

Mafia activity index - some notes

Dec 11, 2018

Background and origin of this work

On May 13, 2018, in Follonica, a small city by the sea in Tuscany, the manager of one of the most well known hotels in the area was killed by a member of Naples's mafia: the Camorra. This murder was totally unrelated to criminal activities but - indirectly - provided the ultimate evidence that organized crime reached also Tuscany. The awe by people in Follonica and local politicians demonstrated at what extent civil society and institutions are unprepared to deal with organized crime. For sure, some 'activity' was spotted but it was left confined in a sort of 'twilight zone' accessible only to law enforcement agencies and prosecutors. Is there a way to explore this twilight zone? I argue that there is and that is data driven.

The proceeding

I proceeded using either corpus analysis techniques and a logit. The dependent variable of the logit is the outcome of the corpus analysis while the independent variables are data from the Interior Ministry regarding the so-called 'spy felonies'. 'Spy felonies' are all those crimes non directly related to mafia, but which could serve as an indicator of mafia activity. The list of them will be provided below.

The dependent variable

I collected the reports to the Parliament by the Dia, the body which coordinates mafia investigations across the Italian law enforcement agencies. I counted the municipalities nominated in those documents, then I aggregated those municipalities by provinces.

Given the different length of those reports, in order to make my data comparable, I divided the aggregate sum (by provinces) of utterances of municipalities by 10 pages. Afterwards, I added a column to my data-set. Whenever the number of quotations per 10 pages was larger than 0, I wrote 1 in my column.

The independent variable

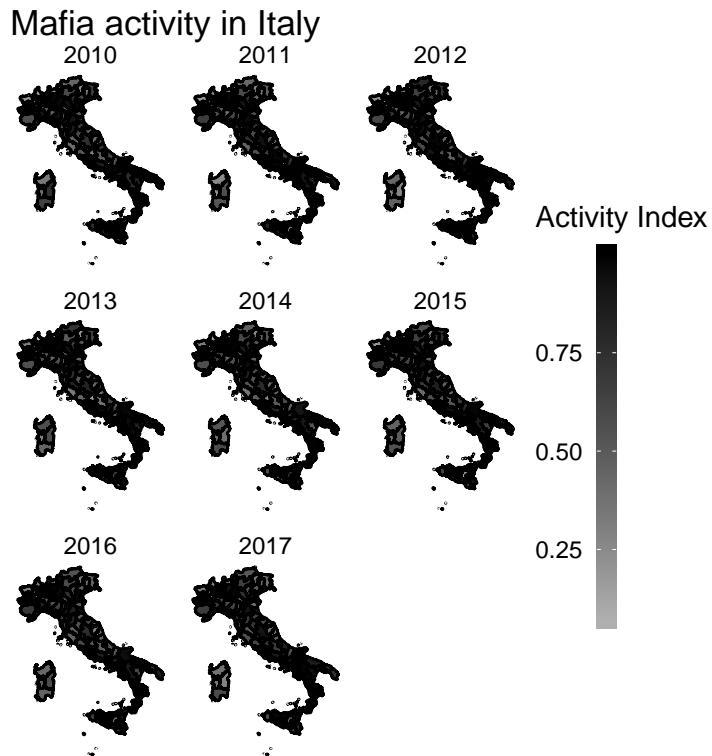
The independent variables are:

- ARSON Arson
- DAMARS Damages after an arson
- DRUG Drug Crimes
- EXTORT Extortion
- MAFIAHOM Mafia Homicide
- MAFIASS Mafia Association
- MENACE Menaces
- MONEYLAU Money laundering
- PROSTI Prostitution
- SMUGGL Smuggling

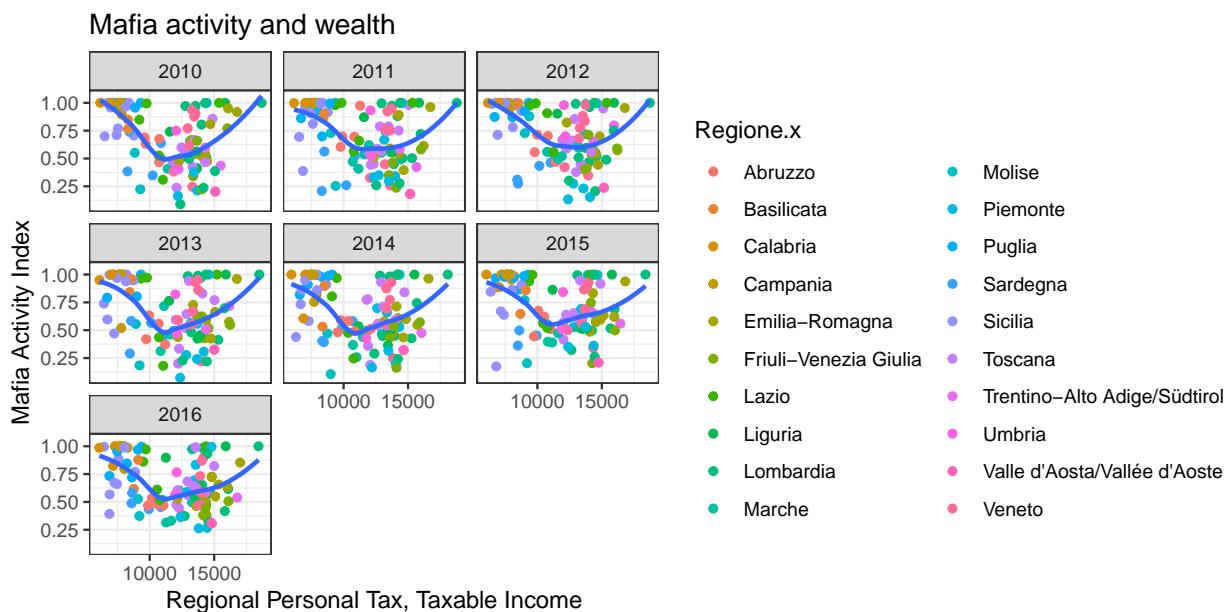
I used either the absolute numbers and the number of crimes per 100 thousands inhabitants ('p100k_' prefix in the tables)

The regression:

I regressed using a logit in order to measure the probability that a province was quoted in one of the Dia reports. This probability is what I defined as the 'Mafia Activity Index'. Here is the map of this index:



The maps show that Mafia is a national issue in Italy. Moreover, it shows that it is widespread in particularly affluent provinces, in the main cities (Rome, Milan, Turin) and, more importantly on the borders with France (Turin and Imperia) and Swiss (Como and Lecco). Moreover I discovered a weak but meaningful ‘U’ relation between the presence of mafia and income. In other words, Mafia appears to be more active in the richest and poorest areas of the country with some exceptions. This infex is being validated using news reports and data from NGOs, the government and so on.



On the y axis we have the mafia activity index, on the x axis we have the mean taxable income used to collect the regional personal tax. What we see is that mafia works as way to move wealth from the richest

parts of Italy (on the right end of the chart) towards the poorest (n the left end). WHat I am doing now is to understand if there is a relation of any kind between the terrotorial economic specialization (such as industry or services) in oreder to understand which economic activity is more prone to mafia infiltration.

The base of the Mafia Activity Indiex is the following regression:

```
summary(logit_unrestricted)

##
## Call:
## glm(formula = factor ~ ., family = binomial(), data = mafia_crime_pop[, -c(1:3, 25, 27)])
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.7519  -0.8878   0.1627   0.8917   2.2887
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.229e-01 7.206e-01 -0.726 0.46806
## population   -1.070e-06 1.632e-06 -0.656 0.51214
## ARSON        9.730e-03 5.419e-03  1.795 0.07260 .
## DAMARS       -1.563e-02 4.268e-03 -3.663 0.00025 ***
## DRUG          3.142e-03 2.638e-03  1.191 0.23355
## EXTORT       -1.040e-02 9.968e-03 -1.043 0.29692
## MAFIAHOM     -1.691e+00 1.220e+00 -1.386 0.16576
## MAFIASS       6.022e-01 5.630e-01  1.070 0.28477
## MENACE        4.898e-03 1.829e-03  2.679 0.00739 **
## MONEYLAU     -3.024e-02 1.297e-02 -2.331 0.01974 *
## PROSTI        -5.810e-02 3.985e-02 -1.458 0.14483
## SMUGGL        2.592e-01 1.009e-01  2.570 0.01017 *
## p100k_ARSON   1.886e-02 1.547e-02  1.219 0.22284
## p100k_DAMARS  4.649e-02 1.541e-02  3.018 0.00254 **
## p100k_DRUG    -1.542e-02 1.041e-02 -1.482 0.13846
## p100k_EXTORT  7.374e-02 3.671e-02  2.009 0.04456 *
## p100k_MAFIAHOM 1.493e+01 8.400e+00  1.777 0.07551 .
## p100k_MAFIASS  1.178e+00 1.799e+00  0.654 0.51283
## p100k_MENACE   -1.769e-02 6.543e-03 -2.704 0.00685 **
## p100k_MONEYLAU 2.106e-01 6.954e-02  3.029 0.00246 **
## p100k_PROSTI   2.591e-01 1.386e-01  1.869 0.06155 .
## p100k_SMUGGL   -6.053e-01 2.901e-01 -2.087 0.03693 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1075.44 on 852 degrees of freedom
## Residual deviance: 782.87 on 831 degrees of freedom
## AIC: 826.87
##
## Number of Fisher Scoring iterations: 12
```

As you can see, there are several variables which have no meaningful relation with dependent variable ($p > 0.05$). But we now, by other sources (news reports, government documents, etc.) know that each of this

variables is good to spot mafia and, therefore, it was included anyway with good results.

What now?

- Is this project newsworthy?
- Can this project be replicated?
- Is this scientifically accurate enough?
- What is we use this data as a training data set in a machine learning perspective? (Can this data-set be the training data for a machine learning model used to measure the activity of organized crime across Europe?)
- How can we make it ‘cool’ for the general audience?
- How do we integrate this kind of journalism with normal reporting?
- How do we organize newsrooms to make the most of this kind of work?
- Who pays for it?
- Can this project be the base for understanding other issues (suche as terrorism)?