BANCA D'ITALIA
EUROSISTEMA

UIF
Unità di Informazione Finanziaria per l'Italia

Quaderni dell'antiriciclaggio

Analisi e studi

The risk of mafia infiltration in Italian municipalities: Statistical and machine learning evidence from financial information

Stefano Iezzi and Claudio Pauselli

aprile 2025

numero 27

BANCA D'ITALIA
EUROSISTEMA

UIF
Unità di Informazione Finanziaria per l'Italia

Quaderni dell'antiriciclaggio

Analisi e studi

The risk of mafia infiltration in Italian municipalities: Statistical and machine learning evidence from financial information

Stefano Iezzi and Claudio Pauselli

n. 27 – aprile 2025

La collana Quaderni dell'antiriciclaggio *ha la finalità di presentare statistiche, studi e documentazione su aspetti rilevanti per i compiti istituzionali dell'Unità d'Informazione Finanziaria per l'Italia.*

La collana si articola in diversi filoni: il filone Statistiche *presenta, con periodicità semestrale, statistiche sulle segnalazioni ricevute e dati sulle attività dell'Unità; il filone* Rassegna normativa *illustra i principali aggiornamenti della normativa e della giurisprudenza in materia AML/CFT; il filone* Analisi e studi *comprende contributi sulle tematiche e sui metodi in materia di contrasto al riciclaggio e al finanziamento del terrorismo. I lavori pubblicati riflettono esclusivamente le opinioni degli autori, senza impegnare la responsabilità delle Istituzioni di appartenenza.*

Unità di Informazione Finanziaria per l'Italia
Per la pubblicazione cartacea: autorizzazione del Tribunale di Roma n. 1942013 del 30 luglio 2013
Per la pubblicazione telematica: autorizzazione del Tribunale di Roma n. 1932013 del 30 luglio 2013

Direttore responsabile
Enzo Serata

Indirizzo
Largo Bastia, 35 – 00181 Roma – Italia

Telefono
+39 0647921

Sito internet
https://uif.bancaditalia.it/

ISSN 2283-3498 (stampa)
ISSN 2283-6977 (online)

Stampato nel mese di aprile 2025

Grafica e stampa a cura della Divisione Editoria e stampa della Banca d'Italia

The risk of mafia infiltration in Italian municipalities: Statistical and machine learning evidence from financial information

Stefano Iezzi and Claudio Pauselli[1]

Abstract

This study investigates the risk of criminal infiltration in Italian local administrations by analysing municipalities' financial data from 2016 to 2021 using statistical and machine learning techniques. If compared with a control group, the budgets of city councils being dissolved for mafia infiltration exhibit higher operating costs, lower spending on education and social services, increased budget rigidity, and misallocation of funds toward sectors like construction and waste management. These municipalities also display lower revenue collection efficiency, likely due to decreased administrative capability or deliberate under-collection of taxes from selected individuals and businesses. A machine learning model achieves 98.2% accuracy (AUC) in predicting infiltration risks. As a validation, the risk of infiltration is found to be correlated with both the presence of organized crime-linked businesses and opacity in public procurement. This research offers a data-driven framework for identifying criminal influence in local administrations, supporting anti-mafia policies and more effective resource allocation.

Sommario

Questo studio analizza il rischio di infiltrazione criminale nelle amministrazioni locali italiane attraverso l'analisi dei dati finanziari municipali dal 2016 al 2021, impiegando tecniche statistiche e di *machine learning*. Rispetto a un gruppo di controllo, i bilanci dei comuni sciolti per infiltrazione mafiosa evidenziano costi operativi più elevati, una riduzione delle spese per istruzione e servizi sociali, una maggiore rigidità di bilancio e una allocazione impropria di fondi verso settori come l'edilizia e la gestione dei rifiuti. Tali comuni mostrano anche una minore efficienza nella riscossione delle entrate, attribuibile a una diminuzione delle capacità amministrative o a una deliberata sotto-riscossione delle imposte da individui e imprese selezionati. Lo studio sviluppa un algoritmo di *machine learning* che, nel valutare il rischio di infiltrazione, raggiunge un'accuratezza del 98,2% secondo la metrica dell'AUC. Secondo gli esercizi di validazione effettuati, il rischio di infiltrazione misurato dal modello risulta correlato con la presenza nello stesso territorio di imprese legate alla criminalità organizzata e con l'opacità nei processi di appalto pubblico dei comuni interessati. Questa ricerca fornisce un approccio basato su dati empirici per l'identificazione dei rischi di contaminazione criminale delle amministrazioni locali, contribuendo al sostegno delle politiche anti-mafia e a un'allocazione delle risorse più efficace per contrastare il fenomeno.

JEL Classification: C53, C63, H79, K42.

Keywords: organized crime, municipalities, mafia infiltration, financial statements, machine learning.

[1] Italy's Financial Intelligence Unit, Banca d'Italia.

Contents

1. Introduction and motivation[1]

Criminal organizations pose a substantial threat to the socio-economic stability of countries worldwide, as their activities hinder economic growth (Peri, 2004; Pinotti, 2015; Mocetti and Rizzica, 2023) and disrupt the functioning of lawful businesses (De Simoni, 2022; Mirenda et al., 2022; Le Moglie and Sorrenti, 2022; Arellano-Bover et al., 2024). Beyond their economic impact, these organizations exert far-reaching influence in the political sphere, infiltrating public institutions to advance their objectives. This phenomenon is especially evident in local administrations (Daniele and Marani, 2011; Daniele and Geys, 2015; Alesina et al., 2019), where control of resources and decision-making processes provide criminal groups with critical leverage.

The infiltration of organized crime into municipalities represents a severe and widespread threat to the rule of law and local democracy. It not only affects public order but also undermines the management of public services, the fairness in administrative decisions, and public trust in democratic institutions. Criminal organizations are particularly attracted to local administrations due to the significant financial resources they control. While it may seem surprising that sophisticated criminal organizations target small local governments, their strategy lies in their need to root themselves in and control their territory, imposing their own rules on civil society (Mete, 2008; Sciarrone, 2009).

Criminal organizations infiltrate local institutions primarily through two main methods: direct and indirect. Direct infiltration occurs when criminal organizations place their members or affiliates into public office by manipulating election results with votes they control. This approach is especially effective in smaller communities, where altering a relatively small number of votes can secure significant influence over local administration. Indirect infiltration, on the other hand, involves corruption and intimidation of legitimately elected officials. This method is more common in larger communities, where mafia-controlled votes are insufficient to achieve outright political dominance. In such cases, these votes serve as leverage in a broader web of corruption and coercion. This network includes colluding politicians, public officials, professionals, and entrepreneurs, all working together to divert administrative actions from their intended public goals to serve the interests of the criminal organization.[2] The ultimate aim is to subvert the administrative process for the benefit of the criminal organization, compromising the integrity and functionality of local governance.

In recent years, research on the infiltration of organized crime in municipalities has gained momentum. The economic consequences of infiltration have been extensively explored by Di Cataldo and Mastrorocco (2022), who demonstrate how organized crime's involvement in local politics can significantly affect the allocation of public resources. Their findings reveal that, under mafia influence, municipalities tend to direct funds into specific sectors, such as construction and waste management, while reducing investments in law enforcement and public transportation. This misallocation of resources can inhibit economic growth and hamper local development.

Scholars have also explored other dimensions of the infiltration phenomenon, particularly its impact on local elections and public procurement. Baraldi et al. (2022) examine how criminal organizations influence local elections by manipulating voter behaviour, while Daniele and Dipoppa (2017) and Alesina et al. (2019) highlight the role of political violence as a tool for influencing election outcomes and politicians' behaviour. Other studies, such as those by Barone and Narciso (2015) and

[1] The views expressed in this paper are those of the authors and do not necessarily represent those of the Financial Intelligence Unit for Italy, or those of Banca d'Italia. We are grateful to Mario Gara, Andrea Silvestrini, and seminar participants at UIF for their very useful comments, and to Giuseppe De Feo for an accurate discussion of the paper. We would also like to thank Marco De Simoni and Marianna Siino for their help in using the data employed for the model validation exercises.

[2] Recently, evidence suggesting of networking efforts by the organized crime has been provided by Arellano-Bond et al. (2024).

Ravenda et al. (2020), focus on the manipulation of public procurement processes, revealing how mafia infiltration leads to misallocation of resources, increased corruption, and inefficiencies in public spending.

Against this background, the aim of this paper is to evaluate the risk of infiltration in Italian municipalities using an extensive set of financial and balance sheet data from the municipal administrations. To achieve this, our focus is on one of the most decisive policies against organized crime: the dismissal of city councils. This impactful measure enables the central government to restore control and reestablish legitimacy in areas where the mafia exerts substantial influence over local politics. If there is evidence of connections between criminal organizations and local politicians, the central government has the authority to dissolve the entire municipal government and appoint a temporary technocratic administration.

Dismissing city councils due to mafia infiltration has proven to foster economic growth and investment in neighbouring areas. Research by Galletta (2017) suggests that this measure leads to reduced public investments in nearby municipalities due to improved law enforcement and decreased misconduct. Additionally, Fenizia and Saggio (2024) highlight that these dismissals boost employment, increase the number of firms, and raise industrial real estate prices in mafia-dominated sectors, indicating that dissolving infiltrated councils weakens mafia influence and restores trust in local institutions.

Our study contributes to this literature in two key ways. Firstly, by comparing municipalities that have been dissolved with a custom-built control sample of municipalities that have never been dissolved, we can statistically examine the unique characteristics of infiltrated municipalities in terms of economic and financial conditions. Secondly, we construct a machine learning model capable of reliably assessing the risk of infiltration in all Italian municipalities, including those which have never been dissolved. Machine learning methods are well-suited to address predictive challenges (Varian, 2014). These techniques are specifically designed to reduce prediction errors on out-of-sample data and to perform effectively on new, unseen information (Athey and Imbens, 2017; Athey and Imbens, 2019; de Blasio et al., 2022).

Our analysis leverages information extracted from local elections and municipalities' budget data spanning the years 2016 to 2021. The best-performing machine learning model demonstrates a very high level of accuracy in correctly detecting dissolved and non-dissolved municipalities. This performance aligns with, and in some cases surpasses, results obtained in related research fields.

Our paper is not the first in the literature to develop quantitative methods for assessing the risk of infiltration in municipalities. Nevertheless, only a few studies have contributed to this evolving field of research, leveraging a limited variety of data sources and analytical techniques to identify patterns and anomalies indicative of infiltration. Specifically, Abbattista et al. (2020) develop a logistic regression model to predict criminal infiltration using socio-economic data and crime statistics at the municipal level. Eboli et al. (2021) apply logistic regression and discriminant analysis models to provide statistical indicators capable of identifying potential infiltration based on a limited number of municipal budget indicators. Very recently, in the domain of artificial intelligence, Campedelli et al. (2024) develop a machine learning model that also utilizes budget indicators. Their work closely aligns with ours in terms of techniques employed, although we use a broader set of financial indicators and a more robust control sample.[3]

A key distinguishing factor of our research is the incorporation of revenue indicators and a wide set of financial indicators from the Plan of Indicators and Expected Budget Results, in addition to the spending-side financial variables used by Campedelli et al. (2024). Our findings suggest that, alongside

[3] There are other contributions that apply machine learning in the broader field of financial fraud detection (Chengwei et al., 2015; Sharma and Panigrahi, 2013; Wyrobek, 2020) and crime prevention (Kleinberg et al., 2018; Campedelli, 2022). More closely related to our paper, some studies attempt to identify companies connected to organized crime through their financial statements (Ravenda et al, 2015; Cariello et al., 2024).

spending indicators, revenue indicators, such as tax autonomy and efficiency in waste tax collection, are among the most important predictors of infiltration risk. By broadening the scope of financial data analyzed, our research provides a more comprehensive understanding of the economic behaviors associated with infiltration. This expanded focus not only validates specific predictors used in prior studies but also introduces new dimensions of fiscal analysis, thereby contributing significantly to the field.

Our study contributes to the relevant literature by not only expanding upon previous research with the use of an unprecedentedly large set of financial information but also by offering significant methodological advancements. Predicting mafia infiltration risk in Italian municipalities involves challenges due to the concentration of dissolved municipalities in economically disadvantaged Southern provinces, which complicates the disentanglement of budgetary impacts from socio-economic factors. To address these challenges, we implement a dual strategy. Firstly, we compare dissolved municipalities with a control sample of never-dissolved ones, selected for their statistical similarity in socio-economic variables, and incorporate these factors into our analytical models. This approach strengthens our study's capacity to accurately identify the impact of infiltration. Furthermore, to mitigate biases from potentially undetected infiltration in non-dissolved municipalities, we construct the control sample using unique data from Italy's Financial Intelligence Unit, drawing from provinces with presumed low infiltration levels in the business fabric. Addressing the issue of mislabelled data is crucial in classification tasks; if left uncorrected, it can significantly impair model performance and reliability, introducing bias into the results.

Furthermore, we validate our machine learning model through two analyses: one linking infiltration risk to the presence of mafia-connected businesses, and another examining its association with opacity in public procurement using data from Italy's Central Anti-Corruption Authority (ANAC). These validation exercises, focused on regions with pervasive criminal activity (Campania, Apulia, Calabria, and Sicily), support the model's discriminatory capability.

The remainder of the paper is organized as follows. Section 2 describes the institutional background related to the law on the dissolution of infiltrated municipalities. Section 3 details all the data employed in the study. Section 4 discusses how the control sample of never-dissolved municipalities is constructed. Section 5 presents the statistical analysis through which we compare dissolved municipalities with the control sample in terms of financial information. Section 6 outlines the machine learning approach, the main results, and a robustness analysis, while Section 7 provides evidence from two validation exercises with external sources. Finally, Section 8 offers some concluding remarks.

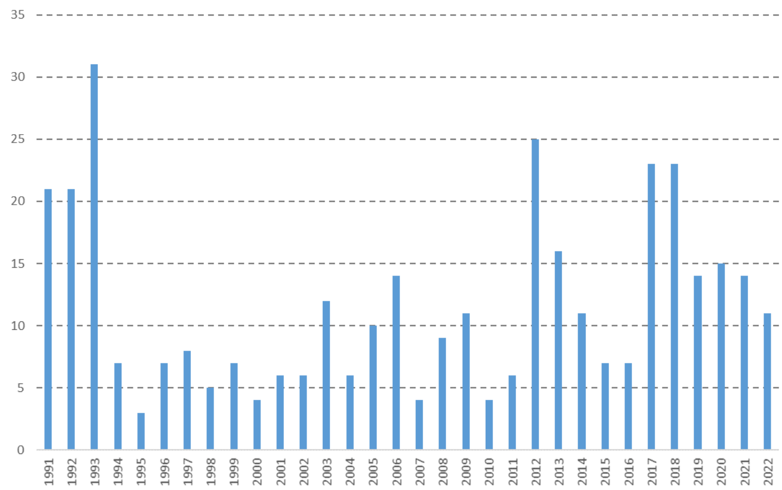
2. Institutional background and the list of infiltrated municipalities

In response to the mafia's growing influence on local governments in the 1980s, the Italian parliament introduced a law in 1991 to dismiss city councils (Decree-Law No. 164, Article 1). The dissolution of local administrations due to infiltration and mafia-related influences — introduced during one of the most challenging periods in the struggle between the State and the mafia and subject to numerous amendments over the years — has been comprehensively regulated by the *Testo Unico delle leggi sull'ordinamento degli enti locali* (Consolidated Laws on the Organization of Local Authorities) since 2000. This measure is adopted when evidence emerges regarding direct or indirect connections of administrators with organized crime or coercion over the administrators themselves. It does not require overwhelming proof of the local administration's connection to criminal organizations but is based on "well-founded suspicion". The law establishes that the national government can decree the dissolution of a local administration when there is evidence of direct or indirect connections between members of the municipal council and criminal organizations, which compromise the free will of the elective body and the proper functioning of the municipal administration. The dissolution of the council results in the removal of councillors, mayors, members of the executive board, and any other position related to the offices held. The aim of this measure is to restore the conditions of "normal" legality in "compromised" situations.

The dissolution due to mafia infiltration occurs in two phases. In the first phase, based on investigations initiated in other areas, suspicions of "influence" by organized crime are identified, which could compromise the entity's ability to make "autonomous" decisions. The Prefect then appoints an Investigative Commission, which establishes itself in the local administration under exam and conducts ad hoc investigations, producing a report. Based on the report, the Prefect decides whether to request dissolution from the Minister of the Interior. The formal decision of dissolution is made by a Presidential Decree (DPR) on the government's proposal. In the second phase, once the dissolution decision is made, the local administration is placed under a commissioner's management for a period lasting between 18 to 24 months or, in any case, until new political elections. Since it is composed of unelected individuals, the commission is responsible only for the municipality's ordinary management; it cannot approve new budgets or decide on investments.

Since the law's first implementation in 1991, the dissolution of local authorities has been widely practiced, particularly in the Southern regions of Italy. From 1991 to 2022, there were 368 dissolutions due to mafia infiltration, involving 267 municipalities (Figure 1); there were 74 municipalities that were dissolved more than once. Eighty-nine percent of the dissolutions occurred in the three Southern regions traditionally linked to organized crime (Campania, Calabria, and Sicily). However, in recent years, dissolution has also been applied in some municipalities in Northern Italy. From 1991, this phenomenon has affected over 4.8 million inhabitants, approximately 8.1 percent of the Italian population. In the South, 10.2 percent of municipalities have been dissolved at least once, and the percentage of the population involved over the total is 32.8 percent.

Figure 1 - Number of municipalities dissolved due to mafia infiltration over the 1991-2022 period



Note: The data is updated as of January 2024. The count includes dissolution orders that were subsequently cancelled by the administrative court, as well as orders of dissolution of the same municipality more than once over time.

3. The data

3.1 Financial budget indicators

Our main source of data consists of the financial statements of Italian municipalities for the period 2016-2021, derived from the harmonized final certificates from the Public Administration Database (BDAP) of the State General Accounting Office. As prescribed by Article 227 of the Unified Text of Local Authorities (Law 267 of 2000), municipal councils approve the final account (statement) of the previous year by April 30th. This final account includes, among other things, the budget statement, income statement, balance sheet, and the plan of indicators and expected results. With the harmonization of accounting systems and budget models for local authorities, as provided by Legislative Decree 118/2011, which became fully operational with the 2016 final account, the budgets of the municipalities became directly comparable and overlapping.

The harmonized certificates contain annual statistics on public spending, both in capital and current accounts, and are disaggregated into 23 specific spending categories known as missions, reflecting the services and functions for which resources are allocated and spent. Current and capital expenditures are further divided into three spending sessions: commitment, payment on an accrual basis, and payments with residuals. Commitment spending corresponds to the financial resources the municipality has planned to spend in the following year. Payments on an accrual basis refer to what the municipality has actually spent during the reference year, while payments with residuals consist of resources that have not been spent.

Following the approach used by Di Cataldo and Mastrorocco (2022), in our analysis we only use spending decisions data, as data on accrual and residual spending are much more fragmented, less reliable, and less homogeneous. Moreover, they often include expenses decided by previous municipal councils. By law, spending decisions cannot contain budget allocations from previous years, which allows us to capture the role of mafia infiltrations in spending decisions.

To standardize expenditure variables relative to the financial size of the municipality, the indicators used in this analysis are calculated as a proportion of the total spending, i.e., as the ratio between mission spending and total spending for all missions. Due to the strong fragmentation of mission spending and the presence of missions with expenditure values close to zero, the 23 missions have been grouped into 9 broader categories (see Table A1 in the Appendix).

Table 1. Expenditure indicators

Category	Indicator	Description
Expenditure commitments by mission	Administration, management and control	Spending decisions by mission, in current, capital, or total, as a percentage of total expenditure
	Public order and safety	
	Education and culture	
	Housing	
	Territory and environment	
	Transport	
	Social sector	
	Economic development and productive activities	
Other expenditure-related indicators	Spending inefficiency	Percentage of operating expenditure to current expenditure
	Rigidity of current expenditure	(Personnel expenses + Passive interest expenses) / (Tax revenue + transfers + Extra-tax revenue) * 100
	Current expenditure per capita	Current expenditure / Population
	Capital expenditure per capita	Capital expenditure / Population
	Total expenditure per capita	(Current expenditure + Capital expenditure) / Population

In addition to these mission-expenditure indicators, two additional indicators are included to measure the efficiency of operating expenses and the degree of rigidity of current spending, as well as three population-based indicators to account for the different heterogeneity among municipalities in terms of spending capacity: current spending per capita, capital spending per capita, and total spending per capita. Table 1 reports the list of all expenditure indicators used for the analysis.

The harmonized final certificates of Italian municipalities also contain data related to the revenues of local authorities. In particular, two budget items are crucial: (i) the volume of collections, which are the revenues actually collected during the fiscal year, and (ii) revenues expected at the beginning of the year. The ratio between collections and expected revenues becomes a critical indicator for evaluating the capacity of a local authority to collect taxes. A higher ratio suggests more efficient tax collection. To assess this efficiency, we calculate four distinct indicators: two that focus specifically on property and waste tax revenues, and two broader ones that encompass total tax revenues and total general revenues. These indicators, being the only ones that utilize information about expected revenues in comparison with collections, serve as the sole measures used to offer a detailed assessment of a municipality's effectiveness in collecting anticipated revenues.

In addition to these revenue indicators, three indicators measuring the degree of financial and tax autonomy of the municipality and the per capita tax burden are added. In particular, the indicator of tax autonomy measures the proportion of the municipality's total revenues (comprising tax revenues, transfers, and extra-tax revenues) that is derived from its own tax collections. A higher percentage implies greater tax autonomy, as it suggests that a larger share of the municipality's budget comes from taxes that it directly controls, as opposed to external transfers or other non-tax sources. This can reflect the municipality's ability to make independent fiscal decisions and respond to local economic conditions without as much reliance on external funding.

Table 2 reports the list of all revenue indicators used for this analysis.

Table 2. Revenue indicators

Category	Indicator	Description
Revenue indicators	Efficiency of property taxes	Ratio of collections to expected revenues
	Efficiency of waste taxes	
	Efficiency of tax revenues	
	Efficiency of total revenues	
	Financial autonomy	(Tax revenues + Extra-tax revenues) / (Tax revenues + transfers + Extra-tax revenues) * 100
	Tax autonomy	Tax revenues / (Tax revenues + transfers + Extra-tax revenues) * 100
	Tax burden per capita	Tax revenues / Population
	Total revenues per capita	Total revenues / Population

In addition to harmonizing the accounting systems and budget models of local authorities, Legislative Decree 118/2011 requires local authorities to adopt a system of indicators, called the "Plan of Indicators and Expected Budget Results", a system of "normalized" statistical indicators built according to common criteria and methodologies (Plan indicators, from now on). The examination of the Plan Indicators allows for the consistent analysis of various aspects of municipal budgets and the identification of budget issues, differentiating structural ones from contingent ones, as well as highlighting efficiencies in various stages of forecasting, adjustment, and final account. The Plan consists of a total of 55 synthetic indicators grouped into 14 categories. The ability to consult the Plan Indicators, both in their temporal evolution and by comparing them with those of other municipalities, is particularly relevant, especially for municipalities in financial difficulty, such as those subject to dissolution due to mafia infiltration. Just like with the harmonization of budgets, the decree regarding the Plan of Indicators was implemented for the first time to the data of the 2016 fiscal year.

For the purposes of this analysis, 48 out of the 55 Plan indicators have been used; the remaining 7 indicators were excluded because they have too much missing data. Table A2 in the Appendix provides a concise description of all the Plan indicators used in the analysis grouped in 14 categories.

3.2 The list of infiltrated municipalities

The data from the municipal budgets is supplemented with information regarding the years of mafia infiltration for dissolved municipalities. The years of infiltration correspond to the period during which the council deemed collusive was in office, meaning from the year when the municipal council took office following the elections until its dissolution due to mafia infiltration. The dates of the Decrees of the President of the Republic (DPR) that decree the dissolution of municipalities are sourced from the Ministry of the Interior, as well as the Historical Archive of local elections, which contains data on the date of local elections.

By convention, the year in which the dissolution of the municipality occurs is considered as a year of infiltration only if the date of the DPR is after June, that is, if for most of the year, the municipal council was still in office in that year. Similarly, the year in which a new infiltrated municipal council takes office is considered a year of infiltration if the municipal elections took place before June. [4] Since dissolved municipalities are very limited in number compared to the total number of Italian municipalities, this approach is conservative as it allows for the retention of the maximum possible number of years of infiltration. Data used in our analysis are updated as of October 2022.

Figure 2 – Infiltrated municipalities, 2016-2021

Municipalities dissolved due to mafia infiltration at least once *Percentage of dissolved municipalities at province level*

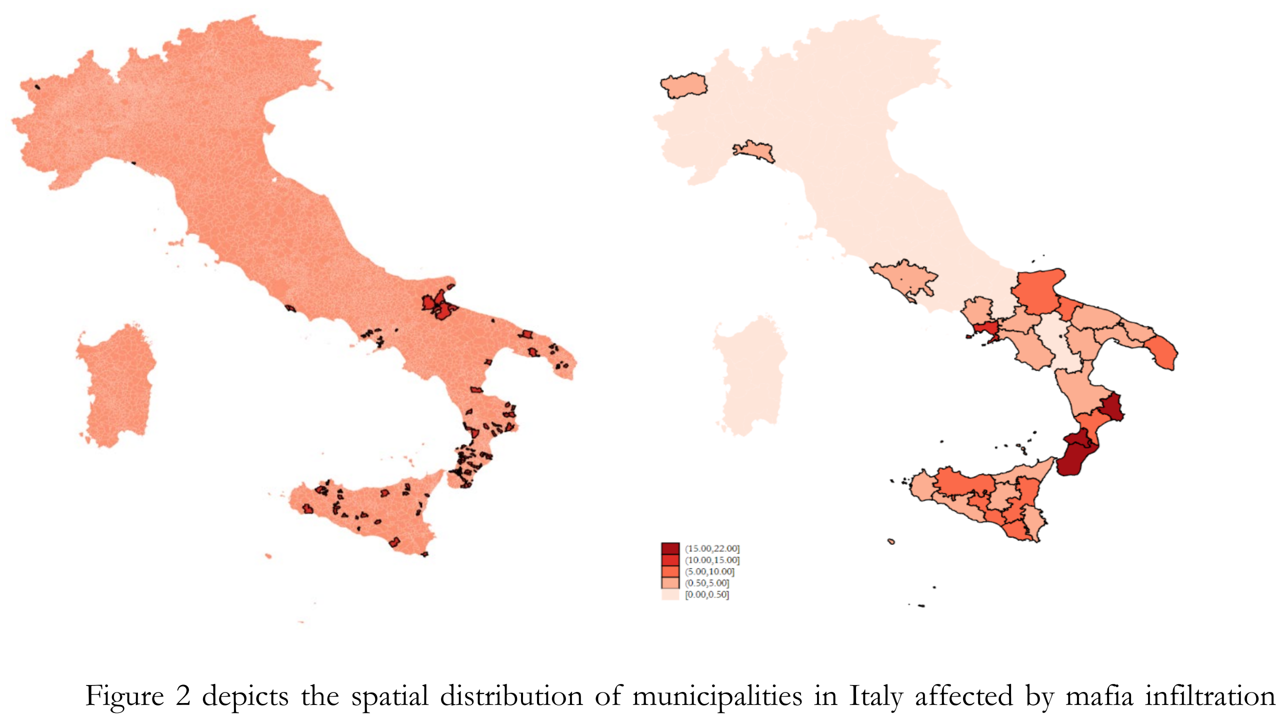


Figure 2 depicts the spatial distribution of municipalities in Italy affected by mafia infiltration from 2016 to 2022. The analysis reveals that the highest proportion of infiltrated municipalities,

[4] When identifying the years in which a municipality is identified as infiltrated, consideration is also given to any dissolutions and resignations of the bodies that occurred just before the publication of the DPR.

approximately 20 percent, is concentrated in the provinces of Reggio Calabria and Crotone. Substantial levels of infiltration are also observed in the provinces of Naples and Vibo Valentia, with additional noteworthy percentages identified in the regions of Sicily and Apulia. In Northern Italy, it is notable that only two provinces, in the regions of Piedmont and Liguria, have reported the presence of infiltrated municipalities, with each province recording just one dissolved municipality.

3.3 Socio-economic control variables

In addition to budget data, we use data at the municipal level concerning population sourced from Italy's National Institute of Statistics (Istat) and the total taxable income of individuals from the Italian Revenue Agency (Agenzia delle Entrate). The population data is used to account for the different sizes of Italian municipalities and to compute per capita indicators, while per capita income data controls for the varying levels of economic well-being among municipalities. Other control variables at the municipal level include the unemployment rate, the percentage of individuals employed in the industrial sector,[5] the percentage of the adult population with tertiary education level, and the number of enterprises, all sourced from Istat. We include two proxy variables to measure the level of crime and illegal economic activity within the territory: the number of crimes committed per 100,000 inhabitants at the province level and a proxy for the shadow economy, defined as the percentage of enterprises in the wholesale, retail, building, hotel, and restaurant sectors, both sourced by Istat. Additionally, we consider the overall level of financial activity at the provincial level as sourced by SARA reports collected by Italy's Financial Intelligence Unit,[6] along with the percentage of cash transactions relative to the total value of financial operations also computed on SARA data.

It is worth noticing that the fact that socio-economic characteristics can be useful for prediction purposes does not imply that they represent determinants of mafia infiltration.

4. Control sample

The substantial imbalance of dissolved municipalities towards the Southern provinces of Italy poses a significant challenge to the robustness and generalizability of our analysis, given that those provinces are also characterized by the lowest level of economic development in Italy. The strong association between economic disparities and the prevalence of mafia infiltration in local administrations complicates the disentanglement of the impact of municipal budgets in predicting infiltration from the influence of socio-economic factors.

To address these challenges, we adopt a dual strategy. Specifically, we compare dissolved municipalities with a carefully selected control sample of never-dissolved municipalities. This control sample is constructed to maximize their statistical similarity with the dissolved municipalities in terms of socio-economic variables. Additionally, we incorporate the same set of socio-economic variables as supplementary features in our regressions and machine-learning models in order to account for the potential residual effects of these variables on the likelihood of infiltration. This multifaceted approach enhances the validity and reliability of our findings, allowing us to draw more accurate conclusions regarding the factors influencing mafia infiltration.

Another critical issue concerning the robustness of our analysis pertains to the so-called "labelling process". While we define the infiltration label using information about the dissolution of municipalities

---

[5] Data on unemployment and occupation are not available for all years of analysis; for these two variables, the averages over the available years are employed.

[6] The Italian anti-money laundering law (Legislative Decree no. 231/2007) mandates banks and other financial intermediaries to report on a monthly basis to the UIF all transactions amounting to €15,000 or more (€5,000 since 2021), after aggregating them according to several criteria, regardless of any money laundering suspect. The reports refer to the operations carried out by the customers of the obliged intermediaries and, as the data are aggregated, they are reported in an anonymous format.

due to mafia infiltration, thus focusing on confirmed cases of infiltration, there is still a possibility that some local administrations labelled as non-infiltrated may have connections to criminal organizations. This introduces the potential for bias in our analysis. To address this concern, we construct our control sample to include only local administrations from provinces with a supposedly low level of infiltration in the economy, based on external sources. This strategic choice significantly reduces the risk of false negatives, specifically cases of undetected infiltration among non-dissolved municipalities.

More specifically, the control sample selection is based on two concurrent criteria.

Statistical Similarity. Non-dissolved municipalities are selected to closely resemble dissolved ones in terms of socio-economic variables, minimizing the influence of socio-economic factors. We use a statistical matching technique, pairing dissolved municipalities with non-dissolved counterparts based on the similarity of socio-economic control variables (computed as average on the period).[7] The similarity between these observations is quantified using the Euclidean distance, computed from the standardized variables' average values across the years. Every dissolved municipality is subsequently paired with the first 100 non-dissolved municipalities that exhibit the lowest Euclidean distance. Each dissolved municipality may be matched with multiple non-dissolved municipalities, and non-dissolved municipalities can be matched with different dissolved municipalities. Robust analysis evaluating the sensitivity of the results to the choice of the number of matched municipalities is presented in Section 6.3.

Low Infiltration Risk. To mitigate the risk of false negatives, non-dissolved municipalities are exclusively drawn from provinces with an assumed incidence of infiltration within the business sector lower than the median computed on all provinces. The incidence of infiltration in the business sector is quantified at a province level using an experimental mapping of companies potentially connected to organized crime (OC). This mapping has been created by Italy's Financial Intelligence Unit (UIF, 2021, pages 47-48).[8] The use of these unique and confidential UIF data represents a significant methodological contribution, as it allows for partial control of the mislabeled data phenomenon that affects all studies in the field of automatic classification and can significantly impair model performance and reliability, introducing bias into the results. Table 3 reports the main descriptive statistics of the incidence, measured as the proportion of firms with potential ties to organized crime as the ratio of the total number of registered firms in the province. The incidence of firms potentially connected to OC is on average equal to 1.8 percent, reaching its minimum in Sondrio (0.4 percent) and its maximum in Reggio Calabria (6.3 percent); there is a remarkable North-South disparity.

Table 3. Incidence of firms potentially linked to OC across the provinces
(percentage of total firms)

Min.	P5	Mean	Median	P95	Max.
0.43	0.58	1.83	1.54	4.28	6.28

Table 4 displays the annual count of municipalities within both the infiltrated group and the control sample. Notably, the infiltrated municipalities exhibit a substantial decline over the years of

[7] Since dissolved municipalities never have populations exceeding approximately 200,000 inhabitants, the control sample, matched on population size and other socio-economic variables, includes only non-dissolved municipalities within the same population range. Consequently, all subsequent analyses exclude the 14 larger municipalities.

[8] The firms included in the mapping have been selected as those whose directors and other corporate officers include i) persons of interest on the basis of information exchanges with the DNA, ii) persons investigated for mafia crimes who appear in business archives, or iii) persons named in information requests from the judicial authorities regarding organized crime.

analysis, with about 50 occurrences in 2016-2017, progressively falling to 15 units by the year 2021. The ratio of infiltrated municipalities and control sample is about 1 over 29.

Table 4 – Number of municipalities across the years

	Dissolved municipalities	Control sample
2016	52	1,086
2017	50	1,102
2018	42	1,090
2019	38	1,119
2020	24	1,106
2021	15	1,068
Total	221	6,571

Table 5 displays the mean values of socio-economic variables across the entire set of non-dissolved municipalities, the control sample (drawn from this set), and the infiltrated municipalities. The results indicate that the disparities between non-dissolved and dissolved municipalities have been significantly mitigated due to the careful selection process applied to the non-dissolved municipalities included in the control sample. However, some differences persist, particularly in terms of population size and per capita income. As explained before, we incorporate all socio-economic variables as controls in the models in order to account for the potential residual role of these variables on the likelihood of infiltration.

Table 5. Main statistics for the socio-economic control variables

(average values over 2016-2021)

	Dissolved municipalities	Non-dissolved municipalities	Control sample of non-dissolved municipalities
Population (unit)	17,417	7,580	7,686
Per capita taxable income (euro)	8,201	12,407	11,216
Unemployment rate (%)	22.6	12.8	15.6
Employees in the manufacturing sector (%)	14.8	24.1	17.4
Number of enterprises	1,036	606	628
Crime rate (x100.000 inhabitants) at provincial level (*)	3,404	3,427	3,102
Proxy of shadow economy (**)	56.4	51.8	53.8
Total SARA reports (millions)	854	2,620	1,079
Cash as percentage of total SARA report	9.3	4.9	6.8

Note: The sample of dissolved municipalities includes 93 observations of municipalities. The control sample includes 1,144 observations. All never-dissolved municipalities are 7,385.
(*) For Sardinia provinces regional rate is used.
(**) Percentage of enterprises in wholesale, retail, building, hotel and restaurant sector.

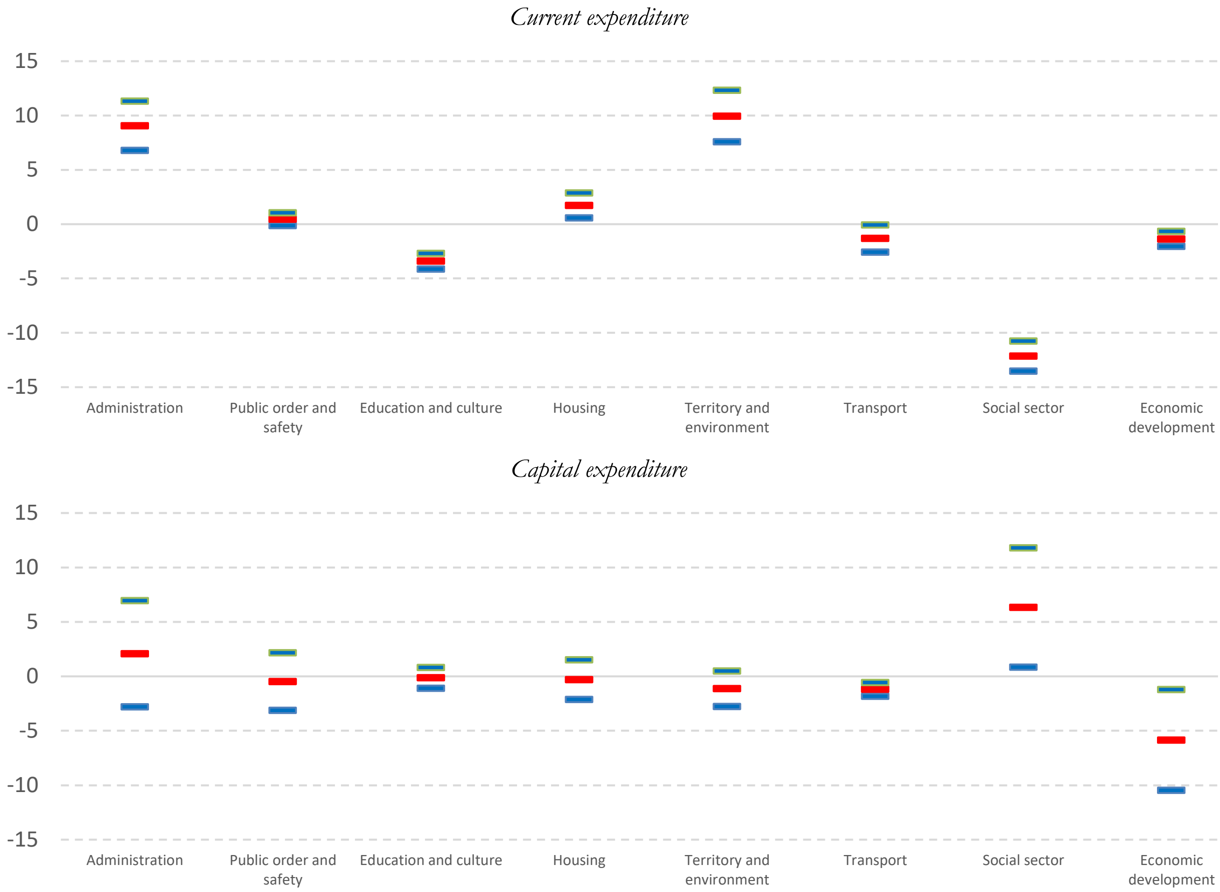
5. Statistical analysis of financial indicators

The objective of this section is to compare the sample of infiltrated municipalities with the control sample using all the financial indicators defined in section 3.1. This will be done through t-tests on mean differences and appropriate regression models.

Tables A2 and A3 in the Appendix display mean and standard deviation of the expenditure and revenue indicators and the Plan indicators, together with a t-test on the mean differences between the infiltrated municipalities during infiltration years and the control sample. The tables also show the results of an inferential exercise with the aim to evaluate the statistical significance of the mean differences between the two groups of municipalities after controlling for the socio-economic characteristics. In particular, a generalized linear model is estimated separately for each indicator, where the response variable is the indicator it-self and the regressors are a dummy variable indicating the infiltration status of the municipality and the socio-economic control variables. In case of indicators bounded in the range [0,1] a logit link function is used, while for the other variables we employ a robust linear regression to deal with potential outliers. The sign and significance of the regression coefficient of the infiltration status, which is displayed in the last column of Tables A2 and A3, is our very last objective.

Results show that the share of expenditure per mission is particularly effective in discriminating between infiltrated and "lawful" municipalities, also after controlling for socio-economic differences. Specifically, infiltrated municipalities have relatively lower current expenses on education and culture, transportation,[9] social sector, while they spend more on administration, management and control, housing, territory and environment (Table A2). Specifically, after controlling for socio-economic differences, infiltrated local administrations allocate 3% and 11% more current resources to, respectively, housing and 'territory and environment' compared to their "lawful" counterparts. Conversely, infiltrated administrations allocate 13% and 2% less to, respectively, the social sector and economic development (Figure 2).

Figure 2. Marginal effects of infiltration status on expenditure share by mission
(average value and confidence interval at 95%; percentage points)



These results align with those found by Di Cataldo and Mastrorocco (2022). Criminal organizations strategically invest in the construction and waste management sectors due to their profitability and low technological and financial barriers to entry, making these sectors ideal for long-

[9] Investments in this chapter of the balance sheet include expenditures for local public transportation (e.g., school buses), public lighting, as well as any improvements in the management of road traffic.

term investment. The mafia seeks to control these sectors through activities deeply rooted in local communities, aiming to expand its influence and protect their markets from competition to increase rents. This involves employing a range of tactics, such as intimidating potential competitors, manipulating public work bids through political connections, and corrupting officials to establish monopolistic control over critical resources and services, thereby benefiting affiliated enterprises. By embedding itself across all levels of these sectors, organized crime strengthens its network and ensures its prosperity (De Feo and De Luca, 2017). Di Cataldo and Mastrorocco (2022), examining municipal public procurement data from 2000 to 2012, also show that infiltrated local governments allocate more resources to construction and waste management tenders. This allows criminal organizations to direct resources to strategic sectors, offer business opportunities to controlled firms, and reinforce territorial control.

Similarly, when examining other expenditure indicators, a notable distinction emerges between infiltrated municipalities and the control sample, with the former exhibiting higher operating expenses with respect to their "lawful" counterparts. Criminal organizations often engage in corrupt practices that inflate operating expenses. For example, contracts may be awarded to firms associated with the organization at inflated prices, or public funds may be siphoned off through fraudulent schemes. This leads to higher operating expenses as funds are misallocated or wasted.

Infiltrated municipalities also exhibit higher spending rigidity with respect to non-dissolved ones: infiltration in the administration may result in inflated personnel expenses through the hiring of unnecessary staff, friends, or associates (nepotism or clientelism). This creates rigidity because these expenses are largely fixed and difficult to reduce; corruption can lead to long-term contracts that are difficult to renegotiate or cancel locking the administration into high personnel or service costs, further increasing spending rigidity; moreover, corrupt administrators may take on unnecessary or inflated debt, often through shady deals or at non-competitive interest rates, leading to increase incidence of passive interests. More generally, infiltrated municipalities often suffer from weakened governance and oversight mechanisms, making it easier for criminal organizations to manipulate the budget; this lack of accountability can lead to persistent inefficiencies and inflated costs, contributing to both higher operating expenses and spending rigidity.

Regarding revenue indicators, significantly lower revenue efficiency is recorded for infiltrated municipalities across all segments (property tax[10], waste tax, total tax and total revenues). These results generally align with those of Di Cataldo and Mastrorocco (2022), who show that municipalities infiltrated by criminal organizations are characterized by a lower efficacy in collecting fiscal revenues, a result that seems to be mostly driven by a reduction in the ability in collecting waste taxes. The observed connection between waste tax collection and mafia infiltrations is widely explored in the literature (D'Amato et al., 2015; Daniele and Dipoppa, 2017; Di Cataldo and Mastrorocco, 2022) and can be explained by two factors. One possibility is a reduced administrative and revenue-collection capability of local governments during infiltration periods or the deliberate action to under-collect taxes from politically connected individuals and businesses. This hypothesis is supported by a significant reduction also in total tax collection. The alternative explanation suggests that citizens may be less willing to pay waste taxes when managed by local administrations linked to criminal groups, especially if waste management is contracted to mafia-controlled companies. Despite increased spending on waste management by infiltrated governments, the service quality may decrease due to higher instances of environmental crimes, leading citizens to resist payment.

Financial and tax autonomy exhibit a noteworthy increase in infiltrated municipalities, which can be attributed to several factors. Corrupt administrations might use extra-tax revenues to boost their financial autonomy in order to compensate for the poor efficacy in tax collection. Moreover, these extra-tax revenues, such as fees, fines, or sales of public assets, can be manipulated more easily and provide

[10] The difference in property tax efficiency between infiltrated municipalities and the control group is not significant after controlling for socio-economic differences.

quicker returns. Ultimately, this approach allows them to reduce their reliance on government transfers without needing to efficiently collect taxes. Moreover, greater financial autonomy gives these municipalities more discretion in collecting and allocating resources, potentially reducing oversight. Criminal organizations can exploit this autonomy to channel funds toward projects or contracts that benefit them, while minimizing scrutiny from higher levels of government.

We also observe a higher per capita tax burden in infiltrated local administrations. This could seem in contrast with the evidence of a low tax collection efficiency, but actually it can just be explained by the need to compensate for it or for reduced revenue from other sources. In fact, infiltrated municipalities may raise taxes for individuals and businesses that are not connected and thus less able to evade them. Additionally, organized crime infiltration can lead to economic stagnation or decline, reducing the overall tax base. With fewer businesses and individuals contributing to the tax system, it becomes necessary to raise per capita tax rates to maintain the funds needed to support the network of connected individuals and businesses.

Finally, we also observe a significant disparity in per capita expenditures and revenues between infiltrated municipalities and their "lawful" counterparts. Infiltrated local administrations tend to have lower per capita spending and revenue levels. The presence of organized crime discourages legitimate businesses and deters new investments. As a result, the overall economic decline in these areas reduces the resources available for public spending, leading to lower per capita expenditures and revenues.

Regarding the Plan indicators, as evidenced by the significance of differences between infiltrated municipalities and the control sample reported in Table A3, several important results emerge that confirm previous findings. Specifically, infiltrated municipalities exhibit higher levels of incidence of passive interest on current revenues (IND6_1 to IND6_3), lower levels of total per-capita investments (IND7_1 and IND7_2), a lower incidence of passive residuals (IND8_1 to IND8_3), and a lower index of non-financial debt clearance (IND9_1 to IND9_5; indicators of the administration's efficiency in paying off non-commercial debts). Additionally, dissolved municipalities show a lower incidence of the free allocation in capital accounts within the budget surplus (IND11_2),[11] which implies a reduced capacity to autonomously finance investment projects, and a lower incidence of the restricted allocation within the budget surplus (IND11_4),[12] potentially because long-term projects are not generally undertaken. Overall, infiltrated municipalities may show significant differences in most of the indicators included in the Plan of indicators, compared to the control sample, due to a general compromise in their capacity to plan and manage resources, further aggravated by instability, corruption, and administrative inefficiencies. Note that, for 9 of the indicators in the Plan of indicators, it was not possible to test whether the difference between the means of infiltrated municipalities and the control sample is significantly different from zero due to an excessive presence of zero values. Nevertheless, all the indicators presented here are used in the machine learning model to compute a risk score of municipal infiltration, as we believe that each indicator, to varying degrees, can contribute to the model's predictive capability.

6. A machine learning approach

6.1 Classification methodology

The objective of this section is to describe the implementation of a machine learning approach that, based on all financial indicators and socio-economic control variables, produces a risk score of infiltration for each Italian municipality. By comparing the risk score with an appropriate threshold value,

[11] The incidence of the free allocation in capital accounts within the budget surplus measures the proportion of the budget surplus available for discretionary use by the administration. It is a key indicator providing important insights into the financial health of a municipality and its capacity to independently finance investment projects.

[12] The incidence of the restricted allocation within the budget surplus provides insights into the specific use of the municipality's resources and their availability. It is a crucial indicator to understand how much of the financial surplus is already committed to binding obligations and how much is freely available for other spending or investment choices.

it is possible to classify municipalities as infiltrated or not. To this end, the sample of infiltrated municipalities observed during the years of infiltration is compared with the control sample, which includes only municipalities that have never been dissolved (see Section 4).

As a first crucial step, dividing the entire sample into training and test sets is essential to develop a machine learning model that is not only accurate, but also generalizes well to new data. To this end, using a very common criterion in classification tasks, 80% of the units are allocated to the training set, while the remaining 20% are allocated to the test set. The partitioning of the dataset into training and test sets is applied using a proportional stratified cluster sampling strategy, where the stratum is based on the infiltration status (infiltrated and non-infiltrated) to ensure that an equal percentage of infiltrated municipalities is maintained in each of the two training/test subsets. Moreover, since data are available for each municipality over multiple years, stratified sampling is combined with cluster sampling. This means that instead of directly selecting individual observations, all observations from the same municipality (cluster) are selected. This approach ensures the exclusive inclusion of each municipality in either the training or testing set, preventing overlaps and thus cross-contamination between training and test sets.

A significant challenge arises from the substantial class imbalance in our dataset, with dissolved municipalities constituting only 1 for every 29 non-dissolved municipalities in the control sample. This imbalance poses issues in terms of reliability of the model estimates and predictions derived from the model. To address this, we employ a replication-based over-sampling strategy on dissolved municipalities: since the number of dissolved municipalities is very limited, in order to avoid certain municipalities being overrepresented (i.e., sampled too many times) with simple random over-sampling, each dissolved municipality is replicated a fixed number of times to achieve perfect balance with the control sample.

To evaluate the model's performance, we use the Area Under the Curve (AUC) metric. AUC is a measure that evaluates a model's ability to distinguish between classes, independent of the cut-off point. The cut-off point is often determined by cost factors associated with false positives and false negatives, which are typically unknown in most real-world applications. AUC is a common metric for classification problems, as it allows for the comparison of a model's performance against a random guess (AUC = 0.5). It involves plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings, where specificity is the true negative rate. The area under this curve, known as the Receiver Operating Characteristic (ROC) curve, indicates the model's ability to distinguish between classes.

In order to assess the performance of our model and select the hyperparameters, a stratified k-fold cross-validation is applied by dividing the training set into 5 equal folds and thereafter evaluating the performance of the model on each fold. Specifically, cross-validation carries out the following ordered steps:

1. Divide the training set into 5 groups without replacement by applying stratified sampling with clustering where infiltration status is the stratum, in order to capture the imbalanced class distribution of the target feature in the data.
2. Take the first group as a validation set and the remaining 4 groups as a whole training set.
3. Only the training set is oversampled using a replication-based over-sampling technique; the validation set is not oversampled, thus it maintains the imbalanced class distribution of the target feature as in the main data.
4. Perform model performance evaluation and score model performance according to the AUC.
5. Repeat from step 2 for every other group (second group, third group, …, fifth group).
6. Take the mean and variance of the distribution of the five scores to estimate the overall performance of the model.

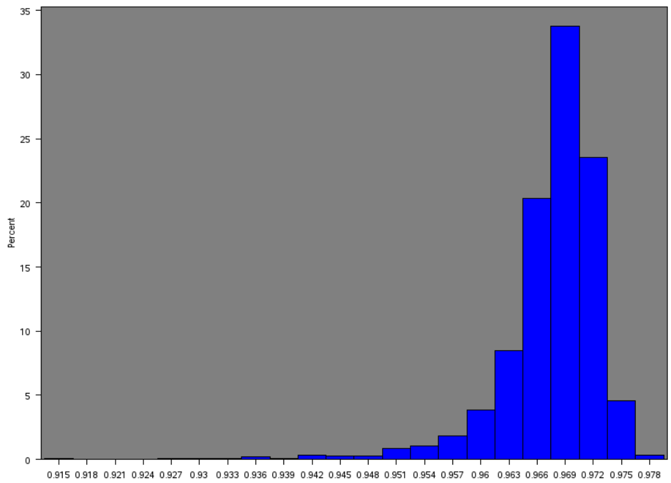
Table 6. Selection ranges of hyperparameters

	Min	Max	Optimal values
Learning rate	0.01	0.30	0.24
Trees	10	330	209
Maximum depth	1	10	2
Sampling rate of rows	0.5	1	0.87
Sampling rate of columns	0.5	1	0.58
Gamma coefficient	0	0.5	0.15
Minimum number of observations required to create a new node	1	8	4

To find the optimal combination of hyperparameters that maximize the AUC we use a sample of 2,500 sets of hyperparameters randomly selected in the ranges shown in Table 6.

The results of the k-fold cross-validation are summarized in Figure 2. The AUC mean distribution across the parameter sets exhibits minimal variation, ranging from 91.47% to 97.75%. This distribution is heavily skewed towards higher values, underscoring the stability and robustness of the results across different parameter combinations.

Figure 2 – AUC mean distribution for different hyperparameter sets



The optimal model, characterized by the highest AUC, is presented in Table 7, showcasing its average performance and the consistency observed across different cross-validation folds. Specifically, the mean AUC attains an impressive 97.75%, with a minimal standard deviation of merely 1.96%. Although our model has been calibrated according to the performance metric AUC, which is independent of the choice of the cut-off point, examining additional performance metrics computed at a standard cut-off point of 0.5 provides further insights. Specifically, in our case, these metrics reveal highly commendable outcomes, indicating a sensitivity of 76.3%, a specificity of 99.3%, and a precision of 79.2%.

Table 7. Performance metrics on validation sets
(percentage values)

	AUC	Sensitivity	Specificity	Precision
Mean	97.75	76.30	99.28	79.22
Coefficient of variation	1.96	9.47	0.36	6.72

Note: Sensitivity, specificity and precision are computed by using a cut-off of 0.5.

6.2 *Model evaluation*

We evaluate the performance of the best model on the test set, which exhibits a dissolved/non-dissolved imbalance ratio of about 1/29. Our achieved performance generally aligns with that obtained on the training set with respect to AUC, reinforcing the confidence that the models are not overly tailored to the training data, thereby mitigating concerns of overfitting (Table 8).[13]

Table 8. Performance metrics – training vs test
(percentage values)

	AUC	Sensitivity	Specificity	Precision	Precision under perfect balance (1)	Precision under original imbalance (2)
Training set (mean)	97.75	76.30	99.28	79.22	-	-
Test set	98.22	60.00	99.62	84.38	99.37	44.16

Note: Sensitivity, specificity and precision are computed by using a cut-off of 0.5.
(1) Precision estimated under the hypothesis of a perfect balance, i.e. 1/1 between positive and negative cases.
(2) Precision estimated under the hypothesis of an imbalance of 1/200 between positive and negative cases.

Table 9. Performance metrics on test set at different cut-offs
(percentage values)

Cut-off	Sensitivity	Specificity	Precision	Precision under perfect balance (1)	Precision under original imbalance (2)
0.05	82.22	97.50	52.86	97.04	14.10
0.10	80.00	98.25	61.02	97.87	18.65
0.25	68.89	99.17	73.81	98.80	29.21
0.50	60.00	99.62	84.38	99.37	44.16
0.75	57.78	99.85	92.86	99.74	65.56
0.90	55.56	99.92	96.15	99.86	78.55
0.95	53.33	100.00	100.00	100.00	100.00

(1) Precision estimated under the hypothesis of a perfect balance, i.e. 1/1 between positive and negative cases.
(2) Precision estimated under the hypothesis of an imbalance of 1/200 between positive and negative cases.

These results show that, using a standard cut-off point of 0.5, the model can identify 60% of the dissolved municipalities (sensitivity) and over 99% of the non-dissolved municipalities (specificity).

[13] We also performed a robustness check by randomly selecting 10 samples from both the training and test sets and applying the optimal hyperparameters identified through cross-validation. The results indicate a non-negligible variability in the sensitivity metric across the samples and the cut-off points, which can be attributed to the limited sample size.

Lowering the cut-off to 0.05 results in an increase of sensitivity to 82% at the small expense of a reduction in the ability to identify non-dissolved municipalities to 97% (Table 9).

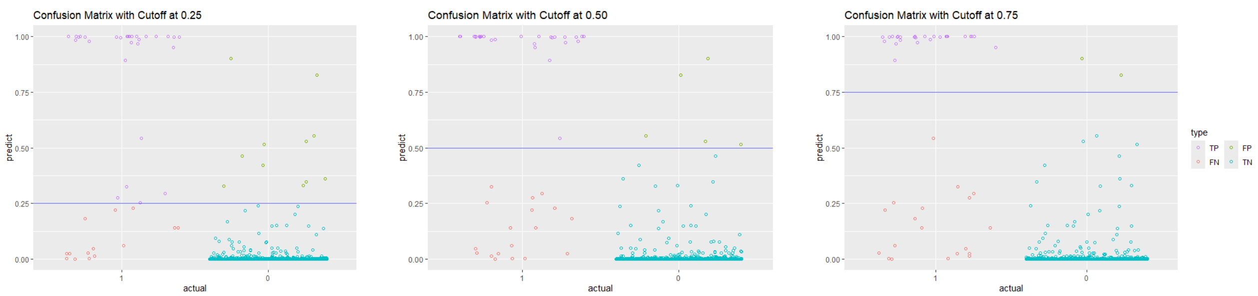
Regarding precision, which measures the percentage of municipalities identified as infiltrated that are actually infiltrated, the model performs at 84% in the test set with a standard cut-off point of 0.5. However, we must acknowledge that the precision rate is strongly affected by the degree of imbalance between positive and negative cases. The 84% precision is obtained on a test set with an imbalance of 1/29. As shown in the last two columns of Table 9, under the hypothesis of perfect balance between positive and negative cases, the precision would drastically rise to 99.37%, whereas if the imbalance were 1/200, which is the case in the entire population of municipalities, the precision would instead drop to 44.16% which is still a very good result for a highly imbalanced sample.

Compared to other similar studies applying machine learning models to identify criminal contexts, our model demonstrates good performance. For example, de Blasio et al. (2022), focusing on corruption in Italian municipalities, attained a maximum sensitivity and specificity of 74% and 86.7%, respectively. Cariello et al. (2024) adopt a machine learning model to identify firms at risk of infiltration, achieving a sensitivity of 76% and a specificity of 74%.

Campedelli et al. (2024), who developed a very similar model to ours, using only spending-side budget indicators to predict infiltration in Italian municipalities, achieved a very good sensitivity of 96.4% at the expense of lower specificity of 87.8% and, most notably, a very low precision of 6.2%. When using a cut-off point of 0.05, we can achieve higher specificity (97.5%) and higher precision (14.1%) but lower sensitivity (82.2%). Therefore, while their model is able to identify a greater number of infiltrated municipalities, it does so at the cost of a higher number of false positives and lower precision. Note, however, that while Campedelli et al. (2024) calibrate their model to optimize sensitivity at a 0.5 cut-off point, we calibrate our model to optimize the AUC, which is a performance metric measured across the entire spectrum of cut-off values.

Figure 3 presents the confusion matrix at cut-off values of 0.25, 0.50, and 0.75, illustrating a high concentration of points around the extreme predicted risk scores of 0 and 1. Specifically, raising the cut-off from 0.5 to 0.75 increases true negatives (TN) by two units and decreases true positives (TP) by one only. Lowering the cut-off to 0.25 increases false positives (FP) by five units and decreases false negatives (FN) by two. The minimal classification changes within the 0.25-0.75 range suggest a stable model performance, indicating that a cut-off value within this range could be optimal for operational purposes.

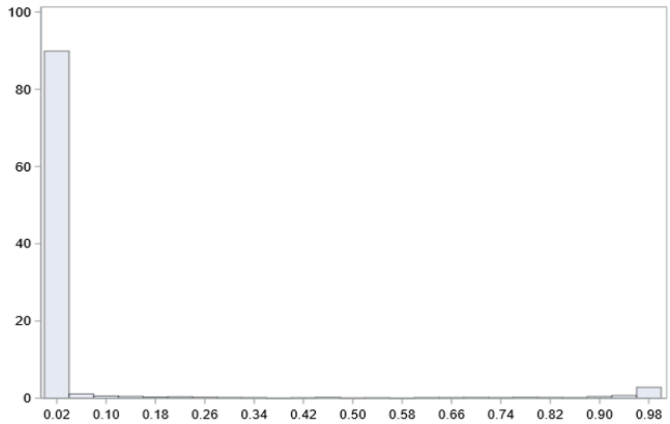
Figure 3. Confusion matrix on test set at different cut-offs



To identify the primary factors influencing the model's ability to predict infiltration risk, we employ the average gain criterion (Figure A1 in the Appendix). The findings highlight that the predictors with the highest level of importance include tax autonomy, efficiency in waste taxes, the share of current expenditures for territory and environment, and the incidence of new current passive residuals.

Figure 4. Infiltration risk score distribution

(all municipalities; year 2021)



Note: The risk score is missing for municipalities where financial data are not available in 2021.

Table 10. Infiltration risk score across regions

(Risk score multiplied by 100; All municipalities; Year 2021)

	N	Mean	Median	75th perc.	90th perc.	95th perc.	99th perc.	Max
North-west								
Piedmont	1,124	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Aosta Valley	74	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Liguria	224	0.0	0.0	0.0	0.0	0.0	1.0	5.0
Lombardy	1,400	0.0	0.0	0.0	0.0	0.0	0.0	4.0
North-east								
Emilia-Romagna	305	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Friuli-Venezia Giulia	185	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Trentino-South Tyrol	89	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Veneto	366	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Centre								
Lazio	351	4.0	0.0	0.0	6.0	27.0	95.0	100.0
Marche	215	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tuscany	260	0.0	0.0	0.0	0.0	0.0	0.0	1.0
Umbria	92	0.0	0.0	0.0	0.0	0.0	1.0	1.0
South and Islands								
Abruzzo	270	0.0	0.0	0.0	1.0	2.0	8.0	24.0
Molise	127	1.0	0.0	0.0	0.0	0.0	8.0	41.0
Campania	511	29.0	3.0	67.0	99.0	100.0	100.0	100.0
Apulia	248	13.0	0.0	6.0	65.0	96.0	100.0	100.0
Basilicata	125	0.0	0.0	0.0	0.0	1.0	5.0	9.0
Calabria	337	46.0	33.0	95.0	99.0	100.0	100.0	100.0
Sicily	130	45.0	28.0	93.0	100.0	100.0	100.0	100.0
Sardinia	338	0.0	0.0	0.0	0.0	0.0	0.0	0.0
All municipalities	**6,771**	**6.1**	**0.0**	**0.0**	**4.6**	**70.2**	**99.7**	**100.0**

Note: In order to enhance the reading of the table, in this table the risk scores – which range from 0 to 1 – are multiplied by 100. The risk score is missing for municipalities where financial data are not available in 2021.

The machine learning model we trained generates an infiltration risk score – ranging from 0 to 1 – for each Italian municipality for which financial information is available. This risk score can be interpreted as an indicative measure of the probability of infiltration. Figure 4 shows the entire distribution of the risk scores for all Italian municipalities for the year 2021, highlighting a thick right tail that represents the municipalities identified as highly at risk of infiltration, confirming the model's high capacity to discriminate between positive and negative cases. Table 10 – which reports risk scores expressed as percentages – breaks down this distribution across the regions: municipalities in the South and Islands exhibit higher predicted infiltration risk levels, with municipalities in Sicily and Calabria showing an average infiltration risk of over 45%, and over 93% for the top 25% of municipalities. The central regions display significantly lower risk levels compared to the South; however, the region of Lazio stands out with an average risk of 4% and values exceeding 27% for the highest fifth percentile. Northern municipalities have an essentially negligible risk, except for a few localized cases in Liguria and Lombardy.

6.3 Robustness check on control sample selection

The model's performance results may be influenced by the method of selecting the control sample. To assess the robustness of the results in this regard, we constructed three alternative control samples:

(A) The first alternative control sample is created using only one of the two criteria employed for selecting the non-dissolved municipalities for the main control sample, that is, the incidence of companies with ties to organized crime in their respective provinces is lower than then the median.

(B) The second alternative control sample was selected using both criteria employed in the main control sample, but the number of non-dissolved municipalities matched to infiltrated municipalities according to socio-economic variables is set to 10 units, instead of 100 as in the main control sample.

(C) The third alternative control sample followed the same approach as the second but raised the number of matched non-dissolved municipalities to 40 units.

We also measure the performance of the model estimated with the control sample but omitting the socio-economic factors as control variables.

Table 11 compares the model's performance when using the main control sample against three alternative samples and when omitting the control variables. The results indicate that selecting the control sample based solely on the incidence of companies with potential ties to organized crime results in a 2.4 percentage point decrease in sensitivity, while AUC and specificity slightly increase. Reducing the number of matched municipalities used in the similarity criterion to 10 generally leads to a minor decrease in performance metrics. When the number of matched municipalities is 40 performance metrics are generally invariant. Additionally, omitting socio-economic variables results in a performance decline of 1 to 4 percentage points according to the performance metric. Overall, we can conclude that the criteria used for selecting the control sample do not significantly impact the machine learning model's performance.

Table 11. Performance metrics on cross validation by using alternative control samples and omitting control variables

(percentage values)

	AUC	Sensitivity	Specificity
Main control sample	97.75	76.30	99.28
(A) Low infiltration risk among firms	99.53	74.76	99.68
(B) Low infiltration risk among firms + statistical similarity (10 units)	97.09	74.11	97.51
(C) Low infiltration risk among firms + statistical similarity (40 units)	97.42	77.17	97.60
Use of control sample without socio-economic variables in the machine learning	96.48	72.55	97.83

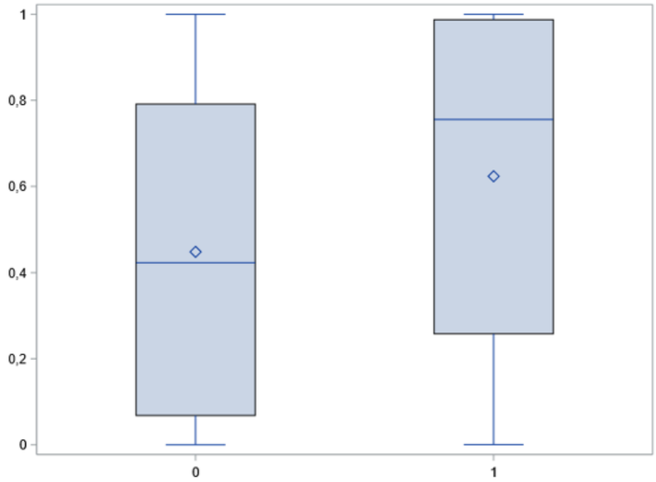
Note: Sensitivity, specificity and precision are computed by using a cut-off of 0.5.

7. Infiltration risk score: validation using external data sources

To validate the effectiveness of the model in accurately measuring the risk of infiltration in Italian municipalities, two distinct analyses are conducted. These analyses are restricted to municipalities in the four regions of Campania, Apulia, Calabria, and Sicily, which are traditionally recognized as the origins of Italian mafias. In these regions, the risk of infiltration is relatively high, making it much harder for the machine to distinguish between infiltrated municipalities and "lawful" ones. This difficulty makes the validation exercises more informative about the model's capacity.

The first analysis aims to verify whether the infiltration risk assessed by the model is statistically associated with the presence of businesses connected to organized crime within the same municipality. Under this hypothesis, in municipalities where at least one business connected to organized crime (OC) exists, the infiltration risk should be significantly higher. To test this hypothesis, data from the mapping of potentially OC-linked businesses are employed (see section 4). Figure 5 presents the statistical distribution of infiltration risk score across municipalities, depicted as a box plot. It compares municipalities with no potentially OC-linked businesses to those with at least one such business, showing a higher skew toward elevated values of infiltration risk score in the latter group.

Figure 5. Infiltration risk in local municipalities according to the presence of OC-linked firms
(0=no firm, 1=at least one firm; average risk score over 2016-2021; municipalities of Campania, Apulia, Calabria and Sicily)



Source: UIF's mapping of firms potentially linked to OC.

Additionally, Table 11 summarizes key statistics for these two groups of municipalities. A Wilcoxon test is performed, showing that the infiltration risk is significantly different between the two groups. Furthermore, a fractional logit model is estimated with the infiltration risk score as the response variable. The regressors include a dummy variable indicating the presence of at least one potentially OC-linked firm in the same municipality, along with the socio-economic control variables. The coefficient associated with the presence of OC-linked firm is positive and statistically significant, providing support for the validity of the model.

Table 11. Statistics on infiltration risk in municipalities according to the presence of OC-linked firms

(Average risk score over 2016-2021; municipalities of Campania, Apulia, Calabria and Sicily)

	No OC-linked firm	At least one OC-linked firm
N.	1.014	547
Mean	0.45	0.42
Median	0.17	0.76
Wilcoxon test p-value		<0.01
Regression coefficient* (p-value)		+0.38*** (<0.01)

Source: UIF's mapping of firms potentially linked to OC.

(*) Fractional logit model with socio-economic control variables and clusters at regional level. Control variables are computed as averages on 2016-2021.

The second exercise aims to verify whether the infiltration risk in municipalities is statistically associated with a higher opacity in the management of public procurement by municipalities. The underlying hypothesis is that the low transparency in public procurement management by infiltrated municipalities allows connected criminal organizations to manipulate tenders to favour their own businesses or those of their allies without attracting the attention of law enforcement. To conduct this analysis, we utilize the open database[14] from Italy's Central Anti-Corruption Authority (ANAC), which collects a wide set of information regarding public tenders filed by the contracting authorities. For each award, we compute five dummy variables indicating whether information regarding the start and the end of the contract, its awarding process, the winning company, and the funding sources used to finance the tender are transmitted to ANAC. Subsequently, we compute the frequency of awards issued by the authority that meet the conditions specified by each dummy variable (Gara et al., 2024). We analyse Italian public contracts with an ordinary CIG[15] published between January 2018 and June 2023 in the ANAC's open database. This data allows us to assess whether municipalities with higher infiltration risks exhibit lower levels of compliance in providing comprehensive information about public tenders.

Table 12. Infiltration risk vs transparency in public procurement process

(Average risk score over 2016-2021; municipalities of Campania, Apulia, Calabria and Sicily)

	Regression coefficient	
Lack of communication of…		
Contract start	0.975	**
Contract end	1.068	*
Awarding process	0.616	***
Winning company	0.618	***
Funding sources	0.601	***

Source: ANAC's open database.

Note: Fractional logit model with socio-economic control variables and clusters at regional level. Control variables are computed as averages on 2016-2021. ***: significance at 1%; ** significance at 5%; * significance at 10%.

[14] https://dati.anticorruzione.it/opendata

[15] Contracts with an ordinary CIG (Codice Identificativo di Gara) are those valued over 40,000 euros or for which an ordinary CIG has been specifically requested. Typically, contracts valued below 40,000 euros and those not subject to reporting obligations by the contracting authorities are assigned a so-called SmartCIG. As a result, information on this latter category of contracts is more limited, primarily covering data contained in the call for tender. However, there are exceptions to this rule. For example, an ordinary CIG is required for tenders funded by the National Recovery and Resilience Plan (PNRR) and the National Cohesion Plan (Piano Nazionale per gli Investimenti Complementari - PNC), regardless of their value.

Table 12 presents the results of the estimate of a fractional logit model where the response variable is the infiltration risk score in the municipality and the regressors include the five variables (one at a time) related to the transmission of information to ANAC, along with the socio-economic control variables. The regression coefficients corresponding to the dummy variables are consistently positive and statistically significant. These findings suggest that municipalities at higher risk of organized crime infiltration tend to exhibit lower levels of transparency in their public procurement processes, serving as an indirect validation of the machine learning model.

These two exercises do not aim to establish a causal relationship between the analysed phenomena (mafia infiltration, presence of OC-linked firms and transparency in public procurement process), but merely to measure the presence of a statistical association. This association provides indirect evidence of the model's effectiveness in appropriately measuring the infiltration risk in local administrations.

8. Conclusions

This study provides valuable insights into the risk of criminal organization infiltration in Italian municipalities by employing an extensive dataset of financial information (drawn from municipality financial statements) alongside statistical and machine learning techniques. By focusing on the policy of dissolving mafia-infiltrated city councils, we analyse and compare the economic and financial behaviours of dissolved municipalities with a carefully constructed control sample of non-dissolved municipalities. This control sample is designed to minimize the influence of socio-economic factors that could explain the financial differences between the two groups. Furthermore, by using confidential data from Italy's Financial Intelligence Unit, the control sample is limited to provinces with a low presence of organized crime in the legal economy, thereby reducing the risk of including infiltrated municipalities in the control group, which could significantly undermine the robustness of the analytical results.

A statistical analysis highlights how infiltrated municipalities exhibit distinct spending patterns, characterized by higher operating expenses, increased spending rigidity, and a notable misallocation of funds toward sectors like construction and waste management. These practices align with the strategic objectives of organized crime to capitalize on profitable, low-barrier sectors for money laundering and territorial control.

Additionally, the study underscores the reduced efficiency in revenue collection among infiltrated municipalities, particularly concerning waste and total taxes. This inefficiency may stem from reduced administrative capability or intentional under-collection of taxes from politically connected individuals and businesses, but also presumably reflects the broader impact of organized crime on civic engagement and trust in local governance.

The financial indicators are finally used to train a machine learning model to distinguish infiltrated municipalities from those in the control sample, with the ultimate goal of measuring the infiltration risk in all Italian municipalities. The model demonstrates a high degree of accuracy in predicting the risk of infiltration in local public administrations, with a performance level, as measured by AUC, of 98.2%. The model's robustness is further validated through external analyses examining the statistical connection between the infiltration risk computed by the model and the presence of organized crime-linked businesses, as well as the opacity in public procurements issued by the municipalities. These findings provide evidence supporting the model's discriminatory capability and its potential utility for policymakers and law enforcement agencies.

The research contributes significantly to the existing literature by offering a rigorous and comprehensive approach to assessing mafia infiltration risk in Italian local administrations. It extends the analytical toolkit available to scholars and practitioners alike, providing robust techniques to detect the pervasive influence of organized crime on local governance.

Moreover, and above all, this research has practical implications for the design and implementation of anti-mafia policies. The risk score computed by the machine learning model and the identification of financial indicators linked to infiltration according to the statistical analysis can guide the allocation of resources for monitoring and intervening in at-risk municipalities. This is particularly vital for enhancing the transparency and accountability of public administrations in regions vulnerable to organized crime.

References

Abbattista G., V. N. Convertini, V. Gattulli, L. Sarcinella (2020), "Use of the Logistic Regression Model for the analysis of mafia infiltration in Italian municipalities", Journal of Computer Engineering, Volume 22, Issue 1, Ser. II (Jan - Feb 2020), 37-41.

Alesina A., S. Piccolo, P. Pinotti (2019), "Organized crime, violence, and politics", The Review of Economic Studies 86(2):457–499.

Arellano-Bover, J, M. De Simoni, L. Guiso, R. Macchiavello, D. J. Marchetti and M. Prem (2024), "Mafias and Firms", CEPR Discussion Paper No. 18982. CEPR Press, Paris & London. https://cepr.org/publications/dp18982

Athey S., G. W. Imbens (2017), "The State of Applied Econometrics: Causality and Policy Evaluation", Journal of Economic Perspectives, 31 (2): 3-32.

Athey S., G. W. Imbens (2019), "Machine Learning Methods That Economists Should Know About", Annual Review of Economics, Vol. 11:685-725.

Baraldi A. L., G. Immordino, M. Stimolo (2022), "Self-selecting candidates or compelling voters: How organized crime affects political selection", European Journal of Political Economy, 71, 102133.

Barone G., G. Narciso (2015), "Organized Crime and Business Subsidies: Where Does the Money Go?", Journal of Urban Economics, 86, 98–110.

Campedelli, G. M. (2022), "Explainable machine learning for predicting homicide clearance in the United States", Journal of Criminal Justice, 79, 101898.

Campedelli, G. M., G. Daniele, M. Le Moglie (2024), "Mafia, Politics and Machine Predictions", CEPR Discussion Paper No. 19322. CEPR Press, Paris & London.

Cariello P., M. De Simoni, S. Iezzi (2024), "A machine learning approach for the detection of firms linked to organised crime in Italy, based on balance sheet data", Quaderni dell'Antiriciclaggio – Analisi e studi, Banca d'Italia, N. 22.

Chengwei L., C. Yixiang, S.H.A. Kazmi, F. Hao (2015), "Financial Fraud Detection Model Based on Random Forest", International Journal of Economics and Finance, vol. 7, no. 7.

D'Amato A., M. Mazzanti, F. Nicolli (2015), "Waste and Organized Crime in Regional Environments: How Waste Tariffs and the Mafia Affect Waste Management and Disposal," Resource and Energy Economics, vol. 41, 185–201.

Daniele V., U. Marani (2011), "Organized crime, the quality of local institutions and FDI in Italy: a panel data analysis", European Journal of Political Economy 27(1):13 42.

Daniele G., B. Geys (2015). "Organised Crime, Institutions and Political Quality: Empirical Evidence from Italian Municipalities", Economic Journal 125(586): F233–F255.

Daniele G., G. Dipoppa (2017) "Mafia, elections and violence against politicians", Journal of Public Economics, Volume 154, Pages 10-33, ISSN 0047-2727, https://doi.org/10.1016/j.jpubeco.2017.08.004.

de Blasio G., A. D'Ignazio, and M. Letta (2022), "Gotham city. Predicting corrupted municipalities with machine learning", Technological Forecasting and Social Change, November, 184, 122016.

De Feo, G., G. De Luca. 2017. “Mafia in the Ballot Box”, American Economic Journal: Economic Policy, 134–67.

De Simoni M. (2022), “The financial profile of firms infiltrated by organised crime in Italy”, Quaderni dell’antiriciclaggio – Analisi e studi, Banca d’Italia, N. 17.

Di Cataldo M., N. Mastrorocco (2022), “Organized Crime, Captured Politicians, and the Allocation of Public Resources”, The Journal of Law, Economics, and Organization, Vol. 38, No. 3.

Eboli, M., A. Toto, A. Ziruolo (2021). “Mafia infiltrations in Italian municipalities: two statistical indicators”, Applied Economics, 53(24), 2693-2712. https://doi.org/10.1080/00036846.2020.1866157

Fenizia A., R. Saggio (2024), “Organized Crime and Economic Growth: Evidence from Municipalities Infiltrated by the Mafia”, American Economic Review, vol. 114, no. 7.

Galletta S. (2017), “Law enforcement, municipal budgets and spillover effects: Evidence from a quasi-experiment in Italy”, Journal of Urban Economics, September, 101, 90–105.

Gara M., S. Iezzi, M. Siino (2024), “Corruption risk indicators in public procurement: A proposal using Italian open data”, Quaderni dell’antiriciclaggio – Analisi e studi, Banca d’Italia, N. 23.

Kleinberg J., H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan (2018), “Human Decisions and Machine Predictions,” The Quarterly Journal of Economics, 133 (1), 237–293.

Mirenda L., S. Mocetti, L. Rizzica (2022), “The economic effects of mafia: Firm level evidence”, American Economic Review, 112 (8), 2748–2773.

Mete V. (2008), “Fuori dal comune. Lo scioglimento delle amministrazioni locali per infiltrazioni mafiose”, Bonanno.

Mocetti S., L. Rizzica (2023). “Organized Crime in Italy: An Economic Analysis”, Italian Economic Journal. 10.1007/s40797-023-00236-4.

Le Moglie M., G. Sorrenti (2022), “Revealing Mafia Inc.”? Financial Crisis, Organized Crime, and the Birth of New Enterprises”, The Review of Economics and Statistics, 104 (1), 142–156.

Peri G. (2004), “Socio-Cultural Variables and Economic Success: Evidence from Italian Provinces 1951-1991”, B.E. Journal of Macroeconomics, 4(1), pp. 1-36.

Pinotti P. (2015), “The Economic Costs of Organised Crime: Evidence from Southern Italy”, The Economic Journal, 125 (586), F203–F232.

Ravenda D., M. G. Giuranno, M. M. Valencia-Silva, J. M. Argiles-Bosch, J. García-Blandón (2020), “The effects of mafia infiltration on public procurement performance”, European Journal of Political Economy, Volume 64, 101923, ISSN 0176-2680, https://doi.org/10.1016/j.ejpoleco.2020.101923.

Ravenda D., J. M. Argilés-Bosch, M. M. Valencia-Silva (2015), “Detection Model of Legally Registered Mafia Firms in Italy”, European Management Review, 12: 23-39.

Sciarrone R. (2009), “Mafie vecchie, mafie nuove - Radicamento ed espansione”, Donzelli Editore, Roma.

Sharma A., P. Panigrahi (2013), “A Review of Financial Accounting Fraud Detection based on Data Mining Techniques”, International Journal of Computer Applications vol. 39, n.1.

UIF (2021). “Rapporto Annuale 2020 – Unità di Informazione Finanziaria per l’Italia – Anno 2020”. UIF N. 13 – 2021. https://uif.bancaditalia.it/pubblicazioni/rapporto-annuale/2021/index.html

Varian H. R. (2014), “Big Data: New Tricks for Econometrics”. Journal of Economic Perspectives, 28 (2): 3-28.

Wyrobek J. (2020), “Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture”. Procedia Computer Science 176, 3037–3046.

Appendix

Table A1 - Spending missions and aggregation into macro-categories

Macro category	Spending mission
Administration, management, and control	1. Institutional, general, and management services 18. Relations with other territorial and local authorities 19. International relations 20. Funds and provisions 50. Public debt 60. Financial advances 99. Services on behalf of third parties
Public order and safety	2. Justice 3. Public order and safety
Education and culture	4. Education and the right to study 5. Protection and enhancement of cultural assets and activities
Housing	8. Territorial planning and housing
Territory and environment	9. Sustainable development and protection of the territory and the environment 11. Civil protection
Transport	10. Transport and the right to mobility
Social sector	12. Social rights, social policies, and family 6. Youth, sports, and leisure policies 13. Health protection 15. Policies for employment and vocational training
Economic development and productive activities	7. Tourism 14. Economic development and competitiveness 17. Energy and diversification of energy sources 16. Agriculture, agri-food , and fisheries

Table A2 – Infiltrated municipalities and control sample - main statistics of budget indicators

	Control sample		Infiltrated municipalities		t-test on mean difference	Regression coefficient	
	Mean	St. Dev.	Mean	St. Dev.			
Spending decisions in current as a percentage of total expenditure							
Administration, management and control	0.374	0.116	0.362	0.102	*	0.384	***
Public order and safety	0.037	0.029	0.052	0.029	***	0.128	**
Education and culture	0.095	0.046	0.065	0.028	***	-0.474	***
Housing	0.015	0.029	0.037	0.039	***	0.785	***
Territory and environment	0.197	0.093	0.318	0.090	***	0.547	***
Transport	0.073	0.049	0.043	0.036	***	-0.211	*
Social sector	0.179	0.121	0.111	0.087	***	-1.241	***
Economic development and productive activities	0.029	0.066	0.013	0.020	***	-0.623	***
Spending decisions in capital as a percentage of total expenditure							
Administration, management and control	0.185	0.207	0.172	0.232		0.135	
Public order and safety	0.073	0.126	0.067	0.145		-0.072	
Education and culture	0.013	0.046	0.011	0.061		-0.098	
Housing	0.037	0.089	0.039	0.118		-0.089	
Territory and environment	0.054	0.104	0.047	0.122		-0.244	
Transport	0.016	0.072	0.003	0.031	***	-1.325	*
Social sector	0.269	0.248	0.368	0.301	***	0.306	**
Economic development and productive activities	0.351	0.249	0.293	0.264	***	-0.270	**
Spending decisions in current and capital as a percentage of total expenditure							
Administration, management and control	0.320	0.110	0.322	0.093		0.336	***
Public order and safety	0.031	0.026	0.046	0.026	***	0.145	**
Education and culture	0.109	0.066	0.083	0.047	***	-0.389	***
Housing	0.048	0.074	0.061	0.069	***	0.376	***
Territory and environment	0.185	0.097	0.301	0.088	***	0.543	***
Transport	0.109	0.080	0.057	0.051	***	-0.324	***
Social sector	0.164	0.108	0.113	0.083	***	-1.033	***
Economic development and productive activities	0.035	0.068	0.018	0.039	***	-0.551	**
Other expenditure-related indicators							
Spending inefficiency	0.747	0.121	0.816	0.082	***	0.392	***
Rigidity of current expenditure	0.297	0.128	0.445	0.217	***	0.089	***
Current expenditure per capita	1120.2	675.9	748.3	214.8	***	-206.300	***
Capital expenditure per capita	452.8	1124.4	165.1	218.6	***	-68.870	***
Per capita personnel expenditure	826.0	518.9	607.7	177.4	***	-79.550	***
Total expenditure per capita	1573.0	1515.9	913.4	339.5	***	-262.500	***
Revenue indicators							
Efficiency of property taxes	0.842	0.174	0.719	0.212	***	-0.035	
Efficiency of waste taxes	0.651	0.241	0.397	0.228	***	-0.145	***
Efficiency of tax revenues	0.748	0.153	0.585	0.161	***	-0.327	***
Efficiency of total revenues	0.790	0.108	0.738	0.113	***	-0.478	***
Financial autonomy	0.640	0.286	0.730	0.184	***	0.918	***
Tax autonomy	0.460	0.236	0.577	0.177	***	0.754	***
Tax burden per capita	440.6	315.6	334.6	134.4	***	28.990	**
Total revenues per capita	2006.8	1460.9	1545.4	852.0	***	-221.300	***

Note: The sample of infiltrated municipalities includes 221 observations of municipalities. The control sample includes 6,571 observations. The analysis period is 2016-2021. Regression coefficient indicates the parameter of infiltration dummy in a logit regression or a robust linear regression. The table reports the results of the two-tailed test on the difference between two means. A significant difference indicates the rejection of the hypothesis that the two means are equal. ***: significance at 1%; ** significance at 5%; * significance at 10%. In the regression coefficient column, the asterisks refer to the significance of the coefficient. A positive coefficient indicates that the indicator is on average higher for infiltrated municipalities.

Table A3. Main statistics of Plan of Indicators

	Control Sample			Infiltrated municipalities			t-test on mean difference	Regression coefficient	
Variable	Mean	Std Dev	N	Mean	Std Dev	N			
IND2_1 (b)	322.48	1,446.16	6,456	2,076.50	3,800.56	216	***	-4.61	***
IND2_2 (b)	289.75	1,326.18	6,443	1,241.38	2,962.21	194	***	-2.49	**
IND2_3 (b)	184.07	853.08	6,468	1,146.83	2,207.36	215	***	16.54	***
IND2_4 (b)	163.25	759.12	6,458	757.19	1,834.76	195	***	16.50	***
IND2_5 (b)	238.69	1,094.30	6,452	1,097.99	2,217.92	212	***	-23.29	***
IND2_6 (b)	211.58	989.48	6,434	699.90	1,734.37	193	***	-20.22	***
IND2_7 (b)	144.26	703.71	6,468	577.69	1,344.04	211	***	-0.68	
IND2_8 (b)	124.93	615.74	6,451	432.23	1,244.86	192	***	0.33	
IND3_1 (c)	50.14	948.60	4,491	14.35	94.48	164	**	N.A.	
IND3_2 (c)	7.23	50.94	4,510	23.57	84.89	172	**	N.A.	
IND4_1 (a)	0.28	0.09	6,217	0.32	0.12	167	***	0.20	***
IND4_2 (a)	0.09	0.07	6,028	0.06	0.05	164	***	-0.32	***
IND4_3 (a)	0.09	0.11	6,111	0.06	0.11	175	***	0.01	
IND4_4 (b)	248.86	455.31	6,126	238.29	1,321.27	201		-72.76	***
IND5_1 (a)	0.21	0.12	6,203	0.26	0.14	180	***	-0.05	
IND6_1 (b)	6.93	39.40	6,263	83.60	189.90	214	***	1.53	***
IND6_2 (a)	0.02	0.09	4,539	0.05	0.11	153	***	0.25	
IND6_3 (a)	0.00	0.05	4,412	0.01	0.04	128		0.23	
IND7_1 (a)	0.21	0.14	6,217	0.15	0.15	168	***	-0.06	
IND7_2 (b)	327.72	737.34	6,163	187.39	1,364.77	208	***	-42.30	
IND7_3 (c)	20.41	90.46	5,309	2.63	21.91	132	***	N.A.	
IND7_4 (b)	347.68	819.41	6,171	95.24	172.57	207	***	-47.17	***
IND7_5 (b)	120.84	552.39	5,758	351.12	1,237.03	205	***	-13.50	***
IND7_6 (c)	0.81	7.02	4,490	0.92	7.43	155		N.A.	
IND7_7 (c)	13.98	90.40	4,763	94.76	336.75	165	***	N.A.	
IND8_1 (a)	0.74	0.19	6,146	0.55	0.20	169	***	-0.73	***
IND8_2 (a)	0.64	0.30	6,112	0.44	0.31	170	***	-0.46	***
IND8_3 (a)	0.05	0.22	4,378	0.03	0.16	150	*	-1.04	*
IND8_4 (a)	0.49	0.21	6,220	0.36	0.18	170	***	-0.09	
IND8_5 (a)	0.44	0.33	6,237	0.34	0.33	180	***	0.22	
IND8_6 (a)	0.07	0.23	4,447	0.03	0.16	144	**	-0.71	
IND9_1 (b)	184.15	871.57	6,374	956.23	2,034.20	214	***	-16.53	***
IND9_2 (b)	183.03	848.39	6,380	827.68	1,825.53	216	***	-22.73	***
IND9_3 (b)	99.52	468.14	6,269	237.83	906.54	175	**	-21.69	***
IND9_4 (b)	94.80	452.86	6,174	139.21	640.90	170		-14.84	***
IND9_5 (b)	514.62	1,909.59	4,623	503.86	2,015.29	183		-1.14	***
IND10_3 (b)	17.09	84.46	6,302	129.36	309.82	208	***	2.58	***
IND11_1 (a)	0.35	0.32	5,890	0.18	0.32	137	***	-0.34	
IND11_2 (a)	0.07	0.13	5,485	0.03	0.08	159	***	-0.92	***
IND11_3 (a)	0.35	0.29	5,607	0.41	0.37	96		-0.04	
IND11_4 (a)	0.19	0.20	5,797	0.13	0.18	142	***	-0.61	***
IND12_4 (c)	0.76	3.43	4,556	3.47	6.79	142	***	N.A.	
IND13_1 (c)	0.49	3.70	4,748	1.50	7.94	138		N.A.	
IND13_2 (c)	0.05	0.47	4,432	0.48	1.67	155	***	N.A.	
IND13_3 (c)	0.07	0.50	4,385	0.37	1.30	149	***	N.A.	
IND14_1 (a)	0.65	0.34	5,958	0.51	0.41	172	***	-0.56	***
IND15_1 (a)	0.16	0.08	6,250	0.16	0.12	159		0.16	**
IND15_2 (a)	0.18	0.09	6,170	0.19	0.13	157		0.21	***

Note: The sample of infiltrated municipalities includes 221 observations of municipalities. The control sample includes 6,571 observations. The number of observations for the single indicator can be different because missing values.

(a) Regression parameter estimated with logit clustered at municipal level.
(b) Regression parameter estimate with robust regression clustered at municipal level.
(c) Regression parameter not estimable because of too many zeroes in the indicator.

Table A4. Plan of Indicators and Expected Budget Results

Category	Indicator	Variable
Current revenues	Incidence of current account assessments on initial current account forecasts	IND2_1
	Incidence of current account assessments on the final current account forecasts.	IND2_2
	Incidence of own revenue assessments on initial current account forecasts	IND2_3
	Incidence of own revenue assessments on the final current account forecasts	IND2_4
	Incidence of current receipts on initial current account forecasts	IND2_5
	Incidence of current receipts on the final current account forecasts.	IND2_6
	Incidence of own revenue cash collections on initial current account forecasts	IND2_7
Treasury Institute Advances	Average Use of Cash Advances	IND3_1
	Advances Closed Only on Accounting Basis	IND3_2
Personnel Expenses	Incidence of personnel expenses on current expenditure	IND4_1
	Incidence of supplementary and incentive salary compared to the total personnel expenses.	IND4_2
	Incidence of flexible personnel expenses compared to the total personnel expenses.	IND4_3
	Per-capita personnel expenses (Absolute value indicator of dimensional balance)	IND4_4
Outsourcing of Services	Indicator of service outsourcing	IND5_1
Passive Interest	Incidence of passive interest on current revenues	IND6_1
	Incidence of passive interest on advances as a percentage of the total expenditure for passive interest	IND6_2
	Incidence of late interest on the total expenditure for passive interest	IND6_3
Investments	Incidence of investments on the total of current and capital expenditure	IND7_1
	Per capita direct investments (in absolute value)	IND7_2
	Per capita contributions to investments (in absolute value)	IND7_3
	Total per-capita investments (in absolute value)	IND7_4
	Share of total investments financed by current savings	IND7_5
	Share of total investments financed by the positive balance of financial items	IND7_6
	Share of total investments financed by debt	IND7_7
Residual Analysis	Incidence of new current passive residuals on the current stock of passive residuals	IND8_1
	Incidence of new capital passive residuals on the capital stock of passive residuals as of December 31	IND8_2
	Incidence of new passive residuals due to an increase in financial assets on the stock of passive residuals due to an increase in financial assets as of December 31	IND8_3
	Incidence of new current active residuals on the current stock of active residuals	IND8_4
	Incidence of new capital active residuals on the capital stock of active residuals	IND8_5
	Incidence of new active residuals due to a reduction in financial assets on the stock of active residuals due to a reduction in financial assets	IND8_6
Non-Financial Debt Disposal	Disposal of commercial debts incurred during the fiscal year	IND9_1
	Disposal of commercial debts incurred in previous fiscal years	IND9_2
	Disposal of debts to other public administrations incurred during the fiscal year	IND9_3
	Disposal of debts to other public administrations incurred in previous fiscal years	IND9_4
	Annual indicator of payment timeliness (as per Article 9, Paragraph 1, of the DPCM of September 22, 2014)	IND9_5
Financial Debts	Sustainability of financial debts	IND10_3
Composition of the Administration Surplus	Incidence of the free share of current part in the administration surplus	IND11_1
	Incidence of the free share in capital/capita in the administration surplus	IND11_2
	Incidence of the reserved share in the administration surplus	IND11_3
	Incidence of the earmarked share in the administration surplus	IND11_4
Administration Deficit	Sustainability of the actual deficit burden on the fiscal year	IND12_4
Off-Budget Debts	Recognized and funded debts	IND13_1
	Debts under recognition	IND13_2
	Recognized and currently funded debts	IND13_3
Multi-Year Earmarked Fund	Utilization of the MYEF (Multi-Year Earmarked Fund)	IND14_1
Revolving and Third-Party Accounts	Incidence of clearing accounts and third-party accounts on income	IND15_1
	Incidence clearing accounts and third-party accounts outgoing	IND15_2

Figure A1. Feature importance