Online Appendix

A Data Description

A.1 Original questions on individual attacks

Variable	Survey question	Code
Module: Respondent attack history (Up to 9 attack events)		
A_{jt}	*Was there any violent event in village j where you lived in year t?*	= 1 if resp. reported a violent event in village j in year t
$Perp_{jt}$	*Who was the perpetrator?*	= F/M/CA if resp. reported a violent event where perpetrator was a foreign armed group, Congolese militia, or Congolese national army
Mot_{jt}	*What was the attack motive?*	= p/s/c if resp. reported a violent event where the motive was pillage, sanction, or conquest
A_{ijt}	*Were you physically assaulted during the attack?*	= 1 if resp. reported a violent event where resp. was physically assaulted
$Theft_{f(i)jt}$	*Was any property of your household stolen during the attack?*	= 1 if resp. reported a violent event where any property of his household was stolen
$Chief_{jt}$	*Was the village chief assaulted during the attack?*	= 1 if resp. reported a violent event where the village chief was attacked

Variable	Survey question	Code
Module: Household information		
$A_{f(i)t}$	*For each of your household members, including yourself, list three episodes he/she was assaulted*	= 1 if any of the household members (excluding resp. himself) reported being assaulted in year t
$Viol_{f(i)t}$	*For each of your household members, including yourself, list three episodes he/she was sexually victimized*	= 1 if any of the household members (excluding resp. himself) reported being sexually victimized in year t

Subscript j indicates that information comes from respondent attack module where respondents are asked about violent events in contemporary villages. The information can vary across different respondents who live in the same village in the same year, but for concise notation we do not add additional individual subscript. Subscript i indicates the action imposed on respondent i. Subscript $f(i)$ indicates the action was imposed on respondent i's other household members, excluding respondent himself. In the next subsection, subscript o indicates the action was imposed on other households in the same village in year t.

A.2 Construction of main attack variables

Variable	Construction	Interpretation
$AA_{f(i)jt}$	$= A_{jt} \times (Mot_{jt} \neq c) \times A_{f(i)t}$	Whether resp. i reported an attack in year t in village j with nonconquest motive, and in which year any of the household members (excluding himself) reported being assaulted
I_{it}^{Victim}	$= \mathbb{1}\big(\exists t' < t, AA_{f(i)j(it')t'} = 1\big)$	Whether resp. i reported an attack on household before year t (j depends on i's living history)

The main definition of attack in this paper focuses on reported violent events with nonconquest motives on other household members, excluding respondent himself. We consider violent events with conquest motives mainly involving combatants during war and less about civilians, and thus conquest motives do not capture the mechanism where civilians participate out of intrinsic preferences. We do not include attack on respondent himself because it might affect participation through additional mechanisms—that is, attack on respondent himself potentially affects the respondent's capabilities, for instance, through handicap (see Section 6.2 and 8 for detailed discussion).

The main explanatory variable in specification 1, I_{it}^{Victim}, is constructed as an indicator whether respondent i reported any attack on household (excluding self) in the past. Subscript $j(it')$ stresses that reported attacks took place in villages where respondents lived in year t.

The fact that the main attack variable is constructed by combining information from different modules might complicate the interpretation in at least two scenarios in the following:

- Suppose a respondent reports two violent events in the same year, both with pillage motive. The first event was perpetrated by a foreign armed group, the second event was perpetrated by Congolese militia. The respondent also reports an attack on his spouse in the same year, and in reality his spouse was attacked in the second event. Our construction of main attack variable, however, would create a "false" attack on respondent's spouse by a foreign armed group. This scenario, however, is unlikely to happen in our data. In total, 1,415 respondents have reported 4,097 nonconquest violent events, and 77.4% are reported in the year when the respondent does not report any other nonconquest violent events.

- Other household members might live in a different locations than the respondent in year t. This is also unlikely because the majority of the households observed in the data are nuclear family households. Out of 1,534 households that have detailed rosters of current family members that live with the respondent (the total number of households is 1,537), 68% of the households do not include family members other than spouse and children. If the respondent reported that his spouse or children were attacked in year t, we assume that his spouse or children were living with the respondent in the same village as well.

A.3 Asset variables

Each respondent in South Kivu is asked to list yearly purchase and sales for farm animals (cows, goats, and pigs) and fields since 1990. For asset stock at birth, we ask how many cows, goats, pigs, and fields the respondent's father had when the respondent was born. We

also ask each respondent to report their asset stock at the survey year in farm animals but not fields.

We adopt the following approach to construct asset stock in cows, goats, and pigs. If respondent is not married, for farm animals and lands, we start from respondent's current asset stock and calculate respondent's asset stock in previous year by subtracting respondent's net purchase of asset this year from current asset stock. We calculate respondent's asset stock in each year backward up to year 1995. If respondent is married, we calculate respondent's asset stock backward up to the year when respondent was first married (89.9% of respondents who have hold marriages are only married once). Before the year respondent was first married, we start from respondent's asset stock at birth and calculate asset stock in following years by adding net purchase of asset up to the year before respondent was first married. The reason for this approach is that respondent is separated from his original household and starts a new household when he is married. For fields, since we did not ask each respondent to report their fields at the survey year, we calculate respondent's stock of fields starting from his stock of fields at birth and adding net purchase of fields in the years that follow. We further assume that when respondent is married, he gets one more piece of field in addition.

The construction of wealth variables above does not take into account the potential effect of attack on asset stocks. We use the following method to account for the loss of properties during a violent event. We first calculate the average loss in farm animals across all recorded violent events, and assume that each household would lose the average amount of farm animals if their household suffers from theft. Then, during the years when respondent reports a violent event with theft on the household, we decrease the total asset by the

assumed amount of loss of farm animals. We assume that violent events would not affect the stock of fields.

We then extract the principal component from the computed asset stock of cows, goats, pigs, and fields as the main wealth variable (see Table I, Table VI). The results are unchanged whether the calculation of the asset stocks account for loss of properties. For investment, we compute the principal component from the purchase of cows, goats, pigs, and fields. For the wealth of birth, we compute the principal component from the amount of cows, goats, pigs, and fields the respondent's father had at the respondent's birth, and the number of wives of his father and whether the respondent is a relative of the village chief.

B Additional background on FDLR and militia

B.1 Re-emergence of the Raia Mutomboki in 2011

On one side, a security vacuum was caused by the process of "regimentation" of the Congolese army (Stearns, 2013; Vogel, 2014). As in 2004–05, the FDLR took advantage of the security vacuum to take control in Shabunda. Under pressure from successive joint military operations by the Congolese and Rwandan military operations from 2009 to 2012 (in particular, the Umoja Wetu, Kimia II, and Amani Leo operations), the FDLR carried out violence against local populations, sparking the remobilization of the Raia Mutomboki. The 2011 movement, however, was considerably larger than the one in 2004–05, spreading to the northern areas of the territory of Shabunda in South Kivu, and eventually into the neighboring territories of Mwenga and Kalehe in South Kivu province, and parts of North Kivu province (Stearns, 2013; Hoffmann and Vlassenroot, 2014; Vogel, 2014).

B.2 Additional qualitative evidence on FDLR attacks

Consider the following account of a member of the Mera village:

> Arriving [in the FDLR area], they accused us of being collaborators of the enemy. We were immediately arrested and the decision was to kill us. We were beaten up, handcuffed, and they wanted to cut my head off. The machete was already on my neck and made an incision, when one of these Hutu had shouted to stop killing us because collaborating with the Tutsi was also their way of saving themselves from these people. We were released and returned to the village. When I arrived in

> the village, I was dominated by anger and the only revenge and I decided to join the armed group with my brother [redacted] who was in [redacted] . . . I then stayed at least six years in this armed group and my exit or my demobilization was determined by the pacification of the country . . . It is through this process that I integrated civilian life and abandoned the army.

B.3 Additional Details on the Origins of the FDLR

The armed group known as the Front de Liberation du Rwanda (FDLR) is an ethnic Hutu group. In July 1994, a rebel movement took power in Rwanda, ending the genocide that had been perpetrated by government supported Hutu dominated militias, the Interahamwe, and the government forces, against the Tutsi. In response to the change of power, two million Rwandans, mostly Hutus, fled into eastern DRC, specifically North Kivu. Among them were the Interahamwe, but also former Rwandan state bureaucrats and armed forces. They formed the Armée de Libération du Rwanda (AliR), predecessor of the FDLR.

In 1996, the Rwandan government launched a military campaign that started the First Congo War (1996–97). One of the goals was to eliminate the insurgent threat coming from the Kivus. While the Rwandan coalition succeeded in defeating Congolese government forces, installing a new president, and occupying large parts of the country, they failed to completely defeat Rwandan rebel activity in eastern DRC.

Conflicts between the new Congolese government and its Rwandan and Ugandan backers in 1998 plunged the DRC into the Second Congo War (1998–2004). During this war, Rwanda backed a rebel group, the Rassemblement Congolais pour la Democratie (RCD), that quickly

controlled the eastern half of the country, where it overtook the apparatus of the state and all urban areas. In the countryside, resistance militia had formed, which the RCD fought through counterinsurgency campaigns. The Congolese state had no formal control over the east in this period (Verweijen and Vlassenroot, 2015, Clark, 2002, Ngonzola-Ntalaja, 2002).

Instead, the Congolese government supported various armed groups and provided them with funds and ammunition to fight the RCD. Among them were the former Rwandan government forces and militia members, AliR, who in 2000 formed the FDLR. The FDLR is, in most areas of DRC, a foreign-armed group. By 2004, all major armed groups, except the FDLR, vacated the east in exchange for benefits precluded in a peace agreement (Sun City peace agreement). The Congolese state struggled to regain control over the eastern provinces, creating a security vacuum. The FDLR took advantage, and expanded their territory in North and South Kivu.

Between 2004 and 2008, the FDLR became notorious as one of the most violent groups among a host of armed actors in the eastern DRC. The Rwandan government continued to support armed groups who fought against the FDLR, while the Congolese state alternatively tolerated or actively supplied the FDLR.

C Appendix Tables

Table C.1: Main Regression with Participation in Congolese Army and Past Attack by Congolese Army

Past attack on HH	Recruiter groups and identities					
	Congolese army			Congolese militia	Foreign armed group	Any armed group
	(1)	(2)	(3)	(4)	(5)	(6)
By foreign armed group	-0.07					
	(0.07)					
By Congolese militia		-0.01				
		(0.15)				
By Congolese army			-0.14	6.16	0.40***	6.49
			(0.09)	(6.11)	(0.15)	(6.10)
Control mean	0.19	0.20	0.18	2.25	0.31	3.11
Obs.	25,060	25,060	25,060	25,060	25,060	25,060

Notes: This table presents OLS estimates of equation 1, focused on Congolese army both as a recruiter and perpetrator. The dependent variable in columns (1)–(3) is an indicator for whether the respondent joins a Congolese army. The explanatory variable in each line is an indicator for whether the respondent's household has been attacked by foreign armed groups, Congolese militia, and Congolese army, respectively. The dependent variables in columns (4)–(6) are indicators for whether the respondent joins a Congolese militia, foreign armed groups, or any armed group, respectively. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced an attack on household by the corresponding armed group before year t. P-value: *** 0.01, ** 0.05, * 0.10.

Table C.2: Balance between Victims vs. Nonvictims

	HH never attacked	HH attacked by foreign group		
		All	Intention Pillage	Intention Not pillage
	(1)	(2)	(3)	(4)
Obs. (All ages, $1995 \leq t \leq 2013$)	26,786	333	296	37
Panel A. Socio-demographic background per individual-year obs				
Age in year t	28.11	34.77***	35.01***	32.86***
Married in year t (%)	37.28	41.44**	42.23*	35.14
Works primarily in mining in year $t-1$ (%)	11.31	14.44	14.84	12.50
Works primarily in agriculture in year $t-1$ (%)	37.94	57.22***	56.13***	62.50***
Works primarily as civil servant in year $t-1$ (%)	5.49	6.95	7.10	6.25
Works primarily in school in year $t-1$ (%)	16.99	8.56***	9.68*	3.13***
Unemployed in year $t-1$ (%)	28.27	12.83***	12.26***	15.63***
Panel B. Economic status per individual-year obs				
Wealth at birth (z-score, only $t = 2012$)	-0.04	0.18**	0.16**	0.25**
Asset stock in year $t-1$ (z-score)	-0.03	0.05**	0.04**	0.09
Investment in year $t+1$ (z-score)	0.02	0.42***	0.44***	0.28
Panel C. Participation history per individual-year obs				
Ever participated in any armed group before year t (%)	6.16	11.11	10.81	13.51
Ever participated in Congolese militia in year t (%)	3.98	6.91	6.76	8.11
Ever participated in foreign armed group in year t (%)	1.67	3.60	3.38	5.41

Notes: This table shows descriptive statistics for observations where respondents' household members are attacked by foreign armed groups versus not in year t. Column (1) shows the mean characteristics of observations in which the household is not attacked by foreign armed groups and column (2) shows those for individual-year observations in which the household is attacked by foreign armed groups. Stars indicate the levels of statistical significance for the difference between the means in column (1) and (2) including village and year fixed effects to isolate within village targeting, with standard errors clustered two-way at the individual respondent and the village*year levels. We examine balance on socio-demographic characteristics in panel A. Attacked individuals tend to be 7 years older at the time of the attack than nonattacked observations, 4 pp. more likely to be already married, 19 pp. more likely to work in agriculture and 15 pp. less likely to be unemployed. This suggests that attackers target more economically active and established households within village. To explore this possibility, Panel B shows balance on economic outcomes. We find that the wealth at birth indicator for attacked households in attacked years is 22% of sd. larger than nonattacked observations. Our imputed asset stock index is somewhat larger for attacked households, and investment in the following year is also significantly larger. Taken together, this suggests that richer households are more likely to be targeted. The last panel shows the means of past participation history of the respondents. We find that attacks do not target households with more history of participation in any armed group. This rules out that attackers target potential participants.

Table C.3: Examining Potential Confounders—ACLED

	Attack radius									
	5 km	10 km	15 km	20 km	25 km	30 km	35 km	40 km	45 km	50 km
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A. Dependent variable: past attack on HH by foreign armed group (%), OLS										
ACLED attack by foreign armed group	2.52	0.20	0.91	0.37	0.20	0.44	1.41*	0.46	0.43	0.04
	(1.75)	(1.12)	(0.99)	(0.86)	(0.84)	(0.78)	(0.73)	(0.72)	(0.73)	(0.73)
Control mean	10.42	10.22	9.71	9.46	9.24	8.60	7.28	7.40	7.22	6.93
Obs.	16,503	16,527	16,604	16,678	16,728	16,829	16,872	16,886	16,928	16,947
Panel B. Dependent variable: participation in Congolese militia (%), OLS										
ACLED attack by foreign armed group	1.58	-1.04	-0.30	-1.14	-0.97	0.71	1.46*	1.35	1.44*	2.11***
	(1.58)	(0.94)	(0.88)	(0.89)	(0.86)	(0.82)	(0.86)	(0.84)	(0.86)	(0.80)
Control mean	3.21	3.41	3.60	3.94	4.04	3.34	3.31	3.41	3.31	2.79
Obs.	16,503	16,527	16,604	16,678	16,728	16,829	16,872	16,886	16,928	16,947

Notes: This table presents additional robustness check of the effect of past attack on participation, using public data from the Armed Conflict Location Event Data Project (ACLED). Panel A regresses the indicator for whether the respondent's household has been attacked by foreign armed groups on whether an attack by foreign armed group was reported (within a radius of X km around village j) in ACLED before year t. Panel B regresses participation in a Congolese militia on whether an attack by foreign armed group was reported in ACLED before year t. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced a corresponding type of attack on household before year t. P-value: *** 0.01, ** 0.05, * 0.10.

Table C.4: Examining Alternative Causal Channels—Additional Table

	Investment (z-score ×100) (1)	Overreporting any violent event (%) (2)
Past attack on HH by foreign armed group	-8.82	-3.73***
	(7.17)	(1.17)
Control mean	6.82	3.89
Obs.	14,874	25,060

Notes: This table presents OLS estimates of the effect of past attack on alternative mechanisms. The explanatory variable is an indicator for whether the respondent's household has been attacked by foreign armed groups. We include observations between 1995 and 2013 above age 15 at year t. Dependent variable in Column (1) is current investment z-score. Dependent variable in Column (2) is whether respondent overreports any violent event on village in year t. Overreporting is defined as whether respondent reports a violent event on village in year t, but less than half of the other contemporary villagers observed in the sample report so and there is no corresponding record in village chief survey. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced a corresponding type of attack on household before year t. P-value: *** 0.01, ** 0.05, * 0.10.

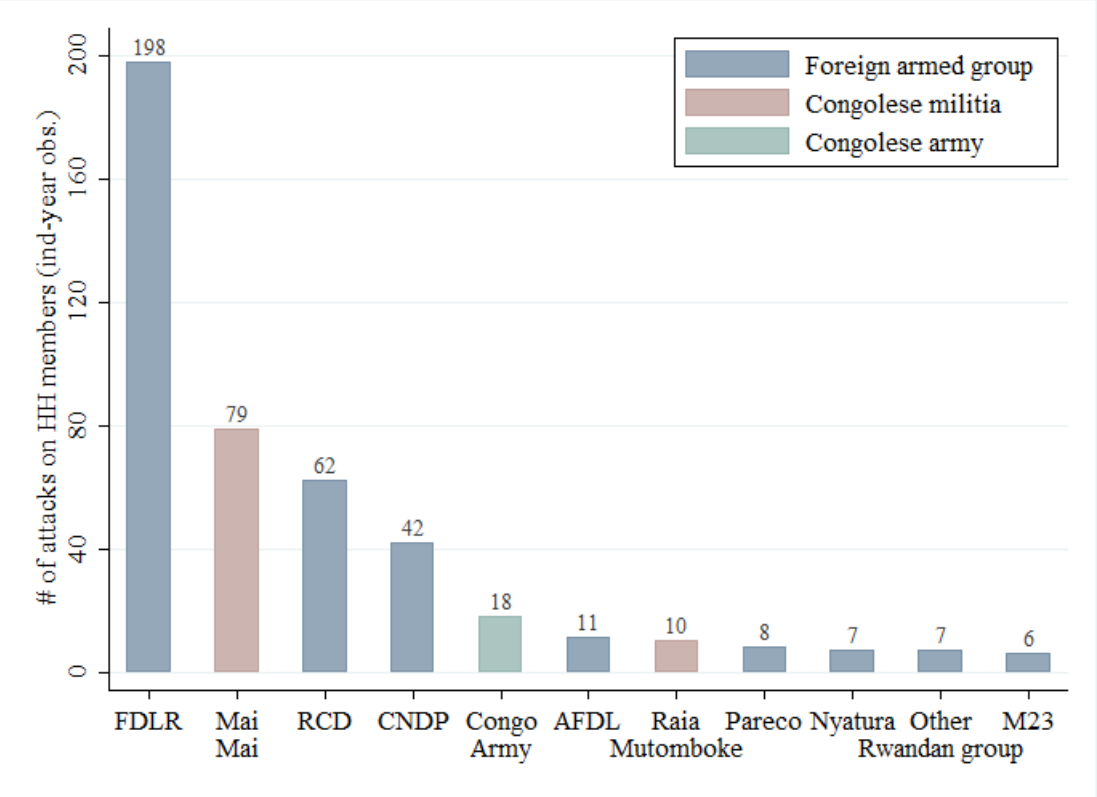
Table C.5: Economic Motives for Becoming a Rebel as Benchmark—Price Shocks

	Participation in t (%)			Occupation in t (%)				
	Any AG	Militia	Foreign	Agr	Mining	Govt	School	Unempl
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Only main occupation and Sud Kivu observations								
$\text{Gold}_j \times \log(P_t)$	-2.22	-2.41*	-0.13	-1.13	3.79**	-0.78	-1.73	-0.15
	(1.39)	(1.29)	(0.33)	(1.88)	(1.86)	(0.96)	(1.21)	(1.68)
Past attack on HH	5.14**	5.22**	-0.70	-2.01	2.75	2.71	-2.84	-0.61
by Foreign armed group	(2.29)	(2.08)	(0.66)	(3.24)	(2.90)	(2.66)	(2.26)	(2.44)
Control mean	4.92	4.24	0.30	53.64	11.10	6.59	8.27	20.40
Obs	12686	12686	12686	11580	11580	11580	11580	11580
Panel B: Including non-main occupations, both South Kivu and North Kivu observations								
$\text{Gold}_j \times \log(P_t)$	-0.92	-1.29	0.14	-2.46*	-0.05	-0.82	-1.95	-0.15
	(1.02)	(0.93)	(0.28)	(1.49)	(1.71)	(1.02)	(1.22)	(1.68)
Past attack on HH	3.53**	3.31**	-0.17	-4.55**	0.88	1.58	-3.57	-0.61
by foreign armed group	(1.53)	(1.39)	(0.45)	(2.26)	(2.25)	(2.81)	(2.36)	(2.44)
Control mean	3.81	2.91	0.43	65.98	20.37	7.33	8.65	20.40
Obs	17578	17578	17578	16472	16472	11580	11580	11580

Notes: This table presents OLS estimates of equation 2. The dependent variables are (a) indicators for whether the respondent joins any armed group, a Congolese militia, or a foreign armed group, respectively, in a given year, and (b) indicators for whether the respondent works in agriculture, mining sector, government office, is still a student or unemployed. The explanatory variable are world (log) price for gold interacting with gold endowment of the village, and an indicator for whether the respondent's household has been attacked by foreign armed groups. We include observations between 1995 and 2013 above age 15 at year t. Panel A includes only main occupations and only limits to observations in South Kivu. Panel B includes non-main occupations and North Kivu observations that have available data on some occupations or gold endowment. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced any attack on household by foreign armed group before year t and did not live in a village endowed with gold. P-value: *** 0.01, ** 0.05, * 0.10.

D Appendix Figures

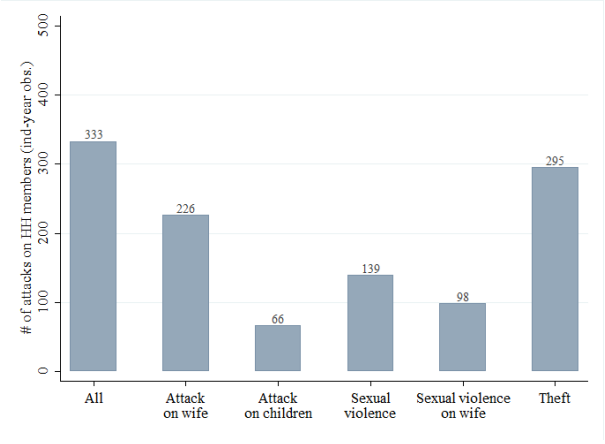
Figure D.1: Descriptive Summaries of Attacks by Perpetrator Identity



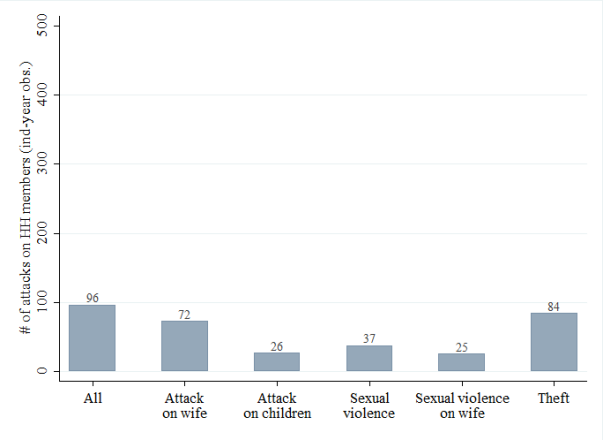
Notes: This figure uses the sample of 475 reported attacks that targeted a households of the respondents in the sample and shows their distribution by detailed perpetrators. Blue bars refer to foreign armed groups; Red bars refer to Congolese militia; Green bar refers to Congolese national army.

Figure D.2: Descriptive Summaries of Attacks by Type

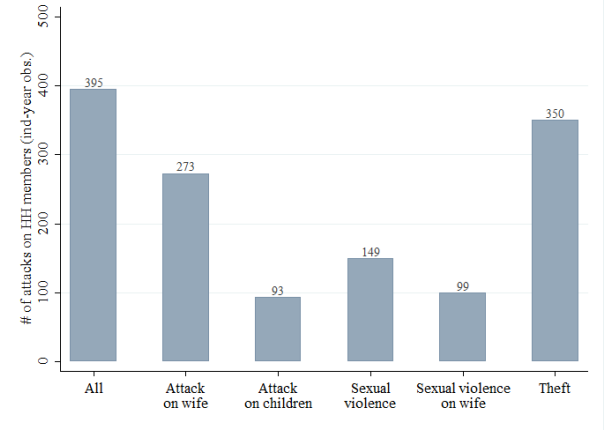
Panel A: Perpetrator: Foreign armed group



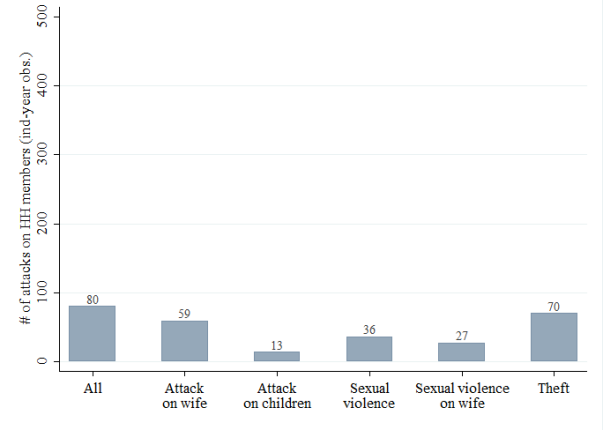
Panel B: Perpetrator: Congolese militia



Panel C: Intention: Pillage

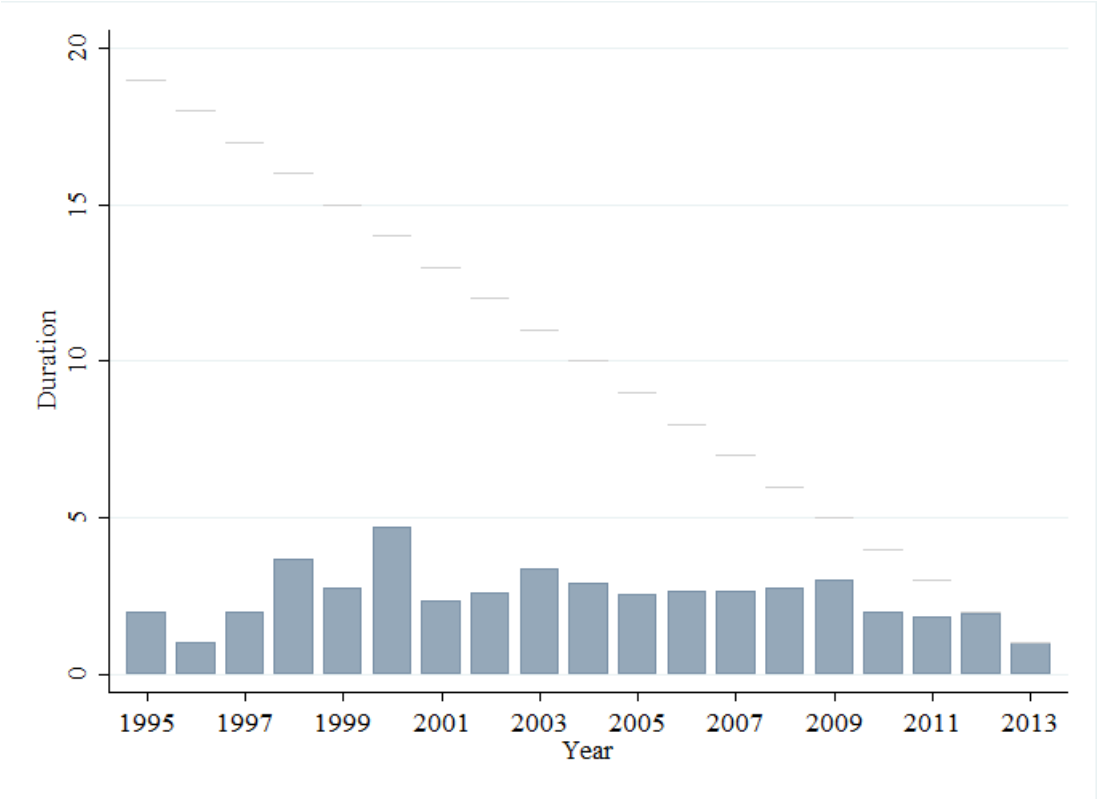


Panel D: Intention: Punishment



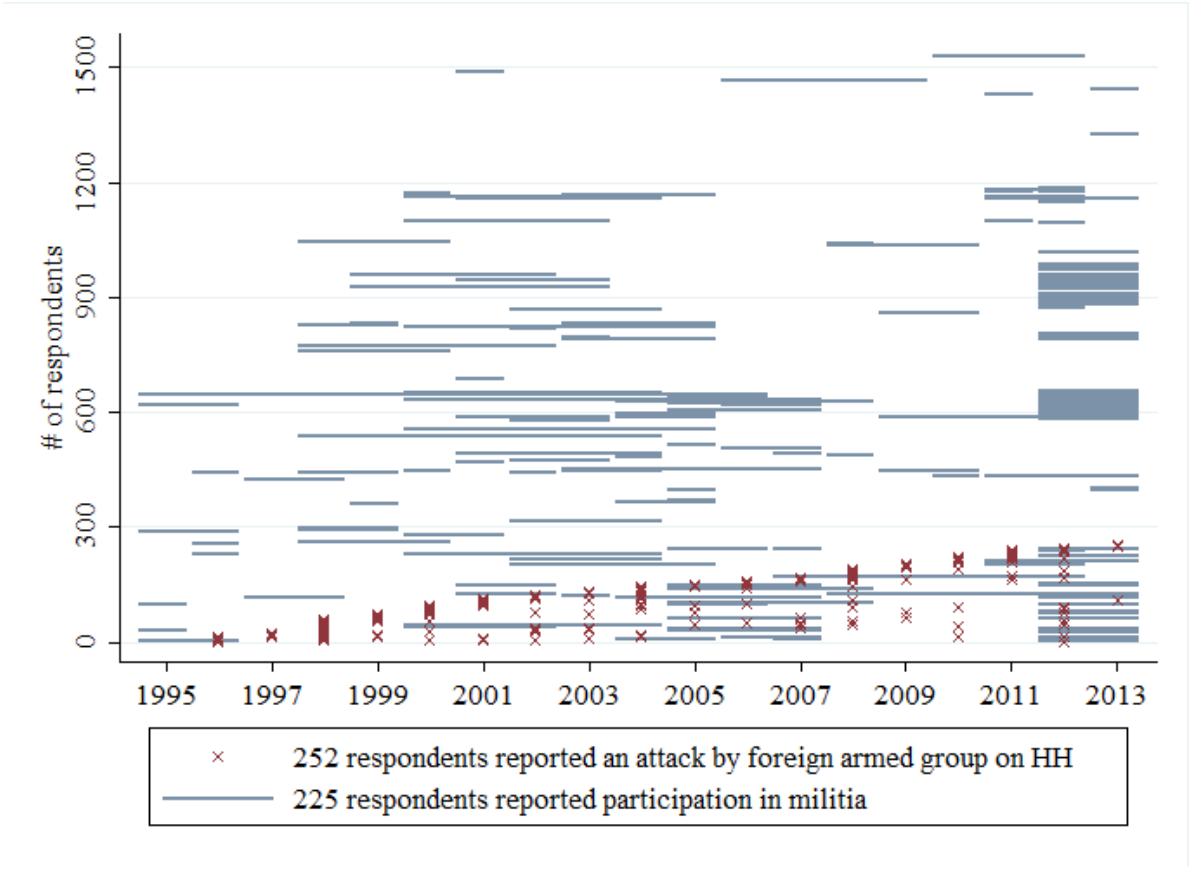
Notes: This figure decomposes different types of attacks on the household by the type of actions that were conducted (not mutually exclusive), respectively: attacks on the spouse, attack on children, attack with sexual violence, attack with sexual violence on spouse, attack in which household property was stolen. We look at four types of attacks respectively: attacks by foreign armed group, attacks by Congolese militia, attacks with intention to pillage, attacks with intention to punish civilians.

Figure D.3: Average Participation Duration in a Congolese militia



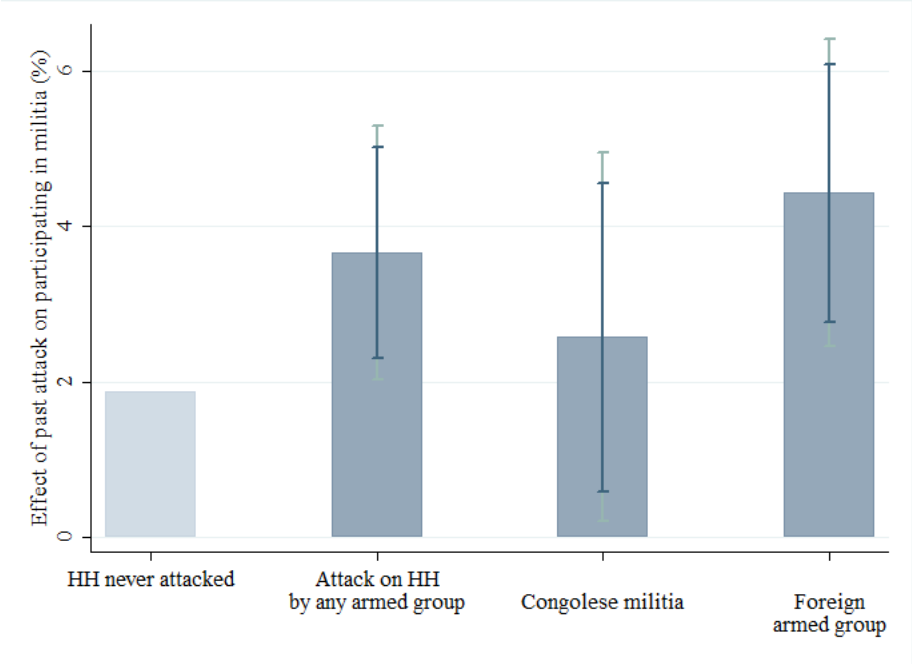
Notes: This figure shows average participation duration in a Congolese militia for each year since 1995. Gray line indicates the maximum year a participant can stay in an armed group.

Figure D.4: Attacks on Household Foreign Armed Group and Participation in a Congolese militia



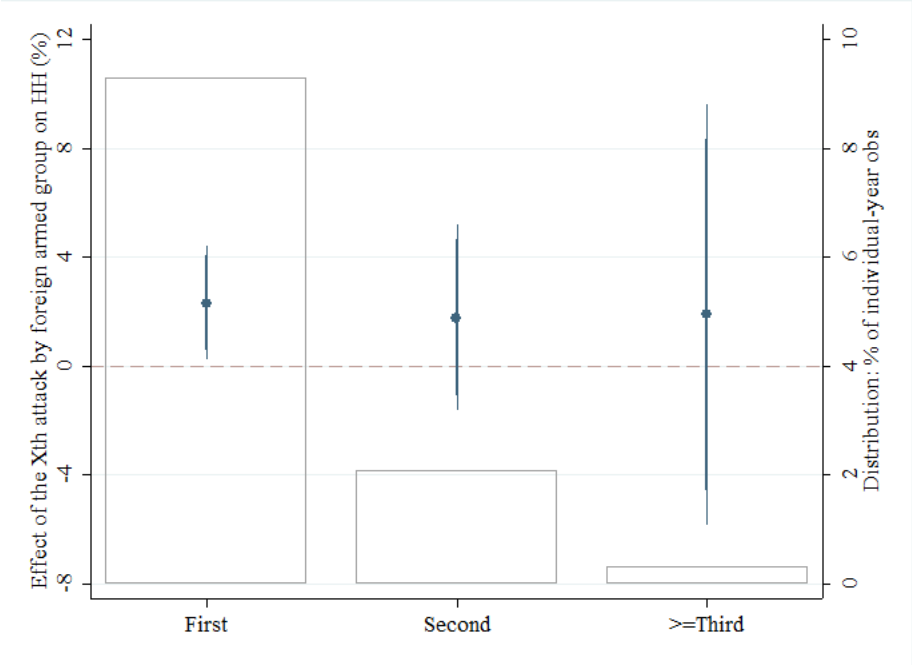
Notes: This figure presents all participation episodes in Congolese militia in the sample from 1995 to 2013 and foreign armed group attacks on the household. Y axis ranks individuals by their date of first reported attack by foreign armed group. X axis shows the years. Red crosses indicate the timing of an attack by a foreign armed group against the household for a given respondent in a given year. Blue lines indicate individual episodes of participation.

Figure D.5: Graphical Representation of the Main Result



Notes: This table presents OLS estimates of equation 1 for different perpetrators (by any armed group, by Congolese militia, and by foreign armed groups respectively) and on participation in a Congolese militia. The dependent variable is an indicator for whether the respondent joins a Congolese militia in a given year. The explanatory variable in each bar is an indicator for whether the respondent's household has been attacked by any armed group, a Congolese militia, foreign armed groups, respectively. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced an attack on household from any armed group before year t. 95% and 90% confidence intervals are shown. See Table II for more discussions.

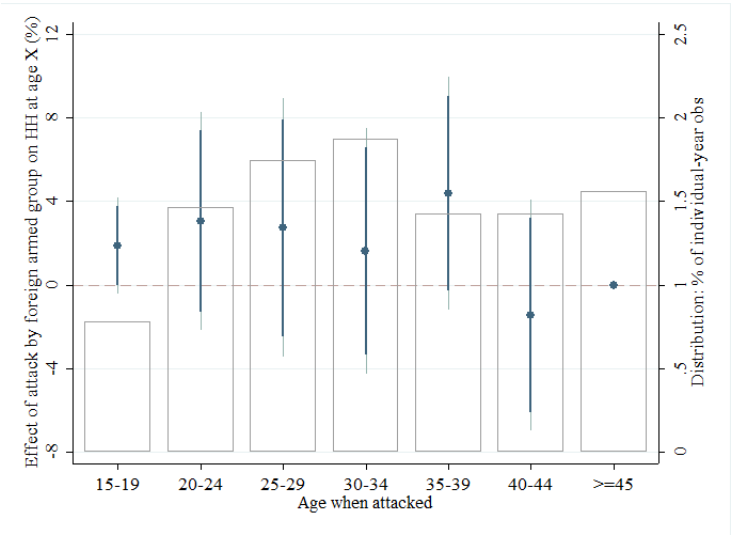
Figure D.6: Order Analysis: The Main Result is Concentrated on the First Attack



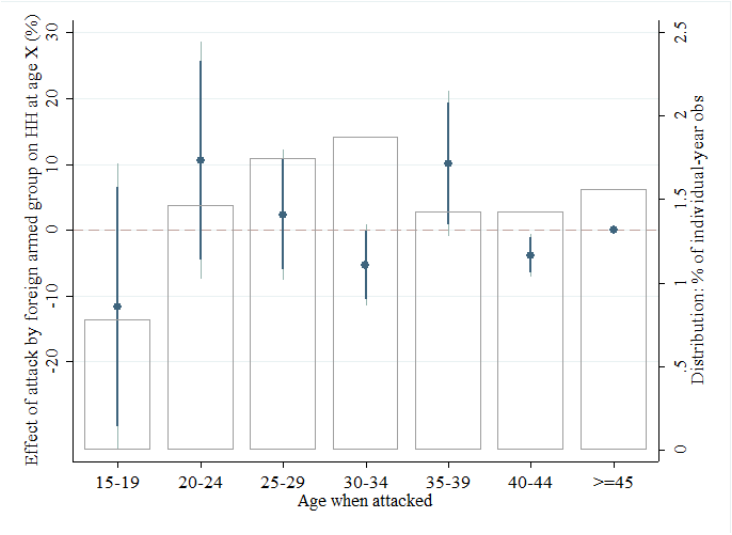
Notes: This table presents OLS estimates of equation 1 for attacks by foreign armed groups and on participation in a Congolese militia. The dependent variable is an indicator for whether the respondent joins a Congolese militia in a given year. The explanatory variables are whether the respondent's household has been attacked by foreign armed groups for the first, second, and third (and more) time. The bars show the fraction (in percentage) of individual-year observations in which respondents have experienced the first, second, and third (and more) foreign armed group attack on household. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. We show 95% and 90% confidence interval. See Table II for more discussions.

Figure D.7: Formation of Preferences: Role of Traumatic Events—Attacks Suffered at an Early Age, Additional Analysis

Panel A: Respondents born before 1980



Panel B: Effect of second attack

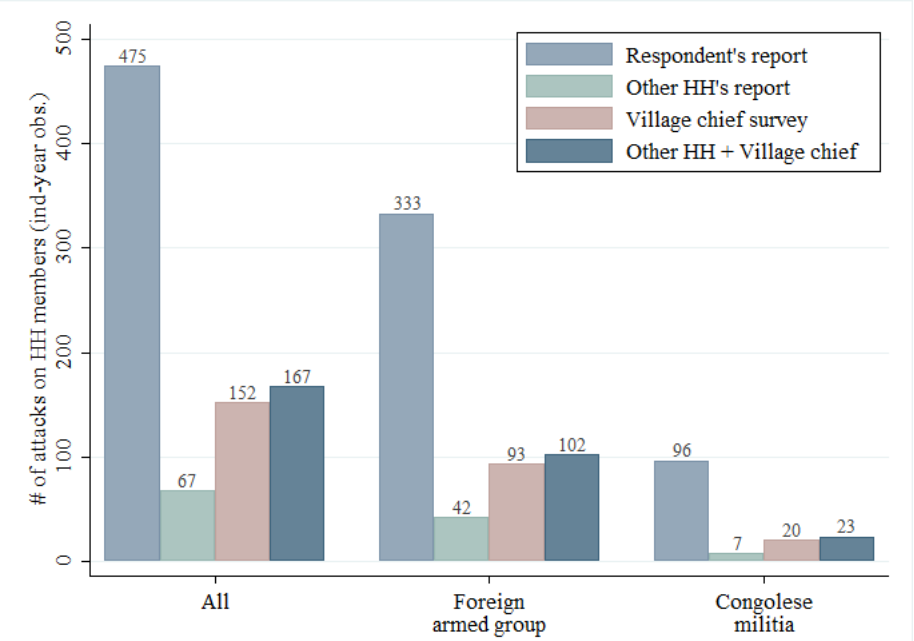


Notes: This figure replicates Figure III, with two deviations. Panel A shows the effect only for respondents who were born before 1980, and thus who are at least 18 years old during First and Second Congo wars and for whom participation into militia was an option. Panel B replicates Figure III using only the second attack. In each panel all coefficients are estimated in one regression. The bars show the fraction (in percentage) of individual-year observations in which respondents have experienced a foreign armed group attack on household at each corresponding age group. We include observations between 1995 and 2013 above age 15 at year t and exclude individuals with no record of foreign armed group attack on household throughout the whole period. We do not include in the regression whether respondents experienced an attack before age 15 because this variable has no variation within individual when we only include observations above age 15. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Brackets show 95% and 90% confidence intervals for each coefficient.

Figure D.8: Times-series of World and Local Gold Prices

Notes: This figure shows the yearly world and local gold prices between 1995 and 2013.

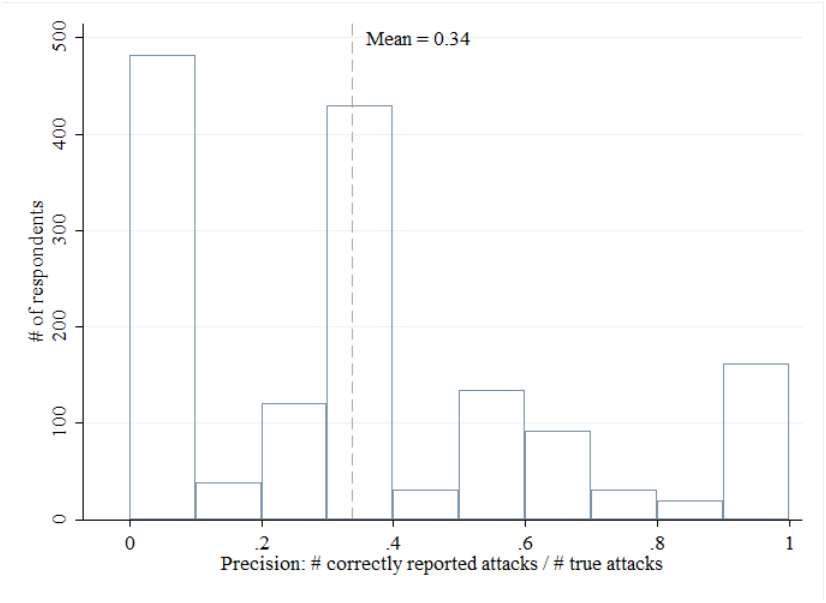
Figure D.9: Cross-validation of Attack Variables



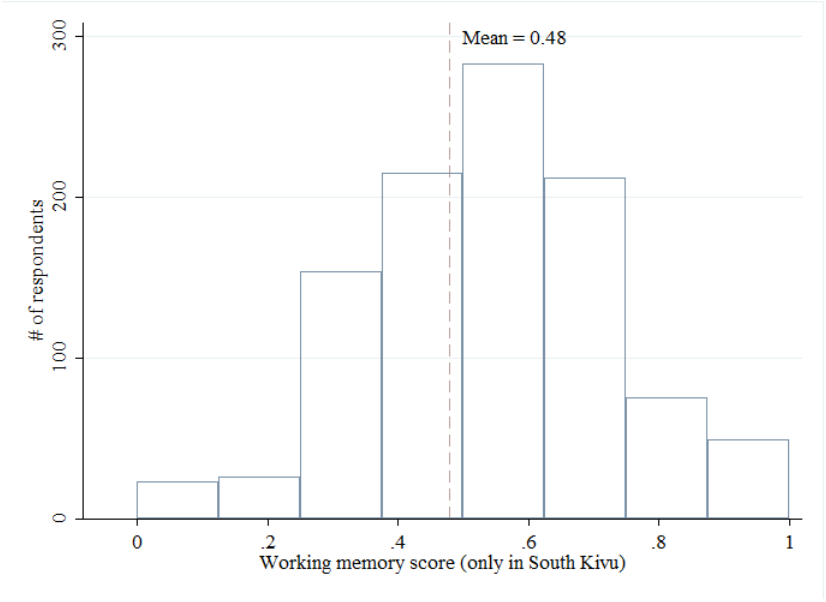
Notes: This figure shows the cross-validation of main attack variables. Light blue bars are reported attacks on household between year 1995 and 2013 from respondents' report. Green bars show attacks on household that are also reported by more than half of other respondents observed in the sample who lived in the same village in the same year. Red bars show attacks on household that are also reported in village chief survey. Dark blue bars show attacks on household that are cross-validated by both village chief survey and more than half of other contemporary respondents.

Figure D.10: Distribution of Report Precision and Working Memory

Panel A: Report precision



Panel B: Working memory



Notes: This figure shows the distribution of report precision and working memory which are used as weights in Table III, Panel B. Report precision in Panel A is defined as the proportion of all the "true" attacks, defined as attack episodes reported by each respondent and also by more than half of other contemporary respondents in the same village or by the village chief. The mean of report precision is 0.34. Working memory in Panel B is constructed from two memory tests described in Section 6.2. The mean of working memory test is 0.48.

E Mathematical appendix

E.1 Model

A unitary household makes decision for infinite time horizon. At time t, the household head decides (1) consumption and investment for all periods (but cannot commit to the plan), and (2) participation into an existing armed group for all periods.[51] We denote C_{it} the consumption at time t for household i, and P_{it} the indicator of whether household head i decides to join an existing armed group at time t. The flow utility is defined as follows:

$$U(C_{it}, P_{it}) = u(C_{it}) + [\theta_{it}(A_{i0}, A_{i1}, ..., A_{i,t-1}) - \phi_t(\overline{\theta}_t, M_t) - \zeta_{it}] \cdot 1(P_{it} = 1). \tag{3}$$

$u(C_{it})$ is a generic concave function satisfying Inada conditions. θ_{it} is the intrinsic utility of joining armed group, depending on household head's past attack history A_{i0}, A_{i1}, ..., $A_{i,t-1}$, which can be interpreted as "revenge utility."[52] ϕ_t is the cost of participation. It depends on two factors: (1) Exogenous constraint M_t. For example, if state force is present, villagers are discouraged to participate in militia. In the meantime, state force might recruit more villagers, so the effect of M_t on ϕ_t can be ambiguous. (2) Average revenge utility $\overline{\theta}_t$. We assume that if villagers enjoy higher utility from revenge on average, the psychological cost for participation is lower, which is designed to capture spillover effect. ζ_{it} is an idiosyncratic utility shock to participation, following some distribution $G(\cdot)$. Revenge (dis)utility can be realized only if the household head participates.

Attack: Attack is realized after household head makes participation decision. We assume

[51] We abstract from the framework endogenous formation of armed groups and leave it for future research.

[52] We assume attack is a continuous variable for simple math derivation. Higher value of A_{it} can be also interpreted as more violence involved. We remain agnostic as to the effect of past attack on intrinsic utility.

each attack event is individually and independently distributed across all t and all i and follows some distribution $F(\cdot)$ with mean a_{it}^1 if household head participates or a_{it}^0 if not.[53]

Death: We consider a possibility of terminating all flow utility after attack is realized. We assume the *expected* survival rate μ is a decreasing function of the attack likelihood a_{it}. For simplicity, we assume $\mu(a_{it}^1) - \mu(a_{it}^0) = \eta(a_{it}^1 - a_{it}^0)$ for some non-increasing function $\eta(\cdot)$.

Income: If household head survives, he earns labor income w through production, or T through appropriation by participating in armed group. We assume $T > w$.

Investment: Household invests the rest of the labor income on capital. Household possesses initial capital K_{it}. Capital depreciates at the rate of δ_{it}. Depreciation rate is higher if household suffers an attack: $\frac{\partial \delta_{it}}{\partial A_{it}} > 0$. Household earns capital gain r from each unit of current remaining capital.

Timing. (1) Household decides whether or not to participate, upon which intrinsic revenge utility θ_{it}, participation cost ϕ_t and idiosyncratic utility shock ζ_{it} are immediately realized. (2) Attack A_{it} is drawn from $F(\cdot \mid a_{it}^0)$ or $F(\cdot \mid a_{it}^1)$ depending on participation status. (3) If household survives, then household decides consumption C_{it} and investment I_{it}. (4) Household enters the next period.

Proposition 1 characterizes the effect of past attack on consumption, capital level, and the present value of participation compared to nonparticipation.

Proposition 1. *Suppose attack A_{it} is a continuous variable, capital depreciation rate δ_{it} is differentiable with regard to A_{ik}, $\forall k \leq t$. Then consumption C_{it} and capital $K_{i,t+1}$ decrease*

[53] We assume idiosyncratic attack likelihood for two reasons. First, individuals with different characteristics might be targeted differently; for instance, see Table C.2. Second, a_{it} can be interpreted as subjective expectation of attack likelihood, which varies for different individuals. We also remain agnostic whether $a_{it}^1 > a_{it}^0$ or not. If $a_{it}^1 > a_{it}^0$, then participation may lead to more targeting from enemies. If $a_{it}^1 < a_{it}^0$, then participation may provide more protection for the household.

in A_{ik}, $\forall k \leq t$, *whether participating or not.*

Proposition 2 decomposes the effect of past attack on participation into intrinsic motivation, wealth effect, and protection effect.

Proposition 2. *Suppose attack* A_{it} *is a continuous variable, capital depreciation rate* δ_{it}, *intrinsic utility* θ_{it}, *attack likelihood* a_{it}^0 *and* a_{it}^1, *and survival function* $\mu(\cdot)$ *are differentiable with regard to* A_{ik}, $\forall k \leq t$. *Then the effect of* A_{ik} *on participation likelihood* p_{it} *can be decomposed into a linear combination of the following three components:*

1. *Intrinsic motivation:* $\frac{\partial(\theta_{it}-\phi_t)}{\partial A_{ik}}$
2. *Wealth effect, positive if marginal utility of contingent consumption weighted by attack likelihood is lower when participating;*
3. *Protection effect*

From the proof of Proposition 2, we are able to rule out wealth effect by controlling in the main regression investment proxy (in the empirical analysis, the principal component of investment in farm animals). Similarly, we can rule out protection effect by controlling subjective attack likelihood (in our empirical analysis, whether respondent overreports any violent event).

E.2 Solution

We solve this problem by writing Bellman equation first. Some notations for simplicity:

$V_t^{N,s}(K)$: Present value if not participating in period t, after household head survives

$V_t^{P,s}(K)$: Present value if participating in period t, after household head survives

$V_t^N(K)$: Present value if not participating in period t, before attack is realized

$V_t^P(K)$: Present value if participating in period t, before attack is realized

$V_t(K)$: Present value before household head makes participation decision

Let β be the discount rate, and suppose household survives. The continuation utility from not participating is as follows:

$$V_t^{N,s}(K_t) = \max_{C_t, I_t} \; u(C_t) + \beta E_t[V_{t+1}(K_{t+1})|P_t = 0]$$
$$\text{s.t.} \;\; C_t + I_t = w + r(1-\delta_t)K_t, \quad K_{t+1} = (1-\delta_t)K_t + I_t.$$

The first order condition is:

$$u'(C_t^N) = \beta E_t[V'_{t+1}(K_{t+1}^N)], \tag{4}$$

where C_t^N, $K_{t+1}^N = (1-\delta_t)(1+r)K_t + w - C_t^N$ is the optimal choice given nonparticipation.

For the continuation utility from participation, similarly,

$$V_t^{P,s}(K_t) = \max_{\{C_t\}} \; u(C_t) + \beta E_t[V_{t+1}(K_{t+1})|P_t = 1]$$
$$\text{s.t.} \;\; C_t + I_t = T + r(1-\delta_t)K_t, \quad K_{t+1} = (1-\delta_t)K_t + I_t.$$

The first order condition becomes:

$$u'(C_t^P) = \beta E_t[V'_{t+1}(K_{t+1}^P)], \tag{5}$$

where C_t^P, $K_{t+1}^P = (1-\delta_t)(1+r)K_t + T - C_t^P$ is the optimal choice given participation.

Let's move backward before attack is realized. Household head would weight the present value by survival likelihood:

$$V_t^N(K_t) = \mu(a_t^N)V_t^{N,s}(K_t)$$
$$V_t^P(K_t) = \theta_t - \zeta_t + \mu(a_t^P)V_t^{P,s}(K_t).$$

And household chooses to participate when $V_t^P(K_t)$ is larger:

$$V_t(K_t) = \max[V_t^N(K_t), V_t^P(K_t)].$$

Hence, at the beginning of each period, the probability of household participating would be:

$$\begin{aligned} p_t(x) &\equiv Pr[V_t^P(x) > V_t^N(x)] \\ &= Pr\big[\zeta_t < \theta_t + \mu(a_t^P)V_t^{P,s}(x) - \mu(a_t^N)V_t^{N,s}(x)\big] \\ &= G\big[\underbrace{\theta_t}_{\text{Intrinsic utility}} + \underbrace{\mu(a_t^P)V_t^{P,s}(x) - \mu(a_t^N)V_t^{N,s}(x)}_{\text{Protection and wealth}}\big] \end{aligned} \tag{6}$$

E.3 Proofs

Before we prove the two propositions, we claim that there is no clear prediction on the effect of participation on consumption or investment. Since value function $V_{t+1}(\cdot)$ is the maximum of $V_{t+1}^P(\cdot)$ and $V_{t+1}^N(\cdot)$, the expectation of $V_{t+1}'(\cdot)$ depends on the likelihood of participating

in $t+1$. Break down $E_t[V'_{t+1}(x)]$:[54]

$$\begin{aligned}
E_t[V'_{t+1}(x)] &= E_t p_{t+1}(x)\frac{\partial E_t(V^P_{t+1})'}{\partial x} + (1 - E_t p_{t+1}(x))\frac{\partial E_t(V^N_{t+1})'}{\partial x} \\
&= E_t p_{t+1}(x) E_t\big[\mu(a^P_{t+1})(1-\delta_{t+1})(1+r)u'(C^P_{t+1}(x))\big] \\
&\quad + (1 - E_t p_{t+1}(x)) E_t\big[\mu(a^N_{t+1})(1-\delta_{t+1})(1+r)u'(C^N_{t+1}(x))\big],
\end{aligned}$$

where $p_{t+1}(x)$ is defined as equation 6, and the second line comes from Envelop Theorem and Euler equations. Take the derivative of $E_t[V'_{t+1}(x)]$:

$$\frac{1}{1+r}\frac{\partial E_t[V'_{t+1}(x)]}{\partial x} = \underbrace{G'(\cdot)\big(E_t[(1-\delta_{t+1})(\mu(a^P_{t+1})u'(C^P_{t+1}) - \mu(a^N_{t+1})u'(C^N_{t+1}))]\big)^2}_{>0} \tag{7}$$

$$+ \underbrace{p_{t+1}E_t[(1-\delta_{t+1})\mu(a^P_{t+1})u''\frac{\partial C^P_{t+1}}{\partial x}]}_{<0} + \underbrace{(1-p_{t+1})E_t[(1-\delta_{t+1})\mu(a^N_{t+1})u''\frac{\partial C^N_{t+1}}{\partial x}]}_{<0} \tag{8}$$

The second and third terms are negative because consumption is a normal good, and thus given the expected participation likelihood, if household has more capital, regardless of participation status, household would consume more in $t+1$ and marginal expected present value would decrease. The first term, however, is positive. This is because the expected difference in marginal utility of contingent consumption (weighted by protection effect of participation, mathematically, $\mu(a^P_{t+1})u'(C^P_{t+1}) - \mu(a^N_{t+1})u'(C^N_{t+1})$) also affects the expected

[54] Derivation of $\frac{\partial E_t p_{t+1}}{\partial x}$ also depends on the difference between marginal utility of contingent consumption, weighted by protection of participation:

$$\frac{\partial E_t p_{t+1}}{\partial x} = (1+r)G'(\cdot)E_t[(1-\delta_{t+1})(\mu(a^P_{t+1})u'(C^P_{t+1}) - \mu(a^N_{t+1})u'(C^N_{t+1}))]$$

participation likelihood $E_t p_{t+1}$ in the same direction. Intuitively, when marginal utility of consumption when participating in $t+1$ is higher (lower) from that of nonparticipating, household will have a higher (lower) likelihood of participation in $t+1$, which in turn increases the marginal expected present value. If the difference in marginal utility of contingent consumption is stark enough, the expected value function would exhibit nonconcavity for some range of capital value, in which case it is possible that $E_t[V'_{t+1}(K^P_{t+1})] > E_t[V'_{t+1}(K^N_{t+1})]$, thus not violating Euler equations.

Therefore, there is no clear prediction on the effect of participation on consumption or investment.[55] One possible scenario of a drastic difference in marginal utility of contingent consumption is that when participation brings a much higher (lower) protection effect—that is, $\mu(a^P_{t+1}) \gg (\ll)\mu(a^N_{t+1})$. If there is no protection effect of participation detected, given proper assumption of utility function and density function of idiosyncratic intrinsic utility ζ_t, it is less likely to observe nonconcavity in the expected value function, and thus household would behave more normally (consumption increases in t when participating t).

Proof of Proposition 1. Rewrite the budget constraint for participation:

$$C^P_t + K^P_{t+1} = T + (1+r)(1-\delta_{it})K_t.$$

[55]The logic on investment is similar. Suppose participants overconsume t so that they have lower capital level in $t+1$, that is, $K^P_{t+1} \leq K^N_{t+1}$, $C^P_t \geq C^N_t + T - w$. With concave utility function, Euler equations and Mean Value Theorem,

$$\exists\ \theta \in [C^N_{t+1}, C^N_{t+1} + T - w], \quad \text{s.t. } E_t[V'_{t+1}(K^P_{t+1})] - E_t[V'_{t+1}(K^N_{t+1})] \leq u''(\theta)(T-w) < 0.$$

If the expected value function exhibits sufficient nonconcavity, the condition above is possible.

Take derivative with regard to A_k:

$$\frac{\partial C_t^P}{\partial A_k} + \frac{\partial K_{t+1}^P}{\partial A_k} \equiv (1+r)[-\frac{\partial \delta_t}{\partial A_k} K_t + (1-\delta_t)\frac{\partial K_t}{\partial A_k}]. \tag{9}$$

Take derivative of Euler equation 5 with regard to A_k:

$$u''(C_t^P)\frac{\partial C_t^P}{\partial A_k} \equiv H(K_{t+1}^P) \cdot \frac{\partial K_{t+1}^P}{\partial A_k}, \tag{10}$$

where $H(x) \equiv \frac{\partial E_t[V'_{t+1}(x)]}{\partial x}$. Solving equation 9 and 10:

$$\frac{\partial C_t^P}{\partial A_k} = \frac{1}{1 + u''(C_t^P)/H(K_{t+1}^P)} \cdot (1+r)[-\frac{\partial \delta_t}{\partial A_k} K_t + (1-\delta_t)\frac{\partial K_t}{\partial A_k}]. \tag{11}$$

Likewise, one can derive $\frac{\partial C_t^N}{\partial A_k}$ by simply replacing superscript P with N. To simplify the question, assume $u''(C_t^P)/H(K_{t+1}^P) = u''(C_t^N)/H(K_{t+1}^N) \equiv \gamma \in (0, +\infty)$.[56] Thus, the effect of attack on consumption, capital, and investment are as follows (superscripts are omitted because we assume participation does not influence the effect of attack on these variables):

$$\frac{\partial C_t}{\partial A_k} = \frac{1+r}{1+\gamma} \cdot [-\frac{\partial \delta_t}{\partial A_k} K_t + (1-\delta_t)\frac{\partial K_t}{\partial A_k}] \tag{12}$$

$$\frac{\partial K_{t+1}}{\partial A_k} = \frac{1+r}{1+1/\gamma} \cdot [-\frac{\partial \delta_t}{\partial A_k} K_t + (1-\delta_t)\frac{\partial K_t}{\partial A_k}] \tag{13}$$

$$\frac{\partial I_t}{\partial A_k} = \frac{r-1/\gamma}{1+1/\gamma} \cdot [-\frac{\partial \delta_t}{\partial A_k} K_t + (1-\delta_t)\frac{\partial K_t}{\partial A_k}] \tag{14}$$

The term $[-\frac{\partial \delta_t}{\partial A_k} K_t + (1-\delta)\frac{\partial K_t}{\partial A_k}]$ is always negative because either attack increases current de-

[56] Intuitively, γ measures the difference between concavity of consumption in t and that of continuation value in the future. $\gamma > 0$ is predicated on expected value function being concave (see the discussion above). If γ is higher, household's utility of consumption today reacts more to the exogenous shock than consumption tomorrow.

preciation rate ($\frac{\partial \delta_t}{\partial A_k} > 0$ if $k = t$), or past attack decreases past capital, which decreases current capital accumulation ($\frac{\partial K_t}{\partial A_k} < 0$ if $k < t$ by induction). Given the assumption $\gamma \in (0, 1)$, the effect of past attack on consumption and capital should be unambiguously negative. The effect on investment, however, depends on the comparison between the return to investment r, versus the concavity of value function $1/\gamma$, which determines the difference in marginal value between consuming today and in the future. □

Proof of Proposition 2. Take differentiation of equation 6 with regard to A_k:

$$\begin{aligned}\frac{\partial p_t}{\partial A_k} &= G'(\cdot)\big[\frac{\partial \theta_t}{\partial A_k} + \frac{\partial\big(\mu(a_t^P)V_t^{P,s} - \mu(a_t^N)V_t^{N,s}\big)}{\partial A_k}\big] \\ &= G'(\cdot)\big[\underbrace{\frac{\partial \theta_t}{\partial A_k}}_{\text{Intrinsic motivation}} + \underbrace{\big(\mu(a_t^P)\frac{\partial V_t^{P,s}}{\partial A_k} - \mu(a_t^N)\frac{\partial V_t^{N,s}}{\partial A_k}\big)}_{\text{Wealth effect}} + \underbrace{\big(\frac{\partial \mu(a_t^P)}{\partial A_k}V_t^{P,s} - \frac{\partial \mu(a_t^N)}{\partial A_k}V_t^{N,s}\big)}_{\text{Protection effect}}\big]\end{aligned}$$

where the second equation comes from Envelop Theorem and Euler equations. By assumption, intrinsic motivation would increase if household head experienced a past attack which leads to higher participation likelihood.

Wealth effect: Use Envelop Theorem and Euler equations, for $k \leq t$, the difference

between contingent present values of participation vs. nonparticipation:[57]

$$\begin{aligned}
&\mu(a_t^P)\frac{\partial V_t^{P,s}}{\partial A_k} - \mu(a_t^N)\frac{\partial V_t^{N,s}}{\partial A_k} \\
&= (1+r)[-\frac{\partial \delta_t}{\partial A_k}K_t + (1-\delta)\frac{\partial K_t}{\partial A_k} - \frac{\partial C_t^P}{\partial A_k}]\mu(a_t^P)u'(C_t^P) \\
&\quad - (1+r)[-\frac{\partial \delta_t}{\partial A_k}K_t + (1-\delta)\frac{\partial K_t}{\partial A_k} - \frac{\partial C_t^N}{\partial A_k}]\mu(a_t^N)u'(C_t^N).
\end{aligned}$$

Substitute equation 12 into the equation,

$$\begin{aligned}
&\mu(a_t^P)\frac{\partial V_t^{P,s}}{\partial A_k} - \mu(a_t^N)\frac{\partial V_t^{N,s}}{\partial A_k} = \\
&\quad (1+r)\underbrace{(\frac{1}{1+1/\gamma})}_{>0}\underbrace{[-\frac{\partial \delta_t}{\partial A_k}K_t + (1-\delta)\frac{\partial K_t}{\partial A_k}]}_{<0}\cdot\big(\mu(a_t^P)u'(C_t^P) - \mu(a_t^N)u'(C_t^N)\big). \qquad (15)
\end{aligned}$$

Therefore, the wealth effect of attack depends on the term of $\mu(a_t^P)u'(C_t^P) - \mu(a_t^N)u'(C_t^N)$, i.e., difference in marginal utility of contingent consumption weighted by protection of participation. Intuitively, given the protection effect of participation, if household thinks participation brings higher consumption and thus lower marginal utility of consumption, it is more likely that he switches from nonparticipation to participation status, and vice versa.

In addition, wealth effect of attack on participation can be expressed in terms of the effect of attack on investment:

$$\mu(a_t^P)\frac{\partial V_t^{P,s}}{\partial A_k} - \mu(a_t^N)\frac{\partial V_t^{N,s}}{\partial A_k} = \frac{1+r}{r-1/\gamma}\big(\mu(a_t^P)u'(C_t^P) - \mu(a_t^N)u'(C_t^N)\big)\cdot\frac{\partial I_t}{\partial A_k}.$$

[57] We assume here that attack does not affect labor income T or w. It is possible that past attack might affect respondents' productivity in different sectors, for instance, through handicap, and thus past attack might affect labor income. Our empirical analysis shows no additional effect of past attack that only affects respondents, and our main analysis focuses on attacks on other household members, which arguably does not affect the labor income of household head in principle. More discussion in Appendix F.

Therefore, if $\mu(a_t^P)u'(C_t^P) - \mu(a_t^N)u'(C_t^N)$ does not vary much, one can control for investment in time t to control for wealth effect of past attack on participation.

Protection effect: The direction is ambiguous. If expected function is well-behaved and concave, it is easy to show $V_t^{P,s} > V_t^{N,s}$ because participation given survival brings higher labor income and thus higher expected continuation value. If the subjective survival likelihood does not vary much with past attack, then protection effect is positive. If, however, the subjective survival likelihood decreases sufficiently when household experienced a past attack, household head would be less likely to participate. In the main analysis, we control for whether respondent overreports any violent event as a proxy to subjective survival likelihood to control for the protection effect.

□

F Attack on respondent himself

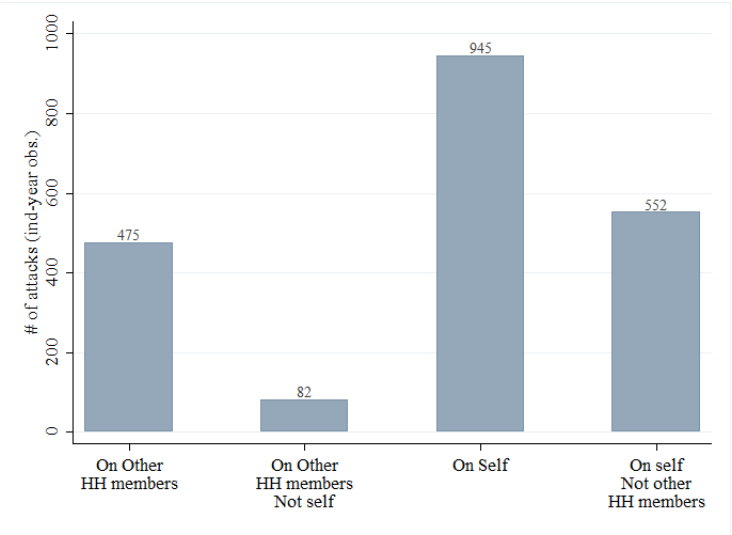
In the main text, we choose attacks on other household members as main attack definition instead of attacks on respondents themselves. The main concern of using attacks on respondents themselves, as we discussed in Section 8, is that attacks on respondents themselves can possibly affect respondents' physical capabilities of participation. The effect of past attack on respondents themselves is likely a mix of the negative effect from decreased physical capabilities and the positive effect from intrinsic preferences.

Figure F.1 shows the distribution of attacks on other household members and attacks on self. Out of 475 attacks where respondent's other household members are attacked, most of them involve respondent themselves being attacked; only 87 of these episodes do not involve respondent themselves being attacked. There are 945 episodes where respondents reported themselves attacked during the violent events; 552 of them do not involve other household members being attacked.

Table F.1 replicates the main results in Table II but breaks down the effect by attacks on other household members and on respondents themselves. We only examine attacks by foreign armed group and participation into a Congolese militia. Column (1) is the benchmark regression. Column (2) breaks down the main attack variable into attacks only involving other household members and attacks involving both other household members and respondents themselves. The effect of attack on both is statistically significant. The effect of attack only on other household members is less precisely estimated due to the insufficient number of such events (only 87), but the coefficient remains similar to the main result (2.92 vs. 2.55).

Column (3) examines the effect of attack on respondents themselves. The effect is a

Figure F.1: Distribution of Attack on Other HH Members and Self



Notes: This figure shows the distribution of attack observations where other household members or respondent himself was attacked.

precise zero. Column (4) breaks down the effect into attacks only involving respondents and attacks involving both other household members and respondents themselves. The effect of attack on both is still statistically significant, but the effect of attack involving only respondents is again a precise zero. Considering the large number of such events (552), we do not think the zero effect is due to imprecision. Column (5) includes attacks only involving other household members, attacks only involving respondents and attacks on both. The main conclusion remains unchanged. All the evidence suggest that attacks only on respondents cannot capture well the intrinsic preferences of participation because it potentially affects respondents' physical capabilities to join an armed group. We thus adopt attacks on other household members as the main definition of attack throughout the paper.

Table F.1: Main Regression, Separating Attacks on Other HH Members and Respondents

	Participation in Congolese militia in year t (%)				
Past attack by foreign armed group	(1)	(2)	(3)	(4)	(5)
On other HH members	2.55**				
	(1.01)				
On other HH members, not on respondents		2.92			2.92
		(2.43)			(2.43)
On respondents			-0.02		
			(0.64)		
On respondents, not on other HH members				-0.05	-0.02
				(0.73)	(0.73)
On respondents and on other HH members		2.37**		2.38**	2.37**
		(1.05)		(1.05)	(1.05)
Control mean	1.89	1.89	1.82	1.82	1.81
Obs.	25060	25060	25060	25060	25060

Notes: This table presents OLS estimates of equation 1, with different attack definitions. The dependent variable is an indicator for whether the respondent joins a Congolese militia. The main explanatory variable is an indicator for whether the respondent's household has been attacked by foreign armed groups. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced an attack on household by foreign armed group before year t. P-value: *** 0.01, ** 0.05, * 0.10.

G Migration

G.1 Difference between migrants and stayers

We first look into whether villagers who migrated are systematically different. Table G.1 compares individual-year observations where villagers moved to a new village in year t versus those where villagers stay in the same village. In total there are 1,223 migration episodes. Migrants are more likely to have participated in any armed group before, but less in a Congolese militia. They tend to be younger, less likely to have married, less likely to have worked in mining or agriculture, and more likely to be in school or unemployed before moving. They do not differ in economic status before migration, although they invest more after they migrate. Migrants are not more likely to experience any attack by foreign armed group in the past, but are less likely to have experienced an attack by Congolese militia.

G.2 Migration as a source of selection bias

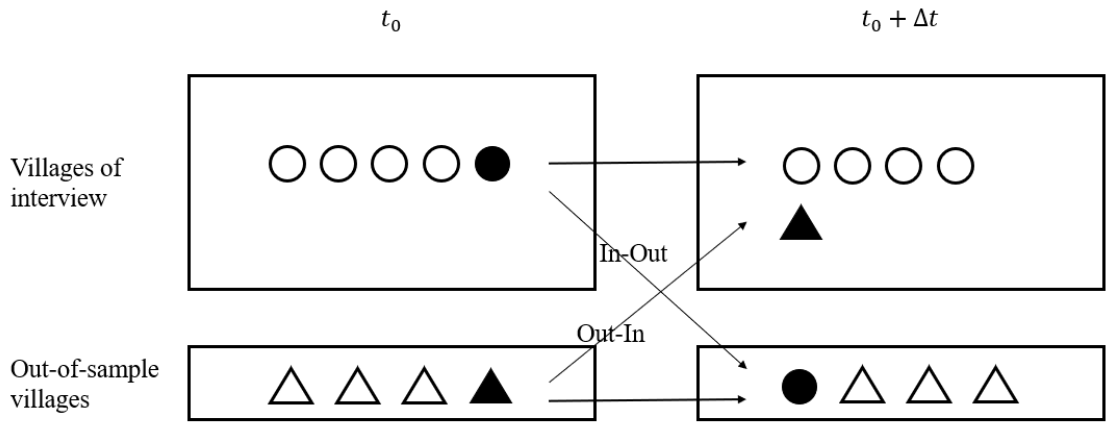
To discuss selection bias resulting from migration, it is necessary to first define the ideal average treatment effect (ATE) of the study. Suppose at time t_0 we have a representative sample from the villages we interview, and we want to estimate the treatment effect of past attack on villagers from the villages of interview. After a period Δt, however, some villagers emigrate to an out-of-sample village (In-Out migration), and some villagers migrate into a village of interview (Out-In migration). A random draw from the interview villages in $t_0+\Delta t$ will not be representative of villagers from the villages we interview at time t_0. Notice that some villagers migrate within villages of interview, but this does not cause the selection bias because they do not alter the composition of villagers from the villages of interview.

Table G.1: Characteristics of Migrants (1,223 migration episodes)

	Nonmigrant (1)	Migrant (2)
Obs. (Age$\geq$ 15, 1995 $\leq t \leq$ 2013)	23,868	1,223
Panel A: Participation history		
Ever participated in any armed group before year t (%)	8.05	8.26*
Ever participated in Congolese militia in year t (%)	5.40	4.91***
Ever participated in foreign armed group in year t (%)	2.02	1.64
Panel B: Socio-demographic background per individual-year obs		
Age in year t	32.33	28.67***
Married in year t (%)	44.31	38.76**
Works primarily in mining in year $t-1$ (%)	14.21	12.48**
Works primarily in agriculture in year $t-1$ (%)	47.18	38.6**
Works primarily as civil servant in year $t-1$ (%)	7.11	7.32
Works primarily in school in year $t-1$ (%)	9.36	12.57**
Unemployed in year $t-1$ (%)	22.14	28.99***
Panel C: Economic status per individual-year obs		
Wealth at birth (z-score, only $t = 2012$)	0.04	0.07
Asset stock in year $t-1$ (z-score)	-0.03	-0.05
Investment in year $t+1$ (z-score)	0.09	0.23***
Panel D: Attack history per indivudla-year obs		
Experienced attack on own HH by any armed group before year t (%)	12.08	11.45
Experienced attack on own HH by foreign armed group before year t (%)	9.32	9.08
Experienced attack on own HH by Congolese militia before year t (%)	3.02	2.13**

Notes: This table shows descriptive statistics for observations where respondents move to a new village versus those where respondents stay in year t. Economic indices are only computed in South Kivu, where the data are available. We indicate the difference between Column 1 and 2 (P-value: *** 0.01, ** 0.05, * 0.10), computed after including village FE and year FE, and clustered two-way at the individual respondent and the village*year level. Construction of economic indices (principal component analysis on following variables): Wealth at birth index—Stock of cows at birth, stock of goats at birth, stock of pigs at birth, stock of lands at birth, relation to village chief, number of father's wives. Asset stock index—stock of cows, stock of goats, stock of pigs, stock of lands. Investment index—purchase of cows, purchase of goats, purchase of pigs, purchase of lands.

Figure G.1: Illustration of Selection Bias



More formally, see Figure G.1:

1. Suppose villages of interview (Group A) constitute proportion $a \in [0, 1]$ of the East Congo population.

2. Within villages of interview, proportion $1-\pi$ of the villagers will never migrate outside (A_1). Proportion π of the villagers will migrate to out-of-sample villages at least once throughout the period (A_2) with probability p.

3. Within out-of-sample villages (Group B), proportion $1-\pi$ of the villagers will never migrate outside (B_1). Proportion π of the villagers will migrate to villages of interview at least once throughout the period (B_2) with probability p.

Assume the real treatment effect of each group is $T(X)$, and past attack does not change

the composition of different subgroups. Treatment effect of interest:

$$ATE(A) = (1-\pi)T(A_1) + \pi T(A_2).$$

Estimate of the treatment effect:

$$\begin{aligned} A\hat{TE}(A) &= \frac{a(1-\pi)}{a(1-\pi)+(1-a)p\pi}T(A_1) + \frac{(1-a)p\pi}{a(1-\pi)+(1-a)p\pi}T(B_2) \\ &= \frac{1}{1+(\frac{1-a}{a}p-1)\pi}ATE(A) + \frac{(1-a)p\pi T(B_2) - a\pi T(A_2)}{a(1-\pi)+(1-a)p\pi}, \end{aligned} \tag{16}$$

There are at least two different ways that selection bias affects the estimation of ATE (through the coefficient term and constant). If, however, we assume that past attack does not affect the composition of migrants and nonmigrants, one can at least estimate the treatment effect on villagers who never migrate outside sample villages without bias ($T(A_1)$).

Table G.2 estimates the effect of past attack by Foreign armed group on participation in a Congolese militia. Column (1) shows the benchmark of the main result. Column (2) implements the main specification 1 within villagers who never migrate outside of the villages of interview. The effect on these villagers are larger and remain statistically significant. This is an unbiased estimate of the effect of past attack on villagers who never migrate outside of the villages of interview if past attack does not affect the composition of migrants ($T(A_1)$). Columns (3) and (4) implements the main specification but controls for different types of migrants. The estimates of $T(A_1)$ remain largely similar.

Results in Columns (3) and (4) also suggest that migrants from outside of the villages of interview, if anything, are negatively selected. Those who are less likely to participate tend

Table G.2: Effect of Past Attack on Migration, on Villagers Who Never Migrated Outside of Sample

	Participation in Congolese militia in year t (%)			
	(1)	(2)	(3)	(4)
Past attack on HH by foreign armed group	2.55**	4.02***	4.12***	4.14***
	(1.01)	(1.48)	(1.44)	(1.44)
Past attack * Migrants			-3.88**	
			(1.75)	
Past attack * Out-In migrants				-2.76
				(1.71)
Past attack * In-Out migrants				-5.86**
				(2.62)
Past attack * Yet-to-be migrants				-5.06***
				(1.70)
Only within nonmigrants		✓		
Control mean	1.89	2.28	1.89	1.89
Obs.	25,060	12,710	25,060	25,060

Notes: This table presents OLS estimates of equation 1, controlling for migration. The dependent variable is an indicator for whether the respondent joins a Congolese militia. The main explanatory variable is an indicator for whether the respondent's household has been attacked by foreign armed groups. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced an attack on household by foreign armed group before year t. P-value: *** 0.01, ** 0.05, * 0.10.

to migrate into the villages of interview. Remember, however, that villagers who migrated from outside of the villages of interview are not representative of Out-In migrants or In-Out migrants, and thus these are potentially a biased estimate of $T(B_2)$ and $T(A_2)$.

G.3 Calibration of real ATE

We conduct a simplified calibration exercise using the results above, assuming migration does not affect migration likelihood (more discussion in the next subsection).

Migration likelihood for each migrant p: On average, each migrant is observed for 16 years in the sample, and moves on average twice. We calibrate $p = \frac{1}{8}$.

Proportion of villagers in the villages of interview a: According to village chief survey, on average, there are 427 villagers in a village of interview in South Kivu. Consider the total population in South Kivu in 2015 to be 5,772,000, and apply the average number of villagers to all 239 villages in South Kivu and North Kivu. We calibrate $a = 1.8\%$.

Proportion of villagers who migrate at least once throughout the observation period π: Out of 1,537 respondents, 1,086 have never migrated outside of the sample once. We calibrate $\pi = 1 - 1086/1537 = 29\%$.[58]

Now we can apply these parameters to Equation 16. We have an imperfect estimate of $T(B_2)$ and $T(A_2)$ from Table G.2, Column 4 ($\hat{T}(B_2) = 4.14 - 2.76 = 1.38, \hat{T}(A_2) = 4.14 - 5.86 = -1.72$), and an estimate of $\hat{ATE}(A) = 2.55$.

Suppose all In-Out migrants do not react to past attack at all; that is, $T(A_2) = 0$. To make the real $ATE(A)$ zero, one needs the effect on Out-In migrants $T(B_2)$ to be at least

[58]This is not the perfect calibration because the denominator does not consider villagers who have emigrated. It is hard to observe emigrants who left the sample; the closest data we have is how many villagers emigrated from each village every year. On average, a village of interview in South Kivu sees 61 in-migrants and 64 out-migrants every year.

3.46. This is different from $\hat{T}(B_2) = 1.38$ with statistical significance (p-value 0.0205). To justify the empirical result, one needs to assume that the representative migrant from outside of the sample reacts more strongly to the past attack, but we only observe the left tail of the distribution.

Suppose all Out-In migrants do not react to past attack at all; that is, $T(B_2) = 0$. To make the real $ATE(A)$ zero, one needs the effect on In-Out migrants $T(A_2)$ to be at least -23.63. This is drastically different from $\hat{T}(A_2) = -1.72$ (p-value 0.0000). To justify the empirical result, villagers who tend to migrate outside of the villages of interview should be strongly discouraged from participation if they have experienced any past attack by foreign group, but we only observe the extremely right tail of the distribution.

G.4 Migration as an alternative channel

Selection bias aside, migration can confound the main result in the following way. Attack history in the past might affect the propensity of migration, and migration leads to a different participation rate through another channel that we did not discuss in the main text.

Table G.3 regresses different migration indicators on whether respondent experienced an attack by foreign armed group before. Column (1) suggests that a past attack leads to lower propensity of migration, although the effect is mainly driven by migration within the villages of interview (see Column (4)), which does not trigger the selection bias discussed above. Column (5) and (6) control for migration history in the past. Although migration within the villages of interview seems to be positively correlated with higher participation rate in a Congolese militia, the main effect of past attack remains largely unaffected, which

suggests that migration does not explain the effect of past attack on participation.

Table G.3: Migration as a Potential Channel on Participation

	Migration (%)				Participation (%)	
	All	In-Out	Out-In	In-In	Militia	Militia
	(1)	(2)	(3)	(4)	(5)	(6)
Past attack on HH by foreign armed group	-1.50*	0.10	-0.50	-1.02*	2.53**	2.41**
	(0.89)	(0.28)	(0.59)	(0.56)	(1.00)	(1.00)
Ever migrated before					1.54*	
					(0.81)	
Ever migrate from Out to In						0.14
						(1.68)
Ever migrate from In to Out						-0.28
						(1.00)
Ever migrate between villages of interview						4.74***
						(1.30)
Control mean	4.80	0.81	2.03	1.45	1.42	1.40
Obs.	25,060	25,060	25,060	25,060	25,060	25,060

Notes: This table presents OLS estimates of equation 1. Column (1) to (4) use different migration indicators as the dependent variable (migration in general, migration from a village of interview to an out-of-sample village, migration from an out-of-sample village to a village of interview, migration between the villages of interview). In Column (5) and (6), the dependent variable is an indicator for whether the respondent joins a Congolese militia. The main explanatory variable is an indicator for whether the respondent's household has been attacked by foreign armed groups. Column (5) and (6) control for whether respondent has migrated in the past. We include observations between 1995 and 2013 above age 15 at year t. All regressions include individual FE, village FE, year FE, age FE, and cluster two-way at the individual respondent and the village*year levels. Control mean is computed among observations where respondents never experienced an attack on household by foreign armed group before year t. P-value: *** 0.01, ** 0.05, * 0.10.