



FINANCIAL STATEMENTS AS AN INFORMATION BASE FOR MONEY LAUNDERING RISK ASSESSMENT: A COMPARISON AMONG ITALIAN DETECTION METHODS

Annalisa Baldissera¹, Claudio Tedesco²

¹*Associate Professor in Business Administration, Department of Law, University of Brescia – Italy*

²*Independent Researcher – Italy*

Abstract

Annual financial statements are one of the information sources used to assess the risk that a firm is involved in money laundering, tax fraud or organised crime infiltration. This paper compares the detection tools based on the financial statements of Italian firms, developed along two distinct strands: academic research and the studies of the Financial Intelligence Unit for Italy (UIF). The five methodological approaches examined – a logistic regression model, two composite indicators, and two machine learning models – serve as tools for identifying firms linked to organised crime, companies issuing invoices for non-existent transactions, and firms interposed in complex tax fraud schemes. The comparison considers four dimensions: detection purpose and target, reference firms, information set, and analytical method. It shows that risk assessment depends on three main choices: which financial statement items to use, which firms to take as reference, and how the data are processed. As these choices vary across tools, financial statements do not provide a uniform information base, and differences in reference firms, validation conditions, and performance measures make reported results not directly comparable. The paper's originality lies in proposing a criterion for variable selection: when variables have equal predictive power, preference should be given to those that firms targeted for detection can alter only at a high cost. This provides a descriptive and conceptual contribution by introducing the cost of manipulating observable characteristics as an additional criterion for selecting predictors.

Keywords

Anti-Money Laundering; Financial Statements; Organised Crime; Fraud Detection; Machine Learning

1. Introduction

Assessing money laundering risk requires information about a firm that is available, prepared according to uniform criteria and comparable over time. Annual financial statements meet these requirements: they are prepared according to formats and principles laid down by law and, in Italy, limited companies (*società di capitali*) are required to file them with the Business Register, which makes large-scale analysis possible. Between 2015 and 2025 they became the information base for detection tools of varying nature and purpose, developed in academia and within Anti-Money Laundering (AML) authorities.

The Italian experience offers a rich field of observation, because these tools have developed along two distinct strands. In the academic literature Ravenda *et al.* (2015) proposed an econometric model to help detect legally registered firms defined as “mafia” firms because they were confiscated by the judicial authorities in relation to their owners’ alleged connections with organised crime; Ambrosini *et al.* (2024)

proposed a machine learning model that identifies, on the basis of financial statements, firms whose directors, or shareholders holding at least 10% of the share capital, have been charged with mafia-related offences. The research series of the Financial Intelligence Unit for Italy (*Unità di Informazione Finanziaria per l'Italia*, UIF), the *Quaderni dell'antiriciclaggio* (Anti-Money Laundering Papers), has published a composite indicator to identify so-called “*società cartiere*”, shell companies that issue invoices for non-existent transactions (Pellegrini *et al.*, 2020); a machine learning model that assigns a score for the risk of links with organised crime to all limited companies active in Italy between 2010 and 2021 (Cariello *et al.*, 2024); and a composite indicator to identify so-called buffer companies, which in complex tax fraud schemes are interposed between shell companies and the firms that benefit from the fraud (De Simoni and Pellegrini, 2025).

Each of these tools documents its own method and results; comparing them shows how risk assessments have been derived from the same information base through different choices. This paper undertakes that comparison by addressing two research questions:

RQ1. How do detection tools referring to Italian firms derive a risk assessment from annual financial statements?

RQ2. How do the tools differ in the way they derive that assessment, and what are the consequences for interpreting their results?

The contribution consists of a structured survey of the tools, a four-dimensional comparison grid, and a proposed criterion for selecting variables, based on the cost that the firm to be detected would incur to alter them.

After situating financial statements within the prevention system (Section 2) and reviewing the literature (Section 3), the paper sets out the method (Section 4), describes the tools (Section 5), compares them (Section 6) and discusses the implications (Section 7); limitations and conclusions follow.

2. Financial statements as an information source in the prevention system

Legislative Decree No. 231 of 21 November 2007 (hereinafter, the AML Decree) requires a set of parties, known as obliged entities, to know their customers through customer due diligence and to submit Suspicious Transaction Reports (STRs) to the UIF. Obligated entities include banking and financial intermediaries, other financial operators, professionals, other non-financial operators and gaming service providers. The UIF, established at the Bank of Italy, is autonomous and operationally independent; it receives the reports and carries out their financial analysis (AML Decree, Articles 3, 6, 17 and 35). Quotations from Italian legislation and documents are translated by the authors.

Within this system, economic information about the customer comes into play at several points. In customer due diligence, obliged entities calibrate their measures to the risk and, to that end, consider among other things the reasonableness of the transaction, the business relationship or the professional service “in relation to the activity carried out by the customer and the extent of the economic resources at his disposal”; they may also obtain, depending on the risk, information on “the customer’s economic and financial situation” (AML Decree, Articles 17 and 18). As regards reporting, suspicion is inferred from the characteristics, size and nature of the transactions, their links or fragmentation, or any other circumstance known by reason of the functions performed, “also taking into account the economic capacity and the activity carried out by the person to whom it refers” (AML Decree, Article 35). Alongside the activity carried out, the legislation thus refers to the economic and financial size of the party concerned. The Decree does not mention financial statements for these purposes; for companies, however, the financial statements are the document that “must give a true and fair view of the company’s assets, liabilities and financial position and of its profit or loss for the year” (Italian Civil Code, Article 2423). This correspondence, adopted here as an interpretive key, underpins the use of financial statements as an information source for risk assessment.

To facilitate the detection of suspicious transactions, the UIF also issues anomaly indicators (AML Decree, Article 6): an operational tool that helps obliged entities identify situations arising in the course of their activities that should be considered when assessing whether the conditions for submitting a suspicious transaction report are met. The UIF Measure of 12 May 2023 contains 34 indicators, each broken down into examples called sub-indicators; the list is not exhaustive, and each obliged entity selects the indicators relevant to its activity. Some sub-indicators refer directly to financial statement figures: sub-indicator 9.1,

for example, concerns transactions of significant overall amount requested by parties with a “limited economic and financial profile (e.g. with limited turnover, minimal or negative share capital or equity), which may also be inferred from tax returns” (UIF, 2023, sub-indicator 9.1). These are textual descriptions of situations, not to be confused with the composite indicators examined in Section 5, which are numerical values computed for each firm.

The amount of information that financial statements make available, and the extent to which it is verified, vary, however, with the size of the firm. The financial statements filed with the Business Register (Italian Civil Code, Articles 2435 and 2478-bis) do not have the same level of detail for all companies. Companies that do not exceed two of three thresholds (total assets of €5.5 million, revenues of €11 million, an average of 50 employees) may prepare them in abridged form: the balance sheet is limited to the main headings, without the breakdown into sub-items, some income statement items may be grouped together, the notes are reduced and no cash flow statement is required (Italian Civil Code, Article 2435-bis). For micro-enterprises, identified by even lower thresholds, the notes may also be omitted (Italian Civil Code, Article 2435-ter). In limited liability companies, moreover, the appointment of a supervisory body or a statutory auditor is mandatory, save in particular cases, only if at least one of three lower thresholds is exceeded for two consecutive financial years: total assets of €4 million, revenues of €4 million or 20 employees (Italian Civil Code, Article 2477). A limited liability company that remains below these thresholds may therefore file abridged financial statements without being required to appoint a supervisory body or a statutory auditor. It follows, for those analysing financial statements from the outside, that the variables that can be computed depend on the regime under which the document was prepared: without a cash flow statement, cash flows cannot be read directly, and some ratios cannot be constructed from the main headings alone.

3. Literature review

A substantial body of literature has highlighted the importance of financial statements in assessing money laundering risks.

A first strand of studies has documented how criminal infiltration is reflected in financial statements. Using a sample of 237 firms seized by the judiciary in organised crime-related prosecutions, De Simoni (2022) finds that infiltrated firms, despite higher revenues than legal businesses, appear less profitable, have low inventory levels and, in the two most frequent types, larger cash holdings, with profiles that vary by type of infiltration. Arellano-Bover *et al.* (2024) find that about 2% of Italian firms appear to have links with criminal organisations, and distinguish three motives for infiltration, which their model associates with firm size: small firms are predominantly used to conduct criminal activities, medium-sized firms primarily gain a competitive advantage from them, and the largest firms mostly remain separate from criminal activities. Other studies have investigated the conditions that precede infiltration. Castelluccio and Rizzica (2025) show that a 10% drop in revenues leads to a 4.9% increase in the probability of “mafia” infiltration; Daniele *et al.* (2025) show that a downgrade to substandard credit status reduces credit availability by 30% over five years and increases the probability of infiltration by 5%, relative to comparable firms. Firms’ economic and financial conditions are thus associated both with infiltration that has occurred and with the vulnerability that precedes it; the signals, however, are not uniform. Three of the studies cited here were published in the UIF’s *Quaderni dell’antiriciclaggio* series, two of them written by UIF analysts together with scholars from universities and research institutes.

Ambrosini *et al.* (2024) offer a theoretical explanation of why financial statements contain such signals, which they call constrained transparency: to access credit, win public contracts or operate within the legal economy, criminal firms must file formally compliant financial statements, while their financial behaviour is shaped by illicit objectives. The resulting irregularities do not necessarily require intentional misstatement: they may arise simply because the firm’s economic model is not oriented towards lawful profit. The same authors stress, however, that criminally connected firms are not a homogeneous group, and expect only a subset of them to leave observable traces in accounting data.

A second set of issues concerns the relationship between a detection tool and the party it assesses. The literature on strategic classification studies settings in which those being classified can change their data, at a cost, to obtain a more favourable outcome, while the classifier’s goal remains to classify accurately with respect to the original data (Hardt *et al.*, 2016). Kleinberg and Raghavan (2020) examine when those being evaluated can be incentivised to invest effort in actually improving their outcomes, as opposed to “gaming” the classifier. Ghalme *et al.* (2021) show that the problem does not disappear when

the classifier is not disclosed: in that case, agents attempt to learn the classifier and game it, and under certain conditions transparency is preferable to opacity. The issue is relevant for tools based on financial statements, because the statements from which all or part of the variables are drawn are prepared by the very party that the tool seeks to classify.

Alongside these studies, detection tools have emerged from two distinct strands of work – the academic literature and institutional practice – which are described in Section 5 and compared in Section 6.

4. Methodology

The paper adopts a comparative research design based on document analysis. The objects of comparison are detection tools described in publicly available publications; the data on which they were built are wholly or partly inaccessible, and the tools have therefore not been reconstructed, nor their performance recalculated. This design is appropriate to the research questions, which concern not the effectiveness of the tools but the way each derives a risk assessment from financial statements – an aspect that the documents describe.

Tools were included if they met three criteria: they return, for each firm, an assessment of its possible connection with criminal phenomena, in the form of a score, a probability, a classification or the value of an indicator; they are explicitly built on financial statement data, possibly supplemented by other sources; and they concern Italian firms. Because all the tools concern Italian firms, the financial statements they use are prepared under the rules described in Section 2, albeit with the changes made over time, and differences between the tools reflect the choices of those who built them rather than membership of different accounting systems.

The identification of relevant detection tools followed two search channels. The first covered the UIF's *Quaderni dell'antiriciclaggio* series, available on the Unit's website, and included all issues published up to No. 36 (July 2026). The search focused on the *Analisi e studi* (Analyses and Studies) strand, which, according to the series' editorial note, includes contributions on the "issues and methods" of combating money laundering. The other strands, devoted to statistics and regulatory and case-law developments, were not considered.

Of the 36 issues, 26 were excluded on the basis of their titles, while the remaining 10 were reviewed in full. Three met the inclusion criteria and proposed detection tools relevant to the analysis (Cariello *et al.*, 2024; De Simoni & Pellegrini, 2025; Pellegrini *et al.*, 2020). Three others examine infiltrated firms empirically but do not propose detection tools and are therefore discussed in Section 3 (Arellano-Bover *et al.*, 2024; Daniele *et al.*, 2025; De Simoni, 2022). The remaining four issues (Nos. 13, 14, 23, and 30) were excluded because they do not assess firms or do not use financial statement data. They address, respectively, cash use at the bank-branch and municipal levels, data-analysis technologies used by authorities, corruption risk in public procurement, and safeguards and prevention tools aimed at protecting legality in the economy.

The second channel was backward citation search: starting from the documents already identified, the works they cite were examined. This led to the two academic tools: Ravenda *et al.* (2015), which Cariello *et al.* (2024) cite as a similar methodology, and Ambrosini *et al.* (2024), cited by Daniele *et al.* (2025). The reference lists of these two works revealed no other tools concerning Italian firms. The same documents confirm the scarcity of precedents: Pellegrini *et al.* (2020) describe the literature on methods for identifying shell companies as "very limited", and De Simoni and Pellegrini (2025) state that, to the best of their knowledge, no quantitative studies have examined the financial statement characteristics of buffer companies.

The tools are read along four dimensions. The first is the purpose and target of detection: for what use and for which users the tool is designed, and what it seeks to identify. The second is the reference firms, that is, the firms regarded as criminal or anomalous on which the tool is built or validated: how they were identified – for example through confiscations, criminal proceedings, court rulings or suspicious transaction reports – and whether they serve to estimate the tool or only to validate its results. The third is the information set: which financial statement figures are used and, where applicable, which other sources. The fourth is the analytical method, which in the cases examined ranges from the sum of indices with predetermined weights to logistic regression and machine learning.

Figures are reported as they appear in the text of the documents, without recalculation or reconciliation, and the comparison concerns the methods as documented, not the validity of the results.

5. The detection tools

The tools are described in chronological order, according to the four dimensions of the grid: purpose and target, reference firms, information set, and method.

5.1 Ravenda *et al.* (2015)

Ravenda *et al.* (2015) propose a model that can contribute to the detection of legally registered “mafia” firms, which they present, to their knowledge, as the first accounting detection model of its kind and which, according to the authors, can aid practitioners and regulators. The reference firms are firms confiscated from organised crime by the judicial authorities. Some were provided by the Italian Agency for the Administration and Destination of Seized and Confiscated Assets (“Agenzia nazionale per l'amministrazione e la destinazione dei beni sequestrati e confiscati alla criminalità organizzata – ANBSC”), the government authority that administers assets seized from organised crime, and were all confiscated by final judgment; the others, confiscated at first instance, were identified in the AIDA database by Bureau van Dijk (Moody's) or through online newspapers. After excluding those confiscated before 2005, 198 remain. Only the financial statements preceding confiscation are considered, and each is matched with that of a legitimate firm of the same year, industry and size, for a total of 852 annual financial statements, on which the model is estimated. The financial statements come from the AIDA database, which according to the authors covered about one million Italian companies with a turnover above €500,000, for the period 2003–2012. The final model includes 19 explanatory variables – among them changes in receivables and inventories, abnormal material expenses, bank leverage and revenue per employee – plus industry and year controls. The method is logistic regression, which estimates the probability that a firm belongs to the group of “mafia” firms. At a probability threshold of 0.50, the model correctly classifies 76.29% of the financial statements of “mafia” firms (sensitivity) and 76.53% of those of legitimate firms (specificity). The Receiver Operating Characteristic (ROC) curve describes, as the threshold varies, the relationship between the share of criminal firms correctly identified and the share of legitimate firms wrongly flagged; the Area Under the Curve (AUC) summarises the tool's ability to distinguish the two groups, and equals 0.5 for a tool that performs no better than chance and 1 for a perfect tool. In Ravenda *et al.*'s model, it is approximately 0.82. Lowering the threshold to 0.35 raises sensitivity to approximately 90% and lowers specificity to approximately 60%. In an out-of-sample validation, conducted by excluding in turn the firms confiscated in each year between 2011 and 2013, together with their control firms, the share of correct classifications ranges between 69.55% and 77.91%.

5.2 Pellegrini *et al.* (2020)

The composite indicator of Pellegrini *et al.* (2020) aims to identify shell companies and, according to the authors, can serve both financial and investigative analysis of tax offences and money laundering and the exclusion of fictitious firms from sectoral or performance analyses. No reference firms are used in its construction: the characteristics of shell companies are derived from Financial Action Task Force (FATF) documents, rulings of the Italian Supreme Court of Cassation and UIF analyses. The firms mentioned in the suspicious transaction reports received by the UIF between 2016 and 2019 serve only to validate the indicator after its construction: very low values correspond more frequently to companies reported for invoicing fraud or intra-EU VAT fraud than to companies reported for other phenomena. The information set consists of five indices computed on financial statement data from the Cerved database, which covers almost all Italian limited companies, for the period 2016–2018: tangible fixed assets to total assets, financial charges to revenues, paid-in capital and reserves to liabilities, an index measuring how closely the costs of purchases, services and use of third-party assets approach revenues, and personnel costs to revenues. Since many shell companies prepare abridged financial statements, the indices are constructed using only the items available in that form, plus paid-in share capital. The indices, available for 342,073 firms, are standardised and summed with equal weights, in the absence of empirical knowledge of the relationships between them. Adopting a “prudent approach”, the authors consider the probability of facing a shell company “reasonably high” for firms falling within the lowest five percentiles of the distribution.

5.3 Cariello *et al.* (2024)

The model of Cariello *et al.* (2024) assigns each limited company a score for the risk of links with organised crime and is proposed as a preliminary screening tool supporting anti-money laundering authorities and law enforcement agencies. The reference firms, 28,570 net of overlaps, come from four lists: businesses seized or confiscated in the main anti-mafia investigations conducted in Italy between

2007 and 2017; seized businesses from the archives of the ANBSC; companies with stakeholders or administrators implicated in organised crime-related proceedings, drawn from a commercial database based on press news; and, for the most part, companies whose stakeholders or administrators, reported in suspicious transaction reports, are under investigation for “mafia”-related crimes, or are mentioned in related judicial documents, according to the databases of the *Direzione Nazionale Antimafia e Antiterrorismo* (National Anti-Mafia and Anti-Terrorism Directorate), or are the subject of information requests to the UIF in connection with organised crime investigations. These firms are used to train the model and to test its results; the model is also validated on independent data, namely the firms registered in the whitelists kept by prefectures for the business activities most exposed to the risk of infiltration, and the firms subject to anti-mafia injunctions (*interdittive antimafia*), measures that impose restrictions or prohibitions because of alleged criminal connections. The information set comprises 32 variables for the period 2010–2021, drawn from financial statements in the Cerved database, from the Bank of Italy’s Central Credit Register (where a customer is reported if its outstanding debt to a bank or financial intermediary is €30,000 or more), from data of the National Institute of Social Security (INPS) on employment, from the Chambers of Commerce and, for provincial value added, from the Italian National Institute of Statistics (ISTAT). The method is a machine learning algorithm of the gradient boosting family (XGBoost), based on the combination of numerous decision trees. Since the financial statements of firms linked to organised crime account for only 1.4% of the total, the model is built and tested on a rebalanced sample in which they represent about half. At a threshold of 0.5, which the authors consider to offer a near-optimal balance between sensitivity and specificity, the model correctly identifies in the test set approximately 76% of the firms linked to organised crime and 74% of the legitimate ones; in 2021 it assigns a score to 931,163 firms.

5.4 Ambrosini et al. (2024)

Ambrosini et al. (2024) ask whether criminal firms can be identified using their own financial statements and propose a tool that they describe as usable for compliance, enforcement and policy design. The reference firms are identified from 155 court documents concerning mafia-related offences between 2008 and 2021: a firm is considered criminal if an individual charged with mafia-related crimes sits on its board of directors or holds at least 10% of its capital, only for the years of infiltration. The firms thus identified for which financial statements are available number 2,082; legitimate firms are those operating in the same municipality and in the same year for which no criminal connection was detected. The sample is split according to the year of the anti-mafia operation that brought each firm to light: firms identified up to 2019 are used to build the model, those identified from 2020 onwards to test it, so as to emulate a realistic application in which past enforcement actions are used to detect new cases. The information set, drawn from the AIDA database for the period 2008–2021, includes financial statement items, financial ratios, variables taken from the literature on criminal firms and from institutional reports, comparisons with industry medians, and annual changes; missing values are neither removed nor imputed, because the authors consider that missingness may itself be correlated with the criminal nature of the firm. The method is the same XGBoost algorithm, with eleven financial statements of legitimate firms for every financial statement of a criminal firm, in both model building and testing. The best model achieves an AUC of 0.758 and detects 12.5% of the criminal firms’ financial statements in the hold-out sample, with a precision – that is, a share of correct flags – of 90.4%.

5.5 De Simoni and Pellegrini (2025)

The composite indicator of De Simoni and Pellegrini (2025) aims to identify buffer companies, which in complex tax fraud schemes are positioned between shell companies and the real operating companies that are the ultimate beneficiaries of the fraud, and is intended for the anti-money laundering functions of financial intermediaries; the authors present it as a preliminary pilot study, which requires additional data and analyses to be validated. The reference firms are 92 buffer companies identified through Court of Cassation rulings and suspicious transaction reports, 39 of which could be used. Comparison with similar firms, selected through propensity score matching, and with shell companies highlights their distinctive features; on this basis, a logistic regression discriminating between buffer companies and other firms estimates the weights of the indicator. The information set, drawn from the Cerved database, is reduced to four signals: the ratio of purchases to revenues, working capital relative to revenues, not being reported in the Central Credit Register, and revenues per employee. The threshold is chosen to correctly classify at least three out of four buffer companies: the resulting sensitivity is 75.68% and specificity 68.36%, with an

AUC of 0.7706. The authors recommend using the indicator in conjunction with liquidity indicators and following it with a second-level analysis, which is necessary to discard false positives.

6. Comparing the tools

Table 1 summarises the five tools along the four dimensions of the grid and indicates the conditions under which their performance is measured.

Table 1 The detection tools compared

Tool	Purpose and target	Reference firms and their role	Information set	Method	Performance validation
Ravenda <i>et al.</i> (2015)	Legally registered “mafia” firms; practitioners and regulators	198 confiscated firms, by final judgment or at first instance; model estimation	AIDA financial statements 2003–2012; 19 explanatory variables plus industry and year controls	Logistic regression	In the matched estimation sample: sensitivity 76.29%, specificity 76.53%, AUC approx. 0.82. Out of sample: 69.55–77.91% correct classifications
Pellegrini <i>et al.</i> (2020)	Shell companies; financial and investigative analysis, sectoral or performance analyses	None in construction; firms reported to the UIF (2016–2019) for validation only	Cerved data 2016–2018; 5 indices built on the items of the abridged form	Sum of standardised indices with equal weights; lowest 5 percentiles	Higher frequency of fraud-related reports among the lowest values
Cariello <i>et al.</i> (2024)	Firms linked to organised crime; AML authorities and law enforcement agencies	28,570 firms from four lists: two of seized or confiscated firms, one from press news, one from STRs cross-checked against anti-mafia databases and from information requests; training and testing	32 variables 2010–2021 from Cerved, Central Credit Register, INPS, Chambers of Commerce, ISTAT	Gradient boosting (XGBoost)	On separate data drawn from a sample rebalanced to 50%, threshold 0.5: approx. 76% of firms linked to organised crime and 74% of legitimate firms correctly identified
Ambrosini <i>et al.</i> (2024)	Mafia-connected firms; compliance, enforcement and policy design	2,082 firms with directors or 10% shareholders charged in mafia-related proceedings (2008–2021); training and testing	AIDA financial statements 2008–2021; items, ratios, variables from the literature and institutional reports, industry medians; missing values as a signal	Gradient boosting (XGBoost)	On firms identified in anti-mafia operations from 2020 onwards, eleven financial statements of legitimate firms for each of a criminal firm: AUC 0.758; 12.5% of criminal financial statements detected, with 90.4% precision

De Simoni and Pellegrini (2025)	Buffer companies in complex tax fraud; AML functions of financial intermediaries	92 buffer companies from Court of Cassation rulings and STRs, 39 usable; weight estimation	Cerved data; 4 signals, including absence from the Central Credit Register	Indicator with weights from logistic regression; threshold set for a sensitivity of at least 75%	Sensitivity 75.68%, specificity 68.36%, AUC 0.7706; further validation declared necessary
---------------------------------	--	--	--	--	---

Source: Authors' elaboration based on Ravenda *et al.* (2015), Pellegrini *et al.* (2020), Cariello *et al.* (2024), Ambrosini *et al.* (2024) and De Simoni and Pellegrini (2025)

As regards their target, three tools seek firms linked to organised crime (Ambrosini *et al.*, 2024; Cariello *et al.*, 2024; Ravenda *et al.*, 2015), while two identify the vehicles of tax fraud, namely shell companies and buffer companies (De Simoni & Pellegrini, 2025; Pellegrini *et al.*, 2020). The declared users also differ, ranging from practitioners and regulators to anti-money laundering authorities and law enforcement agencies, and to the anti-money laundering functions of financial intermediaries. The same financial statements are thus read to answer different questions, at different points of the prevention system: from the preliminary screening that an intermediary may carry out on its own customers to the analysis that an authority may conduct on all the country's limited companies.

The deepest differences concern the reference firms, that is, what each tool treats as a criminal or anomalous firm in order to be built or validated. Ravenda *et al.* (2015) use confiscated firms; Ambrosini *et al.* (2024) firms linked to individuals charged in criminal proceedings; Cariello *et al.* (2024) combine seizures and confiscations with lists derived from press news, from suspicious transaction reports cross-checked against anti-mafia databases and from information requests; De Simoni and Pellegrini (2025) rely on Court of Cassation rulings and suspicious transaction reports. The choice is not neutral, and three of the five documents explicitly acknowledge it. Ambrosini *et al.* observe that drawing on an early stage of criminal proceedings allows for a larger sample but increases the risk of false positives, whereas using seized firms introduces a further selection which, according to the authors, may distort the representation of the phenomenon and limits the sample to firms still operating at the time of the enforcement operation; they also point to undetected criminal connections among the firms classified as legitimate as a problem. Pellegrini *et al.* (2020) warn that the presence of a firm in a suspicious transaction report indicates only a suspicion; Cariello *et al.* note that seized and confiscated firms are only a tiny fraction of the firms linked to organised crime, and that an anti-mafia injunction does not constitute conclusive evidence of infiltration. The use that each tool makes of its reference firms also differs. The indicator of Pellegrini *et al.* is built without them, with indices and weights set in advance, and the reported firms serve only to check, *ex post*, whether the indicator distinguishes them from the others. In Ravenda *et al.* and in De Simoni and Pellegrini, they are used to estimate the weights: the regression determines how much each variable contributes to distinguishing the reference firms from the others. In the machine learning tools, the algorithm learns from part of them and is then tested on the rest.

The information set varies in breadth, source and period. It ranges from the four signals of De Simoni and Pellegrini (2025) and the five indices of Pellegrini *et al.* (2020) to the 19 explanatory variables of Ravenda *et al.* (2015), and up to the 32 variables of Cariello *et al.* (2024) and the sets of Ambrosini *et al.* (2024), which combine several families of variables. Two tools rely on the AIDA database and three on Cerved, and the data cover different periods between 2002 and 2022. Some tools supplement financial statements with sources external to the firm, such as the Central Credit Register (Cariello *et al.*, 2024; De Simoni & Pellegrini, 2025). The abridged form, which Section 2 identified as a source of differences in the variables that can be computed, also enters the documents in different ways. Pellegrini *et al.* design their indicator around it, computing the indices on the items it provides. Cariello *et al.* note that smaller firms may consolidate certain items, particularly receivables and payables. De Simoni and Pellegrini specify that the database they draw on also obtains the composition of receivables and payables from the notes to the financial statements when the statements themselves do not report it. A key finding emerging from the literature review, which is relevant to addressing RQ1, is that financial statements do not operate as a single, uniform information base. What each tool observes depends on the items selected, the database that makes them available, the period and the reporting regime, and the sources with which financial statements are supplemented.

As regards method, the tools use three families of techniques: from the sum of indices with predetermined weights (Pellegrini *et al.*, 2020) to logistic regression, used to estimate a model (Ravenda *et al.*, 2015) or the weights of an indicator (De Simoni & Pellegrini, 2025), and to gradient boosting (Ambrosini *et al.*, 2024; Cariello *et al.*, 2024). Performance is also measured under different conditions. Ravenda *et al.* (2015) report their main results on the estimation sample, in which the financial statement of each confiscated firm is matched, year by year, with that of a legitimate firm, together with an out-of-sample validation in which the share of correct classifications ranges between 69.55% and 77.91%. Cariello *et al.* (2024) test their model on a part of the data not used to train it. This part, however, comes from a purpose-built sample: among all available financial statements, those of firms linked to organised crime account for only 1.4%, but the authors excluded part of the financial statements of legitimate firms so that the two groups are roughly equal in size. Ambrosini *et al.* (2024), by contrast, test their model on firms identified in anti-mafia operations subsequent to those used to build it, in a sample with eleven financial statements of legitimate firms for every financial statement of a criminal firm. De Simoni and Pellegrini (2025) present their results as those of a pilot study, to be validated with further data; Pellegrini *et al.* (2020) compare the distribution of the indicator with the suspicious transaction reports received by the UIF. The measures reported do not coincide: the AUC appears in three of the five documents; sensitivity and specificity depend on the chosen threshold; and precision also depends on the proportion of criminal and legitimate firms in the test sample. The case of the two machine learning tools is telling: Cariello *et al.* and Ambrosini *et al.* use the same algorithm, but test its results on different samples and reference firms and report partly different measures, so the values shown in Table 1 cannot be set side by side to establish which of the two tools performs better. These findings highlight a key point for addressing RQ2: the performance figures reported for the individual tools are not directly comparable and therefore do not provide a basis for ranking the tools.

To the RQ1 – how detection tools referring to Italian firms derive a risk assessment from financial statements – the comparison answers that they do so through three choices: which items to observe, which firms to take as examples of what is sought, and which technique to use to translate the items into an assessment of the firm. With regard to the answer to RQ2 – how the tools differ in the way they derive the risk assessment, and with what consequences for interpreting their results – the analysis shows that the tools differ in each of the choices underlying that assessment. Their results should therefore be interpreted in light of the reference firms, information set, analytical method, and validation conditions specific to each tool, rather than directly compared with those of the others.

7. Discussion

The comparison has shown that the tools derive a risk assessment from financial statements through three choices: which items to observe, which firms to take as examples of what is sought, and which technique to use to translate the items into an assessment of the firm. The first of these choices is affected by a circumstance noted in Section 3: financial statements are prepared by the very party that the tool seeks to classify. The literature on strategic classification shows that those who know the rules of a tool may alter their data to obtain a better outcome, and that the tool's performance may then deteriorate sharply (Hardt *et al.*, 2016). Nor is it enough to keep those rules confidential, since those being assessed still attempt to learn them and game them (Ghalme *et al.*, 2021). An item that currently distinguishes criminal from legitimate firms well may therefore cease to do so once criminal firms learn to alter it: predictive power measured on past data is not, on its own, a sufficient basis for selecting it.

This paper therefore proposes an additional criterion: at equal predictive power, preference should go to variables that the firm to be detected can alter only at a high cost. From this perspective, three groups of data can be distinguished, in decreasing order of cost. The first comprises data that the firm does not produce but that are recorded by third parties, such as debts recorded in the Central Credit Register: the firm cannot alter them through its own reporting choices, but only by changing its relationship with the intermediary. The second comprises variables that the firm can alter only through real operations, such as cash flows from operations; the third, those that it can also alter through mere accounting estimates, with no effect on cash flows, such as changes in receivables, inventories and payables. The distinction between the second and third groups is the one, which Ravenda *et al.* (2015) draw from the earnings management literature and apply to “mafia” firms, between real activities manipulation, which affects cash flows from operations, and accrual-based earnings management, which does not. Classifying variables by the cost of altering them, without measuring that cost, has a precedent in Hardt *et al.* (2016). When applying their

model to real data, these authors did not measure costs either: they assigned each variable to one of three broad categories, according to how costly it was to increase or decrease its value, and acknowledged that the numerical values attached to these categories were largely arbitrary.

Cost, however, must be assessed from the standpoint of the type of firm the tool is intended to detect, rather than that of firms in general. Ravenda *et al.* (2015) observe that real activities manipulation, which may be particularly costly for firms facing intense competition, may not be so for “mafia” firms, which usually face weak market competition and benefit from significant competitive advantages. A variable that is costly for a legitimate firm to alter, because it requires real operations, may therefore be much less costly for a “mafia” firm. Applying the criterion thus requires asking not how costly it is, in general, to alter a variable, but how costly it is for the firm to be detected.

8. Limitations

This study has some limitations. First, the analysis relies solely on published documents; any tools used by authorities or intermediaries but not publicly disclosed therefore fall outside the scope of the comparison. Second, as noted in Section 4, the performance figures are those reported by the authors, and the description of the tools necessarily depends on the information provided in the underlying documents. Finally, the comparison is limited to Italian experiences, and its conclusions have not been tested in other jurisdictions. However, these limitations reflect the study’s reliance on verifiable and publicly available information and its deliberate focus on the Italian context, which provides a consistent basis for the comparison while leaving broader validation and cross-country analysis to future research.

The criterion proposed in Section 7 has not been tested on data: the proposition that variables that are more costly to alter make detection tools more resistant to firms’ adaptation is a reasoned hypothesis rather than an empirical finding. Moreover, the criterion considers only the regular ways in which a firm can alter its data – by changing its relationships with intermediaries, carrying out real operations, or choosing accounting estimates – and does not address the falsification of accounts, such as the recording of fictitious transactions or amounts. Finally, the criterion is of no help when a firm linked to organised crime operates regularly and its financial statements therefore show no anomalies to be detected. This case is not marginal: according to Arellano-Bover *et al.* (2024), when an infiltrated firm is large and already well established, infiltration mostly reflects a pure investment motive, whereby the criminal organisation keeps the firm separate from its illegal activities; this motive accounts for a substantial share of overall infiltration. However, these limitations also point to directions for future research. The proposed criterion could first be tested empirically to assess whether variables that are more costly to alter actually improve the robustness of detection tools over time. Further research could then examine how the criterion might be extended to account for the falsification of financial statements and how financial statement information could be combined with other sources to identify infiltrated firms that exhibit no accounting anomalies.

9. Conclusions

This paper has compared five tools developed in the academic literature and within the UIF that draw on the financial statements of Italian firms to identify those linked to organised crime or involved in tax fraud. In answer to the first research question, the comparison shows that the tools derive a risk assessment from financial statements through three choices: the items to observe, the firms taken as examples of what is sought, and the technique by which the items are translated into an assessment of the firm. In answer to the second, the tools differ in each of these choices, with two consequences. Since the items observed, the sources and the reference periods vary, financial statements do not operate as a single, uniform information base. Since the reference firms, validation conditions and performance measures also vary, the reported performance figures are not comparable, and the results of each tool should be read in the light of how it was built and tested.

The paper also proposes a criterion for selecting variables: at equal predictive power, preference should go to those that the firm to be detected – not just any firm – can alter only at a high cost. The criterion stems from the fact that financial statements are prepared by the very party that the tools seek to classify, and it is a reasoned proposal that has not been tested on data.

References

- [1] Ambrosini, F., Fabrizi, M., & Parbonetti, A. (2024). *Accounting for crime: Detecting Mafia-connected firms using machine learning* [Working paper, revised April 2025]. SSRN. <https://ssrn.com/abstract=4912709>
- [2] Arellano-Bover, J., De Simoni, M., Guiso, L., Macchiavello, R., Marchetti, D. J., & Prem, M. (2024). *Mafias and firms* (Quaderni dell'antiriciclaggio – Analisi e studi No. 24). Unità di Informazione Finanziaria per l'Italia. <https://uif.bancaditalia.it/pubblicazioni/quaderni/2024/quaderno-24-2024/index.html>
- [3] Cariello, P., De Simoni, M., & Iezzi, S. (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 No. 22). Unità di Informazione Finanziaria per l'Italia. <https://uif.bancaditalia.it/pubblicazioni/quaderni/2024/quaderno-22-2024/index.html>
- [4] Castelluccio, M., & Rizzica, L. (2025). *Mafia infiltrations in times of crisis: Evidence from the Covid-19 shock* (Temi di discussione No. 1502). Banca d'Italia. <https://doi.org/10.32057/0.TD.2025.1502>
- [5] Daniele, G., De Simoni, M., Marchetti, D. J., Marcolongo, G., & Pinotti, P. (2025). *A loan you can't refuse: Credit rationing and organized crime infiltration of distressed firms* (Quaderni dell'antiriciclaggio – Analisi e studi No. 32). Unità di Informazione Finanziaria per l'Italia. <https://uif.bancaditalia.it/pubblicazioni/quaderni/2025/quaderno-32-2025/index.html>
- [6] De Simoni, M. (2022). *The financial profile of firms infiltrated by organised crime in Italy* (Quaderni dell'antiriciclaggio – Analisi e studi No. 17). Unità di Informazione Finanziaria per l'Italia. <https://uif.bancaditalia.it/pubblicazioni/quaderni/2022/quaderno-17-2022/index.html>
- [7] De Simoni, M., & Pellegrini, A. (2025). *“Not a dog. Not a wolf. All he knows is what he's not”: Detection indicators for buffer companies involved in complex fiscal frauds* (Quaderni dell'antiriciclaggio – Analisi e studi No. 28). Unità di Informazione Finanziaria per l'Italia. <https://uif.bancaditalia.it/pubblicazioni/quaderni/2025/quaderno-28-2025/index.html>
- [8] Ghalme, G., Nair, V., Eilat, I., Talgam-Cohen, I., & Rosenfeld, N. (2021). Strategic classification in the dark. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th International Conference on Machine Learning* (Vol. 139, pp. 3672–3681). PMLR. <https://proceedings.mlr.press/v139/ghalme21a.html>
- [9] Hardt, M., Megiddo, N., Papadimitriou, C., & Wootters, M. (2016). Strategic classification. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science* (pp. 111–122). Association for Computing Machinery. <https://doi.org/10.1145/2840728.2840730>
- [10] Kleinberg, J., & Raghavan, M. (2020). How do classifiers induce agents to invest effort strategically? *ACM Transactions on Economics and Computation*, 8(4), Article 19. <https://doi.org/10.1145/3417742>
- [11] Pellegrini, A., De Franceschis, P., Bentivogli, C., & Laurenza, E. (2020). *Un indicatore sintetico per individuare le società cosiddette cartiere* [A composite indicator to identify so-called shell companies] (Quaderni dell'antiriciclaggio – Analisi e studi No. 15). Unità di Informazione Finanziaria per l'Italia. <https://uif.bancaditalia.it/pubblicazioni/quaderni/2020/quaderno-15-2020/index.html>
- [12] Ravenda, D., Argilés-Bosch, J. M., & Valencia-Silva, M. M. (2015). Detection model of legally registered Mafia firms in Italy. *European Management Review*, 12(1), 23–39.
- [13] Unità di Informazione Finanziaria per l'Italia. (2023, May 12). *Provvedimento della UIF del 12 maggio 2023* (Measure of the UIF of 12 May 2023 establishing anomaly indicators). https://uif.bancaditalia.it/normativa/norm-indicatori-anomalia/Provvedimento_della_UIF_del_12_maggio_2023_e_allegato.pdf