

# **Are Many Sets of Eyes Better Than One?**

Evaluating Multiple Databases of Armed Actors in Colombia

## **Online Appendix**

Javier Osorio

University of Arizona

# Appendix

<b>A1. Data set characteristics</b>	page 1.
- Definitions	page 2.
- Measurement set characteristics	page 3.
- Detailed list of actors	page 7.
<b>A2. Dimensionality-reduction Jaccard Algorithm</b>	page 8.
<b>A3. Jaccard Similarity Over Time</b>	page 10.
<b>A4. Armed Actors as Dependent Variable - full results</b>	page 11.
- Model specification and control variables	page 11.
- Descriptive statistics	page 13.
- Regression Results for Paramilitary Groups	page 14.
- Regression Results for Guerrilla Groups	page 16.
- Regression Results for FARC	page 18.
- Regression Results for ELN	page 20.
- Regression Results for AUC	page 22.
<b>A5. Armed Actors as Independent Variable - full results</b>	page 24.
- Model specification and control variables	page 24.
- Descriptive statistics	page 27.
- Summary of results	page 28.
- Regression Results for Paramilitary Groups	page 31.
- Regression Results for Guerrilla Groups	page 32.
- Regression Results for FARC	page 33.
- Regression Results for ELN	page 34.
- Regression Results for AUC	page 35.
<b>A6. Model Assessment</b>	page 36.
<b>A7. Relevance of Measurement Sets</b>	page 40.
<b>References</b>	page 41.

## A1. Data Set Characteristics

This section presents in detail the different characteristics of the seven databases analyzed in this study. These data bases are:

- **RC**: Paramilitary presence from Rutas del Conflicto (2019).
- **CL**: Paramilitary and guerrilla groups from Claudia López (2010).
- **IN**: Narcoparamilitary group activity from Indepaz (2019).
- **CD**: Paramilitary and guerrilla attacks from Centro de Estudios Sobre Desarrollo Económico (2019).
- **VI**: Paramilitary and guerrilla violent presence from ViPAA (Osorio et al. 2019).
- **RE**: Paramilitary and guerrilla attacks from Restrepo et al. (2004).
- **UC**: Paramilitary and guerrilla violent events from the UCDP (Sundberg and Melander 2013).

## Definitions

Table 1 below presents the definitions of the dependent variables used in each database.

Table 1: Definitions

Database	Object of study	Definition
INDEPAZ	Narcoparamilitary presence.	"Narcoparamilitaries "are defined as illegal armed organizations primarily focused on maximizing profit for their leaders and members. To foster their presence and expansion, [these groups] engage in public order and subversive activities, as long as they are related to their business" (Gonzalez Posso and Espitia 2017, p. 3). Narcoparamilitary activity "is defined as the group activity mentioned in a news article that indicates the presence of the group in a given date." [...] "This type of record indicates the presence of transit of the armed group." [...] "The type of activity indicates the characteristics of their presence" (Gonzalez Posso and Espitia 2017, p. 3)."
UCDP	Event of organized violence by paramilitaries or guerrilla groups.	"Event of organized violence is defined as "an individual incident (phenomenon) of lethal violence occurring at a given time and place" (Högbladh 2021, p. 4). An event is defined as "an incident where armed force was used by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date" (Högbladh 2021, p. 4).
Restrepo	Bellicose actions perpetrated by paramilitaries or insurgent groups.	"Political violence is defined as violent acts exerted as a means of political-social fighting aimed at maintaining, modifying, substituting, or destroying a model of state or society, or directed to the destruction or repression of a human group, organized or not, identified by social, political, occupation, ethnic, racial, religious, cultural or ideological traits" (Restrepo et al. 2004, p. 404). "In building our database, we included only those actions that, given the descriptions found in the publication, we considered as bellicose by following the criteria of motivation and group action" (Restrepo et al. 2004, p. 404). "We also believe that this classification procedure is more objective [...] as it is based on a clear characterization of 'warfare actions' as defined by the Geneva Convention as bellicose actions. It follows that there is a wide range of events including not only clashes, but also incursions, shootings, attacks on military targets, ambushes, attacks on pipelines and energy and communication infrastructure, etc." (Restrepo et al. 2004, p. 405). "For each event we include general descriptors: date and location [...] whether or not there was a clash and, if so, the groups involved, whether or not there was an attack and, if so, the type of attack and the group(s) responsible, finally killings and injuries, the intensity measures." (Restrepo et al. 2004, p. 405).
Claudia Lopez	Paramilitary or guerrilla presence.	Presence of guerrilla or paramilitary groups at the municipal level. The source does not provide specific definitions. See López (2010).
ViPAA	Violent presence of paramilitary or guerrilla groups involved in events of political violence.	"Political violence is defined as violence used as mean of socio-political struggle with the objective of maintaining, modifying, substituting, or destroying the State or a society, or with the objective to destroy or repress a human group on the basis of their social, political, group, ethnic, racial, religious, cultural or ideological identity, whether they are organized or not" (Centro de Investigación y Educación Popular 2017, p. 14). Political violence includes: human rights violations when the State is the perpetrator, and bellicious actions when perpetrated by insurgents or non-state armed actors such as paramilitaries or criminal groups."
Rutas del Conflicto	Paramilitary presence.	Paramilitary presence at the municipal level. The source does not provide a specific definition. See Rutas del Conflicto (2019).
CEDE	Paramilitary and guerrilla attacks.	Offensive activities by armed groups. The source does not provide a specific definition. See Acevedo and Bornacelly (2014)

## Measurement Set Characteristics

Table 2 presents in detail the characteristics of the different measurement sets analyzed in this study. The assessment considers the characteristics indicated below. These features are associated with a ranking that informs visualizations in Figure ?? in the body of the paper. The categories used to describe the measurement sets are:

- **Variable type:** Indicates whether the variables in each measurement set are dichotomous or count data. The ranking considers the variation in count data about the frequency of incidents more informative than dichotomous variables only indicating the presence of armed actors. Denoted as "Var" in Figure ??, this dimension considers the following categories and their respective values.
  - Dichotomous (1)
  - Count data (2)
- **Newspapers:** Indicates whether each measurement set relies on newspaper articles or not. If they do, it indicates whether the sources are international or local and whether it provides a specific list of sources or not. The assumption is that local sources are more reliable and specific than international ones. Denoted as "News" in Figure ??, this dimension includes the following levels:
  - No news (1)
  - International news - not specific (2)
  - International news - specific (3)
  - Local news - not specific (4)
  - Local news - specific (5)
- **NGOs:** Indicates whether each measurement set relies on information from NGOs or not. If they do, the values reflect if those are international or local NGOs, and whether it presents a detailed list of organizations or not. The assumption is that local NGOs are more reliable and specific than international organizations. Denoted as "NGOs" in Figure ??, this dimension includes the following levels:

- No NGOs (1)
  - International NGOs - not specific (2)
  - International NGOs - specific (3)
  - Local NGOs - not specific (4)
  - Local NGOs - specific (5)
- **Government agencies:** Indicates whether each measurement set relies on information from government agencies or not. If they do, the values reflect if the names of those agencies are listed specifically or not. The assumption is that using information from government agencies contributes to the content of the database. Denoted as "Gov" in Figure ??, this dimension includes the following levels:
    - No government (1)
    - Government - not specific (2)
    - Government - specific (3)
  - **Testimonies:** Indicates whether a measurement set relies on testimonies or interviews as part of the information sources. The assumption is that testimonies and interviews provide rich information that contributes to the content of the database. Denoted as "Test" in Figure ??, this dimension includes the following levels:
    - Does not rely on testimonies (1)
    - Relies on testimonies (2)
  - **Internal Quality Assessment (IQA):** Indicates whether or not the methodology describes the procedures put in place to internally assess the validity and quality of the data produced in each measurement set. The assumption is that a mix of manual and automated procedures yield a higher data quality and validity than only manual protocols or no IQA procedures at all. Denoted as "IQA" in Figure ??, this dimension includes the following levels:
    - No explicit IQA (1)
    - Manual IQA (2)

- Manual and automated IQA (3)
- **External Replicability:** Indicates whether or not it is possible to replicate the output of the data independently from the researchers. The assumption is that data that allows replicability is more transparent. Denoted as "Rep" in Figure ??, this dimension includes the following levels:
  - No replicability (1)
  - Replicability (2)

Table 2: Measurement Set Characteristics

Measurement set	Variable type	International news agencies	International NGOs	National newspapers	Local newspapers	Local NGOs	Local HR NGOs	Government agencies	Internal quality assessment	External replicability
INDEPAZ	Dichotomous	No	No	Yes	Yes	Not specific	Yes	Yes	No	No
UCDP	Count	Yes	Specific list	No	No	Not specific	No	No	Yes	No
Restrepo	Count	No	No	Yes	Yes	Specific list	Yes	No	Yes	No
Claudia Lopez	Dichotomous	No	No	No	No	Specific list	Yes	Yes	No	No
VIPAA	Count	No	No	Yes	Yes	Not specific	Yes	Yes	Yes	Yes
Rutas del Conflicto	Dichotomous	No	No	Yes	Yes	Not specific	No	No	No	No
CEDE	Count		No	No	No	No	No	Yes	No	No



Table 3: Detail of Actors Considered by each Measurement Set

Source	Type of actor	Num. groups	Armed Groups
CEDE	Guerrilla	3	National Liberation Army (ELN), People’s Revolutionary Army (ERP), Revolutionary Armed Forces of Colombia (FARC)
	Paramilitaries	1	United Self-defense forces of Colombia (AUC)
Lopez	Guerrilla	1	Generic
	Paramilitaries	1	Generic
Indepaz	Guerrilla	0	NA
	Paramilitaries	22	Aguilas Negras, Autodefensas Campesinas de Tolima, Bloque Meta, Clan del Golfo, Cordillera, EPL Los Pelusos, FIAC, La Constru, La empresa, Libertadores del Vichada, Llaneros, Los Botalones, Los del ejido, Los Pacheca, Los Policarpa, Los Rastrojos, Los Soto, Narcoparamilitar, Otros , Puntilleros, Renacer, Urabeños
Rutas del Conflicto	Guerrilla	5	Disidencia de las FARC - Columna Ricardo Franco, ELN, FARC, Grupo Guerrillero no identificado, Movimiento Armado Quintin Lame
	Paramilitaries	36	Autodefensas de Cordoba y Uraba, Autodefensas de Meta y Vichada, Carranceros, Autodefensas de Ortega, Autodefensas del Casanare, Autodefensas del Magdalena Medio, Autodefensas del Sur del Cesar, Autodefensas de Los Rojos o el Palmor, Autodefensas de Puerto Boyaca, Autodefensas de Puerto Boyaca, Autodefensas de Santander y del Sur del Cesar, Grupo Paramilitar Andaquies Caqueta, Grupo Paramilitar no identificado , Masetos, Narcoparamilitar, Paramilitares Bananero, Paramilitares Cacique Nutibara, Paramilitares Calima, Paramilitares Catatumbo, Paramilitares Centauros, Paramilitares de Hernan Giraldo, Paramilitares del Bloque, Paramilitares del Fidel Castaño, Paramilitares del Norte del Valle, Paramilitares Elmer Cardenas, Paramilitares Heroes de Granada, Paramilitares Heroes de Tolova, Paramilitares Libertadores del Sur, Paramilitares Metro, Paramilitares Mineros, Paramilitares Montes de Maria, Paramilitares Noroccidente Antioqueño, Paramilitares Norte, Paramilitares Pacifico, Paramilitares Suroeste Antioqueño, Paramilitares Tolima, Paramilitares Vencedres de Arauca
ViPPA	Guerrilla	8	19th of April Movement (M-19), Ernesto Rojas Commandos (CER), Generic mentions of insurgents, National Liberation Army (ELN), People’s Revolutionary Army (ERP), Popular Liberation Army (EPL), Revolutionary Armed Forces of Colombia (FARC), Simon Bolivar Guerrilla Coordinating Board (CGSB)
	Paramilitaries	7	Generic mentions of paramilitaries , Independent paramilitary groups, Peasant Self-defense forces of Cordoba and Uraba (ACCU), Popular Revolutionary Anti-Terrorist Army of Colombia (ERPAC), Social Cleansing, United Gaitan Self-defense forces (AGU), United Self-defense forces of Colombia (AUC)
UCDP	Guerrilla	2	National Liberation Army (ELN), Revolutionary Armed Forces of Colombia (FARC)
	Paramilitaries	1	United Self-defense forces of Colombia (AUC)

## A2. Dimensionality-reduction Jaccard Similarity Algorithm

Calculating Jaccard similarity in the presence of missing data can distort the similarity assessment. As discussed in the body of the article, missing data is a considerable problem in many measurement sets considered in the study, thus representing a challenge for assessing their similarity. To address this challenge, the analysis proposes a dimensionality-reduction algorithm to calculate Jaccard Similarity in the presence of missingness.

Algorithm 1 considers time ( $t$ ), location ( $i$ ), and measurement sets ( $M$ ) as relevant dimensions to calculate the Jaccard similarity. The algorithm uses the available  $M$  and  $i$  in a given  $t$  to generate vectors as data subsets ( $s$ ) and calculates their local Jaccard scores ( $J_{ts}$ ) as indicated in equation ?? in the body of the manuscript.

If the data comprises subsets of different dimensions in a given year, the algorithm calculates the weighted average of the local Jaccard using the proportion of observations in each subset ( $w_s = n_s/N$ ), such that  $J_{tw} = (\sum_s^S J_{ts}w_s)(1/S)$ . If there are no subsets in a given year, the algorithm directly calculates the local Jaccard ( $J_t$ ) as in equation ?. After computing local Jaccard scores for each  $t$ , the algorithm generates an aggregated Jaccard score ( $\bar{J}$ ) averaging local similarity scores over time ( $t = 1, \dots, T$ ) as expressed in equation ?? in the manuscript. In this way, the algorithm only uses measures overlapping at different points in time without being affected by missing data.

---

**Algorithm 1** Dimensionality-Reduction Jaccard Calculator

---

**Input:** Comparison set  $C = \{M_1, M_2, \dots, M_N\}$

```
1: for each actor  $e$  in  $C$  do
2:   for each  $t$  in year do
3:     Subset data by  $t$  and keep all locations  $i$ 
4:     Eliminate empty measurement sets  $M$  (columns with all NAs)
5:     Consider each  $M$  as a data vector
6:     if any available  $M$  has missing cells then
7:       Create data subsets ( $s$ ) using available dimensions of  $i$  and  $M$ .
8:       for each  $s$  do
9:         Assign  $x \leftarrow \mathbb{Z}$  as the number of measurement sets  $M$  available
10:        Calculate the weight  $w_s$  of each  $s$ 
11:        if  $x > 1$  then
12:          Calculate local Jaccard Index  $J_t$  for each  $s$  in time  $t$  using available  $M$ 
13:        else
14:          Assign  $J_t = 0$  as there is no other  $M$ 
15:        end if
16:      end for
17:      For all data subsets that exist in a given year,  $\forall s, \exists t$ ,
18:      calculate the weighted average of Local Jaccard scores of  $\forall s$ 
19:    else
20:      There are no missing cells in  $M$ 
21:      Assign  $x \leftarrow \mathbb{Z}$  as the number of measurement sets  $M$  available
22:      if  $x > 1$  then
23:        Calculate the Local Jaccard  $J_t$  in time  $t$  using available  $M$ 
24:      else
25:        Assign  $J_t = 0$  since there is no other  $M$  to compare to
26:      end if
27:    end if
28:    if Local Jaccard Index by  $t$  then
29:      Save similarity scores and plot  $J_t$  for  $\forall t, \exists T$ 
30:    else
31:      For all measurement sets that exist in the comparison set,  $\forall m, \exists C$ ,
32:      Calculate the average Jaccard  $\bar{J}$ 
33:      Save similarity scores and plot  $\bar{J}$ 
34:    end if
35:  end for
36: end for
```

---

### A3. Jaccard Similarity Over Time

Figure 1 reports the temporal variation of Local Jaccard scores by actor type (Panel a) and at the group level (Panel b). Temporal oscillations in Panel (a) show that guerrilla and paramilitary similarity steadily increases until reaching an inflection point around 2000; then both series decline with paramilitaries experiencing a sharper drop; finally their similarity increases around 2010.

The time series in Panel (d) display a plateau of AUC similarity between 1993 and 2002, after which the level of agreement sharply drops, just to bounce up in 2011. The temporal fluctuation of ELN similarity follows a similar pattern, but at a lower level of agreement. Finally, FARC displays an inverse U shape reaching a similarity peak around 2002 and then experiencing a sustained decline.

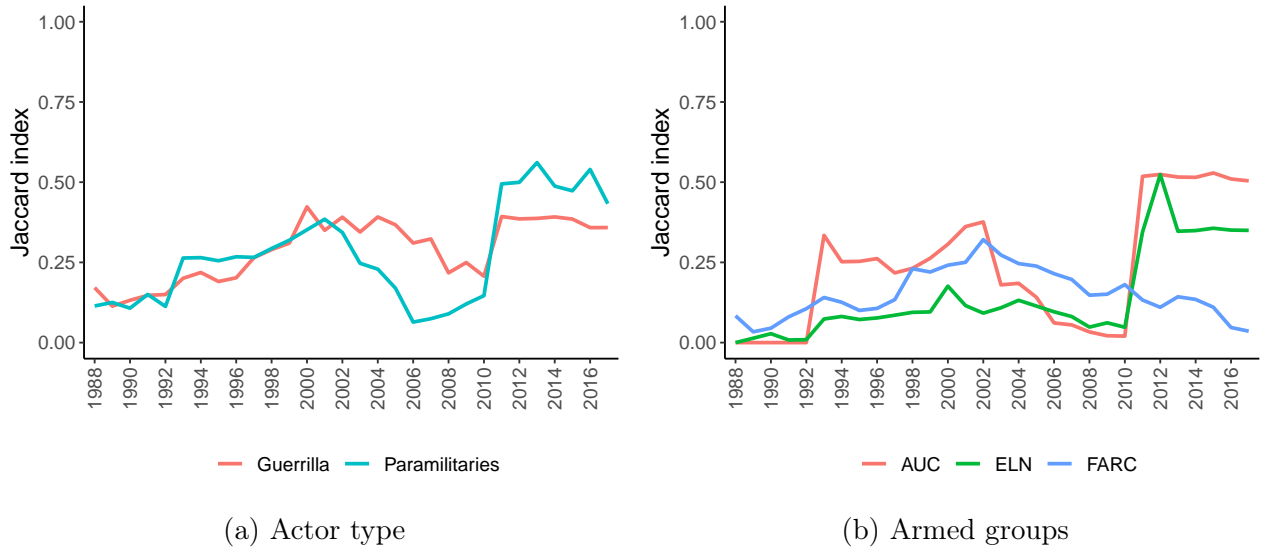


Figure 1: Local Jaccard Similarity Over Time

#### A4. Armed Actors and Dependent Variable - full results

##### Model Specification and Control Variables in Statistical Analysis

Dube and Vargas (2013) use a two-stages least squares (2SLS) model to estimate the differential effects of oil and coffee price shocks on armed actors in Colombia. According to their study, oil shocks are expected to increase armed actor presence by promoting predatory behavior. In contrast, coffee shocks are expected to reduce armed actor presence through a labor substitution effect diverting combatants away from fighting and into legal coffee production.

The first stage of the 2SLS model considers three plausibly exogenous instruments consisting on the coffee export volume of the other three leading coffee exporting countries Brazil, Vietnam, and Indonesia, labeled here as *Top3* variable, interacted with local weather conditions of *rainfall*, *temperature* in Colombia, and the product of those two (*rainfall*  $\times$  *temperature*). For simplicity, let's label rainfall as (R) and temperature as (T). The intuition is that the exogenous international shock prices of the Top 3 coffee producers interacted with local environmental conditions are likely to generate a plausibly exogenous coffee shocks measures as the variation on local coffee production (*Cof<sub>it</sub>*) in Colombia given local coffee prices in the domestic market (*CP<sub>t</sub>*). Together, *Cof<sub>it</sub>* and *CP<sub>t</sub>* constitute the internal coffee production index as coffee shocks.

Based on this intuition, the first stage of the model has the following specification:

$$\begin{aligned} (Cof_{it} \times CP_t) = & \omega_1(Top3 \times R) + \omega_1(Top3 \times T) + \omega_1(Top3 \times R \times T) + \\ & \lambda(Oil_{rt} \times OP_t) + \gamma Coca_{irt} + \phi X_{irt} + \alpha_i + \beta_t + \delta_{rt} + \epsilon_{irt} \end{aligned} \quad (1)$$

where (*Cof<sub>it</sub>*  $\times$  *CP<sub>t</sub>*) is the endogenous coffee shock measures as the internal coffee index comprising the interaction of local coffee production (*Cof<sub>it</sub>*) with domestic coffee prices (*CP<sub>t</sub>*) in a given municipality *i* and year *t*; the instruments are the interaction of international

coffee production from Brazil, Vietnam, and Indonesia interacted with rain ( $Top3 \times R$ ), with temperature ( $Top3 \times T$ ), and with the product of rain and temperature ( $Top3 \times R \times T$ );  $Oil_{rt}$  indicates municipal oil production interacted with international oil prices  $OP_t$ ;  $Cof_{it}$  represents municipal hectares of coffee cultivation interacted with domestic coffee prices  $CP_t$  derived from the first stage;  $Coca_{irt}$  indicates coca cultivation as a dummy variable;  $X_{irt}$  represents time-varying controls including: municipal population (logged) and region-year ( $\delta_{rt}$ ); the model also includes fixed effects by municipality ( $\alpha_i$ ) and year ( $\beta_t$ ), and the error terms ( $\epsilon_{irt}$ ).

The second-stage model specification has the following functional form:

$$y_{irt} = \lambda(Oil_{rt} \times OP_t) + \rho(\widehat{Cof_{it}} \times CP_t) + \gamma Cof_{irt} + \phi X_{irt} + \alpha_i + \beta_t + \delta_{rt} + \epsilon_{irt} \quad (2)$$

where  $y$  indicates the presence of armed actor type or specific group reported in each measurement in municipality  $i$ , region  $r$ , and year  $t$ .  $Oil_{rt}$  indicates municipal oil production interacted with international oil prices  $OP_t$ ;  $Cof_{it}$  represents predicted Coffee shocks as the variation of municipal hectares of coffee cultivation interacted with domestic coffee prices  $CP_t$  derived from the exogenous instruments in the first stage;  $Coca_{irt}$  indicates coca cultivation as a dummy variable;  $X_{irt}$  represents time-varying controls including: municipal population (logged) and region-year ( $\delta_{rt}$ ); the model also includes fixed effects by municipality ( $\alpha_i$ ) and year ( $\beta_t$ ), and the error terms ( $\epsilon_{irt}$ ).

For further details about the model, variables, operationalization, and sources, see Dube and Vargas (2013).

## Descriptive Statistics

Table 4: Descriptive statistics

	Mean	SD	Min	Max	N
Paramilitaries(VI)	0.04	0.20	0.00	1.00	17,532
Paramilitaries(CD)	0.85	0.36	0.00	1.00	12,662
Paramilitaries(CL)	0.30	0.46	0.00	1.00	8,766
Paramilitaries(RE)	0.38	0.48	0.00	1.00	17,712
Paramilitaries(UC)	0.03	0.16	0.00	1.00	16,558
Paramilitaries(All)	0.75	0.43	0.00	1.00	17,784
Guerrilla(VI)	0.19	0.40	0.00	1.00	17,532
Guerrilla(CD)	0.47	0.50	0.00	1.00	12,662
Guerrilla(CL)	0.47	0.50	0.00	1.00	8,766
Guerrilla(RE)	0.41	0.49	0.00	1.00	17,712
Guerrilla(UC)	0.08	0.27	0.00	1.00	16,558
Guerrilla(All)	0.50	0.50	0.00	1.00	17,784
FARC(VI)	0.14	0.34	0.00	1.00	17,532
FARC(CD)	0.39	0.49	0.00	1.00	12,662
FARC(RC)	0.00	0.06	0.00	1.00	13,636
FARC(RE)	0.30	0.46	0.00	1.00	17,712
FARC(UC)	0.06	0.24	0.00	1.00	16,558
FARC(All)	0.34	0.47	0.00	1.00	17,820
ELN(VI)	0.07	0.26	0.00	1.00	17,532
ELN(CD)	0.21	0.41	0.00	1.00	12,662
ELN(RC)	0.00	0.04	0.00	1.00	4,870
ELN(UC)	0.02	0.14	0.00	1.00	16,558
ELN(All)	0.18	0.38	0.00	1.00	17,558
AUC(VI)	0.07	0.25	0.00	1.00	17,532
AUC(CD)	0.85	0.36	0.00	1.00	12,662
AUC(RC)	0.05	0.21	0.00	1.00	10,714
AUC(RE)	0.18	0.38	0.00	1.00	8,856
AUC(UC)	0.03	0.16	0.00	1.00	16,558
AUC(All)	0.62	0.49	0.00	1.00	17,666
Coffee shocks	0.53	1.08	0.00	10.43	17,604
Oil shocks	0.01	0.22	0.00	7.83	17,964
Rainfall x T3CP	6,482.27	3,486.83	493.87	35,372.03	17,964
Temperature x T3CP	73.43	17.94	12.04	111.11	17,964
Rainfall x Temperature x T3CP	142,888.15	93,272.56	11,676.94	944,433.31	17,964
Coca 94 x year	100.03	435.55	0.00	2,005.00	17,964
Population, (log)	-4.35	0.96	-8.83	-1.36	17,964
year 1989	0.06	0.23	0.00	1.00	17,964
year 1990	0.06	0.23	0.00	1.00	17,964
year 1991	0.06	0.23	0.00	1.00	17,964
year 1992	0.06	0.23	0.00	1.00	17,964
year 1993	0.06	0.23	0.00	1.00	17,964
year 1994	0.06	0.23	0.00	1.00	17,964
year 1995	0.06	0.23	0.00	1.00	17,964
year 1996	0.06	0.23	0.00	1.00	17,964
year 1997	0.06	0.23	0.00	1.00	17,964
year 1998	0.06	0.23	0.00	1.00	17,964
year 1999	0.06	0.23	0.00	1.00	17,964
year 2000	0.06	0.23	0.00	1.00	17,964
year 2001	0.06	0.23	0.00	1.00	17,964
year 2002	0.06	0.23	0.00	1.00	17,964
year 2003	0.06	0.23	0.00	1.00	17,964
year 2004	0.06	0.23	0.00	1.00	17,964
year 2005	0.06	0.23	0.00	1.00	17,964
region 2	0.15	0.35	0.00	1.00	17,964
region 3	0.10	0.30	0.00	1.00	17,964
region 4	0.15	0.35	0.00	1.00	17,964
(region 2)*year	294.07	707.58	0.00	2,005.00	17,964
(region 3)*year	202.05	602.16	0.00	2,005.00	17,964
(region 4)*year	292.07	705.58	0.00	2,005.00	17,964
Note: T3CP = Top coffee producers (log), which includes Brazil, Vietnam, and Indonesia.					

## Regression Results for Paramilitary Groups

Table 5 presents the results of the First-stage of the Dube and Vargas (2013) replication using coffee shocks as the endogenous variable and the three different instruments of Top 3 coffee producing countries interacted with rain, temperature, and the interaction of rain and temperature. In line with the expectations, all three different instruments are statistically significant. The models show different sample sizes because they are estimated based on the available data from the measurement sets under study. Despite the different sample sizes, the F-statistic is above the instrument validity threshold of  $F\text{-stat} = 10$  proposed by Angrist and Pischke (2009), but all of them are below the recent  $F\text{-stat} = 104.7$  instrument validity threshold of Lee, David et al. (2020).

Table 5: First Stage - Paramilitaries

DV: Coffee shocks	(1)	(2)	(3)	(4)	(5)	(6)
Top3 $\times$ rain	-0.002*** (0.000)	-0.001*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
Top3 $\times$ temp	-0.135*** (0.000)	-0.107*** (0.000)	-0.208*** (0.000)	-0.135*** (0.000)	-0.128*** (0.000)	-0.135*** (0.000)
Top3 $\times$ rain $\times$ temp	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Oil shocks	0.072 (0.127)	0.099 (0.118)	0.020 (0.761)	0.075 (0.106)	0.070 (0.099)	0.076 (0.098)
Coca	-0.010 (0.342)	-0.011 (0.289)	-0.020 (0.291)	-0.010 (0.355)	-0.010 (0.341)	-0.010 (0.360)
Population	0.100 (0.604)	0.080 (0.591)	0.210 (0.494)	0.095 (0.615)	0.101 (0.580)	0.091 (0.626)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
AP-F-statistic	15.60	15.74	15.48	15.64	15.53	15.75
AP-p-value	0.000	0.000	0.000	0.000	0.000	0.000
J-statistic	3.76	8.30	0.64	1.93	3.35	2.90
J-p-value	0.152	0.015	0.726	0.381	0.187	0.234
Observations	17190	12415	8595	17370	16235	17442
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						

However, results of the Sargan's J test in Model (2) of the First-stage cast doubt about the validity of the instruments. The J test is used to assess the over-identifying restrictions in a statistical model where the null hypothesis considers that the over-identifying restrictions



are valid. However, if the J-statistic is statistically significant and rejects the null hypothesis, then the instruments are weak and therefore cannot be trusted as valid.

Table 6 presents the results of the second-stage model for paramilitary groups using the different measurement sets as the dependent variable. The analysis shows limited support for the expectations of Dube and Vargas (2013). Although the coefficients of coffee shocks show a consistent negative direction, the CL and RE estimates fail to reach statistical significance in Models (3) and (4). Results also provide limited support for the positive effect of oil shocks on the violent presence of armed groups. Oil shocks present the expected positive relationship for VI in Model (1), CL in Model (3), and UC in Model (3). However, oil shocks do not reach statistical significance for CD in Model (2) and All in Model (6), and the oil estimate reports a negative coefficient for RE in Model (4) without reaching statistical significance.

Table 6: Second Stage - Paramilitaries

DV: Paramilitaries	(1) VI	(2) CD	(3) CL	(4) RE	(5) UC	(6) All
Coffee shocks	-0.103** (0.002)	-0.297** (0.002)	-0.084 (0.125)	-0.052 (0.220)	-0.082** (0.002)	-0.102* (0.032)
Oil shocks	0.205** (0.003)	0.357 (0.061)	0.294* (0.030)	-0.107 (0.347)	0.280*** (0.000)	0.284 (0.087)
Population	0.079** (0.007)	0.179* (0.050)	0.086 (0.346)	-0.042 (0.421)	0.019 (0.315)	-0.124* (0.014)
Coca	-0.000 (0.863)	0.012 (0.089)	-0.001 (0.924)	0.006 (0.348)	0.003 (0.147)	0.005 (0.280)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,190	12,415	8,595	17,370	16,235	17,442
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						
CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA;						
UC=UCDP; RE=Restrepo						

In general, the replication analysis provides limited support to the findings presented by Dube and Vargas (2013) when using a variety of measurement sets of paramilitary groups. As the J-statistic in Table 6 indicates, the instruments can be considered valid in most models, with the exception of Model (2). However, the expected negative effect of coffee shocks and the positive effect of oil shocks only find support when using the VI and UC measures of paramilitaries in Models (1) and (5), respectively.

## Regression Results for Guerrilla Groups

Table 7 presents the results of the First-stage of the Dube and Vargas (2013) replication for guerrilla groups at the actor-type level using coffee shocks as the endogenous variable and the three instruments mentioned in the previous model. In line with the expectations, all three different instruments are statistically significant. The models show different sample sizes because they are estimated based on the available data from the measurement sets under study. Despite the different sample sizes, the F-statistic is above the instrument validity threshold of  $F\text{-stat} = 10$  proposed by Angrist and Pischke (2009), but all of them are below the recent  $F\text{-stat} = 104.7$  instrument validity threshold of Lee, David et al. (2020). The p-values of the Sargan J tests across models fail to reject the null hypothesis for the over-identifying assumptions, thus suggesting that the instruments are valid.

Table 7: First Stage - Guerrilla

DV: Coffee shocks	(1)	(2)	(3)	(4)	(5)	(6)
Top3 $\times$ rain	-0.002*** (0.000)	-0.001*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
Top3 $\times$ temp	-0.135*** (0.000)	-0.107*** (0.000)	-0.208*** (0.000)	-0.135*** (0.000)	-0.128*** (0.000)	-0.135*** (0.000)
Top3 $\times$ rain $\times$ temp	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Oil shocks	0.072 (0.127)	0.099 (0.118)	0.020 (0.761)	0.075 (0.106)	0.070 (0.099)	0.076 (0.098)
Coca	-0.010 (0.342)	-0.011 (0.289)	-0.020 (0.291)	-0.010 (0.355)	-0.010 (0.341)	-0.010 (0.360)
Population	0.100 (0.604)	0.080 (0.591)	0.210 (0.494)	0.095 (0.615)	0.101 (0.580)	0.091 (0.626)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
AP-F-statistic	15.60	15.74	15.48	15.64	15.53	15.75
AP-p-value	0.000	0.000	0.000	0.000	0.000	0.000
J-statistic	4.42	0.08	2.78	0.95	2.42	0.13
J-p-value	0.109	0.958	0.249	0.621	0.298	0.937
Observations	17190	12415	8595	17370	16235	17442
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						

Table 8 shows the results of the second-stage model for guerrilla organizations across the different measurement sets. The analysis shows limited support for the expectations of Dube and Vargas (2013). Although the coefficients of coffee shocks show a consistent negative

direction, estimates from the CD, CL, and RE regressions fail to reach statistical significance in Models (2), (3), and (4). In contrast to the expectation of a positive effect of oil shocks on the violent presence of armed groups, results of the second-stage largely contradict this expectation. The coefficient of Oil shocks only presents the expected positive relationship for VI in Model (1), but it is not statistically significant. All the other coefficients of oil shocks present a negative effect, which reaches statistical significance for CD in Model (2), UC in Model (5), and All in Model (6). The other two models (3 and 4) present a negative sign but do not reach statistical significance.

Table 8: Second Stage - Guerrilla

DV: Guerrilla	(1) VI	(2) CD	(3) CL	(4) RE	(5) UC	(6) All
Coffee shocks	-0.336*** (0.000)	-0.051 (0.520)	-0.031 (0.579)	-0.093 (0.123)	-0.143** (0.002)	-0.145* (0.017)
Oil shocks	0.080 (0.282)	-0.175* (0.036)	-0.022 (0.576)	-0.039 (0.331)	-0.185*** (0.000)	-0.098* (0.035)
Population	0.136 (0.087)	-0.091 (0.313)	0.029 (0.799)	-0.138* (0.022)	-0.004 (0.930)	-0.048 (0.376)
Coca	0.015** (0.008)	0.015 (0.068)	-0.004 (0.557)	0.017** (0.002)	0.012*** (0.000)	0.009 (0.086)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,190	12,415	8,595	17,370	16,235	17,442
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01. CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA; UC=UCDP; RE=Restrepo						

Overall, the statistical assessment largely fails to provide support to the argument advanced by Dube and Vargas (2013) when analyzing guerrilla organizations at the actor-type level. Despite the validity of the instruments in the first stage (see J-stat in Table 8), estimates of coffee and oil shocks present contradictory or uncertain estimates.

## Regression Results for FARC

Table 9 presents the results of the First-stage of the Dube and Vargas (2013) replication for the FARC insurgency using coffee shocks as the endogenous variable and the three different instruments. In line with the expectations, all three different instruments are statistically significant. The models show different sample sizes because they are estimated based on the available data from the measurement sets under study. Despite the different sample sizes, the F-statistic is above the instrument validity threshold of  $F\text{-stat} = 10$  proposed by Angrist and Pischke (2009), but all of them are below the recent  $F\text{-stat} = 104.7$  instrument validity threshold of Lee, David et al. (2020). However, the Sargan's J test for the over-identifying restrictions of the First-stage reject the null hypothesis in Models (1) and (3), which raises concerns about the validity of the instruments.

Table 9: First Stage - FARC

DV: Coffee shocks	(1)	(2)	(3)	(4)	(5)	(6)
Top3 $\times$ rain	-0.002*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
Top3 $\times$ temp	-0.135*** (0.000)	-0.107*** (0.000)	-0.151*** (0.000)	-0.135*** (0.000)	-0.128*** (0.000)	-0.135*** (0.000)
Top3 $\times$ rain $\times$ temp	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Oil shocks	0.072 (0.127)	0.099 (0.118)	0.070 (0.238)	0.075 (0.106)	0.070 (0.099)	0.077 (0.096)
Coca	-0.010 (0.342)	-0.011 (0.289)	-0.012 (0.343)	-0.010 (0.355)	-0.010 (0.341)	-0.009 (0.362)
Population	0.100 (0.604)	0.080 (0.591)	0.131 (0.565)	0.095 (0.615)	0.101 (0.580)	0.092 (0.623)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
AP-F-statistic	15.60	15.74	15.41	15.64	15.53	15.74
AP-p-value	0.000	0.000	0.000	0.000	0.000	0.000
J-statistic	5.15	0.44	6.59	1.56	2.06	0.59
J-p-value	0.076	0.802	0.037	0.457	0.357	0.746
Observations	17190	12415	13370	17370	16235	17478
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						

Table 10 reports the results of the second-stage model for the FARC insurgency. The analysis fails to find support for the Dube and Vargas (2013) study. Although the coffee shocks estimates show a consistent negative sign across models, estimates from the CD, RC,

and RE analysis do not reach statistical significance in Models (2), (3), and (4). In contrast to the expectation of a positive effect of oil shocks on the violent presence of armed groups, results of the second-stage largely contradict this expectation. The expected positive effect of Oil shocks on the violent presence of armed actors only reports a positive sign with statistical significance for VI in Model (1) and RC in Model (3). In contrast, estimates from CD, UC, and All present a negative effect and reach statistical significance, thus contradicting the expectation. Model (4) also presents a negative sign for FARC insurgents measured by RE, but it fails to reach statistical significance.

Table 10: Second Stage - FARC

DV: FARC	(1) VI	(2) CD	(3) RC	(4) RE	(5) UC	(6) All
Coffee shocks	-0.263*** (0.000)	-0.093 (0.126)	-0.005 (0.252)	-0.107 (0.141)	-0.091** (0.003)	-0.132* (0.049)
Oil shocks	0.228** (0.006)	-0.197* (0.012)	0.157*** (0.000)	-0.045 (0.248)	-0.191*** (0.000)	-0.105* (0.036)
Population	0.115 (0.086)	-0.067 (0.399)	-0.006 (0.451)	-0.104 (0.051)	0.004 (0.888)	-0.039 (0.489)
Coca	0.012* (0.016)	0.017* (0.019)	0.001 (0.340)	0.018*** (0.001)	0.010* (0.012)	0.013* (0.014)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,190	12,415	13,370	17,370	16,235	17,478
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						
CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA;						
UC=UCDP; RE=Restrepo						

The overall assessment fails to support the argument of Dube and Vargas (2013) when analyzing different measures of the FARC insurgency. As indicated in the J-test in the first-stage Table 9, the proposed instruments are valid in Models (2), (4), (5), and (6). However, none of these models support the expectations from Dube and Vargas (2013) as they present contradictory or uncertain results in the second stage.

## Regression Results for ELN

Table 11 presents the results of the First-stage of the Dube and Vargas (2013) replication using coffee shocks as the endogenous variable and the three different instruments of Top 3 coffee producing countries interacted with rain, temperature, and the interaction of rain and temperature. In line with the expectations, all three different instruments are statistically significant. The models show different sample sizes because they are estimated based on the available data from the measurement sets under study. Despite the different sample sizes, the F-statistic is above the instrument validity threshold of  $F\text{-stat} = 10$  proposed by Angrist and Pischke (2009), but all of them are below the recent  $F\text{-stat} = 104.7$  instrument validity threshold of Lee, David et al. (2020). However, results of the Sargan's J test cast doubt about the validity of the instruments in Models (2) and (5).

Table 11: First Stage - ELN

DV: Coffee shocks	(1)	(2)	(3)	(4)	(5)
Top3 $\times$ rain	-0.002*** (0.000)	-0.001*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
Top3 $\times$ temp	-0.135*** (0.000)	-0.107*** (0.000)	-0.253*** (0.000)	-0.128*** (0.000)	-0.135*** (0.000)
Top3 $\times$ rain $\times$ temp	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Oil shocks	0.072 (0.127)	0.099 (0.118)	0.040 (0.641)	0.070 (0.099)	0.072 (0.127)
Coca	-0.010 (0.342)	-0.011 (0.289)	-0.030 (0.337)	-0.010 (0.341)	-0.010 (0.340)
Population	0.100 (0.604)	0.080 (0.591)	0.294 (0.509)	0.101 (0.580)	0.100 (0.605)
Municipality FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
AP-F-statistic	15.60	15.74	15.72	15.53	15.61
AP-p-value	0.000	0.000	0.000	0.000	0.000
J-statistic	4.24	5.65	0.98	2.42	4.86
J-p-value	0.120	0.059	0.612	0.298	0.088
Observations	17190	12415	4775	16235	17216
Note: *p<0.1; **p<0.05; ***p<0.01.					

Table 12 presents the results of the second-stage model for the ELN guerrilla organization. The analysis finds limited support for the Dube and Vargas (2013) study. Although coffee shocks show a consistent negative sign across models, estimates from RC fail to reach

statistical significance in Model (3). The expectation related to the positive association between oil shocks and the violent presence of armed actors only finds support for the ELN measure by VI data in Model (1). The coefficients Models (2), (4) and (5) present a positive sign for the CD, UC, and All measures of the ELN insurgency, but do not reach statistical significance. In contrast, Model (3) presents a negative sign with high statistical significance for the RC measure.

Table 12: Second Stage - ELN

DV: ELN	(1) VI	(2) CD	(3) RC	(4) UC	(5) All
Coffee shocks	-0.139* (0.024)	-0.139* (0.022)	-0.000 (0.937)	-0.049* (0.031)	-0.188* (0.013)
Oil shocks	0.226** (0.007)	0.039 (0.487)	-0.529*** (0.000)	0.013 (0.613)	-0.013 (0.790)
Population	0.060 (0.173)	-0.029 (0.721)	0.023* (0.045)	-0.009 (0.690)	0.077 (0.244)
Coca	0.003 (0.177)	0.012* (0.023)	-0.002 (0.287)	0.002 (0.246)	0.008 (0.125)
Municipality FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes
Observations	17,190	12,415	4,775	16,235	17,216
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01. CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conficto; VI=ViPAA; UC=UCDP; RE=Restrepo					

Overall, the results provide limited support for the Dube and Vargas (2013) argument when analyzing the ELN guerrilla organization. As indicated in the first-stage post-estimation assessment in Table 11, the instruments are valid only in Models (1), (3), and (5), but only the second-stage results of Model (1) support the Dube and Vargas (2013) argument for for VI, while RC and All provide contradictory or uncertain estimates.

## Regression Results for AUC

Table 13 presents the results of the First-stage of the Dube and Vargas (2013) replication using coffee shocks as the endogenous variable and the three different instruments of Top 3 coffee producing countries interacted with rain, temperature, and the interaction of rain and temperature. In line with the expectations, all three different instruments are statistically significant in the First-stage analysis. The models show different sample sizes because they are estimated based on the available data from the measurement sets under study. Despite the different sample sizes, the F-statistic is above the instrument validity threshold of F-statistic = 10 proposed by Angrist and Pischke (2009), but all of them are below the instrument validity threshold of F-stat = 104.7 by Lee, David et al. (2020). However, results of the Sargan's J test reject the null hypothesis for all models with the exception of Model (5), thus casting doubt about the validity of the instruments.

Table 13: First Stage - AUC

DV: Coffee shocks	(1)	(2)	(3)	(4)	(5)	(6)
Top3 $\times$ rain	-0.002*** (0.000)	-0.001*** (0.000)	-0.002*** (0.000)	-0.003*** (0.000)	-0.002*** (0.000)	-0.002*** (0.000)
Top3 $\times$ temp	-0.135*** (0.000)	-0.107*** (0.000)	-0.121*** (0.000)	-0.208*** (0.000)	-0.128*** (0.000)	-0.135*** (0.000)
Top3 $\times$ rain $\times$ temp	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)	0.000*** (0.000)
Oil shocks	0.072 (0.127)	0.099 (0.118)	0.063 (0.334)	0.026 (0.671)	0.070 (0.099)	0.069 (0.133)
Coca	-0.010 (0.342)	-0.011 (0.289)	-0.013 (0.316)	-0.019 (0.306)	-0.010 (0.341)	-0.011 (0.331)
Population	0.100 (0.604)	0.080 (0.591)	0.137 (0.492)	0.205 (0.497)	0.101 (0.580)	0.100 (0.602)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
AP-F-statistic	15.60	15.74	15.52	15.51	15.53	15.62
AP-p-value	0.000	0.000	0.000	0.000	0.000	0.000
J-statistic	9.14	8.30	8.07	9.13	3.35	10.94
J-p-value	0.010	0.015	0.017	0.010	0.187	0.004
Observations	17190	12415	10505	8685	16235	17324
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						

Table 14 presents the results of the second-stage model for the AUC paramilitary organization. In general, the analysis finds moderate support for the Dube and Vargas (2013)



study. All the coffee shocks coefficients report negative effect across models, but the point estimate from RC in Model (3) fails to reach statistical significance. In line with the expected positive sing of oil shocks on the violent presence of the AUC, the second-stage results report a positive effect across Table 14. With the exception of the CD measure of AUC paramilitaries in Model (2), the coefficient of oil shocks reach statistical significance across all models. However, as indicated in the first-stage results of the J-test in Table 13, only the results of Model (5) can be trusted, since all other models have flawed instruments. Therefore, the results originally presented by Dube and Vargas (2013) only find support when analyzing AUC paramilitaries with the UC measure.

Table 14: Second Stage - AUC

DV: AUC	(1) VI	(2) CD	(3) RC	(4) RE	(5) UC	(6) All
Coffee shocks	-0.220*** (0.000)	-0.297** (0.002)	-0.026 (0.489)	-0.235*** (0.000)	-0.082** (0.002)	-0.261*** (0.001)
Oil shocks	0.500*** (0.000)	0.357 (0.061)	0.081*** (0.001)	0.540* (0.019)	0.280*** (0.000)	0.308* (0.024)
Population	0.108 (0.094)	0.179* (0.050)	0.070 (0.118)	0.105 (0.305)	0.019 (0.315)	0.127 (0.111)
Coca	0.008* (0.016)	0.012 (0.089)	0.011* (0.021)	0.025** (0.009)	0.003 (0.147)	0.007* (0.036)
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region $\times$ Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17,190	12,415	10,505	8,685	16,235	17,324
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01.						
CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conficto; VI=ViPAA;						
UC=UCDP; RE=Restrepo						

## A5. Armed Actors as Independent Variable - full results

### Model Specification and Control Variables in Statistical Analysis

The statistical analysis reported in the lower row of Figure ?? in the manuscript evaluates the effect of using different measures of armed actor presence as independent variables. This statistical analysis uses the following model specification:

$$y_{it} = \alpha_i + \beta A_{it} + \delta X_{it} + \epsilon_{it} \quad (3)$$

where  $y_{it}$  is the homicide rate reported by the Colombian Police (Moreno 2014). The term  $A_{it}$  refers to the type of actor or specific groups reported in each measurement;  $X_{it}$  is a matrix of controls described below;  $\alpha_i$  are municipal fixed effects, and  $\epsilon_{it}$  represents the error terms.

This section presents the list of control variables used in the empirical evaluations conducted in this analysis. All the statistical models in the analysis include a rich set of control variables captured in the matrix  $X$  of the model specification in equation 3. The set of controls includes the following variables:

- The Gross Domestic Product per capita (*GDP pc*) at the department level as reported by DANE (2016). GDP is deflated to express real prices in 2005 (The World Bank 2016). This variable helps to take into account the relations between poverty and violence which could be interpreted as a measure of greed by Collier and Hoeffler (2004) or as state weakness by Fearon and Laitin (2003).
- The control variables also include data about *coca cultivation* and *oil production* with data from (Centro de Estudios Sobre Desarrollo Económico 2019). These variables account for the relationship between natural resources and conflict (Collier and Hoeffler 2004; Ross 2006; Fearon 2005, 2004).
- The model also uses the size of the municipal *population* and its quadratic form *Population sqr.* with data from (Centro de Estudios Sobre Desarrollo Económico 2019).

This curvilinear relationship helps to model the difference between rural areas and large urban centers as they may provide distinct mechanism for violence.

- To account for the effect of Plan Colombia on the dynamics of conflict, the model uses the log of *Plan Colombia Military* funds provided by the U.S. as well as the log of *Plan Colombia Economic* aid for development.
- The controls also include a set of cross-sectional invariant dummy variables *Barco Negotiations* and *Pastrana Negotiations* taking the value of 1 for the years during which the government held peace talks with guerrilla groups in the administration of President Virgilio Barco (1990-91) and President Manuel Pastrana (1998-1999), and 0 otherwise.
- To account for the aggressive military campaign launched by the administration of President Alvaro Uribe (2002-2006), the model includes a dummy variable called *Plan Patriota*, which was the name of this military operation, and takes the value of 1 during his administration, and 0 otherwise.
- The analysis also includes a dummy variable for presidential *elections* that may offer political opportunities for guerrilla attacks or paramilitary assassinations. This variable takes the value of 1 for years when presidential elections took place in Colombia, and 0 otherwise.
- In addition, the model includes dummy variables for the municipalities that report *land mines* in any given year with data from (Centro de Estudios Sobre Desarrollo Económico 2019).
- The variable *Medellin Offensive* accounts for the confrontation period between the state and the Medellin cartel that affected Colombia between 1989 and 1993. This is a cross-sectional invariant variable taking the value of 1 for the 1989-1993 period, and 0 otherwise. Although the Medellin Cartel was located in municipality of Medellín, Antioquia, the cartel conducted violent attacks against the state and civilian targets in different parts of the country. For this reason, this variable only considers temporal variation.

- The model also includes a dummy variable for *reservations* that refer to areas designated as national parks.
- Finally, the statistical analysis includes municipality and year Fixed Effects in all its specifications. This helps to control for unit and time specific unobserved factors.

## Descriptive statistics

Table 15: Descriptive statistics

	Mean	St. Dev.	Min	Max	N
Homicide rate	46.4	67.5	0.0	1,327.9	34,782
Paramilitaries (VI)	0.1	0.3	0.0	1.0	33,000
Paramilitaries (CD)	0.7	0.4	0.0	1.0	27,500
Paramilitaries (CL)	0.3	0.5	0.0	1.0	9,900
Paramilitaries (IN)	0.3	0.5	0.0	1.0	8,800
Paramilitaries (RE)	0.3	0.5	0.0	1.0	26,312
Paramilitaries (UC)	0.01	0.1	0.0	1.0	31,900
Paramilitaries (All)	0.7	0.5	0.0	1.0	34,118
Guerrilla (VI)	0.2	0.4	0.0	1.0	33,000
Guerrilla (CD)	0.6	0.5	0.0	1.0	27,500
Guerrilla (CL)	0.5	0.5	0.0	1.0	9,900
Guerrilla (RE)	0.4	0.5	0.0	1.0	26,312
Guerrilla (UC)	0.1	0.2	0.0	1.0	31,900
Guerrilla (All)	0.6	0.5	0.0	1.0	34,118
FARC (VI)	0.1	0.3	0.0	1.0	33,000
FARC (CD)	0.3	0.5	0.0	1.0	27,500
FARC (RC)	0.003	0.1	0.0	1.0	22,000
FARC (RE)	0.3	0.5	0.0	1.0	26,312
FARC (UC)	0.05	0.2	0.0	1.0	31,900
FARC (All)	0.3	0.5	0.0	1.0	34,118
ELN (VI)	0.1	0.2	0.0	1.0	33,000
ELN (CD)	0.4	0.5	0.0	1.0	27,500
ELN (RC)	0.001	0.03	0.0	1.0	7,700
ELN (UC)	0.01	0.1	0.0	1.0	31,900
ELN (All)	0.4	0.5	0.0	1.0	33,050
AUC (VI)	0.1	0.2	0.0	1.0	33,000
AUC (CD)	0.7	0.4	0.0	1.0	27,500
AUC (RC)	0.05	0.2	0.0	1.0	13,200
AUC (RE)	0.1	0.3	0.0	1.0	16,016
AUC (UC)	0.01	0.1	0.0	1.0	31,900
AUC (All)	0.6	0.5	0.0	1.0	33,694
Coca plantations	0.7	1.8	0.000	9.7	34,782
Oil	52.5	31.1	15.9	98.1	34,782
Mines	0.7	4.9	0.0	232.0	34,782
Elections	0.4	0.5	0.0	1.0	34,782
Population	9.5	1.1	5.1	15.9	34,782
Population sqr.	91.8	21.9	25.6	251.3	34,782
Plan Colombia military	4.2	2.6	0.0	6.7	34,782
Plan Colombia economic	3.1	2.4	0.0	5.5	34,782
Pastrana peace	0.2	0.4	0.0	1.0	34,782
Barco peace	0.1	0.2	0.0	1.0	34,782
Plan Patriota	0.1	0.3	0.0	1.0	34,782
Medellin offensive	0.02	0.1	0.0	1.0	34,782
GDP pc	1,708.4	2,715.1	4.1	38,546.1	34,782
National Parks	0.1	0.3	0.0	1.0	34,782

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA; RE=Restrepo

## Summary of Results

The tables below present the summary of statistical results as well as the full regression tables for the analysis using the different measures of armed actor violent presence as the independent variable. In particular, Table 16 presents the summary of relevant coefficients for paramilitary groups at the actor-type level in Panel (a) and for guerrilla organizations in Panel (b). In addition, Table 17 presents the summary of the organization-level analysis for the FARC insurgency in Panel (a), the ELN guerrilla in Panel (b), and the AUC paramilitary organization in Panel (c). In addition, Tables 18-22 present the full regression results for each correspondent level of analysis using the different measurement sets considered in the study.

Table 16: Summary of Regression Results by Type of Armed Actor

<i>Dependent variable: Homicide rate</i>								
<b>Panel (a) Paramilitaries</b>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Paramilitaries (VI)	16.55*** (1.34)							
Paramilitaries (CD)		9.16*** (0.92)						
Paramilitaries (RE)			24.53*** (0.99)					
Paramilitaries_CL				12.14*** (1.70)				
Paramilitaries_IN					2.61*** (0.75)			
Paramilitaries (RC)						26.00*** (2.38)		
Paramilitaries (UC)							52.06*** (2.52)	
Paramilitaries (All)								9.28*** (0.78)
Controls included	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Controls dropped	0	1	0	2	4	0	0	0
Observations	33,000	27,500	25,668	9,900	8,800	19,800	31,900	33,474
<i>Dependent variable: Homicide rate</i>								
<b>Panel (b) Guerrilla</b>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Guerrilla (VI)	20.21*** (0.99)							
Guerrilla (CD)		15.55*** (0.80)						
Guerrilla (RE)			18.00*** (0.97)					
Guerrilla_CL				10.42*** (1.77)				
Guerrilla (RC)					22.30** (11.23)			
Guerrilla (UC)						25.12*** (1.43)		
Guerrilla (All)							13.35*** (0.79)	
Controls included	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	
Controls dropped	0	1	0	2	3	0	0	
Observations	33,000	27,500	25,668	9,900	6,600	31,900	33,474	

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

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA; RE=Restrepo.

Table 17: Summary of Regression Results by Specific Armed Group

<i>Dependent variable: Homicide rate</i>						
<b>Panel (a) FARC</b>	(1)	(2)	(3)	(4)	(5)	(6)
FARC (VI)	21.71*** (1.10)					
FARC (CD)		15.61*** (0.84)				
FARC (RC)			45.17*** (7.29)			
FARC (RE)				18.73*** (1.04)		
FARC (UC)					25.14*** (1.61)	
FARC (All)						14.69*** (0.81)
Controls included	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Controls dropped	0	1	0	0	0	0
Observations	33,000	27,500	22,000	25,668	31,900	33,474
<i>Dependent variable: Homicide rate</i>						
<b>Panel (b) ELN</b>	(1)	(2)	(3)	(4)	(5)	
ELN (VI)	19.37*** (1.42)					
ELN (CD)		11.13*** (0.86)				
ELN (RC)			86.30*** (22.48)			
ELN (UC)				18.97*** (2.76)		
ELN (All)					10.37*** (0.86)	
Controls included	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	
Controls dropped	0	1	3	0	0	
Observations	33,000	27,500	7,700	31,900	33,050	
<i>Dependent variable: Homicide rate</i>						
<b>Panel (c) AUC</b>	(1)	(2)	(3)	(4)	(5)	(6)
AUC (VI)	22.97*** (1.40)					
AUC (CD)		9.16*** (0.92)				
AUC (RC)			15.87*** (3.06)			
AUC (RE)				33.93*** (1.59)		
AUC (UC)					52.06*** (2.52)	
AUC (All)						5.01*** (0.77)
Controls included	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>	<i>Yes</i>
Controls dropped	0	1	2	2	0	0
Observations	33,000	27,500	13,200	15,624	31,900	33,302

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.0130

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA;

RE=Restrepo.



# Regression Results for Paramilitary Groups

Table 18: Regression Results for Paramilitary Groups

	<i>Dependent variable: Homicide rate</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Paramilitaries (VI)	16.55*** (1.34)							
Paramilitaries (CD)		9.16*** (0.92)						
Paramilitaries (RE)			24.53*** (0.99)					
Paramilitaries_CL				12.14*** (1.70)				
Paramilitaries_IN					2.61*** (0.75)			
Paramilitaries (RC)						26.00*** (2.38)		
Paramilitaries (UC)							52.06*** (2.52)	
Paramilitaries (All)								9.28*** (0.78)
Coca	2.01*** (0.39)	2.05*** (0.40)	1.59*** (0.47)	1.81** (0.81)	0.16 (0.51)	0.71 (0.62)	2.12*** (0.39)	2.19*** (0.38)
Oil	-0.38*** (0.02)	-0.44*** (0.02)	-0.18*** (0.03)	0.13 (0.12)	0.03 (0.03)	-0.06 (0.08)	-0.39*** (0.02)	-0.39*** (0.02)
Mines	1.03*** (0.07)	1.02*** (0.06)	1.58*** (0.09)	2.51*** (0.21)	0.25*** (0.04)	2.91*** (0.16)	1.09*** (0.07)	1.07*** (0.07)
Elections	0.03 (0.69)	-0.01 (0.70)	-0.85 (0.89)	2.76 (3.56)	1.08 (0.86)	0.34 (1.08)	0.54 (0.69)	-1.14* (0.69)
Population	78.03*** (21.44)	9.79 (26.06)	78.53*** (27.13)	221.65* (120.17)	156.65* (81.76)	148.65*** (44.34)	45.57** (22.03)	62.06*** (19.63)
Population sqr.	-4.86*** (1.10)	-1.15 (1.34)	-5.41*** (1.40)	8.66 (6.15)	-10.02** (4.18)	-9.69*** (2.26)	-3.08*** (1.13)	-4.02*** (1.01)
Plan Colombia military \$	0.23 (0.29)	1.02*** (0.32)	0.27 (0.31)	-2.58 (5.84)	-11.67 (7.76)	-0.03 (0.36)	0.01 (0.30)	-0.01 (0.29)
Plan Colombia economic \$	-0.17 (0.35)	0.74** (0.34)	-0.85** (0.38)	0.24 (1.85)	26.86* (14.09)	0.30 (0.43)	0.55 (0.35)	0.61* (0.35)
Pastrana peace	4.09*** (1.26)	0.07 (1.28)	3.97*** (1.38)	-0.64 (2.82)		1.52 (1.55)	1.66 (1.25)	2.61** (1.25)
Barco peace	1.95 (1.37)		-0.69 (1.46)			-0.19 (1.65)	2.07 (1.38)	5.23*** (1.37)
Plan Patriota	-0.36 (1.18)	0.50 (1.12)	-1.93 (1.27)	-17.05*** (2.96)		-16.37*** (2.60)	-1.43 (1.17)	1.49 (1.18)
Medellin Offensive	53.27*** (2.51)	36.01*** (4.75)	48.30*** (2.74)			42.44*** (3.25)	57.32*** (2.50)	52.66*** (2.49)
GDP pc	-0.003*** (0.0003)	-0.004*** (0.0003)	-0.01*** (0.001)	-0.01*** (0.003)	0.001** (0.0005)	-0.004*** (0.001)	-0.003*** (0.0003)	-0.003*** (0.0003)
National park	11.52** (4.68)	5.77 (5.15)	16.52*** (5.75)	-7.28 (15.97)	3.80 (13.33)	21.01** (8.83)	6.64 (4.92)	12.29*** (4.43)
Observations	33,000	27,500	25,668	9,900	8,800	19,800	31,900	33,474

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

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA; RE=Restrepo

## Regression Results for Guerrilla Groups

Table 19: Regression Results for Guerrilla Groups

	<i>Dependent variable: Homicide rate</i>						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Guerrilla (VI)	20.21*** (0.99)						
Guerrilla (CD)		15.55*** (0.80)					
Guerrilla (RE)			18.00*** (0.97)				
Guerrilla_CL				10.42*** (1.77)			
Guerrilla (RC)					22.30** (11.23)		
Guerrilla (UC)						25.12*** (1.43)	
Guerrilla (All)							13.35*** (0.79)
Coca	1.96*** (0.39)	1.93*** (0.40)	1.81*** (0.47)	1.80** (0.82)	1.33 (1.08)	2.17*** (0.39)	2.20*** (0.38)
Oil	-0.34*** (0.02)	-0.48*** (0.02)	-0.37*** (0.03)	0.15 (0.12)	0.20** (0.10)	-0.35*** (0.02)	-0.45*** (0.02)
Mines	0.94*** (0.07)	0.98*** (0.06)	1.49*** (0.09)	2.43*** (0.21)	3.75*** (0.32)	0.97*** (0.07)	1.07*** (0.07)
Elections	0.04 (0.69)	-0.59 (0.69)	-0.60 (0.90)	3.26 (3.56)	36.94*** (6.41)	0.37 (0.69)	-0.66 (0.68)
Population	71.84*** (21.28)	-4.57 (25.93)	95.68*** (27.27)	255.23** (120.17)	265.01** (130.58)	49.87** (22.07)	55.47*** (19.60)
Population sqr.	-4.58*** (1.09)	-0.44 (1.33)	-6.34*** (1.40)	10.43* (6.15)	12.48* (6.69)	-3.33*** (1.13)	-3.68*** (1.01)
Plan Colombia military \$	-0.03 (0.29)	0.62** (0.31)	0.32 (0.32)	-0.86 (5.84)	63.03*** (9.77)	0.21 (0.30)	0.16 (0.29)
Plan Colombia economic \$	-0.54 (0.35)	-0.02 (0.33)	-0.25 (0.38)	-0.28 (1.85)	-17.01*** (3.07)	-0.14 (0.35)	0.03 (0.35)
Pastrana peace	-0.31 (1.28)	3.14*** (1.18)	3.89*** (1.39)	-0.91 (2.82)	-0.31 (3.64)	3.90*** (1.25)	2.71** (1.24)
Barco peace	1.19 (1.36)		2.60* (1.46)			1.87 (1.38)	3.25** (1.34)
Plan Patriota	-2.00* (1.18)	-0.53 (1.10)	-0.82 (1.28)	-17.58*** (2.96)		-3.32*** (1.18)	-0.22 (1.16)
Medellin Offensive	53.65*** (2.50)	34.75*** (4.73)	50.88*** (2.75)			56.10*** (2.51)	52.67*** (2.48)
GDP pc	-0.003*** (0.0003)	-0.004*** (0.0003)	-0.01*** (0.001)	-0.01*** (0.003)	-0.004* (0.002)	-0.003*** (0.00)	-0.004*** (0.00)
National park	12.07*** (4.66)	6.04 (5.13)	15.81*** (5.78)	-7.74 (15.99)	-0.85 (17.07)	6.68 (4.93)	13.22*** (4.42)
Observations	33,000	27,500	25,668	9,900	6,600	31,900	33,474

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

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA; RE=Restrepo

## Regression Results for FARC

Table 20: Regression Results for FARC

	<i>Dependent variable: Homicide rate</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
FARC (VI)	21.71*** (1.10)					
FARC (CD)		15.61*** (0.84)				
FARC (RC)			45.17*** (7.29)			
FARC (RE)				18.73*** (1.04)		
FARC (UC)					25.14*** (1.61)	
FARC (All)						14.69*** (0.81)
Coca	2.04*** (0.39)	1.73*** (0.40)	1.93*** (0.50)	1.88*** (0.47)	2.11*** (0.39)	2.02*** (0.38)
Oil	-0.34*** (0.02)	-0.33*** (0.02)	-0.37*** (0.03)	-0.33*** (0.03)	-0.36*** (0.02)	-0.32*** (0.02)
Mines	0.94*** (0.07)	0.92*** (0.06)	1.11*** (0.08)	1.44*** (0.09)	0.98*** (0.07)	1.00*** (0.07)
Elections	0.07 (0.69)	0.90 (0.69)	5.15*** (1.12)	-1.50* (0.90)	0.24 (0.69)	0.32 (0.68)
Population	70.48*** (21.29)	17.24 (25.94)	69.85** (32.12)	96.14*** (27.28)	52.11** (22.09)	73.93*** (19.58)
Population sqr.	-4.51*** (1.09)	-1.54 (1.33)	-4.65*** (1.64)	-6.37*** (1.41)	-3.45*** (1.13)	-4.64*** (1.01)
Plan Colombia military \$	0.09 (0.29)	0.11 (0.31)	1.99*** (0.37)	0.20 (0.32)	0.27 (0.30)	-0.08 (0.29)
Plan Colombia economic \$	-0.64* (0.35)	-0.62* (0.33)	-0.96** (0.40)	-0.36 (0.38)	-0.13 (0.35)	-0.54 (0.35)
Pastrana peace	0.73 (1.27)	4.77*** (1.17)	3.94*** (1.42)	4.28*** (1.39)	4.00*** (1.25)	4.14*** (1.23)
Barco peace	1.53 (1.36)		2.76* (1.62)	2.75* (1.46)	1.99 (1.38)	1.62 (1.34)
Plan Patriota	-1.71 (1.18)	-2.37** (1.11)	-3.63*** (1.35)	-1.84 (1.28)	-2.67** (1.18)	-2.08* (1.16)
Medellin Offensive	53.45*** (2.50)	35.29*** (4.73)	66.13*** (3.42)	50.80*** (2.75)	55.76*** (2.51)	53.67*** (2.48)
GDP pc	-0.003*** (0.0003)	-0.003*** (0.0003)	-0.004*** (0.001)	-0.01*** (0.001)	-0.003*** (0.0003)	-0.003*** (0.0003)
National park	11.41** (4.67)	4.83 (5.13)	6.34 (6.76)	14.88** (5.78)	6.15 (4.93)	11.78*** (4.41)
Observations	33,000	27,500	22,000	25,668	31,900	33,474
R <sup>2</sup>	0.10	0.10	0.08	0.08	0.10	0.10
Adjusted R <sup>2</sup>	0.07	0.06	0.03	0.03	0.07	0.07

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

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA;

RE=Restrepo

## Regression Results for ELN

Table 21: Regression Results for ELN

	<i>Dependent variable: Homicide rate</i>				
	(1)	(2)	(3)	(4)	(5)
ELN (VI)	19.37*** (1.42)				
ELN (CD)		11.13*** (0.86)			
ELN (RC)			86.30*** (22.48)		
ELN (UC)				18.97*** (2.76)	
ELN (All)					10.37*** (0.86)
Coca	2.09*** (0.39)	2.11*** (0.40)	2.09** (0.95)	2.20*** (0.39)	2.28*** (0.39)
Oil	-0.36*** (0.02)	-0.50*** (0.02)	-0.40*** (0.06)	-0.37*** (0.02)	-0.48*** (0.03)
Mines	1.05*** (0.07)	0.98*** (0.06)	1.64*** (0.19)	1.06*** (0.07)	1.06*** (0.07)
Elections	-0.13 (0.69)	-0.49 (0.70)	-7.81*** (2.55)	-0.08 (0.69)	-1.03 (0.70)
Population	64.73*** (21.35)	1.44 (26.03)	-283.07*** (96.20)	46.94** (22.16)	57.34*** (21.23)
Population sqr.	-4.16*** (1.09)	-0.74 (1.33)	13.82*** (4.93)	-3.17*** (1.13)	-3.82*** (1.09)
Plan Colombia military \$	0.18 (0.29)	0.69** (0.31)	0.33 (5.07)	0.34 (0.30)	0.32 (0.29)
Plan Colombia economic \$	-0.19 (0.35)	0.32 (0.33)	3.31** (1.55)	-0.09 (0.35)	0.36 (0.35)
Pastrana peace	2.78** (1.27)	3.44*** (1.18)	15.36*** (2.86)	5.11*** (1.25)	3.57*** (1.26)
Barco peace	1.63 (1.37)			2.23 (1.38)	4.06*** (1.37)
Plan Patriota	-1.06 (1.18)	0.22 (1.11)		-1.44 (1.18)	0.27 (1.19)
Medellin Offensive	53.64*** (2.51)	34.70*** (4.75)		56.41*** (2.52)	53.76*** (2.51)
GDP pc	-0.003*** (0.00)	-0.004*** (0.00)	-0.01*** (0.001)	-0.003*** (0.00)	-0.003*** (0.00)
National park	11.75** (4.68)	6.40 (5.15)	2.35 (13.40)	6.04 (4.95)	12.27*** (4.65)
Observations	33,000	27,500	7,700	31,900	33,050
R <sup>2</sup>	0.10	0.09	0.07	0.10	0.10
Adjusted R <sup>2</sup>	0.07	0.06	-0.09	0.07	0.07

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

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA;

RE=Restrepo

## Regression Results for AUC

Table 22: Regression Results for AUC

	<i>Dependent variable: Homicide rate</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
AUC (VI)	22.97*** (1.40)					
AUC (CD)		9.16*** (0.92)				
AUC (RC)			15.87*** (3.06)			
AUC (RE)				33.93*** (1.59)		
AUC (UC)					52.06*** (2.52)	
AUC (All)						5.01*** (0.77)
Coca	2.11*** (0.39)	2.05*** (0.40)	1.29** (0.65)	1.77*** (0.56)	2.12*** (0.39)	2.27*** (0.38)
Oil	-0.35*** (0.02)	-0.44*** (0.02)	-0.07 (0.08)	-0.24*** (0.03)	-0.39*** (0.02)	-0.40*** (0.02)
Mines	1.04*** (0.07)	1.02*** (0.06)	2.63*** (0.16)	1.19*** (0.10)	1.09*** (0.07)	1.08*** (0.07)
Elections	0.43 (0.69)	-0.01 (0.70)	-0.16 (1.33)	0.76 (1.27)	0.54 (0.69)	-0.95 (0.70)
Population	74.62*** (21.35)	9.79 (26.06)	73.78 (64.46)	-131.99** (55.68)	45.57** (22.03)	65.36*** (20.98)
Population sqr.	-4.67*** (1.09)	-1.15 (1.34)	-5.74* (3.30)	6.23** (2.87)	-3.08*** (1.13)	-4.19*** (1.07)
Plan Colombia military \$	0.18 (0.29)	1.02*** (0.32)	1.54*** (0.43)	1.73 (2.39)	0.01 (0.30)	0.002 (0.29)
Plan Colombia economic \$	-0.39 (0.35)	0.74** (0.34)	-1.02** (0.43)	-1.43* (0.80)	0.55 (0.35)	0.54 (0.36)
Pastrana peace	2.61** (1.26)	0.07 (1.28)	-1.17 (1.49)	2.11 (1.61)	1.66 (1.25)	3.11** (1.28)
Barco peace	1.59 (1.36)				2.07 (1.38)	5.12*** (1.44)
Plan Patriota	-2.10* (1.18)	0.50 (1.12)	-13.73*** (2.48)	-5.21*** (1.50)	-1.43 (1.17)	0.23 (1.18)
Medellin Offensive	54.01*** (2.50)	36.01*** (4.75)			57.32*** (2.50)	53.67*** (2.50)
GDP pc	-0.003*** (0.0003)	-0.004*** (0.0003)	-0.01*** (0.001)	-0.01*** (0.001)	-0.003*** (0.00)	-0.003*** (0.00)
National park	12.08*** (4.67)	5.77 (5.15)	-0.02 (11.44)	4.68 (7.66)	6.64 (4.92)	11.85*** (4.56)
Observations	33,000	27,500	13,200	15,624	31,900	33,302
R <sup>2</sup>	0.10	0.09	0.04	0.09	0.11	0.09
Adjusted R <sup>2</sup>	0.07	0.05	-0.05	0.02	0.08	0.06

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

CD=CEDE; CL=Claudia Lopez; RC=Rutas del Conflicto; VI=ViPAA; RE=Restrepo

## A6. Model Assessment

The empirical evaluation of using a variety of armed actor measures as independent variables in the previous section reveals the diverging estimates stemming from different measures of armed actors in relation to violence in Colombia. However, since this research is agnostic about the essential superiority or proximity to the "ground truth" of one measure over the other ones, it is necessary to rely on basic methodological prescriptions to assess the quality of the different models. It is not enough to state that the regression estimates are different simply because they come from different measures of armed actors. Instead, this section relies on three methodological dimensions to evaluate the quality of the models derived from different measurement sets: (i) information loss; (ii) the Akaike Information Criterion (AIC) (Akaike 1974), and (iii) the Bayesian Information Criterion (BIC) (Stone 1979). Table 23 reports these three dimensions for the generic type of actors (paramilitaries and guerrilla) under Panel (a) and for the specific armed groups (FARC, ELN, and AUC) under Panel (b).

[Table 23 around here]

First, given the pervasiveness of missing data in some of the measurement sets, the loss of information is a central criteria for assessing the quality of the data and the statistical inferences derived from them. Departing from the assumption that researchers look for generalizable conclusions that are valid to explain a broad scope of cases considered under a given theoretical framework, a high rate of missing data would prevent generalizing the inferences from the model outside the sample, particularly if the missing data is generated through a non random process (Honaker and King 2010). To evaluate this concern, Table 23 reports the percentage of missing data related to each model with respect to the panel of all municipalities of Colombia between 1988 and 2017. According to the calculations, the integrative measure *All* and VI are the measurement sets with the least loss of information, with only 5.51% and 6.85% of missing data respectively. The other measurement sets generate a considerable amount of missing data ranging from 22.37% to 75.16%, thus severely undermining the completeness of the data and the prospects of generalizing results. Researchers seeking

to generate generalizable conclusions would benefit from using the *All* and VI measurement sets over alternative indicators of armed actors. Substantially, data from *All* is the most complete as it integrates the observations of armed actors from all other measurements.

Based on the initial analysis of missing data, the following evaluation of the AIC and BIC metrics focuses exclusively on comparing the performance of *All* and VI models. Panel (a) in Table 23 reports the performance of the AIC and BIC statistics for the type of actor (paramilitaries and guerrilla) and Panel (b) reports those metrics for specific armed groups (FARC, ELN, and AUC). The AIC statistic (Akaike 1974) evaluates the relative quality of the model based on the goodness of fit and its parsimony. In this way, a smaller AIC score indicates a better model than an alternative specification with a larger AIC score. In addition, the BIC statistic (Stone 1979) evaluates the relative quality of a model based on the likelihood function, the number of observations, and the parsimony of the model. A smaller BIC score also indicates a preferable specification when compared to a larger BIC score.

When comparing the integrative measure *All* and VI for paramilitary organizations and guerrilla groups in Panel (a) of Table 23, the AIC and BIC provide inconclusive results. For Paramilitary actors, the AIC favors the integrative measure *All* while the BIC favors VI. The model assessment scores alternate in a similar manner when assessing *All* and VI models for guerrilla groups. The differences between the AIC and BIC scores for *All* and VI is so small, that the alternance in favor of one model over the other is not surprising. This ambivalence at the margin suggests that the fit of the models is very similar. Therefore, it is plausible to consider *All* and VI as the two best measurement sets to track the presence of paramilitary and guerrilla groups in Colombia.<sup>1</sup>

Panel (b) in Table 23 yields to similar conclusions for the specific measurements of the FARC, ELN, and AUC groups. Both the AIC and BIC criteria alternate their favor of the

---

<sup>1</sup>The AIC and BIC scores are indeed smaller for CD, CL, IN, RC, and RE than for the integrative measure *All* and VI. However, those AIC and BIC scores are smaller because they contain fewer observations than the more complete *All* and VI models.

integrative *All* metric or the VI database for each of the armed groups. This ambivalence at the margin is consistent with the assessment for paramilitary and guerrilla type of actors.

In general, conducting statistical analyses with the integrative *All* measure or the VI data is preferable than using alternative measurement sets. The specific differences between *All* and VI models are marginal when considering the percentage of missing values, as well as the AIC and BIC performance statistics. This is applicable for both the type of armed actors as well as the specific armed groups evaluated in the study. Therefore, researchers can consider both metrics as better options for conducting statistical analyses on the dynamics of conflict without incurring in severe problems of missing data that could yield bias inferences and limit the generalizability of their conclusions.

However, despite the similarity of the model fit assessment derived from the AIC and BIC measures, the integrative *All* measure and VI do not yield the same coefficients from the statistical analysis. As indicated in Tables 16 and 17, as well as in the coefficient plots in the lower row of Figure ??, VI tends to generate larger estimates than those derived from the integrative *All* measurement set for both the type of armed actors (paramilitaries and guerrillas) as well as for specific armed groups (FARC, ENL, and AUC).



Table 23: AIC and BIC Criteria

Model	% Missing	Panel (a)						Panel (b)					
		Type of Actor			Guerrilla			FARC			Armed Group		
		Paramilitaries		AIC	Guerrilla		BIC	FARC		BIC	Armed Group		BIC
		AIC	BIC		AIC	BIC		AIC	BIC		AIC	BIC	
1 All	5.51	11.383	361,479	11.383	361,339	361,339	11.389	374,529	11.399	357,301	11.386	359,893	11.386
2 ViPAA (VI)	6.85	11.400	356,799	11.392	356,530	356,530	11.393	356,557	11.399	356,764	11.397	356,679	11.397
3 CEDE (CD)	22.37	11.072	293,299	11.061	293,014	293,014	11.063	293,049	11.069	293,229	11.072	293,299	11.072
4 Restrepo (RE)	27.54	11.236	280,178	11.236	336,511	336,511	11.248	280,470	11.094	240,084	10.599	359,893	10.599
5 Rutas del Conflicto (RC)	59.63	11.093	218,173	11.093	72,575	72,575	11.094	240,084	10.091	84,430	10.507	143,066	10.507
6 Claudia Lopez (CL)	72.05	10.358	108,704	10.360	108,723	108,723							
7 Indepaz (IN)	75.16	8.165	78,342										

AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion.

## A7. Relevance of Measurement Sets

Table 24 presents a summary of Google Scholar citations of each different database. Although these citation counts may not accurately reflect the use of these databases outside academia (e.g. policy, journalism, or civil society sectors), they provide some indication about their relevance in the research community.

Table 24: Citations of Measurement Sets in Google Scholar

Database	Publication year	Source	Number of Citations	Google Scholar links
RE	2004	Restrepo et al. (2004)	238	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22The+Dynamics+of+the+Colombian+Civil+Conflict%22+%2B+Restrepo&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=0%2C3&amp;q=%22Special+Data+Feature%3B+The+Severity+of+the+Colombian+Conflict%3A+Cross-Country+Datasets+Versus+New+Micro-Data%22&amp;btnG=https://scholar.google.com/scholar?cites=13116620271918594088&amp;as_sdt=805&amp;scioldt=0,3&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22The+Dynamics+of+the+Colombian+Civil+Conflict%22+%2B+Restrepo&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=0%2C3&amp;q=%22Special+Data+Feature%3B+The+Severity+of+the+Colombian+Conflict%3A+Cross-Country+Datasets+Versus+New+Micro-Data%22&amp;btnG=https://scholar.google.com/scholar?cites=13116620271918594088&amp;as_sdt=805&amp;scioldt=0,3&amp;hl=en</a>
CL	2010	López (2010)	443	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Introducing+the+UCDP+georeferenced+event+dataset%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=0%2C3&amp;q=%22Panel+Municipal+del+CEDE%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Geograf%C3%ADa+del+Paramilitarismo%22+en+Colombia+%2B+Rutas+del+Conflicto&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Introducing+the+UCDP+georeferenced+event+dataset%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=0%2C3&amp;q=%22Panel+Municipal+del+CEDE%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Geograf%C3%ADa+del+Paramilitarismo%22+en+Colombia+%2B+Rutas+del+Conflicto&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en</a>
UC	2013	(Sundberg and Melander 2013)	877	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Introducing+the+UCDP+georeferenced+event+dataset%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=0%2C3&amp;q=%22Panel+Municipal+del+CEDE%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Geograf%C3%ADa+del+Paramilitarismo%22+en+Colombia+%2B+Rutas+del+Conflicto&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Introducing+the+UCDP+georeferenced+event+dataset%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=0%2C3&amp;q=%22Panel+Municipal+del+CEDE%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Geograf%C3%ADa+del+Paramilitarismo%22+en+Colombia+%2B+Rutas+del+Conflicto&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en</a>
CD	2014	Acevedo and Bornacelly (2014)	193	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Geograf%C3%ADa+del+Paramilitarismo%22+en+Colombia+%2B+Rutas+del+Conflicto&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;as_ylo=2010&amp;as_yhi=2022&amp;q=%22Geograf%C3%ADa+del+Paramilitarismo%22+en+Colombia+%2B+Rutas+del+Conflicto&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en</a>
RC	2019	Rutas del Conflicto (2019)	18	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en</a>
VI	2019	(Osorio et al. 2019)	27	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Osorio+%22ViPAA%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=805&amp;scioldt=807&amp;cites=4593544114437560797&amp;scipsc=&amp;q=%22Informe+sobre+presencia+de+grupos+Narcoparamilitares%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Osorio+%22ViPAA%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=805&amp;scioldt=807&amp;cites=4593544114437560797&amp;scipsc=&amp;q=%22Informe+sobre+presencia+de+grupos+Narcoparamilitares%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en</a>
IN	2019	Indepaz (2019)	68	<a href="https://scholar.google.com/scholar?hl=en&amp;as_sdt=805&amp;scioldt=807&amp;cites=4593544114437560797&amp;scipsc=&amp;q=%22Informe+sobre+presencia+de+grupos+Narcoparamilitares%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en">https://scholar.google.com/scholar?hl=en&amp;as_sdt=805&amp;scioldt=807&amp;cites=4593544114437560797&amp;scipsc=&amp;q=%22Informe+sobre+presencia+de+grupos+Narcoparamilitares%22&amp;btnG=https://scholar.google.com/scholar?hl=en&amp;as_sdt=807&amp;q=Verdad+Abierta+%22Geograf%C3%ADa+del+Paramilitarismo%22&amp;btnG=https://scholar.google.com/scholar?cites=4593544114437560797&amp;as_sdt=805&amp;scioldt=807&amp;hl=en</a>

## References

- Acevedo, K. and I. Bornacelly. 2014. Panel Municipal del CEDE. Technical report, Universidad de los Andes, Bogota, Colombia.
- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6): 716 – 723.
- Angrist, J. and J.-S. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton, New Jersey: Princeton University Press.
- Centro de Estudios Sobre Desarrollo Económico 2019. Catálogo de Datos.
- Centro de Investigación y Educación Popular 2017. Marco Conceptual de Red Nacional de Bancos de Datos. Technical report, Centro de Investigación y Educación Popular - Programa por la Paz (CINEP-PPP), Bogota, Colombia.
- Collier, P. and A. Hoeffler. 2004. Greed and grievance in civil war. *Oxford Economic Papers* 56: 563–595.
- Departamento Administrativo Nacional de Estadística 2016. Producto Interno Bruto Departamental a precios corrientes.
- Dube, O. and J. Vargas. 2013. Commodity Prices Shocks and Civil Conflict: Evidence from Colombia. *Review of Economic Studies* 80(4): 1384–1421.
- Fearon, J. D. 2004. Why Do Some Civil Wars Last So Much Longer than Others? *Journal of Peace Research* 41(3): 275–301.
- Fearon, J. D. 2005. Primary Commodity Exports and Civil War. *The Journal of Conflict Resolution* 49(4): 483–507.
- Fearon, J. D. and D. D. Laitin. 2003. Ethnicity, Insurgency, and Civil War. *American Political Science Review* 97(1): 75–90.
- Gonzalez Posso, C. and C. E. Espitia. 2017. Dejación de Armas y Sometimiento. XIII Informe Presencia de Grupos Narcoparamilitares. Technical report, Instituto para el Desarrollo y la Paz, INDEPAZ, Bogota, Colombia.

- Högbladh, S. 2021. UCDP GED Codebook version 21.1. Technical report, Department of Peace and Conflict Research, Uppsala University, Uppsala, Sweden.
- Honaker, J. and G. King. 2010. What to do about missing values. *American Journal of Political Science* 54(2): 561–581.
- Indepaz 2019. Informe sobre presencia de grupos Narcoparamilitares.
- Lee, David, M. Moreira, J. McCrary, and J. Porter. 2020. Valid t-ratio Inference for IV.
- López, C. (Ed.) 2010. *Y Refundaron la Patria: De cómo Mafiosos y Políticos Reconfiguraron el Estado Colombiano*. Bogotá, Colombia: Corporación Nuevo Arcoiris.
- Moreno, J. 2014. Tasas de homicidios en Colombia.
- Osorio, J., M. Mohamed, V. Pavon, and S. Brewer-Osorio. 2019. Mapping Violent Presence of Armed Actors in Colombia. In *ICC 2019 International Cartographic Conference*, Tokyo, Japan, pp. 1–8. International Cartographic Association.
- Restrepo, J., M. Spagat, and J. Vargas. 2004. *The Dynamics of the Colombian Civil Conflict: A New Data Set*, Volume 21.
- Ross, M. 2006, jun. A Closer Look At Oil, Diamonds, and Civil War. *Annual Review of Political Science* 9(1): 265–300.
- Rutas del Conflicto 2019. Geografía del Paramilitarismo en Colombia.
- Stone, M. 1979. Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society* 41(2): 276–278.
- Sundberg, R. and E. Melander. 2013. Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4): 523–532.
- The World Bank 2016. GDP deflator.