

# International Migration Amid Insurgent Violence

## (Job Market Paper)

**Ryan Ellis\***

January 28, 2025

[\[Link to most recent version\]](#)

### **Abstract**

How does exposure to insurgent violence influence an individual's decision to leave their home country? To answer this question, I combine anonymous mobile call detail records with data on the location and timing of violent events in Afghanistan to identify individual exposures to insurgent actions. I propose a novel method to detect international migration events that combines inferences from users' spatiotemporal movement patterns and episodes of sample attrition. Estimating a migration choice model reveals that a unit increase in the rate of monthly exposures to insurgent violence raises the odds of international migration by 6.4%. I show evidence that this effect is due more to extreme, high-casualty events than to the gradual accumulation of exposures to common, low-intensity events. In contrast, *internal* migration is unaffected across all violence types and intensities. I explore how wealth and employment composition might explain the differential effects. These findings provide policy-relevant insights into how potential migrants respond to insurgent violence.

**Keywords:** Migration, Conflict, Big Data

**JEL Codes:** 015, D91, D74

---

\*School of Economics, Georgia Institute of Technology, Atlanta, GA 30332. I thank Robert Gonzalez, Daniel Dench, Stephen O'Connell, Olga Shemyakina, and Max Rosenthal for their comments and guidance. I am also grateful to Dylan Brewer, Tibor Besedes, Mayra Pineda-Torres, Austin Wright, Jean-François Maystadt, Giovanni Peri, Catalina Amuedo-Dorantes, Xiao Hui Tai, Aidan Milliff, Mohammad Isaqzadeh, and participants at the 2024 Annual Meeting for Empirical Studies of Conflict, the 2024 UNU-WIDER Development Conference, and the 2023 AlianzaMX Summer School on the Economics of Migration.

# 1 Introduction

In 2023, approximately 117.3 million people worldwide lived displaced from home due to conflict, violence, human rights abuses, or persecution (United Nations High Commissioner for Refugees, 2024). Rates of violent conflict continue to rise globally, creating conditions that drive even more people to leave their homes (World Bank, 2020). Migration due to violence is not only a critical humanitarian concern, but also a polarizing political issue in both host and origin countries. For these reasons, an improved understanding of the nature of violent conflict as a migration push factor is critical, especially for migration across international boundaries. Despite its importance, research on the issue remains limited. This gap exists for two primary reasons. First, external migration is challenging to measure in data-scarce environments, such as those where violence-driven migration is often most prevalent. Second, it is difficult to reliably measure individual-level exposure to violence.

This paper studies the impact of exposure to violence on an individual’s decision to migrate in the context of Afghanistan during a peak period of the Taliban insurgency, 2010-2012. In doing so, I introduce a novel approach to the data challenges described above. First, using high frequency call detail records from a major Afghan mobile network operator to approximate the daily locations of individuals, I identify cross-border migration events through a set of spatial and temporal classifying criteria related to phone users’ movement patterns at the time of their attrition from the data. To my knowledge, this paper is the first to use call detail records to infer migration events across international borders. Second, I match user locations with previously classified, geolocated, and time-stamped significant actions event data (SIGACT) to generate an individual-by-day measure of exposure to insurgent violence. I estimate a discrete choice model of migration that is flexible to the inclusion of both internal and international locations.

I find that a unit increase in an individual’s monthly rate of exposure to insurgent violence is associated with an increase in the odds of migrating internationally between 6.3% and 19.3%. The estimated range is illustrative of a conservative bounding exercise in my

classification of migrants with call detail records, which requires careful researcher choice of tuning parameters. When allowing for a larger radius of exposure to violent events (moving from 5km to 20km), the effect on migration shrinks to an odds increase between 0 and 2.6%. I explore possible response heterogeneity by restricting violence to especially high-casualty events (10+ killed or wounded), revealing estimated odds increases that are higher by more than an order of magnitude—from 6.3% for all events, to 90% for high-casualty events—suggesting that the low-level accumulation of exposure may be much less salient to decision-makers than single extreme shocks.

Interestingly, there is no discernible effect on rates of internal migration subject to the same changes in exposure, even in response to high-casualty events. I explore the role of income and other determinants to explain the differential effects, showing that mobility (a proxy for income) increases the odds of internal migration only. The main results are robust to expanding the location choice set to include internal and international destinations. I provide additional consistent evidence via a Cox proportional-hazards model, allowing an interpretation that is sensitive to time.

The majority of previous micro-level work on violence and migration has been limited by national borders, within which only internal migration is observable. When one leaves their country of origin, they are no longer observable in its administrative datasets, censuses, or national surveys. Similarly, data collected by host countries rarely contain the geographic detail necessary to construct meaningful measures of a migrant’s previous life in their country of origin.<sup>1</sup> A related literature makes use of aggregate-level country statistics of the stock and flow of migrants to estimate determinant factors, but is less appropriate for individual or household-level decision models.<sup>2</sup> By joining the benefits of aggregate-level studies that observe true migration events with those of micro-level studies that observe pre-migration

---

<sup>1</sup>Exceptions include surveys produced by scholars for the express purpose of studying immigration in specific contexts, such as the Mexican Migration Project.

<sup>2</sup>See Tai et al. (2022) for a recent example that also uses digital trace data, focused on internal migration. See Shaver et al. (2024) for an aggregate-level approach, and Buggle et al. (2023) for a creative solution using historical records.

histories, this project provides a valuable contribution to our understanding of the role of violence in migration decisions. The classification method for identifying international migration events is itself a further contribution of the paper, as it is generalizable for use across migration contexts.

One recent meta-analysis of both micro-level and aggregate-level migration scholarship notes that a majority of studies find violence to be central among displacement push-factors, while the economic or policy-related pull factors of host locations may be less important (Shaver et al., 2024). Several examples find consistently positive effects of violence on migration across distinct settings (Kirchhoff and Ibanez, eds, 2002; Ibáñez and Vélez, 2008; Orozco-Aleman and Gonzalez-Lozano, 2018; Buggle et al., 2023). While the overall direction of this effect is obvious in a rational choice framework, many studies have documented heterogeneous and even counter-intuitive results. Bohra-Mishra and Massey (2011) show evidence for a nonlinear threshold theory of violence—that at low levels, exposures to violence actually reduce the likelihood of leaving, until a threshold is met and the sign switches. Milliff and Christia (2022) show how measures of social centrality can have a strong moderating influence on other determinants of migration, including exposures to violence. Notably for this project, Basu and Pearlman (2017) find that violent events can influence international movement, but appear to be unrelated to internal migration decisions. This paper provides evidence that supports their conclusion.

Other recent works have turned to nontraditional data in settings where censuses and surveys are scarce or unsatisfactory, borrowing from information and communications technology, demography, physics, and other fields where necessary. Blumenstock (2012) first developed a method to infer internal migration events from call detail records. The methods of measurement developed in that paper have since been updated or refined (Chi et al., 2020; Blumenstock et al., 2023) and used in subsequent research related to violence and migration (including the aforementioned work of Milliff and Christia (2022)). Tai et al. (2022) finds violence plays a vital but nuanced role in internal migration episodes in Afghanistan. The

authors find the odds of migration increase much more for high-casualty events, especially those credited to insurgent groups who were perceived to be more extreme. While these previous works focused exclusively on using call detail records to infer internal migration, this paper presents a novel method for using CDR to identify external migration. This creates a valuable link for migration researchers attempting to match post-migration outcomes to individual characteristics and behaviors in their country of origin.

The remainder of the paper is structured as follows: Section 2 provides historical background on violence in Afghanistan. Section 3 describes the data and reports summary statistics. Section 4 introduces the classifying criteria for cross-border migration events. Section 5 outlines a model for displacement choice, and 6 contains a framework for estimation, as well as the main results. Section 7 discusses the relevance of the contributions to researchers and policymakers, then concludes.

## **2 Setting**

### **2.1 Historical background**

The setting of this study is Afghanistan from the end of 2010 until mid-2012. The half-century prior to this period was characterized by almost constant war, both internally and with outside actors. In 1978, an Afghan civil war was absorbed into the broader international landscape of the Cold War. The Soviet Union invaded, occupied, and waged war in Afghanistan against Islamist rebel groups supported by the United States and Pakistan, among others. Estimates of Afghan civilian deaths are generally around 1.5 to 2 million in the period until 1989, with up to 6 million forcibly displaced as refugees, mostly to Iran and Pakistan, and an additional 5 million displaced internally (Kakar, 1997). The country still has not recovered even half of the 6 million displaced during its Soviet era, and return migration has stalled in recent decades. As of 2021, only Syria had more of its citizens living as refugees in other nations (United Nations High Commissioner for Refugees, 2022). Following

the collapse of the Soviet Union, several groups attempted to gain power in the vacuum left behind. The Taliban, a militant group of Islamist scholars and former anti-Soviet fighters, first took control of Kandahar City in 1995 following a surprise attack. With support from Pakistan’s Inter-Service Intelligence (ISI), they quickly gained popularity and territory, establishing themselves as the leaders of a renamed Islamic Emirate of Afghanistan one year later in 1996. They ruled harshly until 2001, when they were ousted by a U.S.-supported coalition that included Afghanistan’s Northern Alliance.

## 2.2 The Taliban insurgency

During the 2000s, the Taliban regrouped in hiding, and by 2010 were an active and threatening insurgency, prompting the Obama administration to commit 100,000 troops to the country (Stanford University, 2018). This is the approximate window during which the events of this study take place. In the south near the once *de facto* capital Kandahar, in population centers near Kabul, and in the eastern provinces patrolled by the violent offshoot Haqqani network, the Taliban were destructive and relentless. They employed a deliberate strategy to make the state appear weak. They threatened voters, then followed through with violence during elections. They attacked city centers that offered no feasible territorial gain, chosen only to spread terror and sow mistrust. They used surprise violence to erode public faith in local and national government, and to humiliate the cooperating international coalition (Condra et al., 2018).

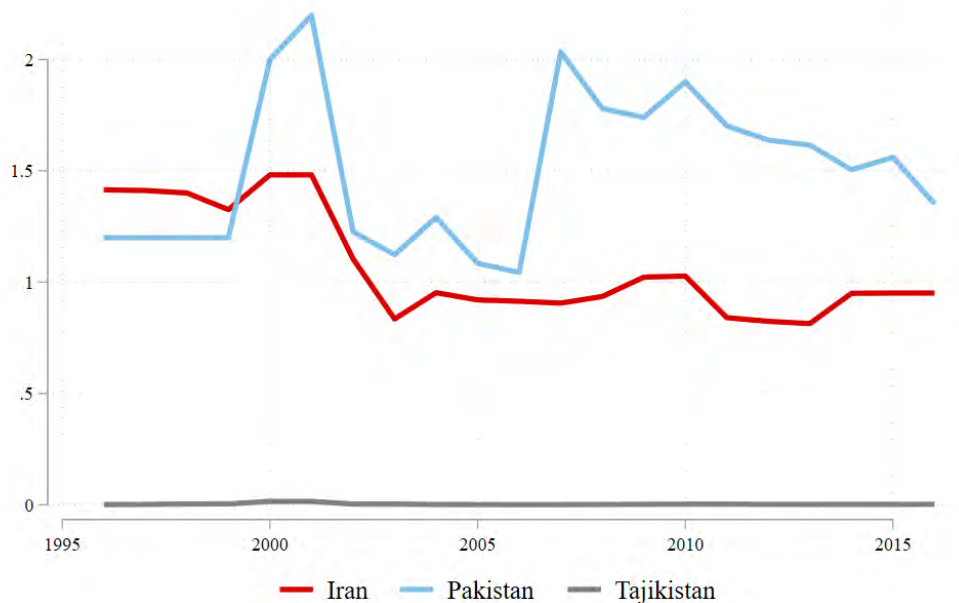
The majority of Afghan refugees from both the Soviet period and the Taliban insurgency live in Pakistan and Iran. The next most common host country that also shares a land border with Afghanistan is Tajikistan, shown for reference in Figure 1.<sup>3</sup> While the stock of refugees began decreasing around 2010, many continued to displace across borders, even as return migration in certain regions outpaced the rate of exits. The border dividing Pakistan and Afghanistan is notable in a migration study, as its establishment in 1893 created an artificial

---

<sup>3</sup>Millions of asylum-seekers from Afghanistan also reside in nations across the world, primarily in Europe, but they fall outside the scope of this study.

divide through the lands of the Pashtun people. Many refuse to recognize the border (called the Durand Line). The shared culture, language, and history of this region are all obvious positive determinants for migration choice, especially if there are different rates of violence on either side of the border.

Figure 1: Millions of Afghan refugees in bordering countries



*Notes:* The stock of Afghan refugees in Pakistan, Iran, and Tajikistan from 1995 to 2015. Data are from the United Nations High Commissioner for Refugees (2022). The units on the vertical axis represent millions.

While the Taliban insurgency is a unique setting, its human and economic costs are generalizable to other conflicts and crises. The World Bank estimates that by 2030, two thirds of the world's population could live in settings characterized by fragility, conflict, and violence (World Bank, 2020). A global pandemic and the return of land invasion-style warfare between nation-states are two recent shocks that threaten the stability and well-being of vulnerable populations. It is urgent that conflict and development research adopt the full array of new tools and methods at its disposal to address these issues. That includes new sources of data, especially in settings where administrative data are scarce. Afghanistan, for

instance, last conducted an official census in 1979. Satellite imagery, social media metadata, cell phone records, and other digital trace data have unique limitations, some more obvious than others, as discussed below. Even still, their scope, consistency, and ubiquitous presence in almost every country make them helpful in diagnosing and solving policy issues.

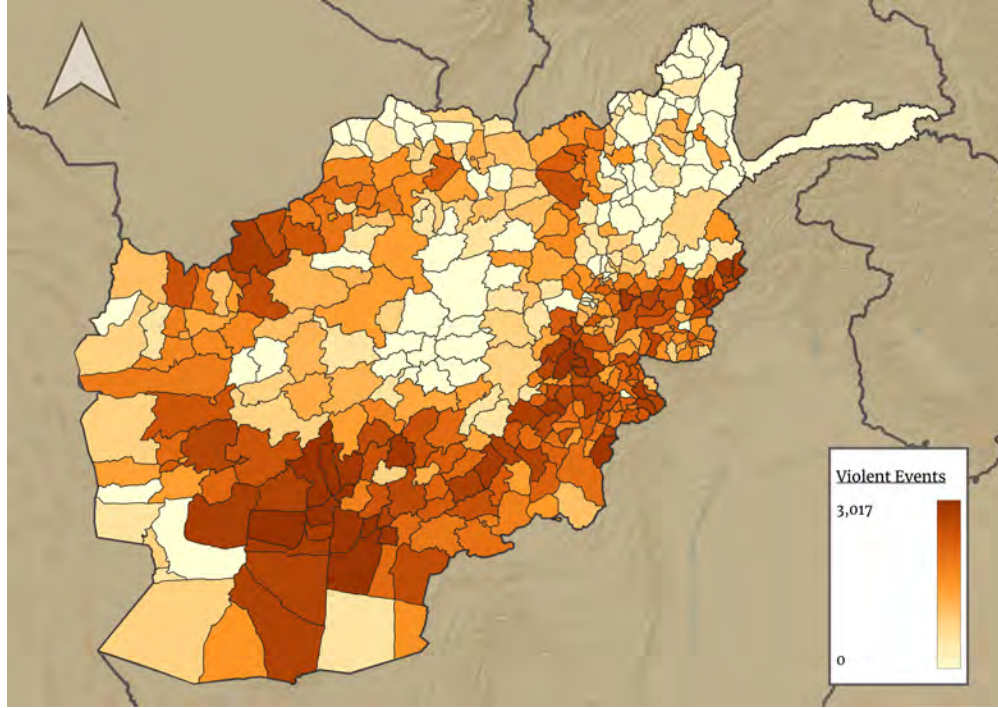
## 3 Data

### 3.1 SIGACT

There are two primary data sources for this study. The first is the violent events dataset, compiled from the U.S. Department of Defense’s Significant Activities database (SIGACT) during *Operation Enduring Freedom*, the department’s name for both the war in Afghanistan and the more broadly defined US war against terrorism. SIGACT data in Afghanistan were recorded by Afghanistan’s own military and police units, as well as by cooperating forces of the North Atlantic Treaty Organization’s (NATO) International Security Assistance Force (ISAF).

Violent events are categorized by type of attack (direct fire, indirect fire, improvised explosive device (IED), surface-to-air attack, and more) and are catalogued with high temporal and spatial precision. They include instances of engagement between insurgents and the ISAF counterinsurgency, as well as events of civilian or built-environment targeting. The systematic recording procedure provides an advantage in accuracy and completeness over databases of violent events compiled from media reports (Condra et al., 2018). SIGACT data from Afghanistan have been used in conflict research since 2015 (Sexton, 2016; Trebbi et al., 2020; Blumenstock et al., 2024; Grasse et al., 2024). Figure 2 illustrates the geographical spread of insurgent violence coded in SIGACT during the sample period of this study, which is concentrated along a ring shaped network of highways connecting population centers, with less activity in the country’s remote center.

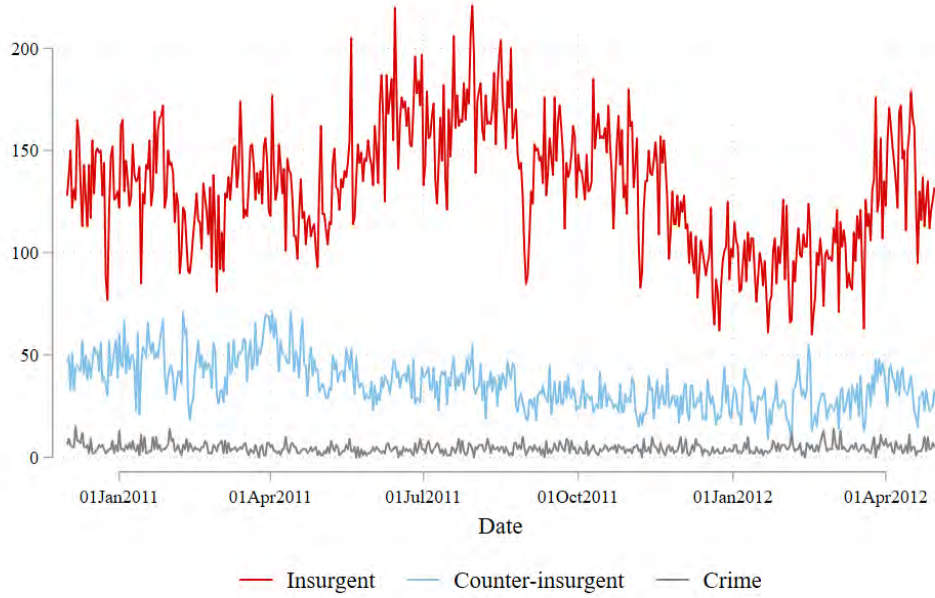
Figure 2: SIGACTs in Afghan districts during the sample period



*Notes:* Darker shaded districts experience more violence than lighter ones. Though SIGACT data contains several event types, I define violent exposure to include only direct fire, indirect fire, successful IEDs, and surface-to-air strikes. Notably, the category only includes violence accomplished by the insurgency.

This project is primarily concerned with insurgent violence, but SIGACT does contain records of counterinsurgent actions and to a lesser extent, civilian crime. Figure 3 shows the relative frequency of activity in SIGACT by actor (insurgent, counterinsurgent, or civilian). For a breakdown of insurgent violence only, see Figure 4: a timeline of the four main insurgent attack types. Direct fire events are the most commonly recorded. Violence types with higher production costs are less common, and it is unclear how civilians might respond differently to seeing surface-to-air missiles, improvised explosives, or mortar shelling, as opposed to gunfire. For simplicity in the main analysis, I group all insurgent actions together as a total violence indicator.

Figure 3: Total SIGACTs by actor during sample period



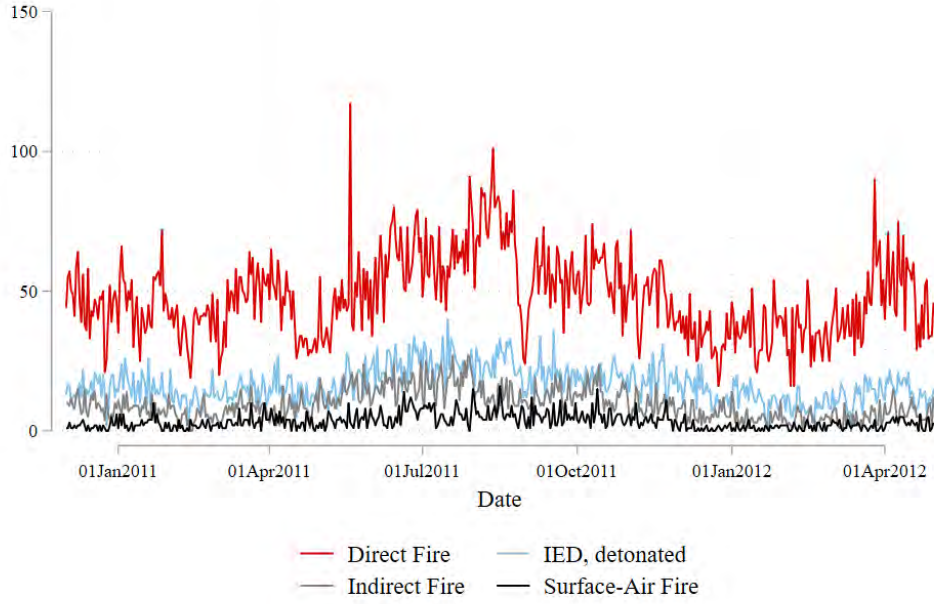
*Notes:* This graph compares all SIGACTs by actor, without the author’s specified violence categorization. Insurgent actions therefore include threats, found and cleared IEDs, and all the categories shown in Figure 4. Crime is committed by non-combatant civilians.

### 3.2 Call detail records (CDR)

The SIGACT event records are spatially and temporally merged to a set of call detail records (CDR) from the largest Afghan cellular service provider, Roshan. Specifically, the records are from subscribers to Roshan’s mobile banking service, M-Paisa. The data are provided in a user-by-day panel, allowing clustering for statistical inference at the individual level. The 17-month panel of over 15,000 users gives just under 3.5 million observations. Importantly, users do not enter or exit the sample in a uniform pattern.

Of course, there is selection into cell phone ownership, and again into mobile-banking, both of which could hamper the external validity of the study. In 2011, 46% of all Afghans had a cell phone, and 9% had a mobile bank account (World Bank, 2015). However, there are reasons to assume a high level of variation among M-Paisa users, even if the sample may not perfectly represent the population as a whole. One might expect an outsized portion

Figure 4: Insurgent violence by type during sample period



*Notes:* Though SIGACT data contains several event types, I define a binary variable for violent exposures that includes only direct fire, indirect fire, successful IEDs, and surface-to-air strikes. Each of these describe acts of violence accomplished by insurgent actors only.

of the mobile banking users to be wealthy early adopters of new technology. In fact, the M-Paisa rollout included contracts with varied groups such as the Afghan National Police, Roshan’s employees, and humanitarian groups that distribute cash aid (Blumenstock et al., 2024). Even with promising variation in user type, the conclusions of this paper should be understood to be susceptible to some amount of selection bias. However, because the call detail records are anonymous and contain no additional user traits, the likely direction and magnitude of the bias is ambiguous. This is a common limitation of digital trace data like call records or social media posts, where data are generated by normal use of a consumer product. As the penetration rate of the product increases in a given research setting, digital trace data become more fully representative.

Call detail records are unique among big data types. Unlike a geotagged Facebook post, which utilizes a phone’s GPS system, the latitude and longitude reported by CDR data are

not the exact location of a user at a given time. Instead, these data are records of calls, texts, or transaction activities matched to the location of the cell tower that transmits the corresponding signal. The user’s location is inferred with various measures of centrality, discussed below. The number and locations of towers is thus vital in understanding the accuracy with which call detail records map to reality. By 2011, cell towers were ubiquitous in all populated areas of Afghanistan. Their distribution is not random, but they provide a level of fine granularity acceptable for large sample generalizations (Blumenstock, 2012; Chi et al., 2020). As shown in Figure 5, cell towers in Afghanistan map onto population centers. The general ring shape of the distribution of cell towers closely follows the highway system linking major cities. In the aggregated panel used for this study, tower coordinates corresponding to each phone use event are averaged over a 1-day period, reported as a daily center of gravity (COG). That is, for cell phone user  $i$ , who makes or receives  $K$  calls on day  $t$ , define her daily center of gravity as,

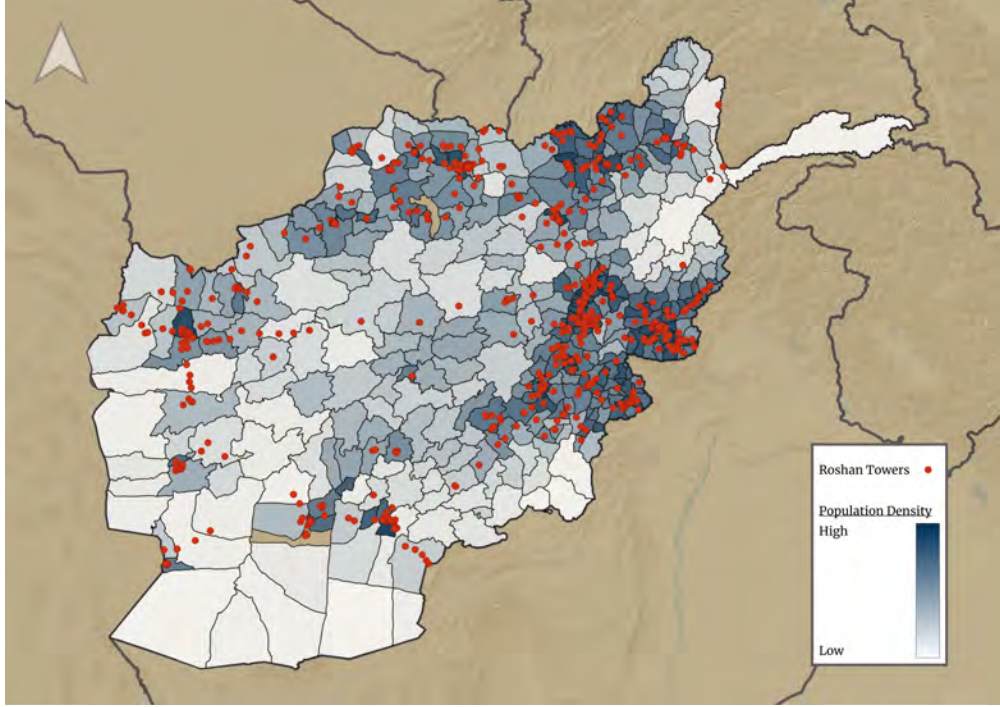
$$\frac{1}{K} \sum_{k=1}^K r_{itk} \quad (1)$$

where  $r_{itk}$  is the vector of latitude and longitude coordinates,  $\begin{pmatrix} lat \\ lon \end{pmatrix}$ , associated with the tower location of each call  $k$ . Thus, each user in the data has exactly one observation per day in sample, and the associated coordinates represent an average location derived from the nearest towers. In the available data, days with no usage are imputed with the previous day’s COG. See Appendix A for further discussion of imputation in the sample, including statistics.<sup>4</sup> I create several versions of the data, subject to restrictions on imputations, and find that results are largely consistent. That said, I drop individuals with a total time in sample less than 31 days, as my preferred definitions of migration events require at least one consecutive month of residency to establish the home location from which one could migrate.

---

<sup>4</sup>The work of collapsing M-Paisa data to daily averages, COG imputations, as well as the merger with SIGACT events, was done upstream of this research use case. See Blumenstock et al. (2024) for more information on the data.

Figure 5: Roshan cell towers in Afghanistan, 2011



*Notes:* Red dots denote cell tower locations, which were obtained from Roshan. Population density was obtained from CIESIN - Columbia University (2016). Shading shows density at the district level, with high-density areas in darker blue.

### 3.3 Measurement

The M-Paisa call detail records are merged at the daily level with SIGACT violent events. For user  $i$  at time  $t$ , define a binary exposure variable  $D_{it}^{\rho}$  equal to one (1) if her daily COG was within some radius  $\rho$  of a SIGACT event. In this study,  $\rho$  takes on only the discrete values of 5 and 20 kilometers.  $D_{it}^{\rho}$  can also be broken into attack types, e.g. direct fire, indirect fire, or IED attack. I find that most specifications provide consistent results across attack types, and for simplicity use an indicator that includes all of them. From  $D_{it}^{\rho}$ , I construct a count-style variable, as users in areas with high target-value can be exposed to multiple events per day. I then normalize the count of exposures by a measure of time to simplify

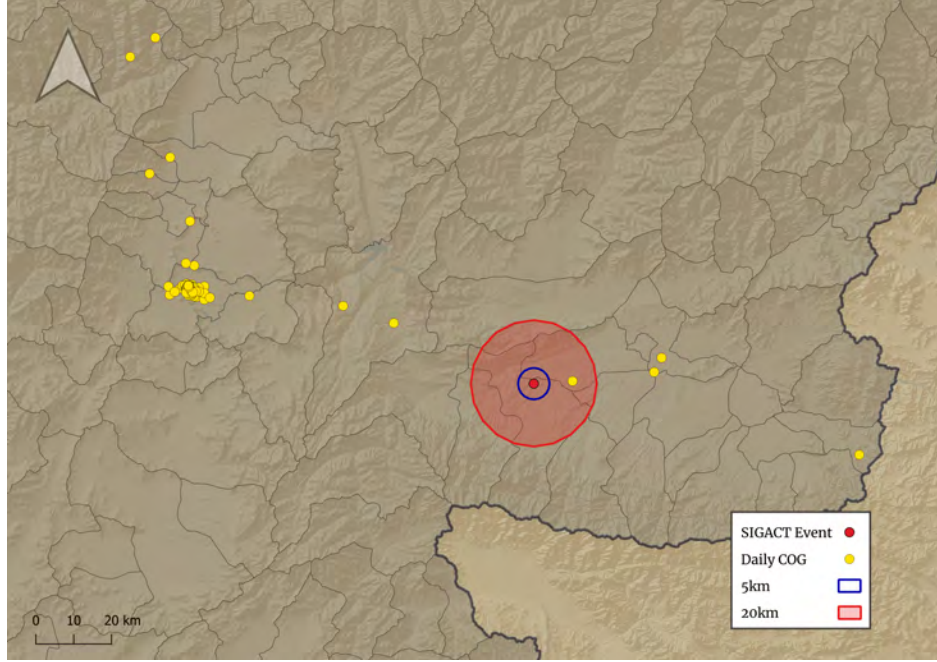
interpretation of regression results. Table 1 describes the high frequency of violent exposure in this setting. 89% of M-Paisa users are exposed within 5km to at least one insurgent event during the sample period. On average, users are exposed 0.06 times per day, approximately twice per month. Figure 6 provides an illustration of how SIGACT events are matched to daily centers of gravity to form the exposure variable.

Table 1: 5km exposures to violence, by individual user

|                                            | Mean  | SD    | Med  | Max |
|--------------------------------------------|-------|-------|------|-----|
| <i>Panel A: Violence (Ever exposed)</i>    |       |       |      |     |
| All violence                               | 0.89  | 0.31  | 1.00 | 1   |
| Direct fire                                | 0.80  | 0.40  | 1.00 | 1   |
| Indirect fire                              | 0.39  | 0.49  | 0.00 | 1   |
| Surface-air                                | 0.48  | 0.50  | 0.00 | 1   |
| IED                                        | 0.75  | 0.43  | 1.00 | 1   |
| <i>Panel B: Violence (Total exposures)</i> |       |       |      |     |
| All violence                               | 13.84 | 27.66 | 6.00 | 399 |
| Direct fire                                | 6.15  | 16.57 | 2.00 | 305 |
| Indirect fire                              | 1.48  | 3.35  | 0.00 | 47  |
| Surface-air                                | 1.14  | 1.69  | 0.00 | 19  |
| IED                                        | 5.06  | 10.50 | 2.00 | 106 |
| <i>Panel C: Violence (Per day)</i>         |       |       |      |     |
| All violence                               | 0.06  | 0.09  | 0.04 | 2   |
| Direct fire                                | 0.03  | 0.06  | 0.01 | 1   |
| Indirect fire                              | 0.01  | 0.01  | 0.00 | 0   |
| Surface-air                                | 0.01  | 0.01  | 0.00 | 0   |
| IED                                        | 0.02  | 0.03  | 0.01 | 1   |

*Notes:* Excludes data from imputation gaps. Data are collapsed to the individual user-level. For restricted samples that further limit the number of imputed observations, see Appendix A. For the same statistics for 20km exposures to violence, see Table 6 in Appendix B.

Figure 6: Identifying exposure to violent events



*Notes:* Yellow dots show a one-month spell of locations for a randomly-selected Roshan user. The red dot represents a violent event from SIGACT. The location of the red dot is for illustrative purposes only, and does not represent an actual event from the data. A single yellow dot sits inside the red circle, which represents a 20km radius from the hypothetical event. This user-day observation would receive a 1 for the binary variable indicating a 20km exposure. As illustrated by the blue circle, that same user-day observation would be coded 0 for the variable indicating a 5km exposure.

One strength of CDR is their precision. As given, the coordinates have a resolution of about one square meter. Researchers have noted that in areas with a high density of cell towers, movements can be falsely recorded among nearby towers (Milliff and Christia, 2022). To account for this, I coarsen the data to a resolution of about one square kilometer to capture changes in centers of gravity that are appropriately distant from one another for the measurement of migration events. The result is similar to imposing a grid of 1km cells over the area of the country to track discrete user locations. I follow Blumenstock (2012) and others, coding each user's monthly modal location as their home center of gravity. Results do not differ substantially with other denominations of home, such as monthly mean location. I encode a unique value to each ordered lat-lon pair visited in the entire dataset,

and calculate the modal pair for each user, every month they are in the sample. I then calculate distances (km) between users’ daily COG and home location (Picard, 2019). Even though the observation is technically an average of tower locations, I refer to it as the *user’s* COG for simplicity. Other variables are generated from the distances and coordinates, such as a binary indicator for being outside one’s home district or province. Another variable, the user’s *radius of gyration* (ROG), is a statistic constructed to represent the geographic extent of one’s movements over time. The ROG is the root mean square distance of all locations a user visits from their home center of gravity.

## 4 Classifying cross-border migration events

### 4.1 Motivation

Researchers using administrative or survey data at the individual or household level are rarely able to identify migrants themselves, for the simple fact that most countries’ statistical agencies collect data only within their national borders. Consider an individual from country  $A$  who migrates to country  $B$ . They might be present in data collected by  $A$  up until the date of migration, but as a resident of  $A$  only. Their status as a migrant is a future development; it is concurrent with their disappearance from the data  $A$  collects. Once in country  $B$ , the same individual could be identified as a migrant in a census or survey within  $B$ , but often with little to no information about their life before arrival. A researcher interested in the determinants of migration (or the welfare of migrants in their host countries) would prefer individual or household data with information from both  $A$  and  $B$ . While many host countries do identify noncitizens in their statistical records, there is rarely enough additional information to meaningfully link individuals to anything more than their country of origin. The American Community Survey, for example, collects data on foreign-born residents, but only identifies one’s place-of-origin from among six continental regions. This project is the first to use CDR data to infer international migrations. Subsequently, the method provides

a potential solution to the data problem described above, allowing researchers to observe both migration events and pre-migration histories at the individual level. This is especially useful in settings like Afghanistan where the last government census was collected in 1979. While researchers do conduct surveys in Afghanistan, I am aware of none that attempt to measure out-migration from interviews at border crossings or airports, or by other means.<sup>5</sup>

## 4.2 General method

There are several challenges to inferring international migration from call detail records. I propose a method based on identifying attrition events that satisfy certain filtering restrictions, described below. Roshan does not record phone activity outside of Afghanistan, so one plausible reason for user attrition is that they have left the country by crossing a land border.<sup>6</sup> Of course, there are many reasons why a user’s activity could disappear from the data. I argue that most or all of these concerns are assuaged by the conditions that follow. In other words, each of the following restrictions serves to decrease the likelihood that an attrition event is due to reasons other than the user crossing an international border.<sup>7</sup>

Let  $r_{iT}$  represent a user’s location on their final day in the sample. I label user  $i$  an external migrant if the following conditions hold:

$$(r_{iT} - COG_{iT}^h) > \delta * ROG_i \quad (2)$$

$$(r_{iT} - r_{cross}) < \gamma \quad (3)$$

$$T_i < T - \epsilon \quad (4)$$

---

<sup>5</sup>The 2011 Afghanistan National Risk and Vulnerability Survey contains a question about the migration of family members.

<sup>6</sup>Even if a mobile network operator provides service on both sides of an international border, the networks themselves are bounded and discontinuous. When a user crosses the border with continuous service, this is accomplished by roaming; they drop from the origin country records no matter what.

<sup>7</sup>Some possibilities include nonpayment of the phone bill, canceling of the contract, or destruction of the phone or SIM card.

$$\frac{1}{S} \sum_{s=T_i-S}^{T_i} \left[ (r_{i,s+1} - r_{cross}) - (r_{is} - r_{cross}) \right] < 0 \quad (5)$$

where  $COG_{iT}^h$  is the user's home location in the month of their attrition,  $\delta$  is a flexible scaling parameter, and  $ROG_i$  is the user's average radius of gyration. The inclusion of  $ROG_i$  accounts for differences in typical mobility across individuals. This condition simply requires a user be suitably far from home, moderated by the limits of their average movement. In equation 3,  $r_{cross}$  is the location of a known border crossing (see Figure 7), and  $\gamma$  is a flexible distance parameter. This condition requires a user be sufficiently near a border crossing point before attrition. In (4),  $T_i$  is the date of the user's final day in the sample,  $T$  is the full sample final date, and  $\epsilon$  is a flexible time parameter. This condition removes the possibility that an attrition event is simply due to the end of the sample. Finally, in (5), I define  $s \in S$  to be a set of lead-up days before attrition. Since each user is closest to a single  $r_{cross}$  on their final sample day, I trace their movement toward the border crossing over  $S$  days, requiring that on average they move toward it, rather than away. This filters remaining individuals who may simply live near a border crossing, or those who have extended periods of non-usage leading up to attrition. To summarize, these conditions imply the following: in order to be classified as an external migrant, a user must be a certain distance from home on their final day in sample, and simultaneously near a known border crossing. Additionally, I disqualify users whose final day is one of the last  $T - \epsilon$  sample dates, i.e. they are not dropping out simply because they recorded no phone usage in the final  $\epsilon$  days of total sample time. The condition described in equation (5) stipulates that users' day-over-day distance to a final border crossing point is decreasing in the  $S$  days leading up to their attrition from the data. Table 2 summarizes mobility and migration outcomes in the Roshan sample, and introduces two classifications of external migrant based on the tuning parameters in the classification scheme above. Internal migrations are identified using the method described in Blumenstock et al. (2023), whereby users' home locations change to one in another province for at least one month.

Table 2: Migration statistics, by individual user

|                           | Mean  | SD    | Med  | Max |
|---------------------------|-------|-------|------|-----|
| <i>Panel A: Mobility</i>  |       |       |      |     |
| Km from home              | 10.91 | 23.48 | 3.02 | 378 |
| Radius of gyration (km)   | 19.09 | 37.06 | 5.73 | 500 |
| Outside home district     | 0.11  | 0.14  | 0.05 | 1   |
| Outside home province     | 0.04  | 0.08  | 0.00 | 1   |
| <i>Panel B: Migration</i> |       |       |      |     |
| Internal migrant          | 0.07  | 0.25  | 0.00 | 1   |
| External migrant (1)      | 0.03  | 0.16  | 0.00 | 1   |
| External migrant (2)      | 0.01  | 0.08  | 0.00 | 1   |

*Notes:* “Km from home” is measured daily. Radius of gyration is a monthly statistic. Variables for “outside home district” and “outside home province” are binary and measured daily. Internal migrations are classified as in Blumenstock et al. (2023). The two types of external migrations represent a (1) lenient and (2) restrictive set of parameters, respectively. Excludes data from imputation gaps. Data are collapsed to the individual user-level. For restricted samples that further limit the number of imputed observations, see Appendix A.

There is substantial researcher choice in setting the tuning parameters of the classification criteria. As a bounding exercise, I choose the following parameter values under two different specifications, one lenient and one restrictive (see Equations 2 - 5). For the *lenient* classification of a migrant,

$$\delta = 0.1, \gamma = 75km \quad (6)$$

and the *restrictive* classification,

$$\delta = 0.2, \gamma = 50km \quad (7)$$

The chosen values of  $\delta$  adjust the moderating effect of the radius of gyration (*ROG*) on the distance from home required for classification. The more restrictive model requires that individuals be further from home than the lenient model. The choice of  $\gamma$  is particularly

setting-specific; values too large risk falsely classifying those simply visiting population centers near border crossing points. I choose 75km in the lenient classification based on the nearness of Jalalabad to the Torkham border crossing. Researchers may also adjust the value of  $\epsilon$ ; here it is set to 1 for both classifications.

### 4.3 Limitations and validation

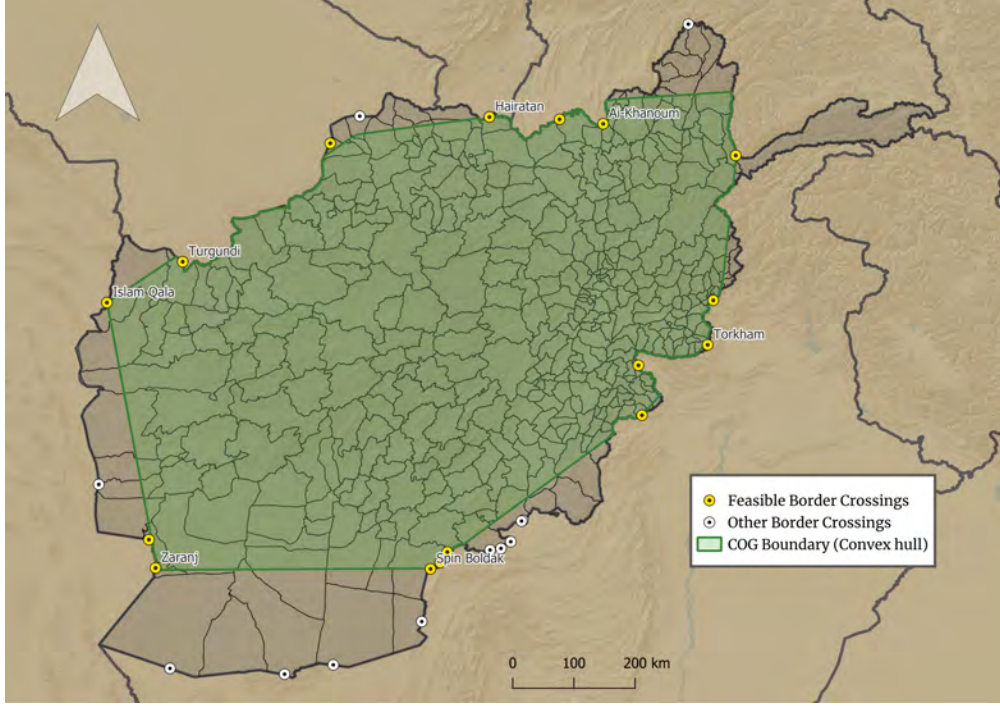
The locations of Roshan towers in 2011 covered almost all of Afghanistan’s population, but a slightly smaller proportion of its total land area. Because user centers of gravity are calculated from a daily average of tower locations, I only observe locations within the convex hull of Roshan towers intersected by the national borders of the country.<sup>8</sup> As illustrated in Figure 7, it is possible that users in the remote southwest of the country receiving service could appear in the data with inaccurate location measurement. However, as seen in Figure 5, population is very sparse in areas without coverage from Roshan. Another potential issue arising from Figure 7 is the fact that not all known border crossings fall within the coverage boundary. The white dots, especially those at the southern border, are mechanically left out of this analysis. However, the vast majority of all migration and commercial activity occurs in a relative few crossing points, specifically those shown with their names. All of these are fully within the coverage area.

The usefulness of the CDR attrition method to infer external migrations is limited by the data available to fully validate it. For example, GPS data do not suffer from the constraints of national borders, and thus offer the best potential validation for the classifying criteria. Obtaining GPS data are outside the scope of this project, but future research considering this method should seek to validate it in this way.

---

<sup>8</sup>The boundary of the convex hull is the simple closed curve with minimum perimeter that contains all the towers.

Figure 7: The intersection of Afghan borders and the convex hull of Roshan towers



*Notes:* The green shaded area is the convex hull of Roshan towers intersected with Afghanistan's national borders. This map illustrates the fact that even if an individual successfully uses their phone outside the green area, their location in the data will be coded as that of the nearest tower to them, on or near the boundary of the shaded region. Also present are known border crossing locations. Yellow dots are border crossings that are feasibly covered by the convex hull of tower locations, and white crossings are not. The external migration results do not change with the inclusion/exclusion of the crossings outside the shaded area.

## 5 A model of displacement choice

### 5.1 Binary choice

A simple model, adapted from Ibáñez and Vélez (2008) and (Chin and Cortes, 2015), is useful to illustrate the individual migration decision. An individual decides to migrate (internally or internationally) based on an inequality in expected utility between their home and a receiving area. Define  $P^M \equiv I(V_{id} - V_{ih})$  where  $I(\cdot)$  is the index function, and indirect utility  $V_i$  is indexed over a home location and a single destination location:

$$V_{ih} = \alpha S_{ih} + \beta w_{ih}, \quad (8)$$

$$V_{id} = \alpha S_{id} + \beta(w_{id} - C_{ihd}). \quad (9)$$

Individual  $i$  will migrate from home location  $h$  to destination  $d$  when  $P^M > 0$ .  $w_i$  represents location-specific wages,  $C_{ihd}$  are the migration costs from  $h$  to  $d$ , and  $S_i$  are individual perceptions of security, also unique to each location choice. First order conditions imply that utility is increasing in destination safety ( $\frac{\partial P^M}{\partial S_{id}} > 0$ ), and thus a difference in  $S_{id} - S_{ih}$  of sufficient magnitude could shift those on the margins into displacement. As expected, migration is decreasing in costs ( $\frac{\partial P^M}{\partial C_{ihd}} < 0$ ) and home wages ( $\frac{\partial P^M}{\partial w_{ih}} < 0$ ).<sup>9</sup>

## 6 Estimation and results

### 6.1 Simple and conditional logit

After identifying potential migration episodes via the methods described above, I estimate a cross-sectional logit regression. Though it cannot account for an expanded choice set, simple logit estimation is canonical in empirical migration work. In this paper, it is useful in identifying the extensive margin of international migration choice. In this specification, I identify the change in a users' odds of migrating in response to an increase in their monthly rate of violent exposures, conditional on individual and geographic controls.<sup>10</sup> In Table 3, the outcome variable is  $Pr(M_i = 1)$  where  $M_i$  is equal to 1 for users classified as migrants by the classification criteria and zero otherwise. I regress on the monthly rate of violence (as opposed to each user's total count of exposures) to control for idiosyncratic patterns of phone usage, and to allow a more natural interpretation of the results. Results are robust to

---

<sup>9</sup>Migration choice models often combine amenities such as safety into the cost term, but in this context it is useful to make explicit the disutility of non-economic concepts like violence.

<sup>10</sup>Perhaps more accurately, these are the odds of classifying a user as a potential migrant with the criteria developed in this paper. Controls include mean ROG, days present in sample, and home province indicators.

alternate modeling choices, as seen in Appendix B. The cross-sectional logit model assumes the error term has a type 1 extreme value distribution, and is estimated as follows:

$$Pr(M_i = 1|z_i) = \Lambda(z_i) = \Lambda(\alpha + \beta \times \textit{violence}_i + \theta \times \mathbf{X}_i) \quad (10)$$

where  $\mathbf{X}_i$  contains the individual-level controls: mean radius of gyration (ROG), total time in sample, share of days with phone usage, and indicators for modal province. I use  $\Lambda(x)$  as a placeholder for the sigmoid function,  $\frac{e^x}{1+e^x}$ .

Under both the lenient and restrictive parameter choices (Equations 6 and 7), the odds of an external migration event are increasing with additional monthly exposures to violence, by 6.4% in the lenient sample, and by 19.4% in the restrictive sample. Table 3 reports full results, including average marginal effects, which can be interpreted as a percentage point change in migration likelihood for the entire sample. The table contains several notable results. First, the strong positive effects of violence on the odds of external migration reported in columns (1) and (3) diminish significantly—from 6.4% to zero in the lenient sample; from 19.4% to 2.6% in the restrictive—with a larger radius of exposure to violence, shown in columns (2) and (4). Next, the effect of violence is indistinguishable from zero for internal migrants at both 5km and 20km radii of exposure. I explore possible explanations for this differential effect in Section 6.3.

The average marginal effects and proportional changes over baseline reported in the lower panel provide an intuitive change in likelihood interpretation that accounts for relative rarity of the different migration event types. Estimated with the lenient migration classification, a unit increase in monthly exposure to violence increases the migration likelihood by 5.4%. When estimated using the restrictive classification, the likelihood increase is 16.3%. The difference in magnitude between the lenient and restrictive columns illustrates the importance of researcher choice in classifying the external migration events. In this case, the tuning variables were chosen to provide conservative upper and lower bounds of the estimated effects.

Table 3: Migration and the Monthly Rate of Violence

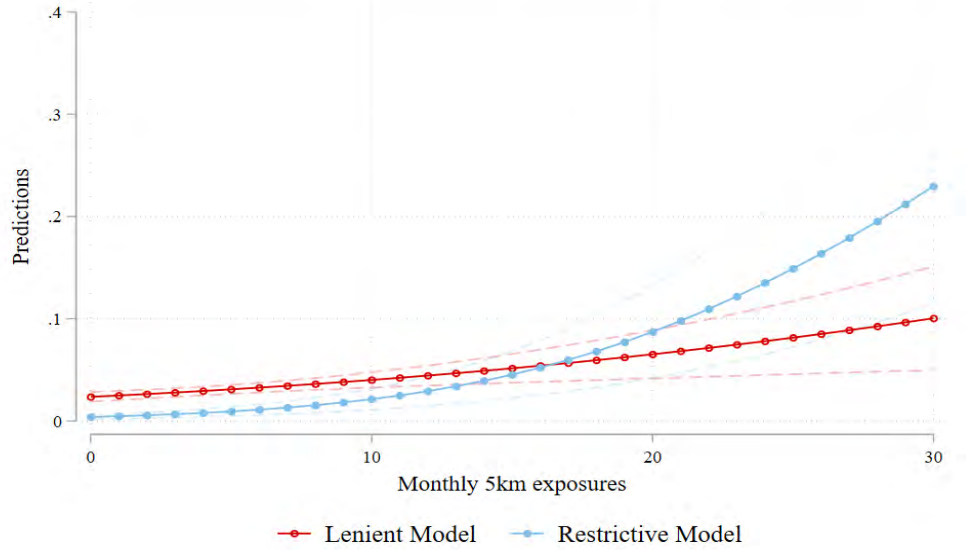
|                         | External (Lenient)  |                  | External (Restrictive) |                     | Internal            |                     |
|-------------------------|---------------------|------------------|------------------------|---------------------|---------------------|---------------------|
|                         | (1)                 | (2)              | (3)                    | (4)                 | (5)                 | (6)                 |
| Monthly 5km exposures   | 1.064***<br>(0.017) |                  | 1.194***<br>(0.023)    |                     | 1.018<br>(0.028)    |                     |
| Monthly 20km exposures  |                     | 1.002<br>(0.003) |                        | 1.026***<br>(0.006) |                     | 1.004<br>(0.006)    |
| Mean ROG (km)           | 1.000<br>(0.003)    | 1.000<br>(0.003) | 0.999<br>(0.004)       | 0.997<br>(0.003)    | 1.018***<br>(0.002) | 1.018***<br>(0.002) |
| Dependent var. mean     | 0.0266              | 0.0266           | 0.0066                 | 0.0066              | 0.0695              | 0.0695              |
| Average marginal effect | 0.0014              | 0.0000           | 0.0011                 | 0.0002              | 0.0011              | 0.0003              |
| $\Delta$ over baseline  | 0.0536              | 0.0016           | 0.1630                 | 0.0243              | 0.0151              | 0.0037              |
| Elasticity              | 0.1039              | 0.0210           | 0.3030                 | 0.2951              | 0.0293              | 0.0469              |
| Mean monthly exposure   | 1.74                | 11.52            | 1.74                   | 11.52               | 1.74                | 11.52               |
| Observations            | 12532               | 12532            | 12532                  | 12532               | 12532               | 12532               |
| Geographic controls     | YES                 | YES              | YES                    | YES                 | YES                 | YES                 |

*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria, while columns (3) and (4) use the “restrictive” version. Columns (5) and (6) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of violent exposures at either 5km or 20km.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

There are plausible reasons to select different parameter values, unique to each migration research setting.

Figure 8: Marginal change in predicted log odds of external migration

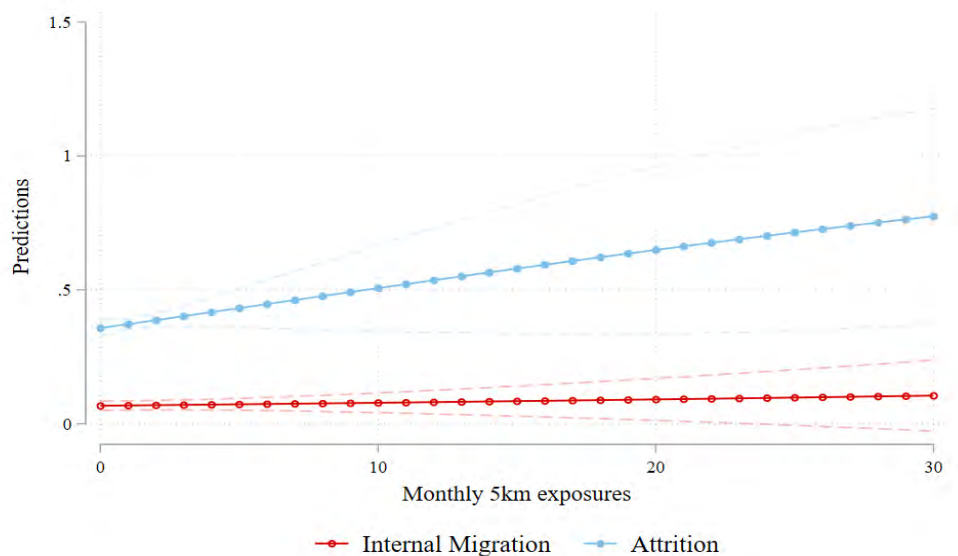


*Notes:* Plots show marginal changes in the predicted log odds of migration events over values of monthly 5km exposures to violence. The red dotted line corresponds to the lenient classification of external migration, and the blue to the restrictive version. Dashed lines show 95% confidence intervals. The mean value of monthly exposures at 5km is 1.74, and the maximum is 60.

To further explore the effects of violent exposures, I plot the marginal predicted log odds of migration over the support of the monthly violent exposures in Figure 8. Both sets of external migration classification criteria generate convex and increasing marginal prediction curves. Importantly, the average monthly rate of exposure in the sample is 1.74; most users are not experiencing violence at rates represented far to the right of these graphs. Still, the curves illustrate a distinctive rise in the propensity to migrate as individual exposures to violent events become more common. Figure 9 plots the marginal predicted log odds for internal migration and general attrition events over the same support. The comparative flatness of these curves reveals a much weaker or nonexistent association with the violent exposure variable, suggesting that violence is not driving internal migration or non-migration

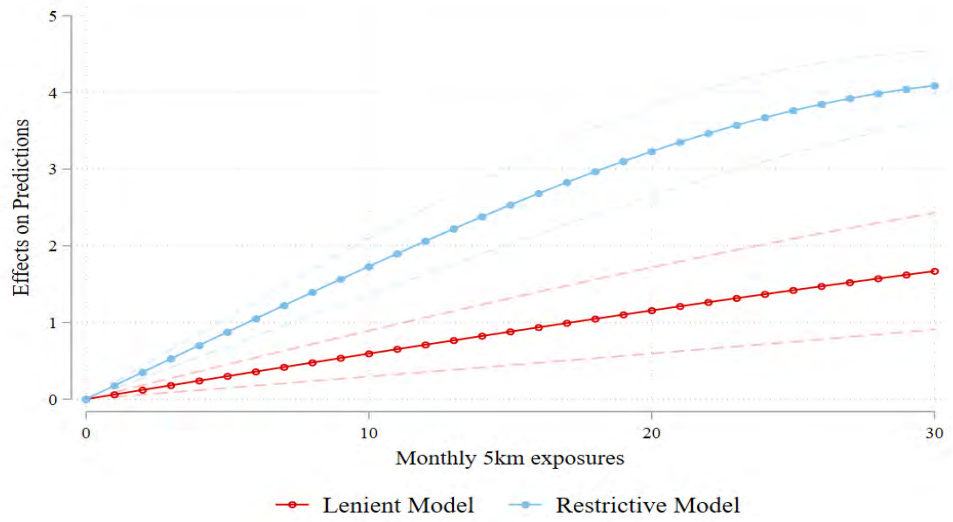
attrition from the sample. The same comparative pattern between external migration, internal migration, and general attrition, can be seen in plots of predicted elasticity of violence, Figures 10 and 11, respectively.

Figure 9: Marginal change in predicted log odds of internal migration and sample attrition



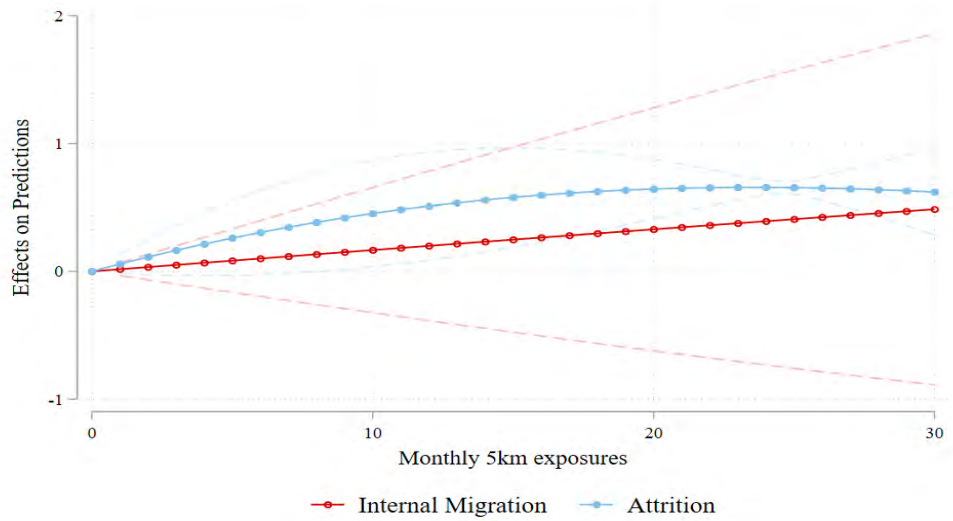
*Notes:* Plots show marginal changes in the predicted log odds of two event types over values of monthly 5km exposures to violence. The red dotted line corresponds to internal migrations, and the blue to sample attrition. Dashed lines show 95% confidence intervals. The mean value of monthly exposures at 5km is 1.74, and the maximum is 60.

Figure 10: Elasticity of external migration to violent exposures



*Notes:* Plots show marginal changes in the elasticity of migration events over values of monthly 5km exposures to violence. The red dotted line corresponds to the lenient classification of external migration, and the blue to the restrictive version. Dashed lines show 95% confidence intervals. The mean value of monthly exposures at 5km is 1.74, and the maximum is 60.

Figure 11: Elasticity of internal migration and sample attrition to violent exposures



*Notes:* Plots show marginal changes in the elasticity of two event types over values of monthly 5km exposures to violence. The red dotted line corresponds to internal migrations, and the blue to sample attrition. Dashed lines show 95% confidence intervals. The mean value of monthly exposures at 5km is 1.74, and the maximum is 60.

I also provide estimates from a conditional logit specification (McFadden, 1974) that relaxes the discrete choice IIA assumption by allowing a more complete set of location choice alternatives (stay, internal, external). Estimation of migration probability is given by the following expression,

$$Pr(M_{ij} = 1|z_{ij}) = \frac{\exp(z_{ij})}{\sum_{i=1}^3 z_{ij}} \quad (11)$$

where  $z_i$  is as in Equation 10 and  $j \in J = (1 = Stay, 2 = Internal, 3 = External)$ . Results are nearly identical, and are reported in Table 4, where  $j = 1$  is the reference alternative.

Table 4: Displacement Choice with 3 Alternatives

|                        | Lenient Model       |                     | Restrictive Model   |                     |
|------------------------|---------------------|---------------------|---------------------|---------------------|
|                        | (1)                 | (2)                 | (3)                 | (4)                 |
| Internal               |                     |                     |                     |                     |
| Monthly 5km exposures  | 0.009<br>(0.025)    |                     | 0.018<br>(0.028)    |                     |
| Monthly 20km exposures |                     | 0.004<br>(0.006)    |                     | 0.005<br>(0.006)    |
| Mean ROG (km)          | 0.017***<br>(0.002) | 0.017***<br>(0.002) | 0.017***<br>(0.002) | 0.017***<br>(0.002) |
| External               |                     |                     |                     |                     |
| Monthly 5km exposures  | 0.063***<br>(0.015) |                     | 0.179***<br>(0.021) |                     |
| Monthly 20km exposures |                     | 0.002<br>(0.003)    |                     | 0.026***<br>(0.005) |
| Mean ROG (km)          | 0.002<br>(0.004)    | 0.002<br>(0.003)    | 0.002<br>(0.005)    | 0.001<br>(0.006)    |
| Avg. monthly exp.      | 1.74                | 11.52               | 1.74                | 11.52               |
| Observations           | 37596               | 37596               | 37596               | 37596               |

*Notes:* The base alternative in this choice framework is to stay home. In columns 1 and 2, external migrations are inferred with the lenient parameters. Restrictive parameters are used for columns 3 and 4. Conditional logit regressions of displacement choice on violent exposures at 5km (columns 1 and 3) and 20km (columns 2 and 4). Regressions include controls for individual's radius of gyration and home province indicators. Errors are clustered at the individual level.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10.

Finally, Cox proportional-hazards models provide additional graphical evidence. Defining failure as either an external migration, internal migration, or attrition event, the survival plots illustrate predicted time to failure over the support of violent exposures. Comparing the graphs in Figure 12 to those in Figure 13, the same violent exposure rates that drive external migrations appear to have no effect on internal migration, and a negligible effect on sample attrition of any kind.

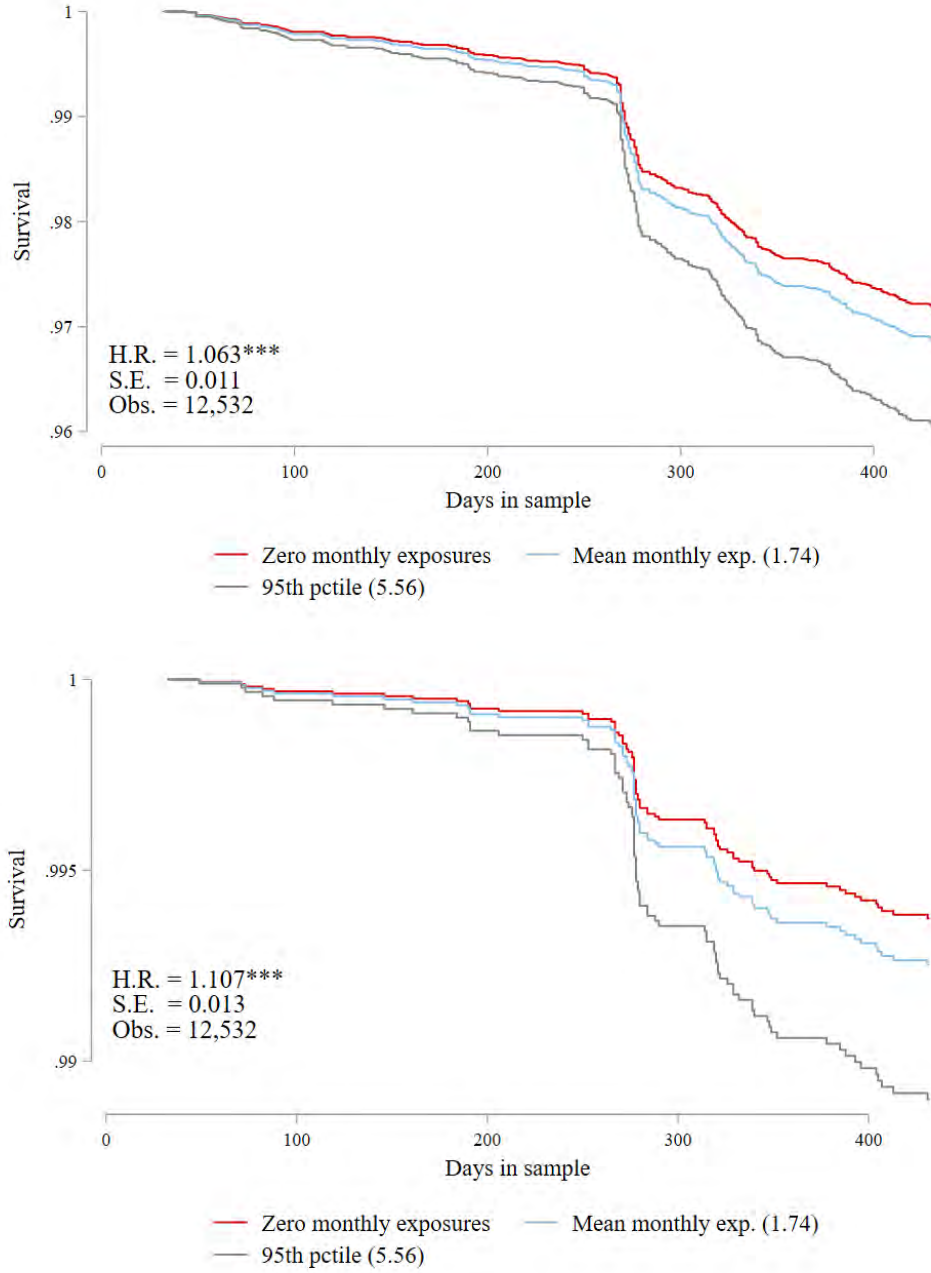
## 6.2 High-casualty events

Authors Tai et al. (2022), whose paper is set in 2013-2017 Afghanistan, find that internal migration is much more responsive to high-casualty events, or to violence claimed by the Islamic State.<sup>11</sup> To test for similar effects along the intensive margin, I use data from the Global Terrorism Database (GTD) to identify high-casualty and high-wounded events (START , National Consortium for the Study of Terrorism and Responses to Terrorism). The GTD is compiled from media reports of terrorist activities, and includes data on the number of deaths and wounded for each attack. Importantly, it also contains an exact date and coordinate pair, allowing the construction of a similar exposure variable when merged with the M-Paisa data in this project. I construct binary variables for 5km and 20km exposures to events from GTD with 10 or more deaths, another set for events with 10 or more wounded, and a third set that containing the union of the first two (10+ killed or 10+ wounded). These events are rare by comparison to the total count of insurgent violence in SIGACT used in the original model. For this reason, the restrictive classification of external migration is unusable in this regression, as there are too few classified migrants with exposures to high-casualty events. Table 5 shows the effect of this joint exposure on external migrants from the lenient classification and internal migrants. See Appendix B for estimates using the disjoint high-wounded and high-death count exposures.

---

<sup>11</sup>The Islamic State established notoriety in Afghanistan following the sample period of this paper. In 2022, the Taliban themselves labeled **I-S** a terrorist group.

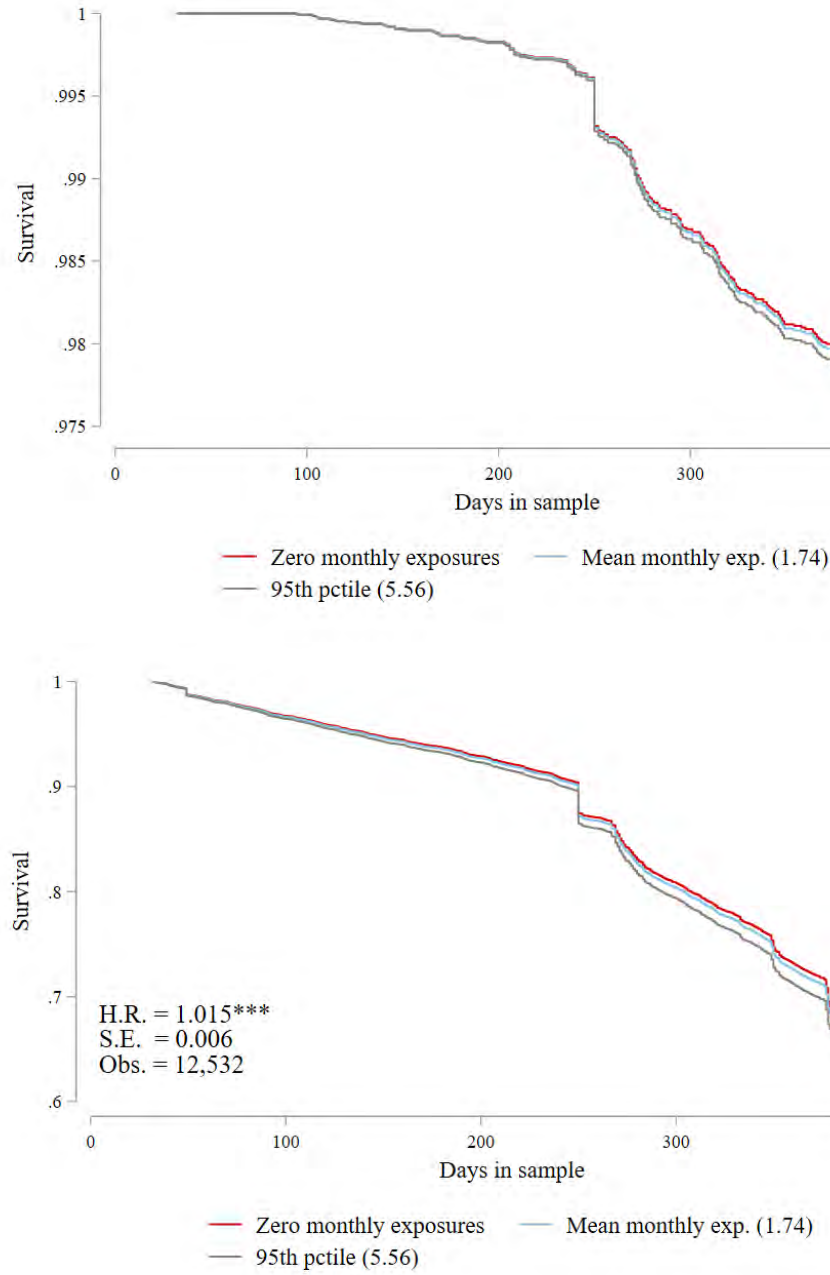
Figure 12: Cox Proportional-Hazards:  
(1) Lenient and (2) Restrictive criteria



*Notes:* Survival graphs from a Cox proportional hazards model. Hazard ratios, standard errors, and observations reported. Cross-sectional regression of individual users. Failure (defined as lenient (1) or restrictive (2) classifications of external migration events) is regressed on monthly exposures to violence. Red, blue, and grey lines represent predicted time until a failure event when monthly violence is at 0, its mean value, and its 95th percentile, respectively. Controls include individual's mean radius of gyration (ROG) and home province indicators. Errors are clustered at the individual level.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

Figure 13: Cox proportional-hazards:  
(1) Internal migration and (2) General attrition



*Notes:* Survival graphs from a Cox proportional hazards model. Hazard ratios, standard errors, and observations reported. Cross-sectional regression of individual users. Failure (defined as (1) internal migration events or (2) general attrition) regressed on monthly exposures to violence. Red, blue, and grey lines represent predicted time until a failure event when monthly violence is at 0, its mean value, and its 95th percentile, respectively. Controls include individual's mean radius of gyration (ROG) and home province indicators. Errors are clustered at the individual level.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

Table 5: Migration and High-Casualty Violence

|                                        | External (Lenient)  |                     | Internal            |                     |
|----------------------------------------|---------------------|---------------------|---------------------|---------------------|
|                                        | (1)                 | (2)                 | (3)                 | (4)                 |
| Monthly High-Casualty Exposures (5km)  | 1.907***<br>(0.440) |                     | 0.268<br>(0.225)    |                     |
| Monthly High-Casualty Exposures (20km) |                     | 4.730***<br>(1.128) |                     | 1.314<br>(0.671)    |
| Mean ROG (km)                          | 1.000<br>(0.002)    | 0.999<br>(0.002)    | 1.017***<br>(0.002) | 1.017***<br>(0.002) |
| Dependent var. mean                    | 0.0266              | 0.0266              | 0.0695              | 0.0695              |
| Average marginal effect                | 0.0149              | 0.0356              | -0.0768             | 0.0160              |
| $\Delta$ over baseline                 | 0.5610              | 1.3383              | -1.1051             | 0.2295              |
| Elasticity                             | 0.0161              | 0.0619              | -0.0318             | 0.0105              |
| Mean monthly exposure                  | 0.03                | 0.04                | 0.03                | 0.04                |
| Observations                           | 12532               | 12532               | 12532               | 12532               |
| Geographic controls                    | YES                 | YES                 | YES                 | YES                 |

*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria. The “restrictive” criteria does not classify enough migrants with varying exposure to high-casualty events for inclusion in this specification. Columns (3) and (4) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of exposures to violent events with at least 10 killed or at least 10 wounded, at either 5km or 20km.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

The marginal effects of 5km high-casualty exposures are greater by more than an order of magnitude than those from the original model’s full set of SIGACT events. The exponentiated coefficient reveals a 90.7% increase in the odds of external migration. Though these events are comparatively very rare, they may be driving migration responses in both models.<sup>12</sup> Interestingly, the estimated effect of exposures within 20km (a log odds increase of

<sup>12</sup>The SIGACT data almost certainly contain every event in GTD, but the merged version of the data

1.554) is much greater than that for exposures within 5km (0.645). One possible explanation relates to the salience of high-casualty events. If violence with high deaths or high numbers of wounded are equally salient within 20km, then the difference in effects between the 5km and 20km exposure variables is due to the greater combined total of events at 20km. The 20km exposure variable contains all 5km exposures too—they are not disjoint. Defining an exact relationship between distance and salience of violence is beyond the scope of this project. However, these results are suggestive of an intuitive pattern: that large events are felt in a larger radius. The effects to migration of smaller events dissipate much nearer to the event itself than those of events in which many are killed or wounded.

The coefficient estimates in 5 may not be directly comparable to those in 3, since the Global Terrorism Database does not contain every event from SIGACTS over the period of study. Another approach maintains all SIGACTS exposures as a control, and assumes that exposures to events in the Global Terrorism database are additive within the same model. Represented graphically in 14, I regress the migration outcome on binned measures of event intensity. Numeric results are reported in Appendix B, Table 13. The regression specification is identical to Eq. 10, but for the addition of indicator variables for the binned exposure counts, and includes all SIGACT exposures and individual-level controls:

$$Pr(M_i = 1|z_i) = \Lambda(z_i) = \Lambda(\alpha + \beta \times \textit{violence}_i + \lambda \times \mathbf{intensity}_{qi} + \theta \times \mathbf{X}_i) \quad (12)$$

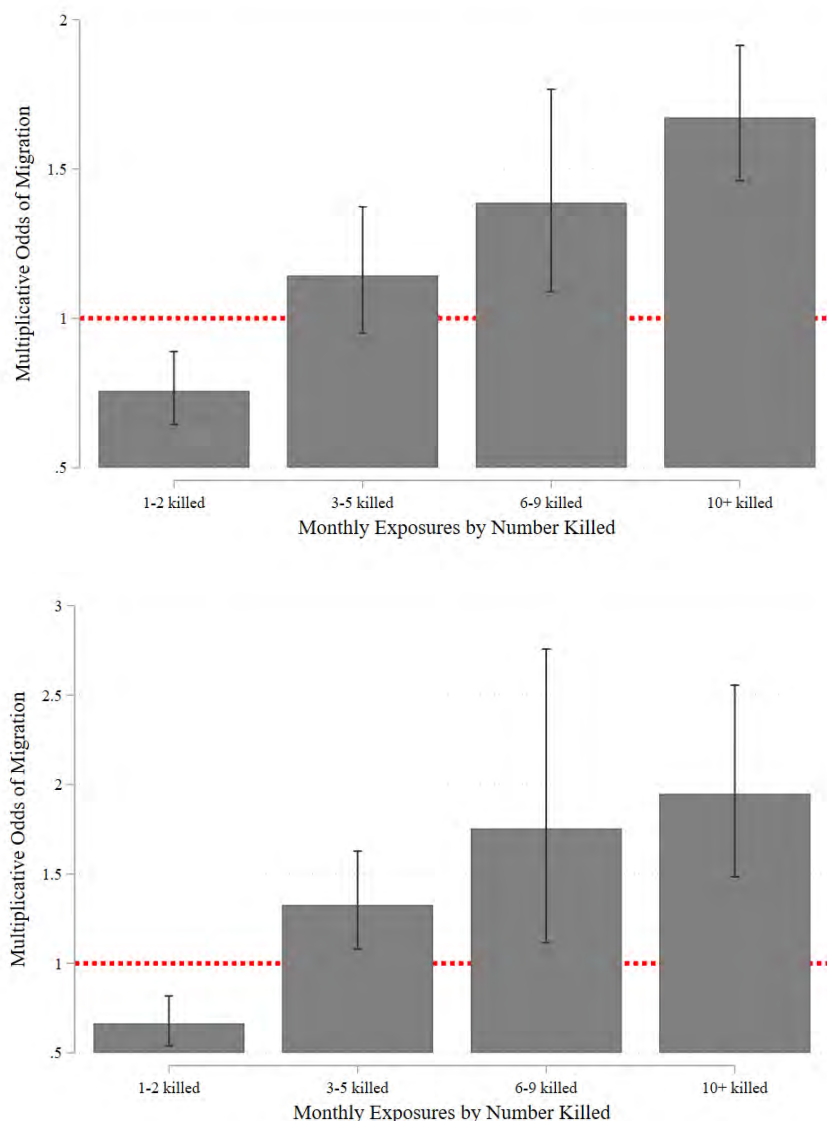
where *intensity* is a vector of monthly rates of exposure indexed by  $q \in Q$ , distinct bins of violent intensity measured by the number of people killed in an event. The increasing odds of migration over these bins supports the notion that high-casualty events are disproportionate drivers of external migration.

This project also allows for a test of the threshold theory of violence and migration, as outlined by Bohra-Mishra and Massey (2011). The authors document a curvilinear migration response to violence in Nepal during the insurgency from 1996-2006, indicating that low levels

---

that I work with does not list details of the events themselves, like the number of casualties.

Figure 14: Increasing Migration Response by Event Intensity  
 (1) 5km exposures and (2) 20km exposures



*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. I use the “lenient” migrant classification, as the “restrictive” criteria does not encode enough migrants with varying exposure to high-casualty events for inclusion in this specification. Regression controls include the SIGACT exposure from the main specification, as well as users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. The coefficients reported by the gray bars are the multiplicative change in odds of external migration. Each bar represents monthly exposure to violence, binned by intensity of exposure, against a baseline bin of zero killed. Intensity measures are from the Global Terrorism Database.

of violence are actually a deterrent to migration, and only once above a certain accumulated threshold does the sign of the effect turn positive. As shown in Figure 8, estimated response curves show no such relationship at low levels in this setting. Following the authors, I add a quadratic term for violent exposures to the original model, Equation 10. The results show no change in the sign of the exposure variable (a requirement of the curvilinear characterization) and no meaningful change in magnitude.<sup>13</sup> Full results are reported in Table 15 in Appendix B.

### 6.3 Asymmetric migration response

Another notable result of this paper is the asymmetric response estimated between internal and external migrations. One potential explanation relates to income. If we assume that mobility is correlated with income, then it may be acceptable to interpret each user’s radius of gyration (ROG) as an income or wealth proxy. There is a clear positive relationship between ROG and internal migration odds, reported in Table 3. It may be that wealthy individuals move to avoid violence as well, but only to comparatively safe internal destinations, judging themselves able to afford additional moves in the future, should they be necessary. In contrast, a less wealthy individual or family may be wary of future violence from the Taliban, perhaps with expectations that their influence would only spread within the country more over time. In this scenario, a cost-minimizer might predict that they are only able to afford a single migration attempt, and so try to ensure future safety from the insurgent threat by moving into a new country entirely. The results of Basu and Pearlman (2017) from the drug war in Mexico speak to a similar phenomenon. A related explanation of the difference in effects combines the influence of income with network size. Work like that of Milliff and Christia (2022) emphasizes network effects over violence as a migration push factor, though testing network impacts is outside the scope of this project.<sup>14</sup> Researchers with access to

---

<sup>13</sup>In Bohra-Mishra and Massey (2011), the authors defined the threshold relationship by a negative coefficient on the linear violence term, plus a positive coefficient on the quadratic term.

<sup>14</sup>Most other work using CDR data should be able to accomplish tests of this idea, but the merged dataset I use in this project lacks the variables necessary to construct networks or measures of user centrality.

more detailed individual data may be able to test these assertions by combining measurements from CDR data with surveys or other statistics.

## 7 Conclusion

This project estimates the impact of exposures to violence on individual migration decisions using call detail records from mobile banking customers in Afghanistan, 2011-2012. I develop a set of spatial and temporal classifying criteria to infer cross-border movement, a unique contribution generalizable to other empirical migration studies. I then estimate a model of location choice that suggests increasing odds of migrating over international borders with additional exposures to violence. When external migrants are classified with lenient criteria, an additional 5km-exposure to violence per month is associated with a 6.3% increase over baseline in migration odds. With a more restrictive classification of migrants, the effects appear much stronger: a 19.3% odds increase. The average marginal effects over the whole sample are small but persistent, about .001 percentage points for both classification schemes (external migration is a rare event).

When focusing solely on extreme violent events with high casualties, estimated effects are much larger, pointing to the importance of particularly shocking and salient violent events in driving migration episodes. Interestingly, the same levels of exposure appear to have no effect on rates of *internal* migration. Related results show that violence is also unrelated to general attrition from the data, helping to validate the classification criteria and proving support to the findings on external migration. An expansion of the discrete choice framework produces near-identical results, and survival plots from a Cox proportional-hazards model provides additional evidence of the effects. Future work should be dedicated to the validation of the migration event detection method and tests of network and income channels that may be driving the disparities in effects between internal and external migrants.

## References

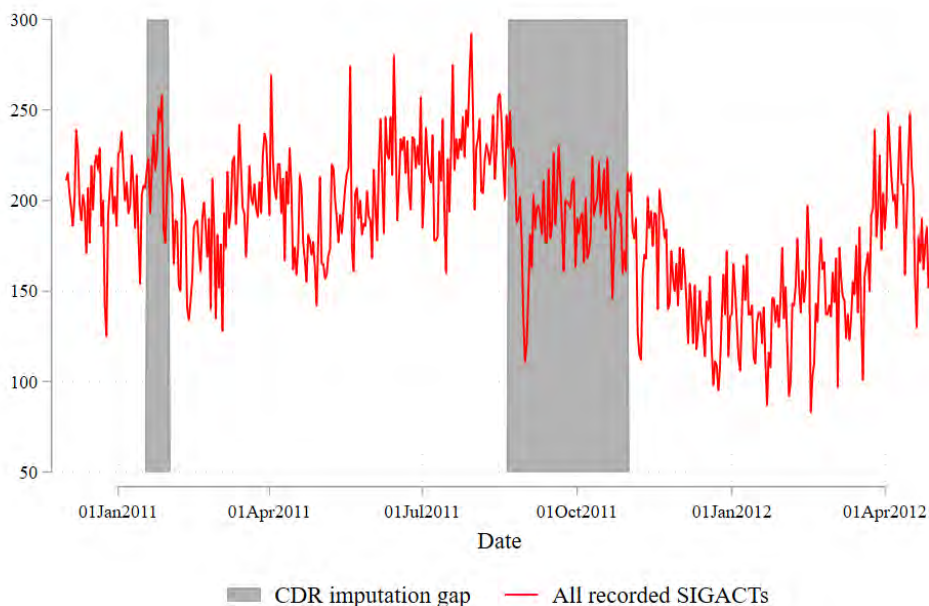
- Basu, Sukanya and Sarah Pearlman, “Violence and migration: evidence from Mexico’s drug war,” *IZA J Develop Migration*, December 2017, 7 (1), 18.
- Blumenstock, Joshua E., “Inferring patterns of internal migration from mobile phone call records: evidence from Rwanda,” *Information Technology for Development*, April 2012, 18 (2), 107–125. Publisher: Routledge .eprint: <https://doi.org/10.1080/02681102.2011.643209>.
- Blumenstock, Joshua E, Guanghua Chi, and Xu Tan, “Migration and the Value of Social Networks,” *The Review of Economic Studies*, December 2023, p. rdad113.
- Blumenstock, Joshua E., Michael Callen, Tarek Ghani, and Robert Gonzalez, “Violence and Financial Decisions: Evidence from Mobile Money in Afghanistan,” *The Review of Economics and Statistics*, March 2024, 106 (2), 352–369.
- Bohra-Mishra, Pratikshya and Douglas S. Massey, “Individual Decisions to Migrate During Civil Conflict,” *Demography*, May 2011, 48 (2), 401–424.
- Buggle, Johannes, Thierry Mayer, Seyhun Orcan Sakalli, and Mathias Thoenig, “The Refugee’s Dilemma: Evidence from Jewish Migration out of Nazi Germany\*,” *The Quarterly Journal of Economics*, May 2023, 138 (2), 1273–1345.
- Chi, Guanghua, Fengyang Lin, Guangqing Chi, and Joshua Blumenstock, “A general approach to detecting migration events in digital trace data,” *PLOS ONE*, October 2020, 15 (10), e0239408. Publisher: Public Library of Science.
- Chin, Aimee and Kalena E. Cortes, “Chapter 12 - The Refugee/Asylum Seeker,” in Barry R. Chiswick and Paul W. Miller, eds., *Handbook of the Economics of International Migration*, Vol. 1 of *Handbook of the Economics of International Migration*, North-Holland, January 2015, pp. 585–658.
- CIESIN - Columbia University, “Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals,” 2016.
- Condra, Luke N., 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.
- Grasse, Donald, Renard Sexton, and Austin Wright, “Courting Civilians During Conflict: Evidence from Taliban Judges in Afghanistan,” *International Organization*, January 2024, 78 (1), 134–169.
- Ibáñez, Ana María and Carlos Eduardo Vélez, “Civil conflict and forced migration: The micro determinants and welfare losses of displacement in Colombia,” *World Development*, 2008, 36 (4), 659–676. Publisher: Elsevier.
- Kakar, Mohammed, *Afghanistan : The Soviet Invasion and the Afghan Response, 1979-1982*, Vol. [Pbk. ed., 1997], Berkeley, Calif: University of California Press, 1997.
- Kirchhoff, Stefanie and Ana Maria Ibanez, eds, *DISPLACEMENT DUE TO VIOLENCE IN COLOMBIA: DETERMINANTS AND CONSEQUENCES AT THE HOUSEHOLD LEVEL* ZEF - Discussion Papers on Development Policy No. 41 2002.
- McFadden, Daniel, “The measurement of urban travel demand,” *Journal of Public Economics*, November 1974, 3 (4), 303–328.
- Milliff, Aidan and Fotini Christia, “A Big-Data Approach to the Determinants of Forced Migration,” *Working Paper*, 2022.

- Orozco-Aleman, Sandra and Heriberto Gonzalez-Lozano**, “Drug Violence and Migration Flows: Lessons from the Mexican Drug War,” *J. Human Resources*, 2018, 53 (3), 717–749.
- Picard, Robert**, “GEODIST: Stata module to compute geographical distances,” *Statistical Software Components*, June 2019. Publisher: Boston College Department of Economics.
- Sexton, Renard**, “Aid as a Tool against Insurgency: Evidence from Contested and Controlled Territory in Afghanistan,” *American Political Science Review*, November 2016, 110 (4), 731–749.
- Shaver, Andrew, Benjamin Krick, Judy Blancaflor, Xavier Liu, Ghassan Samara, Sarah Yein Ku, Shengkuo Hu, Joshua Angelo, Martha Carreon, Trishia Lim, Rachel Raps, Alyssa Velasquez, Sofia De Melo, and Zhanyi Zuo**, “The Causes and Consequences of Refugee Flows: A Contemporary Reanalysis,” *American Political Science Review*, May 2024, pp. 1–9.
- Stanford University**, “Mapping Militant Organizations: Afghan Taliban,” 2018.
- START (National Consortium for the Study of Terrorism and Responses to Terrorism)**, “Global Terrorism Database 1970-2020,” 2022.
- Tai, Xiao Hui, Shikhar Mehra, and Joshua E. Blumenstock**, “Mobile phone data reveal the effects of violence on internal displacement in Afghanistan,” *Nat Hum Behav*, May 2022, 6 (5), 624–634. Number: 5 Publisher: Nature Publishing Group.
- Trebbi, Francesco, Eric Weese, Austin L. Wright, and Andrew Shaver**, “Insurgent Learning,” *PIP*, August 2020, 1 (3), 417–448. Publisher: Now Publishers, Inc.
- United Nations High Commissioner for Refugees**, “UNHCR Data,” 2022.
- , “Global Trends: Forced Displacement in 2023,” Technical Report 2024.
- World Bank**, *The World Bank Group A to Z 2016*, The World Bank, October 2015.
- , “World Bank Group Strategy for Fragility, Conflict, and Violence 2020–2025,” 2020.

## A Imputed CDR

In Roshan’s call detail records, about 45% of user locations are imputed. According to the available description of the data, user-days with no phone use were imputed with the previous day’s location. Though this imputation may be innocuous, I cannot rule out that the imputations represent other processes, errors, or deliberate obfuscation. While there appears to be a random component to the imputation that is consistent with pattern as described (about 30% of observations on a given day), there are two time gaps of 12 and 71 days respectively, during which every location is imputed for every active user. Because imputed locations are *repeated* locations, all variation in mobility is lost in those gaps—as if every subscriber in the data stopped moving entirely. Dropping these observations fully disrupts the time variable, while keeping them introduces measurement error and compromises the internal validity of the study. I choose a specification that allows for this imbalance in the time variable.<sup>15</sup>

Figure 15: Imputed Records in Roshan Data



*Notes:* Grey blocks indicate spans for which all user data was imputed. The red line represents the number of all SIGACTs, uncategorized, over the sample period.

<sup>15</sup>Insurgents during this period were known to extort telecommunications companies in order to gain tactical advantage. Cell towers themselves were targeted and destroyed on more than one occasion

## B Additional Figures

Table 6: 20km violent exposures, by individual user

|                                            | Mean  | SD     | Med   | Max  |
|--------------------------------------------|-------|--------|-------|------|
| <i>Panel A: Violence (Ever exposed)</i>    |       |        |       |      |
| All violence                               | 0.98  | 0.15   | 1.00  | 1    |
| Direct fire                                | 0.96  | 0.19   | 1.00  | 1    |
| Indirect fire                              | 0.68  | 0.47   | 1.00  | 1    |
| Surface-air                                | 0.78  | 0.42   | 1.00  | 1    |
| IED                                        | 0.93  | 0.25   | 1.00  | 1    |
| <i>Panel B: Violence (Total exposures)</i> |       |        |       |      |
| All violence                               | 85.40 | 182.25 | 32.00 | 1978 |
| Direct fire                                | 48.79 | 136.72 | 15.00 | 1593 |
| Indirect fire                              | 8.33  | 18.10  | 3.00  | 191  |
| Surface-air                                | 6.81  | 8.26   | 4.00  | 125  |
| IED                                        | 21.47 | 39.24  | 9.00  | 303  |
| <i>Panel C: Violence (Per day)</i>         |       |        |       |      |
| All violence                               | 0.38  | 0.71   | 0.20  | 8    |
| Direct fire                                | 0.23  | 0.57   | 0.10  | 7    |
| Indirect fire                              | 0.03  | 0.07   | 0.02  | 1    |
| Surface-air                                | 0.03  | 0.03   | 0.03  | 0    |
| IED                                        | 0.09  | 0.12   | 0.05  | 1    |

*Notes:* This table is the 20km exposure analog to Table 1, which contains the 5km exposure statistics. Data are collapsed to the individual user-level.

Table 7: Violence and migration in recent papers

| Authors<br>year                                 | and    | Setting                   | Migration<br>type            | data | Result                                                                                            |
|-------------------------------------------------|--------|---------------------------|------------------------------|------|---------------------------------------------------------------------------------------------------|
| Milliff<br>Christia (2024)                      | and    | Yemen,<br>2010-2012       | Call detail records<br>(CDR) |      | Violence and other determinants mediated by network size and centrality                           |
| Daniele<br>(2023)                               | et al. | Mexico,<br>2006-2017      | Consular ID, Census          |      | Increases in poppy suitability led to internal and international out-migration                    |
| Tai et al. (2020)                               |        | Afghanistan,<br>2013-2017 | Call detail records<br>(CDR) |      | High casualty violence and violence by the Islamic State increased internal migration rates       |
| Orozco-Aleman<br>and Gonzalez-<br>Lozano (2018) |        | Mexico,<br>2007-2012      | Border surveys               |      | Drug violence had a net positive effect on migration to the United States                         |
| Basu and Pearlman (2017)                        |        | Mexico,<br>2006-2010      | Census                       |      | No effect of violence on internal migration, some evidence of small effects on external migration |

Table 8: Migration and high-casualty violence: 10+ wounded events

|                     | External (Lenient) |                     | Internal            |                     |
|---------------------|--------------------|---------------------|---------------------|---------------------|
|                     | (1)                | (2)                 | (3)                 | (4)                 |
| (max) w_tagpermon   | 0.996**<br>(0.394) |                     | -2.616*<br>(1.336)  |                     |
| (max) w_tag20permon |                    | 1.916***<br>(0.507) |                     | -0.842<br>(0.921)   |
| (max) meanrog       | -0.000<br>(0.002)  | -0.001<br>(0.002)   | 0.017***<br>(0.002) | 0.017***<br>(0.002) |
| dy/dx               | 0.0230             | 0.0441              | -0.1525             | -0.0491             |
| $\epsilon$          | 0.020              | 0.063               | -0.051              | -0.026              |
| Avg. monthly exp.   | 0.02               | 0.03                | 0.02                | 0.03                |
| Observations        | 12532              | 12532               | 12532               | 12532               |
| Migrants classified | 333                | 333                 | 871                 | 871                 |
| % classified        | 2.66               | 2.66                | 6.95                | 6.95                |

*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria. The “restrictive” criteria does not classify enough migrants with varying exposure to high-casualty events for inclusion in this specification. Columns (3) and (4) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of exposures to violent events with at least 10 wounded, at either 5km or 20km.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10.

Table 9: Migration and high-casualty violence: 10+ killed events

|                     | External (Lenient)  |                     | Internal            |                     |
|---------------------|---------------------|---------------------|---------------------|---------------------|
|                     | (1)                 | (2)                 | (3)                 | (4)                 |
| (max) k.tagpermon   | 1.087***<br>(0.294) |                     | 0.381<br>(1.296)    |                     |
| (max) k.tag20permon |                     | 3.523***<br>(0.781) |                     | 2.896**<br>(1.199)  |
| (max) meanrog       | -0.000<br>(0.002)   | -0.000<br>(0.002)   | 0.017***<br>(0.002) | 0.017***<br>(0.002) |
| dy/dx               | 0.0251              | 0.0805              | 0.0222              | 0.1688              |
| $\epsilon$          | 0.005               | 0.025               | 0.002               | 0.020               |
| Avg. monthly exp.   | 0.01                | 0.01                | 0.01                | 0.01                |
| Observations        | 12532               | 12532               | 12532               | 12532               |
| Migrants classified | 333                 | 333                 | 871                 | 871                 |
| % classified        | 2.66                | 2.66                | 6.95                | 6.95                |

*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria. The “restrictive” criteria does not classify enough migrants with varying exposure to high-casualty events for inclusion in this specification. Columns (3) and (4) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of exposures to violent events with at least 10 killed at either 5km or 20km.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10.

Table 10: Migration and violence (Poisson)

|                         | External (Lenient)  |                  | External (Restrictive) |                     | Internal            |                     |
|-------------------------|---------------------|------------------|------------------------|---------------------|---------------------|---------------------|
|                         | (1)                 | (2)              | (3)                    | (4)                 | (5)                 | (6)                 |
| Monthly 5km exposures   | 1.051***<br>(0.014) |                  | 1.160***<br>(0.016)    |                     | 1.005<br>(0.025)    |                     |
| Monthly 20km exposures  |                     | 1.002<br>(0.003) |                        | 1.024***<br>(0.005) |                     | 1.000<br>(0.004)    |
| Mean ROG (km)           | 1.000<br>(0.002)    | 1.000<br>(0.002) | 0.999<br>(0.004)       | 0.997<br>(0.003)    | 1.011***<br>(0.001) | 1.011***<br>(0.001) |
| Dependent var. mean     | 0.0266              | 0.0266           | 0.0066                 | 0.0066              | 0.0695              | 0.0695              |
| Average marginal effect | 0.0013              | 0.0000           | 0.0010                 | 0.0002              | 0.0004              | 0.0000              |
| $\Delta$ over baseline  | 0.0493              | 0.0015           | 0.1486                 | 0.0235              | 0.0050              | 0.0003              |
| Elasticity              | 0.0861              | 0.0173           | 0.2592                 | 0.2708              | 0.0088              | 0.0029              |
| Mean monthly exposure   | 1.74                | 11.52            | 1.74                   | 11.52               | 1.74                | 11.52               |
| Observations            | 12532               | 12532            | 12532                  | 12532               | 12532               | 12532               |
| Geographic controls     | YES                 | YES              | YES                    | YES                 | YES                 | YES                 |

*Notes:* This table is the Poisson-estimated analog of Table 3. Results from Poisson regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in likelihood. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria, while columns (3) and (4) use the “restrictive” version. Columns (5) and (6) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of violent exposures at either 5km or 20km.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10.

Table 11: Migration and violence (Probit)

|                         | External (Lenient)  |                  | External (Restrictive) |                     | Internal            |                     |
|-------------------------|---------------------|------------------|------------------------|---------------------|---------------------|---------------------|
|                         | (1)                 | (2)              | (3)                    | (4)                 | (5)                 | (6)                 |
| Monthly 5km exposures   | 1.035***<br>(0.009) |                  | 1.085***<br>(0.016)    |                     | 1.008<br>(0.014)    |                     |
| Monthly 20km exposures  |                     | 1.001<br>(0.002) |                        | 1.013***<br>(0.002) |                     | 1.002<br>(0.003)    |
| Mean ROG (km)           | 1.000<br>(0.001)    | 1.000<br>(0.001) | 1.000<br>(0.001)       | 1.000<br>(0.001)    | 1.010***<br>(0.001) | 1.010***<br>(0.001) |
| Dependent var. mean     | 0.0266              | 0.0266           | 0.0066                 | 0.0066              | 0.0695              | 0.0695              |
| Average marginal effect | 0.0016              | 0.0001           | 0.0012                 | 0.0002              | 0.0009              | 0.0002              |
| $\Delta$ over baseline  | 0.0610              | 0.0022           | 0.1775                 | 0.0282              | 0.0124              | 0.0028              |
| Elasticity              | 0.1663              | 0.0397           | 0.4767                 | 0.4813              | 0.0279              | 0.0410              |
| Mean monthly exposure   | 1.74                | 11.52            | 1.74                   | 11.52               | 1.74                | 11.52               |
| Observations            | 12532               | 12532            | 12532                  | 12532               | 12532               | 12532               |
| Geographic controls     | YES                 | YES              | YES                    | YES                 | YES                 | YES                 |

*Notes:* This table is the Probit-estimated analog of Table 3. Results from Probit regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in likelihood. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria, while columns (3) and (4) use the “restrictive” version. Columns (5) and (6) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of violent exposures at either 5km or 20km.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10.

Table 12: Migration and IED exposures

|                         | External (Lenient) |                    | External (Restrictive) |                  | Internal            |                     |
|-------------------------|--------------------|--------------------|------------------------|------------------|---------------------|---------------------|
|                         | (1)                | (2)                | (3)                    | (4)              | (5)                 | (6)                 |
| Monthly IED, 5km        | 1.076<br>(0.175)   |                    | 0.834<br>(0.191)       |                  | 0.929<br>(0.061)    |                     |
| Monthly IED, 20km       |                    | 1.093**<br>(0.048) |                        | 1.083<br>(0.064) |                     | 1.019<br>(0.027)    |
| Mean ROG (km)           | 1.000<br>(0.003)   | 0.999<br>(0.003)   | 0.997<br>(0.004)       | 0.998<br>(0.003) | 1.018***<br>(0.002) | 1.018***<br>(0.002) |
| Dependent var. mean     | 0.0266             | 0.0266             | 0.0066                 | 0.0066           | 0.0695              | 0.0695              |
| Average marginal effect | 0.0017             | 0.0020             | -0.0012                | 0.0005           | -0.0043             | 0.0011              |
| $\Delta$ over baseline  | 0.0637             | 0.0769             | -0.1742                | 0.0764           | -0.0615             | 0.0159              |
| Elasticity              | 0.0438             | 0.2269             | -0.1115                | 0.2086           | -0.0423             | 0.0464              |
| Mean monthly exposure   | 0.62               | 2.64               | 0.62                   | 2.64             | 0.62                | 2.64                |
| Observations            | 12532              | 12532              | 12532                  | 12532            | 12532               | 12532               |
| Geographic controls     | YES                | YES                | YES                    | YES              | YES                 | YES                 |

*Notes:* This table is an analog of Table 3 where only IED events are included. Results from Logit regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria, while columns (3) and (4) use the “restrictive” version. Columns (5) and (6) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of IED exposures at either 5km or 20km.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

Table 13: Migration and High-Casualty Violence

|                              | (1)                 | (2)                 |
|------------------------------|---------------------|---------------------|
| SIGACTs 5km                  | 1.061***<br>(0.014) |                     |
| 1-2 killed                   | 0.757***<br>(0.062) |                     |
| 3-5 killed                   | 1.143<br>(0.107)    |                     |
| 6-9 killed                   | 1.387***<br>(0.171) |                     |
| 10+ killed                   | 1.672***<br>(0.115) |                     |
| SIGACTs 20km                 |                     | 1.003<br>(0.003)    |
| 1-2 killed                   |                     | 0.664***<br>(0.071) |
| 3-5 killed                   |                     | 1.326***<br>(0.139) |
| 6-9 killed                   |                     | 1.754**<br>(0.405)  |
| 10+ killed                   |                     | 1.947***<br>(0.270) |
| (max) meanrog                | 1.000<br>(0.003)    | 1.000<br>(0.004)    |
| Dependent var. mean          | 0.0266              | 0.0266              |
| $\Delta$ over baseline (1-2) | -0.2355             | -0.3395             |
| $\Delta$ over baseline (3-5) | 0.1129              | 0.2336              |
| $\Delta$ over baseline (6-9) | 0.2764              | 0.4652              |
| $\Delta$ over baseline (10+) | 0.4343              | 0.5517              |
| Observations                 | 12532.00            | 12532.00            |
| Geographic controls          | YES                 | YES                 |

*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria. The “restrictive” criteria does not classify enough migrants with varying exposure to high-casualty events for inclusion in this specification. Columns (3) and (4) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variables of interest: users’ monthly rate of exposures to violent events in the Global Terrorism Database, binned by the number of killed.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .

Table 14: Do migrants experience more violence?

|                        | External Migrant (Lenient) |                 | p-value |
|------------------------|----------------------------|-----------------|---------|
|                        | NO                         | YES             |         |
| N                      | 12,199 (97.3%)             | 333 (2.7%)      |         |
| Monthly 5km exposures  | 1.724 (2.749)              | 2.502 (3.668)   | <0.001  |
| Monthly 20km exposures | 11.462 (21.377)            | 13.548 (17.457) | 0.078   |

|                        | External Migrant (Restrictive) |                 | p-value |
|------------------------|--------------------------------|-----------------|---------|
|                        | NO                             | YES             |         |
| N                      | 12,449 (99.3%)                 | 83 (0.7%)       |         |
| Monthly 5km exposures  | 1.724 (2.736)                  | 4.846 (5.883)   | <0.001  |
| Monthly 20km exposures | 11.400 (21.215)                | 29.135 (24.268) | <0.001  |

|                        | Internal Migrant |                 | p-value |
|------------------------|------------------|-----------------|---------|
|                        | NO               | YES             |         |
| N                      | 11,661 (93.0%)   | 871 (7.0%)      |         |
| Monthly 5km exposures  | 1.741 (2.812)    | 1.799 (2.307)   | 0.546   |
| Monthly 20km exposures | 11.447 (21.610)  | 12.452 (16.279) | 0.179   |

*Notes:* This table shows a summary of violent exposure rates by migration status. Under both the lenient and restrictive classification, those coded as external migrants have higher rates of exposure to the insurgency. The difference between internal migrants and non-migrants is not statistically different from zero.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.10.

Table 15: Test of Bohra-Mishra and Massey (2011)

|                        | External (Lenient) |          | External (Restrictive) |          | Internal |          |
|------------------------|--------------------|----------|------------------------|----------|----------|----------|
|                        | (1)                | (2)      | (3)                    | (4)      | (5)      | (6)      |
| Monthly 5km exposures  | 1.055              |          | 1.271***               |          | 1.101    |          |
|                        | (0.040)            |          | (0.024)                |          | (0.107)  |          |
| $(5kmExposure)^2$      | 1.000              |          | 0.997***               |          | 0.995    |          |
|                        | (0.001)            |          | (0.001)                |          | (0.005)  |          |
| Monthly 20km exposures |                    | 1.058*** |                        | 1.143*** |          | 1.057*** |
|                        |                    | (0.012)  |                        | (0.019)  |          | (0.022)  |
| $(20kmExposure)^2$     |                    | 0.999*** |                        | 0.999*** |          | 0.999*   |
|                        |                    | (0.000)  |                        | (0.000)  |          | (0.000)  |
| dy/dx                  | 0.0012             | 0.0013   | 0.0015                 | 0.0008   | 0.0056   | 0.0032   |
| $\epsilon$             | 0.089              | 0.633    | 0.410                  | 1.513    | 0.156    | 0.591    |
| Avg. monthly exp.      | 1.74               | 11.52    | 1.74                   | 11.52    | 1.74     | 11.52    |
| Observations           | 12532              | 12532    | 12532                  | 12532    | 12532    | 12532    |
| Migrants classified    | 333                | 333      | 83                     | 83       | 871      | 871      |
| % classified           | 2.66               | 2.66     | 0.66                   | 0.66     | 6.95     | 6.95     |

*Notes:* Results from logistic regression on a cross-section of all users. The outcome of interest is a dichotomous variable equal to 1 if a user is ever categorized as a migrant, zero (0) otherwise. Coefficients are reported as the multiplicative change in odds. Columns (1) and (2) are estimated using data from a “lenient” implementation of the migrant classification criteria, while columns (3) and (4) use the “restrictive” version. Columns (5) and (6) estimate internal migrations as defined in Blumenstock et al. (2023). Regression controls include users’ mean radius of gyration (ROG), total time in sample, share of days with phone usage, and modal home province indicators. Standard errors are clustered at the province level. Rows in the lower panel report average marginal effects, proportional change, elasticity, and mean monthly exposure, each with respect to the independent variable of interest: users’ monthly rate of violent exposures at either 5km or 20km. To be consistent with the results of Bohra-Mishra and Massey (2011) and the general threshold theory of violence and migration, the coefficient of the linear exposure variable would report a value less than one (in multiplicative odds), and the corresponding quadratic would be greater than one.

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ .