Harvest and Conflict: Measuring Agricultural Income's Effect on Displacement in Afghanistan

Alex Sobczynski

December 9, 2023

Abstract

This paper utilizes remote sensed MODIS satellite data and highly precise mobile device data to investigate the interplay of conflict and climate shocks in household decision making in Afghanistan. Focusing on the 2021 Taliban takeover, which saw rapid internal displacement, I delve into the factors influencing households' choices to leave their homes amidst violence and political change. A critical element is the reliance on agriculture, particularly the wheat harvest which occurs differentially across districts, as the primary source of income for smallholder farmers in rural areas. Displacing during harvest time, marked by significant financial loss, poses a unique challenge for these households, affecting both their economic stability and subsistence. Event studies indicate no response to the Taliban takeover in districts undergoing a wheat harvest, while samples taken over before and after the harvest period have significant outmovement in the weeks following the takeover. A fixed effects model shows households with higher estimated 2021 agricultural incomes displace further and at a higher rate. Conversely, households with higher production choose to displace less and closer to home, where production may proxy for historic income and land quality. A measure of historic crop performance and more rigorous price interpolation methods should be integrated into future work to support this hypothesis. The impact of expected income on displacement outcomes for those taken over before the harvest are null. My findings contribute valuable insights for policymakers, shedding light on the intricacies of decision-making in conflict and climate-affected regions and proposing considerations for future research into remote sensing methods and resource distribution strategies.

1 Introduction

Conflict and climate crises force households to make difficult decisions to leave their homes or face dangerous situations. Ensuring safety can come at the cost of sacrificing assets and careers. Measuring these decisions is particularly challenging in conflict settings due to ongoing violence. Further, because conflict occurs disproportionately in low income regions with low state capacity, there is a lack of infrastructure to measure households' economic and social decision making. In the case of Afghanistan, agriculture via subsistence farming and cash crops is the primary industry, especially in rural areas. The primary crop across Afghanistan is wheat. In 2023, Afghani farmers produced more than 19 times more wheat than any other crop USDA (2023). Smallholder farmers' assets are typically only reaped once or twice a year. For most farmers, the wheat harvest is their most valuable asset, thus displacing during harvest time can represent a significant financial loss. The next opportunity for income is typically months away. Given this mechanism, those that displace after harvest time should be wealthier in expectation than those who displace before. Because farmers depend on the wheat harvest for both income and subsistence, a lost harvest can also impede a household's health.

Income from the harvest varies across the country with price, land quality, and seasonal weather shocks. While an increase in income enables households to displace more optimally, it may also deter them from displacing if they have the means to buy favor with the Taliban and improve their standing within communities. Households with more assets may also delay or accelerate the displacement process. If a household is wealthy enough to displace to a safer yet further location, they may take more time to prepare whereas a lower income household would choose instead to displace for an anticipated short period of violence, then return after immediate threats have dissipated.

In 2021, the Taliban takeover occurred primarily over a period of five months from April to September with variation across districts. Though the Taliban has maintained a presence in Afghanistan for well over a decade and had de facto control over several of Afghanistan's

398 districts [1], the 2021 takeover occurred rapidly with relatively little push back from government or international forces. Over this period, the UNHCR estimates that 3.5 million people were displaced internally UNHCR (2022). Because of challenges in monitoring the rapidly developing process on the ground, however, there is no way to know exactly how precise estimated trends and magnitudes are. Though some households had more notice about the forthcoming takeover, it is unlikely that households were able to anticipate the takeover early enough to smooth their consumption and, most notably, this process likely would not have occurred deferentially across districts. The takeover and wheat harvest periods both occur from late Spring to early Fall and vary across districts. Additionally, 2021 was a unique year in that Afghanistan was experiencing a La Niña event that disproportionately affected the country and resulted in poor crop performance. While the cost of food products generally decreases after the harvest begins in the earliest harvesting districts, the interaction of the takeover and climate event resulted in a 14% increase from April to December 2021 (NSIA, 2022). This climate event translated into regionally varying shocks to wheat price. The three elements of takeover, harvest, and price provide an ideal environment to examine displacement outcomes with respect to historic and season specific agricultural income.

This paper leverages two highly granular data sets to identify individual movement and income patterns remotely. The first is spectroscopy data acquired using satellites. These data estimate how productive a plot is at 500 meter resolution. The second is mobile device data which gathers precise coordinates of devices to a second by aggregating locations from premium apps. Section 2 provides a review of the displacement literature as well as some publications outside of economics on remote sensing and GPS methods. Section 3 offers an overview of all data sets used in the analysis including how they've been processed from their raw form and considerations about their representativeness of the effects estimated in the analysis. Section 4 presents the results of the analysis including event studies of movement following the takeover for a sample split by stage of the harvest and an exercise estimating

[1] Several of Afghanistan's district borders are contested, but for synchronicity of existing data, I use the 398 district configuration.

income's effect on displacement outcomes. Section 5 discusses the results and provides a validation of the exogeneity of wheat harvest and takeover time. Section 6 concludes.

2 Literature Review

There is a well established literature examining the impact of climate crisis and conflict on internally displaced people and refugees' economic outcomes. A breadth of work examines the interaction of the resettlement process with integration into new labor markets. Kondylis (2010) looks at the welfare impacts and labor market outcomes of displaced households in the years following the 1992-95 conflict in Bosnia and Herzegovina. Delacretaz (2023) builds a model of refugee matching into sites of refuge and eventual labor markets considering a wide set of family structures. There is some literature about the empirical relationship between wealth and displaced populations. Arias et al. (2014) examine how asset accumulation is impeded by conflict among displaced households in Colombia. In a labor migration setting, Bijwaard and Wahba (2014) finds that movement into higher income countries is differential over income with the highest and lowest income groups most likely to return. From macroeconomic perspective, there is also a substantial literature on conflict traps and the interaction of poverty and conflict in a negative feedback loop over time (Collier, 2003; Hegre, 2017). I am not aware of any research that uses the harvest as a measure of wealth in a conflict context.

Mobile device data and call detail records are both utilized increasingly within migration research (Beine et al., 2021; Blumenstock, 2012; Chi et al., 2020; Ciacci et al., 2020; Luca et al., 2022; Sekara et al., 2019). Choi (2020) finds that the call detail records and mobile device data are very closely correlated, though mobile device data has a slightly larger spread. With particular relevance to the 2021 Taliban takeover, Tai et al. (2022) uses call detail records to measure responsiveness to violence over the period.

Remote sensing methods have been used in development for agricultural monitoring for decades. In a central European context, Panek and Gozdowski (2020) finds that the maximum Normalized Difference Vegetation Index (NDVI) for a season can explain approximately 80% of variation in yield using MODIS data. Similarly, Lopresti (2015) identifies a strong correlation in Argentina with $R^2 = .52$. Though the correlation between NDVI nearing the harvest (often proxied for the maximum NDVI) is strong across the board, it does vary by region and there is no existing research for Afghanistan. These studies require ground truth data for validation. Stoy PC (2022) uses higher resolution data from the Landsat satellite and flyover data to a 10 cm resolution to show that both yield and quality of wheat via Grain Protein Content can be estimated with spectroscopy data, most notably NDVI and the Enhanced Vegetation Index.

3 Data

3.1 Mobile Device Data

The central component of the analysis is mobile device data which precisely identifies approximately 600,000 unique devices over the January 2021 to December 2022 period of which approximately 91,000 have identifiable dwell locations and are observable after the takeover. These data were purchased from Veraset whose data are gathered by accumulating check-in time and locations from premium apps. Precise locations are scrambled by 30-100 meters from the initial location for security purposes, but provide otherwise highly accurate location by second. Because data is collected from apps that are rarely opened by some users, many users are only observed a few times over the two year period. Even though there is not consistent cell coverage over Afghanistan, because of the way app data is collected, the time of a ping is recorded at use, stored, and sent when the device next pings a cell phone tower. This ensures that the dataset is not biased towards locations with high cell reception. In order for a ping to be recorded, however, a device needs to be in proximity of a cell tower at some point during the sample period. To increase the likelihood that displaced devices are identified within a short time frame of their actual displacement, devices with a ping frequency of less than 10 are removed from the data. The cleaning process also includes collapsing these data to location-day by device to avoid over representing active users. Dwell locations are identified as the modal grid cell of each device in the January to April 2021 period. Distance from the dwell location is used to qualify displacement over the period. The sample size is further reduced by removing devices which are not observable in the pre-period.

One concern with mobile device is representativeness of the population. According to the World Bank, 57% of the Afghanistan population had a cell phone in 2021 World Bank (2021). Considering that approximately half of Afghanistan's population is under the age of 18, it is possible that the vast majority of heads of households own a cell phone and that their movements are representative of the households. It may also be the case that higher income people who are less constrained by agricultural income are more likely to be using premium apps and entering the sample. Because the distribution of pings across districts is approximately the same as the population across districts, it is likely that displaced devices' distribution across space is representative of population density. These maps are presented in Figure 1. An unpublished study funded by the UNHCR also finds that displacement outcomes measured by Veraset data are highly correlated with those identified by an on-the-ground rapid assessment.

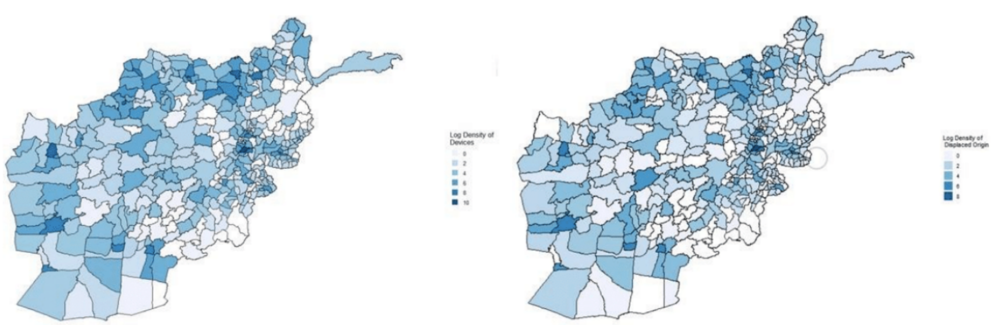


Figure 1: Log density of devices in the January-April 2021 (Left) and log density of eventually devices also in January-April 2021 (Right)

3.2 Conflict and Takeover Data

The FDD Long War Journal compiles takeover dates, categorizing them into three distinct phases: Government control, Contested status, and Taliban control. While de jure control remains with the government, the Taliban typically exercises influence over local services in areas marked as contested. In my analysis, I prioritize the official Taliban takeover date over the date when a district is classified as contested.

Violence data is sourced from the Armed Conflict Location Event Data Project (ACLED). These data are detailed accountings of violent events including number of fatalities, if civilians were targeted, and who the actors were. Longitude, latitude, and timestamp values are given for each event, thus fatalities can be pinpointed to their exact grid cell from the mobile device data. There were 426 documented events from January 1, 2021 through September 1, 2021.

3.3 Remote Sensed Wheat Production

To estimate production, remote sensed data from the MODIS satellite is acquired via Google Earth Engine to form a Normalized Difference Vegetation Index (NDVI) at a 500 meter resolution in 8 day intervals. NDVI measures greenness by finding the ratio of the difference to the sum of near infrared and red wavelengths detected via spectral sensors:

$$NDVI_i(x,y) = \frac{NIR_i(x,y) - Red_i(x,y)}{NIR_i(x,y) + Red_i(x,y)} \tag{1}$$

MODIS's NIR band at 500 meter resolution catches 841-876 nm wavelengths while the Red band catches 620-670 nm. The NDVI is a scale from -1 to 1 where values roughly above .2 indicate vegetation, with crop quality improving as NDVI approaches 1. In several empirical settings, NDVI has been proven good measure of wheat production. It should be noted that much of the literature relating NDVI to crop quality and production is done in temperate settings rather than arid climates like Afghanistan. More input use and higher soil quality in experimental plots result in max NDVI values for wheat over .7 while 90% of irrigated wheat plots in Afghanistan have NDVI values below .7. Rainfed plots have max NDVI values that are even lower. Rainfed plots also have longer growing periods and more variable harvest times. Because spectroscopy equipment may be subject to malfunction, the maximum NDVI values may induce measurement error. Thus, to collapse the distribution of NDVI over time into a single indicator of production, the 95th percentile of NDVI over the year is taken as a proxy for production. These data for 2021 are mapped in Figure 3. Comparing the NDVI and wheat price maps, northeastern regions have lower prices at harvest and higher NDVI and the reverse in the southwest, illustrating an effect of supply on price and supporting NDVI as a valid measure of production.

Because NDVI is only a measure of greenness and several crops with varying max NDVI values and harvest times are observed in the full sample, wheat area needs to be identified separately. Identifying crop type with remote sensed data alone is particularly challenging for a plethora of reasons. Phenological patterns for the same crop are different across regions and also vary with planting time. NDVI patterns also vary temporally because of seasonal differences. Especially due to the La Niña event in 2021, past years' wheat patterns cannot be extrapolated to 2021. Machine learning techniques, primarily random forests, are current best practice in the remote sensing literature, though they require ground truth. Because there are no existing wheat maps for the 2021 season in Afghanistan, I use a map developed in Tiwari (2020) which is also available on the MAIL website. They use qualitative data gathered by the Ministry of Agriculture, Irrigation, and Livestock (MAIL), spectroscopy data, and a random forest approach to identify where wheat was grown in the 2016-2017 season at 8 meter precision.

The primary concern with using wheat maps from 5 years before the sample period is that land use has changed over time. Though the wheat hectares in 2021 were not exactly the same as in 2016, NDVI can identify if a plot is fallow and, because the NDVI is taken at a 500 meter resolution, it is likely that slight land use changes are still captured. To align with the NDVI grid and account for computing restrictions, wheat maps had to be broken into chunks, reprojected into the World Geodetic System, converted to a 500 meter resolution qualified as a single cell overlap, and stitched back into a binary map. The wheat maps identify rainfed and irrigated wheat separately. All maps were eventually converted to a 2.5 km resolution to account for the grid cell size for the mobile device data. As a result, the scaling from 8 meter to 2.5 km identifies some plot as being rainfed and irrigated. As Figure 2 shows, these are primarily plots in rainfed areas that had a lone irrigated plot. The overlapping plots are dropped from the sample. Regions of high NDVI are also visible in the 2017 wheat map.

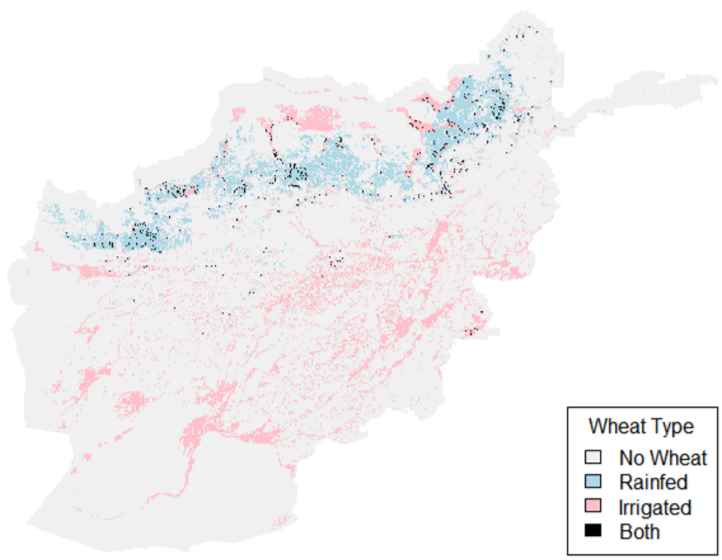


Figure 2: Rainfed and Irrigated Plots in 2017 at 2.5 km resolution

3.4 Wheat Pricing and Harvest

2020 and 2021 wheat flour price data comes from the National Statistical and Information Authority (NSIA) annual CPI reports. Farm gate wheat prices, while available for earlier years, are not available for 2020 or 2021, so the assumption is made that wheat processing costs do not vary substantially across space. Though these data were collected by the Afghanistan government during a tumultuous period politically, the NSIA has been generating CPI reports including monthly price data for dozens of products since 2007 and has continued into the present day. Three prices per product are recorded by enumerators each month at urban markets and averaged into a price estimate. These data are available monthly for 20 of 34 provincial capitals. Because most wheat is grown in rural areas, inverse weighting (IDW) with a β parameter of 2 and weights λ_i:

$$\lambda_i(x_0) = \frac{d^{-\beta}(x_0, x_i)}{\sum_{i=0}^{n} d^{-\beta}(x_0, x_i)} \tag{2}$$

generates estimates of price at a 500 meter resolution for each month across the country. The root mean square errors from IDW for each month are consistently comparable to a kriging approach, but the small sample size inhibits spatial variations in a kriged mask. IDW is thus favored. Optimal character recognition software was used to convert the NSIA data from .pdf to .csv format.

Afghanistan's agro-ecological zones vary from mountainous regions to arid deserts and growing practices vary dramatically across districts. Harvest time data combines two sources from Afghanistan's MAIL crop calendars that were generated in partnership with the Food and Agriculture Organization. One dataset is province level and covers the whole country while another includes 187 districts surveyed because of their high wheat production. 160 districts could be identified using a fuzzy match analysis, with the rest likely unidentifiable because of the merge of districts from 421 to 398. Districts that do not have crop calendar data from the district map default to the time for their province. Province level times are close to times recorded within their respective districts' further supporting the accuracy of these data. Because of seasonal changes across years, these ranges for harvest time are only approximations of the actual harvest time in the 2021 season as on-the-ground data is not available.

To identify expected income among those who have not yet harvested, expected price is taken to be the 2020 interpolated price for the month of the end of the harvest in each district. I assume that farmers will take the previous year's price to be predictive of the 2021 price. Though one may think COVID-19 resulted in wheat price shocks, price trends in 2020 are comparable to years prior. The interpolated price for each district's month of end of harvest is displayed in Figure 3.

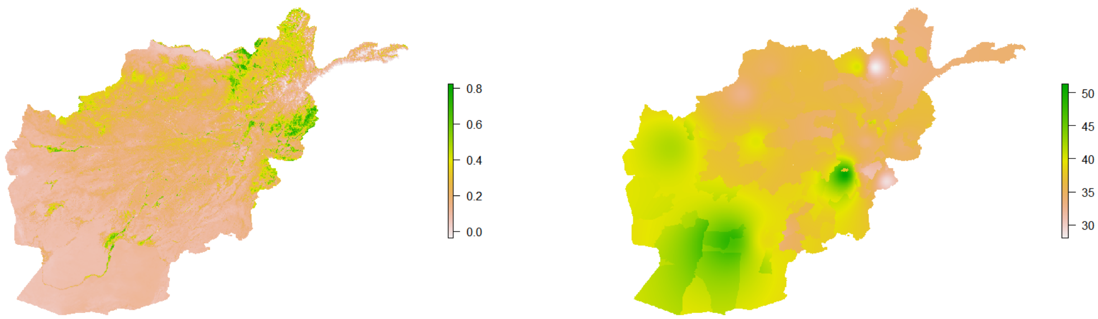


Figure 3: 2021 95th percentile Normalized Difference Vegetation Index at 500 m resolution (Left) and 2020 Interpolated Price of Wheat Flour During Month of Harvest (Right)

4 Results

4.1 Differential Movement Across Seasons

To first substantiate that the wheat harvest has an effect on propensity to displace, I use an event study framework to examine how distance from dwell location changes following the takeover in a split sample between those before, in, and after the harvest period. To maintain as much of my sample as possible, my unit is pings from all users where some users appear more than others in each sample. Pings into each of the samples is based on if the timestamp for each individual ping is before, in, or after the harvest. Note that the data was cleaned to remove duplicate pings from a user in a grid cell in a given day. Subsequent analyses limit the sample to those users whose dwell location is in a wheat plot to identify predicted income, but these models include all users. Wheat is the primary crop across most of the country and because the majority of the population is financially dependent on the agricultural industry, I consider these models to measure a wider community, rather than household level, effect. The model is as follows:

$$Distance_{it} = \gamma_d + \lambda_t + \sum_{t=-29}^{30} \beta_t D_t + X\Phi + \epsilon_{it} \tag{3}$$

where distance is the Euclidean distance from one's dwell location, γ_d are district fixed effects, λ_m are day fixed effects, D_t are indicators for days since the Taliban takeover, Φ is a set of user specific controls including distance to provincial capital, distance to Kabul, and number of fatalities in the grid cell prior to takeover.

Figure 4 displays the results of the before (n = 7,534), in (n = 6,793), and after (n = 89,966) samples' event studies respectively. Full size plots are in the appendix. The time window for these event studies is the 30 days before and after the district takeover. In the event studies for the before and after samples, the trend of the estimates is as expected for the full sample; the devices remain close to their dwell locations and distance grows over the course of the month following the takeover. The confidence intervals in the after harvest study are much tighter as a result of the larger sample size. The after harvest estimates are more modest, peaking at approximately 20 km from dwell location in the second week after takeover. In the pre-period of the before harvest event study, confidence intervals are smaller than after the takeover. This behavior is likely an artifact of a smaller relative sample size as time since takeover increases. The sample decreases from 261 to 21 from t = -29 to t = 30. It is also possible that reactions to the takeover are noisier among poorer households who haven't harvested yet.

Most notably, there is no substantial movement in the sample in the middle of the harvest and far more noise than in the before harvest sample. Despite a similar decrease in sample size, the estimates are noisy throughout the pre- and post- period. Movement in an ordinary year is expected during the harvest as smallholders bring their goods to market

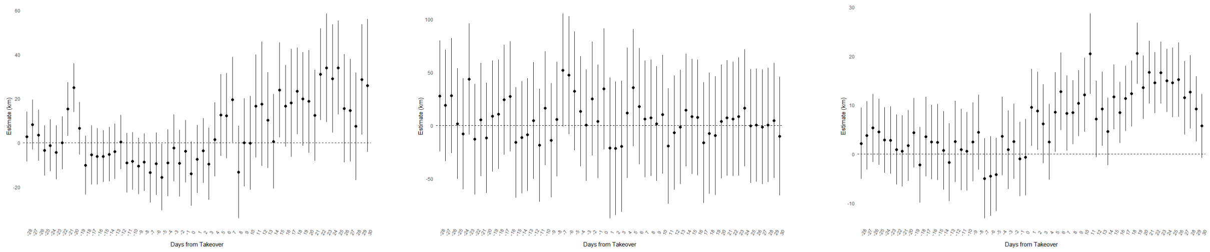


Figure 4: Before (Left; n = 7534), In (Center; n = 6793), and After (Right; n = 89966) Harvest estimates for event studies over a 60 day window with district-day fixed effects and 95% confidence intervals

and wage laborers migrate. However, there is no visible response at all to the Taliban takeover. Further, I may expect the joint impact of travel because of seasonal migration and out movement due to the takeover to result in positive, yet insignificant estimates in the post-period. One would also hypothesize the joint seasonal and conflict effect to result in more noise. In actuality, the confidence intervals are the same size in the post period if not slightly smaller. This is even more surprising because the number of pings per day of the study decreases from t = -29 to t = 30 as it does in the before harvest sample. Similarly sized confidence intervals over the period despite the sample decrease further support the conclusion that people choose not to displace in response to violence when there are assets at stake.

Despite the samples' inclusion of all devices regardless of their proximity to a wheat plot, there is an observable effect of the wheat harvest on responsiveness to the takeover. This supports that the harvest has a community effect outside of farmers themselves, even in conflict settings. While one may expect less device movement in the before harvest period than in the after harvest period because communities have fewer assets, this effect is not observable in these models. District and time fixed effects further decrease the likelihood of confounding variables. Poor power in the post takeover period for the before harvest sample impedes my ability to draw concrete conclusions on if the effect is due to the process of the harvest or the effect of having not yet completed the harvest.

4.2 Isolating the Farmer Effect

To isolate how agricultural seasonality affects smallholders, I restrict the sample to only those who live in an plot where wheat is grown. Because of decreased sample sizes for before, in, and after harvest samples to 1691, 948, and 15,855 respectively, leads and lags are aggregated from days to weeks. Predicted income given NDVI and price data is also included as a control in this sample. Results are in Figure 5. The before harvest wheat area trend is approximately the same for both significance and magnitude of estimates and size of noise. The suggests that the community and farmer effects are comparable for those districts taken over before the harvest. The in harvest wheat area sample has slightly more movement in the pre- period and less in the post- period, suggesting that movement in response to the takeover is even less for people living on wheat plots during harvest than those in districts undergoing harvest generally. Lastly, the sign and significance of estimates across the period for the after harvest wheat area sample are approximately the same, but magnitudes are lower. This may partly result from the aggregation of intervals from days to weeks or that wheat farmers move less than non farmers. Because of the inclusion of controls for distance from a city in the full sample, it is unlikely that the farmer effect is confounded by being in a rural location.

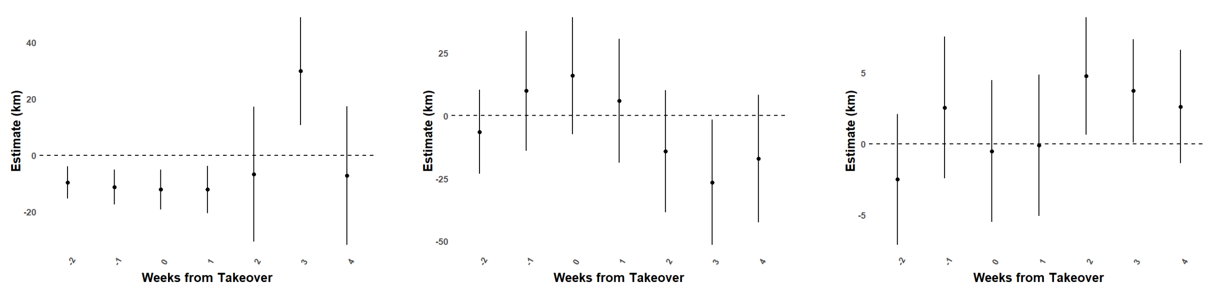


Figure 5: Before (Left; n = 1691), In (Center; n = 948), and After (Right; n = 15885) event studies for Wheat Area Only Sample

4.3 Effect of Agricultural Income on Displacement Outcomes

There are two ways that this paper measures displacement outcomes. First is the maximum distance displaced from one's dwell location after the takeover period. Second is a binary decision to displace where displacement is qualified as being more than 20 km from one's home following the Taliban takeover. To identify farmers, the sample is restricted to devices whose dwell location grew wheat in 2017. Wheat area is further restricted to have a 95th percentile NDVI value of at least .2, indicating that the land is still used for agricultural purposes. Increasing the NDVI threshold does not substantially effect the results. 35.00% of the 420 devices with a wheat area dwell location and a takeover time before the end of the harvest and 19.73% of the 2965 devices with a wheat area dwell location and a takeover time after the harvest were displaced at least 20 km after the takeover. It is possible that the displacement disparity is an artifact of the data – because ping rate goes down over time, there might be denser data for those who were taken over earlier. Because the difference in mean takeover time between the two samples is only 12.2 days, however, this is unlikely. Further, the difference in mean time of max distance from dwell location is 46.6 days suggesting that those taken over after the harvest are either travelling over a longer period or displacing later. Though propensity to displace differs for those who have and have not harvested, the distance displaced does not unless district fixed effects are included. These results of these OLS and LPM regressions are displayed in Table 1. Conversely the difference in propensity to displace disappears when district fixed effects are included. Though the event studies with district and time fixed effects indicate that movement patterns over time differ for those in and out of the harvest, the significance of the harvest indicator on time invariant displacement outcomes is unstable.

To examine the effects of agricultural income among those who have already harvested,

Table 1: Effect of Harvest Time of Displacement Outcomes

	Dependent variable:			
	Max Distance Displaced (km)		1 if Displaced > 20 km	
	(1)	(2)	(3)	(4)
1 if TO after Harvest	5.945	27.294*	-0.075^{***}	0.048
	(7.440)	(14.417)	(0.025)	(0.049)
NDVI	8.952	-54.174^{*}	0.133*	-0.251^{**}
	(23.067)	(29.707)	(0.079)	(0.101)
Fatalities	2.242	2.890	0.015*	0.008
	(2.427)	(3.284)	(0.008)	(0.011)
Dist to Closest City	-0.075	-3.026^{***}	0.002***	-0.007^{*}
	(0.178)	(1.169)	(0.001)	(0.004)
Ping Frequency	0.250***	0.251***	0.001***	0.001***
	(0.046)	(0.046)	(0.0002)	(0.0002)
Dist to Kabul	0.061***	4.176***	0.0002***	0.015***
	(0.009)	(1.117)	(0.00003)	(0.004)
Constant	25.736**		0.162***	
	(11.762)		(0.040)	
Observations	3,385	3,385	3,385	3,385
District Fixed Effects	No	Yes	No	Yes
R^2	0.019	0.016	0.041	0.024
Adjusted R^2	0.018	-0.011	0.039	-0.003
Residual Std. Error (df = 3378)	117.853		0.404	
F Statistic	11.051***	8.939***	23.942***	13.499***
	(df = 6; 3378)	(df = 6; 3293)	(df = 6; 3378)	(df = 6; 3293)

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

the following models

$$Displace_i = \beta Price_i \times NDVI_i + \gamma_d + X\Theta_i + \varepsilon_i$$

$$Max\ Distance_i = \beta Price_i \times NDVI_i + \gamma_d + X\Theta_i + \varepsilon_i$$

are estimated where γ are district fixed effects and Θ are a set of controls including NDVI, the number of fatalities in a grid cell in the period prior to takeover, distance to the closest city, the frequency of pings over the 2021 sample period, and the distance to Kabul. The distance variables control for how rural the setting is. Ping frequency is likely correlated with wealth, but because the information about the data collection process (apps data is taken from, types of devices these apps are compatible with, etc.) is not entirely known, it's challenging to form a confident hypothesis. Ping frequency does capture some individual effect uncorrelated with location. Because takeover time is district specific, district fixed effects control for variation in time of takeover. In Table 1, Columns 1-3 measure effects on maximum distance from dwell location after takeover and columns 4-6 model propensity to displace. Predicted asset value (NDVI x Price) is significant to a p value of at least .01 and positive for both outcome measures and across specifications. Columns 3 and 6 include district fixed effects. It is possible that the correlation between NDVI and production is heterogeneous across districts, but the similar estimates between the OLS and fixed effects model (85.118 and 75.491) suggest that this may not be the case. This interpretation relies on the assumption that the correlation between the interaction term and displacement outcomes is capturing an asset effect.

In addition to the positive effect of asset value on distance displaced and propensity to displace, the effect of NDVI is negative and very significant across specifications, suggesting that people with higher quality land and 2021 yields are less likely to move far or displace. It is possible that people with consistently high yielding plots have better community connections and prefer to displace closer to home or to not displace. They may also have more non-seasonal assets to pay the Taliban in exchange for safety without displacing.

Table 2: Outcomes for Those in Districts Taken Over After Harvest

	Dependent variable:					
	Max Distance Displaced Post-Takeover (km)			1 if Displaced > 20 km		
	(1)	(2)	(3)	(4)	(5)	(6)
NDVI x 2021 Price	1.550**	85.118***	75.491***	0.008***	0.156**	0.172**
	(0.760)	(19.380)	(24.680)	(0.002)	(0.063)	(0.080)
NDVI		−2,943.421***	−2,677.767***		−5.297**	−6.270**
		(668.334)	(852.828)		(2.187)	(2.770)
2021 Price		−25.614***	−24.273		−0.062***	−0.166
		(6.980)	(106.254)		(0.023)	(0.345)
Fatalities		19.676	33.650*		0.007	−0.048
		(12.874)	(17.919)		(0.042)	(0.058)
Dist to Closest City		0.205	−4.703***		0.004***	−0.010*
		(0.267)	(1.756)		(0.001)	(0.006)
Ping Frequency		0.208***	0.227***		0.001***	0.001***
		(0.049)	(0.049)		(0.0002)	(0.0002)
Dist to Kabul		0.066***	5.464***		0.0003***	0.017***
		(0.015)	(1.832)		(0.00005)	(0.006)
Constant	34.634***	917.595***		0.121***	2.213***	
	(7.843)	(239.878)		(0.026)	(0.785)	
Observations	2,965	2,965	2,965	2,965	2,965	2,965
District FE	No	No	Yes	No	No	Yes
R^2	0.001	0.028	0.020	0.003	0.036	0.027
Adjusted R^2	0.001	0.026	0.0005	0.003	0.034	0.008
Residual Std. Error	121.064	119.559		0.397	0.391	
	(df = 2963)	(df = 2957)		(df = 2963)	(df = 2957)	
F Statistic	4.159**	12.189***	8.347***	9.460***	15.681***	11.526***
	(df = 1; 2963)	(df = 7; 2957)	(df = 7; 2907)	(df = 1; 2963)	(df = 7; 2957)	(df = 7; 2907)

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

While the effects of agricultural income on displacement outcomes are significant for those taken over after the harvest, there is no specification that finds a significant effect of predicted income from 2020 or 2021 wheat flour prices on distance displaced or propensity to displace. It is possible this results from a small sample size and noisy data or it may be the case that expected income is not salient for those who have not harvested. Because of ping sparsity over the period, it is not possible to identify with accuracy when a device leaves its dwell location and if rates of movement differ between the samples. For the those that have already harvested, I can assume that displaced households will have income from the harvest. For those who have not yet harvested, however, I cannot know if they chose to wait to displace or were travelling to their destination over a long period. Further, because the sample of devices in areas taken over before the harvest that qualify as displaced before the end of the harvest is not large enough for sufficient statistical power, there is no way to check.

5 Discussion

5.1 Recent versus Accumulated Asset Effects

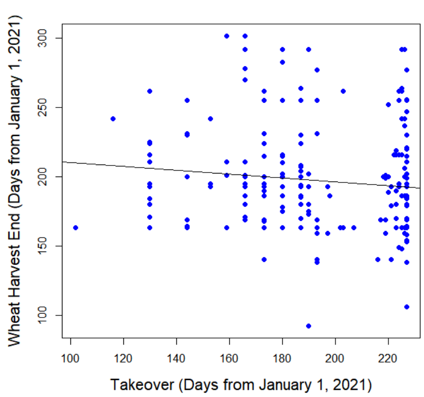
If NDVI is less time variant than asset value, then the NDVI estimate can be thought of as an effect for farmers with consistently good harvest and potentially more wealth from previous seasons while the interaction represents a year specific effect of income. An extension to this paper would validate this work by using historic max NDVI estimates as a benchmark for agricultural production and land quality. As the La Niña shock effected districts disproportionately, it would become possible to identify the effect of accumulated versus recent assets on migration behavior separately. The results are clear that asset value and production have an effect on displacement outcomes, but the mechanism with respect to seasonal shocks to production and existing assets cannot be determined from these findings. There would also be an opportunity to include a behavioral component to the analysis whereby the effect of a decrease in production for one season on displacement patterns is measured controlling for expected income. Though potentially observable in a conflict setting, these questions are perhaps more topical in the case of climate refugees whereby people are making decisions to leave their homes permanently.

5.2 Correlation between Harvest and Takeover Times

One concern with using the harvest as an approximation of assets is its potential endogeneity with respect to takeover time. One may imagine that Taliban forces preferred to wait until the harvest had ended to takeover a district because of welfare considerations or taxation opportunities. For the past decade, for example, opium has been taxed at a rate of 10% by the Taliban in a traditional Islamic tithe known as *Ushr* UNODC (2019). These taxes are less common for wheat, however, as the profit margins are far lower.

Conversely, local militias may be weaker during and before harvest because of exertion or poor nutrition during the starving season, suggesting takeover before the end of the harvest. A systematic correlation would suggest that there is some sorting occurring between those districts that are taken over before or after the wheat harvest. Because some of the districts were taken over in late 2020 and there was a subsequent seven month period of no takeovers, I've removed those 76 districts taken over in late October 2020 from the sample. Because the takeovers of these districts were less violent and attitudes more pro-Taliban in these districts, we can also think of them and their takeover process as systematically different. For the remaining 322 districts, the correlation is significant and negative where an additional day until takeover is correlated with harvest time .138 days earlier ($p = .0028$). The scatterplot in the left side of Figure 6 displays the variation in harvest and takeover time with a line of best fit. The mechanism of this correlation is not clear. Because most of the Afghan population is involved in the agricultural industry, it may be the case that the Taliban is waiting until local Taliban fighters have completed their seasonal work to transition into active fighting. If this were the mechanism inducing the negative correlation, we can think of the farmers in districts taken over before and after the takeover as the same in expectation, but this paper does not delve into this question. When the sample is restricted to the 87 districts with detected wheat, however, the correlation between takeover time and end of harvest becomes insignificant and the sign of the estimate becomes positive. This finding is plotted to the left in Figure 6. Though there are some endogeneity concerns for the full sample, the lack of correlation for the sample restricted to observable wheat production supports the findings in Sections 4.2 and 4.3.

The aversion to takeover districts during the harvest is also observable in the differential dropoff in sample size when restricting by sample size. The shares of the original sample that lives in an irrigated wheat area are 22% of the after harvest subset, 13% of the in harvest subset, and 17% of the before harvest subset, illustrating that the Taliban was more likely to take over a district during the wheat harvest if it were less dependent on wheat. Further, because the share of users who live in a wheat area and are taken over in the before harvest period is higher than in the during harvest period, it is likely that the Taliban is not simply cognizant of seasonality as it correlates with wealth from harvest, but also of the harvest process as a whole. Restricting the sample to districts where wheat plots are detected does not change the relative size of these shares: 53% of the after harvest subset, 29% of the in harvest subset and 49% of the before-harvest subset.



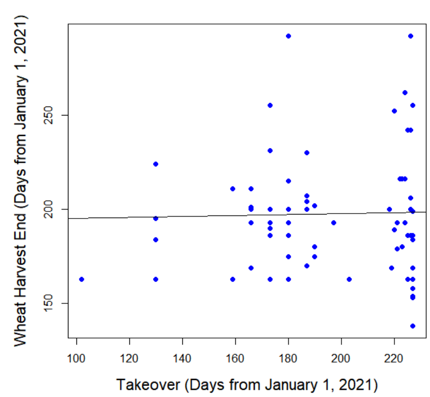


Figure 6: Correlation between Harvest and Takeover Time for districts taken over in 2021 (Left; n = 322) and Correlation between Harvest and Takeover Time for districts with detected wheat (Right; n = 87)

6 Conclusion

The initial aim of this paper was to construct an empirical model gauging the willingness to pay for safety by leveraging the forgone income from harvest. The ambiguity surrounding the null results raises questions about whether the absence of an effect reflects the reality or stems from a lack of statistical power. The challenge of pinpointing the exact displacement timing of devices further complicated the analysis.

In instances where perfectly measured data yields null results, several factors may be at play. Firstly, households taken over before the harvest may not perceive the harvest as time-sensitive, feeling comfortable leaving their homes for an extended period. As demonstrated by Tai et al. (2022), households tend to be more responsive to violence than a takeover in general. In this context, violence typically did not persist with the same intensity as in the initial weeks of the takeover period. Although households choose to displace, they may be doing so for a shorter duration, aligning with the observed shorter time after takeover of the maximum distance displaced. Unfortunately, the sparsity of data over time inhibits a rigorous analysis.

In this scenario, the expected value of the harvest may not significantly impact decision-making because households anticipate harvesting regardless of their displacement choice. This narrative aligns with the observed greater maximum displacement distance after harvest compared to the before-harvest sample, despite no difference in the propensity to displace when district fixed effects are included. While many displaced to evade Taliban rule, these findings align with UNHCR surveys and existing literature, indicating that households primarily displace temporarily in response to an immediate threat rather than a political or rule change.

Furthermore, beyond their relevance in conflict settings, these findings establish a foundation for measuring the impact of seasonal income on displacement patterns following climate events. Given the close link between income and climate events, leveraging remote sensing and mobile device data could enhance the understanding not only of the relationship between weather shocks and migration, as currently explored in the literature but also of the connections between season-to-season changes in production and income and short- versus long-term migration decisions. Employing remote sensing and near-real-time methods to measure income and movement could prove invaluable for practitioners in informing resource distribution, particularly in regions where on-the-ground information gathering is expensive, tedious, and often unreliable. However, further research is necessary to validate and understand the limitations of these measures.

References

Arias, M., Ibáñez, A., and Querubin, P. (2014). The Desire to Return during Civil War: Evidence for Internally Displaced Populations in Colombia. *Peace Economics, Peace Science and Public Policy*, 20(1):209–233.

Beine, M., Bertinelli, L., Cömertpay, R., Litina, A., and Maystadt, J. (2021). A gravity analysis of refugee mobility using mobile phone data. *Journal of Development Economics*, 150:102618.

Bijwaard, G. E. and Wahba, J. (2014). Do high-income or low-income immigrants leave faster? *Journal of Development Economics*, 108:54–68.

Blumenstock, J. E. (2012). Inferring patterns of internal migration from mobile phone call records: Evidence from Rwanda. *Information Technology for Development*, 18(2):107–125.

Chi, G., Lin, F., Chi, G., and Blumenstock, J. (2020). A general approach to detecting migration events in digital trace data. *PLOS ONE*, 15(10).

Choi, J. (2020). Comparison of CDR and GPS data for estimating the individual activity space.

Ciacci, R., García-Hombrados, J., and Zainudeen, A. (2020). Mobile phone network and migration: Evidence from Myanmar.

Collier, Paul; Elliott, V. L. H. H. H. A. R.-Q. M. S. N. (2003). *Breaking the Conflict Trap : Civil War and Development Policy.* World Bank and Oxford University Press.

Delacretaz, D., K. S. D. . T. A. (2023). Matching mechanisms for refugee resettlement. *American Economic Review*, 113(10):2689–2717.

Hegre, Håvard, N. H. M. R. R. F. (2017). Evaluating the scope and intensity of the conflict trap: A dynamic simulation approach. *Journal of Peace Research*, 54(2):243–261.

Kondylis, F. (2010). Conflict displacement and labor market outcomes in post-war bosnia and herzegovina. *Journal of Development Economics*, 93(2):235–248.

Lopresti, M.F., D. B. C. D. A. (2015). Relationship between modis-ndvi data and wheat yield: a case study in northern buenos aires province. *Argentina Information Processing in Agriculture*, 2(2):73–84.

Luca, M., Barlacchi, G., Oliver, N., and Lepri, B. (2022). Leveraging mobile phone data for migration flows. In *Data Science for Migration and Mobility*, pages 71–93.

NSIA (2022). Consumer price index yearbook 2021.

Panek, E. and Gozdowski, D. (2020). Analysis of relationship between cereal yield and ndvi for selected regions of central europe based on modis satellite data,. *Remote Sensing Applications: Society and Environment*, 17.

Sekara, V., Omodei, E., Healy, L., Beise, J., Hansen, C., You, D., Blume, S., and Garcia-Herranz, M. (2019). Mobile phone data for children on the move: Challenges and opportunities. In *Guide to Mobile Data Analytics in Refugee Scenarios*, pages 53–66.

Stoy PC, Khan AM, W. A. S. N. P. S. (2022). The spatial variability of ndvi within a wheat field: Information content and implications for yield and grain protein monitoring. *PLoS ONE*, 17(3).

Tai, X. H., Mehra, S., and Blumenstock, J. E. (2022). Mobile phone data reveal the effects of violence on internal displacement in Afghanistan. *Nature Human Behaviour*, 6(5):624–634.

Tiwari, Varun, M. M. A. Q. F. M. E. W. L. B. B. V. K. R. B. R. Y. W. (2020). Wheat area mapping in afghanistan based on optical and sar time-series images in google earth engine

cloud environment. *Frontiers in Environmental Science*, 8.
UNHCR (2022). Asia and the pacific regional trends - forced displacement and statelessness 2022.
UNODC (2019). Afghanistan opium survey 2019.
USDA (2023). Afghanistan production summary.
World Bank (2021). Mobile cellular subscriptions (per 100 people) - afghanistan.

Figures

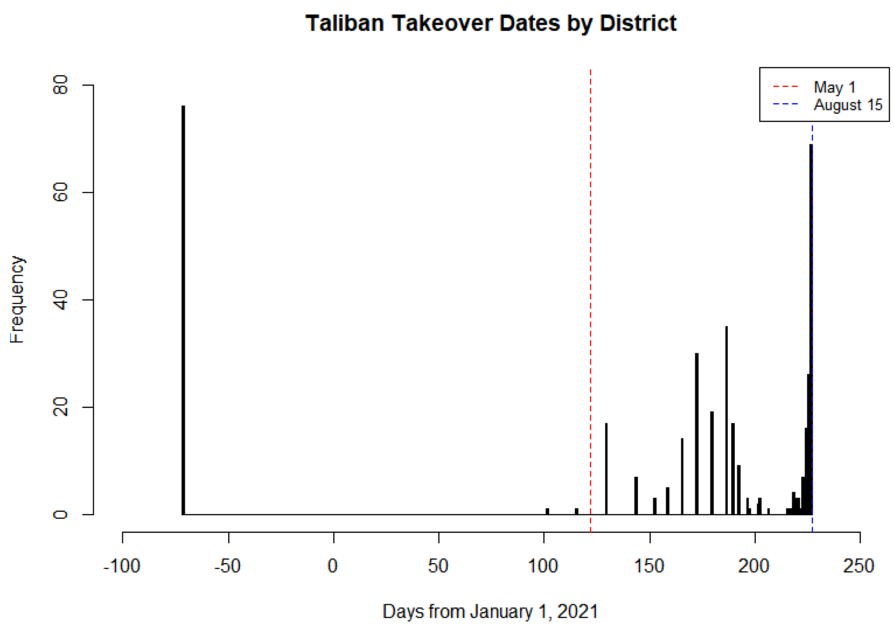


Figure 7: Histogram of Takeover Dates by Days from January 1, 2021

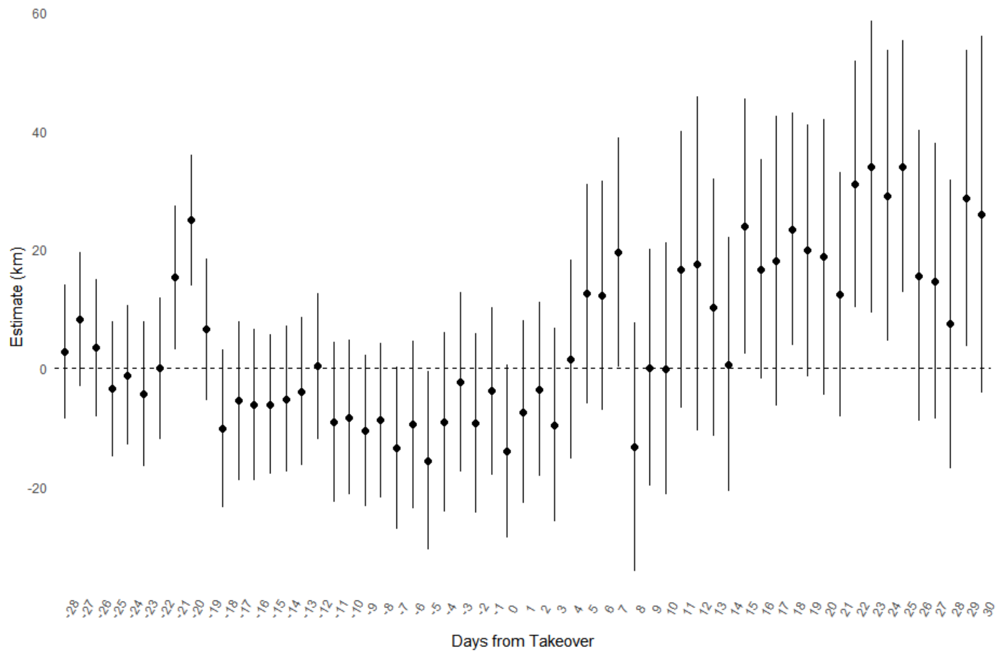


Figure 8: Event study from Taliban Takeover with all users before the Harvest; n = 7534, CI = 95%

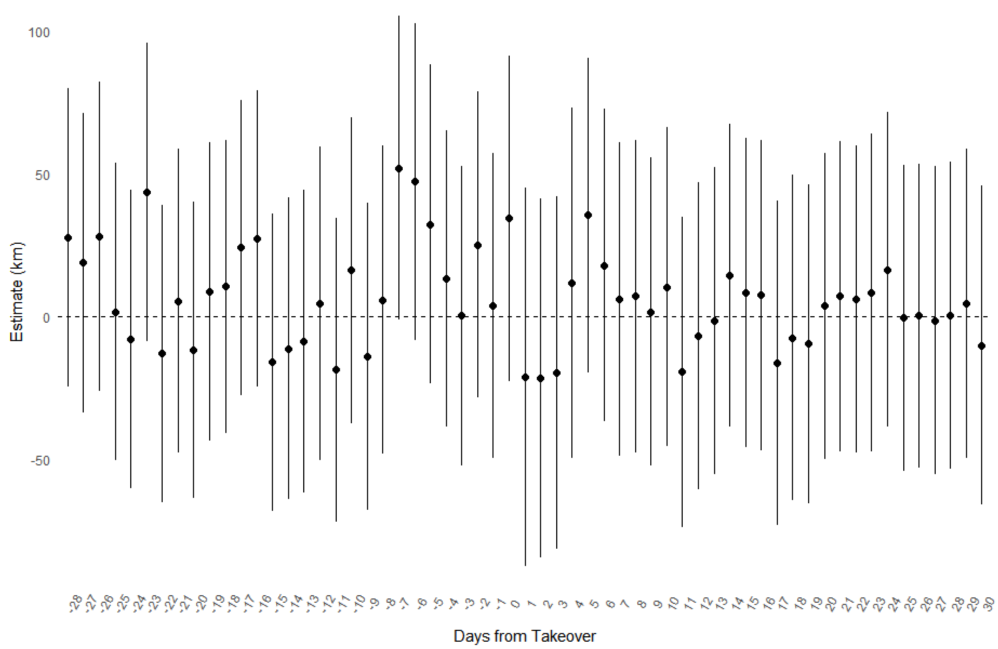


Figure 9: Event study from Taliban Takeover with all users during the Harvest Period; n = 6793, CI = 95%

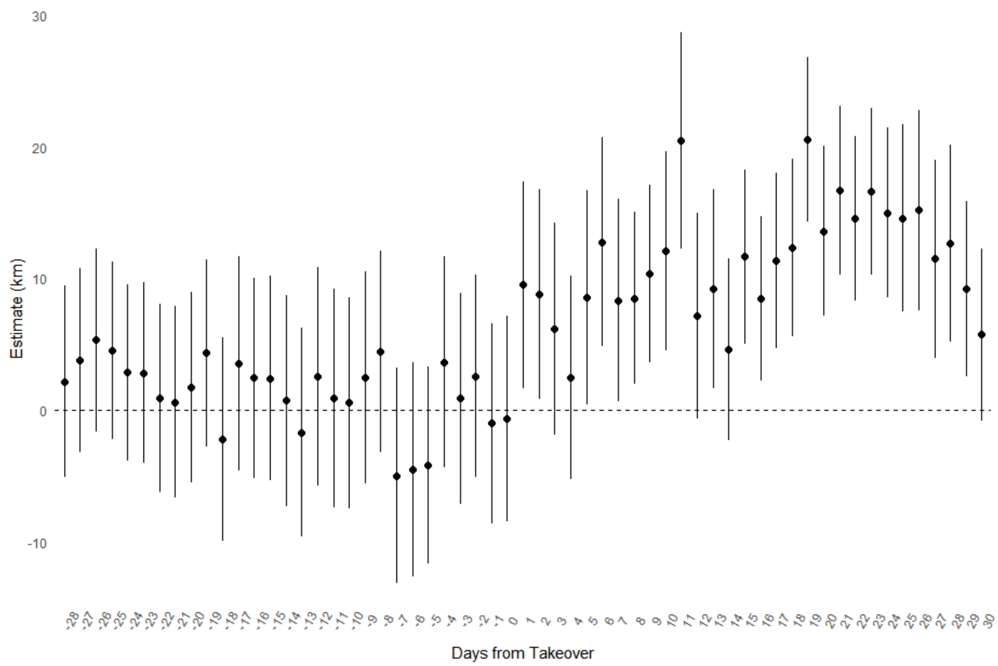


Figure 10: Event study from Taliban Takeover with all users after the Harvest; n = 89966, CI = 95%