Spring 5-2021

Identifying armed group presence using Hidden Markov Models

MAURICIO VELA BARON
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST.LOUIS

School of Arts & Sciences
Department of Mathematics and Statistics

Identifying Armed Group Presence Using Hidden Markov Models

by

Mauricio Andres Vela Baron

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2021
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Firstly, I would like to thank my parents, whom without this would have not been possible.I would also like to thank my thesis committee: Dr. Nan Lin, Dr. Enamorado and Dr. Figueroa-Lopez for all their help and advice.

Mauricio Andres Vela Baron

Washington University in Saint Louis

May 2021

Dedicated to my parents and Viole.

ABSTRACT OF THE THESIS

Identifying Armed Group Presence Using Hidden Markov Models

by

Mauricio Andres Vela Baron

Master of Arts in Statistics

Washington University in St. Louis, 2021

Identifying armed group presence is often helpful for conflict studies to examine patterns of conflict. Armed group presence is often used as the main variable of interest in several studies, and in some cases, this variable is ignored. Many of these studies use expert data or proxy variables to analyze armed group presence. This paper proposes Hidden Markov Models (HMMs) as a method to identify armed group presence. HMMs permit identifying armed group presence at a sub-national level and in long panel data sets. A HMM is used in this paper to identify paramilitary and FARC presence in Colombia. The armed groups' presence predictions are used to analyze the effect of economic shocks on violence. It shows that economic shocks can have differential effects depending on the level of armed group presence in the territory.

Chapter 1

Introduction

1.1 Motivation

Studies on intractable conflicts often rely on indicators of armed group presence. Identifying which actor occupies a territory within a country is an essential task. These studies frequently include this variable into their central hypothesis (Prem, Saavedra, and Vargas, 2020) or may even be the main objective of the study (Thill et al., 2017). However, identifying the presence of certain armed groups in a particular territory is a difficult task. These studies often use data from experts (Bickel, 2018; Raleigh et al., 2010), or data of violent actions by armed groups (Fergusson, Vargas, and Vela, 2013; Namen, Prem, and Vargas, 2019; Prem, Saavedra, and Vargas, 2020; Thill et al., 2017). The former expert data may not always be available or, if available, may be limited to some years or some territories, and it is not useful when having sub-national panel data. Expert data is generally limited to one specific year and, in most cases, not disaggregated into small sub-national units such as municipalities or counties. Therefore, using expert data is not always a possibility when analyzing armed group presence, especially when the study aims to understand the conflict's dynamics within a country. Armed group presence is generally dynamic so that in one year, an armed group has a strong presence in a territory, but in some other period, this is no longer the case.

The second approach, using data of violent actions by armed groups such as terrorist attacks, clashes, homicides, kidnappings, etc., solves the previously mentioned problems as this data is usually very detailed. With these data, scholars can identify the exact time and location of violent events. However, these data may help the study measure violence exposure. Still, it is not useful if the goal is to identify armed group presence or define which armed actor controls a specific territory. For instance, a terrorist attack by a particular illegal armed group does not necessarily mean that this armed actor has a strong presence in the territory under attack. Instead, high levels of terrorist attacks may imply the opposite strong-armed group presence; terrorist attacks may be a violent tactic if the illegal actor does not have a considerable presence in the territory (Calle and Sánchez-Cuenca, 2015).

The purpose of this research is to propose the use of Hidden Markov Models (HMMs) as an approach to uncover the presence of armed groups. HMMs are statistical models very well used in data sequence generation, pattern recognition, classification, and pattern segmentation problems. They have been applied to different fields of study, including bioinformatics, epidemiology, environment, finance, ecology, among others (Albert and Chib, 1993; Green and Richardson, 2002; Juang and Rabiner, 1991; Langrock et al., 2012, e.g.). In HMMs, the distribution that generates the observation depends on the state of an underlying and unobserved Markov process. In the case of conflict, the unobserved latent variable is the armed group presence, and we can assume that it satisfies the Markov property.

Although this variable is not observed, we can infer it from armed groups' observable violent actions. For instance, clashes between armed groups or between an armed group and the

State are more likely if the armed group presence is medium, not a consolidated presence but neither an absent level of presence. Attacks on civilians and terrorist attacks are also more common in territories with low armed group presence (Huang and Kennedy, 2008).

In this way, HMMs may be helpful to predict armed group presence. These models can detect underlying latent structures and detect unobservable states. Therefore it can be used to detect armed group presence. In the next section is a literature review of other methodologies used to identify armed group presence. The following two sections describe the data and classification of violent events. Then the methodology and the results are presented. The following section includes an estimation including covariates, and finally, predicted armed group presence is used to estimate the effects of economic shocks on violence.

1.2 Literature Review

Previous studies have proposed methods to identify armed group presence. Thill et al. (2017) uses theory to measure the presence of armed groups and their control, and their approach relies on geo-coded event data that records information about battles between combatants. They categorized the armed actor who initiates each battle and the combatant who gains control after the battle. With the combinations of these categories, they define whether an armed group is present or not. Their method has a significant flaw. Information on government and rebel forces' positioning is usually difficult to obtain, so they assume that major towns are under the government's complete control.

Using event data, Sundberg and Melander (2013) created conflict polygons defined as geographical areas within which conflict events occur. Asal et al. (2016) also use GIS to assess hypotheses about civil conflict at the sub-national level. However, these two approaches alone do not identify the level of armed group presence, but only which areas are under dispute. Tao et al. (2016) seeks to solve this limitation by applying network analysis on a transportation network to identify the polygons and areas with armed group presence. They used road data and hexagon-fishnet-based artificial road data to simulate off-road movement. With this geospatial approach, they can consider on and off-road movements and define if an armed group is present and the area they control. They assume that the government or rebel group holds an area if its troops can reach and take action within a specific response or reaction time. However, again they must know ex-ante which locations already are under an armed group control.

Using group-based trajectory models García-Sánchez (2010) identified distinctive trajectories described by armed groups' violent actions based on the spatial analysis of conflict-related variables. He uses semi-parametric group-based modelling applied to the longitudinal data that identifies clusters based on the optimal number of trajectories and their shape. The model allows identifying municipality clusters in which violent actions perpetrated by armed groups followed significantly different patterns. The information is then classified to categorize each municipality according to the intensity of armed actors' violent events. Using HMMs, however, has the advantage of not only using geo-located but other contextual variables as well.

Finally, other studies have sought to identify other dimensions of conflict, such as forecasting when a state will fail (King and Zeng, 2001) or to identify ongoing conflicts and new ones Ward and Gleditsh, 2002. Similarly, some studies use HMM to forecast conflict in Israel (Shearer, 2007), Lebanon (Schrodt, 2000) and the Balkans(Schrodt, 2006). Recent literature in Political Science has also analyzed the possibility of latent heterogeneity in the covariate effects. For instance, Ferrari (2020) propose a hierarchical Dirichlet process of generalized linear models in which the latent heterogeneity can depend on context-level features. To the extent of my knowledge, this is the first paper to use HMMs to uncover armed group presence.

Chapter 2

Data and Methodology

2.1 Colombian case

For this research, I will be using the Colombian case. The current Colombian conflict started with the launch of a communist insurgency in the 1960s. During this time, several left-wing guerrilla groups were formed to overthrow the state. The guerrilla groups included the National Liberation Army (ELN), Revolutionary Armed Forces of Colombia (FARC), Popular Liberation Army (EPL), and the 19 of April Movement (M-19). Some of these groups demobilized after different peace processes. These guerrilla groups were mainly rural and focused their attacks on fixed government positions and public infrastructure. During the 1980s, a new armed actor emerged: the paramilitary groups. These paramilitary groups started to play a central role in the Colombian state's counterinsurgent strategy. However, these groups were distinguished by the use of violence and vigilante justice, defended regional elite's and drug dealer's private interests with the use of the force, committed several violations to human rights, and were linked to illegal activities from cocaine exports, contraband, illegal mining, extortion, and others. The Colombian conflict has been one of the bloodiest and longest civil conflicts in the world. More than six million victims between 1984 and 2019 have been counted from this conflict.

I will use event-counting data on violent actions. The information I will use comes from *Centro de Investigación y Educación Popular* (CINEP) and is constructed from coding events and includes more than 30,000 war-related episodes in all Colombian municipalities from 1998 to 2018. This data is a well-known dataset on conflict-related issues that has been used by several authors studying conflict in Colombia. This database is very useful for the posterior analysis. It includes every violent action during the Colombian conflict since 1998 with detailed information about the perpetrator, exact location, day and a brief description of the event.

2.2 Violent events classification

I first organized the CINEP's database according to the perpetrator of each violent event. Some of the violent actions with unidentified perpetrator were identified by comparing the data with the database of violent events from the *Comisión de la Verdad.* For this paper, only the two biggest actors in the Colombian conflict will be considered: the FARC guerrilla and the paramilitary groups, which account for more than 60% of the whole violent events during the time frame between 1998 and 2018. Both of these groups demobilized in separate peace agreements during 2006 and 2016, respectively. However, the remains of these groups did not demobilize and created dissident groups. Violence committed by these dissident groups will be considered as perpetrated by the FARC or paramilitary groups.

The most complicated part of organizing the database was the classification of the type of violence. Each event was classified into several categories according to the type of violence by using the description in the dataset of each violent episode. Each of these categories is expected, according to the theory, to require a different level of armed group presence. Table 2.1 shows the six categories that will be used in the following analysis with a brief explanation, the kind of violence in that category, and the expected level of armed group presence to commit that type of violence.

Table 2.1: Type of Violent Events

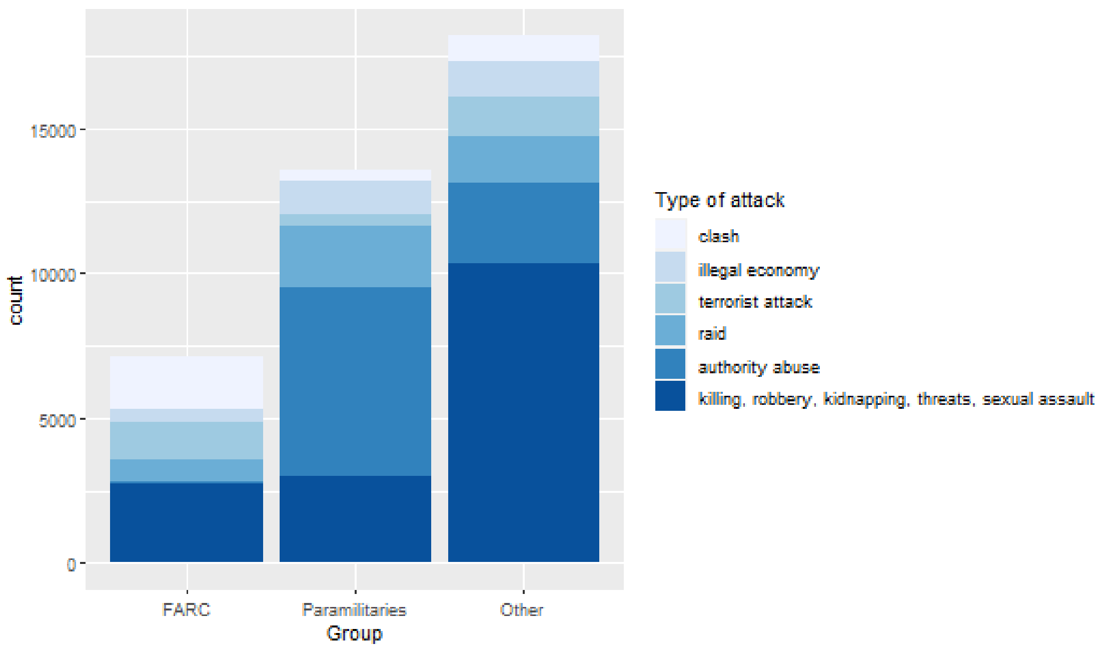
Category	Explanation	Subcategories	Expected Illegal group presence
Authority Abuse	Violation of human rights as an abuse of authority	Extrajudicial execution, arbitrary detention, territorial control, illegal patrolling, social cleansing operations	High
Raid	Army incursions to municipalities	Armed incursions, armed raids, armed invasions, takeovers of municipalities, displaced communities, threatened collectives, political violence, systematic and selective crimes, massacres, confinements, cylinder bombs against populations, illegal roadblocks, massive looting	High
Clash	Combats against two or more armed groups	Attacks, dispute, actions against other armed groups, ambushes, combat, combat, cylinder bombs against other armed groups	Medium-High
Illegal Economy	All crimes related to illicit drug trafficking, illegal mining, contraband and extortion to the population	cocaine trafficking, poppy cultivation, illicit mines, drug trafficking, illicit crops, extortion, intimidation, threats to traders or transporters, strategic drug corridors	Medium
Killings	Non-systematic crimes against civilians	Robbery, kidnapping, killings, threats, sexual assaults	Low
Terrorist Attacks	Hit and Run attacks	Attacks against infrastructure, bombs, explosive devices, use of grenades against stations and buildings, explosive charges, planned detonations	Low

The idea of this classification is to firstly depart from the binary approach to violence, i.e. terror versus restraint or terrorism versus conventional tactics, and to develop a more complex variable of violence type. Previous studies often rely on the control-collaboration model by Kalyvas (2006). According to this theory, the kind of violence depends on the territorial control by an armed group, i.e. when a rebel group can pursue, establish and defend a territory. The model posits that the use of indiscriminate violence is inversely related to the level of territorial control. Consequently, terrorism and other forms of indiscriminate violence occur when the armed group has no control over the territory, and irregular fighting, especially attacks on civilians, diminishes when the armed group controls the territory.

However, recent studies show that the type of violence depends not only on the level of territorial control (Kalyvas, 2012). Furthermore, selective violence against civilians can also occur when the territory is not controlled (Thill et al., 2017). Besides, lethal violence is not a good proxy for the overall pattern of violence and that it is necessary to analyze different forms of violence (Gutiérrez-Sanín and Wood, 2017). For these reasons, this study's proposed classification tries to capture the required level of armed group presence and not the level of territorial control.

The first of these categories is called authority abuse. Authority abuses are all violent-related episodes where an armed actor use violence to impose authority illegally. As shown in figure 2.1 most of these actions were committed by paramilitary groups, but FARC groups also inflicted some illegal authority abuses. This type of crimes have the purpose of building sanctuaries of impunity to create conditions for the reproduction of their illicit business; and local powers to maintain their authoritarian social orders (CNMH, 2018).

Figure 2.1: Violent events by type and perpetrator



Authority abuses occur due to the State's inability to replace the illegal economies that allow the armed groups to reproduce. The purpose of these types of crimes is to replace the authoritarian social order built in these regions. Examples of these types of crimes are social cleansing operations and illegal control enforcement to impose "good conduct" codes. This type of violence requires strong illegal group presence as they require strong vigilance over the population (Kaplan, 2013) and, in some cases, huge violent operatives as in the massive extrajudicial cases executions. This type of violence may have the purpose to govern the local population (Mampilly, 2011).

Authority abuse may be associated with high territorial control, but it is not always the case. Authority abuse may also be present in the absence of territorial control. When faced with competition from rival groups, insurgents are likely to become increasingly coercive toward civilians, as they are trying to maintain their control over crucial resources (Douglass, 2012). Additionally, authority abuse is not always associated with a massive level of violence. It is not always selective, and on some occasions, it is used for purposes other than curbing collaboration with the enemy (Vargas, 2009). For instance, illegal detentions do not necessarily imply lethal violence. Illegal detentions can also be indiscriminate against a sub-population-group (e.g. ethnic groups) and may not have the purpose to curb support for the enemy but to impose particular behaviour to the population. Therefore, authority abuse is associated with high armed group presence but not necessarily high territorial control.

The second classification is called raids. This type of violence is also expected to require high levels of armed group presence. The reason is that it includes all incursions to a municipality by a particular illegal armed group. These events require strong military operatives, as the main goal is to attack and occupy a municipality. This classification also includes selective crimes and massacres. Illegal armed groups often use this type of crimes to intimidate civilians into supporting their side in the battle (Kalyvas, 1999). Both paramilitary and FARC armed groups used this type of violence intensively, as shown in figure 2.1.

This classification of violence may not always imply territorial control, but it does always mean huge illegal armed presence. Selective violence, for instance, can be more likely to be used in areas primarily controlled by the enemy (Bhavnani, Miodownik, and Choi, 2011). Additionally, municipality takeovers can last a few hours or days before the illegal armed group leaves the municipality. This type of violence has also been frequent in areas where armed groups had little control and places where the two sides approached parity (Vargas, 2009).

The third classification includes all types of clashes. These are mainly all violent events related to combats or attacks against another armed group. This type of events may require medium or high armed group presence that depends on the combat or attack's intensity. Clashes are more the type of conventional violence present in an armed conflict. Commonly, clashes occur as a mean for expanding and establishing territorial control (Calle and Sánchez-Cuenca, 2015), and usually require more fighters and arms than other irregular tactics but less than a raid.

The fourth classification incorporates all crimes related to illegal drug trafficking, illegal mining, contraband and extortion. Economic activities such as the contraband of stolen petrol are regular in territories with high or medium armed group presence. The reason to associated these crimes with medium armed group presence is that when the armed group successfully commit these crimes, these crimes usually are not reported (e.g. drug trafficking without the authorities' knowledge). Also, it is less likely that these crimes are notified when there are no negative spillovers of violence (e.g. killings associated with drug trafficking). Armed groups are more successful in the illegal economy when they have a high presence in the territory. Instead, when the illegal activity is harder to perform, crimes related to drug trafficking or other unlawful activity occur, and individuals committing these crimes or illicit activities are captured, reported, or denounced.

The fifth category is named killings but encompasses all crimes such as robbery, killings, sexual assaults, threats and kidnappings. Non-systematic killings of civilians do not require substantial armed group presence (Tyner, 2016). Furthermore, these crimes can occur with few armed group members in the territory and often operating clandestinely. It is essential to notice that these crimes can also happen during an authority abuse, a raid, or related to the illegal economy. Still, in such cases, the main event is the raid, illegal economic or authority abuse, and the event is not duplicated. Most of these crimes are again not related to the level of territorial control. While it can be the case that, when an armed group controls the territory, these crimes are reduced. It can also be the case that armed groups rely on crimes such as kidnappings and expropriation of assets in controlled territories (Vargas, 2009).

Finally, terrorist attacks are the last category. These are all violent events related to terrorism and indiscriminate violence. Most of these are hit-and-run attacks. While these types of events can be sophisticated, as in large-scale bombs, they do not require a strong armed presence in the territory attacked. The intelligence and planning occur typically outside the territory. Also, terrorism is associated with low territorial control levels and low armed group presence (Carter, 2015).

2.3 Data

The data used in this paper will consist of every violent event committed by paramilitary groups or by the FARC between 1998 and 2018. The data will be analyzed at the municipality level and grouped by each quarter of a year. The data will contain the number of days during each quarter that each municipality experienced each type of violent event and the number of days having no violent events.

The following two maps, Figures 2.2 and 2.3, show the territorial distribution of the violent events. The blue colours denote the total number of events in a territory committed by the paramilitary groups or FARC. The dots describe the type of events in each of the municipalities in Colombia. As shown, there is not a clear distribution of violent events across the territory. In some cases, there can be a strong armed presence in one municipality and none in the nearby municipality. Figure C.1 in the appendix also shows the number of events from paramilitary or FARC groups by type and year for every municipality.

Figure 2.2: Violent events from paramilitary groups

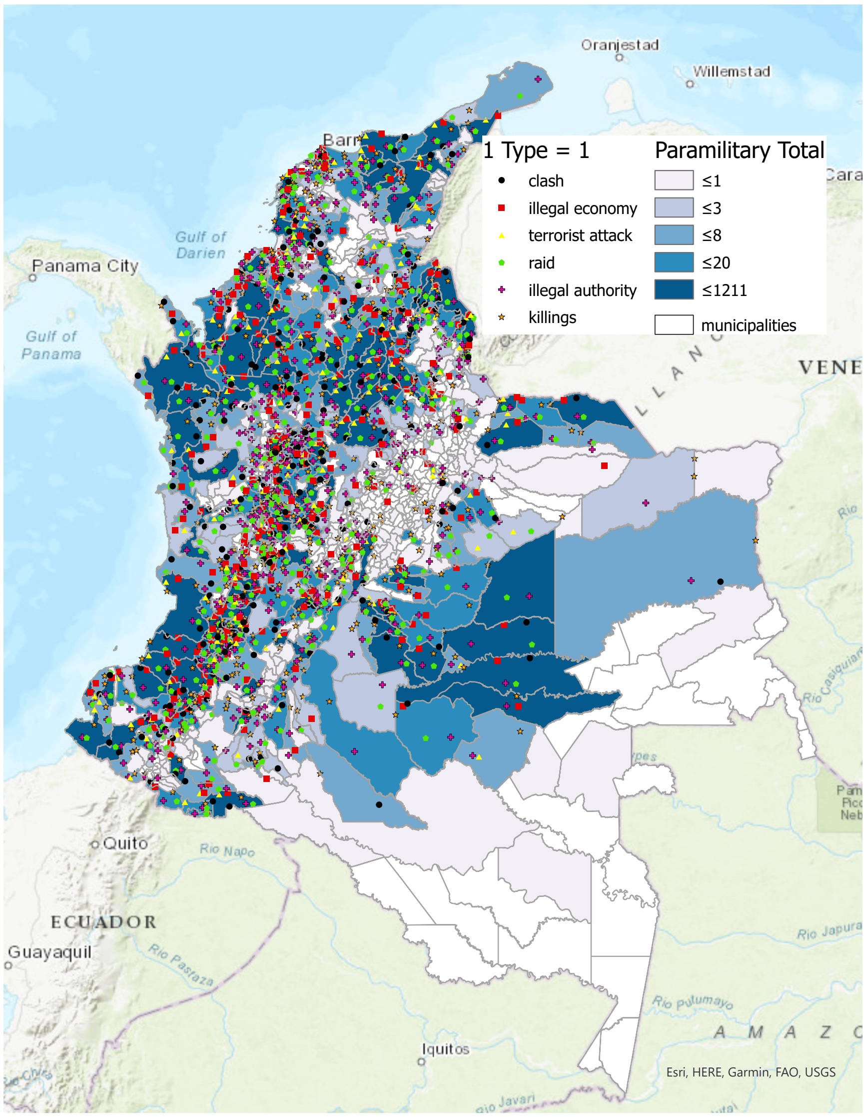
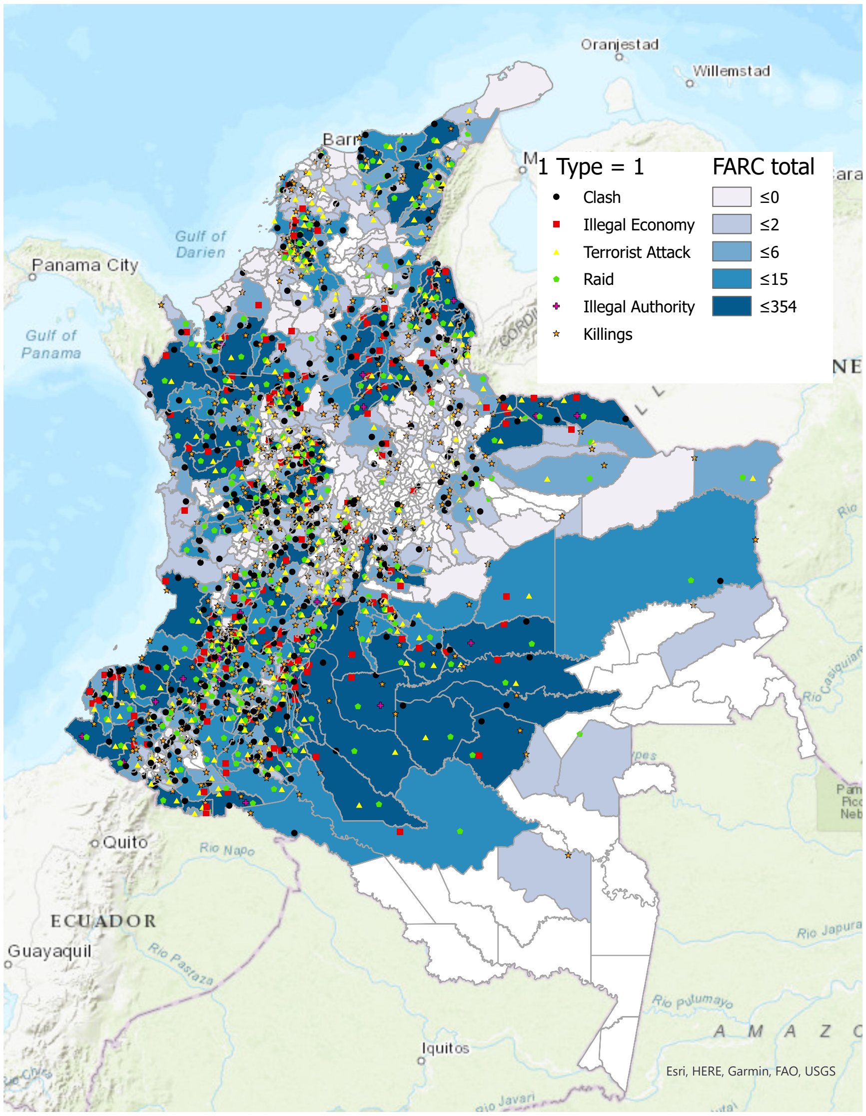


Figure 2.3: Violent events from FARC



One characteristic of the data is that it is auto-correlated for every type of event within each municipality. If one municipality experiences a specific kind of violent event, the violent event is likely repeated in the next quarter. For this reason, a HMM is useful at it considers this possible auto-correlation. Figure C.2 in the appendix shows the autocorrelation plots for each type of event.

Finally, tables 2.2 and 2.3 show transitions from one type of event to another. Each number represents the number of occasions a municipality transitions from experiencing a certain type of violent event in a quarter to another kind or having no events in the next quarter. As shown in the tables, it is prevalent to see municipalities continue having the same event type in one period and the following next period. It is less common to see transitions from events that require a high level of armed presence to events that require less armed presence. For example, there are zero cases of municipalities experiencing authority abuse to experiencing terrorist attacks in the following period. Moreover, there are many periods of times not occurring any violent event, which also needs to be addressed for the posterior analysis.

Table 2.2: Transition of violent events (paramilitary groups)

	To						
from	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	None
Authority Abuse	799	101	52	58	166	0	1431
Raid	92	150	18	26	35	6	429
Clash	55	28	46	9	24	3	206
Illegal Economy	53	22	10	44	50	4	323
Killings	119	47	29	48	170	2	766
Terrorist Attacks	2	7	3	3	7	23	96
None	1358	403	208	318	755	91	71405

Table 2.3: Transition of violent events (FARC)

	To						
from	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	None
Authority Abuse	40	10	12	12	13	0	18
Raid	6	86	30	4	44	20	306
Clash	6	33	146	20	107	22	692
Illegal Economy	1	35	20	52	32	17	167
Killings	2	25	102	44	228	70	1030
Terrorist Attacks	0	18	41	6	69	74	436
N	16	279	691	159	960	418	73694

2.4 Methodology

A HHM consists of an unobserved parameter process $\{P_t : t = 1, 2, \ldots\}$ satisfying the Markov property on the conditional probabilities, called the transition probabilities,: $\Pr(P_{t+1}|P^{(t)}) = \Pr(P_{t+1}|P_t)$. Armed group presence, P_t, could be modelled as a discrete state with several m categories ranging from null armed group presence to strong-armed group presence and satisfying the Markov properties. Modelling armed group presence as a Markov process is realistic as the level of armed group presence in time t highly depends on the level in $t-1$ but not necessarily in $t-k$. Figure 2.4 shows how the random variable of illegal group presence could behave in a way that past and future are dependent only on the present.

Figure 2.4

The HMM methodology allows the hidden process characteristics, such as the transition probabilities from one state to another, to be estimated from the directly observable data. The Markovian property of the hidden process can accommodate serial dependence between observations collected in successive time intervals.

A HMM also includes a state-dependent process $\{X_t : t = 1, 2, \ldots\}$ in which the distribution of X_t depends only on the current P_t and not on previous states or observations. In this study, X_t represents violent actions committed by armed groups, which can also be categorized according to the type of violent action. The q categories can be classified into authority abuse, clashes, crimes related to the illegal economy, raids, killings, robberies, kidnappings, threats and sexual assaults, terrorist attacks and no violent event. As explained above, each of these actions is more likely depending on the level of armed group presence.

Consequently, a HMM includes $X^{(t)}$ and $P^{(t)}$ representing the histories from time 1 to time t and where: $\Pr(P_t|P^{(t-1)}) = \Pr(P_t|P_{t-1})$, $t = 2, 3 \ldots$ and $\Pr(X_t|x^{(t-1)}, P^{(t)}) = \Pr(X_t|P_t)$, $t \in \mathbb{N}$. The first probability governs the level of armed group presence transition. The probabilities of transition gives you the probability from going from the different m states of armed group presence to another. For instance, the probability from having a territory with high armed group presence H to low presence L, $\Pr(P_t = H|P_{t-1} = N)$, would be expected to be a very low probability. The second probability is that of guessing each level of armed group presence from the type of violent action. The probability of observing a particular type of attack from a level of armed group presence is known as emission probability. The structure is represented in figure 2.5.

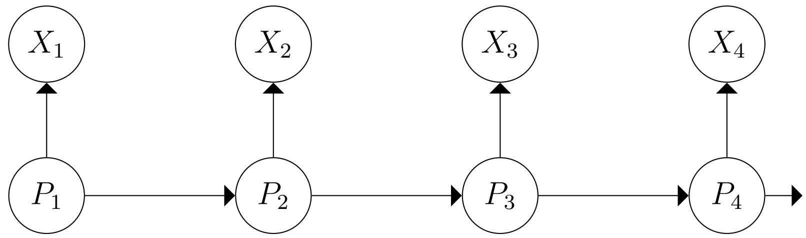


Figure 2.5

In this way, this research posits that a HMM can uncover the level of armed group presence in a territory by assuming that armed group presence is a hidden factor underlying the general trend of violent actions by armed groups. Therefore, the hidden factor can be discovered from the data of violent actions by armed groups.

The previous simple HMM needs to be extended to estimate the model correctly to account for the counts of days with violent actions by armed groups in each category in a given time and territory. If we consider time frames of a quarter, we could likely observe, for example, more than one attack or more than one clash in a given territory. Therefore, X_t is no longer a categorical series but a series of counts of days with violent actions o no violent actions by armed groups in the q category: $\{x_{jt} : t = 1, 2 \ldots T; j = 1 \ldots q\}$ with $\sum_{j=1}^{q} x_{jt} = n$. For this reason, a multinomial-HMM could be a better model to account for this fact. A multinomial-HMM could be more suited because our interest lies in the count of days with various types of violent actions by armed groups in the territory within a time frame of a quarter. The emission probability would be: $\Pr(X_t = x_t|P_t = i) = \binom{n}{x_1 \ldots x_{tq}}\pi_{i1}^{x_{t1}} \ldots \pi_{q1}^{x_{tq}}$.

The likelihood function for such HMM is given by $L_T = \delta P_1(x_1)\Gamma P_2(x_2) \ldots \Gamma P_T(x_T)1'$ where Γ is the transition matrix , δ is the initial state probability vector, and $P_t(X_t) = \mathrm{Diag}(\Pr(X_t = x_t|P_t = i), \ldots \Pr(X_t = x_t|P_t = m))$. The parameters of the model can then be estimated by maximizing the likelihood as a function of $m(q-1)$ of the $m(m-1)$ diagonal transition probabilities and of the success probabilities, π_{ij} for $j = 1, 2, \ldots q-1$ and all i. The likelihood maximization is done using the EM algorithm (see Zucchini, Macdonald, and Langrock (2017) and Visser and Speekenbrink (2010)). Confident intervals can also be obtained using the finite-differences approximation (Visser, Raijmakers, and Molenaar, 2000) (see Appendix).

Chapter 3

Results

3.1 First Estimations

The first model fitted is a simple multinomial-HMM with three states. The Multinomial-HMM will be used to predict armed group presence at each municipality in each quarter. There are 24 parameters to be estimated: six transition probabilities of specifying the Markov chain and six probabilities for each of the three states.

The number of states, or the HMM order, were defined by comparing the Akaike information criterion (AIC) and Bayesian information criterion (BIC) of different models with different m states. Tables C.1 and C.2 in the appendix display the AIC and BIC values for the models using paramilitary-related events or FARC-related events. From the tables, we see that of the four models considered, the three-state model is chosen by both the AIC and BIC in both paramilitary and FARC events. Each state will represent a level of armed group presence: Level 1, low armed group presence, Level 2 medium armed group presence and Level 3 high armed group presence.

Table 3.1 display the starting values for the parameters of the response models. These initials values reflect the theoretical expected π's in the emission probabilities of the multinomial distribution for each of the three states. The parameters will be estimated separately for both violent events committed by paramilitary groups and FARC. As displayed in the table, it is expected that no days with violent events might occur with higher probability in all states as it is not common to see that many violent events in most of the days. However, when being in a higher state or higher level of armed group presence, this probability is expected to diminish. The opposite is expected for the probability for the number of days with authority abuse in each quarter. A higher π is expected in state 3 than in state 2 or 1. In the state with middle armed group presence, state 2, a higher probability is expected for violent events related to the illegal economy or killings.

Table 3.1: Initial π's

	π_i						
State	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	No event
St1	0.001	0.002	0.02	0.035	0.04	0.08	0.80
St2	0.03	0.04	0.08	0.15	0.13	0.02	0.55
St3	0.2	0.15	0.09	0.10	0.05	0.01	0.40

The matrix below also shows the starting values for the parameters of the transition models. It is expected that there is a low transition between states. For that reason, a probability of 0.9 is expected when remaining in the same state among quarters and only a probability of 0.1 to transit to the next state.

$$\begin{pmatrix} 0.9 & 0.1 & 0 \\ 0.05 & 0.9 & 0.05 \\ 0 & 0.1 & 0.9 \end{pmatrix}$$

3.1.1 Paramilitary presence

The following results show the estimated parameters when using paramilitary-related events. The initial state probabilities are shown in the matrix δ below. Initial state probability estimates indicate that state 1 is the starting state for the process, and the probabilities are lower of starting in state 2 and even lower in state 3. The 95% confident intervals of all further estimations are shown in the Appendix. For this model, the log-likelihood is -1405.69.

$$\delta = \begin{pmatrix} 0.628 & 0.301 & 0.071 \end{pmatrix}$$

The estimated probabilities of the transition matrix reflect what I anticipated about the transition among states. There is a high probability of remaining in the same state. However, this probability is lower when being in a higher state, which means a more paramilitary presence.

$$\Gamma = \begin{pmatrix} 0.837 & 0.163 & 0.000 \\ 0.155 & 0.618 & 0.227 \\ 0.000 & 0.331 & 0.669 \end{pmatrix}$$

Finally, the estimated response parameters are shown in table A.1. Conditionally on being on state 1, low paramilitary presence, the probability of having all days of the quarter with no violent days is almost 99%. Compare to state 1, the probability of having all days with no violence is only 87% when having a high paramilitary presence. The estimated π_j's for illegal economy and killings when being in state 2 is 0.013 and 0.035, respectively, much higher than the other two states' estimated parameters. The estimated parameters also reflect that when being in a state with a high paramilitary presence, the occurrence of authority abuse, raids and clashes is very high compared to the other two states. The occurrence of terrorist attacks and killings is lower when having a low paramilitary presence. However, conditionally to the occurrence of any violent event, the probability of having terrorist attacks or killings is much higher when being in state 1.

Table 3.2: Estimated π's

	π_i						
State	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	No event
St1	0.001	0.0009	0.0006	0.003	0.003	0.0004	0.991
St2	0.016	0.001	0.006	0.016	0.035	0.001	0.925
St3	0.096	0.004	0.009	0.003	0.014	0.002	0.872

The following map in figure 3.1 shows the predicted states in each municipality from 2005 to 2009 based on all quarters' mode in these four years. During this period, most of the main paramilitary blocks demobilized, but many of these groups persisted, as explained above. The darker blue indicates a higher level of paramilitary presence in that municipality. The patterns point out the strong paramilitary presence in southern Meta (south-east) where the Bloque Centauros was established; in Uraba and Magdalena Medio (north-east) where the Bloque Córdoba and Autodefensas Campesinas del Magdalena Medio operated; in the northern part of the country where the Bloque Norte had their prominent presence; in the region of Catatumbo (north-east) where the Bloque Catatumbo was consolidated, and in Valle del Cauca (south-west) where the Bloque Calima operated.

Also, some data from experts coming from López Hernández and Ávila Martínez (2010) can be used to compare the predicted level of paramilitary presence. Table 3.3 displays the number of municipalities between 2005 and 2009 according to the predicted states using the HMM and according to the expert classification. Unfortunately, the expert data has only two categories. The table shows that 69% of the municipalities are classified as having no paramilitary presence, 2.8% have paramilitary presence according to the expert data but not according to the HMM, and 1.7% have high paramilitary presence according to the HMM but not according to the expert data. In general, the state predictions are very good when comparing them to the expert data. The HMM seems to show some more municipalities with a medium paramilitary presence that experts may consider as having no paramilitary presence. Expert data, with only two categories, is more rigorous when deciding when a municipality has a paramilitary presence.

Figure 3.1: Predicted paramilitary presence

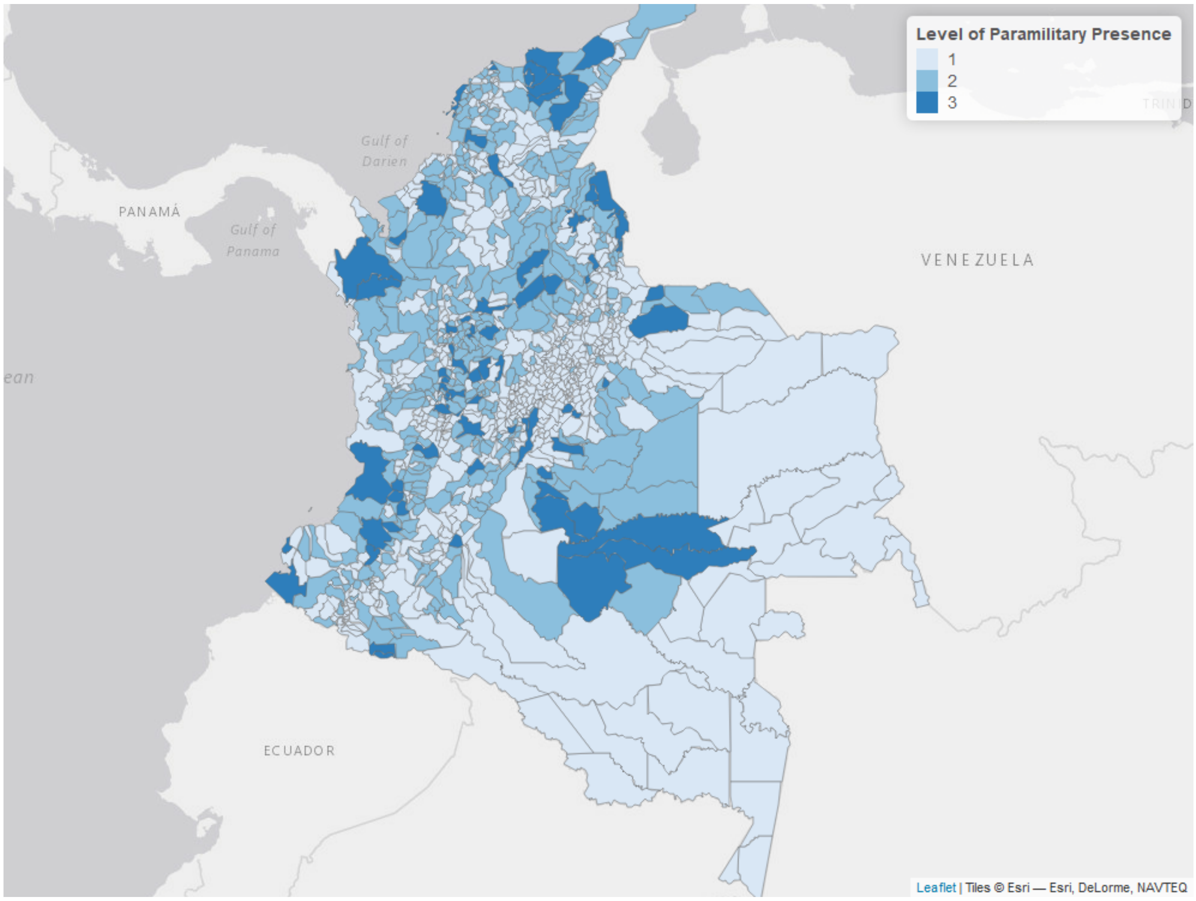


Table 3.3: Comparing paramilitary presence with expert data

	Expert Data	
HMM	None	Presence
Low	769 (69.9%)	31 (2.8%)
Middle	169 (15.3%)	182 (16.6%)
High	19 (1.7%)	29 (2.6%)

3.1.2 FARC presence

The exercise is repeated but now using only violent events caused by members of the FARC guerrilla. A multinomial-HMM with three states is fitted again with the same violent event classification by counting the number of days during each quarter and municipality. The initial state probabilities are shown in the vector below:

$$\delta = \begin{pmatrix} 0.596 & 0.315 & 0.089 \end{pmatrix}$$

Similarly to the HMM with paramilitary data, the first state has a higher initial probability, and state 3 has the lowest. The estimates for the transition matrix, Γ, are similar to the one of the transition matrix estimated with the paramilitary violent events showing a high probability of remaining in the same state. However, the transition probabilities from one state to another are higher for the FARC presence. These transitions make sense as the FARC were more mobile than the paramilitaries, and during the period studied, the FARC were hit harder by public forces. The Log-likelihood of this model was -832.38.

$$\Gamma = \begin{pmatrix} 0.859 & 0.141 & 0.000 \\ 0.377 & 0.439 & 0.185 \\ 0.000 & 0.298 & 0.702 \end{pmatrix}$$

Finally, the estimated π's in table A.2 show some expected results. In situations with low FARC presence, the probability of having many days with no violent events is high, 99%. But conditionally on being on state 2 or 3, higher FARC presence, this probability is lower. As the FARC did not commit that many crimes related to authority abuse, we see an estimated

π for authority abuse close to zero and greater for state 2. However, the estimated π for raids are greater when having a high FARC presence. The same occurs for clashes. Again, as happened with the paramilitary data, the probability of experiencing killings conditionally on any violent event happening is higher when being in state 1, low FARC presence. With FARC violent events, terrorist attacks are more common conditionally on experiencing any violent event in state 2. The reason may be that the FARC committed much more terrorist attacks than the paramilitary groups, and most of them were against public infrastructure and occurred in territories where they had more presence. Contrary to paramilitary groups, the FARC main goal was to overthrow the State and not capture and permeate the public institutions. So it makes sense that terrorist attacks were common in places where they had a significant presence.

Table 3.4: Estimated π's

				π_i			
State	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	No event
St1	0.0000	0.0005	0.0003	0.0002	0.0006	0.0003	0.999
St2	0.0003	0.0012	0.0008	0.0027	0.0036	0.0073	0.984
St3	0.0000	0.0429	0.0022	0.0009	0.0084	0.0035	0.942

Finally, the map in figure 3.2 shows the predicted FARC presence from 2005 to 2009. The main FARC blocks are clearly shown in a darker blue: the Bloque Caribe in the north of the country, Bloque Oriental in the south-eastern part of the country, Bloque Noroccidental in the north-western part of the country, Bloque Magdalena Medio in the north-eastern part of the country, and Bloque Sur in the south-eastern part of the country.

The comparison with expert data in table 3.5 shows that the model correctly predicted, according to the expert data, the low FARC presence in most municipalities and the high FARC presence in some of the municipalities. However, 10% of the municipalities were predicted as having a medium FARC presence, but the expert data classify those municipalities as having low FARC presence. Again the reason for this discrepancy is the use of only two categories by the expert data. Therefore, experts classify only municipalities with a solid FARC presence.

Figure 3.2: Predicted FARC presence

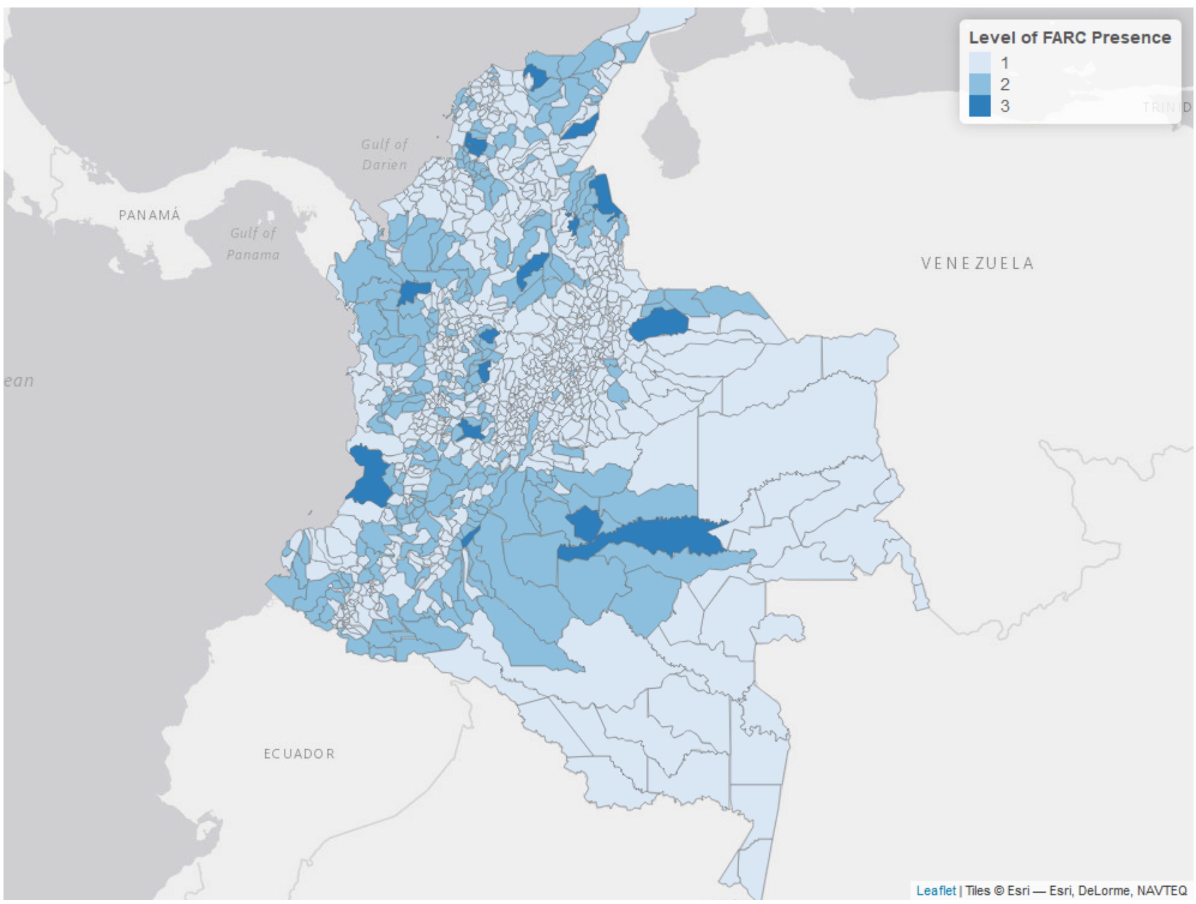


Table 3.5: Comparing FARC presence with expert data

	Expert Data	
HMM	None	Presence
Low	937 (85.2%)	2 (0.1%)
Middle	118 (10.7%)	11 (1.0%)
High	11 (1.0%)	20 (1.8%)

3.2 Including Covariates

Another aspect to consider when identifying armed group presence through HMM is the potential inclusion of covariates, which values are regarded as given. It is possible to include covariates by allowing some of the state-dependent distribution parameters to depend on covariates. For instance, cocaine production in the Colombian context is a huge predictor of violent actions by armed groups. Figure 3.3 shows a HMM including covariates.

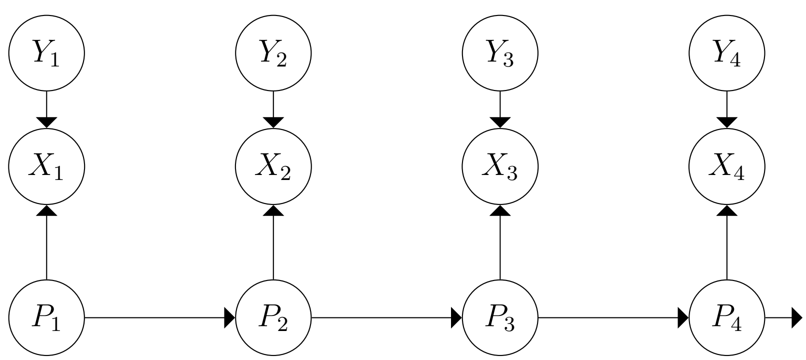


Figure 3.3

Previous studies of the Colombian conflict have demonstrated several factors that affect the occurrence and the type of violent events. For instance, Dube and Vargas (2013) present micro-level evidence that commodity price shocks of coffee and oil increase the severity of rebel violence in some municipalities. They show that a sharp fall in coffee prices increased violence differentially in municipalities cultivating more coffee. In contrast, a rise in oil prices increased municipal revenue and violence in oil regions. For that reason, the following model includes coffee and oil price shocks as variables influencing the level of violence.

Wright (2015) demonstrates that the level of violence and the type of violence are affected by these economic shocks. When governments benefit from economic conditions, insurgents favour irregular tactics, such as terrorist attacks. When rebels are strengthened by revenue booms, for instance, from cocaine shocks, they prefer conventional tactics, such as raids and clashes. The reason is that as the cultivation of coca bushes expands and market prices for cocaine rise, local rebel income surges, permitting sub-units to better arm and recruit fighters. In contrast, if coffee prices rise, non-combatants have economic options outside of joining illegal armed groups. Resource gains by illegal armed groups may also lead to public goods provision, increasing sympathy for illegal armed groups, manifested in higher violent events such as authority abuse (Wood, 2003). As for oil shocks, counterinsurgent capacity increases with positive oil shocks directly—through the allocation of royalties and tax income to fight rebel groups.

Weinstein (2007) also posits that organizations with access to economic endowments engage in broader repertoires and targeting and higher frequency of violence against civilians than those with social endowments. As a result, groups with economic endowments from cocaine resort to all forms of violence at high levels with little discrimination, such as raids and clashes, while armed groups with high social endowments resort to low levels of highly selective, primarily lethal violence.

To investigate explanatory variables y'_t for the emission probability, we can use a link function with a multinomial logit in the following way: $\log \frac{\Pr(X_{mt}=j|P_{mt}=i)}{\Pr(X_{mt}=r|P_{mt}=i)} = \beta_{ij} y'_{mt}$, where j is the one of the possible q categories of violent events, r is the reference category, m is the municipality, and i is one of the possible three states of illegal group presence. For simplicity, X_t is no longer a series of counts but a categorical series. For every municipality in quarter t, the municipality is classified into seven exclusive categories: authority abuse, clashes, crimes related to the illegal economy, raids, killings, robberies, kidnappings, threats and sexual assaults, terrorist attacks or no event.[1].

There are four covariates included: first, the coffee shock, which is the interaction of the natural log of the internal coffee price and land devoted to coffee production; second, oil shock denoted by the interaction of natural log of oil prices and oil production level; third, cocaine shock interacting cocaine crops with the natural log of cocaine prices; and finally, a variable of the distance to the nearest capital city which could serve as a proxy for strong state presence. I draw on data collected by Dube and Vargas (2013) and data from *Centro de Estudios sobre el Desarrollo Económico* (CEDE).

$$\beta_{ij} y'_{mt} = \alpha_{ij} + \beta_{1ij} \text{Coffee}_m \times \text{CoPr}_{mt} + \beta_{2ij} \text{Oil}_m \times \text{OiPr}_{mt} + \beta_{3ij} \text{Coca}_{mt} \times \text{CocPr}_{mt} + \beta_{4ij} \text{Dist}_{mt}$$

3.2.1 Paramilitary presence

The results of using violent events committed by paramilitary groups and including covariates are presented in this section. The initial state probabilities δ show that state 1 has the

[1]The classification is done by arranging municipalities in each quarter according to the number of events in each of the categories. The municipality is classified according to the event requiring more illegal group presence that occurred during that quarter.

highest probability for starting the process, but it is lower than the estimated probability of the HMM without covariates. The estimated probabilities of the transition matrix, Γ, reflect a similar pattern as before: low transition probabilities between different states and a high probability of remaining in the same state. The log-likelihood for this model was -947.04

$$\delta = \begin{pmatrix} 0.512 & 0.253 & 0.235 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 0.936 & 0.064 & 0.000 \\ 0.071 & 0.774 & 0.155 \\ 0.000 & 0.118 & 0.882 \end{pmatrix}$$

Table 3.6 shows all the coefficients estimated from the emission probability with the confident intervals in parenthesis. Each row shows the coefficients in the logit function for each type of violent event with respect to the category of "No violent event" and, separately, for each of the three states. Each column displays each of the coefficients for each of the four covariates.

When the predictor variables in the model are evaluated at zero, the probability of having no violent events are 99%, 91%, and 33% conditionally on being on state 1, 2, and 3, respectively. The multinomial log-odds of experiencing authority abuses to experiencing no violent events would be expected to decrease by 11.9, 0.3, and 0.06 for one unit increase in the coffee shock (1.02 standard deviations) conditionally on being on state 1, 2 or 3, respectively. As predicted by the theory, these results show that we expect less violent events, in this case for authority abuses, when there is a positive shock of coffee. However, the effect is lower when having a strong paramilitary presence. Similar results are suggested for raids, clashes and crimes related to the illegal economy, as shown in the coefficients in the second column of the table. However, the coefficients are positive when being in state 3, high paramilitary presence, for illegal economy and raids. The contrary occurs for killings and terrorist attacks: a positive shock of coffee increases the log odds of experiencing those types of events to experiencing no violent events by 11.6 and 11.001, respectively, when having a low paramilitary presence. Still, the effect is much lower when having a high paramilitary presence and negative when having a medium paramilitary presence.

The coefficients in the column for oil shocks reflect that, in general, increases in oil prices are associated with higher odds of experiencing violent events in municipalities with oil production. However, in most cases, the impact is more significant when having a medium paramilitary presence. The coefficients show different patterns for terrorist attacks and crimes related to the illegal economy: a unit increase in oil shocks is associated with lower odds of these types of events when having a medium or high paramilitary presence.

For cocaine shocks, the coefficients show that a positive shock in cocaine prices in those municipalities with cocaine crops are associated with higher odds of experiencing violent events. The impact is stronger in municipalities with a lower paramilitary presence. This result reflects that when there are positive shocks of prices of cocaine, the effect of violence is powerful in municipalities with cocaine crops but with and low paramilitary presence. The reason could be the possible expansion of these groups during these positive shocks. Finally, the coefficients for distance to the nearest capital reflect that those municipalities further away from capital cities suffer more from different type of violent events when there is a low paramilitary presence. When there is a high or medium paramilitary presence, distance does not affect that much the occurrence of violent events.

Table 3.6: Estimated coefficients (No event as reference level)

State	Type of Event	α	Coffee×Log Coffee Price	Oil×Log Oil Price	Coca×Log Coca Price	Distance
1	Authority Abuse	-19.966	-11.924	6.04	3.948	0.886
		(-20.942,-19.980)	(-12.771,-11.072)	(3.941,8.141)	(-0.135,8.031)	(-0.422,2.194)
2		-3.304	-0.300	12.42	0.022	-0.003
		(-4.772,-1.836)	(-0.594,-0.006)	(11.064,13.776)	(-0.043,0.087)	(-0.158,0.152)
3		0.056	-0.059	5.09	0.033	-0.004
		(0.017,0.095)	(-0.100,-0.018)	(3.818,6.362)	(0.021,0.045)	(-0.114,0.107)
1	Raid	-18.019	-11.692	2.72	3.896	0.882
		(-18.669,-17.369)	(-12.890,-10.494)	(0.502,4.938)	(-1.672,9.464)	(-0.347,2.111)
2		-4.583	-0.221	7.503	0.008	-0.001
		(-5.843,-3.323)	(-0.302,-0.140)	(-0.489,15.495)	(-0.007,0.023)	(-0.176,0.174)
3		-1.63	0.035	2.116	0.031	-0.002
		(-2.751,-0.509)	(0.028,0.042)	(-1.349,5.581)	(0.023,0.039)	(-0.193,0.189)
1	Clash	-19.88	-11.788	-1.291	3.859	0.885
		(-21.530,-18.231)	(-10.606,-12.970)	(-1.803,-0.779)	(3.341.4.378)	(-0.472,2.240)
2		-5.542	-0.461	10.175	0.002	-0.001
		(-6.509,-4.575)	(-1.187,0.265)	(9.402,10.948)	(0.001,0.003)	(-0.183,0.181)
3		-2.598	-0.615	-5.198	0.008	0.002
		(-3.533,-1.663)	(-1.304,0.074)	(-6.240,4.156)	(0.003,0.013)	(-0.206,0.202)
1	Illegal Economy	-12.202	-11.157	5.595	3.968	0.89
		(-13.938,-10.466)	(-13.938,-8.376)	(4.087,7.103)	(-0.250,8.186)	(-0.384,2.164)
2		-4.812	-0.358	-2.934	0.014	-0.001
		(-6.802,-2.822)	(-1.176,0.460)	(-4.781,-1.087)	(-0.065,0.093)	(-0.203,0.202)
3		-1.985	0.697	-5.195	0.025	-0.002
		(-2.243.-1.727)	(0.479,0.915)	(-6.645,-3.745)	(0.018,0.032)	(-0.237,0.234)
1	Killings	-12.055	11.589	7.78	3.864	0.86
		(-13.980,-10.131)	(9.928,13.251)	(5.259,10.303)	(-0.240,7.968)	(-0.293,1.753)
2		-3.681	-0.448	10.895	0.024	-0.003
		(-5.554,-1.808)	(-0.519,-0.377)	(9.113,12.677)	(-0.064,0.112)	(-0.291,0.285)
3		-0.792	0.222	2.847	-0.079	-0.004
		(-1.479,-0.105)	(0.039,0.405)	(-0.374,6.068)	(-0.161,0.002)	(-0.307,0.299)
1	Terrorist Attacks	-12.608	11.001	5.694	3.837	0.857
		(-14.484,-10.732)	(8.559,13.443)	(-0.365,11.753)	(-0.948,8.622)	(-0.302,2.016)
2		-5.329	-0.503	-1.363	0.036	-0.005
		(-6.874,-3.784)	(-0.670,-0.336)	(-3.231,0.505)	(-0.019,0.009)	(-0.288,0.278)
3		-3.554	0.029	-4.473	-0.033	-0.003
		(-4.667,2.441)	(-0.016,0.074)	(-9.334,0.388)	(-0.117,0.051)	(-0.216,0.210)

Figure 3.4 displays the map for the predicted state in all municipalities in Colombia during the period between 2005-2009. Again it shows the mode during all the quarters of that period. Compared to the previous HMM with no covariates, municipalities in state 2 or 3 are more common. The reason may be given by lower transition when being in state 2 or 3. If a municipality has a strong or medium paramilitary presence and has some quarters with no violent events, it is still considered a high or medium paramilitary presence. Table 3.7 shows the comparison with expert data. There are 33.7% of municipalities in the state with a medium paramilitary presence, but experts consider them not having a paramilitary presence. Only around 3% of municipalities are located in low paramilitary presence by the HMM and high paramilitary presence by the expert data or the other way around.

Figure 3.4: Predicted paramilitary presence using covariates

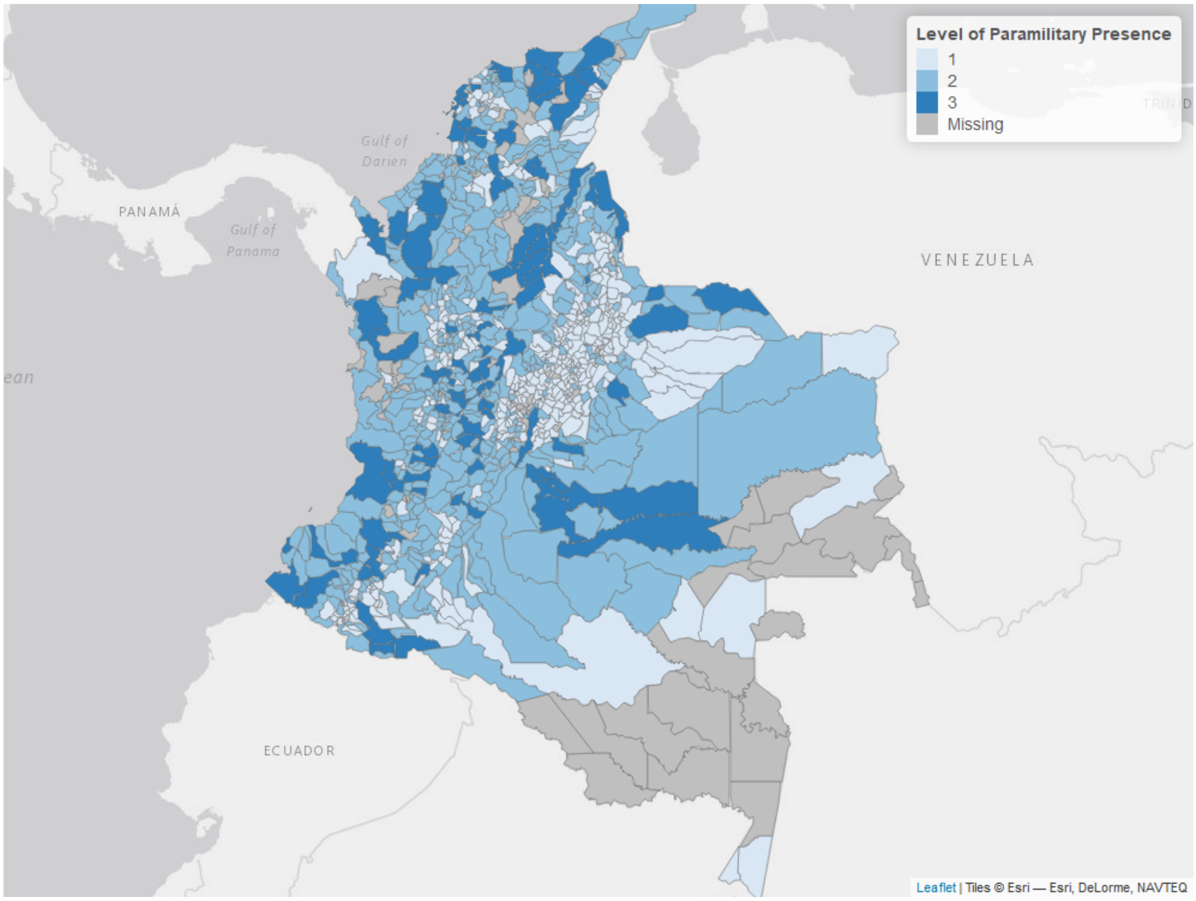


Table 3.7: Comparing paramilitary presence with expert data

	Expert Data	
HMM	None	Presence
Low	422 (41.3%)	16 (1.5%)
Middle	345 (33.7%)	126 (12.3%)
High	15 (1.4%)	98 (9.6%)

3.2.2 FARC presence

The same covariates are included in the HMM for the data with FARC violent events. The initial state probabilities are shown below. In contrast to previous estimations, the initial probabilities of state 2 or 3 are much higher. The estimated transition probabilities in matrix Γ also show a very low transition between states, much lower than the HMM with no covariates. Therefore, a municipality with a high FARC presence is expected not easily to transit to another state. The log-likelihood is -771.23.

Initial state probabilities

$$\delta = \begin{pmatrix} 0.417 & 0.329 & 0.253 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 0.978 & 0.022 & 0.000 \\ 0.131 & 0.853 & 0.016 \\ 0.000 & 0.065 & 0.935 \end{pmatrix}$$

Estimated coefficients are shown in table 3.8 with confident intervals in parenthesis. The intercepts α's show that when the predictor variables in the model are evaluated at zero, the probability of having no violent events is 99% 96% 64% for state 1, 2 and 3, respectively. This result reflects that having a high FARC presence implies a high probability of experiencing violent events.

We see a similar pattern of the coefficients for the coffee shocks as with the paramilitary data. A municipality with suitable land for coffee planting has lower probabilities of presenting violent events of any type when there is a positive shock of coffee price. However, the effect of coffee shocks seems to be much higher in territories with no FARC presence. For example, a unit increase in the coffee shock decreases the log odds of experiencing crimes related to the illegal economy, killings and terrorist attacks in 21.7, 3.9, and 5.7, respectively. However, conditioned on having a strong FARC presence, the log odds of crimes related to the illegal economy, killings and terrorist attacks increase by -0.65, 0.7 and 0.3, respectively. These results suggest that the predicted theory about more unconventional tactics, such as terrorism and killings after positive coffee shocks, occurs mainly in municipalities with high paramilitary presence.

As predicted by previous studies, oil shocks have positive associations with the odds of presenting different types of violence. The positive association, however, is higher in municipalities with a low FARC presence. This association means that municipalities with oil production are affected by violence of any type, especially if they have a low FARC presence. However, in contrast to the theory, unconventional tactics do not increase more relative to other conventional tactics after a positive oil shock with any FARC presence.

The coefficients in column 4 for cocaine shocks show that an increase in the interaction of cocaine crops with the log of cocaine prices is associated with positive increases in the odds of violence of all types. According to the theory, conventional war tactics, such as raids and clashes, increase more relatively to irregular tactics such as terrorism. But these results suggest that it mainly occurs in municipalities with a high FARC presence. Terrorist attacks are reduced in municipalities with a high paramilitary presence after these shocks. Finally, the coefficients for the distance to the nearest capital city show that distance is associated with higher odds of violence but only in municipalities with low FARC presence, except for crimes related to authority abuses.

Table 3.8: Estimated coefficients (No event as reference level)

State	Type of Event	α	Coffee×Log Coffee Price	Oil×Log Oil Price	Coca×Log Coca Price	Distance
1		-8.595	-12.023	10.337	0.031	-0.026
	Authority Abuse	(-10.039,-7.151)	(-23.823,-0.223)	(0.987,19.687)	(-0.503,0.565)	(-0.213,0.161)
2		-5.469	-11.756	7.705	0.042	-0.003
		(-6.451,-4.487)	(-21.874,-1.638)	(0.217,15.193)	(-0.598,0.682)	(-0.325,0.319)
3		-3.073	-5.38	0.101	-0.006	0.003
		(-4.243.-1.903)	(-8.865,-1.795)	(-0.645,0.847)	(-0.005,0.017)	(-0.247,0.253)
1		-10.36	-9.091	1.101	0.028	0.007
	Raid	(-11.133,-9.587)	(-17.536,-0.646)	(-0.092,2.294)	(0.005,0.051)	(-0.402,0.418)
2		-5.394	-4.884	0.797	0.006	0.002
		(-6.733,-4.055)	(-9.419,-0.349)	(-0.025,1.619)	(-0.131,0.143)	(-0.185,0.189)
3		-2.707	-0.612	0.067	0.008	0.007
		(-4.083,-1.331)	(-1.474,0.250)	(-0.489,0.623)	(-0.232,0.248)	(-0.201,0.215)
1		-7.569	-8.844	6.885	0.013	0.005
	Clash	(-8.107,-7.031)	(-16.529,-1.159)	(0.234,13.536)	(-0.127,0.153)	(-0.307,0.317)
2		-4.931	-17.972	0.874	0.033	0.001
		(-6.427,-3.436)	(-33.231,-2.713)	(-0.336,2.084)	(-0.033,0.067)	(-0.266,0.268)
3		-2.261	-1.263	0.400	0.005	0.002
		(-3.772,-0.750,)	(-2.929,0.403)	(-0.389,1.189)	(0.005,0.006)	(-0.280,0.285)
1		-7.913	-21.665	4.757	0.033	0.001
	Illegal Economy	(-9.468,-6.358)	(-38.968,-4.362)	(0.379,9.136)	(,-0.032,0.098)	(-0.186,0.189)
2		-5.996	-3.641	0.891	0.010	0.006
		(-6.633,-5.359)	(-6.422,-0.860)	(0.149,1.633)	(-0.029,0.048)	(-0.338,0.350)
3		-3.693	-0.65	-1.036	0.093	0.002
		(-4.564,-2.822)	(-1.375,0.075)	(-2.074,0.102)	(0.084,0.103)	(-0.242,0.246)
1		-6.761	-3.933	0.123	-0.017	0.001
	Killings	(-8.739,-4.783)	(-8.512,0.646)	(-0.066,0.312)	(-0.172,0.138)	(-0.188,0.190)
2		-4.843	-4.132	0.995	0.023	0.004
		(-6.055,-3.631)	(-7.939,-0.324)	(0.037,1.953)	(-0.014,0.060)	(-0.302,0.310)
3		-1.634	0.700	0.296	0.008	0.001
		(-2.419,-0.854)	(-0.774,2.174)	(-0.326,0.917)	(-0.013,0.030)	(-0.229,231)
1		-7.483	-5.723	1.270	0.016	0.004
	Terrorist Attacks	(-9.329,-5.637)	(-10.905,-0.541)	(0.555,1.985)	(0.009,0.023)	(-0.350,0.358)
2		-5.335	-3.662	1.093	0.021	0.001
		(-6.829,-3.841)	(-4.011,0.349)	(0.127,2.059)	(0.005,0.037)	(-0.202,0.204)
3		-2.228	0.269	-0.006	-0.042	-0.002
		(-3.770,0.686)	(-0.204,0.742)	(-0.573,0.561)	(-0.057,0.141)	(-0.227,0.231)

Figure 3.5 displays the map for the predicted FARC presence in all municipalities between 2006 and 2009. As the transition from high FARC presence to lower FARC presence is lower, we see the map with darker blue compared to the previous HMM with no covariates. We again see the same regions where the main FARC blocks operated during this period. Comparing the predicted FARC presence with expert data, we see now 18% of municipalities classified with high FARC presence but with no presence according to expert data. Table 3.9 also shows that of those municipalities denoted with high FARC presence by the experts, almost all were classified with high FARC presence by the HMM.

Figure 3.5: Predicted FARC presence using covariates

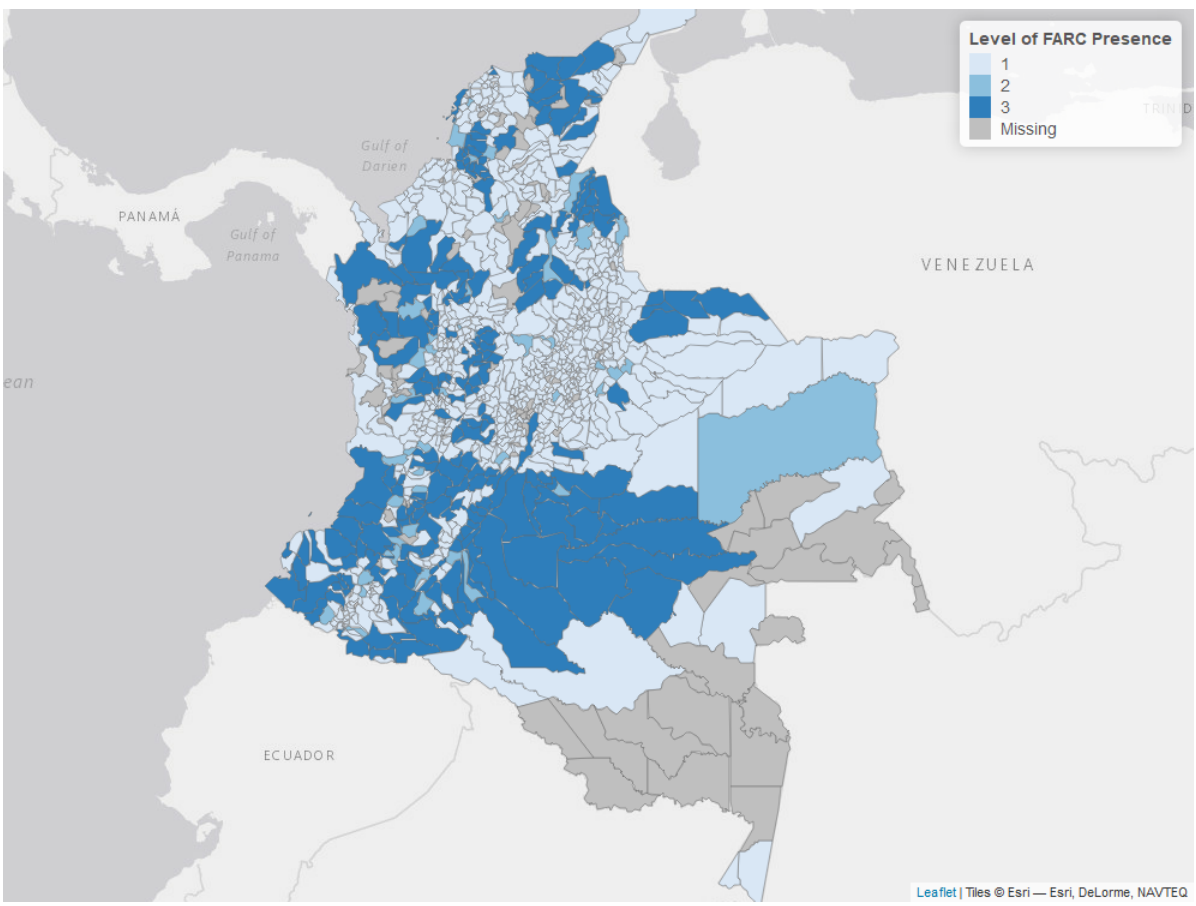


Table 3.9: Comparing FARC presence with expert data

	Expert Data	
HMM	None	Presence
Low	756 (74.0%)	2 (0.1%)
Middle	45 (4.4%)	8 (0.7%)
High	189 (18.5%)	22 (2.1%)

3.3 Using the predicted data

In this section, I replicated the results from Dube and Vargas (2013). As mentioned above, the idea is to identify whether coffee and oil shocks affect the level of violence. They use a dependent variable called attacks which includes firing upon another group, incursions into a village, killing civilians, bombing pipelines, bridges, and other infrastructure targets, destroying police stations or military bases, and ambushing military convoys. In other words, it is a variable mixing raid, terrorist attacks and killings as defined in this paper.

I included the variable of FARC or paramilitary presence into their linear regressions depending on the dependent variable used. These variables were converted to an indicator equal to 1 if FARC or paramilitary presence is medium or high and otherwise equal to 0. The idea is to check whether their findings are conditioned on the level of armed group presence.

As shown in table 3.10, in column 1 and 2, the dependent variable is guerrilla attacks. Their coefficients suggested that positive coffee shocks were associated with a decrease in guerrilla attacks. However, the second column indicates that this effect is very strong but occurs mainly in municipalities with low FARC presence. However, the effect is much lower in municipalities with medium or high FARC presence. The same happens on the effect on paramilitary attacks, as seen in column 4. The effect occurs mainly in municipalities with lower paramilitary presence. This result suggests that the opportunity-cost effect of reducing labour supplied to appropriation may take place primarily in municipalities with low illegal armed presence, where civilians may be less coerced by illegal armed groups.

In column one, the coefficient of oil shocks was not statistically different from zero, and the coefficient of column 3 suggested that oil shocks are positively associated with paramilitary attacks. Including the interaction with FARC presence and paramilitary presence, the coefficients indicate that the oil shocks exert different effects for FARC and paramilitary violence. Oil shocks seem to increase guerrilla attacks in municipalities with high FARC presence but not in municipalities with low FARC presence. As suggested by previous studies, oil shocks may increase irregular tactics such as attacks on oil pipelines which the FARC can easily commit in territories with higher armed group presence. In contrast, oil shocks seem to increase paramilitary attacks more in municipalities with low paramilitary presence. One potential explanation for these results is the irregular alliances between the private sector and paramilitary groups. Oil shocks lead to increased counterinsurgent capacity, and in some cases, through private contracts between oil-exporting firms and paramilitary units. Therefore, oil shocks may increase the paramilitary economic capacity and expand their influence through new territories with lower paramilitary presence.

Table 3.10: The effect of the coffee and oil shocks on violence

	(1)	(2)	(3)	(4)
Dependent variable	Guerrilla Attacks		Paramilitary Attacks	
Coffee int. x log coffee price	-0.611**	-0.924*	-0.160***	-0.545**
	(0.249)	(0.476)	(0.061)	(0.262)
Coffee int. x log coffee price x Armed group presence		0.730**		0.427*
		(0.355)		(0.227)
Oil production x log oil price	0.700	-3.734	0.726***	1.231**
	(1.356)	(2.369)	(0.156)	(0.610)
Oil production x log oil price x Armed group presence		4.543**		-0.513***
		(1.940)		(0.207)
Observations	17,604	17,604	17,604	17,604

Notes: Standard errors clustered at the department level are shown in parentheses.Variables not shown include municipality fixed effects, year fixed effects, log of population, armed group presence (FARC or paramilitary depending on the dependent variable) and linear trends by region and municipalities cultivating coca in 1994. The interaction of the internal coffee price with coffee intensity is instrumented by the interaction of the coffee export volume of Brazil, Vietnam, and Indonesia with rainfall, temperature, and the product of rainfall and temperature. *** is significant at the 1% level; ** is significant at the 5% level; * is significant at the 10% level are.

Chapter 4

Conclusions

The main idea behind this research is to provide a statistical tool to identify armed group presence. This information on armed group presence is generally not available and is very useful in conflict panel studies. However, some studies seem to ignore the level of armed group presence, which can significantly vary across the territory. Other studies instead use violent events as proxies of armed group presence. However, a better solution is to use that information to estimate armed group presence. This research proposes that a Hidden Markov Model can be used for this purpose by modelling the distribution of violent actions by armed groups depending on a current hidden state, in this case, the armed group presence, and assuming that the unobserved parameter process satisfies the Markov property. This model also allows for several extensions of the basic model, such as the inclusion of covariates. This model's estimation will enable having complete historical panel data of armed group presence within a country. An extension to this model could also estimate the presence of different illegal armed groups jointly. In some cases, an armed group's presence may be positively associated with the presence of another group.

Using the prediction of armed group presence, this study shows that significant results can be obtained when including this variable in different conflict studies. In particular, it shows that most effects of economic shocks occur in municipalities with a low presence of armed groups. Also, some economic shock effects differ if the FARC or the paramilitary groups commit the violence. A further study may analyze the differential tactics and goals of paramilitary and guerrilla groups and how these tactics differentially affect the level of violence and the type of violence used after economic shocks.

References

Albert, James H. and Siddhartha Chib (1993). “Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts.” In: *Journal of Business and Economic Statistics* 11.1, pp. 1–15.

Altman, Rachel Mackay (2007). “Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting.” In: *Journal of the American Statistical Association* 102.477, pp. 201–210.

Asal, Victor et al. (2016). “Political Exclusion, Oil, and Ethnic Armed Conflict.” In: *Journal of Conflict Resolution* 60.8, pp. 1343–1367.

Bhavnani, Ravi, Dan Miodownik, and Hyun Jin Choi (2011). “Violence and control in civil conflict: Israel, the West Bank, and Gaza.” In: *Comparative Politics* 44.1, pp. 61–80.

Bickel, David R. (2018). “Bayesian revision of a prior given prior-data conflict, expert opinion, or a similar insight: a large-deviation approach.” In: *Statistics* 52.3, pp. 552–570.

Calle, Luis de la and Ignacio Sánchez-Cuenca (2015). “How Armed Groups Fight: Territorial Control and Violent Tactics.” In: *Studies in Conflict & Terrorism* 38.10.

Carter, David B. (2015). “When terrorism is evidence of state success: Securing the state against territorial groups.” In: *Oxford Economic Papers* 67.1, pp. 116–132.

CNMH, Centro Nacional de Memoria Histórica (2018). *Paramilitarismo, Balance de la contribución del CNMH al esclarecimiento histórico*. Bogotá.

Douglass, RW (2012). “Hearts, minds, or bodies: the strategy of selective violence in civil war.”

Dube, O. and J. F. Vargas (2013). “Commodity Price Shocks and Civil Conflict: Evidence from Colombia.” In: *The Review of Economic Studies* 80.4, pp. 1384–1421.

Fergusson, Leopoldo, Juan F. Vargas, and Mauricio A. Vela (2013). “Sunlight Disinfects? Free Media in Weak Democracies.” In: *SSRN Electronic Journal*.

Ferrari, Diogo (2020). “Modeling Context-Dependent Latent Effect Heterogeneity.” In: *Political Analysis* 28.1, pp. 20–46.

García-Sánchez, Miguel (2010). “Violent contexts, electoral participation and vote choices in Colombia: A hierarchical approach.” In: *Annual Meeting of the Midwest Political Science Association, Chicago*.

Green, Peter J. and Sylvia Richardson (2002). “Hidden Markov models and disease mapping.” In: *Journal of the American Statistical Association* 97.460, pp. 1055–1070.

Gutiérrez-Sanín, Francisco and Elisabeth Jean Wood (2017). "What Should We Mean by "Pattern of Political Violence"? Repertoire, Targeting, Frequency, and Technique." In: *Perspectives on Politics* 15.01, pp. 20–41.

Huang, Ruihong and Christina Kennedy (2008). "Uncovering Hidden Spatial Patterns by Hidden Markov Model." In: *Geographic Information Science*. Springer Berlin Heidelberg, pp. 70–89.

Juang, B. H. and L. R. Rabiner (1991). "Hidden markov models for speech recognition." In: *Technometrics* 33.3, pp. 251–272.

Kalyvas, Stathis N. (1999). "Wanton and senseless?" In: *Rationality and Society* 11.3, pp. 243–285.

Kalyvas, Stathis N (2006). *The Logic of Violence in Civil War*. Cambridge, UK: Cambridge Univ. Press.

– (2012). "Terrorism and Political Violence Micro-Level Studies of Violence in Civil War: Refining and Extending the Control-Collaboration Model." In: *Terrorism and Political Violence* 24, pp. 658–669.

Kaplan, Oliver (2013). "Nudging armed groups: How civilians transmit norms of protection." In: *Stability* 2.3.

King, Gary and Langche Zeng (2001). "Improving Forecasts of State Failure." In: *World Politics* 53.4, pp. 623–658.

Langrock, Roland et al. (2012). "Flexible and practical modeling of animal telemetry data: Hidden Markov models and extensions." In: *Ecology* 93.11, pp. 2336–2342.

López Hernández, Claudia Nayibe and Ariel Fernando Ávila Martínez (2010). *Y refundaron la patria–: de cómo mafiosos y políticos reconfiguraron el estado colombiano*. Debate.

Mampilly, Zachariah Cherian (2011). *Rebel Rulers: Insurgent Governance and Civilian Life During War*. Cornell University Press.

Namen, Olga, Mounu Prem, and Juan F. Vargas (2019). "The Human Capital Peace Dividend." In: *SSRN Electronic Journal*.

Prem, Mounu, Santiago Saavedra, and Juan F. Vargas (2020). "End-of-conflict deforestation: Evidence from Colombia's peace agreement." In: *World Development* 129, p. 104852.

Raleigh, Clionadh et al. (2010). "Introducing ACLED: An Armed Conflict Location and Event Dataset." In: *Journal of Peace Research* 47.5, pp. 651–660.

Schrodt, Philip A (2000). "Early Warning of Conflict in Southern Lebanon using Hidden Markov Models." In: *American Political Science Association*.

Schrodt, Philip A. (2006). "Forecasting Conflict in the Balkans using Hidden Markov Models." In: *Programming for Peace*. Ed. by R. Trappl. Vol. 2. Springer. Chap. 8, pp. 161–184.

Shearer, Robert (2007). "Forecasting Israeli-Palestinian Conflict with Hidden Markov Models on JSTOR." In: *Military Operations Research* 12.4, pp. 5–15.

Sundberg, Ralph and Erik Melander (2013). *Introducing the UCDP Georeferenced Event Dataset*. Tech. rep. 4, pp. 523–532.

Tao, Ran et al. (2016). "A hybrid approach to modeling territorial control in violent armed conflicts." In: *Transactions in GIS* 20.3, pp. 413–425.

Thill, Jean-Claude F et al. (2017). *Territorial Control and Violence Against Civilians Boko Haram in Nigeria Exploratory Spatial Flow Data Analysis View project Territorial Control and Violence Against Civilians Boko Haram in Nigeria*. Tech. rep.

Tyner, Evan (2016). *Do Territorial Control and the Loss of Territory Determine the Use of Indiscriminate Violence by Incumbent Actors? An Examination of the Syrian Civil War in Aleppo over 45 weeks*. Tech. rep. 1.

Vargas, Gonzalo (2009). "Terrorism and Political Violence Urban Irregular Warfare and Violence Against Civilians: Evidence From a Colombian City Urban Irregular Warfare and Violence Against Civilians: Evidence From a Colombian City." In: *Terrorism and Political Violence* 21, pp. 110–132.

Visser, Ingmar, Maartje E. J. Raijmakers, and Peter C. M. Molenaar (2000). "Confidence intervals for hidden Markov model parameters." In: *British Journal of Mathematical and Statistical Psychology* 53.2, pp. 317–327.

Visser, Ingmar and Maarten Speekenbrink (2010). "depmixS4: An R Package for Hidden Markov Models." In: *Journal of Statistical Software* 36.7.

Ward, Michael D. and Kristian S. Gleditsh (2002). "Location, Location, Location: An MCMC Approach to Modeling the Spatial Context of War and Peace." In: *Political Analysis* 10.3, pp. 244–260.

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

Wood, Elisabeth Jean (2003). *Insurgent Collective Action and Civil War in El Salvador - Elisabeth Jean Wood, Wood, Professor Elisabeth Jean Wood, Wood Elisabeth Jean - Google Books*. Cambridge: Cambridge University Press.

Wright, Austin L (2015). "Economic Shocks and Rebel Tactics: Evidence from Colombia." In: *SSRN Electronic Journal*.

Zucchini, Walter, Iain L. Macdonald, and Roland Langrock (2017). *Hidden Markov models for time series: An introduction using R, second edition*. CRC Press, pp. 1–370.

Appendix A

Confident intervals

Confident intervals can be obtained using the finite-differences approximation, $f'(x) = \lim_{\delta x \to 0} \frac{f(x+\delta x)-f(x)}{|\delta x|}$, for the second partial derivative of the log-likelihood function, $L(O|\lambda)$ resulting in the following entries of the Hessian matrix: $\hat{h}_{ij} = \frac{\partial^2}{\partial \lambda_i \lambda_j} f(\lambda) = \frac{L(\lambda+\delta_i e_i+\delta\lambda_j e_j)-L(\lambda+\delta_i e_i)-L(\lambda+\delta_j e_j)+L(\lambda)}{|\delta\lambda_i \delta\lambda_j|}$, with λ_i and λ_j being the ith and jth parameters of the log-likelihood function, L the log-likelihood function and e_i the ith unit vector.

A.1 First Estimations-Paramilitary data

$$\delta = \begin{pmatrix} [0.599 - 0.658] & [0.276 - 0.329] & [0.006 - 0.136] \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} [0.826 - 0.848] & [0.153 - 0.174] & [0.000 - 0.000] \\ [0.142 - 0.169] & [0.604 - 0.632] & [0.216 - 0.239] \\ [0.000 - 0.000] & [0.289 - 0.3737] & [0.627 - 0.711] \end{pmatrix}$$

Table A.1: Estimated π's

	π_i						
State	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	No event
St1	[0.001-0.002]	[0.0008-0.001]	[0.0006-0.0007]	[0.002-0.004]	[0.003-0.004]	[0.000-0.000]	[0.991-0.992]
St2	[0.013-0.019]	[0.001-0.003]	[0.005-0.007]	[0.015-0.017]	[0.033-0.037]	[0.001-0.002]	[0.919-0.931]
St3	[0.081-0.111]	[0.003-0.005]	[0.007-0.011]	[0.002-0.004]	[0.009-0.019]	[0.002-0.004]	[0.857-0.907]

A.2 First estimations- FARC data

$$\delta = \begin{pmatrix} [0.584 - 0.608] & 0.315[0.291 - 0.339] & 0.089[0.081 - 0.097] \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} [0.840 - 0.878] & [0.135 - 0.146] & [0.000 - 0.000] \\ [0.366 - 0.388] & [0.426 - 0.452] & [0.174 - 0.193] \\ [0.000 - 0.000] & [0.286 - 0.310] & [0.687 - 0.717] \end{pmatrix}$$

Table A.2: Estimated π's

	π_i						
State	Authority Abuse	Raid	Clash	Illegal Economy	Killings	Terrorist Attacks	No event
St1	[0.000-0.000]	[0.0005-0.0005]	[0.0003-0.0003]	[0.0001-0.0002]	[0.0006-0.0007]	[0.0002-0.0003]	0.999[0.997-0.999]
St2	[0.0003-0.0004]	[0.001-0.0014]	[0.0008-0.0009]	[0.0024-0.003]	[0.0033-0.0039]	[0.0069-0.0077]	[0.980-0.988]
St3	[0.000-0.000]	[0.0416-0.0442]	[0.0014-0.0030]	[0.0008-0.0009]	[0.0079-0.089]	[0.0031-0.0039]	[0.930-0.954]

Appendix B

Longitudinal Structure and zero inflation

The third aspect to consider is the longitudinal nature of the data. Conflict is often analyzed with longitudinal data, and the idea behind using HMM is to identify armed group presence in each of the K territories in each time t. The past estimations did a complete pooling where it was assumed that the transition matrix parameters were equal for all subjects (municipalities in this case). For instance, $\gamma_{12}^1 = \gamma_{12}^2 \ldots \gamma_{12}^K = \gamma_{12}$ with K being the total number of municipalities, and the same for all the $q \times m$ parameters of the emission probability (π_{ij}).

One can do a partial pooling where the transition probability parameters are pooled but not those of the state-dependent distributions. If all the $q \times m \times K$ parameters of the state-dependent distribution were assumed to be fixed, the number of parameters in the state-dependent distributions will increase by K times. An alternative approach is to use random effects such that variability across territories is allowed. In this case, some parameters are random effects and are independently drawn from a distribution that is common to all component series (Altman, 2007). The inclusion of random effects can permit the estimation of population-level effects and greater flexibility in modelling the correlation structure. Therefore, we assume that the parameters of the HMM follow a given group-level distribution. The mean and the variance of the group-level distribution of a given parameter express the overall mean parameter value in a group of subjects and the parameter variability between the subjects in the group.

Therefore, the observations of each subject are distributed according to the same distribution $X_{kt} \sim Q(\Pi_k)$ but also the subject-specific parameters, Π_k, are realizations of a common group level distribution W with parameter set Θ: $\Pi_k \sim W(\Theta_k)$. So the subject level parameters are assumed to be random draws from a given distribution, in this case a a normal group level distribution will be used. In the case of the multinomial model, subject k's probabilities of observing categorical outcome j within state m, π_{kij} are modelled using m batches of $q-1$ random intercepts $\alpha_{ki} = (\alpha_{ki2}, \alpha_{ki3} \ldots \alpha_{kiq})$ so that $\pi_{kij} = \frac{\exp\{\alpha_{kij}\}}{1+\sum_{j=2}^{q}\exp\{\alpha_{kij}\}}$. Each batch of random intercept, α_{ki} comes from a state i specific population level multivariate normal distribution with mean $\bar{\alpha}_i$ of length $q-1$,[2] and covariance Φ_i denoting the covariance between the $q-1$ state specific intercepts over subjects.

Another important aspect to consider is the excess of zeros in the violent event data. In many quarters many territories have many days with zero activity from armed groups. Therefore, another specification to account for potential overdispersion problems or zero inflation could be using a zero-inflated model. A zero-inflated Poisson distribution will be used for the state one emission probability distribution using the total number of violent events in each municipality for accounting for this potential problem. For those municipalities predicted to be in state two, a multinomial-HMM is estimated as before to predict state 2 or 3. The zero-inflated Poisson distribution in the following way $P(X_{it} = x_{it} | P_{it} == 1) = \theta_{it} 1_{x_t=0} + \{1-\theta_{it}\}\frac{\lambda_{it}(m)^x \exp \lambda_{it}(m)}{y!}$

[2] Or $\bar{\alpha}_i + X_K^{T\prime}\beta_{ij}$ when including covariates.

B.1 Paramilitary violent events

Using the previously describe HMM model with the paramilitary data, the following are the estimated parameters for the zero-inflated Poisson model using two states to predict whether the municipality has a low paramilitary presence. The log-likelihood for this estimation was -2554, and the AIC was 5121.07. As with previous models, the initial probability is high at the first state, 91%. The estimated $\hat{\theta}$, and $\hat{\lambda_m}$ of the zero-inflated Poisson distribution are 0.905, and 0.69, respectively, and $\hat{\lambda}$ for state 2 is 4.66. The estimated transition probabilities show again that the transition probability of switching from state 1 to state 2 and from state 2 to state 1 is low, 10% and 16% respectively.

$$\delta = \begin{pmatrix} 0.914 & 0.086 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} 0.893 & 0.107 \\ 0.163 & 0.836 \end{pmatrix}$$

$$\lambda = \begin{pmatrix} 0.693 & 4.66 \end{pmatrix}$$

The next step is to estimate the random effect HMM with 2 states. The state-dependent probability distribution is considered to be a categorical distribution instead of a multinomial distribution to facilitate the estimation: $Pr(X_t = j|P_t = i) \sim Cat(\pi_i)$ where es $j = 1, 2, \ldots, q$ are the q possible categorical outcomes for each type of violent event (each type of violent event will have 3 possible categorical outcomes: low (L), medium (M) or high (H)) and where $\pi_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{iq})$ with $\sum_j = 1$. Therefore, we will have 36 parameters in the emission distributions (3 for each categorical outcomes, 2 for each of the two possible states and 6 for each type of violent event data).

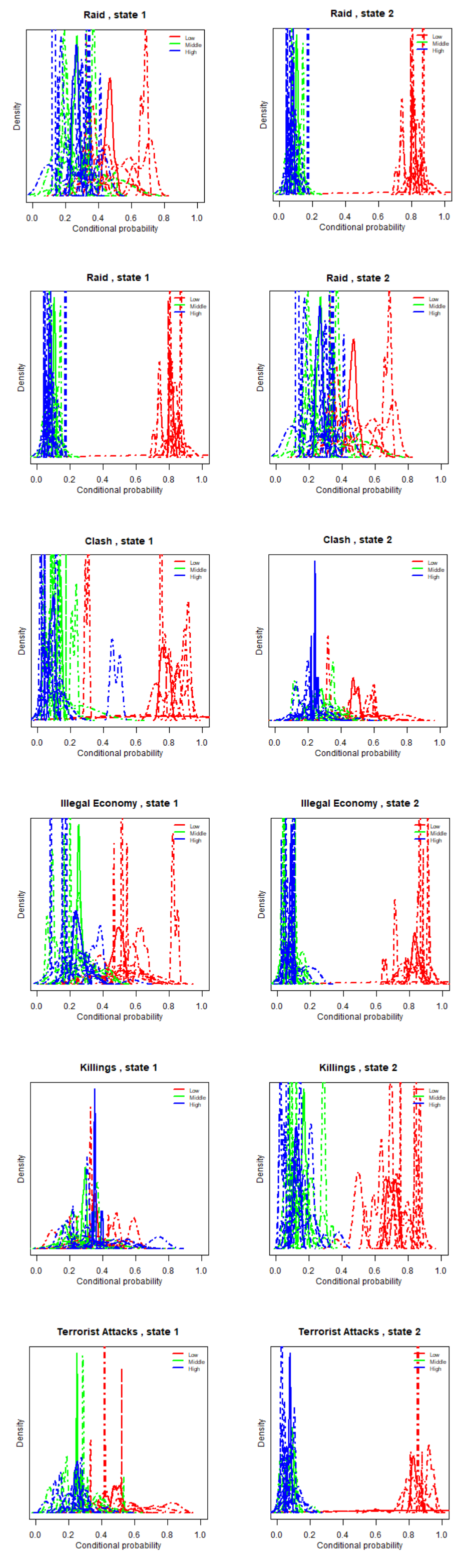
The resulting model indicates two well-separated states. The Γ transition matrix shows that the probability for switching from state 1 to state 2 is 12%, and the probability of switching from state 2 to state 1 is 37%. The estimated transition probabilities are The log-likelihood of is -80.95, and the AIC is equal to 213.9. The group-level emission probabilities show that the probability of having high authority abuse is higher conditionally on being in state 2. The probability of having high raids levels is only 9% when being in state 1 and 23% when being in state 2. A similar result is obtained with clashes, while for the illegal economy, killings, and terrorist attacks, the opposite occurs. The probability of having low crimes related to the illegal economy, killings or terrorist attacks is around 4% when being in state 1. Figure B.1 shows the probabilities for each type of event for all the municipalities (dashed line) and the group level (solid line).

$$\Gamma = \begin{pmatrix} 0.872 & 0.128 \\ 0.379 & 0.621 \end{pmatrix}$$

Table B.1: Group Level Emission probabilities

Type of event	State	L	M	H
	State 1	0.429	0.265	0.287
Authority Abuse	State 2	0.500	0.202	0.304
	State 1	0.777	0.132	0.092
Raid	State 2	0.477	0.280	0.243
	State 1	0.836	0.081	0.082
Clash	State 2	0.493	0.257	0.240
	State 1	0.465	0.272	0.265
Illegal Economy	State 2	0.814	0.107	0.078
	State 1	0.348	0.292	0.352
Killings	State 2	0.713	0.171	0.123
	State 1	0.490	0.251	0.254
Terrorist Attacks	State 2	0.834	0.087	0.078

Figure B.1: Conditional probabilities by type of event



Combining the results from the zero-inflated Poisson HMM and the HMM with random effects, I selected municipalities to be in three possible states. Figure B.2 shows the map according to these three possible classifications. Table B.2 compares the number of municipalities in each category with the expert classification. As seen, the classification is pretty good.

Figure B.2: Predicted paramilitary presence

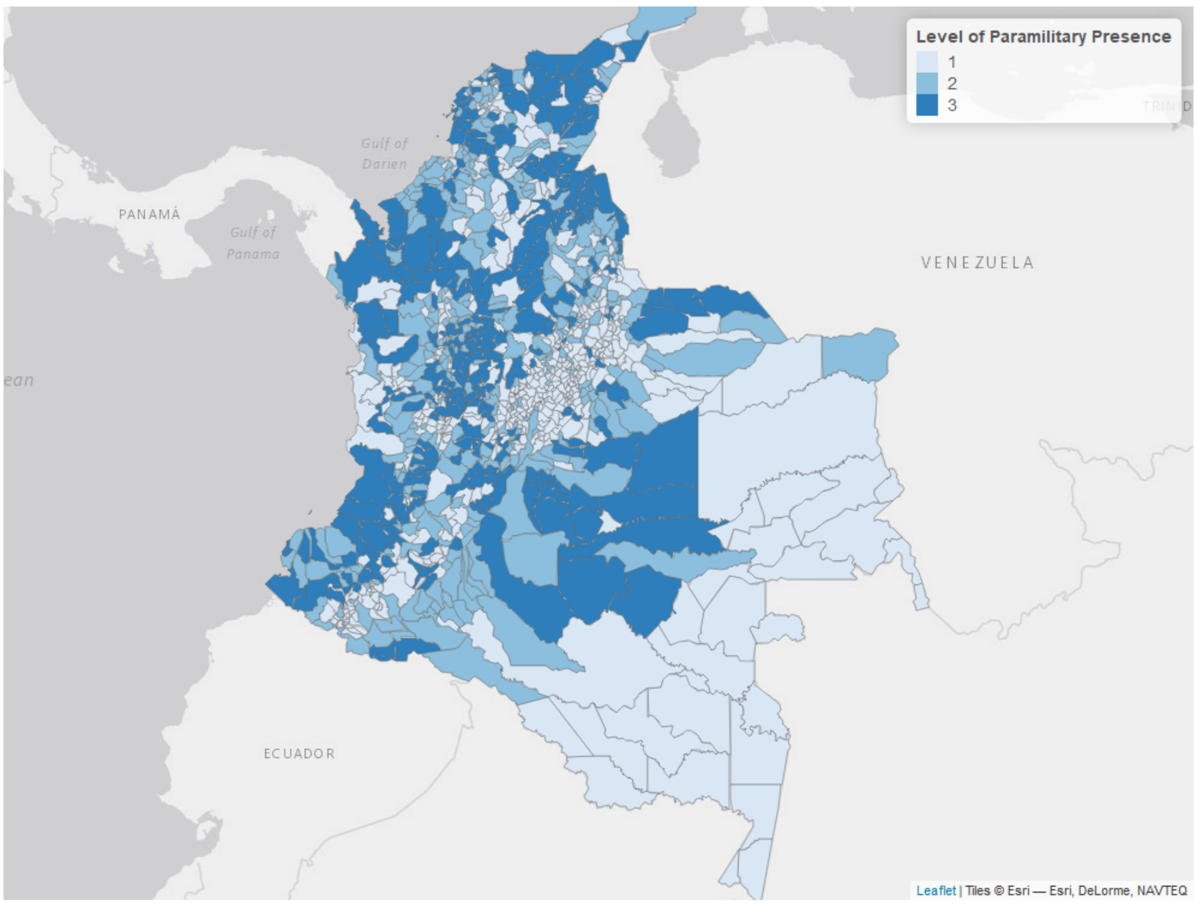


Table B.2: Comparing paramilitary presence with expert data

	Expert Data	
HMM	None	Presence
Low	627 (55.8%)	11 (0.9%)
Middle	257 (22.9%)	133 (11.8%)
High	8 (0.7%)	86 (7.6%)

Appendix C

Additional Figures and Tables

Figure C.1: Line Plot by type of violent event

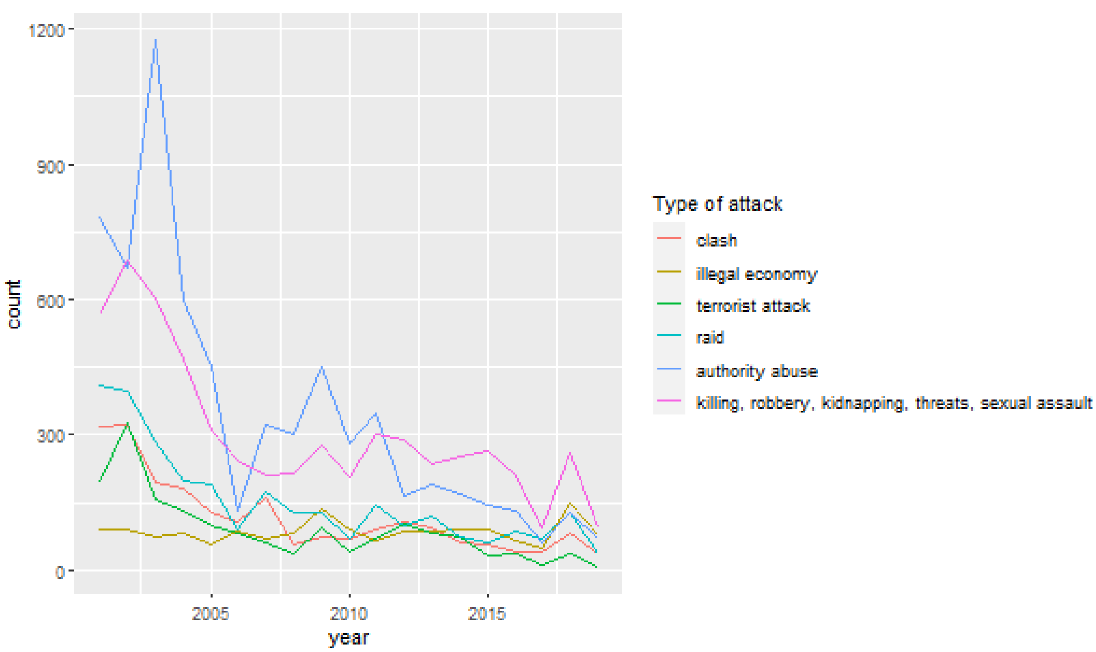


Figure C.2: Auto-correlation plot by type of event

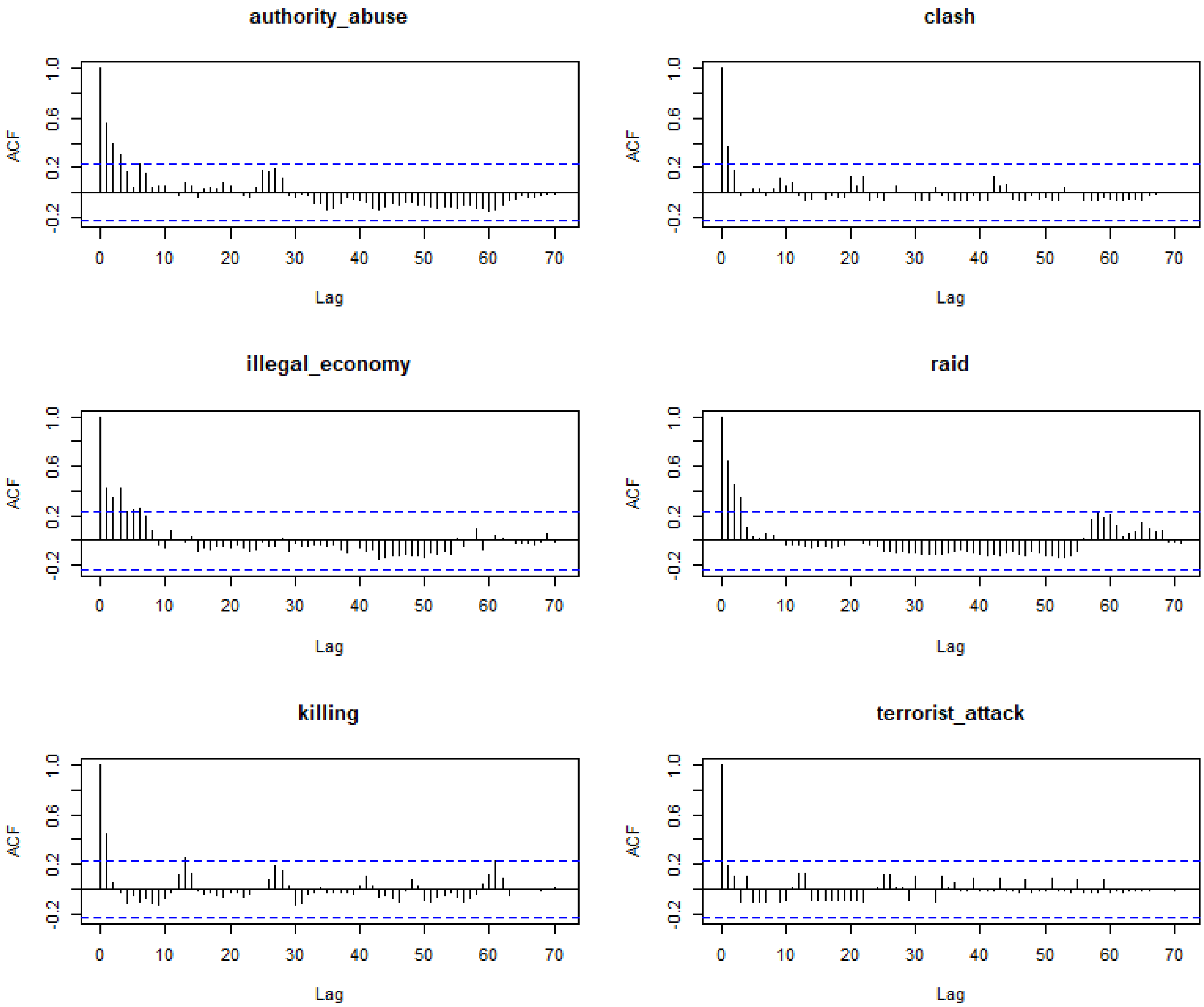


Table C.1: AIC and BIC for paramilitary violent events

N. States	-LogLIK	AIC	BIC
2	1465.98	2933.97	2945.93
3	1405.69	2814.38	2836.4
4	1488.06	2883.12	2915.79
5	1499.94	2895.88	2975.82

Table C.2: AIC and BIC for FARC violent events

N. States	-LogLIK	AIC	BIC
2	864.59	1731.19	1743.15
3	832.38	1670.78	1691.79
4	885.42	1773.74	1795.57
5	972.61	1880.41	1949.82