaign" (p. 58). See also Arce and Sandler (2007).

of attacking simultaneously in many districts versus the military value of attacking separately in each district at the most opportune moment for that district. In Appendix N, we discuss this hypothesis further and support it with regression evidence. We also check implications of the signaling model just outlined above: for example, it appears that (both in Afghanistan and in a cross-country sample) insurgents are less likely to launch simultaneous attacks relative to stand-alone attacks in areas where they have a limited presence and are thus potentially more vulnerable.

Our main objective in the paper is to use the fact that insurgent groups do launch simultaneous attacks in order to identify the number of such groups and their geographic extent. We do not formalize the above signaling model of attacks—the framework is standard. Instead, in Section 2, we build an econometric model of simultaneous attacks based on the assumption that all groups launch such attacks to at least some degree.[7]

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, p. 258)). The Indian Mujahideen, responsible for the 2008 Mumbai attacks, typically carry out simultaneous attacks (Subrahmanian, Mannes, Roul, and Raghavan (2013, Chapter 6)). 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) described 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.

APPENDIX 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

$$\min_{\hat{\Gamma}_D} \operatorname{Tr}(\hat{\Gamma}_L)$$

$$\text{s.t.} \quad \hat{\Gamma}_L = \bar{\Gamma} - \hat{\Gamma}_D, \quad \hat{\Gamma}_D \text{ diagonal}, \tag{C.1}$$

[7]In the model presented below, there are disorganized individual insurgents who attack randomly, and thus even a particularly weak insurgent group would have an incentive to launch the occasional simultaneous attack, in order to distinguish themselves from these "lone wolf" actors.

$$\hat{\Gamma}_D \succ 0, \qquad \hat{\Gamma}_L \succ 0.$$

Here, $\mathrm{Tr}(\cdot)$ 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.

Saunderson, Chandrasekaran, Parrilo, and Willsky (2012) showed 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 C.1.

If $N = 200$, the semi-definite program corresponding to (C.1) 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 semi-definite programming algorithms. We thus compute $\hat{\Gamma}_L$ using (C.1), and will use it as the basis for producing an estimate of insurgent group presence in the next two subsections.

C.1. *Recoverability of Low-Rank Matrix*

We are interested in the conditions under which the $\hat{\Gamma}_L$ resulting from (C.1) 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:

EXAMPLE C.1: 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 (C.1) 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) gave such a characterization. First, Saunderson et al. (2012) defined 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 (C.1) applied to $\Gamma = \Gamma_D + \Gamma_L$ returns $\hat{\Gamma}_L = \Gamma_L$. Next, they defined the "coherence" $\mu(\mathcal{U})$ of a subspace $\mathcal{U}$ of $\mathbb{R}^n$ as

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

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

THEOREM C.1—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 (C.1) 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 provided a further result, regarding the "realizability of random subspaces." They argued 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 (C.1) 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.

APPENDIX D: SPECTRAL CLUSTERING ESTIMATOR

Spectral clustering is based on the "graph Laplacian" matrix

$$L = D - \Gamma_L, \tag{D.1}$$

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.[8] Spectral clustering focuses 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 give 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

[8]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 1. 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 1, and thus the null space of the resulting block of L must be rank 1. 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.

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 provide 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 $\hat{J}$.

A similar problem affects the approach presented in the main text, which is based on using 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 (C.1) 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 labeling some eigenvalues as "zero" eigenvalues, despite the fact that, due to random noise, all eigenvalues will probably be nonzero.[9] 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.[10] The difference $\lambda_{k+1} - \lambda_k$ is defined as the kth eigengap. Ng, Jordan, and Weiss (2002) argued that a large eigengap indicates that perturbation of the eigenvectors of L would not change the clusters produced by spectral clustering. von Luxburg (2007) thus suggested 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.[11] The intuition here is that if there truly are $\hat{J}$ eigenvalues that are zero, then these appear to be nonzero 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.

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. von Luxburg

[9]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.

[10]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 three 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.

[11]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.

(2007) suggested 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 converge 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.[12]

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.

APPENDIX E: 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 $\operatorname{Var}(\epsilon_j) = 1$ 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

$$\Gamma_L = \begin{bmatrix} \tilde{\sigma}_1\tilde{\sigma}_1\sum_j \alpha_{1j}\alpha_{1j} & \tilde{\sigma}_1\tilde{\sigma}_2\sum_j \alpha_{1j}\alpha_{2j} & & & \\ \tilde{\sigma}_2\tilde{\sigma}_1\sum_j \alpha_{2j}\alpha_{1j} & \tilde{\sigma}_2\tilde{\sigma}_2\sum_j \alpha_{2j}\alpha_{2j} & & & \\ \dots & & \tilde{\sigma}_i\tilde{\sigma}_i\sum_j \alpha_{ij}\alpha_{ij} & & \\ \tilde{\sigma}_1\tilde{\sigma}_i\sum_j \alpha_{ij}\alpha_{1j} & & & \dots & \tilde{\sigma}_i\tilde{\sigma}_{i'}\sum_j \alpha_{ij}\alpha_{i'j} \\ \dots & & & & \end{bmatrix}. \tag{E.1}$$

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

The transformation to a correlation matrix in this case will be

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

where $D(\cdot)$ 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

$$\Gamma_L = \sigma^2 \begin{bmatrix} \sum_j \tilde{\alpha}_{1j}\tilde{\alpha}_{1j} & \sum_j \tilde{\alpha}_{1j}\tilde{\alpha}_{2j} & & & \\ \sum_j \tilde{\alpha}_{2j}\tilde{\alpha}_{1j} & \sum_j \tilde{\alpha}_{2j}\tilde{\alpha}_{2j} & & & \\ \ldots & & \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{ij} & & \\ \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{1j} & & & \ldots & \sum_j \tilde{\alpha}_{ij}\tilde{\alpha}_{i'j} \\ \ldots & & & & \end{bmatrix}, \tag{E.2}$$

which is exactly the same as (2.1), 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 colors 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 neighbors. 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 (2.2). 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, and Chawla (2004)). We leave this for future work.

APPENDIX F: CLUSTERING DETAILS AND ESTIMATE FOR α

As a first step, 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 $\hat{\Gamma}_L$.

For many k-means algorithms, however, 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'.[13] The cosine distance between two districts with the same group present will be zero asymptotically, while it will be 1 when the districts have different groups present.

For the particular data that we will be considering, 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) provided 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 substantially change any of the results discussed below. Suppose 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 $\bar{\Gamma}$.[14]

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

[13]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.

[14]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.

it is reasonable to use the approximation

$$\sum_{i'\neq i}\alpha_{ij}\alpha_{i'j}\simeq\sum_{i'}\alpha_{ij}\alpha_{i'j}$$
$$=\alpha_{ij}\sum_{i'}\alpha_{i'j}$$
$$=\alpha_{ij}a_j, \tag{F.1}$$

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.[15]

APPENDIX 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 $i\in\{1,\dots,25\}$, 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

$$x_{it}=\sum_j\alpha_{ij}\epsilon_{tj}+u_{it}, \tag{G.1}$$

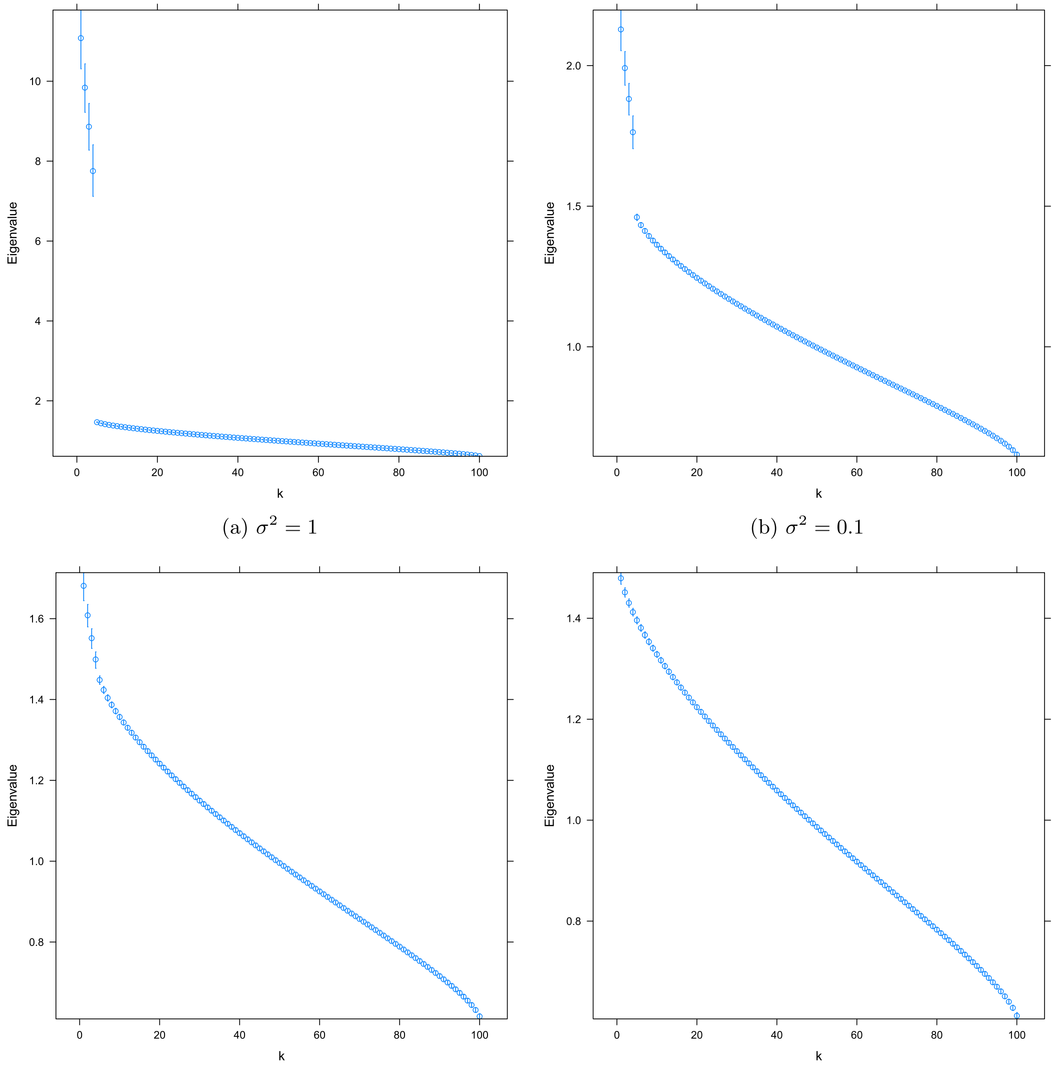
where $u_{it}\sim N(0,1)$, i.i.d.

We then consider eigenvalues associated with the (N by N) covariance matrix of attacks. We perform 100,000 simulations for each of $\sigma^2=1$, $\sigma^2=0.1$, $\sigma^2=0.05$, and $\sigma^2=0$, generating a total of 400,000 simulated sample covariance matrices.[16]

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.1(d) shows the distribution of eigenvalues under $J=0$.

[15]While it would be possible to use nonlinear 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 nontrivial undertaking.

[16]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 distributions 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.

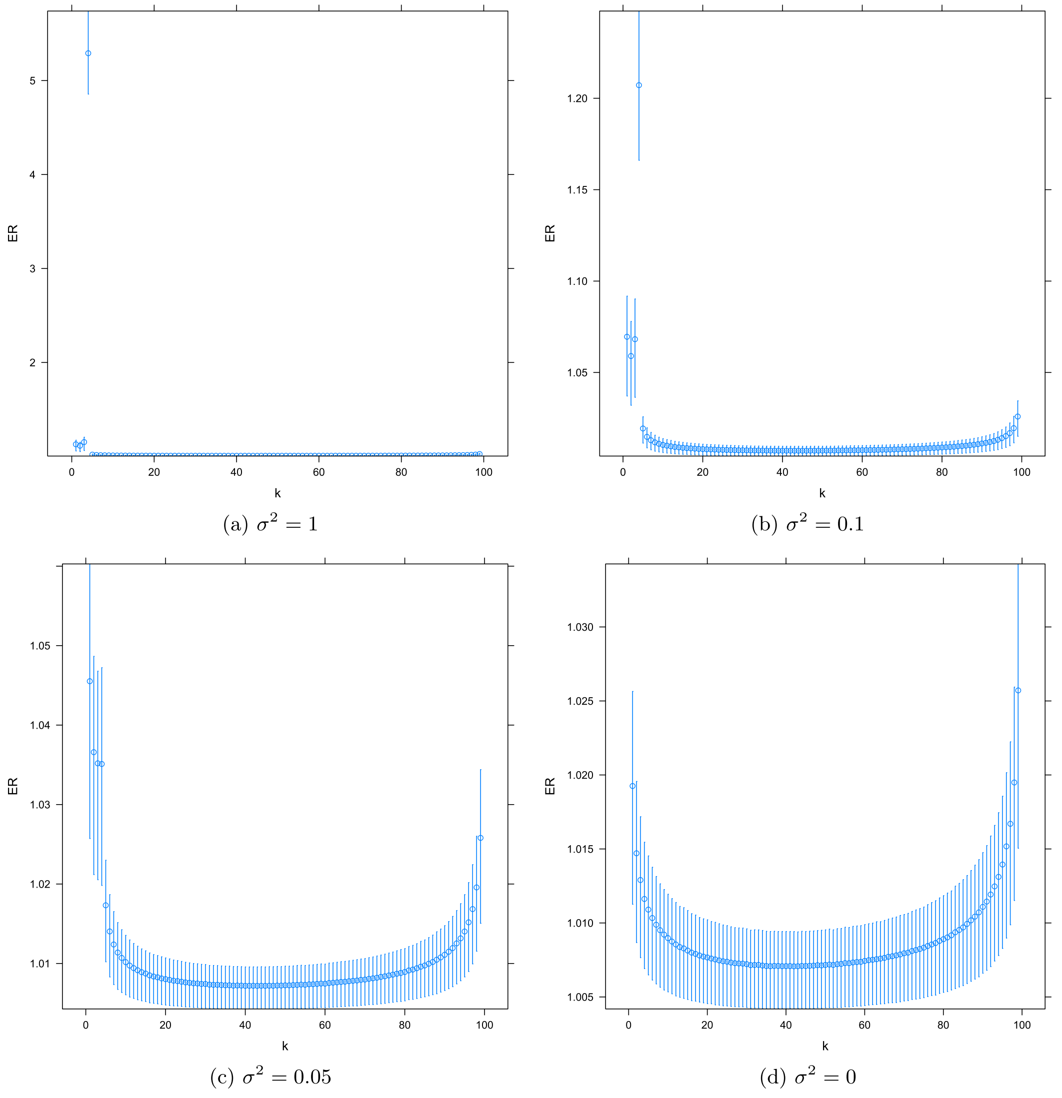


(a) $\sigma^2 = 1$ (b) $\sigma^2 = 0.1$

(c) $\sigma^2 = 0.05$ (d) $\sigma^2 = 0$

FIGURE G.1.—Eigenvalues. Points indicate means over 100,000 simulations. Bars show interquartile range.

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.2(d) 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.2(d) thus illustrates why it is important to have some maximum number of possible groups, $J_{\max}$. The eigenratios associated with the very smallest eigenvalues (towards the right-hand side of Figure G.1(d)) 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



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 0.1$

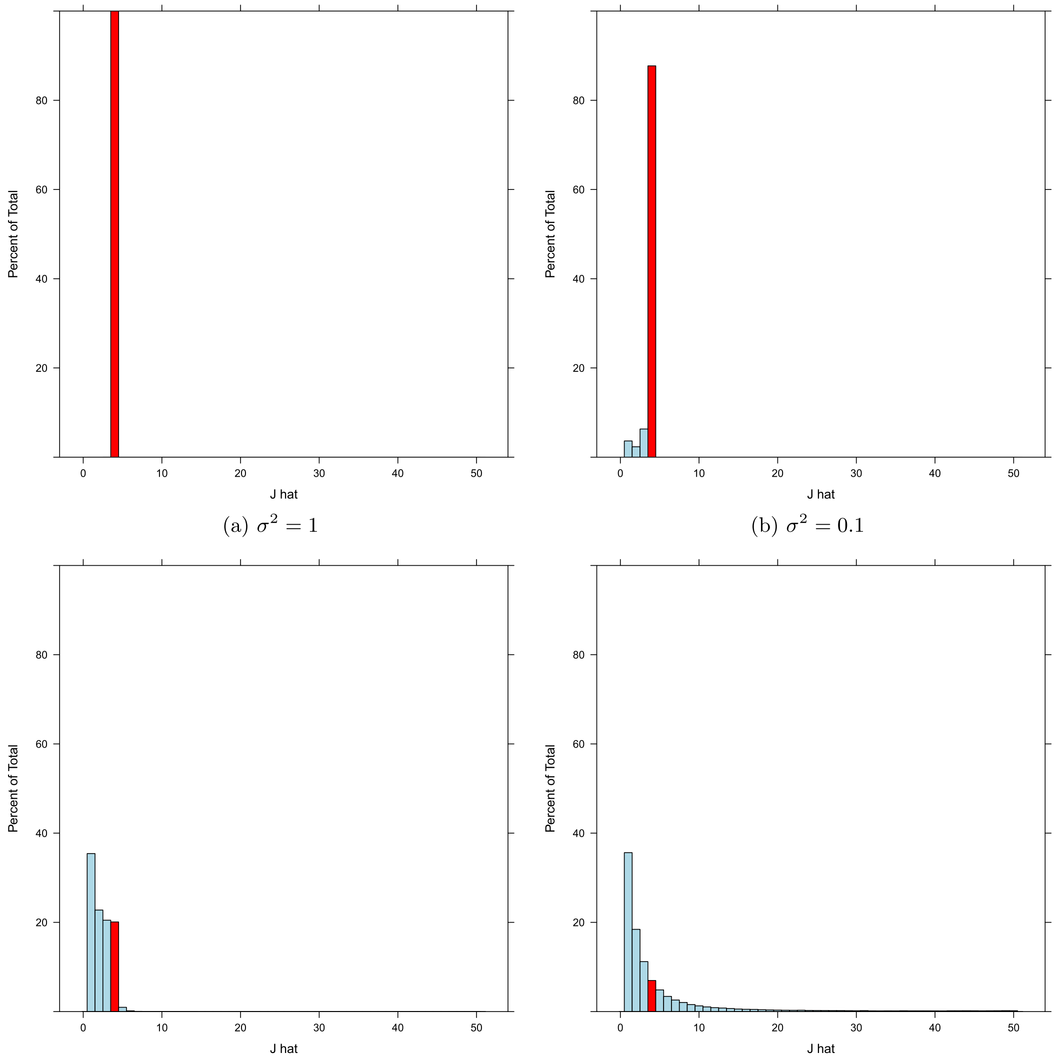
(c) $\sigma^2 = 0.05$ (d) $\sigma^2 = 0$

FIGURE G.2.—Eigenratios. Points indicate means over 100,000 simulations. Bars show interquartile range.

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

In both of Figures G.3(c) and G.3(d), $\hat{J} = 1$ is the modal estimate. Figure G.3(d) 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.3(d) shows the case with no group structure at all, and thus would not change regardless of the true value of J. The bias of



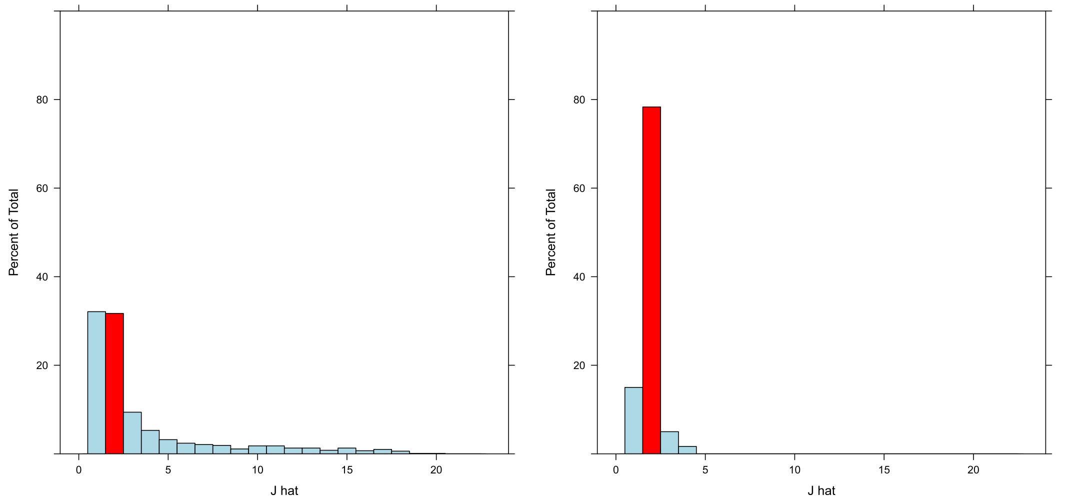
(a) $\sigma^2 = 1$ (b) $\sigma^2 = 0.1$

(c) $\sigma^2 = 0.05$ (d) $\sigma^2 = 0$

FIGURE G.3.—Estimated number of groups ($\hat{J}$). Histograms of estimated number of groups, over 100,000 simulations. True value $J = 4$ shown in red.

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 nontrivial.

Figure G.3(c) provides a possible explanation for why estimates of $\hat{J} = 1$ appear so frequently in Table V. 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.1(d). The evidence provided in Table V should thus mainly be taken as an indication that the null hypothesis of no group structure should be rejected. Figure G.3(c)



(a) Eigenratio (b) Hierarchical Splits

FIGURE G.4.—Estimated number of groups ($\hat{J}$). Histograms of estimated number of groups, over 100 simulations. True value $J = 2$ shown in red.

shows how estimates $\hat{J} = 1$ occur frequently when there is actually a group structure with $J > 1$.[17]

G.1. *Comparison With Hierarchical Splits*

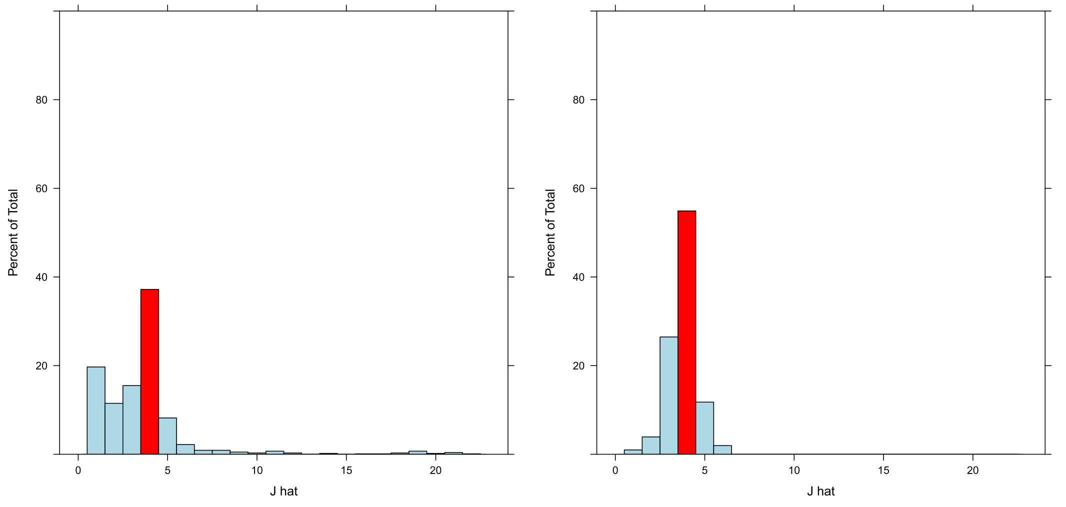
To compare our eigenratio type estimator with the estimator based on hierarchical splits, we need to simulate data with discrete attacks, as the permutation test used requires integer numbers of attacks to permute. Let the number of attacks by group j in district i at time t be drawn from a Poisson(λ_{ijt}) 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 nonzero α_{ij} entries drawn from a Uniform$(0, 0.25)$ distribution, as well as $J = 4$ with the nonzero 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 hierarchical 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 H: NNMF CONSISTENCY

Conditions under which $\hat{\Gamma}_L$ will converge to Γ_L have been discussed in Appendix C.1. We now consider conditions under which a nonnegative matrix factorization of Γ_L will

[17] In the empirical literature, "low" estimates for the number of factors (compared to other methods) were obtained in Choi, Kang, Kim, and Lee (2017) and Wu (2012, Chapter 2). Figures G.3(b) and G.3(c) appear in line with results reported (using actual data) in the supplement to Baurle (2013).



(a) Eigenratio (b) Hierarchical Splits

FIGURE G.5.—Estimated number of groups ($\hat{J}$). Histograms of estimated number of groups, over 100 simulations. True value $J = 4$ shown in red.

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 relabeling 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) discussed uniqueness of symmetric nonnegative factorizations at some length. They concluded 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% nonzero entries are still reconstructed successfully. The Γ_L matrices considered in this paper would generally be expected to have a relatively sparse factorization.

APPENDIX I: 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 are weakly exchangeable: within a given district, permuting the time indices does not change the joint distribution of the attacks.[18] 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 exhibit 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.

[18]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.

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 are 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.[19] This might be the case, for example, if there were particular days that, for whatever reason, generated large amounts of random 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.[20] 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.

I.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

[19]This gives the disorganized militants the same structure as 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.

[20]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,

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

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) provided an accessible introduction to permutation tests.

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 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 are equal to 1 if there were r attacks on day t in district i, and equal to zero otherwise. A valid permutation will satisfy the constraints

$$\sum_r x_{ti}^r = 1, \quad \forall t, i, \tag{I.1}$$

$$\sum_{t=1}^T x_{ti}^r = \sum_{t=1}^T x_{ti}^{r,\text{actual}}, \quad \forall i, r, \tag{I.2}$$

$$\sum_{i=1}^N \sum_r r x_{ti}^r = \sum_{i=1}^N \sum_r r x_{ti}^{r,\text{actual}}, \quad \forall t, \tag{I.3}$$

$$\begin{aligned}\sum_{t=1}^T \Big(\sum_r r x_{ti}^r\Big)\Big(\sum_r \sum_{i=1}^N r x_{ti}^{r,\text{actual}}\Big) \\ = \sum_{t=1}^T \Big(\sum_r r x_{ti}^{r,\text{actual}}\Big)\Big(\sum_r \sum_{i=1}^N r x_{ti}^{r,\text{actual}}\Big), \quad \forall i,\end{aligned} \tag{I.4}$$

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 covariance 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.[21] 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.

[21]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.

TABLE I.I
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

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

APPENDIX J: 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

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

where $\mathrm{SCov}(p_{it}, p_{i't})$ 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

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

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

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

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.

APPENDIX K: CODING OF ATTACK VERSUS DEFENSE

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 counterinsurgency.

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 include 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 data set as simultaneous attacks, but this would be evidence of coordination by the government, rather than by the insurgents.

The easiest data set 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 data source, but as shown in Section 5, this data set 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 data set because a small number of incidents (about 0.1%) are coded as doubtful because the attack could have 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, 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.

APPENDIX L: SINDHUDESH LIBERATION ARMY

The first recorded attack in the GTD under the SDLA banner is recorded on November 2nd, 2010 when incident (ID number `201011020003`) states:

> On Tuesday, in Hyderabad, Sindh, Pakistan, a portion of rail track was damaged when unidentified militants detonated an improvised explosive device, wounding four people. Another bomb was found and defused be security forces at the scene. A two-page pamphlet issued by Sindhu Desh Liberation Army (SDLA) 'chief commander' Darya Khan was found on the spot. The pamphlet enlisted 19 points, mentioning issues of Sindh and targeting what it called Punjabi imperialism.[22]

When reading the entries for the GTD simultaneous incidents of Karachi (`201102110009`) and Hyderabad (`201102110005`) on February 11th, 2011, which are explicitly listed as *not being part of multiple incidents*, the GTD appears uninformed by these events. Consider `201102110005` notwithstanding explicit claiming by the SDLA (possibly discarded as not credible):

> On Friday morning, in Bengali Colony of Hussainabad in Hyderabad, Sindh, Pakistan, unidentified militants blew up railway tracks, causing no casualties but damaging the tracks. A few pamphlets were found at the blast sites carrying the name of an unknown group, Sindhu Desh Liberation Army (SDLA).

Subsequently, the GTD identifies two attacks in Sindh in the month of November 2011. None of these attacks is again classified as part of a multiple incidents event (i.e., coordinated attacks). It is only on February 25th, 2012, about a year after our methodology singles out SDLA activity in Sindh, that coordination of SDLA is finally detected in the GTD, with multiple entries (12 entries, listed explicitly as being part of multiple incidents).[23]

The BFRS data mention in their comment section the SDLA only on 4 of the 41 attacks taking place on the month of November 2011 in Sindh. Of these 41 attacks, we can observe that only 10 are isolated incidents, while 31 attacks occur in bundles of 2 in a day (5 multiple incidents) or 3 attacks in a day (7 multiple incidents). By this time, our methodology is picking up SDLA coordinated activity since early 2011, information clearly missed both in the GTD and in the BFRS.

[22]The next day the GTD records incident `201011030021`:

> On Wednesday, near Nawabshah, Sindh, Pakistan, unknown assailants detonated an improvised explosive device on the Karachi–Lahore railroad. The blast damaged an eight-inch long portion of up-track and caused rail traffic to be suspended for over three hours. A bomb disposal squad later discovered a second bomb and successfully defused it. No casualties were reported. Sindhu Desh Liberation Army (SDLA), has claimed the responsibility for the blasts. The organization's purported chief commander, Darya Khan, has threatened that it would continue to carry out such attacks in future in order to get their "right to liberation" recognized by the United Nations.

[23]Incident `201202250003`, on February 25th, 2012, states that

> Explosives planted along railway tracks detonated in Jamshoro district, Sindh province, Pakistan. The tracks were damaged, but there were no human casualties. This was one of 12 explosive devices planted on railroad tracks in Sindh province on February 25, 2012. Sindhu Desh Liberation Army (SDLA) claimed responsibility, stating that people were fighting nationally and internationally for Baloch independence.

In addition to the one above, incident GTD ID's are all those listed `201202250012`–`201202250022`. On May 2nd 2012, the SDLA followed suit with 21 coordinated bomb attacks on the same day on banks around Sindh province. In the month of May 2012, SLDA activity caused 9 deceased and 30 wounded victims.

APPENDIX M: ADDITIONAL FIGURES AND TABLES FOR SECTIONS 5 AND 6

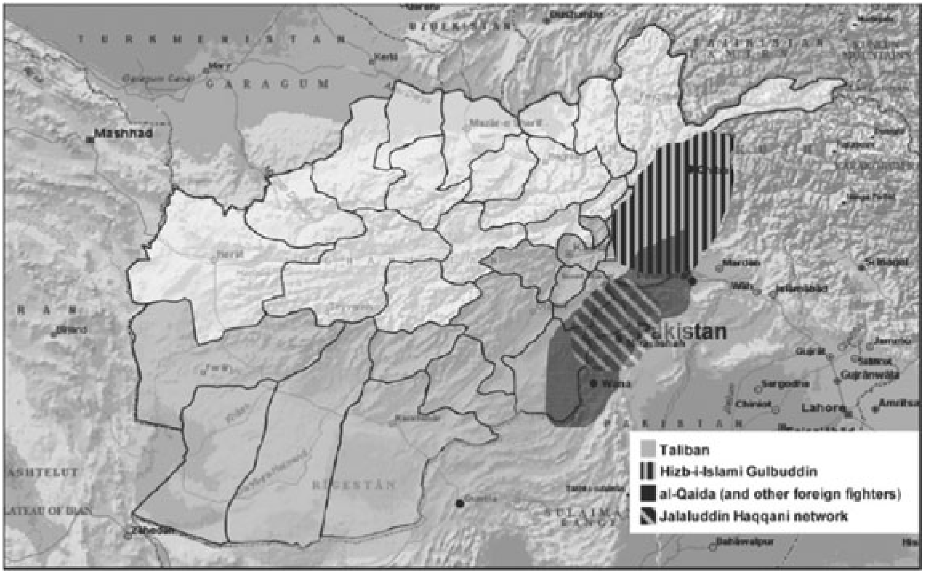


FIGURE M.1.—From Jones (2008). Figure used with permission. Seth G. Jones, 'The Rise of Afghanistan's Insurgency: State Failure and Jihad', International Security, 32:4 (Spring, 2008), pp. 7–40. ©2008 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology. Figure 1.

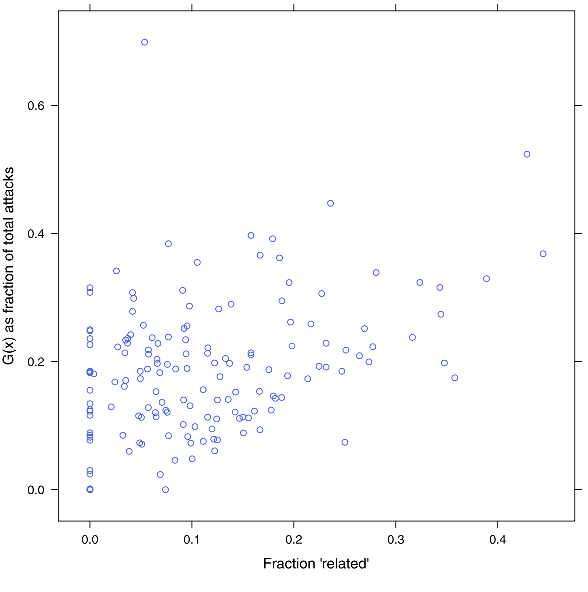


FIGURE M.2.—Overdispersion and "related" attacks.

TABLE M.I

ESTIMATION OF $\hat{J}$ BASED ON HIERARCHICAL SPLITS[a]

		Pakistan (to Apr '11)
Split at (1)?	Randomly shuffled data (mean)	138.49
	Std. dev.	7.82
	Actual data	159.00
	p-value	0.01
Split at (2)?	Randomly shuffled data (mean)	45.82
	Std. dev.	4.69
	Actual data	58.00
	p-value	0.01
Split at (3)?	Randomly shuffled data (mean)	32.36
	Std. dev.	4.09
	Actual data	50.00
	p-value	0.00
Split at (4)?	Randomly shuffled data (mean)	11.00
	Std. dev.	2.24
	Actual data	13.00
	p-value	0.24
Split at (5)?	Randomly shuffled data (mean)	16.69
	Std. dev.	2.76
	Actual data	21.00
	p-value	0.08
Split at (6)?	Randomly shuffled data (mean)	13.12
	Std. dev.	2.45
	Actual data	15.00
	p-value	0.27
Split at (7)?	Randomly shuffled data (mean)	11.60
	Std. dev.	2.42
	Actual data	13.00
	p-value	0.34

[a]A test statistic Q is computed as described in Section 2.3, based on a within-month covariance matrix as described in Section 2.5. Figure 5 shows the order of the potential splits. Data used are the Pakistan BFRS data set for May 2008–April 2011. This is 6 months less data than is used in Table I, which uses data through to October 2011.

TABLE M.II

"RELATED" ATTACKS AND OVERDISPERSION[a]

	I	II	III	IV
(Intercept)	0.01	0.02		
	(0.00)	(0.02)		
Overdispersion	0.35	0.35	0.35	0.36
	(0.03)	(0.03)	(0.03)	(0.03)
FKMS Controls		Yes	Yes	Yes
Country FE			Yes	Yes
Year FE				Yes
N	2006	2005	2005	2005

[a]Robust standard errors in parentheses. 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).

TABLE M.III

DEPENDENT VARIABLE IS HERFINDAHL FRAGMENTATION OF TERRORIST GROUPS[a]

	I	II	III	IV
(Intercept)	0.52	0.06		
	(0.01)	(0.08)		
Overdispersion	−0.27	−0.32	−0.27	−0.27
	(0.04)	(0.03)	(0.03)	(0.03)
Max Possible Overdispersion		0.51	0.51	0.46
		(0.05)	(0.06)	(0.06)
FKMS Controls		Yes	Yes	Yes
Country FE			Yes	Yes
Year FE				Yes
N	1144	1143	1143	1143

[a] Robust standard errors in parentheses. 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, Kruer, Meierrieks, and Schneider (2011). (We thank Daniel Meierrieks for providing the Freytag et al. (2011) data set.)

APPENDIX N: 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.[24] We examine these 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. As discussed in Section 5.2, our method appears to pick up an organized group operating across Sindh almost a year before this would have been visible by examining the group identification in the best available data sets. 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 are not available for more recent years, so we use

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

data from the GTD for this analysis. We use data from the Nov. 2011–Dec. 2016 period, and run a clustering of these data into four groups.[25] The resulting group structure is shown in Figure N.1, and the underlying covariance matrix is shown in Figure N.2. 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 N.I. 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 include fewer attacks than the BFRS data, and have very few incidents in Punjab. We thus do not see any group corresponding to the Punjabi ethnicity, which is the main difference between Figures 7 and N.1. Overall, however, the data show 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. Ninety-six percent 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.[26]

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 groups by qualitative analysts.

The GTD is again useful here, because it reports group identities where available. Groups in the GTD are defined based on perpetrator information, where "the perpe-

[25]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 have 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 is unfortunately not possible to track changes in the group structure in Pakistan across time.

[26]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) cited a Taliban commander describing how simultaneous attacks prevent a concentration of security forces.

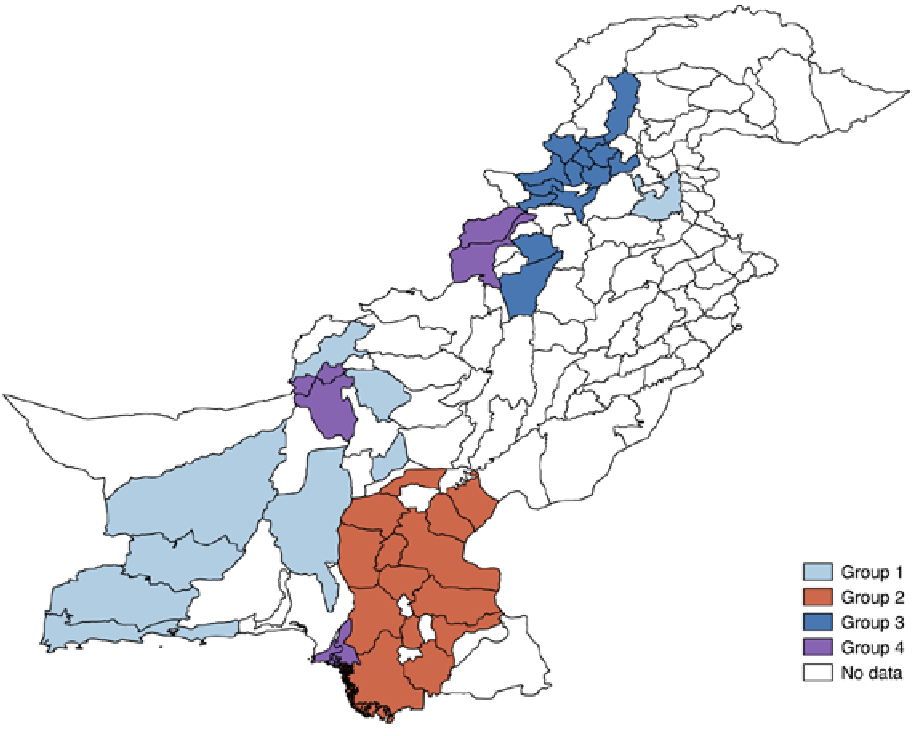


FIGURE N.1.—Pakistan groups with post-Nov 2011 GTDB data.

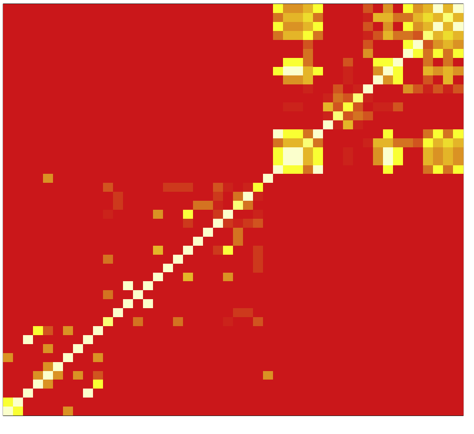


FIGURE N.2.—Covariance matrix for post-Nov 2011 GTDB data. Cells of cross-district covariance matrix, colored from low covariance (red) to high covariance (white). Ordering of rows and columns is the default order for GIS maps of Pakistan, which places districts in the same province together. Three groups are clearly visible. The GTDB data contain very few attacks in Punjab: no group corresponding to Punjab is visible. These data are clustered to produce Figure N.1.

TABLE N.I

PAKISTAN COMPARISON USING POST-NOV 2011 GTDB DATA[a]

	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

[a]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 N.1. 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 7(c). Districts shown as white ("no data") in either Figure 7(c) or N.1 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 N.1 assigns each district to one group. Rows may not sum exactly to 1 because of rounding. Standard errors in parentheses.

trator attributions recorded for each attack reflect what is reported in open-source media accounts, which does not necessarily indicate a legal finding of culpability" and teams of researchers at START (http://www.start.umd.edu/gtd/using-gtd/) are responsible for the verification and consistency of the entries.

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 sources, 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 four 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.

TABLE N.II

WAS THE ATTACK CLAIMED BY A TERRORIST GROUP?[a]

	OLS I	OLS II	OLS III	Logit IV	Logit V	Logit 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
N	87,901	13,441	13,441	87,901	13,441	13,441

[a]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).

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 group's 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 N.II 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 N.III shows that this appears to indeed be the case.

Check 7. Our theory of simultaneous attacks sketched in Appendix B 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. 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 N.III
IS THE ATTACK PART OF MULTIPLE ATTACKS?[a]

	OLS (1)	OLS (2)	Logit (3)	Logit (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
N	89,338	14,156	89,338	14,156

[a] Observations are individual terrorist attacks in all countries. Dependent variable is binary: whether or not the attack is part of a set of simultaneous (same-day) attacks. Attack types are an exhaustive set of dummy variables.

Table N.IV 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 N.IV 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 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 N.3 and Table N.V show that this indeed appears to be the case. Furthermore, Figure N.4 shows that the months in which the Taliban are most active are those months that have the greatest fraction of simultaneous attacks.

TABLE N.IV

RELATIONSHIP BETWEEN NUMBER OF ATTACKS AND OVERDISPERSION[a]

	OLS 1	OLS 2	IV 3	IV 4
(Intercept)	1.503 (0.040)		0.258 (0.319)	
Overdispersion	2.870 (0.207)	1.709 (0.121)	11.066 (1.679)	11.869 (4.658)
FKMS Controls		Yes		Yes
Country FE		Yes		Yes
Year FE		Yes		Yes
N	1940	1939	1466	1466

[a]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.

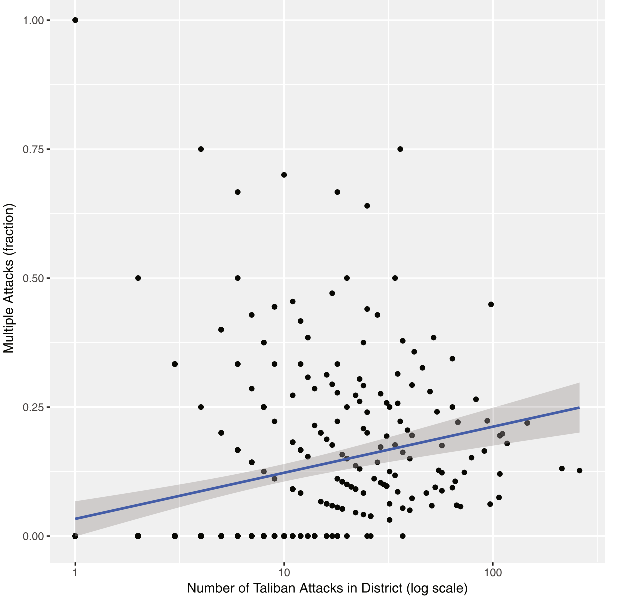


FIGURE N.3.—Fraction of Taliban multiple attacks.

TABLE N.V

FRACTION OF TALIBAN MULTIPLE ATTACKS[a]

	OLS I	OLS II	Logistic III	Logistic IV
(Intercept)	0.235	0.254	−1.016	−4.630
	(0.169)	(0.242)	(0.586)	(1.420)
log(Num Attacks)	0.037	0.039	0.312	0.477
	(0.009)	(0.010)	(0.051)	(0.077)
log(Population)	−0.022	−0.022	−0.264	−0.073
	(0.013)	(0.016)	(0.049)	(0.102)
log(Area)	−0.002	−0.002	−0.082	−0.217
	(0.008)	(0.012)	(0.035)	(0.063)
Night Lights 92, 00, 12		Yes		Yes
Province FE		Yes		Yes
N	341	341	341	341

[a]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

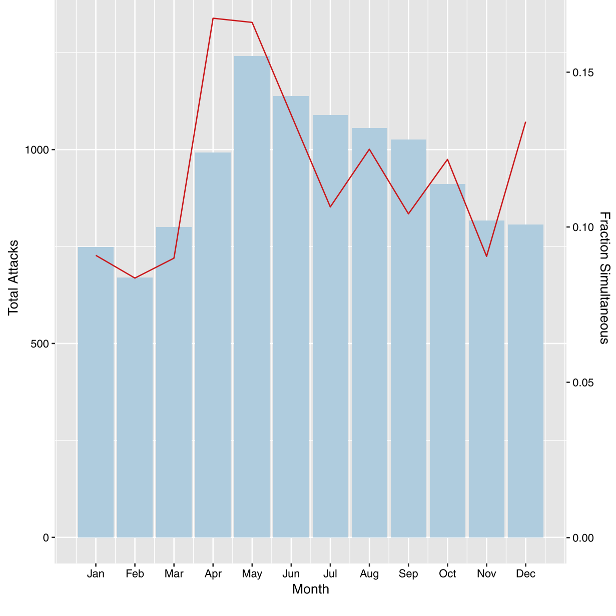


FIGURE N.4.—Seasonality in multiple attacks in Afghanistan. Blue bars show number of attacks. Red line shows fraction of attacks coded as simultaneous.

REFERENCES

AHN, S., AND A. HORENSTEIN (2013): "Eigenvalue Ratio Test for the Number of Factors," *Econometrica*, 81 (3), 1203–1227. [13]

ANDERSON, C. A. (1974): "Portuguese Africa: A Brief History of United Nations Involvement," *Denver Journal of International Law & Policy*, 133. [4]

ANDERSON, T. W. (1963): "Asymptotic Theory for Principal Component Analysis," *Annals of Mathematical Statistics*, 122–148. [8]

ARCE, D. G., AND T. SANDLER (2007): "Terrorist Signalling and the Value of Intelligence," *British Journal of Political Science*, 37, 573–586. [3]

BANSAL, N., A. BLUM, AND S. CHAWLA (2004): "Correlation Clustering," *Machine Learning*, 56 (1–3), 89–113. [10]

BARNO, D. (2006): "Challenges in Fighting a Global Insurgency," *Parameters*, 2006, 15–29. [3]

BAURLE, G. (2013): "Structural Dynamic Factor Analysis Using Prior Information From Macroeconomic Theory," *Journal of Business & Economic Statistics*, 31 (2), 136–150. [15]

CHOI, W. G., T. KANG, G. Y. KIM, AND B. LEE (2017): "Global Liquidity Transmission to Emerging Market Economies, and Their Policy Responses," *Journal of International Economics*, 109 (C), 153–166. [15]

DELOUGHERY, K. (2013): "Simultaneous Attacks by Terrorist Organisations," *Perspectives on Terrorism*, 7 (6), 79–90. [3]

EISENSTADT, M., AND J. WHITE (2005): "Assessing Iraq's Sunni Arab Insurgency," The Washington Institute for Near East Policy, *Policy Focus* No. 50. [1]

FERNANDES, C. (2008): "Hot Spot: Asia and Oceania," ABC-CLIO. [4]

FREYTAG, A., J. J. KRUER, D. MEIERRIEKS, AND F. SCHNEIDER (2011): "The Origins of Terrorism: Cross-Country Estimates of Socio-Economic Determinants of Terrorism," *European Journal of Political Economy*, 27, S5–S16. [24,25,31]

GOOD, P. (2002): "Extensions of the Concept of Exchangeability and Their Applications," *Journal of Modern Applied Statistical Methods*, 1 (2), 243–247. [17]

——— (2005): *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. New York: Springer. [17]

GOTELLI, N. J., AND A. CHAO (2013): "Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity From Sampling Data," in *Encyclopedia of Biodiversity* (Second Ed.), Vol. 5, ed. by S. A. Levin. Waltham, MA: Academic Press, 195–211. [25]

HENDERSON, A. (2005): "The Coalition Provisional Authority's Experience With Economic Reconstruction in Iraq: Lessons Identified," United States Institute of Peace, *Special Report* No. 138. [1]

HORN, B. (2013): "Chaos in Kandahar: The Battle for Building 4," *Canadian Military Journal*, 13 (3), 25–33. [26]

HORNIK, K., I. FEINERER, M. KOBER, AND C. BUCHTA (2012): "Spherical k-Means Clustering," *Journal of Statistical Software*, 50 (10). [10]

HUANG, K., N. SIDIROPOULOS, AND A. SWAMI (2014): "Non-Negative Matrix Factorization Revisited: Uniqueness and Algorithm for Symmetric Decomposition," *IEEE Transactions on Signal Processing*, 62 (1), 211–224. [16]

JENKINS, AND B. MICHAEL (2014): "The Dynamics of Syria's Civil War," *RAND Perspectives*, 115. [2]

JONES, S. G. (2008): "The Rise of Afghanistan's Insurgency," *International Security*, 32 (4), 7–40. [23]

KILCULLEN, D. (2009): *The Accidental Guerrilla: Fighting Small Wars in the Midst of a Big One*. Oxford University Press. [3]

KRISHNA, K., AND M. NARASIMHA (1999): "Clustering High Dimensional Data: A Survey on Subspace Clustering, Pattern-Based Clustering, and Correlation Clustering," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 29 (3), 433–439. [10]

LEDERMANN, W. (1940): "On a Problem Concerning Matrices With Variable Diagonal Elements," *Proceedings of the Royal Society of Edinburgh*, 60 (1), 1–17. [4]

MILLER, G., AND J. WARRICK ((2012), July 23): "In Syria Conflict, U.S. Struggles to Fill Intelligence Gaps," *Washington Post*. [2]

NG, A. Y., M. JORDAN, AND Y. WEISS (2002): "On Spectral Clustering: Analysis and an Algorithm," in *Advances in Neural Information Processing Systems, 14*, ed. by T. K. Leen, T. G. Dietterich, and V. Tresp. Cambridge, MA: MIT Press. [7]

PESARIN, F. (2001): *Multivariate Permutation Tests*. New York: Wiley. [17]

SAUNDERSON, J., V. CHANDRASEKARAN, P. PARRILO, AND A. WILLSKY (2012): "Diagonal and Low-Rank Matrix Decompositions, Correlation Matrices, and Ellipsoid Fitting," *SIAM Journal on Matrix Analysis and Applications*, 33 (4), 1395–1416. [5,6]

SCHMITT, E., AND T. SHANKER ((2004), October 22): "Estimates by U.S. See More Rebels With More Funds," *New York Times*. [2]

SHAPIRO, J. N. (2013): *The Terrorist's Dilemma Managing Violent Covert Organizations*. Princeton University Press. [3]

SHAPIRO, J. N., AND D. A. SIEGEL (2015): "Coordination and Security: How Mobile Communications Affect Insurgency," *Journal of Peace Research*, 52 (3), 312–322. [3]

SHI, J., AND J. MALIK (2000): "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 (8), 888–905. [6]

SMITH, B. (2012): "Syria: No End in Sight?," House of Commons Library Research Paper 12/48. [2]

SPENCE, M. (1973): "Job Market Signaling," *The Quarterly Journal of Economics*, 87 (3), 355–374. [3]

SUBRAHMANIAN, V. S., A. MANNES, A. ROUL, AND R. K. RAGHAVAN (2013): *Indian Mujahideen: Computational Analysis and Public Policy*. Springer. [4]

VON LUXBURG, U. (2007): "A Tutorial on Spectral Clustering," *Statistics and Computing*, 17 (4), 395–416. [7, 8]

WU, Y. (2012): "Essays in Panel Data Analysis," Unpublished PhD Thesis, University of Southern California. [15]

YAO, J., S. ZHENG, AND Z. BAI (2015): *Large Sample Covariance Matrices and High-Dimensional Data Analysis*. Cambridge University Press. [8]

Co-editor Joel Sobel handled this manuscript.

Manuscript received 27 May, 2016; final version accepted 30 June, 2018; available online 22 August, 2018.