

Financing sources of rebel groups

Working paper

Álvaro Ámoros Rodríguez

28/6/2021

Introduction

Summary

For this project, I will use the publicly available Global Terrorism Database in order to understand the economic activities on which rebel organizations rely on in order to finance their fight against the state. Previous research regarding political violence has pointed out that the lack or not of territorial control is a key element in order to understand the different types of violence in which rebel organizations engage. As it happens with violence, I argue that territorial control is determinant when defining the different types of financing activities that rebel groups use. While rebel groups that are able to gain and maintain some sort of territorial control, as guerrillas for example, can rely on a broad spectrum of economic activities, like producing and smuggling illegal goods or taxing local populations, rebel organizations without territorial control, as terrorist groups, will be significantly restricted in the activities they can conduct. Hence, for groups without territorial control, activities that require a low level of sophistication, like robberies or kidnappings, will be more common. As in the GTD there is no information about the territorial control of organizations, I will use the dataset created by Ignacio Sanchez Cuenca, which includes information about territorial control of many of the organizations that are listed in the GTD. Regarding the methodological part, I will use text analysis in order to extract information about the economic nature of the incidents registered in the GTD. To test the relation between territorial control and economic activities I will use odds ratios, as well as a logistic regression analysis. Finally, as until now, the only method used to assess the territorial control of a rebel organization has been through qualitative analysis, I will apply a selection three model to predict if a organizations has territorial control or not, based on the GTD variables. This can be useful for further research, as coding information about territorial control annually through qualitative analysis is a long a tedious process, and many of the organizations of the GTD are missing in Ignacio Snachez Cuenca database

Theory

Luis de la Calle and Ignacio Sánchez Cuenca (2011) have pointed out that the type of violent actions rebel groups engage on are strongly determined by their ability to establish territorial control inside the borders of the state they are fighting against. The lack or not of territorial control would be ultimately a result of the asymmetry in forces between rebels and state. In underdeveloped and weak states, rebel contestants will be able to assert territorial control more easily and form guerrillas, on the contrary, in more developed and strong states rebels will need to operate as clandestine organizations, hence forming terrorist groups. According to the authors, the main features which distinguish terrorist organizations from other type of rebel groups, is precisely the lack of territorial control, which signals an extreme asymmetry in forces between the state and the rebels. Terrorist groups will respond to this asymmetry relying on very specific forms of violence, which generally do not lead to direct confrontation with the state, such as the use of IEDs, targeted assassination or kidnappings.

But the actions of rebel groups are not restricted to violent attacks, actions targeted on obtaining resources are also part of their normal activities. As it happens with violent actions, I claim that the type of economic actions in which rebel groups engage will be determined by their relative position of power in front of the state. Hence, rebels with territorial control, as guerrillas, will be able to engage in more sophisticated activities that can be easily develop in their own territories, as, for example, producing or smuggling drugs. This type of activities, besides probably been more profitable, also entitle a lower level or risk when compared, for example, to bank robberies. Freeman (2011) and Horgan & Taylor (1999) have showed that bank robberies are among the most common sources of financing for clandestine organizations. This is so because they require a very low level of internal organizational and sophistication, and can be carried out in a small window of time, without necessary engagements with state forces, hence becoming a predilect method for week and undercover organizations.

Other type of economic activities commonly associated to rebel groups are extortion and kidnapping. Both require a higher level of sophistication when compared to robberies. Extortion is the simplest of both, but it still requires some method to contact and communicate with the victim, as well as some safe way to recollect the payments. Kidnappings are even more complex, they require precise information about the victim, as well as a safe place to keep it, and a channel of communication with whomever is supposed to pay the ransom.

The main hypothesis of this paper is that there is a correlation between the degree of power asymmetry between the state and rebel groups, and the sophistication of the economic activities the later carry out. Robberies will be the most common type of economic activity for terrorist groups and will be scarcely used by groups with territorial control due to its high risk. Extortion will be the second most common activity for terrorist groups as it requires a more sophisticated structure. Finally, Kidnappings with economic objectives will be among the least used activities for terrorist groups, as they require a save place to keep the victim during negotiations, as well as safe channels of communication.

Data

The Global Terrorism Dataset registers over 180.000 incidents related to the actions of rebel groups all over the world from 1970 to 1997. The dataset codes as terrorist organizations a broad set of non-state violent political organizations. To distinguish terrorist organizations from other type of rebel groups, I will use the dataset provided by Ignacio Sánchez Cuenca, which includes information about the territorial control of over 300 groups. I will merge the two datasets and keep only those observations for which there is information about their territorial control, as well as a summary description of the incident (to be able to conduct text analysis). As almost none incident happened before 1997 includes summary information, I will drop the observations prior to that year as well. The result is a dataset with 41.988 incidents committed by 193 different organizations, of which 63 are clandestine ones and 130 have some sort of territorial control inside the borders of the state they are fighting against. Of the overall number of incidents, 37.848 have been carried out by organizations with territorial control, while 3.999 have been conducted by organizations without. All the registered incidents happened between 1997 and 2017.

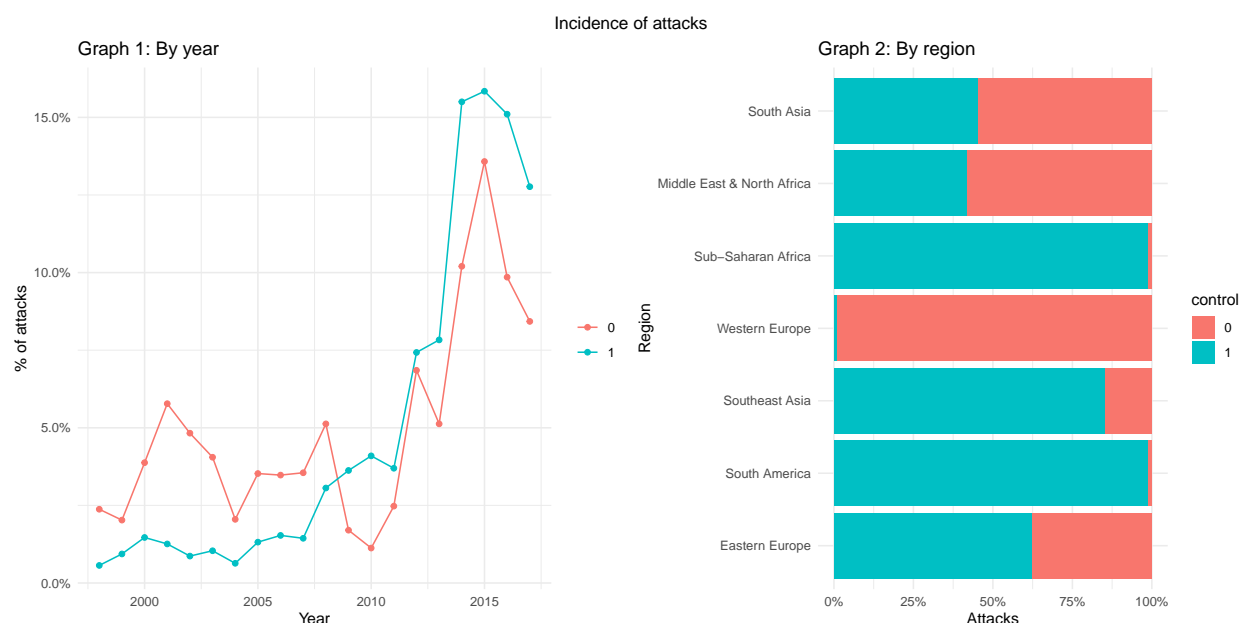
Analysis

Satistical analysis

Graph 1 shows the distribution of incidents over time, for both groups with territorial control and without. From 1997 to 2003 the number of attacks committed by rebel groups with territorial control is quite stable. After 2003 there is a constant increase in those types of attacks, probably boosted by the US invasion of Iraq. After the Arab spring (2010-2012) there is an extreme increase in attacks of terrorist groups with territorial control, that peaks in 2015, coinciding with the most violent period of the Syrian Civil War, and starts dropping after 2015 and the defeat of ISIS. Attacks by terrorist groups without territorial control show

a similar trend. There is a first peak of attacks in 2002 after the US invasion of Afghanistan, followed by a decline that stopped in 2003 with the invasion of Iraq. The late 2000s seemed to be the period with less terrorist attacks, but a violent increase in attacks happens again after the Arab spring, peaking in 2015 and then starting to decrease after the fall of ISIS.

Graph 2 shows the proportion of attacks committed by rebel groups with and without territorial control by region (regions with less than 100 incidents registered were dropped). In the regions of South America, Southeast Asia, and Sub-Saharan Africa, most of the attacks are committed by rebels with some kind of territorial control. The high proportion of attacks committed by organizations with territorial control in those regions, is partly due to the weakness of the states in the region. South Asia, Eastern Europe, and the Middle East and Northern Africa suffer of attacks from both types of organizations in similar proportions. Finally, in western Europe, there is almost exclusively cases of attacks from terrorist groups without territorial control.



Text analysis

To analyse the text contained in the summaries describing the incidents, I will generate a matrix with all the words contained in the *summary* variable. I will drop all *stop words*, all numbers and dates, as well as all geographical locations or references. The result is a dataset with over 1 million words as rows, and the 146 columns of the GTD. To be able to compare the different incidences of words in summaries describing incidents of both groups with territorial control and without, I will calculate the *odds ratio* of each word with relation to the variable of territorial control.

Tables 1 and 2 show the words with the highest and the lowest odds ratio respectively for terrorist groups without any kind of territorial control. To avoid certain words that do not add any valuable insight, I have added two restrictions to the tables. First, to avoid the appearance of certain words with very low frequency but very high odds ratios, I will require all words shown in tables 1 and 2 to appear at least thousand times in the overall Matrix. Second, to avoid the words with the lowest odds ratio being names of failed states or regions (as Afghanistan, Somalia or Mogadishu), in which any rebel group will almost automatically be able to hold some kind of territorial control, or the names of the groups operating in those states (as ISIS or the Taliban), I will require each word to appear at least 100 times in each type of rebel group (with territorial control and without territorial control).

Table 1 shows the five words with the highest odds ratios for terrorist groups without territorial control. The word with the highest odds ratio is *bomb*, which usually appears in summaries describing incidents with explosions provoked by IEDs. This supports previous research that identifies bombings as the preferred type of violent action of terrorist groups without territorial control. The second word with the highest odds ratio is *stated*, most of the summaries containing this word, describe or include some kind of reference of statements given by the rebel group after the registered incident. This points out the importance that publicizing their actions has for terrorist groups. As many of their actions do not pursue any direct military goals, but rather search to generate some kind of psychological impact in the public or in certain state institutions, it is fundamental for terrorist organizations to publicize their attacks. The third word, *damaged*, appears mainly in summaries related to bombing in which some property has been damage. The fourth word, with and odds ratio of 1.9 is *army*, and is included mainly in summaries that describe incidents in which a national army was targeted. These last results are rather strange, and suggest that terrorist groups without territorial control engage more frequently in actions against state armies that their rebel counterparts with territorial control. The last word in the list is *carried*, which usually appears in summaries were a statement about the nature of the attack is quoted.

Table 2 shows the 5 words with the lowest odds ratios, hence negatively correlated with the summaries describing incidents committed by terrorist groups without territorial control. The word with the lowest odds ratio is *attacked*, which is usually included in summaries which describe the assault of specific locations or facilities. This supports the theoretical expectations as those types of actions should be more common among rebel groups with territorial control as guerrillas. The next three words are *attributed*, *sources* and *unknown*, all three are usually included in summaries that describe incidents for which non group has claimed its responsibility. This supports the idea, present in Table 1, that groups with territorial control do not have any specific interest in publicizing their actions, as their attacks pursue specific military objectives. The last word *village* points out to the rural nature of rebel groups with territorial control.

This first text analysis helps to build a general picture of the different types of actions in which groups with and without territorial control engage. The findings point out that the most significant difference in actions committed by both types of rebel groups lies on the type of violence (bombs vs attacks), while the second most significant difference points out to the objectives of those attacks. Groups with territorial control as guerrillas, aim their actions towards specific military objective, hence they are not especially interested in any form of publicity and do not generate specific statements claiming their attacks. On the other hand, terrorist groups, as they lack the capacity to carryout operations with military objectives, can rely only in the psychological impact of their attacks, hence they are committed with publicizing them and reclaiming their authorship.

Table 1: Highest ratio. Groups without territorial control

Word	N. Control	N. No Control	Odds Ratio
bomb	582	1608	3.4
stated	215	931	2.2
damage	275	1277	2.0
out	354	1662	2.0
carried	286	1406	1.9

Table 2: Lowest ratio. Groups without territorial control

Word	N. Control	N. No Control	Odds Ratio
attacked	224	5976	0.35
attributed	652	15402	0.40
number	127	2977	0.40
sources	682	15714	0.41

Word	N. Control	N. No Control	Odds Ratio
unknown	282	6461	0.41

Financing

Besides strictly violent actions, rebel groups engage in specific activities with the aim of obtaining economic resources. The lack of territorial control will be a key factor in determining the kind of actions this groups can undertake. Due to their lack of territorial control, and clandestine nature, terrorist groups will rely on the least sophisticated activities to obtain resources. Of those, the simplest one is robbery, as it does not require any complex internal structure or any specific information. The next activity in the continuum of sophistication would be extortion, which requires a system to contact the victims, communicate with them, and a safe method to recollect the payments. The third most sophisticated activity would be kidnapping with economic aim, as it requires specific information about the victim, a method of communication with whomever is going to pay the ransom, a safe place in which to retain the victim etc...

Of those activities, the only one registered in the GTD is Kidnapping. But looking closer to the summaries of incidents coded as kidnapping, one can see that not all kidnappings have economic aims. In some cases, the victim is immediately executed without asking for any kind of ransom, in others, the victim or victims are used to bargain for military or political demands. For the incidents of robbery and extortion, there is no specific variable.

To capture these different types of incidents, I will combine text analysis with other variables already contained in the GTD. For the least sophisticated of all economic activities, robberies, I will rely exclusively on text analysis. I will code as robbery all incidents in which their summaries include one of the subsequent words: “*bank*”, “*robbery*”, “*theft*”, “*break-in*”, “*burglary*”, “*heist*”, “*looting*” and “*larceny*”. Regarding extortion, the word is used in many summaries to describe incidents of plane extortion and to describe actions of retaliation against civilians that refused to collaborate with the extortionists, but it is used as well in situations of kidnappings in which *extortion money* is demanded. To distinguish regular extortion from kidnaps, I will code as extortion only those incidents in which the words “*extortion*”, “*shakedown*”, “*payment*” and “*money*” appear, but that are neither coded as kidnappings or as robberies. Finally, to capture kidnappings with economic aims, I will code as kidnappings all the incidents that are defined as kidnappings in the variable *attacktype1*, and in which some of the words “*extortion*”, “*shakedown*”, “*payment*” and “*money*” appear.

Table 3 includes the frequencies and odds ratios of all the words previously mentioned for incidents conducted by terrorist groups without territorial control. The results show that the word *bank* has an extremely high odds ratio of 19.33, indicating that bank robberies are an economic activity deeply linked to the lack of territorial control and to the lack of sophistication of armed groups. The rest of the words have odds ratios below 1, indicating that they are more common in summaries of incidents conducted by groups with territorial control, although some of those words have significantly low overall frequencies in the word matrix. After the word *bank* the next three words with the highest odds ratios refer exclusively to economic related incidents and they are *money*, *payment*, and *extortion*. Finally, the last four words, with the lowest odds ratios, are related to kidnapping.

Table 3: Economic Incidents

Word	Territorial control	No territorial control	Odds ratio.
bank	272	133	19.33
payment	3	38	0.86
money	9	118	0.76
extortion	7	126	0.56
kidnap	2	42	0.56
abducted	125	2313	0.51

Word	Territorial control	No territorial control	Odds ratio.
kidnapping	42	964	0.42
abduction	24	638	0.36

Regression analysis

Table 4 shows the regression outputs for the variable's robbery, extortion and kidnapping- (with economic aims), coded as explained in the previous section. The dependent variable is territorial control, coded as 1 if the group does have control and as 0 other ways. In Model 1 a *graded logistic regression* without any controls is applied. Robbery is significant with a 99% confidence and has a negative effect on territorial control, this is driven almost exclusively by the incidents involving bank robberies, that, as showed in table 3, are very distinctive of terrorist organizations. Kidnapping with economic aims is also significant at a 99% level of confidence but has a positive coefficient on territorial control. In model 2 controls for both year and country effects are added, robbery keeps its negative and significant coefficient, while kidnapping also maintains its negative coefficient, but drops its significance to 95%.

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: mié, jul 28, 2021 - 19:57:04

Table 4:		
	<i>Dependent variable:</i>	
	control	
	(1)	(2)
robbery	−1.100*** (0.160)	−1.100*** (0.280)
extortion	0.320 (0.260)	0.360 (0.330)
kidnap	0.700*** (0.240)	0.720** (0.280)
Constant	2.200*** (0.017)	5.400*** (0.510)
Observations	41,847	41,847
Log Likelihood	−13,167.000	−4,685.000
Akaike Inf. Crit.	26,343.000	9,598.000
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Predictions

The analysis conducted until now has been limited by the availability of data regarding the territorial control of groups. Although our statistical results point out that territorial control is a key variable when understanding the behavior of rebel groups, the only data base available with such information is the one created by Ignacio Sanchez Cuenca through qualitative methodologies. Actually, many of the over 180.000 incidents registered in the GTD could not be used due to the lack of information on territorial control. To try to address this lack of information, I will use a set of machine learning techniques learned during the

HarvardX Data Science Certification, in order to predict if different incidents were committed by groups with territorial control or without.

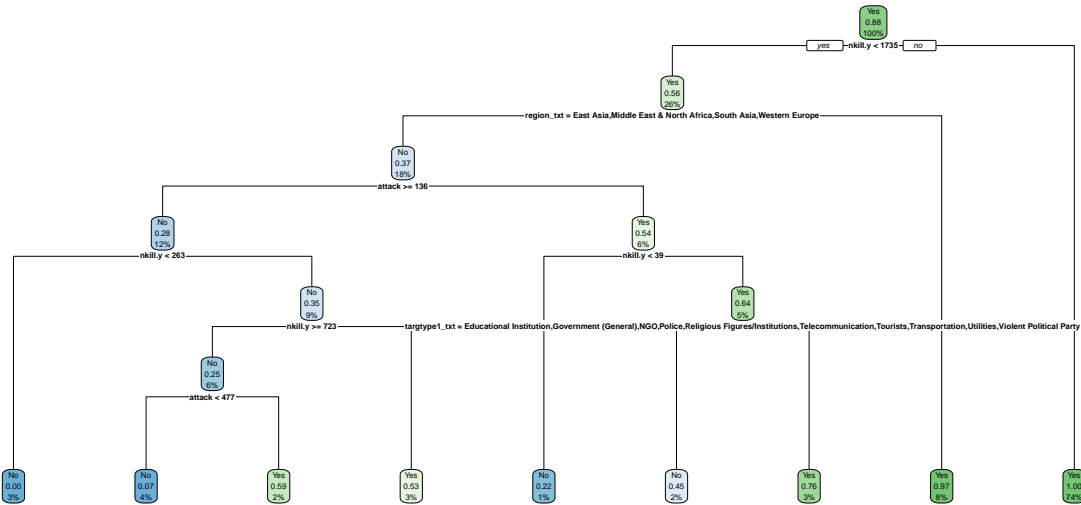
Decision tree

Due to the amount of variables in the data set (148 ones the variables produced through text analysis have been added to the original Global Terrorism Database), the prediction models available are constrained by the *curse of dimensionality*. To fit this high dimensionality, and taking into account that the outcome variable is categorical, I will use a decision tree model to predict territorial control. Although this model has some shortcomings, as a tendency towards overfitting, it has the advantage of being highly interpretable when compared to other machine learning algorithms.

To prepare the GTD data for training, first I will drop many of the columns in the data set that do not add any valuable information. Many of the columns are almost exclusively filled by NAs, and others like even ID or group name do not add any valuable information. Once the irrelevant columns are dropped, there are 36 variables left in the GTD. All cases with missing values in some of those 36 variables will be also deleted. I will continue by dividing the data in train and test sets, with the former containing a random selection of 80 % of the cases, and the former 20 %. A seed will be also set to 1 in order to obtain the same results when repeating the exercise. The results is a train set with 25.231 observations, and a test set with 6.309.

As mentioned before, decision trees have a tendency to overfit models, to avoid that I will add a set of restrictions. First, in order for a split in a node to be attempted I will add a restriction of at least 1000 observations in the node, and this split should result in at least one third of those observations ending in each node ($1000 / 3$). Finally I will add a complexity parameter restriction of 0.01, meaning that each split should increase the precision in at least 0.01 percent. With this restrictions we avoid having an over sized tree, which in an extreme case could have one end node for any observation, this would be an extreme case of overfitting and our model would be useless to predict outcomes in new data sets.

The decision tree in the graph shows how the algorithm is splitting the data. There are 9 end leaves which classify the 25.231 cases of the train set. The variables used to classify the incidents are the overall number of killings attributed to the group involved (nkill.y), the overall number of attacks attributed to the group (attack), the region in which the incident happened, and the target type of the attack. The accuracy of the model is of 0.951. Although this may seem a relatively high precision, if we look closer at the confusion matrix we can see that it is strongly unbalanced, it has a sensitivity of 0.983 and a specificity of only 0.706. This means that the algorithm predicts 98% correctly when they do have territorial control, but only 70% of the cases that do not have territorial control are predicted correctly. As there are significantly more cases with territorial control than without the overall precision of the algorithm is maintained high

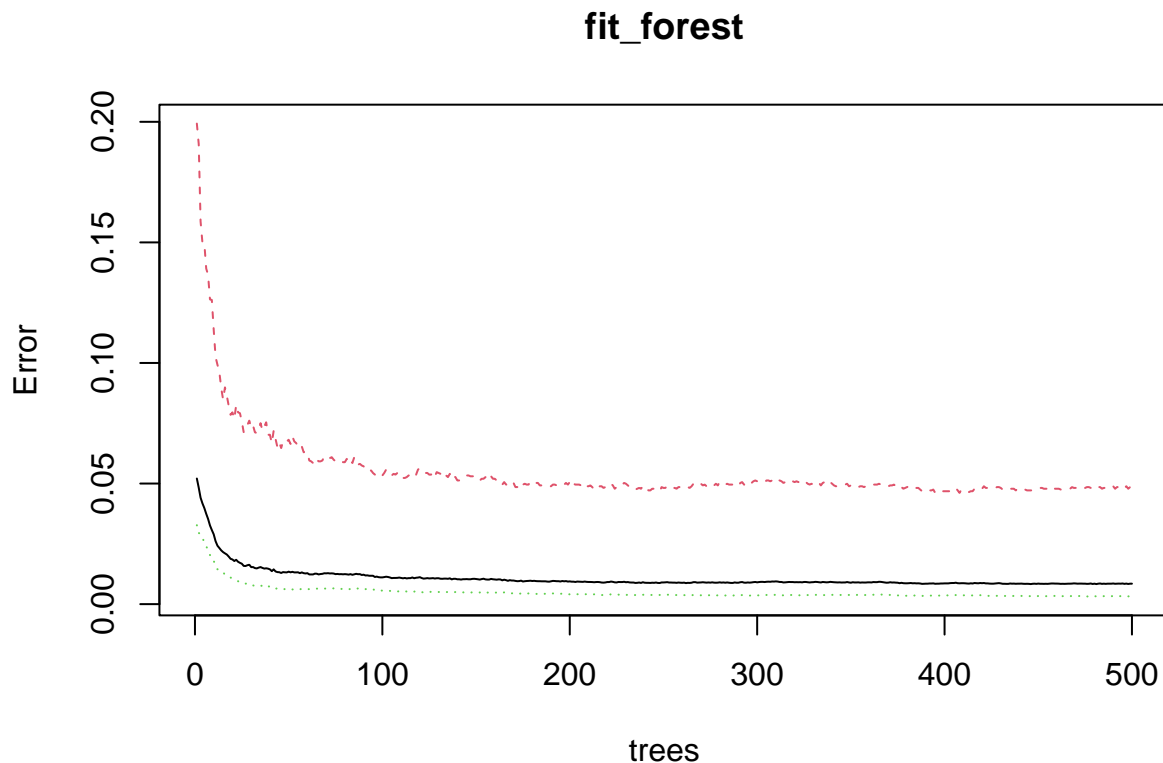


```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   No  Yes
##           No  513  95
##           Yes  214 5487
##
##           Accuracy : 0.951
##           95% CI : (0.945, 0.956)
##           No Information Rate : 0.885
##           P-Value [Acc > NIR] : < 2e-16
##
##           Kappa : 0.741
##
##           McNemar's Test P-Value : 1.91e-11
##
##           Sensitivity : 0.983
##           Specificity : 0.706
##           Pos Pred Value : 0.962
##           Neg Pred Value : 0.844
##           Prevalence : 0.885
##           Detection Rate : 0.870
##           Detection Prevalence : 0.904
##           Balanced Accuracy : 0.844
##
##           'Positive' Class : Yes
```


##

Random forest

Random forest algorithms are commonly used to improve the results of random tree predictions. The idea is straightforward, instead of relying on only one tree, random forest algorithms calculate a set of different trees. The algorithm ensures that those trees are different to each other through a bootstrap process, in which different random samples of the data are used for each tree. After a set of different trees has been generated, the results are simply averaged. The randomForest package takes care of all this and automatizes the process. As each tree in the forest is already different, overfitting should not be an issue here, hence no restrictions are needed. The fit-forest graph shows how the error in the predictions decreases as the tree number increases until reaching around 30 trees, where it stabilizes. The confusion matrix yields a better result with a 0.975 accuracy, with sensitivity reaching 0.99 and specificity increasing to 0.856. Although these are significantly better balanced results the specificity is still low compared to the sensitivity.



```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   No  Yes
##           No  622  53
##           Yes  105 5529
##
##           Accuracy : 0.975
##           95% CI : (0.971, 0.979)
##           No Information Rate : 0.885
```

```

##      P-Value [Acc > NIR] : < 2e-16
##
##              Kappa : 0.873
##
## Mcnemar's Test P-Value : 4.96e-05
##
##      Sensitivity : 0.991
##      Specificity : 0.856
##      Pos Pred Value : 0.981
##      Neg Pred Value : 0.921
##      Prevalence : 0.885
##      Detection Rate : 0.876
##      Detection Prevalence : 0.893
##      Balanced Accuracy : 0.923
##
##      'Positive' Class : Yes
##

```

Conclusions

Regarding the statistical analysis

The main substantive objective of this project, to understand if robberies committed by rebel organizations are linked to their lack of territorial control, has been accomplished. Incidents involving banks registered in the GTD are mostly committed by terrorist groups without territorial control, as the odds ratio of 19.33 showed in the text analysis. The logistic regressions support these results, as committing robberies decrease the chances of a group having territorial control with a 99% of confidence. These results keep their significance once country and year fixed effects have been introduced. Kidnaping with economic aims, coded partly using text analysis, shows a positive effect on territorial control. Although this is contradictory with some of the previous research, all the odds ratios for words related to kidnapping have coefficients below 1, indicating that they are more prevalent among summaries describing actions of groups with territorial control. Furthermore, kidnapping with economic aims yields a positive result in the logistic regression analysis, and maintains its significance at 95% confidence even after controlling for country and fixed effects. Regarding extortion, the odds ratios of all words linked to extortion have been greater than the ones of kidnapping, but all of them are still below one. In the regression analysis there was no statistically significant effect found. This suggests that as an activity which requires an intermediate level of sophistication, and is in between kidnappings and robbery, plain extortion is used by both groups with territorial control and without. Finally, the analysis of the words with the highest and the lowest odds ratios, suggest that a key element distinguishing actions of rebel groups with territorial control and terrorist groups, is that the latter have a strong interest in publicizing their activities, while this is rather irrelevant for the former.

Regarding the prediction models

Although the classification tree model yielded a significantly low specificity, it is still a useful model due to its high interpretability. Being able to plot the tree and see how the algorithm classifies data is useful to understand how this method could be applied on other data sets. The classification tree model also provides relevant information on key characteristics that distinguish rebel groups with territorial control and from those without. The model also shows that it is able to classify groups with territorial control and without with a fairly restricted set of variables, which are available in almost any data set of political violence. The problem of lack of specificity is partially solved in the random forest model, as it increases significantly the balance of the predictions and reached an overall precision of 0.97. These results show the potential that classification models have in order to predict the territorial control of armed groups, the extension of this

methodology to tho the whole of the GTD, as well as to other political violence data sets, could be a useful technique in order to overcome the necessity of qualitative analysis when classifying rebel groups.

Bibliography

Introduction to Data Science Data Analysis and Prediction Algorithms with R Rafael A. Irizarry 2021-07-03

Luis de la Calle & Ignacio Sánchez Cuenca. A THEORY OF TERRORISM. April 2021. Carlos III-Juan March Institute of Social Sciences