Econ-S410: Seminars on Econometrics

25th April 2022

Weak States: Causes and Consequences of the Sicilian Mafia

Replication and extension of the study by D. Acemoglu, G. De Feo, G. De Luca

M. Dendooven, F. Funghi, L. Terlinden, E. Verhaeghe

ULB Solvay Brussels School Economics & Management

Table of Contents

Introduction

The spread of organized crime is a documented culprit for the loss of economic growth, competition, employment and social welfare in many developed Western countries. In Italy, country of origin of some of the most powerful criminal societies such as the Mafia, 'Ndrangheta and Camorra, the economic impact of organized crime has been estimated in a 16% reduction of GDP per capita, mainly given by a net loss of economic activity due to the replacement of private capital with less productive public investments1.

In this work, we are going to analyse and possibly expand a paper on this topic by D. Acemoglu, G. De Feo and G. De Luca entitled Weak States: Causes and Consequences of the Sicilian Mafia.

In chapter 1 we are going to present an overview of the paper, showing the main econometric specifications and the most important findings of the authors. Chapter 2 will be briefly devoted to listing some of the strengths and weaknesses of the paper. Chapter 3 will be finally dedicated to test some alternative econometric tools and hypothesis to try to extend the original work.

1. Overview of the paper

As organized crime is rapidly expanding also in Northern Europe2, which until a few decades ago was immune to this phenomenon, it is increasingly necessary for policymakers and law enforcement authorities to understand the factors that allow organized crime to grow and infiltrate the economic and social fabric, even to the point of replacing state authority in some extreme cases.

1.1 Assessing the origins of the 19th-century expansion of the Mafia

The literature on the subject is not necessarily scarce, but it lacks a comprehensive view of the phenomenon, often limiting itself only to measuring some specific aspects of organized crime, such as the impact on the corruption index, drug trafficking and money laundering activities. For this reason, Acemoglu, De Feo and De Luca attempted to empirically establish what were the environmental and social determinants that allowed one of the strongest and most notorious criminal syndicate, the Sicilian Mafia, to grow unhindered between the 19th and 20th centuries.

To do so, the authors tested a prevailing historical hypothesis that links the rise of the Mafia to the event of the Great Drought of 1893, which would have in turn led to the rise of a socialist workers' movement (*Fasci dei Lavoratori*) in Sicily. The conjecture is that the Mafia's use of force was sought by landowners, who feared the claims of the Fasci movement, and that this

[1] Pinotti, P. (2015). The economic costs of organised crime: Evidence from Southern Italy. *The Economic Journal*, *125*(586), F203-F232.
[2] European Parliament (2013). Organized crime in the European Union. *Library of the European Parliament.*

deployment has led the Mafia to remain permanently entrenched in the territory. The authors' first step is to assess if the drought influenced agricultural production, and subsequently on the rise of Fasci movement, through the following regression:

$$(1) \qquad Fasci_i = \gamma^{Fasci} \cdot relative\ rain_i^{1893} + X_i' \cdot \beta^{Fasci} + \varepsilon_i^{Fasci}$$

where $Fasci_i$ denotes the dummy variable designating the presence of the Fasci movement in municipality i, $relative\ rain_i^{1893}$ is the relative rainfall in the spring of 1893, X_i' is a vector of covariates (geographical controls, Fasci determinants, etc.) , and ε_i^{Fasci} is a random error term. Equation (1) represent the first stage of a 2SLS model.

The second stage aims to measure if the presence of the Fasci movement influenced the rise of the Mafia:

$$(2) \qquad Mafia_i = \alpha^{Mafia} \cdot Fasci_i + X_i' \cdot \beta^{Mafia} + \varepsilon_i^{Mafia}$$

where $Mafia_i$ it's an index of strength of the Mafia (from 0, absent, to 3, very strong presence) in municipality i, as reported by a police inspector's investigation in 1900, and α^{Mafia} is the coefficient of interest measuring the impact of Fasci on Mafia.

By the end of the 2SLS analysis, the authors were able to conclude that the drought of 1893 has a strong predictive power for the location of the Peasant Fasci, while other climatic and social events included in the controls did not show the same significance. In addition, the results of the second stage show that the presence of the Fasci movement in a municipality may explain up to 38% of the strength of the Mafia in 1900.

1.2 Medium- and long-term impact of Mafia

The second part of the paper was dedicated to the measurement of the medium- and long-term impact of Mafia presence on some key socio-economic variables, like the literacy rate, the healthcare quality, and the development expenditure. To conduct this analysis, the authors bypassed the causal channel linking the Mafia with the Fasci movement, and instead resorted to directly using the *1893 rainfall* variable as a source of variation of Mafia in 1900, using the following regression:

$$(3) \qquad Mafia_i = \gamma^{Mafia} \cdot relative\ rain_i^{1893} + X_i' \cdot \beta^{Mafia} + \varepsilon_i^{Mafia}$$

Equation (3) is used as a first stage of a 2SLS, whereas the second stage is expressed in this form:

$$(4) \qquad Y_i = \alpha^{y} \cdot Mafia_i + X' \cdot \beta^{y} + \varepsilon_i^{y}$$

With Y_i being the dependent variable measuring the socio-economic outcome of interest.

This second part of the study also showed some interesting results, especially regarding the medium-term effect in the first three decades of the 1900s. For instance, the authors show a strong effect size of Mafia on the reduction of the literacy rate, the increasing of infant mortality and the reduction of provision of public goods. At the same time, they found that Mafia had an important effect on the reduction of political competition, potentially explaining the large impact of the criminal syndicate on local economic development.

The long-term analysis shows some less clear-cut results. According to the authors, this is primarily due to several historical events not taken in account in this paper, such as the rise of fascism, that had a role in modifying the strength of the Mafia in the 20th century. A more detailed analysis of the influence of the Mafia on political party in the second half of the 1900s is offered in another paper by two of the authors[3].

2. Strengths and weaknesses of the paper

This paper represents one of the most comprehensive studies in the literature regarding the link between organized crime and economic development.

The authors managed to conduct an analysis with fragmentary data from over two centuries ago (such as those regarding drought intensity) through advanced statistical methods and interpolation processes on the available data sets. At the same time, they managed to collect a variate enough database from historical sources, that allowed them to include numerous control variables in the study, such as the average presence of sulphur mines, citrus groves, and olive and vineyards, which have already shown in precedent literature to be correlated with the presence of Mafia. Another area of merit is the presence of numerous robustness checks and falsification tests, e.g. proving that no rainfall level other than that of 1893 shows a link to the presence of Fasci or the intensity of the Mafia.

Some of the weaknesses of the paper that we found are mainly related to some inherent characteristics of the database. For example, we believe that data on Mafia presence may not be completely reliable and may be subject to different forms of bias. More specifically, the spatial distribution of Mafia in 1885 was gathered by a parliamentary enquiry, and it is possible to speculate that data from some municipalities may show underestimated Mafia levels for reasons of electoral and public opinion concern. At the same time, the *Mafia1900* intensity variable is obtained from the work of a police inspector, Cutrera, who compiled its report using information based on its own experience, so possibly influenced by a selection bias. Starting with underestimated data and finishing with overestimated data is creating a two-sided bias that can overestimate the impact of the cause (Here the drought of 1893). There is no way for us to

[3] De Feo, G., & De Luca, G. D. (2017). Mafia in the ballot box. *American Economic Journal: Economic Policy, 9(3), 134-67.*

make improvements to this problem as no additional data is available for this period. We have no choice but to trust the authors.

One last small remark concerns the title and premises of the paper: the role of a *weak state* on the spread of organized crime concerns just a small section of the study, and perhaps it could have been interesting to observe the effect of the transition to a strong, centralized state (the *fascist state*) on the strength of Mafia in Sicily.

3. Extending the paper

As mentioned above, this paper is very complete and is of high quality. Nevertheless, some clarifications could be made. Moreover, we soon discovered that other analysis techniques could be applied to further confirm the research question.

3.1 Geographical Difference-in-difference strategy

When we observed the average relative rainfall of the year 1893 for each municipality in Sicily on the map, we could directly notice that one half of the island was heavily impacted by the drought and the other "barely" not. Knowing this, we had the idea to run a difference-in-difference on the two parts of the island to show that drought had a high chance to have a causal impact on the mafia level on the island.

<u>Difference-in-difference conditions</u>

We used the median of the longitudes of all the municipalities to fix the line that would cut the island in two and define our control and treatment group. To make sure that the longitude chosen to fix the separation between the two groups is the right one, we tried to move the longitude limit to the left and to the right and see the results. But the representation of the two groups is less precise statistically (size equivalence and relative rainfall less interesting).

For the DID being valid, two crucial conditions must be met: first, the two groups must be similar and have had the same trend concerning the observed variable in the years before the treatment. Secondly, in one group the treatment must occur (treatment group), in the other not (control group). We have verified those two conditions with our data.

For the first point, we show on **table number 1** (see Appendix) that the size of the two groups is similar in number of municipalities. But for the point of the trend, our DID shows some weakness. Because the data on the presence of the mafia on the island was collected only twice at a given time (once before the treatment, once after). We cannot safely infer that the mafia pattern was the same for both groups before the drought. This finding prevents us from assuring that the difference in difference on mafia presence that will show up between the two groups after the treatment is largely due to the treatment.

For the second point, we use the function sum for the variable "Relative rainfall in spring 1893 interpolated from weather stations within 30km" for the two groups. The results are showed on the **table number 2**, We can observe that for the treatment group, the mean of the average relative rainfall for all his municipalities is 44%. With a STD of 24%. This result allows us to conclude that this side of the island is clearly undergoing a drought in the spring 1893 and is valid to be the treatment group. The results of the control group are more problematic. Indeed, with an average of 82% and a minimum of 6%, we assume that at least a small part of the control group is undergoing the "treatment" at least partially. This weakens the validity of the DID.

To conclude, even though some conditions of the DID are not perfectly respected, we decided to apply it on our data. We just must be aware that the results obtained might be biased.

Application of the difference-in-difference

Our DiD model take the following form:

$$Mafia_{it} = \beta_0 + \beta_1 Treat_{it} + \beta_2 After_{it} + \beta_3 Treat_{it} * After_{it} + X_i' \cdot \beta^{Mafia} + \varepsilon_i$$

The dependent variable $Mafia_{it}$ represents the density of Mafia across time in 1885 and 1900. $Treat_{it}$ is a dummy variable using the longitude separation, equals to 0 if the municipalities are in the control group and 1 if they are in the treatment group. $After_{it}$ is a dummy variable equals to 0 if it's before the treatment (1885) and 1 if it's after the treatment (1900). ε_i is the error term. β_3 is the coefficient that we are interested in because it really shows the difference of the spread of the Mafia due to the treatment itself

Adding the geographic control variables is a way to really capture the impact of the treatment on the mafia variable. When we don't control for other variables, the coefficient of the interaction may be over-/understated. This is even more crucial in our case where the mutual trend of the two groups is just an assumption and is not statically proven. For example, we added the variable "Citrus groves" as we've known from relevant literature that "areas where citrus fruits are the main crop are more likely to house the Mafia, which reflects not just the productivity of citrus farming but its greater vulnerability to vandalism and disruptions of irrigation"[4].

On **table number 3,** we can observe the results of the DID without control variables. The coefficient β_3 is equal to 0,63. The interpretation is that on average, only due to the drought, the West side of the island will have 0,63 more mafia presence on a scale from 0 to 3. That's significant.

[4] Dimico, A., Isopi, A., & Olsson, O. (2017). Origins of the Sicilian mafia: *The market for lemons. The Journal of Economic History, 77(4),* 1083-1115.

On **table number 4,** we can observe the results of the DID with the control variables. The coefficient $\beta 3$ is equal to 0,71. Interestingly, when we add up the control variables, we see that the treatment effect is even higher than without. Our interpretation is that without the drought (treatment), the East side of the island (control group) had a greater chance of having a higher growth of mafia presence than the west side.

On the **graph number 1**, we see the evolution of the mafia presence in the two periods for both groups. The results are not continuous as shown on the graph and only both ends of the graphs are correct (1885 and 1900). We can observe that the part between the green and the blue line on the 1900 x-axis is the part explained by the treatment.

We can conclude that the dependent variable (the mafia presence) was positively correlated with the treatment (the drought). Of course, these results must be taken "with a pinch of salt" given that the DID had the defects mentioned above.

3.2 Expanded 2SLS strategy

In the initial report, the relative rainfall in the spring of 1893 was used to explain the rise of the Peasant Fasci. We now want to try to add another explanatory variable to perform our analysis. In the dataset provided by the authors, we find that *ind_dir_ratio1884*, the ratio of indirect over direct local taxes collected by the local council, might be a good explanatory variable, uncorrelated with the relative rainfall. Our hypothesis is that municipalities with high rate of indirect taxation might be more vulnerable to workers discontent and can contribute to the rise of Peasant Fasci. We deploy in one case the indirect taxation as the only explanatory variable, while in another case we use both the taxation and relative rain. The resulting regression is the following:

$$Fasci_i = \gamma^{Fasci} \cdot relative\ rain_i^{1893} + \delta^{Fasci} \cdot taxation\ ratio_i^{1884} + X_i' \cdot \beta^{Fasci} + \varepsilon_i^{Fasci}$$

with δ^{Fasci} being the new coefficient of interest linking the indirect to direct taxation ratio to the presence of the Fasci.

From **table 5**, we can see that even that the First stage F statistic is quietly significant, the coefficients are really weak, and we can conclude that taxation had no clear impact to the rise of Peasant Fasci. Even though the coefficients of taxation are more significant in the second equation, it's still not satisfying to estimate Peasant Fasci. We decide to give up taxation.

For this reason, we had to find a more relevant instrument that can predict Peasant Fasci. We know from the study that there are some crops that suffer the drought more than others and our strategy is to select the share of crops that suffer the most from the drought, the grains. The variable "**Share of cultivated land devoted to grains in 1853**" is another instrument which is more relevant because it considers the share of crops that are the more vulnerable across the districts and municipalities. Moreover, this variable is independent because the share of cultivated land is a geographical variable. Note that some municipalities are more composed with crops of citrus and vinery, they are less vulnerable to the drought than municipalities which have as crops wheat and that sort of grains. Therefore, our new strategy is to use the share of cultivated land devoted to grains as an instrument.

$$Panel\ A\ (2IV):\ FasciGrains_i = \gamma^{Fasci} \cdot \text{share } grains_i^{1853} + Rain_i^{1893}\tau_i + X_i'\beta^{Fasci} + \varepsilon_i^{Fasci}$$

$$Panel\ B\ (1IV):\ FasciGrains_i = \gamma^{Fasci} \cdot \text{share } grains_i^{1853} + X_i'\beta^{Fasci} + \varepsilon_i^{Fasci}$$

In table 6, We clearly see that our First stage F statistic is much higher and the coefficients are more significant than in our previous analysis with taxation. There is two Panels A and B in which there are 2 different equations. There are 2 instrumental variables in the first equation with relative rainfall and the share of cultivated grains and the other with one instrumental variable. In the first column, we clearly see that one unit increase in share of cultivated grains can lead to one unit increase in Peasants Fasci. In the first column, the instrumental variable is alone without controls, in the second column, we compute the determinants of Fasci and in the third and fourth column we add respectively the determinants of Mafia and Geographic controls. We use the same controls as in the work except the share of cultivated land devoted to grains in 1853 that we use as an instrumental variable

3.3 Second stage using Ordered logit model

The variable Mafia is a variable which takes the values 0, 1, 2 or 3. It measures the intensity of the Mafia in the several districts. Our strategy is to measure the probability to have a high number of Mafia activity. We want to explore more the field of probabilities in this analysis, we'll strengthen the analysis especially by calculating the marginal effects of our equation below. Therefore, we use the ordered logit model to measure the probability of a high activity of the Mafia in function of several variables.

Suppose $Mafia_i$ is a continuous latent variable which is a linear function of the explanatory variables

$$Mafia1900_i^* = \beta_1 + \beta_2 FasciGrains_{ij} + \beta_3 RelativeRain1893_i + u_i$$

Where $FasciGrains_{ij}$ is our variable which has been estimated by our instrumental variable from 4 different manners. In the first column, the instrumental variable is alone without controls, in the second column, we compute the determinants of Fasci and in the third and fourth column we add respectively the determinants of Mafia and Geographic controls. We use the same controls as in the work except the share of cultivated land devoted to grains in 1853 that we use as an instrumental variable.

This why we'll compare the 4 different manners with the controls, and we'll also create another latent variable, with the original variable Peasants Fasci to see the difference. The equations will be like this:

$$Mafia1900_i^* = \beta_1 + \beta_2 PeasantFasci_i + \beta_3 RelativeRain1893_i + u_i$$

Our thresholds α_j are known it can be 'mapped' on an ordered multinomial variable as follows:

$Mafia1900_i = 0 \; if \; Mafia1900_i^* \leq 0{,}5$

$Mafia1900_i = 1 \; if \; 0{,}5 < Mafia1900_i^* \leq 1{,}5$

$Mafia1900_i = 2 \; if \; 1{,}5 < Mafia1900_i^* \leq 2{,}5$

$Mafia1900_i = 3 \; if \; 2{,}5 \leq Mafia1900_i^*$

Therefore, our log-likelihood is the following

$$\ln L = \sum_{n=1}^{N}\sum_{j=1}^{J} I[y_n = j] \ln\left(F(\alpha_j - \beta_1 - \beta_2 X_i) - F(\alpha_{j-1} - \beta_1 - \beta_2 X_i)\right)$$

Where:

$F(\alpha_j - \beta_1 - \beta_2 X_i) = \frac{1}{1+e^{\left(\alpha_j - \beta_1 - \beta_2 X_i\right)}}$ and $F(\alpha_j - \beta_1 - \beta_2 X_i) = \frac{1}{1+e^{\left(\alpha_j - \beta_1 - \beta_2 X_i\right)}}$

In table 7, Our chi test is relevant for all our 5 different order logits. the Likelihood is greater for the binary variable Peasants Fasci than our single instrumented variable Fascigrains0. The p value for this variable is too high and it's not significant. However, the other columns are more relevant and can predict the probability for one municipality to be in a certain group of Mafia, given the Fascigrains and the rainfall1893. The column 5, which regroup all of our controls is the most statistically significant with a level of chi test of 78.13 and the higher Likelihood. We can interpretate from that result that, for example, with one unit change in Fascigrains3, municipalities are 29% less likely to be part of the group 0 (No mafia) and municipalities are 22% more likely to be in an environment with high intensity of mafia.

Concerning the robustness of Peasant Fasci, from the **figure 3**, we can look at the marginal effects of the variable Mafia1900 based on the intensity of Peasants Fasci (0,1). We clearly see that in fact the probability to have no Mafia with no presence of Peasants Fasci is 40.82%. In contrary the probability to have a very high intensity of Mafia with a presence of Peasant Fasci is 41.92%. This model allows us to say that Peasants Fasci is a relevant variable.

3.4 Probit model

Hypothesis 2: we consider that the inspector Cutrera had a good intuition of the presence of the mafia, but not enough to be able to be very precise

Unlike the "ordered logit" method, we will have to modify the mafia variable so that it becomes binary. We will assign a value of 0 to municipalities with a mafia presence of {0,1} on the Cutrera scale, indicating a rather low presence; and a value of 1 for municipalities with a mafia presence of {2,3}, indicating a rather high presence.

Now, the OLS regression function has a binary dependent variable, which allows the use of a "probit" method, characterized by the function:

With:

$$Mafia1900_i = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$$

With:

$$z(i) = \alpha * Fasci + \beta X' + \varepsilon$$

where X' is a matrix of other explanatory variables.

This method allows us to establish a kind of percentage of mafia presence according to the level of presence of in a municipality.

3.5 *Cloglog et Gompit*

Beyond the "probit" model, 2 other models are possible: the "cloglog" model and the "gompit" model:

Cloglog :

$$Mafia1900_i = 1 - e^{-e^z}$$

Gompit :

$$Mafia1900_i = e^{-e^z}$$

With:

$$z(i) = \alpha * Fasci + \beta X' + \varepsilon$$

These two models bring nuance: Cloglog is a probit where the median individual has a higher probability of getting the value 1, while Gompit is a probit with a lower probability for the median individual to get the value 1.

Knowing which of the three models best fits the distribution gives information about the sensitivity of the presence of mafia to Fasci's presence: if the Gompit model fits best, then it implies that the situation of the farmers must be particularly severe for the chances of mafia presence to be close to 100%, while if the Cloglog model fits best, the relatively low drought intensity leaves room for a statistically greater presence of mafia.

3.1 How to determine the best model and results

We will use AIC/BIC in order to find out which model fits the best the data's ,

$$AIC = 2k - 2\,ln(L)$$

$$BIC = -2\,ln(L) + ln(n)k$$

With k the number of variables, n the number of observations and L the maximum likelihood function. The model obtaining the lowest value on these 2 statistics will be considered as the most adequate. As k and n are the same between the two models, the difference will only be in the maximum likelihood.

After running the code, we can see that the best model is the Cloglog model: it produces the lowest AIC/BIC results (**Table 8 ,9 ,10**). This outcome was predictable: the hypothesis is that the Fasci movement was tackled by the government by using Mafia. The consequence is that even a low level of presence was already seen as a potential danger for government and needed to be countered. We also find that a raise by one unit of the Fasci variable increase by 16% the probability of the presence of Mafia (**Table 11)**.

4. Conclusion

This article not only describes a fascinating subject but is also a statistical success. The authors have gathered a very large panel of data and have tried to be as precise as possible in their choice of statistical techniques. We tried to complete the work of the authors in order to bring different angles of view to make it more complete. The DID technique allowed us to confirm the authors' hypothesis in an alternative way. We thought that there was a room to exploit with predicted probabilities, it was a way to look at the probability for each municipality to be part of a certain group of density of Mafia (0,1,2,3). Therefore, we used the ordered logit and the probit/gompit/cloglog comparison that allowed us to find coherent result with the hypothesis of the authors using other statistical methods like ML.

References

Acemoglu, D., De Feo, G., & De Luca, G. D. (2020). Weak states: Causes and consequences of the Sicilian Mafia. *The Review of Economic Studies*, *87*(2), 537-581.

De Feo, G., & De Luca, G. D. (2017). Mafia in the ballot box. *American Economic Journal: Economic Policy, 9(3), 134-67.*

Dimico, A., Isopi, A., & Olsson, O. (2017). Origins of the Sicilian mafia: *The market for lemons. The Journal of Economic History, 77(4),* 1083-1115.

European Parliament (2013). Organized crime in the European Union. *Library of the European Parliament.*

Pinotti, P. (2015). The economic costs of organised crime: Evidence from Southern Italy. *The Economic Journal*, *125*(586), F203-F232.

Appendix

Table 1 - Frequency of municipalities in each group (1 = treatment group)

Longitude of the main centre of the municipality	Freq.	Percent	Cum.
0	168	50.45	50.45
1	165	49.55	100.00
Total	333	100.00	

Table 2 - Description of the variable "Relative rainfall in spring 1893 interpolated from weather stations within 30km" for the control and treatment group

```
. sum sp3m1893_n30c if treatment == 1
```

Variable	Obs	Mean	Std. dev.	Min	Max
sp3m1893~30c	**290**	**.4605537**	**.2513125**	**.135082**	**1.049212**

```
. sum sp3m1893_n30c if treatment == 0
```

Variable	Obs	Mean	Std. dev.	Min	Max
sp3m1893~30c	**304**	**.8190527**	**.1760885**	**.0613658**	**1.283307**

Table 3 - DID without geographic controls

Linear regression	Number of obs	=	606
	F(3, 602)	=	152.29
	Prob > F	=	0.0000
	R-squared	=	0.3396
	Root MSE	=	.9402

Mafia	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
time	.4495074	.0947671	4.74	0.000	.2633932	.6356216
treatment	.7069264	.1040432	6.79	0.000	.5025948	.9112581
did	.6292805	.1547043	4.07	0.000	.3254549	.9331061
_cons	.2142857	.0388352	5.52	0.000	.1380168	.2905547

Table 4 - DID with geographic controls

```
Linear regression                               Number of obs     =        560
                                                F(18, 541)        =      37.11
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4180
                                                Root MSE          =     .89002
```

Mafia	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
time	.3395541	.0937374	3.62	0.000	.1554202	.523688
treatment	.2911381	.1777664	1.64	0.102	-.0580589	.640335
did	.70899	.1537007	4.61	0.000	.4070667	1.010913
lnpop1861	.1626049	.0677513	2.40	0.017	.029517	.2956927
lnsurface	-.0135737	.0648368	-0.21	0.834	-.1409364	.113789
centreheight	-.0003794	.000379	-1.00	0.317	-.0011239	.0003651
maxheight	-.0003735	.0003573	-1.05	0.296	-.0010754	.0003285
slope2	.001188	.0007101	1.67	0.095	-.0002069	.0025829
pa_pdist1856	-.0034969	.0013456	-2.60	0.010	-.0061401	-.0008537
port2_pdist1856	-.000445	.0024506	-0.18	0.856	-.0052589	.0043689
roads1799	.0427146	.0848425	0.50	0.615	-.1239465	.2093757
ave_temp	-.068649	.0661262	-1.04	0.300	-.1985445	.0612465
var_sp3m_n30	-1.032135	.6017317	-1.72	0.087	-2.214152	.1498819
sp3m_ave_n30	7.96e-06	.0016505	0.00	0.996	-.0032342	.0032501
Citrus_groves	5.311924	2.405557	2.21	0.028	.586547	10.0373
sulfurproduction1868_70	.0057341	.0020532	2.79	0.005	.001701	.0097673
Vineyards	.8032862	.3277086	2.45	0.015	.159549	1.447023
Olives_groves	-1.513875	.6644951	-2.28	0.023	-2.819182	-.2085681
_cons	.8034332	1.433783	0.56	0.575	-2.013031	3.619897

Table 5 Impact on Peasant Fasci, taxation as IV

Dependent variable : Peasant Fasci		Fasci controls	Fasci and Mafia controls	Fasci, Mafia, Geographical controls
Panel A : 2 IV Tax and rainfall as IV				
Rainfall	-1.02 (0.086)	-0.97 (0.088)	-0.97 (0.088)	-0.79 (0.134)
Ind. Tax	0.0027 (0.0334)	0.0042 (0.0034)	0.0056 (0.0034)	0.0064 (0.0037)
R squared	0.3717	0.43	0.4612	0.4814
First stage F statistic	70.71	22.01	15.01	8.39
Panel B : Tax as IV				
Ind. Tax	0.0066 (0.0039)	0.0066 (0.0038)	0.0066 (0.0038)	0.0086 (0.0038)
R squared	0.01	0.146	0.204	0.4064
First stage F statistic	2.92	6.40	5.49	6.73

Table 6 - Impact on Peasant Fasci, share of cultivated grains as IV

Dependent variable : Peasant Fasci		Fasci controls	Fasci and Mafia controls	Fasci, Mafia, Geographical controls
Panel A : 2 IV Share of Cultivated grains and rainfall as IV'S				
Rainfall	-0.94 (0.086)	-0.94 (0.088)	-0.93 (0.088)	-0.782 (0.135)
Share Cultivated grains	0.29 (0.0334)	0.047 (0.135)	0.268 (0.16)	0.2622 (0.19)
R squared	0.375	0.4147	0.4447	0.4632
First stage F statistic	72.60	23.99	15.49	8.29
Panel B : Share Cultivated grains as IV				
Share Cultivated grains	0.63 (0.13)	0.37 (0.16)	0.7 (0.184)	0.287 (0.2)
R squared	0.0855	0.1391	0.2	0.3889
First stage F statistic	25.33	7.16	5.93	6.65

Table 7 - Probabilities of municipalities to be part of a certain group of Mafia density

Dependent variable Mafia 1900	PeasantsFasci (and rainfall1893)	Fascigrains0 variable (and rainfall1893)	Fascigrains1 variable (and rainfall1893)	Fascigrains2 variable (and rainfall1893)	Fascigrains3 variable (and rainfall1893)
Outcomes					
Pr. Mafia = **0**	-0.12 (0.6)	0.02 (0.74)	-0.32 (0.67)	-0.16 (0.7)	-0.29 (0.52)
Pr. Mafia = **1**	-0.04 (0.17)	0.005 (0.2)	-0.1 (0.2)	-0.04 (0.19)	-0.08 (0.15)
Pr. Mafia = **2**	0.06 (-0.31)	-0.01 (-0.364)	0.17 (-0.35)	0.08 (-0.35)	0.15 (-0.27)
Pr. Mafia = **3**	0.10 (-0.46)	-0.013 (-0.57)	0.26 (-0.53)	0.125 (-0.54)	0.22 (-0.4)
LR chi2	75.16	70.82	76.82	72.92	78.13
Likelihood	-300.07	-302.24	-299.24	-301.19	-298.59

Table 8 - CloglLog

```
. cloglog Mafia1900bin peasants_fasci lnpop1861 lnsurface centreheight maxheight slope2 pa_pdist1856 p
> ort2_pdist1856 roads1799 ave_temp var_sp3m_n30 sp3m_ave_n30

Iteration 0:   log likelihood = -138.10399
Iteration 1:   log likelihood = -107.78605
Iteration 2:   log likelihood = -106.66441
Iteration 3:   log likelihood = -106.65944
Iteration 4:   log likelihood = -106.65944

Complementary log-log regression                Number of obs     =        253
                                                Zero outcomes     =        135
                                                Nonzero outcomes  =        118

                                                LR chi2(12)       =     136.27
Log likelihood = -106.65944                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
   Mafia1900bin | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
---------------+----------------------------------------------------------------
 peasants_fasci |   .8151505   .2623418     3.11   0.002     .3009701    1.329331
      lnpop1861 |   .8972583   .2165461     4.14   0.000     .4728357    1.321681
      lnsurface |  -.3353258   .1885797    -1.78   0.075    -.7049353    .0342836
   centreheight |   .0011463   .0013899     0.82   0.410    -.0015778    .0038704
      maxheight |  -.0003964   .0011691    -0.34   0.735    -.0026878    .0018949
         slope2 |   .0020157   .0023746     0.85   0.396    -.0026384    .0066699
   pa_pdist1856 |   -.015517    .002807    -5.53   0.000    -.0210187   -.0100153
port2_pdist1856 |   .0036937   .0069103     0.53   0.593    -.0098502    .0172376
      roads1799 |   .1885107   .2583544     0.73   0.466    -.3178546    .6948761
       ave_temp |  -.1608058    .251106    -0.64   0.522    -.6529645     .331353
   var_sp3m_n30 |  -5.616678   1.858259    -3.02   0.003    -9.258798   -1.974558
   sp3m_ave_n30 |  -.0153217   .0056751    -2.70   0.007    -.0264447   -.0041986
          _cons |   .6866181    5.13919     0.13   0.894    -9.386009    10.75925
------------------------------------------------------------------------------

. |
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
           . |        253  -174.7947  -106.6594      13    239.3189   285.2529
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

Table 9 - Gompit

```
| gen newMafia1900bin = (Mafia1900bin==0)

.
. cloglog newMafia1900bin peasants_fasci lnpop1861 lnsurface centreheight maxheight slope2 pa_pdist185
> 6 port2_pdist1856 roads1799 ave_temp var_sp3m_n30 sp3m_ave_n30

Iteration 0:   log likelihood = -201.42825
Iteration 1:   log likelihood = -179.94382
Iteration 2:   log likelihood = -179.72049
Iteration 3:   log likelihood = -179.72027
Iteration 4:   log likelihood = -179.72027

Complementary log-log regression               Number of obs     =        307
                                               Zero outcomes     =        172
                                               Nonzero outcomes  =        135

                                               LR chi2(12)       =      61.68
Log likelihood = -179.72027                    Prob > chi2       =     0.0000

--------------------------------------------------------------------------------
newMafia1900bin | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
----------------+---------------------------------------------------------------
 peasants_fasci |  -1.064751   .2767844    -3.85   0.000    -1.607238   -.5222632
      lnpop1861 |  -.1464195   .1561304    -0.94   0.348    -.4524295    .1595905
      lnsurface |   .4439972   .1332791     3.33   0.001      .182775    .7052195
   centreheight |   -.000922   .0009511    -0.97   0.332     -.002786    .0009421
      maxheight |   .0001062   .0008006     0.13   0.894    -.0014629    .0016753
         slope2 |  -.0007582   .0015904    -0.48   0.634    -.0038752    .0023589
    pa_pdist1856 |   .0056772   .0019001     2.99   0.003      .001953    .0094014
port2_pdist1856 |    .008012   .0052987     1.51   0.131    -.0023734    .0183973
      roads1799 |  -.0640906   .2161868    -0.30   0.767     -.487809    .3596278
       ave_temp |   .0648104   .1661907     0.39   0.697    -.2609174    .3905382
   var_sp3m_n30 |   1.848402   1.237816     1.49   0.135    -.5776729    4.274476
   sp3m_ave_n30 |   .0046188   .0039338     1.17   0.240    -.0030913    .0123289
          _cons |  -5.160408   3.509286    -1.47   0.141    -12.03848    1.717666
--------------------------------------------------------------------------------

.
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |          N   ll(null)  ll(model)      df        AIC        BIC
-------------+---------------------------------------------------------------
           . |        307  -210.5611  -179.7203      13   385.4405   433.8896
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.
```

Table 10 - AIC/BIC probit

```
. estat ic

Akaike's information criterion and Bayesian information criterion
```

Model	N	ll(null)	ll(model)	df	AIC	BIC
.	253	-174.7947	-108.0054	13	242.0108	287.9449

```
Note: BIC uses N = number of observations. See [R] BIC note.
```

Table 11 - Margins Cloglog model

```
. margins, dydx(peasants_fasci )

Average marginal effects                                Number of obs = 253
Model VCE: OIM

Expression: Pr(Mafia1900bin), predict()
dy/dx wrt:  peasants_fasci

-------------------------------------------------------------------------------
               |            Delta-method
               |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
---------------+---------------------------------------------------------------
peasants_fasci |   .1627781    .049218     3.31   0.001     .0663125    .2592437
-------------------------------------------------------------------------------
```

Figure 1

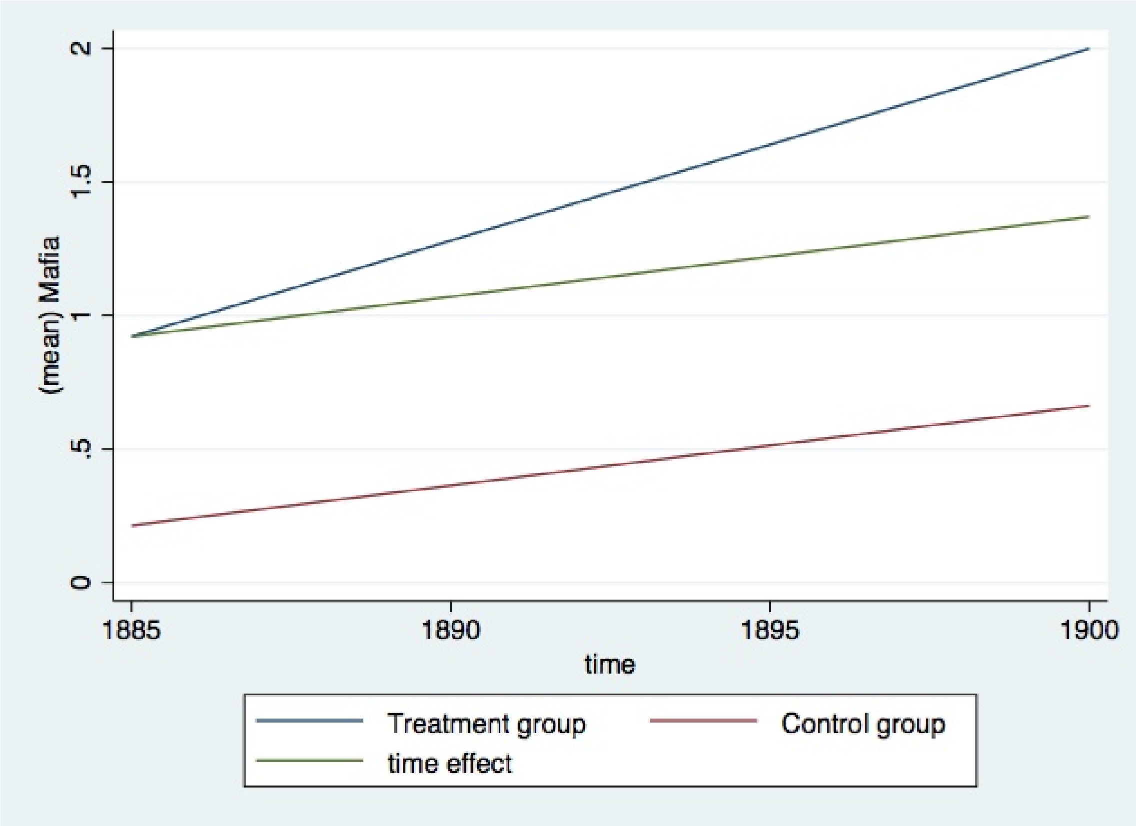

(mean) Mafia
0
.5
1
1.5
2
1885
1890
1895
1900
time
Treatment group
Control group
time effect

Figure 2

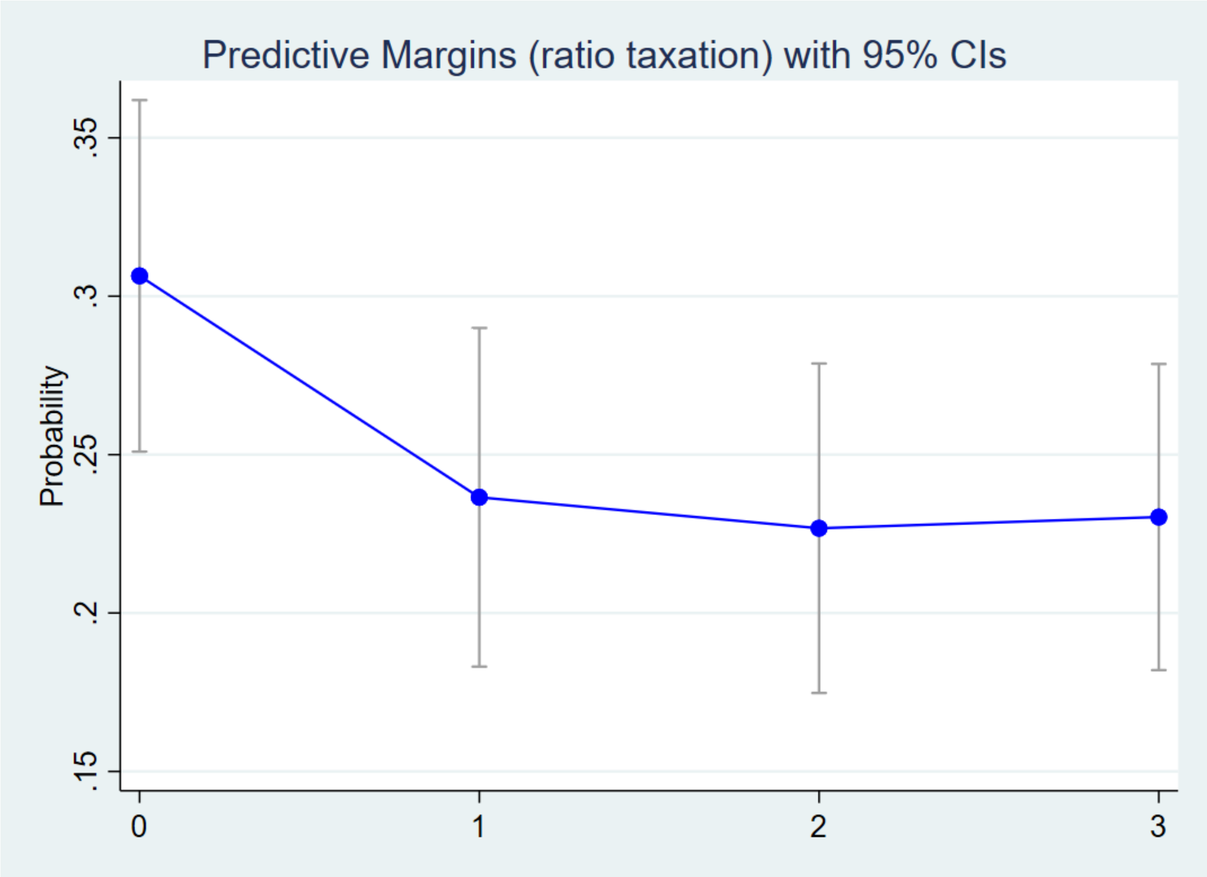
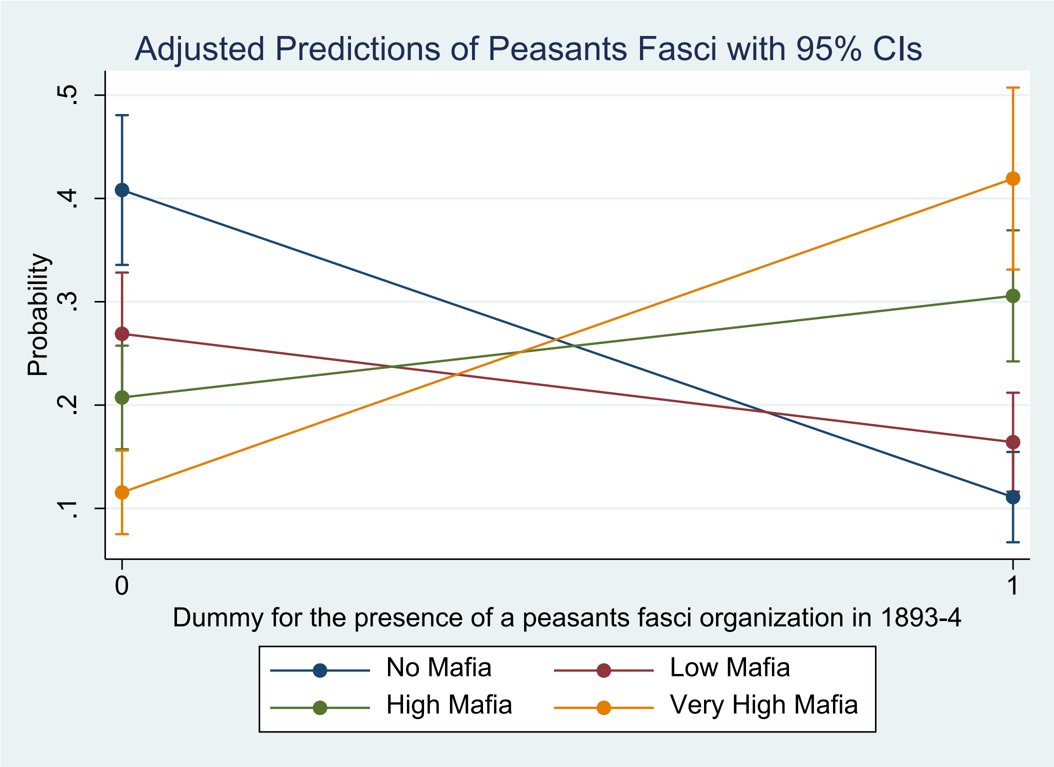

Predictive Margins (ratio taxation) with 95% CIs
Probability
.15
.2
.25
.3
.35
0
1
2
3

Figure 3



Adjusted Predictions of Peasants Fasci with 95% CIs
Probability
.1
.2
.3
.4
.5
0
1
Dummy for the presence of a peasants fasci organization in 1893-4
No Mafia
Low Mafia
High Mafia
Very High Mafia