Rebel Capacity and Combat Tactics*

Konstantin Sonin and Austin L. Wright†

January 4, 2020

Abstract

Classic and modern theories of rebel warfare emphasize the role of unexpected attacks against better equipped government forces. We test implications of a simple model of combat and information-gathering using highly detailed data about Afghan rebel attacks, insurgent-led spy networks, and counterinsurgent operations. Timing of rebel operations responds to changes in the group's access to resources. Results are supplemented with numerous robustness checks as well as a novel IV approach that uses machine learning and high frequency data on local agronomic inputs. Main effects are significantly enhanced in areas where rebels have the capacity to spy on and infiltrate military installations. Consistent with the model, shocks to labor scarcity and government surveillance operations have the opposite effect on attack timing.

*For advice and comments, we thank Karen Alter, Peter Deffebach, Quy-Toan Do, Kai Gehring, Garance Genicot, Matthew Gudgeon, Susan Hyde, Michael Kofoed, Daniel Krcmaric, Sarah Langlotz, Francois Libois, Nicola Limodio, Andrew Little, John Loeser, David Mansfield, Luis Martinez, Aila Matanock, Michaela Mattes, Owen Ozier, Kyle Pizzey, Michael Povilus, Robert Powell, Martin Ravallion, Mehdi Shadmehr, Alexander Stephenson, Oliver Vanden Eynde, Juan Vargas, Duncan Walker, Liam Wren-Lewis, Andrew Zeitlin, and Yuri Zhukov and seminar participants at Columbia University, Georgetown University, Northwestern University, UC Berkeley, University of Pennsylvania, and The World Bank Research Group. Faculty at the United States Military Academy (West Point) provided exceptionally detailed feedback. This material is based in part upon work supported by the Pearson Institute for the Study and Resolution of Global Conflicts. Eli Berman, Ethan Bueno de Mesquita, and Jacob Shapiro are owed a particular debt of gratitude for their support of this and related projects. Maria Ballesteros, Connie Guo, Lana Kugli, Michelle Liang, Mariya Milosh, Yonatan Litwin, Paul Soltys, and Julian Waugh provided excellent research assistance. All errors remain our own.

†Harris School of Public Policy, The University of Chicago, Chicago, IL 60637. Emails: ksonin@uchicago.edu and austinlw@uchicago.edu.

1 Introduction

Intrastate conflicts have replaced interstate wars as the main source of human loss and population displacement. Generally, it is well understood that resource endowments shape how rebels recruit, retain, and deploy their fighters (Weinstein, 2007). Fluctuations in rebel-held economic resources affect the scale of insurgent activity (Dube and Vargas, 2013), their control of strategic territory (Kalyvas, 2006), and how they treat civilians (Wood, 2014). These factors, in turn, impact whether civilians cooperate with rebels or collude with government forces (Condra and Shapiro, 2012) and the ability of the government to engage in development and reconstruction (Sexton, 2016). Shocks to rebel capacity impact even the finest aspects of internal warfare.

One central question that remains largely unexplored is whether rebel capacity influences actual military tactics, i.e., *how* rebels fight. In particular, classic theories of insurgency note that the rebel's main advantage in an asymmetric conflict is the 'element of surprise' (Galula, 1964). If guerrillas aim to undermine their more powerful rivals, their attacks must be unpredictable and, as such, difficult for government forces to anticipate and thwart (Thompson, 1966; Powell, 2007a,b). Yet the strategic value of random combat may diminish as rebels accumulate resources. As their capacity grows, rebels may shift from guerrilla tactics to conventional warfare, where frontal assaults are less random and more costly to the rebel forces (Bueno de Mesquita, 2013; Wright, 2016). Additional resources might also allow rebels to gather precise information about defensive weaknesses, including when troops and military installations are vulnerable to attack.

In our model of irregular warfare, rebels allocate scarce resources to gather information about vulnerability of the targets and conduct attacks based to this information. As government defensive resources are scarce as well, information about the presence of protection at one point in time is simultaneously information about (the absence of) protection during other periods of the day.[1] With higher quality information, rebels concentrate their attacks during time periods when targets are unprotected. This suggests a simple theoretical mechanism linking rebel capacity and patterns of attack timing. Positive economic shocks enable

[1]For example, following the Afghan surge in 2010, insurgents in Uruzgan province were able to monitor and assess weaknesses at one of the Coalition's forward operating bases. Rebels had amassed an arsenal of rockets, which they deployed strategically at particular times during the day. Despite rapid responses and a technological advantage, Coalition forces were unable to neutralize the insurgent threat and carefully timed attacks continued during the subsequent fighting season. See "Despite drones and blimps, rocket attacks in Afghanistan prove hard to stop", *The Christian Science Monitor*, 8/21/2012, https://tinyurl.com/y8e5x466.

rebels to acquire relatively high quality intelligence and their attacks become clustered in particular windows of time, i.e., exhibit low entropy. Empirically assessing the theoretical model requires unusually rich data on when attacks occur as well as largely unobservable details about rebel spy operations. We use newly declassified military records provided by the United States government, which document hundreds of thousands of combat engagements in Afghanistan during Operation Enduring Freedom, including 28,678 instances of indirect fire attacks (typically, rocket or mortar fire). The data include within-day timestamps, often to the minute, which we collapse to the hour. This allows us to study temporal clustering in attack patterns, consistent with a deviation from random attack timing. We quantify the randomness of combat timing by developing an extension of bootstrapping methods in statistics applied to the canonical Kolmogorov-Smirnov test (Abadie, 2002). We then study the association between attack clustering and rebel capacity using granular data on opium revenue (Peters, 2009).

We find consistent evidence that positive shocks to rebel capacity decrease randomness in the timing of violent attacks. As predicted by the model, rebel attacks become clustered during particular time windows following opium windfalls. In other words, insurgents begin to concentrate their combat operations during specific periods of the day as they accumulate resources. This finding survives a battery of robustness checks, including when we account for trends in violence during the fighting, harvest, and planting seasons, implement alternative fixed effects approaches, account for regional variation in crop yield and farmgate prices, exclude late-harvesting districts, and introduce a number of alternative measures of state capacity such as close air support missions, bomb neutralization, counterinsurgent surveillance activity, and safe house raids. The rich nature of our conflict microdata allow us to rule out other potential identification concerns, including rarely documented coercive tactics used by insurgents which may influence opium production. Spatial correlation-corrected randomization inference and coefficient stability tests confirm these benchmark results as well.

We also introduce results from a novel set of instrumental variables approaches. These approaches employ non-parametric estimation and machine learning techniques to identify agronomic factors which reliably predict opium productivity. Our IV approach flexibly leverages daily data on precipitation and temperature fluctuation by district as well as soil quality characteristics. These factors allow us to study variation in opium revenue that is driven by factors outside of combatant control and, because we use only weather conditions during the growing season months before fighting actually occurs, we sidestep potential concerns about violations of the exclusion restriction. We use supplemental data on pre-invasion irrigation networks, newly released data on military reconstruction projects, and sample randomization tests to rule out other potential concerns about instrument validity. Overall, our IV results yield strong evidence consistent with our main result: positive revenue shocks lead to more temporal clustering in rebel attack patterns.

The model emphasizes the critical role of the quality of information that insurgents gather about troop and facility weakness. As insurgents gather enough sufficiently precise information about target vulnerabilities, their attack patterns shift as they optimally calibrate the timing of their attacks. Our military records include previously unreleased information about rebel spy operations—surveillance of troop movement and base activity—observed by military forces. Our data suggests that insurgents were able to conduct surveillance operations in 70 of Afghanistan's 398 districts in 2006 (see Figure 1). These records enable us to test whether the ability to surveil, breach, or infiltrate targets significantly enhances the overall impact of revenue shocks in a manner consistent with our theoretical model. Ultimately, we find strong evidence that the baseline effect we observe is substantially greater in areas where the Taliban have the capability to acquire actionable intelligence about when troops, convoys, and bases are susceptible to attack. Additional data on rebel battlefield casualties and government-led surveillance operations allow us to assess the impact of policy interventions relevant to our theoretical model. We find robust evidence that labor scarcity (due to battlefield losses) and increasing costs of attack clustering (due to government surveillance) lead rebels to randomize the timing of their attacks more, yielding less temporal clustering.

The richness of our data finally enable us to further explore when and how economic shocks impact the production of violence. Opportunity costs, outside options, and reservation wages are central to most economic theories of political violence (Berman et al., 2011; Dal Bo and Dal Bo, 2011). Unique features of the Afghanistan case enable us to identify areas where reservation wages grow relatively faster due to the presence of bureaucratic corruption. To assess the extent of local administrative corruption, we rely on proprietary surveys conducted for the North Atlantic Treaty Organization by a local Afghan firm. Consistent with the theoretical models in Berman et al. (2011) and Dal Bo and Dal Bo (2011), our results suggest revenue shocks from opium production have a larger impact on rebel tactics in places where reservation wages are relatively lower. Savings technologies in rebel organizations may enable insurgents to anticipate and account for revenue volatility in their expenditures during the fighting season after particularly profitable harvests. We find no evidence, however, that rebels engage in consumption smoothing.

This paper contributes to the political economy of conflict. Prior work provides compelling evidence that insurgents respond strategically to local economic shocks (Dube and Vargas, 2013; Berman et al., 2017; Vanden Eynde, 2018), form alliances during war (Konig et al., 2017), and calibrate their use of violence against civilian populations (Condra and Shapiro, 2012; Condra et al., 2018). Recent work also links exogenous economic shocks to terrorism financing and recruitment activity on the dark net (Limodio, 2019). Our evidence highlights how sophisticated institutions commonly associated with states and government forces—structured tax collection schemes, combat coordination, and surveillance operations—influence the timing of insurgent violence.

Understanding how economic shocks and government countermeasures impact rebel tactics remains an important policy issue. The United States alone has spent more than two trillion dollars on combat operations in Afghanistan. Trebbi et al. (2017) present evidence that government counter-IED measures may have been quickly thwarted by rebels. Our evidence, on the other hand, suggests labor constraints and government surveillance may meaningfully impact how rebels fight.

The rest of the paper is organized as follows. Section 2 provides a brief overview of the related literature. Section 3 introduces our theoretical model. Section 4 details the empirical strategy. Section 5 presents the main results and robustness checks. The final section concludes.

2 Economics of Political Violence

Quantitative work on the economics of conflict has largely focused on the conditions that trigger warfare. Fearon and Laitin (2003), Collier and Hoeffler (2004), Miguel et al. (2004), and Bazzi and Blattman (2014) study the proximate causes of conflict. Related work explores incentives to use violence as a means to predate or capture economic resources (Le Billon, 2001; Ross, 2004; Hidalgo et al., 2010). Others have focused on the link between income shocks and levels of political violence at a subnational level (Dube and Vargas, 2013; Jia, 2014; Wright, 2016; Vanden Eynde, 2018; Gehring et al., 2019). A meaningful gap exists, as Berman and Matanock (2015) point out, between our understanding of *when* and *how* rebels engage in armed combat. We advance this agenda by examining the relationship between economic shocks to rebel capacity and the ways in which insurgents fight.

Violence is economically and politically costly. Gould and Klor (2010) find that terror attacks harden in-group biases and cause Israelis to adopt less accommodating political positions.[2] Those affected by these attacks are also more likely to vote for right-wing parties. Similarly, Condra et al. (2018) find that insurgent attacks around elections in Afghanistan are calibrated to avoid civilian casualties and substantially reduce voter turnout. Violence may be triggered by and reinforce ethnic divisions (Esteban and Ray, 2011; Esteban et al., 2012), even in institutions designed to maintain impartiality (Shayo and Zussman, 2017). Insecurity might also undermine political instability through a realignment between international actors and domestic power brokers (Padro i Miquel and Yared, 2012). Civil conflict is also economically disruptive (Abadie and Gardeazabal, 2003), even at the microlevel (Besley and Mueller, 2012). Our results help us better understand how insurgents respond to rent shocks and may enable governments to better anticipate and defend against attacks.

More generally, our paper provides insights into the working mechanism of insurgency. State capacity is central to economic theories of conflict (Besley and Persson, 2011; Gennaioli and Voth, 2015; Besley and Persson, 2010; Powell, 2013; Carter, 2015; Esteban et al., 2015). Yet the resources available to the state's competitors also influence when conflicts emerge (Dube and Vargas, 2013), how internal wars are fought, and whether they end in withdrawal. Recognizing this gap, Bueno de Mesquita (2013) theorizes that the relative strength of the rebellion influences leaders' choice to adopt irregular tactics. We advance this literature by developing a microlevel theory of the relationship between resource endowments and combat tactics.

Technology of conflict has long been an active area of theoretical modelling (Kress, 2012). Optimal allocation of attacking and defensive resources have been studied in the Colonel Blotto-type games starting with Borel (1953). Our game is not a "Blotto game" in the sense of Roberson (2006), which assumes that the side that allocates more resources to one battlefield wins for sure, yet is "Blotto" in the sense of Blackett (1958) and Powell (2007b), which allow for the probability of success to depend on allocated resources more generally.[3] A contribution of our model is the signal about the relative vulnerability of different targets that rebels receive before launching attacks: it allows to analyze comparative statics with respect to the quality of information.

In a setting in which the maximand is the sum of harm to individual sites, Powell (2007a) demonstrates that in the unique equilibrium, the defender minimaxes the attacker regardless of whether or not the attacker can observe the defender's allocation before choosing where to launch the attack. In Powell (2007a), the defense has private information about the relative

[2] See also Berrebi and Klor (2008) and Getmansky and Zeitzoff (2014).

[3] See Golman and Page (2009), Roberson and Kvasov (2012) and Konrad and Kovenock (2009) for recent advances and Kovenock and Roberson (2012) for an excellent survey.

vulnerability of two sites, and allocated protection is public information, which creates the secrecy vs. vulnerability trade off. Goyal and Vigier (2014) consider a zero-sum Tullock-type conflict, in which the defense chooses a network to protect and the attacker chooses where to launch an attack. In particular, they show that a star network with all defence resources allocated to the central node is optimal in many circumstances. In Konig et al. (2017), the network is given; the data on the Second Congo War is used to estimate the impact of selective dismantling of fighting groups and weapons embargoes on conflict intensity. We estimate the potential effects of abnormal battlefield losses (a negative shock during the fighting season partially offsets revenue windfalls), which is close to selective dismantling in Konig et al. (2017).

3 Theory

In our model, a rebel group chooses the number of attacks and the precision of information about targets' vulnerability, and then allocates attacks basing on this information. This is the simplest possible model consistent with our empirical approach that utilizes randomization inference and the bootstrap Kolmogorov-Smirnov method developed by Abadie (2002).

3.1 Setup

Consider a rebel group that attacks the government facilities using a certain technology (e.g., mortars). The group uses information of the quality $\theta \in \left[\frac{1}{2}, 1\right]$ to allocate the total of a attacks across the targets. The government allocates resources to defeat attacks.

There are n time slots to protect.[4] The government has resources to defend $r < n$ time slots. Formally, the government strategy is a probability distribution $G(\cdot)$ over n-tuples $(g_1, ..., g_n)$ such that $\sum_{i \in n} g_i = r$ and $g_i \in \{0, 1\}$ for each $i \in n$. If an attack happens during the period when the target is defended, it does not succeed; if an attack is in an unprotected time slot, it succeeds with probability $p \in (0, 1)$. Since any deterministic choice of protection will result in rebels attacking outside of the time slots when the targets is defended, any reasonable placement of protection should be randomized. After the government allocates protection, rebels gather intelligence about the targets' vulnerability during different

[4] We consider targets to be "time slots" rather than space objects as this corresponds with our empirical data. Therefore, it is natural for rebels to maximize the probability of at least one successful attack, rather than the expected number of destroyed targets, which would be natural if targets were space objects. Our formal results extend to this alternative setting as well.

time-periods. Specifically, rebels receive noisy signals $(s_i)_{i\in n} \in \{0,1\}^n$ that are determined according to

$$P(s_i = 0|g_i = 0) = P(s_i = 1|g_i = 1) = \theta.$$

Then the rebels strategy is a mapping $F(a_1, ..., a_n; \cdot)$ from the n-tuples of signals about periods' vulnerability into a probability distribution on n-tuples $(a_1, ..., a_n)$ such that $\sum_{i\in n} a_i = a$ and a_i is a non-negative integer for each $i \in n$.

The rebel group maximizes the probability of at least one successful attack net of the cost of attacks and information gathering. As gathering intelligence requires resources, we assume that the quality of information θ is a function of revenues in rebels' disposal. [5] Thus, the periods of low rents correspond to lower quality of information available to rebels. The government is interested in minimizing the probability of at least one successful attack.

Timing

1. The government chooses allocation of resources across n time slots.
2. Rebels receive information $(s_i)_{i\in n} \in \{1,0\}^n$ and choose the distribution of attacks across slots.
3. Pay-offs are received.

Definition 1 *Given the resources available to rebels, an equilibrium is rebels' choice of a c.d.f.* $F^*(s_1, ..., s_n; \cdot)$*, which is a function of signals about each time slot, into a probability distribution over* a *attacks on each of the* n *time slots, and the government's choice of a c.d.f.* G^* *over* $r-$*combinations of* n *time slots. Given* G^*, F^* *maximizes the probability of at least one successful attack net of the cost of attacks and information gathering; given* F^*, G^* *minimizes it.*

3.2 The Attack Timing Game

We start backwards. It is straightforward to establish that the government allocates resources into r time slots chosen randomly and uniformly across all possible combinations. The rebels' optimal strategy depends on the signals that they observe. Information gathering results in x "vulnerable" ($s_i = 0$) and $n - x$ "defended" ($s_i = 1$) time slots. Let $q(x)$ denote the *ex*

[5] In the working paper version, we explicitly model the allocation of resources between additional number of attacks and information precision, taking into account the marginal costs of these activities.

post probability that time i with $s_i = 0$ is vulnerable when the total number of vulnerable signals is x :

$$q(x) = P(g_i = 0 | s_i = 0, \#\{s_j = 0\} = x).$$

Importantly, although signals are conditionally independent, when resources are scarce, a signal about vulnerability of one period is informative about the vulnerability of other periods. Indeed, if one time slot is more likely to be vulnerable, other time slots are less likely to be vulnerable as probability of being one of $r-1$ protected time slots among $n-1$ periods is smaller than to be one of r among n. It is of course possible that the number of "vulnerable" signals x is not equal to the total number of attacks a; in fact, x can be any integer between 0 and n with a non-zero probability. In the two extreme cases, $x = 0$ (there are no time slots that are more likely to be vulnerable) and $x = n$ (all potential times an attack could occur are equally vulnerable), there is no information to update upon. In all other cases, $1 \leq x \leq n-1$, signals are informative:

$$P(g_i = 0 | s_i = 0) > P(g_i = 0 | s_i = 1).$$

The number of "vulnerable" signals is a random variable, the sum of two binomial distributions with different probabilities of success: $n - r$ vulnerable time slots produce signal 0 with probability θ, while r defended time slots produce signal 0 with probability $1 - \theta$.[6]

Proposition 1 *There exists a unique equilibrium in the attack timing game. The government protects r time slots chosen randomly and uniformly across all possible combinations and rebels follow the signals that they receive. If $a \leq x = \#\{s_i = 0\}$, then a attacks are distributed uniformly over x "vulnerable" time slots. If $a > x$, there is an optimal number of attacks $\overline{a}(x)$ such that $\min\{a, x\overline{a}(x)\}$ attacks are distributed uniformly over x "vulnerable" time slots. The remaining $a - \min\{a, x\overline{a}(x)\}$ attacks are distributed uniformly across $n - x$ "defended"time slots.*

Critically, with time slots labeled "vulnerable" and "defended" after the informative signals are received, the rebels' optimal strategy is a function of the probability $q(x)$ that a given time period that is signaled to be vulnerable is indeed vulnerable. The intuition is as follows. Consider the rebels' choice of one attack across two time slots with probabilities of being vulnerable q_1 and q_2, respectively, with $q_1 > q_2$. If there are no attacks already planned

[6]The probability distribution of a sum of two or more binomial random variables, i.e. sums of independent Bernoulli trials, with different success probabilities is sometimes called Poisson binomial distribution (Hillion and Johnson, 2017).

on during these times, then an attack timed during the first period provides a higher marginal probability of success (see also Powell (2007b)).

The fact that the vulnerable time slot has a higher probability of success does not mean that all attacks should be concentrated during times when the target is vulnerable. Suppose that there are already m attacks planned during the first time period, $m \geq 1$, and no attacks planned on the second time slot. One more attack during the first time slot results in $q_1 (1-p)^m$ of marginal probability of success as with probability $1-(1-p)^m$ one of the m attacks that are already planned succeeds. An attack that occurs during the second time window contributes $q_2 p$. This explains why, given some sufficient capacity, rebels will launch some attacks during the second, less likely to be vulnerable, period regardless of the initial distribution of probabilities q_1 and q_2.

The term $\min\{a, xa(x)\}$ appears in the statement of Proposition 1 because it is possible that the threshold $a(x)$ is such that $xa(x)$, the desired number of attacks during vulnerable time slots, does not exceed the capacity a. For example, if $n = 5$, $a = 3$, and intelligence about defensive weaknesses suggests two specific time windows are likely to be vulnerable, the optimal strategy might call for $\overline{a}(2) = 2$, i.e. two attacks should be launched during each of the two vulnerable time slots. With the capacity of launching only three attacks, the rebels would have to choose the period for the double attack randomly over the two vulnerable time windows.

3.3 The Rebel's Demand for Precise Information

The rebels' equilibrium strategy described in Proposition 1 depends on the quality of information θ. A higher precision of information leads to a higher temporal concentration of attacks: more attacks are launched during time periods that intelligence gathering indicated as "vulnerable". For any number of "vulnerable" signals x, the optimal strategy requires to launch (weakly) more attacks during the vulnerable time windows. This, in turn, leads to a decrease in the natural measure of randomness of attacks, the entropy $\sum_{i=1}^{n} \frac{a_i}{a} \ln \frac{a_i}{a}$, which is maximized when attacks are distributed uniformly across time intervals.

Proposition 2 *For any number of attacks a, the higher is the precision of information that rebels receive, θ, the higher is the temporal clustering (concentration) of attacks, i.e., the lower is the expected number of unique periods attacked and the larger is the expected number of attacks (both successful and total) per time slot attacked.*

The critical element of Proposition 2 is that for any number x of time slots that are signaled to be vulnerable after the information is collected, the probability $q(x)$ that a time window marked vulnerable is indeed vulnerable is (weakly) increasing in the precision of information θ, and thus the threshold $\overline{a}(x)$ is (weakly) increasing in θ for any x. As a consequence, more precise information leads to a higher concentration of attacks: more attacks are launched during a smaller number of time slots. In the extreme case of perfect information ($\theta = 1$), attacks are uniformly randomized over all $n - r$ vulnerable time slots. In the opposite extreme, when the signals are completely uninformative ($\theta = \frac{1}{2}$), all time windows are equally likely to be vulnerable. In this case, the optimal strategy for rebels is to launch a attacks chosen randomly and uniformly across all time slots.

In equilibrium, the optimal choice of rebels depends, in addition to precision of information θ, which in turn is a function of the rebels' resources, on the number of potential time slots for attacks n, the resources in the disposal of the government r, and the efficiency of weapons p. Proposition 2 establishes that a lower precision of information, a result of a fall in revenues, leads to attacks becoming less temporally concentrated. This increase in entropy as a result of decrease in resources is the central empirical implication of our model which is tested, using the technique suggested in Abadie (2002), in Sections 4-5. (See Subsection 4.4 for the methodology of detecting temporal patterns of combat.) Proposition 3 formally states the comparative statics results with respect to the government's resources and efficiency of rebels' weapons, which we evaluate empirically in Section 6.

Proposition 3 *More resources in the government's disposal (a higher r) and less efficient weapons, i.e., a lower p, result in a higher temporal concentration of attacks (lower entropy).*

Proposition 3 states that an increase in government resources has the same effect as a fall in rebels' resources, but the mechanism is slightly different. A higher r, the amount of protection at the government disposal, decreases the probability of rebels' success as the *ex ante* probability that each time window is protected increases. This has the same effect as the fall in rebels' revenues, which results in less information gathering and, therefore, less temporally concentrated attacks.

4 Empirical Design

In this section, we discuss the setting of our investigation, review our microdata, and introduce our identification strategy.

4.1 Context

We study the relationship between combat activity and rebel capacity in Afghanistan. We focus specifically on the well-documented link between opium production and Taliban tax extraction. In the primary opium producing regions, seeds are planted in late fall and early winter. The growing season ranges from February through April, with most opium latex harvested and packaged in May and June. Taliban commanders and veteran fighters return from Pakistan in June to collect taxes from opium farmers (*ushr*, typically a flat 10% fee mandated by the Quran).[7] Taxes can be paid in currency, opium blocks, or other goods, such as motorcycles, offroad vehicles, and weaponry. The Taliban also benefits from protection fees levied on opium traffickers as they pass through rebel-held territories.[8]

Taxes are collected by fighters and receipts are distributed to farmers to prevent double taxation. Fighters pass their collections to district-level commanders (equivalent in scale to US counties). Taxes are subsequently passed upward to provincial and regional commanders, who keep ledgers of their annual revenue and are subject to audit by the Taliban's Central Finance Committee, based out of southwestern Pakistan. Most proceeds remain with the district commander, for conducting operations in the subsequent fighting season which typically lasts until September. These funds can be used to purchase weapons and ammunition, as well as covering the salaries of fighters and rebel informants. The Central Finance Committee (CFC) retains the authority to demote or relocate field commanders to less desirable fronts if audit irregularities are found. The remaining revenue is split between supporting operations conducted in resource-poor districts where local taxes alone are insufficient for supporting rebel attacks and developing Taliban infrastructure in Pakistan (including small-scale hospitals for wounded fighters).[9]

We focus primarily on the period from 2006 to 2014. The industrial organization of the insurgency, most notably the taxation and command structures oriented around administrative districts, emerged in 2006. These institutions are central to our argument that revenue influences combat tactics. Our military records track insurgent operations until the end of

[7]For a detailed account of the industrial organization of the Taliban, see Peters (2009).

[8]It is worth noting, as highlighted by Mansfield (2016), taxation by the Taliban is not perfectly uniform. In areas where the Taliban are not sufficiently powerful, the amount collected may be reduced (reflecting a local-level bargain with tribal leaders). Microlevel data on tax collection does not exist. Given the approach we detail later, we anticipate that this concern would bias our estimated effects downward. In Table A-6, we use historical territorial control (from the height of the Taliban) to evaluate this conjecture. Indeed, we find that our main effect is attenuated downward in places the Taliban could not have taxed fully prior to the US-led invasion.

[9]We observe variation in the opium tax base of each district, not receipts from the CFC. This type of redistribution will bias our point estimates towards zero.

2014, when the NATO Operation Enduring Freedom was transitioned to Mission Resolute Support.

4.2 Conflict Microdata

Our investigation exploits newly declassified military records which catalogue combat engagements and counterinsurgent operations during Operation Enduring Freedom in Afghanistan. These data were maintained by and retrieved through proper declassification channels from the U.S. Department of Defense. The data platform was populated using highly detailed combat reports logged by NATO-affiliated troops as well as host nation forces (Afghan military and police forces). Data of this type differ substantially in coverage and precision from media-based collection efforts (Weidmann, 2016). As Weidmann (2016) points out, these tactical reports represent the most complete record of the war in Afghanistan. Additional details on data collection are discussed in Shaver and Wright (2017).[10]

The detailed nature of our conflict microdata allows us to track insurgent activity by the hour. Although this data tracks dozens of types of violence, the majority of enemy action events are characterized as indirect fire, direct fire, and IED explosions. Indirect fire consists of mortars and other weapons that can be deployed without close contact with military forces. Direct fire attacks are primarily line-of-sight, close combat events. IEDs consist of explosives that have been emplaced and are detonated through a variety of trigger mechanisms (pressure plate, cable-to-battery, radio signal, laser beam, etc.). Subsets of these data are also studied in Callen et al. (2014), Beath et al. (2013), and Condra et al. (2018), and are highly comparable to tactical data collected in Iraq (Berman et al., 2011).[11]

Our military records include information about counterinsurgent operations, including find and clear missions that neutralized emplaced IEDs, discoveries of weapon caches, such as small arms, ammunition, and bomb making materials, and provision of close air support,

[10] These data are similar to the War Logs released by WikiLeaks in 2010. The most important differences are that the data we study here are more complete (2009 versus 2014) and have been formally declassified.

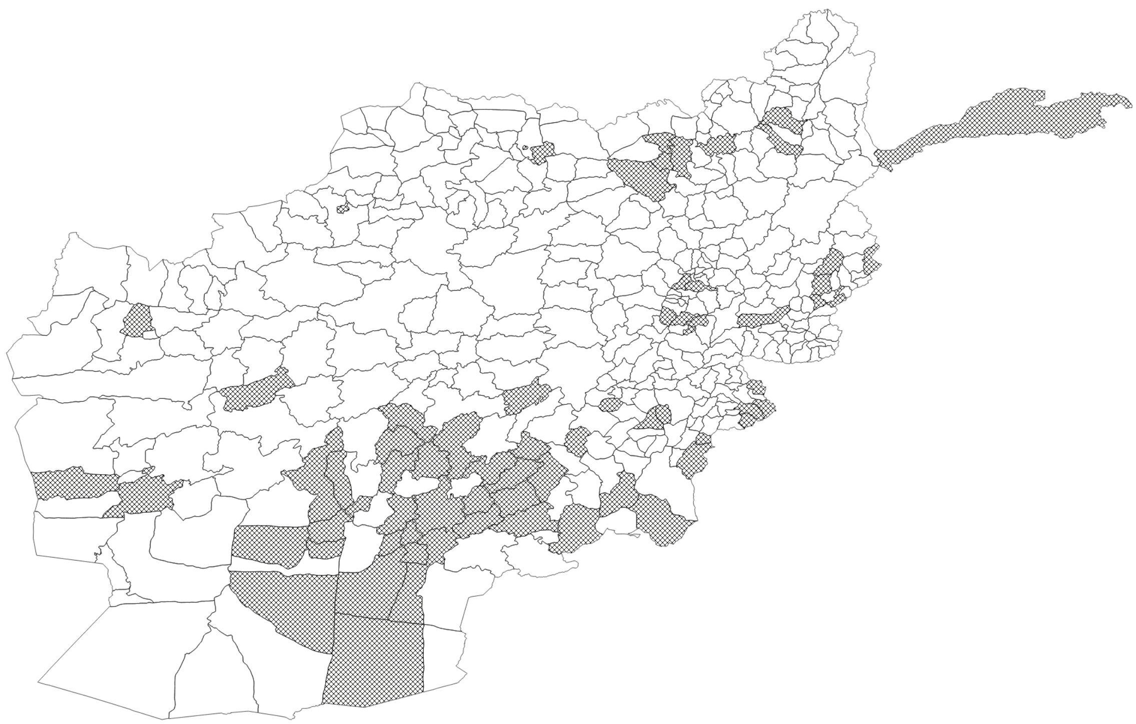
[11] It is important to note that insurgents have differential control over the exact timing of each of these types of combat. Indirect fire events can be initiated at any time against stationary targets. For example, a large number of mortar fire attacks were launched from hillsides and targeted military outposts. As such, insurgents maintained a high level of control over when these events occurred. Direct fire attacks are typically frontal assaults on convoys or forward deployed troops conducting combing operations in remote villages. Although insurgents decide if to attack, conditional on having the opportunity, the timing of these attacks is partially determined by the movement of troops, the timing of which may be determined by a non-stochastic process. Similarly, roadside bombs may be planted hours or days before they are triggered. Some of these bombs may also be detonated by unintended targets, which further reduces insurgent control over the timing of these attacks. For these reasons, we focus our main analysis exclusively on the timing of indirect fire attacks. We discuss this in more detail in the Online Appendix.

which was used primarily to harden mobile targets and extract coalition forces that were pinned down at a fixed location by insurgents. We supplement this information with measures of insurgent capture and detention, counterinsurgent surveillance operations, and safe house raids yielding actionable intelligence about rebel operations. This information enables to address operational factors that may confound the relationship we are trying to estimate. In particular, it might be the case that counterinsurgents strategically assign additional fighting capacity to districts with a substantial rebel presence *and* where opium production is high. Shifts in government resources, like the ability to identify and neutralize weapon caches or to use aerial bombardment to assist troops engaged by insurgents, might also influence the temporal patterns of insurgent attacks. These operations also lead to battlefield losses for insurgents (in the form of fighter casualties), which we utilize to study the impact of labor scarcity on combat tactics.

Opium production may be influenced by combat operations *and* more direct attempts by insurgents to coerce the local population (Lind et al., 2014). In particular, insurgents may use violent and non-violent tactics to intimidate civilians, such as killings of government collaborators and the posting of 'night letters' and other non-lethal shows of force. Fortunately, our military records include information about these tactics as well, enabling us to address potential concerns about residual endogeneity in estimated levels of opium production.

Another unique feature of our conflict data is information on rebel surveillance operations, military installation breaches, and insider attacks. Enemy surveillance operations are logged whenever security forces become aware of attempts by insurgents to track troop and vehicle movement and day-to-day activities on military bases. This kind of insurgent spy activity can be, and likely is, used to identify troop and infrastructure vulnerabilities. We illustrate the location of these spy operations in Figure 1. In addition to surveying force activity from outside of military compounds, insurgents can infiltrate these installations and monitor activity from within. These security breaches might also result in direct confrontations between insurgent and counterinsurgent forces. Records of insider attacks also reveal Taliban attempts to 'turn' Afghan security forces, and launch deadly attacks from within operational units. We use these unique pieces of information to more explicitly test the observable implications of our model regarding intelligence gathering and the quality (precision) of information about government vulnerabilities.

Figure 1: Military records indicate the location of rebel surveillance operations conducted in Afghanistan (2006).



Notes: Data on insurgent spy operations drawn from SIGACTS military records. Cross hatch pattern indicates insurgents conducted at least one detected surveillance operations during 2006, the first year of our sample. District boundaries are drawn from the ESOC Afghanistan map (398 districts).

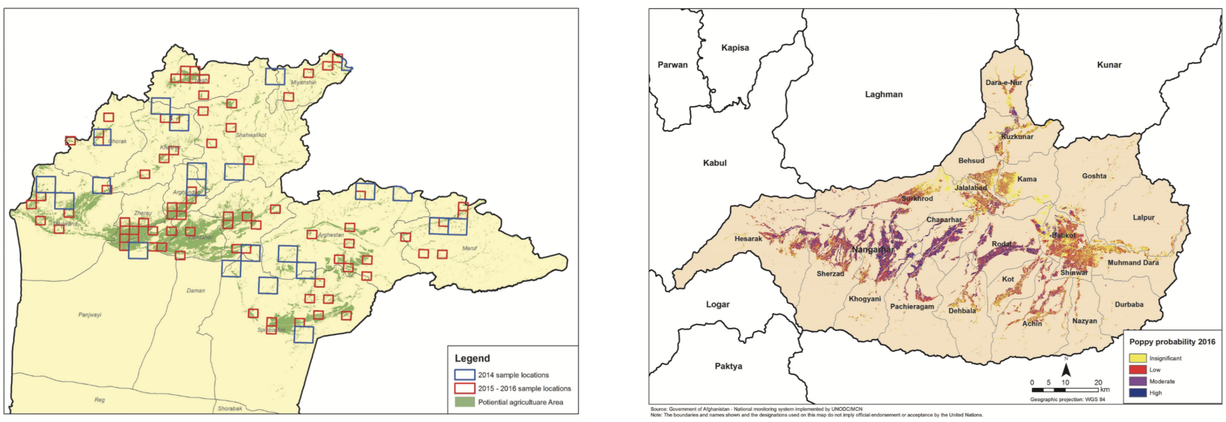
4.3 Opium Cultivation and Prices

We supplement our military records with data on opium production and prices. Opium production estimates are derived from ground-validated remote sensing techniques, which use high resolution satellite imagery to track changes in vegetation during the spring harvest. UNODC-Afghanistan randomly spatially samples potential agricultural zones within provinces and acquires pre-harvest and post-harvest imagery (see Figure 2, panel (a)). These images are then examined for changes in vegetative signatures consistent with the volatile wetness of opium plants after lancing. From this sampling technique, officers estimate the spatial risk of opium production. This enables them to calculate granular estimates of opium production (see Figure 2, panel (b)). These gridded estimates are then compiled as the annual amount of opium production (in hectares) for each district. We correct for changes in the administrative boundaries of districts over time using the Empirical Studies of Conflict

(ESOC) administrative shapefile. To translate production into yields, we compile additional details about annual yield (kilograms per hectare) from UNODC-Afghanistan annual reports. These figures are available at the national level as well as by region.

Opium price data is compiled at national and regional levels. These prices are tracked monthly at various locations across the country via a farmer and market spot price survey system. In Figure 3, panel (a), we introduce the monthly time series for the two price-making regions, Kandahar and Nangahar. Our main specification utilizes the simple average between these prices in June, when taxes are collected (vertical lines added for clarity). In supplemental results, we study the regional price time series in Figure 3, panel (b). We rely on UNODC-Afghanistan documentation to assign districts to price zones. We extract exact prices using WebPlotDigitzer software.

Figure 2: UNODC Methodology for Estimating Annual District Drug Production.



(a) Sampling Satellite Imagery

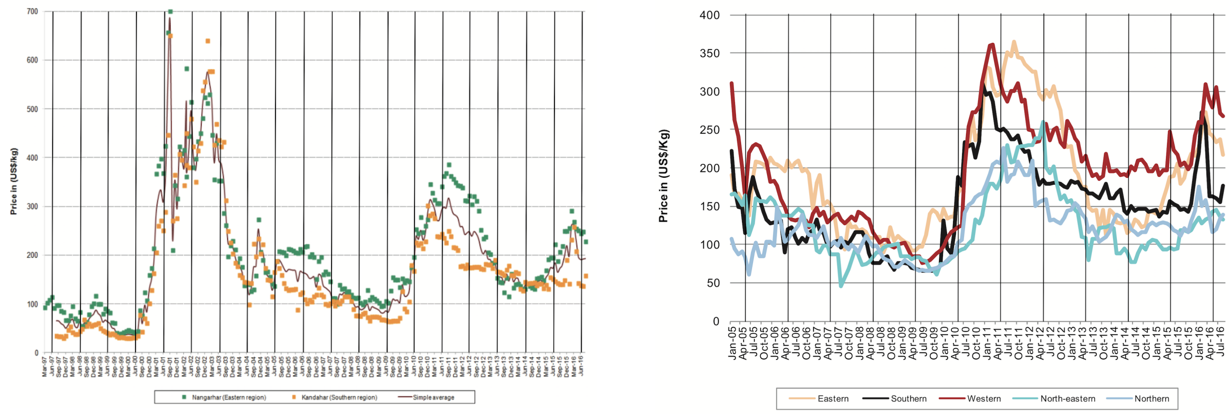
(b) Impute Production from Imagery/Field Obs.

Notes: Methodological figures and details drawn from the 2016 UNODC-Afghanistan Drug Report. Panel (a) demonstrates the sampling design used when acquiring high resolution satellite imagery (location: Kandahar). Panel (b) illustrates the subsequent production estimation, which combines low and high resolution imagery (location: Nangahar).

4.4 Detecting Temporal Patterns of Combat

We now introduce a method for detecting temporal clustering of attack patterns. We focus on timing rather than spatial allocation of attacks due to the tractability of estimation. Most importantly, while we know time windows when attacks do not occur, it is much more difficult to identify counterfactual targets that were not attacked since there exists no publicly

Figure 3: Time series data on opium prices collected at national (a) and regional (b) levels.



(a) National price time series

(b) Regional prices time series

Notes: Figures on national and regional price time series drawn from the 2016 UNODC-Afghanistan Drug Report. Underlying data compiled from farmer and market spot price surveys conducted throughout the year. In Panel (a), the simple average is calculated (Nangahar, Kandahar). In Panel (b), we assign districts to regions according the UNODC documentation. Prices were precisely extracted using WebPlotDigitzer.

available comprehensive data on the spatial allocation of military assets across time (e.g., bases, patrol movement, unmanned infrastructure). Conceptualizing these counterfactuals is central to our measurement strategy. Our approach to quantifying patterns of combat timing employs randomization inference and the bootstrap Kolmogorov-Smirnov method developed by Abadie (2002). The central intuition of this approach is to use randomization inference to better understand the degree of temporal clustering we observe in insurgent combat operations (by hour). The method is executed in several steps.

1. Fit a local polynomial regression to the observed distribution of violence by hour. We specify a conservative bandwidth of 1. This empirical distribution of fitted values is stored.[12]

2. Identify the sequence of district-hours during which indirect fire engagements occur. For each district-hour, we know the sum of the number of attacks.

3. Randomly shuffle the sequence above. This is equivalent to a randomization or permutation test.

[12]Some conflict events lack a time stamp (∼3%). Because we cannot assign these events an hour, they are excluded from the calculation of the empirical distribution.

4. Fit a local polynomial regression to the randomly shuffled distribution of violence by hour. The simulated distribution of fitted values is stored.

5. Execute the bootstrap Kolmogorov-Smirnov test. This test is composed of four elements.

   (a) Compute the T^{KS}_{dfi} for the fitted values of the empirical and simulated distributions, where:
   $T^{KS}_{dfi} = \left(\frac{n_1 n_0}{n}\right)^{\frac{1}{2}} \sup_{y \in \mathbb{R}} \left|F_{1,n_1}(y) - F_{0,n_1}(y)\right|.$

   (b) Resample observations with replacement from observed and simulated distributions. Split the resampled set into two distributions and calculate $T^{KS}_{dfi,b}$. Store $T^{KS}_{dfi,b}$.

   (c) Repeat prior two steps 1,000 times.

   (d) Calculate and store the likelihood parameter of the tests as $\sum_{b=1}^{1000} \frac{1\{T^{KS}_{dfi,b} > T^{KS}_{dfi}\}}{1{,}000}$, where the numerator is an indicator function.

6. Repeat steps 2 through 5 10,000 times. Evaluate the central tendency (mean) of the likelihood parameters.

7. Replace zero values with the minimum observed non-zero rank value and calculate the log.

To clarify, we identify the hour of each attack within a given district-year (fighting season). We then reshuffle the hour vector and compare the empirical distribution to the randomly reshuffled vector. This process is repeated many times per district-year. The result of the technique is a single likelihood parameter, which we call a p-value, for each unit of observation. We estimate these parameters for district-years with a minimum of five conflict events.[13] Higher p-values indicate that the distribution of rebel attacks by hour cannot be distinguished from randomness. Lower p-values reveal attack patterns that are more easily differentiated from randomness; i.e., they are more predictable.

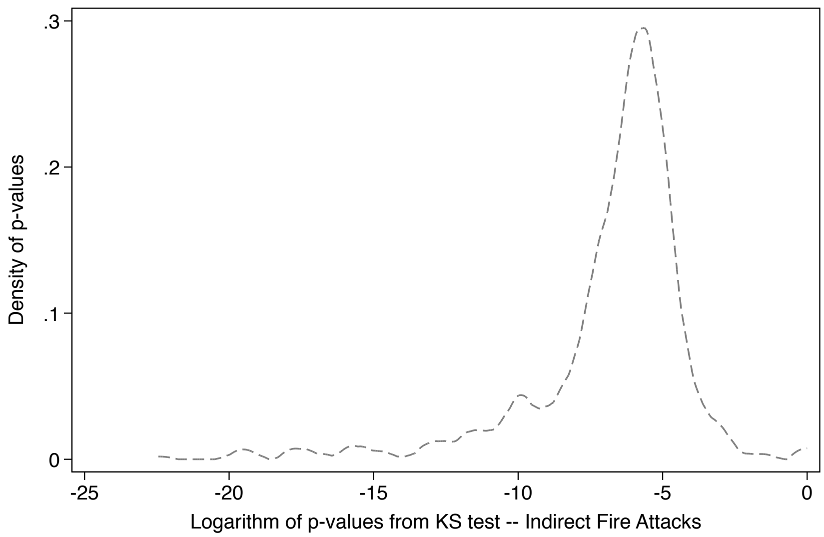
It is important to note here that this parameter does not clarify what exact time windows experience differential attack intensity. One benefit of our approach is that we do not need to

[13] We set the lower threshold at five events to ensure convergence of the simulations. A conflict vector that is too short (i.e., fewer than five) does not permit sufficient randomization when the hour vector is reshuffled. Our results are highly consistent if we raise this threshold upward.

make additional assumptions about potential temporal weaknesses of military installations across seasons in the same district or across districts in the same season (e.g., dawn or dusk hours). Instead, the parameter flexibly captures a uniform measure of temporal clustering even when the underlying distributions differ across locations or periods.

Estimation of our likelihood parameters using this technique requires tens of billions of simulations, so we use several supercomputers. In Figure 4, we plot the calculated p-value (log) distribution for indirect fire attacks. Most district-year p-values above -10. This distribution is characterized by a long left-side tail. This suggests that specific district-fighting seasons exhibit very clear evidence of temporal clustering (i.e., non-randomness).

Figure 4: Distribution of p-values from randomization test of combat in Afghanistan



4.5 Empirical Strategy

We study the relationship between rebel capacity and randomized combat by examining whether the within-day distribution of violence for each district's fighting season is associated with rent extraction from opium. If the results follow our expectations, rebel capacity and the likelihood parameter of our simulation test above should be negatively correlated.

To test the relationship between randomization of attack timing and insurgent capacity, we estimate the following ordinary least squares regression:

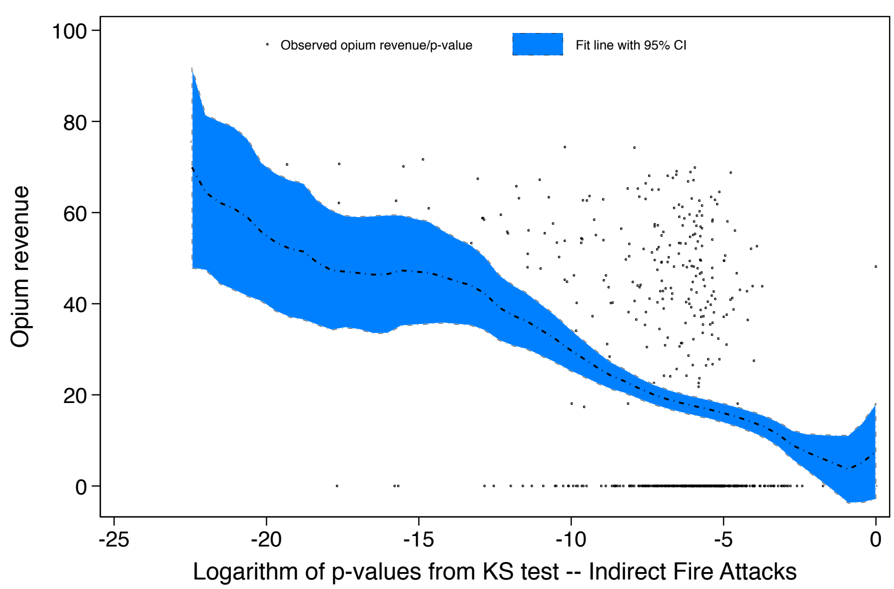
$$log(pval_{d,y}) = \alpha + \beta_1 log(production_{d,y} + 1) \times log(price_y) + F_y + \Lambda X^V_{d,y} + \epsilon \tag{1}$$

Where $pval_{d,y}$ is the p-value for a given district, d, and fighting season, y (year). F_y captures fighting season fixed effects (equivalent of year in the study sample) and $X^V_{d,y}$ captures a vector of additional covariates, which we incorporate to address potential concerns about omitted variables. These covariates include the intensive margin of insurgent operations during the fighting, harvest, and planting seasons as well as supplemental measures of state capacity and insurgent intimidation and alternative fixed effects specifications. We expect the first coefficient β_1 to be negative.

5 Results

We begin by visualizing the data. In Figure 5, we plot the *p*-value distributions for indirect fire attacks against the corresponding opium revenue of each district-year (fighting season). Confidence regions (95%) are shaded in blue. Notice the consistently negative relationship between revenue and randomness of combat. This non-parametric correlation is consistent with our intuition that high capacity rebels produce patterns of violence that are less random and exhibit temporal clustering.

Figure 5: Bivariate relationship between opium revenue and *p*-value of randomization test of combat (indirect fire attacks) in Afghanistan



5.1 Baseline Results

We next turn to our regression-based evidence. In Table 1, we estimate Equation 1. Column 1 is a sparse model that demeans our bivariate correlation by fighting season. We find that a strong negative relationship between opium revenue and combat randomness, confirming our visual evidence. In Columns 2 through 4 we sequentially add controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. We do this to address potential concerns that our bivariate result is driven by a strong relationship between opium revenue and the intensity of violence during the fighting season (which covaries negatively with our randomness parameter). Wright (2016) finds evidence of this type of positive level shift in violence in Colombia.[14] It is also possible that there is a mechanical correlation between the intensity of combat activity and the test statistic if, for example, it is easier to detect clustering in the presence of a large number of events. Accounting for these channels leaves our results are largely unaffected, although our point estimate becomes marginally more precise in Column 2. In Column 3, we attempt to rule out concerns that our estimates are substantially biased by the endogenous relationship between opium production and insurgent violence during the harvest season, which might influence subsequent conflict during the later fighting months (Lind et al., 2014). It might also be the case that farmers are coerced into planting opium through violence exposure. We account for this potential source of bias by including a planting season violence trend in Column 4. This baseline evidence suggests a precise, consistent link between rebel capacity and randomization of indirect fire attacks.

Robustness Checks We conduct several other baseline robustness checks in Table 2. In Column 2, we introduce a province $\times$ fighting season fixed effect, which soaks up any residual mechanism design effects due the spatial sampling procedure utilized by the UNODC to estimate local opium production. This fixed effect also absorbs troop rotation schedules which coincide with the province-year, which includes force movement into and out of regional command posts. Although the magnitude and precision of our main effect declines marginally, the negative relationship persists even in this very demanding specification. Another potential concern one might have is that the strategy we use to account for the intensive margin of

[14]Theoretically, it is also possible that revenue shocks via the opium channel increase opportunity costs, which would cause a level reduction in violence. Gehring et al. (2019) present evidence consistent with this pattern, though the authors focus on violence in the subsequent calendar year rather than the fighting season that immediately follows post-harvest tax collection. We repeat this analysis, focusing on the fighting season, and find a large positive reduced form (Table A-7) and second stage result (Table A-8).

Table 1: Impact of rebel capacity on within-day randomization of indirect fire attacks

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0581***	-0.0588***	-0.0549***	-0.0549***
	(0.0137)	(0.0135)	(0.0125)	(0.0125)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.154	0.171	0.187	0.187

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects. Column 2-4 add controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

violence during the fighting season (i.e., partialling out the violence in levels) is incomplete. Another solution is to inversely weight our model along this margin. This implies that our corresponding estimate is less vulnerable to vertical (conflict) outliers. We present this result in Column 3, which is highly consistent with our baseline result.

In our baseline specification, we rely on a national time series in prices and year-by-year variation in opium yields (kilograms per hectare). Yet the interaction of weather conditions and soil suitability may lead to heterogeneous crop yields by region and year. Regional prices might also differ substantially. These two concerns represent classical error-in-variables, which we can correct with regional price and yield data compiled from UNODC records. In Column 4, we replicate our baseline specification in Column 1 using opium revenues calculated using regional yield rates and regional price data. It is also the case that some provinces have later harvests which occur after fighting has begun in other parts of the country. In Column 5, we use crop calendar maps produced by the UNODC to classify late harvest districts and exclude them from our sample in Column 1. Again, our results are highly consistent in both model specifications. In Column 6, we add a district fixed effect, to absorb in time-invariant district-specific unobservable characteristics that may influence opium production and combat tactics. In the main specification, we omit unit fixed effects because our main analysis is an unbalanced panel. Because of this structure our estimating sample is slightly reduced because we exclude singletons from the analysis. The main effect remains substantively large and precise. In Figure 6, we address potential concerns about spatial correlation in opium production. To do this, we introduce two spatial correlation-adjusted randomization inference tests, where randomization is stratified by province (A) and district (B). In each case, values are randomly drawn within each spatial strata across years and strata fixed effects are added to the model specification. Again, our benchmark effect is in the tail of each zero-centered distribution, suggesting that our main effects are unlikely to have occurred by random chance even in the presence of spatial correlation.

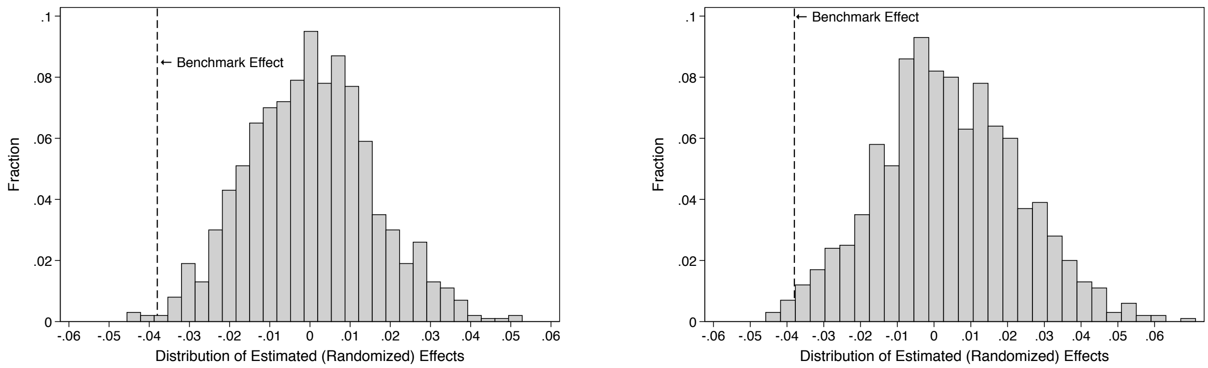
Table 2: Impact of rebel capacity on within-day randomization of indirect fire attacks, robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)
Opium Revenue	-0.0549***	-0.0450**	-0.0531***		-0.0543***	-0.0381**
	(0.0125)	(0.0220)	(0.0116)		(0.0126)	(0.0166)
Opium Revenue (Regional)				-0.0555***		
				(0.0128)		
MODEL PARAMETERS						
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes	Yes	Yes	Yes
ADDITIONAL PARAMETERS						
Province × Fighting Season FE	No	Yes	No	No	No	No
Weighted Least Squares	No	No	Yes	No	No	No
Regional Yield Adjust.	No	No	No	Yes	No	No
Early Harvest Only	No	No	No	No	Yes	No
District Fixed Effect	No	No	No	No	No	Yes
MODEL STATISTICS						
No. of Observations	600	600	600	600	588	563
No. of Clusters	154	154	154	154	150	117
R^2	0.187	0.379	0.188	0.184	0.184	0.503

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Additional Potential Threats to Identification In the online appendix, we address several other potential threats to identification. First, in Tables A-2 and A-3, we address potential concerns that the accumulation of rebel capacity which is observable to counterinsurgents may induce a strategic response during the subsequent fighting season. Heterogeneity in counterinsurgent investments correlated with our measure of revenue may complicate estimation. In principle, these factors may be bad controls due to their simultaneous or post

Figure 6: Randomization Inference to Evaluate Robustness of Benchmark Effect Adjusting for Spatial Correlation



(a) Full Panel Reshuffling, Province Strata

(b) Full Panel Reshuffling, District Strata

Notes: We randomly reshuffle the opium revenue vector in our data using two approaches ($\times$ 1000). We utilize observed opium revenue for the full panel of district-years and vary the spatial strata used for spatially constrained randomization ((a) province, (b) district). The model specification is equivalent to Table 1, Column 4, with strata fixed effects.

treatment allocation. Nonetheless, we confirm that our results remain robust if we account for six measures of counterinsurgent capacity: close air support missions, cache discoveries, IED neutralizations, detention of insurgent forces, counterinsurgent surveillance operations, and safe house raids yielding actionable intelligence assets (salary and recruitment logs, hard drives, forensic materials, etc.).[15] Second, opium production might also be influenced by coercive tactics used by the Taliban to intimidate civilians. This is problematic for estimation if coercion and combat tactics are correlated. In typical settings, these coercive tactics are unobserved or unrecorded. In our context, the military records we study track attempts to intimidate the civilian population, using methods like 'night letters' and shows of non-lethal force as well as deliberate killings of government collaborators (like informants and security force recruits). We present these results in Table A-4, which confirm the robustness of our main results.[16] Finally, to evaluate the cultivation component of revenue, we add one

[15]Another potentially relevant type of state capacity is opium eradication. Although government investments in eradication vary across years, UNODC reports suggest that verified opium eradication reduces national output by roughly 1 to 3% annually. Due to the small scale of these operations, we do not anticipate eradication meaningfully impacts Taliban revenue or tactics.

[16]In the online appendix, we present results from a Oster (2017) coefficient stability test, using the additional observables introduced in Tables A-2 through A-4. These results are introduced in Table A-5. Using the empirically-grounded specification suggested in Oster (2017), we set the δ parameter to 1 and R^{max} equal to 1.3$\times$ the R^2 observed in the fully controlled specification. Notice that the test produces two solutions,

hectare to non-producing districts before we evaluate the logarithm. It is possible that our core results are sensitive to this choice. Relatedly, districts that do not produce opium (zero revenue) may have an outsize influence on our estimated effects and statistical precision. We rule out these concerns in Tables A-9 and A-10, where we first exclude non-producing districts from our estimating sample and, conditional on excluding non-producers, replicate the Table 1 without adding a hectare.[17] The estimated effect of opium revenue once we exclude non-producers and/or adjust the measure of cultivation is larger than the main estimate and remains statistically precise. These results suggest our core estimates are unlikely to be driven by measurement of revenue or inclusion of district-years where no opium production takes place.

5.2 Instrumental Variables Approach

We next introduce an instrumental variables approach. We begin by estimating opium suitability using a combination of degree-day, precipitation-day, and soil quality characteristics. This approach treats production across the full district-year panel of available data as an outcome of interest and uses these agronomic characteristics to produce fitted values of expected productivity given observed district-year input availability. We then standardize these predictions based on exogenous agronomic conditions and weight these values by aggregate production in the prior year, which is correlated with present year prices.

5.2.1 Agronomic Suitability for Opium Production

We leverage data high resolution data on temperature, rainfall, and soil quality to construct a measure of opium suitability.

We gather daily, district-level temperature (Kelvin) and precipitation (mm) measures from reanalysis data. Our climatic data are drawn from the National Centers for Environmental Prediction (NCEP) and the Department of Energy, which prepared the baseline climate reanalysis by using state-of-the-art assimilation techniques. These data are derived from reanalysis (climate modeling) of underlying meteorological data. These techniques and the data generation processes are fully described in Saha et al. (2010), to which we direct interested readers. We then construct parameters capturing the number of days within each growing season these data fall within a particular set of binned ranges. Using this tech-

both of which are below zero, consistent with our main specification.

[17]Results in these two tables are highly consistent (though not exactly the same) because the constant term in Table A-9 absorbs nearly all of the level shift before the logarithm is evaluated.

nique enables us to flexibly account for non-linear relationships between weather conditions and agricultural productivity. We supplement this data with information from Food and Agriculture Organization's Harmonized World Soil Database, which we extract using the district-level cross section. We include nutrient availability, nutrient retention, rooting conditions, oxygen availability, excess soil salts, toxicity, and packedness and workability (which impacts the ability to manage fields). For each district, we calculate the percentage of land mass where these soil features present no or slight limitations to productivity (Class 1 under the FAO guidelines). Because various combinations of weather and soil conditions may produce high and low productivity zones in a complex system, we interact these measures with our degree-day and precipitation-day measures. We merge these data with our panel data on opium production and produce a standardized fitted value of opium productivity given these exogenous parameters. We use the least squares estimation equation below.

$$\begin{aligned} log(production_{d,y}+1) = \alpha + \sum_{i=1}^{7}(\vartheta_i Precip-Day_{d,y}) + \sum_{i=1}^{7}(\zeta_i Precip-Day^2_{d,y}) \\ + \sum_{i=1}^{10}(\eta_i Temp-Day_{d,y}) + \sum_{i=1}^{10}(\rho_i Temp-Day^2_{d,y}) + \sum_{i=1}^{7}(\mu_i SoilQual_d) \\ + \tau_{ij}\sum_{i=1}^{7}(Precip_{d,y}) \times \sum_{j=1}^{7}(SoilQual_d) + \upsilon_{ij}\sum_{i=1}^{7}(Precip^2_{d,y}) \times \sum_{j=1}^{7}(SoilQual_d) \\ + \phi_{ij}\sum_{i=1}^{10}(Temp-Day_{d,y}) \times \sum_{j=1}^{7}(SoilQual_d) + \psi_{ij}\sum_{i=1}^{10}(Temp-Day^2_{d,y}) \times \sum_{j=1}^{7}(SoilQual_d) \\ + \gamma X_y + \varepsilon_d \end{aligned} \tag{2}$$

Where $log(production_{d,y}+1)$ is the production (log) for a given district, d, and growing season, y (year). X_y captures growing season fixed effects. $Precip - Day_{d,y}$ and $Temp - Day_{d,y}$ capture the effect of our precipitation-day and degree-day (temperature-day) parameters. See Figure A-1 for the binned ranges used in the analysis. We also include the square of this counts. $SoilQual_d$ captures the soil quality features noted above. We then fully interact these base terms. From this regression, we produce $log(\widehat{production}_{d,y}+1)$, which is our unstandardized fitted value. Denote this value as $\Lambda_{d,y}$. We standardize this value using the following expression:

$$suitability_{d,y} = \frac{\Lambda_{d,y} - \bar{\Lambda}_{d,y}}{var(\Lambda_y)^{-1}} \tag{3}$$

$suitability_{d,y}$ is demeaned and standardized with respect to the standard deviation of the fitted values. This approach is most similar to Mejía and Restrepo (2014), who use land features and soil characteristics to predict coca production in Colombia. The primary difference between our two methods is the use of high frequency climatic inputs as well as the use of interactions to capture heterogeneous climatic effects via soil quality conditions. In supplemental results, we use machine learning for vector reduction, in line with Rozenas and Zhukov (Forthcoming). National price variation is not driven by any single district, which means that all districts are effectively price-takers. Although not strictly necessary in our case for identification, UNODC reports suggest the primary driver of current prices is aggregate (national-level) production in the prior year. Naturally, increased aggregate production from the prior year drives down national prices in the subsequent year. This gives us an opportunity to instrument for the price component of revenue (once we invert the value), which we implement below.[18]

5.2.2 Plausibility of Identifying Assumptions

We anticipate the exclusion restriction is plausibly satisfied for several reasons. First, we use growing season agronomic conditions, which are observed months ahead of the beginning of the fighting season. Given our non-parametric specification, which uses degree-day and precipitation-day measures and soil suitability, it would be difficult to articulate a clear pathway through which our measures of weather conditions directly impact combat timing. Second, even if such a violation was econometrically plausible and induced non-trivial bias, we expect this bias would be largely absorbed by accounting for variation in the levels of combat activity during the fighting season, which is one of our benchmark parameters. For a violation of the exclusion restriction to significantly bias our results, our growing season agronomic inputs would have to have a persistent effect, that passes through our non-parametric specification and subsequent dimensionality reduction, and is not absorbed by our included covariates. We believe this is unlikely to be a substantial source of bias.

One plausible channel through which the independence assumption might be violated is if, conditional on observing fluctuations in agronomic suitability, government forces strategically reallocate aid projects. If these aid projects effectively reduce opium production (via 'alternative livelihoods' programming) and enhance local employment, our second stage effects are biased towards zero due to increasing reservation wages. It is also possible, however, that rebels are able to capture some of the rents associated with aid projects. This could bias our estimates upward, since the instrument potentially captures variation in opium revenue

[18]The base term of aggregate production is absorbed in our fixed effects.

correlated with both revenue channels. We address these concerns in the Online Appendix using declassified data from the Commander's Emergency Response Program (CERP). In Table A-11, we replicate the benchmark model and include measures of aid intensity overall and with respect to agricultural and irrigation projects only (dollars of aid delivered during the growing season, ln). The main effects are slightly larger when incorporating overall aid and slightly attenuated when limiting our focus to agricultural aid, neither of which are statistically different from our benchmark result.

In Table A-12, we find strong evidence of a positive correlation between our suitability index and annual revenue from the opium trade. Because we are accounting for potential heteroskedasticity by district, we produce Kleibergen-Paap F statistic for our excluded instrument, which is well above 10. This suggests our instrument is relevant and strong. In Table A-15, we investigate whether the monotonicity assumption is plausible with respect to irrigation. This channel is relevant since our instrument captures exogenous variation in suitability. How revenue responds (via production) to suitability may be a function of technologies used to enhance productivity under otherwise poor agronomic conditions. For example, it is possible that districts with varying levels of irrigation access (via canals and other technology) will comply with the instrument at different rates. To assess monotonicity, we gather data collected by FAO prior to the US invasion documenting irrigated sites. We use this data to classify districts along the 75, 90, and 95 percentiles for the main estimating sample and full panel. We find no clear evidence of weakened compliance with the instrument via the irrigation mechanism.[19]

5.2.3 IV Results

We next turn to our main IV results. We report our second stage estimates in Table 3. These results suggest the effect of revenue on temporal clustering of combat operations is substantially larger than our least squares estimates. One potential explanation for the increasing magnitude of the effect relative to the baseline OLS specification is the presence of cross-cutting dynamics with respect to counterinsurgent operations that we do not observe. If counterinsurgents observe variation in opium production (or some endogenous subset), they may strategically reallocate their forces or adjust their combat strategies. If these changes in government activity also make it more difficult to acquire intelligence or mobilize fighters, the per-unit change in rebel tactics due to opium revenue will be lower (i.e., biased

[19]The corresponding second stage estimates are also statistically indistinguishable across these model specifications (relative the benchmark).

towards zero). While we observe some types of military operations, others that are perhaps more relevant to rebel attack timing are unobserved (e.g., secret or top secret missions to eliminate leaders). Our instrument may capture the subset of variation in opium production that counterinsurgents find most difficult to anticipate, since it is unrelated to more readily available information about rebel activity. Stated differently, our instrument perhaps helps us identify opium production that is least predictable from a tactical military perspective. Exploiting variation in revenue that is exogenous to government countermeasures therefore increases the magnitude of the estimated effect of revenue on tactics.[20]

Table 3: Impact of rebel capacity on within-day randomization of indirect fire attacks, instrumental variables approach (second stage)

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0793***	-0.0787***	-0.0741***	-0.0741***
	(0.0271)	(0.0268)	(0.0247)	(0.0247)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.136	0.155	0.173	0.173
IV Specification				
IV Type	Benchmark	Benchmark	Benchmark	Benchmark
Kleibergen-Paap F Statistic	190.9	188.4	188.9	187.4

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year instrumented using an opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We next address three potential concerns about our IV approach. First, it is possible that our estimation of opium suitability is overly saturated by a large number of agronomic inputs. We do not anticipate that this represents an inferential concern. We allow for a large number of input parameters because it enables our model of opium productivity to be flexibly non-linear. This approach is common precisely because it limits the number of

[20]Naturally, we do not have data on unobserved missions, which could be used to further evaluate this potential mechanism for the shift in our estimated coefficients.

functional form assumptions needed when predicting output at a granular level. This is also ideal since we lack a well-developed agricultural model of poppy plants that is specific to the context we study. That said, one alternative to our baseline approach is to use LASSO estimation to reduce the number of input parameters in our model. By design, this limits against oversaturation of the suitability model by eliminating a large number of potentially irrelevant agronomic inputs. We reproduce our second stage results using LASSO-based instrument in Table 4. Our F statistics suggest the IV remains strong. Compared with our benchmark IV results, the main effect is slightly attenuated, though the two point estimates are statistically indistinguishable.

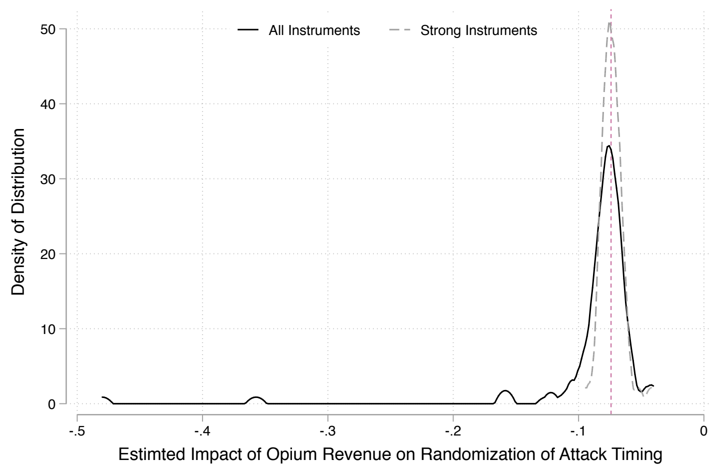
Table 4: Impact of rebel capacity on within-day randomization of indirect fire attacks, instrumental variables approach (second stage) using LASSO selection in suitability index estimation

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0759***	-0.0757***	-0.0701***	-0.0701***
	(0.0270)	(0.0266)	(0.0248)	(0.0248)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.142	0.160	0.178	0.178
IV Specification				
IV Type	Lasso	Lasso	Lasso	Lasso
Kleibergen-Paap F Statistic	114.7	113.2	117.9	118.2

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year instrumented using an opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A second potential concern is that our suitability model uses all available data on opium production. Put differently, we have not withheld any elements of the raw data when constructing our index. Without holding some of our sample back, it is difficult to assess whether some subset of the full panel is driving our estimated productivity. To address this concern, we repeatedly estimate suitability (100 times), randomly holding back a subset (75%) of the full panel. We then reproduce our second stage estimates. These results are visualized in Figure 7. Notice that the second stage coefficient of our benchmark IV approach, noted as a vertical line, is the median estimate of the full distribution of coefficients from the resampling technique. Indeed, all second stage estimates based on sufficiently strong first stages (where the F statistic is at least 10) are within .025 of the main coefficient. This suggests that our main result is representative of the range of results that could be estimated given a large number of arbitrary samples used to produce our suitability measure.

Figure 7: Distribution of second stage IV estimates using random resampling technique to produce alternate suitability instruments. Main effect noted with vertical line.



Yet a third possibility is that we could simply use the raw inputs for our suitability index to instrument for opium revenue. This would effectively bypass the intermediate step where we use raw productivity and agronomic inputs to estimate the district-season production frontier (opium suitability). The central motivation for our main specification, which uses a single estimate of opium suitability as an instrument for revenue, is to avoid the use of many, potentially weak instruments. Alternatively, we could limit our first stage to the baseline degree-day and precipitation-day agronomic inputs as instruments, and reestimate our second stage results. We do this in Table 5. This approach allows us to address potential concerns about misspecified standard errors in the first stage of the benchmark model, which uses an estimated quantity as one component of the instrument,[21] and enables us to explore the plausibility of the estimated effects of agronomic conditions on opium production (via revenue). In Table 5, notice that our point estimates, while slightly attenuated relative

[21] We thank Yuri Zhukov for suggesting this clarification.

our benchmark IV specification, are statistically indistinguishable from the baseline results. The relevant F statistics are above 40, suggesting that our first stage is sufficiently strong to estimate meaningful second stage effects. In the Online Appendix, we report the first stage effects of agronomic inputs on opium revenue. These results (see Figure A-1 below) are highly consistent with qualitative evidence from the UNODC, which suggests opium in Afghanistan grows optimally in the presence of water access and warm conditions. Freezing conditions during the growing season, associated with ground frost and snow cover, also negatively impact productivity (see coefficient estimated for degree-days below 275 Kelvin (equivalent to approximately 35 degrees Fahrenheit or 1 degree Celsius).[22]

Table 5: Impact of rebel capacity on within-day randomization of indirect fire attacks, instrumental variables approach (second stage) using agronomic inputs as instrumental variables

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0681***	-0.0679***	-0.0617***	-0.0617***
	(0.0203)	(0.0197)	(0.0188)	(0.0188)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.150	0.168	0.185	0.185
IV Specification				
IV Type	Agronomic	Agronomic	Agronomic	Agronomic
Kleibergen-Paap F Statistic	44.56	42.48	44.48	44.43

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year instrumented using degree-day (temperature-day) and precipitation-day instruments. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

By leveraging 'as if' random variation in high frequency, microlevel opium suitability, the battery of instrumental variable approaches above give us more confidence that opium revenue influences temporal clustering in rebel attacks. Our IV estimates help us sidestep potential concerns about endogenous variation in revenue driven by the production of violence prior to the beginning of the fighting season.

[22]See Kienberger et al. (2017) and the 2014 UNODC annual Opium Survey for more details regarding suitable agronomic conditions for opium production in Afghanistan.

6 Mechanisms

Having addressed numerous potential threats to identification, we now turn our attention to a more direct test of our model's empirical implications. Our results could be working through at least two plausible mechanisms. First, revenue from the opium trade could give field commanders more flexibility to recruit and arm fighters. This would enable them to engage in combat operations where the timing of attacks is less random and, therefore, easier for state rivals to anticipate and engage in strategic adjustment. As such, battlefield losses are likely. These losses are more easily absorbed if a given rebel division has more, better armed combatants. Second, increasing capacity could make it easier for insurgents to deploy spies or buy information about troop movement patterns from civilians. Responding strategically to intelligence reports about time windows within which troops and bases are vulnerable could lead to temporal clustering. Importantly, these are not rival mechanisms. They can operate simultaneously.

6.1 Intelligence Gathering by Rebels

Our data provides us with a unique opportunity to investigate the second mechanism more precisely. In particular, we expect that conditions that make intelligence gathering easier (i.e., reduced frictions) should enhance the negative effect of opium revenue on randomness of combat. That is, in places where rebels have the ability to gather information about troop and base vulnerabilities, temporal clustering should be even more responsive to rents extracted from the opium trade.

We begin with the simplest test of this mechanism. We gather administrative data on the distribution of ethnic groups across districts. We identify districts where 95% or more of population settlements are classified as Pashto speaking. Pashtuns are Taliban coethnics and form the primary base of civilian support for the insurgency. In principle, we expect it will be easier for insurgents to acquire information about government activity in districts dominated by their coethnics. We present these results in Table 6. Notice that the additive effect of opium revenue in coethnic zones is much larger (roughly 80%) than non-coethnic districts. This is evidence consistent with our story, but relying on coethnicity as a test of our argument may conflate intelligence frictions with a range of other factors as well.

We overcome this concern by turning to more sophisticated tests of our argument. In particular, our military records include previously unreleased information about rebel surveil-

Table 6: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to potential intelligence gathering via coethnics

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0150**	-0.0178**	-0.0146*	-0.0146*
	(0.0137)	(0.00717)	(0.00757)	(0.00766)	(0.00766)
Coethnicity		0.246	0.288	0.378	0.378
		(0.333)	(0.329)	(0.351)	(0.352)
Coethnicity × Revenue		-0.0582***	-0.0554***	-0.0552***	-0.0552***
		(0.0175)	(0.0170)	(0.0168)	(0.0168)
MODEL PARAMETERS					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
MODEL STATISTICS					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.193	0.204	0.218	0.218

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

lance of troop movement and base activity as well as data on security breaches, which occur when insurgents are able to effectively infiltrate the outer perimeter of targets and observe activity from within bases and outposts. These yield another potential test of our information mechanism by tracking incidents where the Taliban are able to 'turn' security recruits and use them to launch attacks from within army units. In each of these cases, we are able to employ much more direct evidence of the ability of the insurgency to gather precise information about target defenses. We test our mechanism using these data in Columns 2 through 5 in Table 7. In Column 2, we interact our measure of the spy network operations with revenue.[23] In the Online Appendix, we repeat this specification with measures of infiltration and insider attacks (see Tables A-23 and A-24). We find strong evidence that our main effect is enhanced in districts where the insurgents have a demonstrated capability to conduct surveillance, infiltrate security installations, and launch insider attacks. These findings yield evidence consistent with the mechanism implied by our theoretical argument and suggest that intelligence gathering may be a primary pathway through which shocks to

[23] To avoid potential concerns about the endogeneity of intelligence gathering, we identify the cross section of districts with these characteristics using only the first year of our sample, 2006.

rebel capacity influence the timing of violent attacks.

Table 7: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to potential intelligence gathering by spies

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0207***	-0.0224***	-0.0194***	-0.0194***
	(0.0137)	(0.00669)	(0.00661)	(0.00663)	(0.00664)
Surveillance		0.605*	0.758**	0.952**	0.953**
		(0.320)	(0.337)	(0.405)	(0.400)
Surveillance × Revenue		-0.0671***	-0.0669***	-0.0678***	-0.0678***
		(0.0193)	(0.0191)	(0.0190)	(0.0191)
MODEL PARAMETERS					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
MODEL STATISTICS					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.206	0.219	0.233	0.233

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6.2 Labor Scarcity

We suggest that revenue shocks may impact combat tactic through the ability to mobilize fighters. Unfortunately, we lack sufficiently granular census data on the size of Taliban subunits to fully assess the direct effects of budget constraints on recruitment and deployment. However, we can investigate this mechanism through an alternative approach: battlefield losses experienced during combat operations. The intuition of this test is in line with the mechanism we theorize. As insurgents accumulate resources, they can mobilize more fighters, enabling them to adopt tactics that would otherwise be too costly from a human capital perspective. If attacks are predictably launched, government forces can deploy relatively more effective countermeasures, potentially neutralizing some subset of fighters through casualties. For relatively weak insurgents, these costs may be too great, leading them to conduct attacks that are randomized and more difficult to thwart.

Although we cannot directly examine whether revenue shocks lead to expansion of rebel forces, we can study whether labor constraints (via lost fighters) lead to a change in tactics. The intensity of potential battlefield losses is increasing in combat activity, which our model accounts for in levels via fighting season activity for indirect fire attacks. However, because battlefield losses are most common in direct combat (when rebels are engaged in close combat with military units), we introduce an additional parameter to our benchmark model capturing the overall intensity of combat activity across the three primary categories of combat. Adding this parameter allows us to partial out any variation in labor scarcity (casualties) due to the intensity of combat operations, yielding plausibly abnormal variation in battlefield losses. We present these results in Table 8. Notice that battlefield losses (positive coefficient) lead to an increase in randomization. The estimated effect on labor scarcity suggests a one standard deviation increase in losses leads to a .12 standard deviation increase in randomization.

Table 8: Effects of rebel capacity and battlefield losses on within-day randomization of indirect fire attacks

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0191***	-0.0222***	-0.0209***	-0.0201***	-0.0200***
	(0.00574)	(0.00660)	(0.00686)	(0.00700)	(0.00698)
Battlefield Losses		0.0456**	0.0486**	0.0485**	0.0486**
		(0.0230)	(0.0211)	(0.0207)	(0.0207)
Model Parameters					
Fighting Season (FS) Fixed Effect	Yes	Yes	Yes	Yes	Yes
FS Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
FS Combat Operations (all, levels)	Yes	Yes	Yes	Yes	Yes
Model Statistics					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.496	0.501	0.504	0.506	0.506

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Fighting season parameters are notated with the abbreviation FS. Battlefield losses in our sample have a mean of 4.605 and standard deviation of 10.547. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A natural concern is that battlefield losses, even after conditioning on combat activity (in levels), are not sufficiently ‘as if’ random to draw causal inferences. We attempt to address this concern in the online appendix, where we add our measures of counterinsurgent operations (from Tables A-2 and A-3), which are the most plausible omitted variables in this regression. We introduce these results in Table A-25. Once we account for these omitted factors, our results get larger in magnitude, giving us more confidence in the main results in Table 8.

6.3 Government Surveillance

We conclude the discussion of the mechanisms by returning to the role of surveillance. We argue that revenue generation by insurgents eases constraints on intelligence gathering, making it easier for rebels to identify time periods during which government forces are vulnerable. If, on the other hand, government forces are capable of and conduct operations to monitor rebel activity, they may be more likely to identify patterns in rebel attacks. Information about when insurgents are likely to launch attacks as well as other features of militia operations enables government forces to directly thwart rebel operations and disrupt their combat planning. Our data gives us another unique opportunity to explore this channel: we observe where and how frequently government forces are actively monitoring insurgent tactics and procedures during the fighting season. We replicate the model above, accounting for the confounding correlation between combat activity and the allocation of surveillance assets. These results are presented in Table 9. Notice that, as our theoretical model anticipates, increased government monitoring is associated with more attack randomization (less temporal clustering).

7 Extensions

In this section, we study several extensions which clarify the industrial organization of rebellion and yield potentially actionable insights about insurgent operations.

7.1 Reservation Wages

To study the relative impact of reservation wages, we take advantage of an important feature of the context of our study: differential taxation. At baseline, we anticipate Taliban fighters collect a flat (typically 10%) tax on opium proceeds. Qualitative documents suggest the bureaucracy established to manage these transactions is sophisticated and repeated taxation ('double taxation') and excessive taxation are prohibited. However, the Taliban cannot fully

Table 9: Effects of rebel capacity and government surveillance missions on within-day randomization of indirect fire attacks

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0191***	-0.0203***	-0.0190***	-0.0181***	-0.0181***
	(0.00574)	(0.00601)	(0.00623)	(0.00646)	(0.00644)
Government Surveillance Operations		0.0916***	0.0917***	0.0932***	0.0930***
		(0.0209)	(0.0215)	(0.0210)	(0.0211)
Model Parameters					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
FS Combat Operations (all, levels)	Yes	Yes	Yes	Yes	Yes
Model Statistics					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.496	0.501	0.504	0.506	0.506

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Fighting season parameters are notated with the abbreviation FS. Government Surveillance in our sample have a mean of .425 and standard deviation of 3.22. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

control how much local politicians—agents of the formal government—extract as informal taxes (Giustozzi, 2009; Mansfield, 2016). In the absence of these secondary taxes, gains from the opium trade translate into simultaneous growth in rebel capacity and growth in the local economy. As such, resources available to compensate fighters and spies *and* local reservation wages may be increasing in revenue. This will lead to a cross-cutting effect that biases our estimate towards zero, since increasing reservation wages offset some of the impact of revenue growth on mobilization, recruitment, and, by extension, combat tactics. To better disentangle these effects, we need to identify locations where informal taxes absorb some of the growth in the local economy which influences reservation wages.

We do this by taking advantage of proprietary military surveys provided to the authors by the North Atlantic Treaty Organization (NATO).[24] These surveys enable us to quantify

[24] We utilize the Afghanistan Nationwide Quarterly Research (ANQAR) survey. ACSOR, an Afghan subsidiary of the international firm D3, was contracted to design and field the survey. ACSOR hired and trained local enumerators in household and respondent selection, how to correctly record answers to questions, culturally sensitive interview methods, and secure storage of contact information. The administrative district (the cross section unit in this study) is the primary sampling unit. These sampling units are selected via probability proportional to size systematic sampling approach. After districts have been sampled, secondary

the level of corruption present in local administrative agencies across Afghanistan.[25]

We anticipate that revenue growth will have larger effects in places with more severe corruption, where reservation wages grow at a slower rate due to informal taxation. We present these results in Table 10. Consistent with our expectations, in districts with corruptible politicians where reservation wages remain relatively lower, we see larger marginal effects over the baseline condition (with relatively less informal taxation).

Table 10: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to variation in reservation wages (via informal taxation by corrupt officials)

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0177**	-0.0269***	-0.0268***	-0.0270***
	(0.0137)	(0.00688)	(0.00745)	(0.00750)	(0.00740)
Corruptible Officials		-0.215	-0.676*	-0.669	-0.681*
		(0.358)	(0.390)	(0.405)	(0.405)
Corruptible × Revenue		-0.0430***	-0.0329**	-0.0291**	-0.0288**
		(0.0152)	(0.0133)	(0.0125)	(0.0124)
Model Parameters					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
Model Statistics					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.164	0.182	0.196	0.196

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

sampling units composed of villages and settlements are randomly selected. A random walk method is used to identify target households and a Kish grid is used to randomize the respondent within each selected household. After the sampling set has been identified and before fielding a survey wave, ACSOR engages with local elders to secure permission for enumerators to enter sample villages. See Condra and Wright (Forthcoming) for more details and diagnostics.

[25]We split districts into low versus high corruption categories based on the percentage of respondents across all available waves which describe corruption in government as a very serious problem. If more than one third report corruption as very serious, we classify the district as corrupt (and local administrators as corruptible). This equates to the tenth percentile of the overall distribution. Results are consistent if we use a lower threshold.

7.2 Savings Technologies

We next consider whether local rebel units utilize savings technologies to engage in consumption smoothing. Revenue from the opium trade can be volatile, with droughts, crop diseases, and other unpredictable factors negatively impacting productivity from season to season. In response, units within the Taliban may engage in consumption smoothing, holding back some of their fighting capacity for the next season in expectation of uncertainty regarding future revenue. To assess this dynamic, we use our full panel data on revenue to calculate income volatility for each district. We then split districts into high and low volatility clusters.[26] If rebels engage in consumption smoothing, we would expect combat tactics to be less responsive per unit of income growth in high volatility districts. We estimate these heterogeneous effects in Table 11. Notice that the high income volatility does not lead to a marginally smaller effect on combat. This suggests that rebels do not actively engage in smoothing across seasons; rebels likely exhaust available resources during the fighting season immediately following tax collection.

8 Conclusion

Rebel tactics are an overlooked feature of internal warfare. Understanding how these conflicts are fought and the strategic responses of armed actors to revenue shocks is on equal footing with thoroughly examined questions about the causes of civil war and the factors that influence when hostilities end.

We argue that shocks to rebel capacity influence the timing of attacks. We develop a simple model of combat during an irregular insurgency. Rebels optimize when they conduct attacks after observing imperfect signals of government defensive maneuvers. As insurgents accumulate resources through taxation, their budget constraint is relaxed and they can recruit more fighters, acquire more armaments, and gather more intelligence about the vulnerability of troops and bases. The labor supply of fighters, coupled with surplus capital, allows rebels to conduct more attacks. Infiltrating military installations and conducting surveillance of troop movement enhances the quality of information rebels have about specific times when attacks will yield the highest probability of success. Thus, revenue shocks will lead to clustering in the temporal distribution of violence. Our model also suggests that

[26]We classify districts into these categories using the median level of volatility observed among opium producing districts in the full panel.

Table 11: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to variation in income volatility

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0583***	-0.0588***	-0.0550***	-0.0551***
	(0.0137)	(0.0146)	(0.0143)	(0.0121)	(0.0121)
High Revenue Volatility		-0.890*	-0.851*	-0.864*	-0.866*
		(0.495)	(0.477)	(0.489)	(0.488)
High Revenue Volatility × Revenue		-0.00215	-0.00253	-0.00210	-0.00198
		(0.0336)	(0.0324)	(0.0340)	(0.0343)
Model Parameters					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
Model Statistics					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.164	0.180	0.196	0.196

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

these tactical shifts will be greatest when armed actors have the institutional capacity to pin-point when such attacks are least likely to be neutralized by government defenses.

We test the empirical implications of our theoretical model using data collected during Operational Enduring Freedom in Afghanistan. The temporal precision of our military records enable us to develop a sophisticated methodology for differentiating the within-day timing of rebel operations from randomized combat. This method produces a likelihood parameter that quantifies the degree of temporal clustering present in the allocation of insurgent attacks at the district level, broken down by fighting season. We couple this novel approach with high resolution estimates of opium production and market prices. We leverage the industrial organization of the Taliban, including their highly institutionalized taxation system, to estimate the impact of local revenue from the drug trade on combat tactics in the subsequent fighting season.

Our evidence suggests that the randomness of attacks significantly decreases as rebels accumulate more resources from the opium trade. As rebels accumulate fighting capacity, their attacks become temporally clustered around particular time windows in the day. This core result survives a number of robustness checks, including accounting for trends in rebel violence during the fighting, harvest, and planting seasons, which may influence the intensity of opium cultivation from year to year. The richness of our microdata also enables us to rule out additional concerns about the positive covariance between rebel capacity and strategic reallocation of government forces.

Our data also yields a unique opportunity to assess the underlying mechanism suggested by our model: intelligence gathering. Temporal clustering occurs because rebels use their resources to improve the quality and precision of their information about target vulnerability. This type of mechanism is largely unobserved and difficult to disentangle from alternative explanations. Our military records, however, include detailed information about where rebels were observed conducting surveillance operations as well as instances where insurgents were able to infiltrate government installations and conduct insider attacks. The capacity to gather intelligence substantially enhances the baseline effect of revenue shocks on combat operations.

References

Abadie, Alberto, "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 2002, *97* (457), 284–292.

_ **and Javier Gardeazabal**, "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, March 2003, *93* (1), 113–132.

Bazzi, Samuel and Christopher Blattman, "Economic shocks and conflict: evidence from commodity prices," *American Economic Journal: Macroeconomics*, 2014, *6* (4), 1–38.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development: Evidence from a Field Experiment in Afghanistan," in "APSA 2013 Annual Meeting Paper" 2013.

Berman, Eli and Aila M. Matanock, "The Empiricists' Insurgency," *Annual Review of Political Science*, 2015, *18*, 443–64.

_ **, Michael Callen, Joseph H. Felter, and Jacob N. Shapiro**, "Do Working Men Rebel? Insurgency and Unemployment in Afghanistan, Iraq, and the Philippines," *Journal of Conflict Resolution*, 2011, *55*, 496–528.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig, "This Mine Is Mine! How Minerals Fuel Conflicts in Africa," *American Economic Review*, June 2017, *107* (6), 1564–1610.

Berrebi, Claude and Esteban F Klor, "Are voters sensitive to terrorism? Direct evidence from the Israeli electorate," *American Political Science Review*, 2008, *102* (3), 279–301.

Besley, Timothy and Hannes Mueller, "Estimating the Peace Dividend: The Impact of Violence on House Prices in Northern Ireland," *American Economic Review*, April 2012, *102* (2), 810–33.

_ **and Torsten Persson**, "State capacity, conflict and development," *Econometrica*, 2010, *78*, 1–34.

_ **and** _ , "The Logic of Political Violence," *The Quarterly Journal of Economics*, 2011, *126* (3), 1411–1445.

Billon, Philippe Le, "The Political Ecology of War: Natural Resources and Armed Conflict," *Political Geography*, 2001, *20*, 561–584.

Blackett, D. W., "Pure strategy solutions of blotto games," *Naval Research Logistics Quarterly*, June 1958, *5* (2), 107–109.

Bo, Ernesto Dal and Pedro Dal Bo, "Workers, Warriors, and Criminals: Social Conflict in General Equilibrium," *Journal of the European Economic Association*, 2011, *9* (4).

Borel, Emil, "La theorie du jeu les equations integrales a noyau symetrique. Comptes Rendus del Academie, 173, 1304-1308; English translation by Savage, L. (1953). The theory of play and integral equations with skew symmetric kernels.," *Econometrica*, 1953, *21*, 97–100.

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

Callen, Michael, Mohammad Isaqzadeh, James D. Long, and Charles Sprenger, "Violence and Risk Preference: Experimental Evidence from Afghanistan," *American Economic Review*, 2014, *104* (1), 123–148.

Carter, David, "When terrorism is evidence of state success: securing the state against territorial groups," *Oxford Economic Papers*, 2015, *2014 (advance access)*, 1–17.

Collier, Paul and Anke Hoeffler, "Greed and grievance in civil war," *Oxford Economic Papers*, 2004, *56* (563-595).

Condra, Luke and Austin Wright, "Civilians, Control, and Collaboration during Civil Conflict," *International Studies Quarterly*, Forthcoming.

Condra, Luke N. and Jacob N. Shapiro, "Who Takes the Blame? The Strategic Effects of Collateral Damage," *American Journal of Political Science*, 2012, *56* (1), 167–187.

_ , James D. Long, Andrew C. Shaver, and Austin L. Wright, "The Logic of Insurgent Electoral Violence," *American Economic Review*, November 2018, *108* (11), 3199–3231.

Dube, Oeindrilla and Juan Vargas, "Commodity price shocks and civil conflict: evidence from Colombia," *Review of Economic Studies*, 2013, *80*, 1384–1421.

Esteban, Joan and Debraj Ray, "Linking Conflict to Inequality and Polarization," *American Economic Review*, June 2011, *101* (4), 1345–74.

_ , Laura Mayoral, and Debraj Ray, "Ethnicity and Conflict: An Empirical Study," *American Economic Review*, June 2012, *102* (4), 1310–42.

_ , Massimo Morelli, and Dominic Rohner, "Strategic Mass Killings," *Journal of Political Economy*, 2015, *123* (5), 1087–1132.

Fearon, James and David Laitin, "Ethnicity, insurgency and civil war," *American Political Science Review*, 2003, *97* (1), 75–90.

Galula, David, *Counterinsurgency Warfare: Theory and Practice*, New York: Praeger, 1964.

Gehring, Kai, Sarah Langlotz, and Stefan Kienberger, "Stimulant or Depressant? Resource-Related Income Shocks and Conflict," 2019.

Gennaioli, Nicola and Hans-Joachim Voth, "State Capacity and Military Conflict," *The Review of Economic Studies*, 2015, *82* (4), 1409–1448.

Getmansky, Anna and Thomas Zeitzoff, "Terrorism and voting: The effect of rocket threat on voting in Israeli elections," *American Political Science Review*, 2014, *108* (3), 588–604.

Giustozzi, Antonio, *Empires of Mud: wars and warlords in Afghanistan*, New York: Columbia University Press, 2009.

Golman, Russell and Scott E. Page, "General Blotto: games of allocative strategic mismatch," *Public Choice*, Mar 2009, *138* (3), 279–299.

Gould, Eric D. and Esteban F. Klor, "Does Terrorism Work?*," *The Quarterly Journal of Economics*, 2010, *125* (4), 1459–1510.

Goyal, Sanjeev and Adrien Vigier, "Attack, Defence, and Contagion in Networks," *Review of Economic Studies*, 2014, *81* (4), 1518–1542.

Hidalgo, Daniel, Suresh Naidu, Simeon Nichter, and Neal Richardson, "Economic determinants of land invasions," *Review of Economics and Statistics*, 2010, *92* (3), 505–523.

Hillion, Erwan and Oliver Johnson, "A proof of the SheppOlkin entropy concavity conjecture," *Bernoulli*, 11 2017, *23* (4B), 3638–3649.

Jia, Ruixue, "Weather shocks, sweet potatoes and peasant revolts in historical China," *The Economic Journal*, March 2014, pp. 92–118.

Kalyvas, Stathis N., *The Logic of Violence in Civil War*, New York: Cambridge University Press, 2006.

Kienberger, Stefan, Raphael Spiekermann, Dirk Tiede, Irmgard Zeiler, and Coen Bussink, "Spatial risk assessment of opium poppy cultivation in Afghanistan: integrating environmental and socio-economic drivers," *International Journal of Digital Earth*, 2017, *10* (7), 719–736.

Konig, Michael D., Dominic Rohner, Mathias Thoenig, and Fabrizio Zilibotti, "Networks in Conflict: Theory and Evidence From the Great War of Africa," *Econometrica*, July 2017, *85*, 1093–1132.

Konrad, Kai A. and Dan Kovenock, "Multi-battle contests," *Games and Economic Behavior*, 2009, *66* (1), 256–274.

Kovenock, Dan and Brian Roberson, "Conflicts with Multiple Battlefields," in Michelle R. Garfinkel and Stergios Skaperdas (eds.). Oxford Handbook of the Economics of Peace and Conflict, Oxford University Press, New York. pp. 503-531.," 2012.

Kress, Moshe, "Modeling Armed Conflicts," *Science*, 2012, *336* (6083), 865–869.

Limodio, Nicola, "Terrorism Financing, Recruitment and Attacks: Evidence from a Natural Experiment," *Chicago Booth Research Paper No. 32*, 2019.

Lind, Jo Thori, Karl Ove Moene, and Fredrik Willumsen, "Opium for the masses? Conflict-induced narcotics production in Afghanistan," *Review of Economics and Statistics*, 2014, *96* (5), 949–966.

Mansfield, David, *A State Built on Sand: How Opium Undermined Afghanistan*, Oxford: Oxford University Press, 2016.

Mejía, Daniel and Pascual Restrepo, "Bushes and bullets: illegal cocaine markets and violence in Colombia," 2014. Working paper.

Miguel, Edward, Shanker Satyanath, and Ernest Sergenti, "Economic shocks and civil conflict: an instrumental variables approach," *Journal of Political Economy*, 2004, *112* (4), 725–753.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2017, *0* (0), 1–18.

Padro i Miquel, Gerard and Pierre Yared, "The Political Economy of Indirect Control," *The Quarterly Journal of Economics*, 2012, *127* (2), 947–1015.

Peters, Gretchen S., *The Taliban and the Opium Trade*, New York: Columbia University Press,

Powell, Robert, "Allocating Defensive Resources with Private Information about Vulnerability," *American Political Science Review*, 2007, *101* (4), 799–809.

_ , "Defending against Terrorist Attacks with Limited Resources," *American Political Science Review*, 2007, *101* (3), 527–541.

_ , "Monopolizing Violence and Consolidating Power," *Quarterly Journal of Economics*, 2013, pp. 807–859.

Roberson, Brian, "The Colonel Blotto game," *Economic Theory*, September 2006, *29* (1), 1–24.

_ **and Dmitriy Kvasov**, "The non-constant-sum Colonel Blotto game," *Economic Theory*, 2012, *51* (2), 397–433.

Ross, Michael, "What do we know about natural resources and civil war?," *Journal of Peace Research*, 2004, *41*, 337–356.

Rozenas, Arturas and Yuri Zhukov, "Mass Repression and Political Loyalty: Evidence from Stalins Terror by Hunger," *American Political Science Review*, Forthcoming, p. 115.

Saha, Suranjana, Shrinivas Moorthi, Hua-Lu Pan, Xingren Wu, Jiande Wang, and et al., "The NCEP Climate Forecast System Reanalysis," *Bulletin of the American Meteorological Society*, 2010, *91* (8), 1015–1057.

Sexton, Renard, "Aid as a Tool against Insurgency: Evidence from Contested and Controlled Territory in Afghanistan," *American Political Science Review*, 2016, *110* (4), 731–749.

Shaver, Andrew and Austin Wright, "Data on Combatant Activity During Afghanistan War Advance Scientific Investigation of Insurgency," 2017.

Shayo, Moses and Asaf Zussman, "Conflict and the Persistence of Ethnic Bias," *American Economic Journal: Applied Economics*, October 2017, *9* (4), 137–65.

Thompson, Robert, *Defeating Communist Insurgency: The Lessons of Malaya and Vietnam*, New York: Praeger, 1966.

Trebbi, Francesco, Eric Weese, Austin Wright, and Andrew Shaver, "Insurgent Learning," *NBER Working Paper 23475*, 2017.

Vanden Eynde, Oliver, "Targets of Violence: Evidence from India's Naxalite Conflict," *The Economic Journal*, 2018, *128* (609), 887–916.

Weidmann, Nils B, "A closer look at reporting bias in conflict event data," *American Journal of Political Science*, 2016, *60* (1), 206–218.

Weinstein, Jeremy M., *Inside Rebellion: The Politics of Insurgent Violence*, Cambridge: Cambridge University Press, 2007.

Wood, Reed M., "Opportunities to Kill or Incentives for Restraint? Rebel Capabilities, the Origins of Support, and Civilian Victimization in Civil War," *Conflict Management and Peace Science*, 2014, *31* (5), 461–480.

Wright, Austin L., "Economic Shocks and Rebel Tactics: Evidence from Colombia," *HiCN paper #232*, March 2016.

APPENDIX [FOR ONLINE PUBLICATION ONLY]

List of Figures and Tables

Table/figure numbers are noted in left column. Page numbers are noted in the right column.

Supplemental Figures

Supplemental Tables

A1 Theory

Proof of Proposition 1

For event H, the complement is denoted H'. We introduce auxiliary random variables T_i, S_i taking two values 0 and 1; $T_i = 1$ means that i-th time slot is defended, $T_i = 0$ means that there is no defense at time i; $S_i = 1$ means that the test of time slot i shows it as defended, and $S_i = 0$ means that the slot i tests as vulnerable. In the absense of index i, S and T correspond to any time slot.

Recall that $P(S|T) = P(S'|T') = \theta$.[27] Finally, let C denote the event that attack is successful during a time slot with an attack. We assume that $P(C|T') = p$ and $P(C|T) = 0$.

Recall that, the parameters are: n is the number of time slots, r is the number of defenses allocated by the government, and a is the number of attacks. The optimal strategy for the government is to allocate r defenses uniformly over n time windows. Then for each window $t(r) = P(\text{time slot is protected}) = \frac{r}{n}$.

We will show that for rebels the optimal strategy looks as follows. If $a < x$, the attacks are allocated uniformly at random among the vulnerable time slots. If $a > x$ then there is a threshold value, a function of parameters of the model and x, that determines how many more put in the vulnerable time windows; above that, they start to put into the windows that tested "defended", again uniformly.

First, observe that cases $x = 0$ and $x = n$ are trivial: there is no information to infer, so the optimal strategy for rebels is to allocate attacks uniformly across n time slots. In what follows, we will assume that $0 < x < n$.

After n slots are tested, a (vector) signal $s = (s_1, .., s_n)$ is obtained and value $N = x$, the number of slots that tested vulnerable is produced. In other words, a (random) partition of a set J into two sets, $J^-(x)$ - slots tested vulnerable, and its complement $J^+(x)$, is obtained. Define $N_1 = |J^-(x) \cap L|$, the number of defended sites that tested vulnerable, i.e., the number of *"false positives"*; and $N_2 = |J^-(x) \cap U|$, the number of correct vulnerable signals. The total number of sites that tested vulnerable is $N \equiv N_{n,r} = N_1 + N_2$.

N_1 and N_2 are independent binomial random variables. Denote $p_i(j)$ the p.m.f. of $N_i, i = 1, 2$, and $p(j|l,p), j = 0, 1, ..., l$ – the p.m.f. of a binomial random variable with l trials and probability of success p. Then $p_1(j) = p(j|r, 1-\theta)$ and $p_2(j) = p(j|n-r, \theta)$.

Now the p.m.f. of N, $g_{n,r}(x)$, $0 \le x \le n$, can be calculated by the standard discrete convolution formula.

$$g_{n,r}(x) = \sum_{0 \le j \le r,\ 0 \le x-j \le n-r} p_1(j) p_2(x-j).$$

[27] All the results go through with the arbitrary parameters $\alpha = P(S|T)$, $\beta = P(S'|T')$ subject to $\alpha + \beta > 1$, i.e., that the test is informative. A standard interpretation for α and β as $\alpha = P(\text{positive test}|\text{disease}) =$ sensitivity; and $\beta = P(\text{negative test}|\text{no disease}) =$ specificity, two important characteristics of any statistical test.

We start with the following basic equations:

$$\begin{aligned}P(C_i &= 1|T_i = 1) = 0,\\ P(C_i &= 1|T_i = 0, u_i) = p(u_i),\end{aligned}$$

where u_i is the number of attacks launched against target i, and $p(u)$ is the success function, the probability of at least one successful attack on a target which faces u attacks.

As we assumed that the success is independent across attacks, $p(u) = 1 - (1-p)^u$. The function $p(u)$ is increasing and upward concave, and the function $\Delta p(u) \equiv p(u+1) - p(u)$ is decreasing. The diminishing effect of each extra attack will play an important role in determing the optimal strategy.

We start with a straightforward lemma.

Lemma A1 *The posterior probability of signal distribution is uniform conditional on the number of signals "vulnerable"* x *:*

$$P(s_1, ..., s_n) = P(N = x)/\binom{n}{x}. \tag{A1}$$

The posterior probability that target i *is vulnerable conditional on the full vector signal* $(s_1, ..., s_n)$ *is equal to the conditional probability that target* i *is vulnerable conditional only on the individual signal* s_i *and the total number of "vulnerable" signals* x*.*

$$P(T_i = 0|s_1, ..., s_n) = P(T_i = 0|s_i, N = x). \tag{A2}$$

Proof. (A1) is straightforward. The left side of formula (A2) can be written as

$$P(T_i = 0)P(s|T_i = 0)/P(s) = P(s_i|T_i = 0)P(s_{-i}|T_i = 0)/P(s),$$

where s_{-i} is vector s without coordinate s_i. Using (A1), we can replace $P(s) = P(N = x)/\binom{n}{x}$. The right side of formula (A2) can be written as

$$\frac{P(T_i = 0, s_i, N = x)}{P(s_i, N = x)} = \frac{P(T_i = 0)P(s_i|T_i = 0)P(N = x|s_i, T_i = 0)}{P(N = x)P(s_i|N = x)}.$$

Let $s_i = 0$. Then, on the left-hand side, using (A1) for a problem with $n-1$ targets and k attacks, we have

$$P(s_{-i}|T_i = 0) = P_{n-1,k}(N = x-1)/\binom{n-1}{x-1}.$$

In the right-hand side we have $P(s_i|N = x) = x/n$ and

$$P(N = x|s_i = 0, T_i = 0) = P_{n-1,k}(N = x-1).$$

Finally, since $\binom{n}{x} = \binom{n-1}{x-1}n/x$, we obtain that after all reductions the left and the right sides of formula (A2) coincide. The proof for the case $s_i = 1$ is similar. ■

The formula (A2) is at the heart of the intuition behind our main results. The optimal strategy $\pi(\cdot|x)$ of the attacker depends on function $p(u)$ and on the ratio $\rho(x|\theta)$, reflecting the relative vulnerability of targets with $s = 0$ and $s = 1$, which, in turns, depends on parameters n, r and θ.

The next Lemma critical ratio that determines the threshold $\overline{a}(x)$ for each x is defined by the following equation:

$$\rho_{n,r}(x) = \rho_{n,r}(x|\theta) \equiv \frac{p^-(x)}{p^+(x)} = \frac{P(C|VN(x))}{P(C|DN(x))}.$$

Our next goal is to establish that $\rho_{n,r}(x) \geq 1$.

Lemma A2 *(a) The probabilities $p^-(x) \equiv P(T'|S', x)$ and $p^+(x) \equiv P(T'|S, x))$ for $0 < x < n$ are given by formulas*

$$\begin{aligned} p^-(x) &= \frac{r}{x} * \theta * \frac{g_{n-1,r}(x-1)}{g_{n,r}(x)}, \\ p^+(x) &= \frac{r}{n-x} * (1-\theta) * \frac{g_{n-1,r}(x)}{g_{n,r}(x)}. \end{aligned}$$

(b) The ratio $\rho_{n,r}(x|\theta)$, $0 < x < n$, is given by the formula

$$\rho_{n,r}(x|\theta) = \frac{\theta}{1-\theta}\frac{n-x}{x}\frac{g_{n-1,r}(x-1)}{g_{n-1,r}(x)}. \tag{A3}$$

Proof. (a) For the sake of brevity, we denote $P(\cdot|N = x) \equiv P(\cdot|x)$ and event $D(x) = (N = x) = D$. By definition $p^-(x) \equiv P(T'|S', x) = P(T'S', x)/P(S', x)$, and $p^+(x) \equiv P(T'|S, x) = P(T'S, x)/P(S, x)$. We have

$$\begin{aligned} P(T'S'D) &= P(T')P(S'D|T') = P(T')P(S'|T')P(D|T'S'), \\ P(T'SD) &= P(T')P(SD|T') = P(T')P(S|T')P(D|T'S). \end{aligned} \tag{A4}$$

We also have $P(T') = \frac{r}{n}$, $P(S'|T') = \theta$, and $P(S|T') = 1 - \theta$. Therefore, to prove the result, it is sufficient to show that

$$\begin{aligned} P(D|T'S') &= g_{n-1,r}(x-1), \\ P(D|T'S) &= g_{n-1,r}(x), \\ P(S'D) &= g_{n,r}(x) * \frac{x}{n}, \\ P(SD) &= g_{n,r}(x) * \frac{n-x}{n}. \end{aligned}$$

To show that $P(D|T'S') = g_{n-1,r}(x-1)$, observe that if $N = x$, and a particular slot has no defense, T', and produced vulnerable signal, S', then in the remaining $n-1$ slots there are r defenses with $x-1$ vulnerable signals. Similarly, to show that $P(D|T'S) = g_{n-1,r}(x)$, observe that if $N = x$, and a particular slot has no defense, T', and produced "defended" signal, S, then in the remaining $n-1$ slots there are r defenses with x vulnerable signals. To demonstrate that $P(S'D) = g_{n,r}(x) * \frac{x}{n}$, note that $P(S'D) = P(D)P(S'|D)$, $P(D) = g_{n,r}(x)$ and $P(S'|D) = \frac{x}{n}$, the probability for one vulnerable signal among x to be in a particular slot. Similarly, $P(SD) = P(D)P(S|D)$, and $P(S|D) = \frac{n-x}{n}$, the probability for one defended signal among $n-x$ to be in a particular slot.

(b) is a straightforward corollary to (a). ■

Lemma A3 [28]

(a) $\rho(x) > 1$.

(b) Functions $\rho_{n,n-1}(x) = \frac{\theta^2}{1-\theta^2}$ *for all* x*, while functions* $\rho_{n,r}(x|\theta)$ *for* $r < n-1$ *are monotonically increasing in* x *for* $0 < x < n$.

(c) Functions $\rho_{n,r}(x|\theta)$ *are monotonically decreasing for all fixed* r, $0 < x < n$ *when* n *is increasing.*

Proof. (a) Let $D_{n,r}$ be a sum of n Bernoulli random variables, r of which have parameter $1-\theta$, and $n-r$ of which have parameter $\theta > 1-\theta$, $0 \le r \le n$. Let

$$\begin{aligned} g_{n,r}(x) &= P(D_{n,r} = x),\ 0 \le x \le n; \\ f_{n,r}(x) &= \frac{g_{n,r}(x-1)}{g_{n,r}(x)},\ 1 \le x \le n; \end{aligned}$$

$$\rho_{n,r}(x) = \frac{n+1-x}{x}\frac{\theta}{1-\theta} f_{n,r}(x), 1 \le x \le n.$$

We assume also that $f_{n,r}(0) = \rho_{n,r}(0) = 0$.

Denote $c = \frac{\theta^2}{1-\theta^2}$. Then $\theta > \frac{1}{2}$ implies $c > 1$.

It is easy to check that

$$\begin{aligned} \rho_{n,0}(x) &= 1,\ \text{for } 1 \le x \le n\,; \\ \rho_{n,n}(x) &= c,\ \text{for } 1 \le x \le n\,; \\ \rho_{n,r}(n) &= \frac{rc + n - r}{n} = 1 + \frac{r}{n}(c-1) \text{ for } 0 \le r \le n\,. \end{aligned} \tag{A5}$$

Using the total probability formula, we obtain a recursive relationship:

$$\begin{aligned} f_{n,r}(x) &= \frac{g_{n,r}(x-1)}{g_{n,r}(x)} = \frac{(1-\theta)g_{n-1,r}(x-1) + \theta g_{n-1,r}(x-2)}{(1-\theta)g_{n-1,r}(x) + \theta g_{n-1,r}(x-1)} \\ &= \frac{1-\theta+\theta f_{n-1,r}(x-1)}{\frac{1-\theta}{f_{n-1,r}(x)} + \theta} \text{ for } 1 \le r,\ x \le n-1\,. \end{aligned} \tag{A6}$$

[28] We thank Isaac Sonin and Ernst Presman for their help with proving this lemma.

This implies that

$$\begin{aligned}
\rho_{n,r}(x) &= \frac{n-x+1}{x}\,\frac{\theta}{1-\theta}\,\frac{1-\theta+\theta\frac{x-1}{n-x+1}\frac{1-\theta}{\theta}\rho_{n-1,r}(x-1)}{\frac{1-\theta}{\rho_{n-1,r}(x)}\;\frac{n-x}{x}\frac{1-\theta}{\theta}+\theta} \\
&= \frac{n-x+1+(x-1)\rho_{n-1,r}(x-1)}{\frac{n-x}{\rho_{n-1,r}(x)}+x} \quad \text{for } 1\le r,\ x\le n-1\,. \qquad \text{(A7)}
\end{aligned}$$

(A5) is the induction base. By induction, the numerator in (A7) is strictly increasing, and the denominator is strictly decreasing, and hence the right side in (A7) is strictly increasing and depends only on c. When c grows from 1 to ∞ (that is, θ is increasing from 0 to $\frac{1}{2}$), it is strictly increasing from 1.

(b) Now we can show that for fixed $n\ge 2,\ 1\le r\le n-1,\ c>1$, function $\rho_{n,r}(x)$ is strictly increasing in x.

Again, we use induction by n. Let $\rho(x)=\rho_{n-1,r}(x),\ H(x)=n-x+x\rho(x)$. Then, by (A7),

$$\rho_{n,r}(x)=\rho(x)\frac{H(x-1)}{H(x)},\quad \rho_{n,r}(x+1)-\rho_{n,r}(x)=\frac{C(x)}{H(x+1)H(x)}\,,$$

where

$$\begin{aligned}
C(x) &= \rho(x+1)H^2(x)-\rho(x)H(x+1)H(x-1) \\
&= \rho(x+1)\left[(n-x)^2+2x(n-x)\rho(x)+x^2\rho^2(x)\right] \\
&= -\rho(x)[n-x-1+(x+1)\rho(x+1)]\,[n-x+1+(x-1)\rho(x-1)] \\
&= \rho(x+1)\left[(n-x)^2+2x(n-x)\rho(x)+x^2\rho^2(x)\right].
\end{aligned}$$

We can check, using (A6) and (A7), that

$$\rho_{2,1}(2)=\frac{1+c}{2}>\rho_{2,1}(1)=\frac{2c}{1+c},$$

$$\rho_{3,1}(3)=\frac{2+c}{3}>\rho_{3,1}(2)=\frac{1+2c}{2+c}>\rho_{3,1}(1)=\frac{3c}{1+2c},$$
$$\rho_{3,2}(3)=\frac{1+2c}{3}>\rho_{3,2}(2)=c\frac{2+c}{1+2c}>\rho_{3,2}(1)=\frac{3c}{2+c},$$

and hence the proof is complete for $n=2$ and $n=3$.

For $n\ge 4$ and $2\le x\le n-2$, we have $x(n-x)-n\ge 0$, and by induction, it follows that

$$\begin{aligned}
C(x) &> \rho(x)+\rho^2(x)\rho(x+1)-\rho(x)\rho(x+1)-\rho^2(x) \\
&= \rho(x)(\rho(x+1)-1)(\rho(x)-1)>0
\end{aligned}$$

for $2\le x\le n-2$.

To prove (c), it remains to show that $\rho_{n-r,r}(2) - \rho_{n-r,r}(1) > 0$ and $\rho_{n-r,r}(n) - \rho_{n-r,r}(n-1) > 0$. Using the total probability formula,

$$\begin{aligned}\rho_{n,r}(n-1) &= \frac{2}{n-1}\,\frac{\theta}{1-\theta}\,\frac{P(D_{nr}=n-2)}{P(D_{nr}=n-1)} \\ &= \frac{1}{n-1}\,\frac{(n-r)(n-r-1)+2(n-r)rc+r(r-1)c^2}{n-r+cr}. \end{aligned} \tag{A8}$$

Then, using (A6), we obtain

$$\rho_{n,r}(n) - \rho_{n,r}(n-1) = \frac{(n-r)r(c-1)^2}{n(n-1)(n-r+cr)} > 0\,.$$

Similarly,

$$\rho_{n,r}(2) = \frac{(n-1)(c(n-r)+r)}{c^2(n-r)(n-r-1)+2(n-r)rc+r(r-1)}\,,$$

$$\rho_{n,r}(1) = \frac{nc}{r+c(1-r)}\,,$$

$$\rho_{n,r}(2) - \rho_{n,r}(1) = \frac{(n-r)r(c-1)^2}{[c^2(n-r)(n-r-1)+2(n-r)rc+r(r-1)](r+c(1-r))}\,.$$

■

Let $B^-(s) = \{i : s_i = 0\}$ and $B^+(s) = \{i : s_i = 1\}$. Then, using (A2), we obtain that, given a strategy $\pi = (u_1, ..., u_n)$ and any signal s with $N(s) = x$, the expected value of a strategy π is

$$w(\pi|x) = 1 - \Pi_{j=1}^{n}(1 - P(C_j|u_j, s_j, x)).$$

Let $U^- \equiv U^-(\pi|s) = \{u_j, j \in B^-(s)\}$ and $U^+ \equiv U^+(\pi|s) = \{u_j \in B^+(s)\}$ be two possible sets of the values of u_j at vulnerable and non-vulnerable targets. Formula (A2) immediately implies that all strategies obtained by permutations of sets (U^-, U^+) among corresponding targets have the same value.

We prove later that the upward concavity of function $p(u)$ implies that the optimal strategy has the property that the number of attacks against any pair of targets with the same signal is the same or almost the same: as the number of attacks is integer, there might be a difference of one attack between two targets with the same signals. Let us assume for simplicity that both equalities hold: $u_i = u^-, i \in B^-(s)$ and $u_i = u^+, i \in B^+(s)$.

Let us do the following transformation:

$$\begin{aligned}\ln(1 - w(\pi|x)) &= \sum_{j=1}^{n}(1 - P(T_i = 0|s_i, x)p(u_i)) \\ &= n - \sum_{j=1}^{n} P(T_i = 0|s_i, x)p(u_i) \\ &= n - p^-(x)\sum_{i\in B^-(s)} p(u_i) - p^+(x)\sum_{i\in B^+(s)} p(u_i). \end{aligned}$$

Now, the same strategy π that maximizes the function $v(\pi|x) = p^{-}(x)\sum_{i\in B^{-}(s)} p(u_i) - p^{+}(x)\sum_{i\in B^{+}(s)} p(u_i)$ is maximizing the strategy value $(\pi|x)$.

The following lemma concludes the proof of Proposition 1. To obtain the optimal strategy, we use the necessary equilibrium condition: with an optimal allocation it is impossible to increase the payoff by moving an attack from one target to another. Given $N = x$, $0 \le x \le n$ the allocation of attacks depends on the number a of attacks available. The optimal strategy has the following structure. Initially, all attacks are launched one by one into each of x vulnerable slots until the threshold level

$$d(x) = \min_{i\ge 1}\left\{ i|\rho(x|\theta)\,(1-p)^{i} < 1 \right\} \tag{A9}$$

is reached in each of them or the attack resources are exhausted. Afterwards, the attacks are added one by one to non-vulnerable slot until there is an attack on each of them. Then, attacks are added one by one until each of the vulnerable slots has $d(x)+1$ attacks on each of them, then back to non-vulnerable slots until each has at least 2 attacks, etc. This "fill and switch" process stops when the attacker runs run out of resources. If $x = 0$ or n then just all slots are filled sequentially. The outcome of this process will be an allocation in which either all "vulnerable" slots will have the same number of attacks launched against them, or all "not vulnerable" slots, or both. If all "not vulnerable" slots have no attacks allocated to them, then the number of attacks against each "vulnerable" slot does not exceed $d(x)$. (Trivially, after the process is complete, the attacker will have to uniformly randomize the distributions of attacks over sets of slots with the same sign: otherwise, the uniform distribution of protection by the defender would not be a best response.) The threshold $\overline{a}(x)$ in the statement of Proposition 1 is a function of both x (via $d(x)$) and the total number of attacks available, a.

Lemma A4 *(i) Let $\pi(x) = (u_i, i = 1, 2, ..., n)$ be an optimal strategy . Then $|u_{i_1} - u_{i_2}| \le 1$ when the signals at time slots i_1, i_2 have the same sign.*

(ii) Let $\pi(x) = (u_i, i = 1, 2, ..., n)$ be a strategy, $0 < x < n$, u^- be the number of attacks in some vulnerable slot, u^+ be the number of attacks in some protected slot, and $d = d(x)$ is defined by formula (A9). Then, if $u^- - u^+ > d(x)$ or, if $u^+ \ge 1$ and $u^- - u^+ < d(x) - 1$, then strategy π is not optimal, or, equivalently, if π is optimal, and $u^+ = 0$, then $1 \le u^- \le d(x)$, and if $u^+ \ge 1$, then $u^- - u^+ = d(x)$ or $d(x) - 1$.

Proof. (i) Let J be a subset of slots, and recall that $C_j = 1$ when slot j is destroyed. The conditional independence of testing and attacks's successes, and total probability formula imply the following formula for the conditional probability of the destruction of a particular slot with $u \ge 1$ attacks

$$P(C|u, F) = P(C|u, T = 0)P(T = 0|F) = p(u)P(T = 0|F),$$

where F is any event generated by testing (signals).

Using the above formula and the definitions of $\rho(x)$, $p^-(x)$ and $p^+(x)$, we have:

$$\begin{aligned} P(C|u, S=1, x) &= P(T=0|S=1, x)P(C|u, T=0) = p^+(x)p(u), \\ P(C|u, S=0, x) &= P(T=0|S=0, x)P(C|u, T=0) = p^-(x)p(u) = \rho(x)p^+(x)p(u). \end{aligned} \tag{A10}$$

Suppose that the statement is not true and let us say $u_{i_1} = u, u_{i_2} = j, u - j \geq 2$ and $S_{i_1} = S_{i_2} = 1$. The concavity of function $p(\cdot)$ implies that $p(u+1) + p(j-1) > p(u) + p(j)$. Then, using the formulas in (A10), we have

$$\begin{aligned} P(C &= 1|u+1, S=1, x) + P(C|j-1, S=1, x) = p^+(x)[p(u+1) + p(j-1))] > \\ > p^+(x)[p(u) + p(j)] &= P(C|u, S=1, x) + P(C|j, S=1, x). \end{aligned} \tag{A11}$$

Thus, $v(\pi|x)$ is not maximized and our initial strategy is not optimal. The proof for $S_{i_1} = S_{i_2} = 0$ is similar with $p^+(x)$ replaced by $p^-(x) = \rho(x)p^+(x)$.

(ii) Let $d(x) = d$. We will show that if $u^- - u^+ > d$ for some pair of vulnerable and protected slots, then a transfer of one attack from a vulnerable slot from this pair to a protected slot will increase the value of a strategy . Similarly, if $u^+ \geq 1$ and $u^- - u^+ < d-1$ for such pair, then the inverse transfer will increase the value. As always, we assume that $a+b > 1$ and then $\rho(x) > 1$ for $0 < x < n$, and hence $u^- \geq u^+$. Let $u^- = u, u^+ = j, P(\cdot|N = x) = P(\cdot|x)$, and denote the incremental utilities for vulnerable and protected slots as

$$\Delta C^-(u|x) = P(C|u+1, S', x) - P(C|u, S'^+(j|x) = P(C|j+1, S, x) - P(C|j, S, x).$$

Then, formula (A10) implies that their difference for $0 \leq j \leq i$ is

$$\begin{aligned} \Delta(u, j|x) &= \Delta C^-(u|x) - \Delta C^+(j|x) \\ &= p\,(1-p)^u\,\rho(x)p^+(x) - p\,(1-p)^j\,p^+(x)) \\ &= p\,(1-p)^j\,p^+(x)[\rho(x)\,(1-p)^{u-j} - 1]. \end{aligned}$$

The definition of $d = d(x)$ in (A9) implies that $\Delta(u, j|x)$ is positive if $j = 0, u < d$, or if $j \geq 1, u - j < d$. Similarly, $\Delta(u, j|x)$ is negative if $j = 0, u \geq d$, or if $j \geq 1, u - j \geq d$. These inequalities imply the claim. Also, note that if $p = 1$, then $d(x) = 1$ for all $0 < x < n$, and if p is decreasing to zero, then $d(x)$ tends to infinity. ■

This concludes the proof of Proposition 1. ■

Proof of Proposition 2

By (A7) in the proof of Lemma A3, the critical ratio $\rho_{n,r}(x|\theta)$ is given by the recursive formula

$$\rho_{n,r}(x) = \frac{n - x + 1 + (x-1)\rho_{n-1,r}(x-1)}{\frac{n-x}{\rho_{n-1,r}(x)} + x} \text{ for } 1 \leq r,\ x \leq n-1\,.$$

For each r, one can use the above formula and induction on n to show that $\rho_{n,r}(x|\theta)$ is an increasing function of θ for each x, $x \leq n-1$. Indeed, $\rho_{n,r}(x|\theta)$ is monotonically increasing in $\rho_{n-1,r}(x|\theta)$. Then, the induction step completes the argument.

Now, take any θ_1, θ_2 such that $\theta_1 < \theta_2$. As $\rho_{n,r}(x|\theta_1) < \rho_{n,r}(x|\theta_2))$, for the thresholds $d(x|\theta_1)$ and $d(x|\theta_2)$ defined by (A9), one has $d(x|\theta_1) \leq d(x|\theta_2)$. This means that attacks are more temporally concentrated with θ_2 than with θ_1, i.e., $\overline{a}(x|\theta_1) \leq \overline{a}(x|\theta_2)$. ■

Proof of Proposition 3

Observe that $\rho_{n,r}(x)$ does not depend on p. By (A9), $d(x) = \min_{i \geq 1} \left\{ i | \rho(x|\theta) \left(1 - p\right)^i < 1 \right\}$. When p increases, $d(x)$ goes down, which results in less temporal clustering of attacks. Similarly, when r increases, so does $\rho_{n,r}(x)$ by (A6) and (A7), which results in a higher $d(x)$, and more temporally clustered attacks. ■

A2 Empirical Appendix

In this brief empirical appendix, we introduce supplemental results.

A2.1 Descriptive Statistics

Table A-1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
OUTCOME OF INTEREST					
Indirect fire, likelihood parameter	-7.025	3.918	-39.686	0	600
REBEL CAPACITY					
Opium revenue	20.506	25.403	0	75.518	600
Opium revenue, regional yield/prices	20.017	24.7	0	73.587	600
TRENDS IN VIOLENCE					
Indirect fire trend, fighting season	19.187	21.153	5	253	600
Indirect fire trend, growing/harvest season	10.788	15.194	0	161	600
Indirect fire trend, planting season	7.663	10.925	0	109	600

Notes: summary statistics are calculated for the sample studied in the main estimating equation.

A2.2 Additional Threats to Inference

In this subsection we detail additional potential threats to inference.

It is possible that revenue from opium production may attract additional counterinsurgent investments. Consistent with our theoretical model, to defeat rebels accumulating resources, security forces may have deployed additional resources that could complicate estimation of the effect of rebel capacity on the randomness of combat. We focus on six measures of counterinsurgent capacity: close air support missions, cache discoveries, IED neutralizations, detention of insurgent forces, counterinsurgent surveillance operations, and safe house raids yielding actionable intelligence assets (salary and recruitment logs, hard drives, forensic materials, etc.). We sequentially add these covariates to Equation 1. We present these results in Tables A-2 and A-3. For comparison, the most conservative specification from Table 1 (Column 4) is included as Column 1 in both of these tables. Notice that these measures of counterinsurgent capacity improve the explanatory power of our models, although the magnitude of the main effect is slightly attenuated.

Table A-2: Impact of rebel capacity on within-day randomization of indirect fire attacks, accounting for state capacity measures (part i)

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0549***	-0.0420***	-0.0346***	-0.0274***
	(0.0125)	(0.00989)	(0.00785)	(0.00945)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes	Yes
ADDITIONAL PARAMETERS				
Weapon Caches Cleared	No	Yes	No	No
Close Air Support	No	No	Yes	No
IEDs Cleared	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.187	0.243	0.299	0.320

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-3: Impact of rebel capacity on within-day randomization of indirect fire attacks, accounting for state capacity measures (part ii)

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0549***	-0.0539***	-0.0528***	-0.0477***
	(0.0125)	(0.0120)	(0.0116)	(0.0103)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes	Yes
Additional Parameters				
Coalition Surveillance	No	Yes	No	No
Safe House Raids	No	No	Yes	No
Detention of Susp. INS	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.187	0.188	0.193	0.221

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Opium production might also be influenced by coercive tactics used by the Taliban to intimidate civilians. In general, these tactics are difficult to track. These coercive tactics represent a problematic omitted variable if they are indeed correlated with production and further correlated with the sophistication of combat tactics used during the fighting season. These are plausible concerns. To address them, we incorporate additional information from our military records which tracks attempts to intimidate the civilian population, using methods like 'night letters' and shows of non-lethal force as well as deliberate killings of government collaborators (like informants and security force recruits). Our main effect is only slightly attenuated when we account for these rebel intimidation tactics in Columns 2 and 3 of Table A-4.

Table A-4: Impact of rebel capacity on within-day randomization of indirect fire attacks, accounting for rebel intimidation tactics

	(1)	(2)	(3)
Opium Revenue	-0.0549***	-0.0492***	-0.0484***
	(0.0125)	(0.0121)	(0.0104)
MODEL PARAMETERS			
Fighting Season Fixed Effect	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes
ADDITIONAL PARAMETERS			
Taliban Intimidation	No	Yes	No
Collaborator Killings	No	No	Yes
MODEL STATISTICS			
No. of Observations	600	600	600
No. of Clusters	154	154	154
R^2	0.187	0.204	0.227

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-5: Estimating treatment effect bounds using the Oster coefficient stability test

		Panel A: Baseline Regression Diagnostic Information	
		(1)	(2)
Treatment Variable	Outcome Variable	Baseline effect (Std. error), [R^2]	Controlled effect (Std. error), [R^2]
Opium revenue	Temporal clustering	-0.0552*** (.0125) [0.128]	-.0246*** (.0081) [0.374]

		Panel B: Oster Coefficient Stability Test Results	
		(3)	(4)
Treatment Variable	Outcome Variable	Effect for R_{max} $((\beta_{R_{max}}$ - $\beta_{ctrl})^2)$ [R_{max}]	Alt. Effect for R_{max} $((\beta_{R_{max}}$ - $\beta_{ctrl})^2)$ [R_{max}]
Opium revenue	Temporal clustering	-0.007 (.0003) [.486]	-0.232 (.0431) [.486]

Notes: Bounds for treatment effects are estimated using the Oster coefficient stability test (Oster, 2017). Unobservables are assumed to have as much explanatory power as the observables. R_{max} set at 1.3 (the threshold suggested in (Oster, 2017)). Model specifications are drawn from least and most conservative main specifications. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-6: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to historical territorial control by Taliban

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0258***	-0.0288***	-0.0232**	-0.0231**
	(0.0137)	(0.00936)	(0.00995)	(0.00994)	(0.00985)
Historical Taliban Control		-0.0634	0.0974	0.109	0.111
		(0.401)	(0.404)	(0.406)	(0.407)
Hist. Control × Revenue		-0.0328***	-0.0307***	-0.0325***	-0.0325***
		(0.0120)	(0.0116)	(0.0121)	(0.0120)
MODEL PARAMETERS					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
MODEL STATISTICS					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.156	0.172	0.188	0.188

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-7: Reduced form effects of rebel capacity on production of indirect fire attacks

	(1)	(2)	(3)
Suitability × Agg. Production	0.135***	0.0938***	0.0903**
	(0.0360)	(0.0338)	(0.0352)
MODEL PARAMETERS			
Fighting Season Fixed Effect	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes
Growing Season Activity (levels)	No	Yes	Yes
Planting Season Activity (levels)	No	No	Yes
MODEL STATISTICS			
No. of Observations	3582	3582	3582
No. of Clusters	398	398	398
R^2	0.696	0.760	0.767

Notes: Outcome of interest is the (log) of indirect fire attacks during the post-harvest fighting season. We add one to all observations before evaluating the logarithm. The quantity of interest is opium revenue for a given district-year. The regression is reduced form, using the primary instrument detailed in the IV section of the text. All regressions include district and fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-8: Second stage effects of rebel capacity on production of indirect fire attacks

	(1)	(2)	(3)
Opium Revenue	0.0365***	0.0261**	0.0252**
	(0.0117)	(0.0104)	(0.0106)
Model Parameters			
Fighting Season Fixed Effect	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes
Growing Season Activity (levels)	No	Yes	Yes
Planting Season Activity (levels)	No	No	Yes
Model Statistics			
No. of Observations	3582	3582	3582
No. of Clusters	398	398	398
R^2	0.557	0.681	0.692

Notes: Outcome of interest is the (log) of indirect fire attacks during the post-harvest fighting season. We add one to all observations before evaluating the logarithm. The quantity of interest is opium revenue for a given district-year. The regression displayed is the second stage, using the primary instrument detailed in the IV section of the text. All regressions include district and fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-9: Impact of rebel capacity on within-day randomization of indirect fire attacks, excluding non-producers

	(1)	(2)	(3)	(4)
Opium Revenue	-0.139***	-0.125***	-0.116***	-0.113***
	(0.0436)	(0.0386)	(0.0365)	(0.0361)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	254	254	254	254
No. of Clusters	70	70	70	70
R^2	0.151	0.185	0.213	0.217

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects. Column 2-4 add controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-10: Impact of rebel capacity on within-day randomization of indirect fire attacks, adjusting measure of revenue

	(1)	(2)	(3)	(4)
Opium Revenue	-0.139***	-0.125***	-0.116***	-0.113***
	(0.0436)	(0.0386)	(0.0365)	(0.0361)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	254	254	254	254
No. of Clusters	70	70	70	70
R^2	0.150	0.185	0.213	0.217

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects. Column 2-4 add controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A2.3 Additional IV Results

Aid delivery during growing season

Table A-11: Impact of rebel capacity on within-day randomization of indirect fire attacks, instrumental variables approach (second stage), accounting for aid delivery

	(1)	(2)	(3)
Opium Revenue	-0.0741***	-0.0743***	-0.0725***
	(0.0247)	(0.0247)	(0.0234)
Model Parameters			
Fighting Season Fixed Effect	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes
Type of Aid	No	All CERP	Agri./Irri. CERP
Model Statistics			
No. of Observations	600	600	600
No. of Clusters	154	154	154
R^2	0.173	0.173	0.180
IV Specification			
IV Type	Benchmark	Benchmark	Benchmark
Kleibergen-Paap F Statistic	187.4	187.1	177.9

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year instrumented using an opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Baseline suitability specification

Table A-12: Impact of suitability instrument on opium revenue, instrumental variables approach (first stage, main estimating sample)

	(1)	(2)	(3)	(4)
Suitability × Agg. Production	24.17***	24.20***	23.71***	23.72***
	(1.749)	(1.763)	(1.725)	(1.732)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.465	0.467	0.472	0.473

Notes: Outcome of interest is opium revenue for a given district-year. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-13: Impact of suitability instrument on opium revenue, instrumental variables approach (first stage, full panel sample)

	(1)	(2)	(3)	(4)
Suitability × Agg. Production	20.54***	20.42***	20.04***	20.01***
	(1.093)	(1.083)	(1.073)	(1.076)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	3582	3582	3582	3582
No. of Clusters	398	398	398	398
R^2	0.304	0.304	0.310	0.311

Notes: Outcome of interest is opium revenue for a given district-year. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-14: Impact of suitability instrument on within-day randomization of indirect fire attacks, instrumental variables approach (reduced form, main estimating sample)

	(1)	(2)	(3)	(4)
Suitability × Agg. Production	-1.917***	-1.904***	-1.758***	-1.758***
	(0.666)	(0.660)	(0.594)	(0.595)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.129	0.141	0.160	0.160

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Monotonicity assumption

Table A-15: Impact of suitability instrument on opium revenue, instrumental variables approach (first stage, main estimating sample) accounting for potential variation in instrument compliance via irrigation mechanism

	(1)	(2)	(3)	(4)
Suitability × Agg. Production	23.72***	20.82***	23.52***	23.75***
	(1.732)	(4.735)	(2.492)	(2.104)
Irrigated Area: 75th pctl and above		7.407**		
		(3.224)		
Suitability × Agg. Production × 75th		1.905		
		(4.893)		
Irrigated Area: 90th pctl and above			7.344	
			(5.400)	
Suitability × Agg. Production × 90th			-2.039	
			(3.722)	
Irrigated Area: 95th pctl and above				13.45*
				(7.676)
Suitability × Agg. Production × 95th				-4.879
				(4.781)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes	Yes
MODEL STATISTICS				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.473	0.494	0.481	0.487

Notes: Outcome of interest is opium revenue for a given district-year. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-16: Impact of suitability instrument on opium revenue, instrumental variables approach (first stage, full panel sample) accounting for potential variation in instrument compliance via irrigation mechanism

	(1)	(2)	(3)	(4)
Suitability × Agg. Production	20.01***	16.84***	19.50***	19.89***
	(1.076)	(1.953)	(1.313)	(1.201)
Irrigated Area: 75th pctl and above		3.939**		
		(1.564)		
Suitability × Agg. Production × 75th		4.595**		
		(2.113)		
Irrigated Area: 90th pctl and above			0.186	
			(2.074)	
Suitability × Agg. Production × 90th			2.076	
			(1.975)	
Irrigated Area: 95th pctl and above				2.859
				(2.879)
Suitability × Agg. Production × 95th				0.0556
				(2.231)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	Yes	Yes	Yes	Yes
Growing Season Activity (levels)	Yes	Yes	Yes	Yes
Planting Season Activity (levels)	Yes	Yes	Yes	Yes
MODEL STATISTICS				
No. of Observations	3582	3582	3582	3582
No. of Clusters	398	398	398	398
R^2	0.311	0.322	0.312	0.312

Notes: Outcome of interest is opium revenue for a given district-year. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted). All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

LASSO-based suitability specification

Table A-17: Impact of suitability instrument on opium revenue, instrumental variables approach (first stage, main sample, LASSO selection)

	(1)	(2)	(3)	(4)
Suitability$^{LASSO}\times$ Agg. Production	22.87***	22.88***	22.39***	22.39***
	(2.135)	(2.150)	(2.063)	(2.059)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.429	0.431	0.435	0.436

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted) where inputs are selected via LASSO. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-18: Impact of suitability instrument on opium revenue, instrumental variables approach (first stage, full panel sample, LASSO selection)

	(1)	(2)	(3)	(4)
Suitability^{LASSO}× Agg. Production	18.05***	17.92***	17.51***	17.49***
	(1.186)	(1.168)	(1.150)	(1.150)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	3582	3582	3582	3582
No. of Clusters	398	398	398	398
R^2	0.257	0.258	0.264	0.266

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted) where inputs are selected via LASSO. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

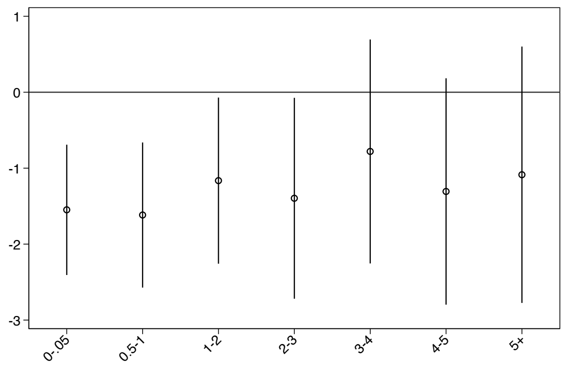
Table A-19: Impact of suitability instrument on within-day randomization of indirect fire attacks, instrumental variables approach (reduced form, main sample, LASSO selection)

	(1)	(2)	(3)	(4)
Suitability^{LASSO} $\times$ Agg. Production	-1.735***	-1.731***	-1.569***	-1.569***
	(0.653)	(0.645)	(0.589)	(0.589)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.111	0.124	0.143	0.143

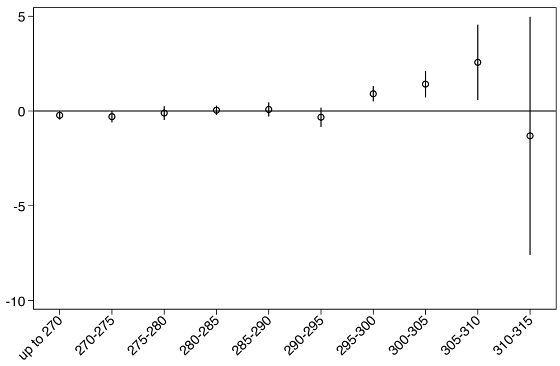
Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is the opium suitability index interacted with the prior year's national-level production (inverted) where inputs are selected via LASSO. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Agronomic inputs specification

Figure A-1: Estimated first stage effects of agronomic inputs on opium revenue.



(a) Precipitation-Day Instruments



(b) Temperature-Day Instruments

Notes: Results from Column 4 of Table A-20, where the outcome is opium revenue (as measured in the main specification) and the right hand side variables include the first stage instruments from our third IV strategy (raw agronomic inputs: precipitation-day (mm) and temperature-day (Kelvin) binned measures). Panel (a) plots the precipitation-day coefficients. Panel (b) displays the temperature-day coefficients.

Table A-20: Impact of agronomic instruments on opium revenue, instrumental variables approach (first stage, main sample, multiple IV approach)

	(1)	(2)	(3)	(4)
Precip Days, 0-.05	-1.528***	-1.533***	-1.549***	-1.548***
	(0.436)	(0.436)	(0.433)	(0.434)
Precip Days, 0.5-1	-1.583***	-1.592***	-1.618***	-1.617***
	(0.489)	(0.486)	(0.483)	(0.484)
Precip Days, 1-2	-1.090**	-1.112**	-1.165**	-1.164**
	(0.547)	(0.554)	(0.552)	(0.554)
Precip Days, 2-3	-1.409**	-1.379**	-1.397**	-1.396**
	(0.674)	(0.670)	(0.666)	(0.669)
Precip Days, 3-4	-0.726	-0.747	-0.780	-0.780
	(0.747)	(0.748)	(0.745)	(0.746)
Precip Days, 4-5	-1.214	-1.264*	-1.306*	-1.306*
	(0.767)	(0.758)	(0.753)	(0.754)
Precip Days, 5+	-1.166	-1.187	-1.083	-1.086
	(0.874)	(0.868)	(0.857)	(0.854)
Temp Days, up to 270	-0.207*	-0.216*	-0.227**	-0.227**
	(0.117)	(0.116)	(0.113)	(0.112)
Temp Days, 270-275	-0.284*	-0.292*	-0.298*	-0.299*
	(0.158)	(0.156)	(0.153)	(0.154)
Temp Days, 275-280	-0.0866	-0.0924	-0.108	-0.109
	(0.187)	(0.186)	(0.182)	(0.181)
Temp Days, 280-285	0.0702	0.0605	0.0423	0.0414
	(0.127)	(0.127)	(0.125)	(0.124)
Temp Days, 285-290	0.108	0.0990	0.0817	0.0814
	(0.193)	(0.193)	(0.189)	(0.188)
Temp Days, 290-295	-0.300	-0.304	-0.327	-0.327
	(0.258)	(0.258)	(0.257)	(0.257)
Temp Days, 295-300	0.916***	0.912***	0.909***	0.910***
	(0.205)	(0.204)	(0.206)	(0.205)
Temp Days, 300-305	1.451***	1.454***	1.420***	1.418***
	(0.351)	(0.352)	(0.350)	(0.357)
Temp Days, 305-310	2.609**	2.628***	2.566**	2.567**
	(1.002)	(1.006)	(1.007)	(1.007)
Temp Days, 310-315	-1.431	-1.765	-1.332	-1.312
	(3.158)	(3.179)	(3.130)	(3.181)
Model Parameters				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
Model Statistics				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.455	0.457	0.459	0.459

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantities of interest are the estimated effects of various precipitation-day and temperature-day binned classifications. Precipitation is in millimeters and temperature is in Kelvin. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-21: Impact of agronomic instruments on opium revenue, instrumental variables approach (first stage, full panel sample, multiple IV approach)

	(1)	(2)	(3)	(4)
Precip Days, 0-.05	-0.768***	-0.748***	-0.742***	-0.738***
	(0.218)	(0.218)	(0.219)	(0.219)
Precip Days, 0.5-1	-0.795***	-0.766***	-0.766***	-0.757***
	(0.246)	(0.246)	(0.247)	(0.247)
Precip Days, 1-2	-1.054***	-1.008***	-1.009***	-1.001***
	(0.282)	(0.283)	(0.283)	(0.283)
Precip Days, 2-3	-1.048***	-1.031***	-1.023***	-1.015***
	(0.267)	(0.267)	(0.268)	(0.268)
Precip Days, 3-4	-0.225	-0.196	-0.204	-0.207
	(0.314)	(0.315)	(0.316)	(0.316)
Precip Days, 4-5	-0.675**	-0.638**	-0.652**	-0.651**
	(0.294)	(0.295)	(0.294)	(0.294)
Precip Days, 5+	-0.738**	-0.706*	-0.646*	-0.654*
	(0.371)	(0.370)	(0.370)	(0.371)
Temp Days, up to 270	-0.225***	-0.224***	-0.229***	-0.231***
	(0.0663)	(0.0664)	(0.0660)	(0.0660)
Temp Days, 270-275	-0.194**	-0.192**	-0.198***	-0.202***
	(0.0752)	(0.0753)	(0.0749)	(0.0749)
Temp Days, 275-280	-0.332***	-0.336***	-0.339***	-0.343***
	(0.0936)	(0.0936)	(0.0927)	(0.0929)
Temp Days, 280-285	-0.0596	-0.0581	-0.0743	-0.0803
	(0.0745)	(0.0748)	(0.0750)	(0.0748)
Temp Days, 285-290	-0.0811	-0.0786	-0.0916	-0.0934
	(0.0793)	(0.0797)	(0.0791)	(0.0789)
Temp Days, 290-295	-0.127	-0.126	-0.137	-0.136
	(0.109)	(0.109)	(0.109)	(0.109)
Temp Days, 295-300	0.381***	0.376***	0.384***	0.389***
	(0.121)	(0.120)	(0.120)	(0.120)
Temp Days, 300-305	0.857***	0.843***	0.793***	0.772***
	(0.193)	(0.192)	(0.191)	(0.192)
Temp Days, 305-310	1.320***	1.340***	1.268***	1.263***
	(0.407)	(0.401)	(0.396)	(0.394)
Temp Days, 310-315	8.374***	8.261***	9.125***	9.382***
	(3.033)	(2.911)	(2.863)	(2.956)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	3582	3582	3582	3582
No. of Clusters	398	398	398	398
R^2	0.204	0.206	0.213	0.214

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantities of interest are the estimated effects of various precipitation-day and temperature-day binned classifications. Precipitation is in millimeters and temperature is in Kelvin. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-22: Impact of agronomic instruments on within-day randomization of indirect fire attacks, instrumental variables approach (reduced form, main sample, multiple IV approach)

	(1)	(2)	(3)	(4)
Precip Days, 0-.05	0.0462	0.0442	0.0524	0.0525
	(0.0795)	(0.0788)	(0.0793)	(0.0794)
Precip Days, 0.5-1	0.0443	0.0408	0.0543	0.0547
	(0.0849)	(0.0845)	(0.0855)	(0.0857)
Precip Days, 1-2	0.0484	0.0393	0.0666	0.0668
	(0.0898)	(0.0873)	(0.0877)	(0.0877)
Precip Days, 2-3	0.0784	0.0905	0.100	0.100
	(0.117)	(0.113)	(0.111)	(0.112)
Precip Days, 3-4	0.107	0.0986	0.115	0.115
	(0.143)	(0.143)	(0.148)	(0.148)
Precip Days, 4-5	0.115	0.0950	0.116	0.116
	(0.109)	(0.108)	(0.108)	(0.108)
Precip Days, 5+	-0.0768	-0.0850	-0.138	-0.139
	(0.210)	(0.210)	(0.202)	(0.202)
Temp Days, up to 270	0.0237	0.0203	0.0260	0.0259
	(0.0178)	(0.0172)	(0.0172)	(0.0172)
Temp Days, 270-275	0.0420*	0.0388*	0.0421*	0.0420*
	(0.0233)	(0.0230)	(0.0227)	(0.0230)
Temp Days, 275-280	0.0369	0.0346	0.0425*	0.0424*
	(0.0248)	(0.0243)	(0.0249)	(0.0251)
Temp Days, 280-285	0.0193	0.0154	0.0247	0.0245
	(0.0217)	(0.0216)	(0.0221)	(0.0226)
Temp Days, 285-290	0.0206	0.0169	0.0258	0.0257
	(0.0288)	(0.0283)	(0.0263)	(0.0263)
Temp Days, 290-295	0.0832**	0.0817**	0.0937**	0.0937**
	(0.0402)	(0.0400)	(0.0416)	(0.0416)
Temp Days, 295-300	0.00547	0.00387	0.00505	0.00512
	(0.0344)	(0.0339)	(0.0339)	(0.0341)
Temp Days, 300-305	-0.171**	-0.171**	-0.154**	-0.154**
	(0.0750)	(0.0738)	(0.0739)	(0.0747)
Temp Days, 305-310	-0.607*	-0.600**	-0.568**	-0.568**
	(0.308)	(0.303)	(0.284)	(0.284)
Temp Days, 310-315	3.420***	3.285***	3.063***	3.067***
	(1.013)	(0.988)	(0.907)	(0.914)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	600	600	600	600
No. of Clusters	154	154	154	154
R^2	0.149	0.160	0.183	0.183

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantities of interest are the estimated effects of various precipitation-day and temperature-day binned classifications. Precipitation is in millimeters and temperature is in Kelvin. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A2.4 Additional Heterogeneous Effects: Intelligence Gathering

Table A-23: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to potential intelligence gathering via security base breaches

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0564***	-0.0574***	-0.0537***	-0.0537***
	(0.0137)	(0.0143)	(0.0141)	(0.0131)	(0.0131)
Infiltration		0.888***	0.740***	0.760**	0.761**
		(0.239)	(0.257)	(0.307)	(0.307)
Infiltration × Revenue		-0.0418***	-0.0350**	-0.0308*	-0.0308*
		(0.0131)	(0.0146)	(0.0160)	(0.0161)
Model Parameters					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
Model Statistics					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.157	0.173	0.188	0.188

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-24: Heterogeneous effects of rebel capacity on within-day randomization of indirect fire attacks with respect to potential intelligence gathering via insider attacks

	(1)	(2)	(3)	(4)	(5)
Opium Revenue	-0.0581***	-0.0461***	-0.0470***	-0.0433***	-0.0433***
	(0.0137)	(0.0110)	(0.0111)	(0.00921)	(0.00920)
Insiders		-2.026	-1.924	-1.868	-1.861
		(1.479)	(1.385)	(1.526)	(1.511)
Insiders × Revenue		-0.0932***	-0.0899***	-0.0903***	-0.0909***
		(0.0302)	(0.0296)	(0.0335)	(0.0344)
Model Parameters					
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	No	Yes
Model Statistics					
No. of Observations	600	600	600	600	600
No. of Clusters	154	154	154	154	154
R^2	0.154	0.256	0.263	0.278	0.278

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A2.5 Additional Heterogeneous Effects: Labor Scarcity

Table A-25: Effects of rebel capacity and battlefield losses on within-day randomization of indirect fire attacks, accounting for counterinsurgent operations from Tables A-2 and A-3]

	(1)	(2)	(3)
Opium Revenue	-0.0191***	-0.0200***	-0.0215***
	(0.00574)	(0.00698)	(0.00670)
Battlefield Losses		0.0486**	0.0828***
		(0.0207)	(0.0274)
Model Parameters			
Fighting Season (FS) Fixed Effect	Yes	Yes	Yes
FS Activity (levels)	No	Yes	Yes
Growing Season Activity (levels)	No	Yes	Yes
Planting Season Activity (levels)	No	Yes	Yes
FS Combat Operations (all, levels)	Yes	Yes	No
FS COIN Operations (all, levels)	No	No	Yes
Model Statistics			
No. of Observations	600	600	600
No. of Clusters	154	154	154
R^2	0.496	0.506	0.542

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects as well as controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Additional parameters are noted in the table footer. Fighting season parameters are notated with the abbreviation FS. Battlefield losses in our sample have a mean of 4.605 and standard deviation of 10.547. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A2.6 Varying Control over Attack Timing

As we describe in the main text, insurgents have varying levels of control over the timing of their attacks. Indirect fire events can be initiated at any time against stationary targets like bases or military outposts. The timing of direct fire and IED attacks cannot be controlled unilaterally by rebels. Close combat, for example, is often characterized by attacks on convoys and non-stationary targets. Troops and vehicles are rotated from locations on non-random schedules. The timing of these attacks, therefore, is consistently less random as a consequence of government strategy, not rebel tactics. Similarly, IEDs may be emplaced hours or days before they are triggered by a passing convoy. Rank ordered, rebels have the least control over the timing of roadside bomb attacks.

We replicate the visual evidence we introduce in our main results in Figures A-2 and A-3. Notice that direct fire, over which insurgents maintain some limited control over timing, is negatively correlated with revenue but the slope is consistently flatter than for indirect fire attacks (over which rebels have unilateral timing control). For IEDs, the coefficient is effectively zero. We introduce the regression-based evidence in Tables A-26 and A-27. Although direct fire attacks are consistently negatively correlated with revenue, the magnitude of the main effect (when compared to our indirect fire results) is consistently weaker. Similarly, although our point estimates are consistently negative with respect to roadside bombs, our precision is substantially diminished. These results are consistent with our main effects since rebels lack the ability to control the timing of these other attack types.

Figure A-2: Bivariate relationship between opium revenue and p-value of randomization test of combat (direct fire attacks) in Afghanistan

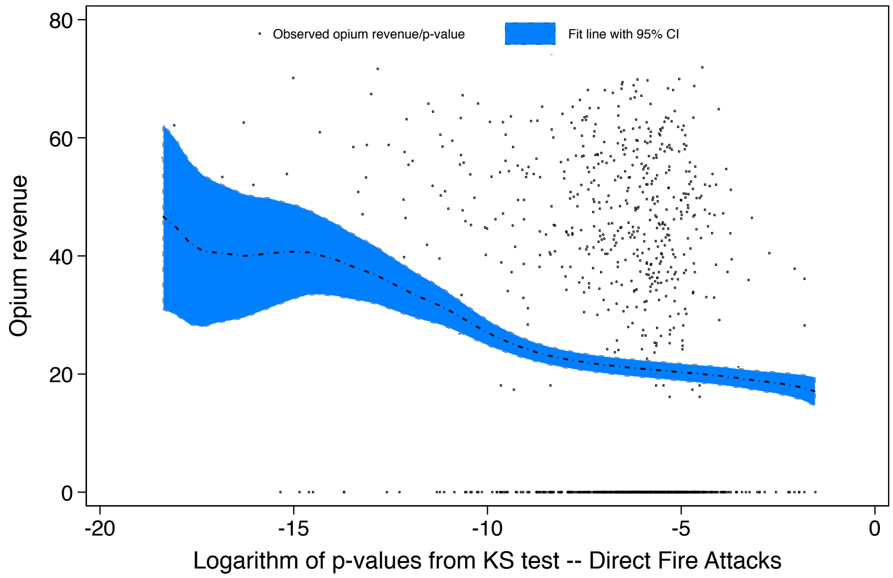


Figure A-3: Bivariate relationship between opium revenue and p-value of randomization test of combat (IED attacks) in Afghanistan

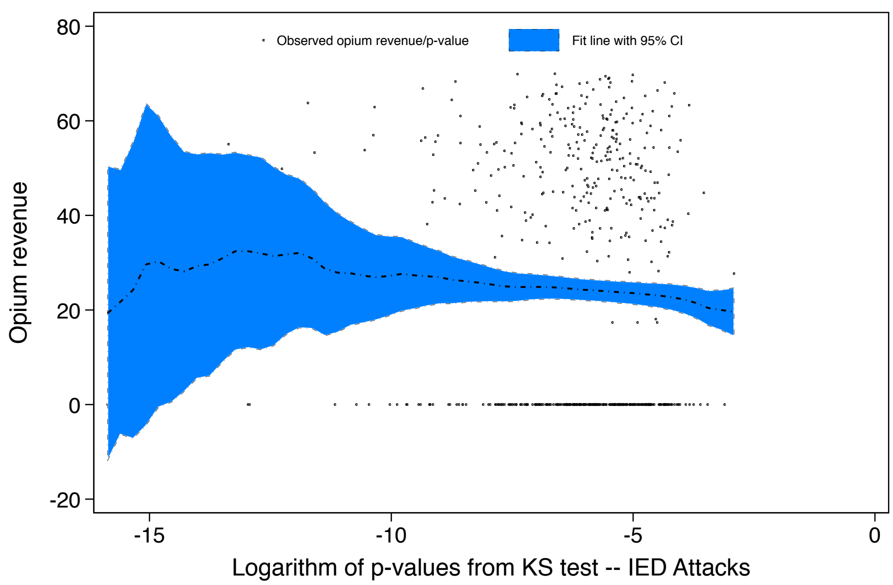


Table A-26: Impact of rebel capacity on within-day randomization of direct fire attacks

	(1)	(2)	(3)	(4)
Opium Revenue	-0.0309***	-0.0110***	-0.0121***	-0.0119***
	(0.00780)	(0.00292)	(0.00328)	(0.00327)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	1128	1128	1128	1128
No. of Clusters	236	236	236	236
R^2	0.0963	0.448	0.450	0.464

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects. Column 2-4 add controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A-27: Impact of rebel capacity on within-day randomization of IED attacks

	(1)	(2)	(3)	(4)
Opium Revenue	-0.00849**	0.000559	-0.000232	-0.00000219
	(0.00385)	(0.00259)	(0.00280)	(0.00270)
MODEL PARAMETERS				
Fighting Season Fixed Effect	Yes	Yes	Yes	Yes
Fighting Season Activity (levels)	No	Yes	Yes	Yes
Growing Season Activity (levels)	No	No	Yes	Yes
Planting Season Activity (levels)	No	No	No	Yes
MODEL STATISTICS				
No. of Observations	653	653	653	653
No. of Clusters	161	161	161	161
R^2	0.0700	0.179	0.186	0.186

Notes: Outcome of interest is the (log) p-value of the randomness test. The quantity of interest is opium revenue for a given district-year. All regressions include fighting season fixed effects. Column 2-4 add controls for the intensive margin of fighting during the fighting, harvest, and planting seasons respectively. Heteroskedasticity robust standard errors clustered by district are reported in parentheses. Stars indicate *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.