Potentiality of Benford's law in deriving a firm-level evasion indicator for Italy

Raffaella Coppier[a], Elisabetta Michetti[b], Luisa Scaccia[c]*

[a,b,c] Department of Economics and Law, University of Macerata, Macerata, Italy

Abstract

This paper investigates the potential of Benford's Law (BL) to detect corporate tax evasion, with a focus on Italy, where tax non-compliance remains a persistent economic challenge. BL predicts the expected distribution of leading digits in naturally occurring numerical data, and significant deviations from this pattern may indicate irregularities, including tax evasion. We apply BL to financial statement data for the years 2014–2022. A firm is classified as potentially non-compliant if its financial data deviate from BL in at least one year. This approach yields a firm-level indicator that can support tax audit targeting and serve as a proxy for tax evasion in empirical research. Such a proxy may help address the scarcity of firm-level data and enable the study of how tax evasion affects firm growth and market distortions. To evaluate its validity, we compare the distribution of potentially non-compliant firms — identified via BL — across regions, sectors, and firm sizes with aggregated data from the Italian Ministry of Economy and Finance and the Revenue Agency, including Synthetic Tax Reliability Indices and other official estimates of tax evasion and the shadow economy. The findings reveal a strong alignment between BL-based results and official indicators, highlighting the potential of BL as a cost-effective, data-driven tool for identifying and studying tax evasion at the firm level.

JEL Code: C61, C73, E32, E37, H26

Keywords: Tax evasion, targeting firms' auditing, Benford's law, balance sheets and income statements.

1 Introduction

Tax evasion represents a significant challenge in Italy, with both individuals and corporations engaging in practices aimed at under-reporting income and avoiding tax obligations.

*Corresponding author: Luisa Scaccia, Department of Economics and Law, University of Macerata, via Crescimbeni 14, 62100 Macerata (Italy). Phone number (+39)0733-2583245, fax number (+39)0733-2583205, e-mail: luisa.scaccia@unimc.it.

Despite efforts to combat tax evasion through legislative reforms and enhanced enforcement measures, it remains a pervasive issue that undermines the integrity of the tax system and deprives the government of much-needed revenue. In particular, corporate tax evasion poses a substantial threat to the Italian economy. In 2021, the value of the underground economy deriving from legal activities nearly reached 174 billion euros[1](ISTAT, 2023). More than 91 billion came from under-reporting of added value, which saw an increase of 11.7 billion euros (equivalent to 14.6%) compared to 2020, with an impact on GDP that has increased from 4.8% to 5.0%, returning to pre-Covid levels.

Addressing tax evasion and encouraging tax compliance are therefore primary objectives of economic policy. The literature on tax compliance has investigated different drivers of taxpayers' behaviour. On the one hand, the "classical" literature pioneered by Allingham and Sandmo (1972) identifies tools to combat evasion by enforcement; on the other hand, a more recent strand of literature identifies two other motivations that explain taxpayers' compliance: tax morale and social norms (Traxler and Spichtig, 2011; Di Gioacchino and Fichera, 2020) focusing on both individual factors and group considerations. Regarding the enforcement, the principal tool available to tax authorities to enforce compliance is firms' auditing. However, visiting or summoning the taxpayer to verify the information reported, the invoices, and other related documents, is quite a costly operation. For this reason, in Italy (CGIA-MESTRE, 2019) as well as in many other countries, in order to increase the efficiency of tax audits, governments segment taxpayers by firm size, so that large firms are subject to higher audit probabilities from tax administrations than small firms (Bachas et al., 2019). Such categorisation could, though, lead to an industrial structure in the long term, characterised by a few large firms and numerous small firms, with these last ones being marked by inefficient resource allocation and non-compliance with tax regulations. Firms may in fact intentionally opt to remain small in order to avoid stricter control (Coppier et al., 2022, 2023, 2024; Almunia and Lopez-Rodriguez, 2018; Boonzaaier et al., 2019; Gourio and Roys, 2014; Guner et al., 2008; López and Torres, 2020; Onji, 2009; Ramaswamy, 2021), thereby further exacerbating the impact of tax evasion on industry concentration (Barros and Martins, 2025; Martin et al., 2026; Benkraiem et al., 2024). These considerations underscore the need to maximise efficiency and minimise costs through other approaches than stratification, such as a strategical targeting of corporate auditing operations.

The Italian Revenue Agency (Agenzia delle Entrate) already employs a range of risk assessment tools to guide the selection of firms for tax audits. These tools include sector-specific indicators of economic consistency, anomaly detection systems, and data-driven models based on financial, transactional, and behavioural variables. Despite the sophistication of these existing instruments, it remains of great interest to investigate the potential of alternative or complementary methods.

[1]The underground economy from legal activities mainly refers to the hidden value added through intentionally incorrect reporting of revenue and/or costs (under-reporting of value added) or generated through the use of irregular labour.

This paper contributes to the debate on this topic by employing the Benford's law, hereafter BL, (Benford, 1938) as a possible tool to better direct corporate auditing operations. This law establishes the frequency distribution of leading digits in a collection of numbers from many real-life situations, and departures from this theoretical distribution may identify possible data manipulation. It has been shown that this result applies to a wide variety of data sets and prior work has utilised BL to identify accounting fraud and data manipulation in corporate firms (e.g. Alali and Romero, 2013; Durtschi et al., 2004; Cunjak Mataković, 2019; Jianu and Jianu, 2021), to assess taxpayers compliance (e.g. Nigrini, 1996), to verify macroeconomic data quality (Rauch et al., 2015; Moul and Nye, 2015). A more recent application for detecting accounting data manipulation can be found in Grammatikos and Papanikolaou (2021). Cerqueti et al. (2024) investigate, instead, the informative content of BL for risk assessment, while Parnes (2022) shows that most banks' off-balance sheet items have partial conformity to BL in their first leading digits but scarce compliance in their second leading digits. Despite its proven usefulness, BL has not yet been applied to detect evasion in the Italian context. Only Arezzo and Cerqueti (2023) considered it in the somehow related issue of the analysis of labour inspections' reliability, while Mir et al. (2014) and Ausloos et al. (2017) provided an application on the yearly aggregated tax income data of all Italian municipalities and regions.

We propose using Benford's Law (BL) on publicly available balance sheet and income statement data to obtain a potential indicator of tax non-compliance at the firm level. We extract these data from the AIDA[2] BvD database, the Italian subset of the ORBIS database, which collects financial, biographical, and sectoral information on Italian firms. For each sampled firm, we compute the empirical distribution of the leading digit (i.e., the first digit from the left) for all accounting figures in the balance sheet and income statement for a given year. We then calculate a discrepancy measure between the observed distribution and the theoretical one predicted by BL. A large discrepancy is interpreted as a possible signal of tax non-compliance, potentially flagging the firm for further auditing.

Evaluating the accuracy of the proposed procedure in detecting tax-evading firms at the individual level is inherently challenging, given the sensitive nature and limited public availability of firm-level tax compliance data. To overcome this limitation, we aggregate the firm-level BL-based proxy to obtain the estimated share of potentially non-compliant firms across different geographical areas, economic sectors, and firm sizes. We then assess the validity of the proxy by verifying whether these aggregated distributions align with well-established stylized facts about tax evasion in Italy, such as its higher prevalence in certain regions, industries, and among smaller firms.

The aim is to provide a reliable tool that can help public authorities in efficiently ad-

[2]The acronym AIDA stands for "Analisi Informatizzata delle Aziende", which means "Computerized Analysis of Companies". It is a firm-level database provided by Bureau van Dijk (Moody's). It is widely used for financial analysis and benchmarking of Italian firms and serves as an essential resource for academic research and market analysis in Italy. The database is accessible through Bureau van Dijk's ORBIS platform: https://www.bvdinfo.com.

dressing auditing activities and contrasting tax evasion. In addition, a firm-level proxy of tax compliance could also be employed within applied economic analysis. Researchers often struggle to empirically test theories related to the market distortions caused by tax evasion or to understand its repercussions on law-abiding firms, largely because of the lack of granular data on corporate tax behaviour. The proxy based on BL data may offer a potential solution by serving as a measure of firms' likelihood to evade taxes. This tool can facilitate investigations into the broader economic consequences of tax evasion, including its influence on competitive balance, efficiency in resource distribution, and firm productivity.

The paper is organized as follows. Section 2 briefly introduces BL, describes the sampling strategy used to draw a sample of firms and details the methodology developed to obtain a firm-level proxy for tax evasion and to assess its reliability in identifying non-compliant firms. Section 3 presents the results and evaluates the performance of the proposed approach in terms of its ability to replicate well-known macro-level characteristics of firms involved in tax evasion in Italy. Section 4 discusses the findings, draws conclusions, and outlines possible directions for future research.

2 Overview of Benford's Law, data sampling and analytical approach

After a brief introduction to BL, this section outlines the procedure used to obtain a representative sample of Italian firms and details the methodological tools employed in the analysis to derive a proxy variable for firm-level tax compliance and to assess its performance.

2.1 Benford's Law: core properties and prerequisites for application

BL describes the frequency distribution of leading digits in empirical datasets, across numerous natural and human phenomena. Its origins can be traced back to Newcomb (1881) and, independently, to Benford (1938) and it specifically states that the leading digit d in numbers from real-world data is not uniformly distributed, as might be intuitively expected, but instead adheres to a logarithmic law given by:

$$p_d = \log_{10}\left(1 + \frac{1}{d}\right), \qquad d = 1, 2, \ldots, 9, \tag{1}$$

where p_d is the probability that a number from the dataset has its first digit equal to d. Other versions of BL exist beyond the one expressed in Equation (1), focusing on the h-th digit, for $h > 1$, or combinations of digits (Hill, 1995a; Nigrini, 2012).

BL displays several distinctive properties that explain its wide empirical applicability. One key property is scale invariance (Pinkham, 1961): multiplying every number in a dataset by a positive constant (for example converting dollars to euros) does not change the distribution of leading digits. In addition, BL is base-invariant (Hill, 1995b), meaning that a dataset that is Benford in base 10 will follow the BL when its numbers are represented in binary, hexadecimal or any other base — the numerical probabilities are given by the same logarithmic formula in Eq. (1), adapted to that base. Mathematical results show that invariance under scaling or under change of numeric base singles out the logarithmic (Benford) distribution, which is why BL is so robust across many naturally occurring datasets. In addition, Hill (1995c) shows that Benford behaviour naturally emerges when observations are drawn from a compound population — in other words, when the underlying data-generating distribution itself is sampled from an ensemble of distributions. Under fairly general mixing conditions, the idiosyncratic digit-patterns of individual parent distributions are washed out, and the distribution of leading digits converges to the logarithmic law described by Benford. This result gives a further theoretical basis for the empirical ubiquity of BL and explains why mixtures of heterogeneous processes often display the characteristic skew toward small leading digits.

A wide range of empirical datasets, drawn from disciplines as varied as economics, physics, and demographics, have been shown to follow Benford's law (Nye and Moul, 2007; Moul and Nye, 2015; Mir, 2014; Sambridge et al., 2010; Ausloos et al., 2015; Sandron and Hayford, 2002), with more comprehensive surveys available in Hill and Berger (2009). Despite this broad evidence, there is still no universally accepted rule that determines in advance which datasets should adhere to the law. The search for necessary and sufficient conditions remains an open problem in the literature. Still, certain empirical features are commonly observed in datasets that do exhibit Benford behaviour: (i) they are large enough for digit frequencies to stabilize, (ii) their values extend across multiple orders of magnitude, and (iii) they are not subject to artificial bounds such as fixed lower or upper limits imposed by human conventions. When these characteristics are present, conformity with the logarithmic digit distribution becomes substantially more likely. In addition Cerqueti and Maggi (2025) have recently introduced generalized extensions of BL that are less restrictive than the original one, thereby increasing the likelihood that empirical datasets conform to the law.

2.2 Sampling procedure and dataset creation

The present study makes use of a sample of Italian firms drawn from the database AIDA, which compiles detailed information on Italian companies, including their financial performance, balance sheets, and other relevant economic data. In particular, we utilize a stratified random sampling approach, with strata defined by the 20 Italian regions and the 21 ATECO sectors of activity[3], resulting in a total of 420 strata. We apply a proportional allocation strategy, with a sampling fraction of 10%, a value chosen as a reasonable compro-

[3]The list of the economic activities according to ATECO 2007 classification can be found at the following link:
https://www.istat.it/it/files/2022/03/Struttura-ATECO-2007-aggiornamento-2022.xlsx

mise between ensuring a sufficiently large sample that adequately represents each stratum and keeping the computational burden manageable. In practice, to obtain the sample, we perform the following steps.

1. Search in AIDA all the Italian industrial, commercial, and service companies, whose legal status is classified as active, established before 2014 and with available financial statements for the years 2014 to 2022. These last two related constraints are imposed to allow for the creation of a compliance proxy based on multiple years of observations, as explained in Section 2.3. This search, conducted on 05/11/2024, resulted in a population of 539,024 companies[4].

2. For each firm in the previously defined population, download from AIDA data on its tax identification number, the region where its legal headquarters is located, and the six-digit ATECO code for its economic activity.

3. Import data into the statistical software R, remove firms with missing values for any of the considered variables, recode the economic activity into 21 sectors, and finally use the `dplyr` package (Wickham et al., 2023) to draw a sample stratified by region and sector.

4. Import the tax identification numbers of the sampled firms into AIDA, and for these firms, download balance sheets for the years 2014 to 2022.

The final dataset obtained in this way consists of 53,898 firms and 2,190 variable columns.

2.3 Construction of the firm-level potential evasion indicator

Separately for each firm in the sample and for each year, we consider all the items in the financial statements, and verify conformity with first-digit BL. Although BL can be applied to different digit configurations — such as the second digit, the first-two digits, or more complex joint digit tests — empirical applications using these extensions remain relatively limited. For instance, Diekmann (2007), Shikano and Mack (2011), and Nigrini (2012) show that higher-digit tests may offer enhanced sensitivity by uncovering patterns that might escape the basic first-digit analysis. The first-two digit test is, instead, more focused and primarily detects anomalous duplication of digits and possible biases in the data (Aris et al., 2017). The last-two digit test helps identify rounding patterns, which could suggest manipulative behaviour, such as rounding up earnings to meet targets (Das and Zhang, 2003; Tran et al., 2023).

In our setting, however, the use of digit configurations beyond the first digit is constrained by the structure of firm-level accounting data. In fact, applying tests based on the

[4]Note that in AIDA, data on firms' financial statements are only available for the past ten years. By choosing 2014 as the starting year, we consider the maximum time span for which data are available. The year 2023 is instead excluded, as financial statements for only a very small percentage of firms were available in AIDA at the time of data extraction.

second digit or higher-order digit combinations would substantially reduce the number of usable items: single-digit values must be discarded when analysing second digits, two-digit values must be excluded when analysing the third digit, and so forth. Given that the number of accounting items available for each firm is already modest, this loss of information would severely compromise the reliability of the diagnostic indicators. Nevertheless, we performed a pilot analysis for a single year using second and higher-order digit tests, which confirmed their limited ability to identify potentially non-compliant firms in our context. Moreover, joint-digit tests (e.g., first-two digits) proved infeasible due to the insufficient number of valid items per firm. For these reasons, and in line with the extensive empirical literature relying primarily on the first digit law, we base our analysis on the distribution of first digits, which maximises data availability and provides the most robust and informative indicator of potential non-compliance at the firm level in our setting.

To evaluate the conformity of financial statement data with BL, the use of an appropriate discrepancy measure is essential. Commonly used measures include the Chi-squared test, Kolmogorov-Smirnov (KS) test, Mean Absolute Deviation (MAD), and Euclidean Distance (ED). Each has distinct strengths and limitations. The Chi-squared test assesses conformity by comparing observed and expected frequencies, but its sensitivity to sample size can lead to false positives in large datasets. The KS test focuses on the maximum difference between cumulative distributions, making it suitable for detecting deviations, but it may lack precision for small datasets. MAD, a straightforward metric based on the average absolute differences between observed and expected frequencies, is easy to compute and interpret, but it does not account for variations across different digits. ED, which measures the root sum of squared differences, provides a comprehensive view of overall deviation but, like MAD, may obscure specific patterns in leading digits.

The choice of discrepancy measure depends on factors like dataset size and the specific characteristics of the data under analysis. As, in the present case, we are considering individual financial statements, generally characterized by a limited number of items[5], we can safely use the Chi-squared statistic without worrying about the sensitivity issues typically associated with large datasets (Cerqueti and Maggi, 2021; Kossovsky, 2021). Therefore, for each firm $j = 1, 2, \ldots, J$ and each year $t = 1, 2, \ldots, T$, we compute

$$\chi^2_{jt} = \sum_{d=1}^{9} \frac{(n_{jtd} - n_{jt}p_d)^2}{n_{jt}p_d},$$

where n_{jtd} is the observed number of times the digit d appears as the first digit within the financial statement values of firm j in the year t, n_{jt} is the number of items in the financial statement of firm j in the year t, and p_d is the probability that the first digit is d according to Equation (1).

[5]Each financial statement contains 243 line items (including sub-items and totals), of which 163 come from the balance sheet and 80 from the income statement. However, most of these items may not be applicable for a given firm, thus substantially reducing the number of items used in assessing conformity with BL.

We calculate the Chi-squared statistic for all firms with at least 30 non-missing items in their financial statements for each individual year. This approach avoids computing a discrepancy measure based on very small absolute frequencies. Additionally, instead of relying on tabulated values of the Chi-squared statistic, we use simulated p-values[6]. In fact, the asymptotic distribution of the Chi-squared statistic is valid only when the expected theoretical frequencies are sufficiently large, typically exceeding a minimum threshold, often set at 5. Given that the theoretical probability of observing the digit 9 is approximately 0.046, at least 110 items in each financial statement would be necessary for the asymptotic distribution of the Chi-squared statistic to be reliable. Since this condition is often not met, we prefer to use simulated p-values rather than asymptotic ones. The scarcity of non-missing items also hinders the use of BL versions for digit combinations[7], as previously discussed.

Since discrepancy measures are known to have low power when the number of items used for their computation is small, we choose to classify as potentially non-compliant with tax payment all firms that do not conform to BL at the 5% level of significance in at least one of the considered years. Conversely, firms that conform to BL in all the considered years are classified as potentially tax compliant, while those for which the Chi-squared statistic cannot be calculated in at least one year (due to fewer than 30 items in the financial statement for that year) are simply considered unclassified and excluded from further analysis. We note that while pooling items from financial statements of subsequent years may bias the BL conformity assessment due to autocorrelation of items over time, the Chi-square statistics computed for individual years should not suffer from this issue. Any correlation of these statistics over time would simply be attributed to either manipulation or non-manipulation of statements over time. In fact, items that typically correspond to the largest figures, such as total assets, revenues, and equity, are likely to exhibit the same first significant digit in subsequent years due to their size and the general persistence of financial trends. If financial statements from subsequent years are pooled, the first digit of these items will appear much more frequently compared to the first digit of smaller items, where the first digit is more likely to change from one year to the next. This could lead to a significant departure from the theoretical BL distribution, even if no manipulation has occurred. In contrast, the Chi-square statistics computed for each individual year are much less affected by the autocorrelation of large items over time.

[6]Simulated p-values are obtained by generating a large number of random samples under the null hypothesis (in our case under conformity with the BL, i.e. using the probability distribution in (1) to draw the samples) and calculating the test statistic for each sample. The p-value is then estimated as the proportion of simulated test statistic values that are equal to or more extreme than the observed test statistic on the original sample of data.

[7]Pooling the items from financial statements of subsequent years to increase the amount of data and thereby enable the use of BL versions for digit combinations is not advisable. Temporal autocorrelation of items may cause a significant divergence between the observed and theoretical digit distributions, even in the absence of financial statement manipulation.

2.4 Assessment of the indicator's performance

Once a proxy variable for firm-level tax compliance has been derived, as described in Section 2.3, it is essential to assess its ability to distinguish between compliant and non-compliant firms.

As a preliminary step, we evaluate whether BL exhibits sufficient statistical power in our setting. Specifically, for each year we test whether the share of firms classified as non-conforming exceeds the nominal significance level of the underlying digit-distribution test. This is done through a one-sample proportion z-test, comparing the observed proportion of non-conforming firms with the 5% threshold implied by the null hypothesis of full conformity with BL. Evidence that the observed proportion is significantly higher than 5% indicates that the BL-based procedure can detect deviations from conformity with a probability exceeding the nominal type-I error rate, providing an initial validation of its diagnostic power at the firm level.

A second preliminary check concerns the stability of the Benford-based indicator with respect to the number of accounting items available for each firm. Since financial statements differ in detail across firms — depending on size, legal form, and reporting requirements — the set of numerical items used in the Benford analysis is not perfectly uniform. This heterogeneity may in principle affect the results and the power of the conformity test: firms with very few available items might exhibit higher sampling variability and lower test power, while firms with more detailed accounts might yield more stable estimates but potentially excessive sensitivity. To assess the extent of this issue, we examine the distribution of firms according to the number of usable items and verify whether it varies systematically across regions, sectors, or firm size, as such patterns could distort aggregate comparisons. In addition, we check whether the proportion of financial statements classified as conform changes with the number of usable items, ensuring that differences in data availability do not mechanically drive the classification. These diagnostics are meant to ensure that heterogeneity in item availability does not systematically bias the resulting non-compliance indicator.

After verifying both the statistical power of the BL-based procedure and its robustness to variation in the number of accounting items per firm, the next step is to assess its capacity to discriminate between compliant and non-compliant firms. If firm-level data on tax compliance were available, this assessment could be carried out by constructing a confusion matrix and computing standard performance metrics, such as the share of correctly classified units. However, no firm-level data on actual compliance exist. To address this limitation and evaluate the performance of the compliance indicator, we examine whether it can reproduce — at an aggregate level — well-established stylized facts regarding the characteristics of firms involved in tax evasion in Italy.

To this end, we compute the percentage of potentially non-compliant firms, according to the BL-based indicator, at different aggregation levels: geographic (macro-areas, regions, and provinces), firm size, and economic activity (sectors and macro-sectors). We then compare these percentages with official aggregate indicators of tax evasion or of the shadow economy. Specifically, we consider statistics such as the relative tax gap[8], the non-observed economy (NOE[9]), both estimated by the Italian Ministry of Economy and Finance (MEF), the estimated share of evaded amounts produced by CGIA-Mestre, and the Synthetic Tax Reliability Indices[10] (ISA), developed jointly by MEF and the Italian Revenue Agency. The ISA indices are particularly relevant for our purposes, as they yield a measure of fiscal reliability that is directly comparable to our BL-based proxy. Using a statistical–economic methodology, ISA exploit information over multiple tax periods to produce a summary score that assesses the normality and internal consistency of taxpayers' business practices and identify "reliable" taxpayers (typically, those with a score of at least 8 on a 0–10 scale), who benefit from substantial reward mechanisms. ISA apply to almost all economic activities, with very few exceptions[11], and data on reliable and non-reliable taxpayers are available by legal form (natural persons, partnerships, corporations), by territory, and by economic activity.

When estimating the percentage of potentially non-compliant firms at an aggregate level, some methodological clarification is required. After restricting the sample to firms for which BL conformity can be meaningfully assessed (i.e., firms with at least 30 usable items), the original equal sampling fraction across region–sector strata no longer holds. Because the number of classifiable firms in each population stratum is unknown, the effective sampling fraction becomes heterogeneous across strata, and strict self-weighting with respect to the original population cannot be preserved. Nevertheless, representativeness is maintained with respect to the reduced target population of classifiable firms. The exclusion

[8]The tax gap represents the difference between the total tax and contribution revenue that should be collected under full compliance with tax laws and the amount actually collected, reflecting losses due to non-compliance, evasion, or errors. The relative tax gap is calculated as the ratio between the tax gap and potential revenue.

[9]The main components of the NOE are represented by the hidden economy and the illegal economy; statistical hidden economy and informal economy complete its spectrum. The hidden economy includes all those activities that are voluntarily concealed from tax, social security, and statistical authorities. It is generated by incorrect reporting of turnover and/or costs of production units (in order to generate under-reporting of value-added) and the labour input used in the production processes (i.e., the use of irregular labour). The illegal economy is defined by the set of productive activities involving illegal goods and services, or activities related to legal goods and services but carried out without the proper authorization or title. The statistical hidden economy includes all those activities that escape direct observation due to informational inefficiencies (sampling and non-sampling errors, coverage errors in archives, etc.). Finally, the informal economy includes productive activities carried out in poorly or unorganized contexts, based on labour relations defined within personal or family relationships and not regulated by formal contracts.

[10]Data downloadable at
https://www1.finanze.gov.it/finanze/pagina_dichiarazioni/public/studisettore.php

[11]They do not apply to certain types of activities, such as: non-commercial entities or entities engaged in non-profit activities; public administrations or government-related organizations; businesses operating under special tax regimes, such as agricultural activities subject to lump-sum taxation or flat-rate schemes; companies in their initial year of activity or those that have permanently ceased operations during the year; entities with limited or exceptional accounting situations, such as those in liquidation or bankruptcy procedures; specific professional categories or activities explicitly excluded by the relevant tax authorities' regulations.

of firms with fewer than 30 items is driven solely by measurement feasibility and does not involve any post-sampling selection on the compliance outcome. As a result, within each region–sector stratum, the remaining firms still constitute an (approximately) unweighted simple random subsample of the classifiable population in that stratum. For this reason, unweighted estimators remain appropriate for any aggregation aligned with the stratification variables — regions and sectors — and for higher-level groupings such as macro-areas or macro-sectors. For aggregations that do not coincide with the stratification structure — such as provinces (which are nested within regions) or firm-size classes (which cut across strata) — greater caution is required. These correspond to domain estimates in a stratified design.

While the estimators remain unbiased for the domain parameters, their precision may differ across domains, and they do not benefit from the automatic variance reduction typical of estimators aligned with the stratification scheme. For this reason, it is essential to report sample sizes and precision measures for each domain. In particular, we report 95% confidence intervals for all estimated proportions. Given that the effective sampling fractions for the classifiable population cannot be recovered and may vary across strata, variance estimation proceeds under the standard large-population (infinite-population) approximation, which omits the finite population correction. This approach is conservative when the true sampling fraction is moderate, and statistically appropriate under incomplete information on the size of the classifiable populations.

The comparison between the percentages of potentially non-compliant firms according to the BL-based indicator and those derived from official statistics is carried out on both a graphical and tabular basis. In addition, we compute Pearson and Spearman correlation coefficients to assess the strength and monotonicity of the association between the BL-based estimates and official aggregate indicators. For the Pearson correlation, statistical significance is evaluated using the standard t-test under the null hypothesis of zero correlation. For the Spearman correlation, significance is assessed using exact permutation-based methods. All correlations and associated significance tests are computed using the `cor.test()` function in R. This combined approach provides a robust assessment of the indicator's ability to reproduce well-established stylized facts about tax evasion patterns across different aggregation levels.

3 Results

This section presents the main empirical results. We begin with a descriptive overview of the sample data and then assess whether the proposed indicator replicates well-known stylized facts on tax evasion — across geography, firm size, and sectors — as documented by official aggregate measures.

3.1 Descriptive statistics

The methodology described in Section 2.3, applied on the sample of 53,898 firms, allows to classify 16,692 of them as potentially tax compliant, 24,830 as potentially tax non-compliant, while 12,376 cannot be classified. The percentage of non-compliant firms, among those that can be classified, is 59.8% (see Table 1). While this may seem high, it is not surprising when considering that the relative tax gap for Corporate Income Tax (IRES), Regional Tax on Productive Activities (IRAP), and Value-Added Tax (VAT) amount to 18.8%, 15.9%, and 13.6%, respectively, in 2021. The gap for Personal Income Tax (IRPEF) from self-employment and business activities is even higher, reaching 66.8% in the same year (MEF, 2024). An even more interesting comparison can be done using ISA for the year 2022. According to the ISA, 63.5% of Italian corporations are classified as having unreliable tax returns[12]. Certain economic sectors show particularly high rates of unreliability: 80.7% for laundries, 79.7% for car rentals, and 75.2% for restaurants. Even the most compliant sectors, such as pharmacies or manufacture of paper and cardboard products, still show 34.6% and 45.1% of unreliable tax returns, respectively. The overall percentage of 63.5% is used in our approach to calibrate the definition of the proxy variable for non-compliance. Specifically, the number of years considered is adjusted to counterbalance the lack of power due to data scarcity and to ensure that the percentage of potentially non-compliant firms aligns with the share of corporations with unreliable tax returns as reported by the ISA. To this end, we used the maximum number of years for which financial statement data were available in AIDA. This results in a percentage of potentially non-compliant firms that is slightly lower than the corresponding ISA figure.

Table 2 highlights the importance of selecting an appropriate time frame for defining compliant and non-compliant firms. The data show that the percentage of potentially non-compliant firms obviously increases as the number of years of observed financial statements grows. However, this increase occurs at a diminishing rate. Notably, when financial statements spanning nine consecutive years are considered, the percentage of potentially non-compliant firms converges closely to the proportion of unreliable tax returns identified by the ISA.

Table 3 shows, instead, the percentage of firms whose financial statements do not conform to BL, separately for each year. It can be observed that this percentage remains quite stable over time, and for each year, it is significantly higher than the 5% value, expected under the null hypothesis that no statements were manipulated. As discussed in Section 2.4, this provides evidence that the BL-based procedure has some power to detect anomalous

[12] In considering ISA data, we restrict our analysis to corporations, as these represent the vast majority of firms included in the AIDA database. AIDA provides detailed financial statement information primarily for limited liability companies (such as SRLs and SPAs), which are legally required to file complete annual accounts. In contrast, partnerships are less consistently represented, often reporting only limited data, while sole proprietorships and natural persons are largely absent from the database unless they operate as registered businesses. As a result, focusing on corporations ensures consistency in data coverage and reliability for our firm-level analysis.

N. of years of non-conformity	N. of firms	% of Firms
0	**16692**	**40.20%**
1	15543	37.43%
2	6294	15.16%
3	2053	4.94%
4	681	1.64%
5	181	0.44%
6	54	0.13%
7	17	0.04%
8	4	0.01%
9	3	0.01%
Subtotal	**24830**	**59.80%**
Total	**41522**	**100.0%**

Table 1: Distribution of firms by the number of years of non-conformity with BL. The bolded values in the first row indicate the number and percentage of firms classified as potentially tax compliant, while the bolded values in the subtotal row refer to those classified as potentially non-compliant. The total number of firms is reported in the last row. A total of 12,376 unclassified firms are excluded from the analysis.

Years considered	Non-compliant	Compliant	% of non-compliant
2022	4322	40332	9.68%
2021 to 2022	7866	35392	18.18%
2020 to 2022	11180	31298	26.32%
2019 to 2022	13951	27930	33.31%
2018 to 2022	16405	25057	39.57%
2017 to 2022	18673	22523	45.33%
2016 to 2022	20730	20376	50.43%
2015 to 2022	22890	18469	55.34%
2014 to 2022	24830	16692	59.80%

Table 2: Compliant and non-compliant firms, based on conformity to BL, as the number of considered years increases. Absolute values and percentages of non-compliance.

Year	Non-conform to BL	Conform to BL	% of non-conformity
2014	4412	39833	9.97%***
2015	4400	40316	9.84%***
2016	4195	38903	9.73%***
2017	4196	38983	9.71%***
2018	4126	39107	9.54%***
2019	4146	39406	9.52%***
2020	4381	39677	9.95%***
2021	4243	40076	9.57%***
2022	4322	40332	9.68%***

Table 3: Conformity and non-conformity to BL by year: absolute values and percentages. The last column shows the significance level for the proportion test, which compares the observed non-conformity percentage with the one expected in the absence of statement manipulation. Significance levels are coded as follows: *** p-value< 0.001, ** p-value< 0.01, * p-value< 0.05, + p-value< 0.1.

N. of items	N. of firms (2014)	% of non-conformity (2014)
0-29	8318	—
30-49	22896	11.0%
50-74	18918	8.9%
75-99	2246	9.0%
100-130	185	7.0%

Table 4: Distribution of firms by number of items in 2014 financial statements and corresponding BL conformity rates within each item-number class.

financial statements.

Finally, we examine the heterogeneity in the number of items in each firm's financial statements and its potential impact on the BL conformity test. Since the number of items per firm varies only slightly from year to year, we present here the results for 2014 only. Results for other years are not reported but are very similar, as the number of items remains largely stable for each firm over time. It can be observed that about 15% of firms have fewer than 30 items in their financial statements, making it impossible to assess conformity with BL for them. In addition, the vast majority of firms (nearly 95%) have fewer than 75 items, and none exceeds 130 items. Interestingly, financial statements with fewer than 50 items exhibit a higher percentage of non-conformity with BL. This finding is counter-intuitive, as larger numbers of items are generally expected to increase the power of the test, making it easier to reject the null hypothesis of conformity. The observed pattern is not due to the asymptotic approximation of the chi-square test, since we rely on simulated p-values. Rather, we attribute this result mainly to firm size: smaller firms typically have fewer items in their financial statements and are also more likely to engage in tax non-compliance (see Section 3.3). This relationship explains why higher non-conformity is observed among firms with fewer items.

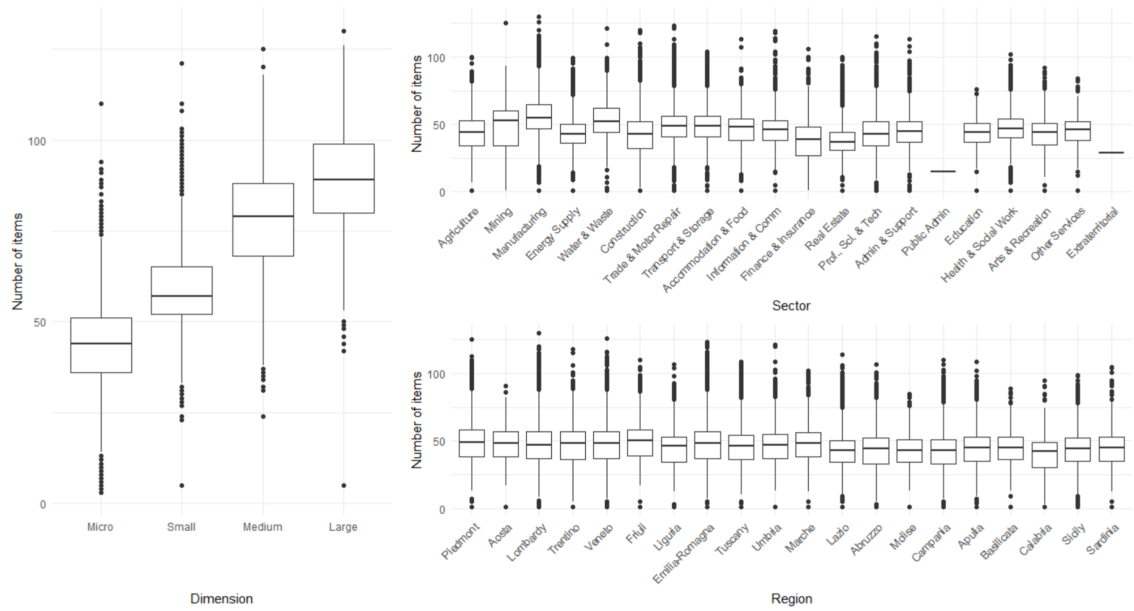

Figure 1: Distribution of firms by number of items in 2014 financial statements by firm's size (left panel), sector of economic activity (top-right panel) and geographical location (bottom-rigth panel).

The limited number of items in the financial statements of small firms is evident in Figure 1 (left panel), which displays the distribution of firms by item count across size classes. As expected, the number of items increases with firm size, with median values of 44, 57, 79, and 89 for micro, small, medium, and large firms, respectively. More similar distributions are observed across the various sectors used in the sample stratification (Figure 1, top-right panel). Slightly higher item counts are found in the Manufacturing and Water & Waste sectors, whereas firms operating in Finance & Insurance and Real Estate tend to report fewer items in their financial statements. These differences reflect the underlying complexity of sector-specific accounting structures: manufacturing and utility firms typically manage a broader set of operational and cost components, while financial and real-estate firms use more compact reporting schemes with fewer line items. The distribution of firms by number of items across regions (Figure 1, bottom-right panel) is instead quite homogeneous, reflecting the fact that each region contains a broadly similar mix of sectors and firm sizes. In any case, as previously discussed, heterogeneity in the number of items is not expected to drive BL conformity by itself. The observed variability in item counts is in fact quite limited: even the most detailed financial statements contain at most around 130 usable items, meaning that the difference between the smallest admissible statements (30 items) and the largest ones is on the order of one hundred items. Such magnitudes are far from those at which the chi-square test becomes overly sensitive to negligible deviations from Benford's distribution. Consequently, modest differences in item availability are unlikely to

Zone	BL % of non compliance	ISA % of unreliability	Relative VAT gap	Non-observed economy	% of evaded amount
North-West	59.1% (58.2%-59.9%)	61.6%	17.3%	9.2%	10.3%
North-East	57.5% (56.5%-58.5%)	60.8%	17.2%	9.7%	11.1%
Centre	60.5% (59.5%-61.4%)	66.4%	18.1%	12.3%	13.6%
South	62.5% (61.4%-63.7%)	64.5%	24.6%	17.2%†	19.0%
Islands	62.6% (60.7%-64.5%)	65.7%	22.9%		19.0%

Table 5: Left side: percentage of potentially non-compliant firms according to BL (95% confidence intervals are provided in brackets) and percentage of unreliable firms according to ISA, by geographic partition. Right side: relative VAT gap, NOE as a % of regional value added (MEF, 2024), and amount evaded for every 100 euros of revenue collected (CGIA-MESTRE, 2023). † this value refers to South and Islands as a whole.

inflate the power of the test or to penalize firms with more articulated financial statements. Any observed association between the number of items and BL conformity arises from other variables that simultaneously influence both the number of items and the likelihood of non-compliance.

3.2 Firm's geographical location and tax evasion

In Italy, tax evasion is well known to follow a clear geographical pattern, with higher levels generally observed in the southern regions compared to the northern ones. This disparity is attributed to a combination of factors, including economic differences, cultural norms, and variations in the effectiveness of tax enforcement. While the northern regions, with their more developed economies and stronger fiscal capacity, are not immune to the issue, evasion tends to be concentrated among small firms. In contrast, in the southern regions, tax evasion is more widespread and systemic.

Table 5 reports, for each geographical partition, the percentage of potentially non-compliant firms identified using BL, alongside the percentage of unreliable tax returns according to ISA. Different indicators, often used to proxy tax evasion, are also provided, such as the relative VAT gap and the percentage of the NOE estimated by MEF (2024) for the year 2021, and the percentage of evaded amounts estimated by CGIA-MESTRE (2023) for the year 2020. Despite the differing nature of all these measures and the fact that they are not directly comparable, as they reflect various facets of a complex phenomenon, they exhibit a similar spatial pattern. Such a pattern is correctly recovered by the proposed BL methodology, suggesting its potential to distinguish between compliant and non-compliant firms and to capture the geographical distribution of tax evasion among firms. This is characterized by a prevalence in the South and Islands, which exhibit high and similar values across all the indicators considered, followed by the Center. In contrast, less vicious behaviour is observed in the North-East and North-West, with only very slight differences between the two areas. Interestingly, according to ISA, the percentage of unreliability of the Centre is absolutely comparable with the one observed in the South and the Islands, and even slightly larger.

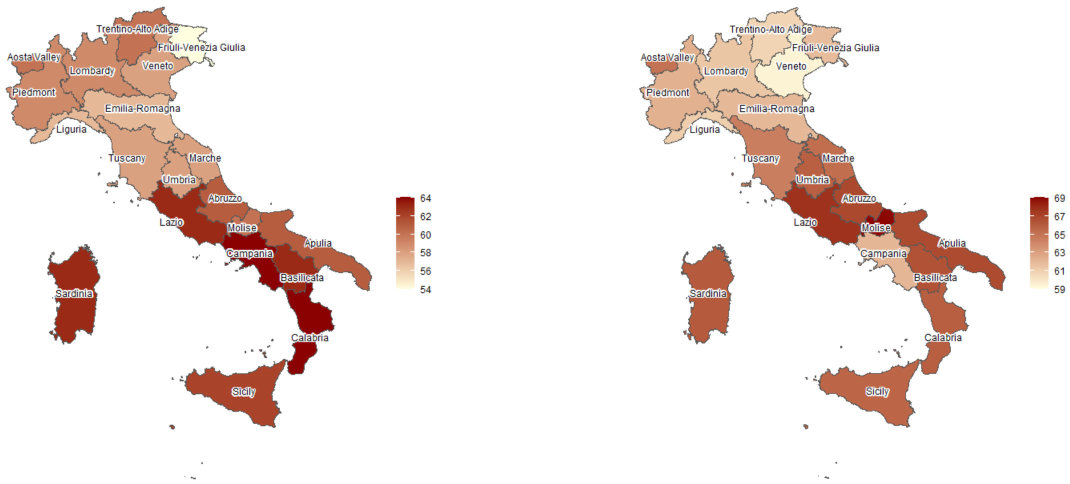

Figure 2: Left panel: percentage of potentially non-compliant firms by region according to BL. Right panel: percentage of unreliability by region according to ISA.

A more detailed comparison is presented in Figure 2, contrasting the regional patterns of evasion estimated using the BL methodology[13] and the ISA. Several notable features emerge. Both indicators consistently highlight two regions — Basilicata, and Lazio — as particularly problematic, ranking them among the top five regions with the highest values in both assessments. In contrast, Friuli-Venezia Giulia and Liguria rank among the five least affected regions, while Aosta Valley exhibits notably high levels of irregularities among northern regions according to both measures. Globally, the two indicators have a Pearson correlation coefficient of 0.51 and a Spearman correlation coefficient of 0.55, indicating a good agreement. Additionally, the BL results demonstrate a positive Pearson correlation of 0.57 with traditional measures of the NOE at the regional level (CGIA-MESTRE, 2023), suggesting that the BL methodology aligns reasonably well with these established metrics, despite capturing distinct dimensions of tax evasion. This agreement highlights the potential of the BL-based approach as a complementary tool for identifying regions where tax evasion is particularly pervasive.

An interesting comparison can also be made between our results and those in Mir et al.

[13]The region with the smallest number of observations is Aosta Valley, with 86 firms in the sample after excluding those for which no conclusion about compliance could be reached. Segment sample sizes are therefore sufficiently large to ensure stable estimation of the percentage of non-compliant firms by region according to BL.

(2014) and Ausloos et al. (2017). In particular, Mir et al. (2014) examine the conformity with BL of the yearly aggregated income taxes for Italian municipalities and for three regions in the South — Campania, Calabria and Sicily — over the period 2007-2011. As the authors state, the reason for focusing attention on these regions is that their local administrations are widely perceived as suffering from weak institutional capacity, inefficiencies in public service provision, and a notable influence of organized crime groups. Given these structural weaknesses, particularly in tax collection and enforcement, the authors expected evidence of tax evasion to emerge in the form of departures from BL. Contrary to this expectation, however, their data for these areas generally conform to the law, as well as data on Italian municipalities as a whole. The only clear exception concerns Campania in 2007 and 2008, where significant deviations from the expected Benford distribution are observed. Overall, their results contrast with ours, which align well with the known factual information on tax evasion across Southern regions. Mir et al. (2014) justify their surprising and unexpected findings by noting that they base their analysis on a single annual tax record for each municipality. Each municipal figure therefore aggregates receipts from hundreds or thousands of residents and is the result of summation, rounding, reallocations and other accounting operations applied to many individual entries. As a consequence, idiosyncratic departures from BL that might be detectable at the micro (taxpayer) level tend to be averaged out or masked by aggregation, substantially reducing the power of first-digit tests on municipal totals. In our case, the analysis is performed at the micro (firm) level, and thus does not suffer from aggregation problems.

Ausloos et al. (2017) use the same dataset as Mir et al. (2014) but test conformity with BL separately for each region and year, as well as for the five-year average. They find that municipal tax data in Liguria and Sardinia deviate from BL in four out of five years and those of Campania in two out of five years. Non conformity with BL is also observed for the three regions on the data averaged over the quinquennium. As for Campania and Sardinia, the finding in Ausloos et al. (2017) are broadly consistent with ours. The same, however, does not hold for Liguria, which in our analysis ranks among the most virtuous regions, in agreement with the ISA. Conversely Ausloos et al. (2017) report the smallest discrepancies with BL for Abruzzo and Calabria — a result they regard as unexpected, especially when considered alongside other apparently compliant regions. As for Mir et al. (2014), aggregation of taxpayers' data at municipal level might play a role in this.

When analysed at a finer, provincial resolution, the data provide further meaningful insights. Figure 3 displays a clear positive correlation between the share of potentially non-compliant firms (according to BL) and the share of firms flagged as unreliable by the ISA within each province. At this level of granularity, the Pearson and Spearman correlation coefficients between the two indicators are 0.26 and 0.29, respectively, both statistically significant at the 1% level. Each point in the graph represents an Italian province and is colour-coded by macro-area. The data tend to cluster around the bisector of the first quadrant. Provinces from the North-East and North-West predominantly appear in the

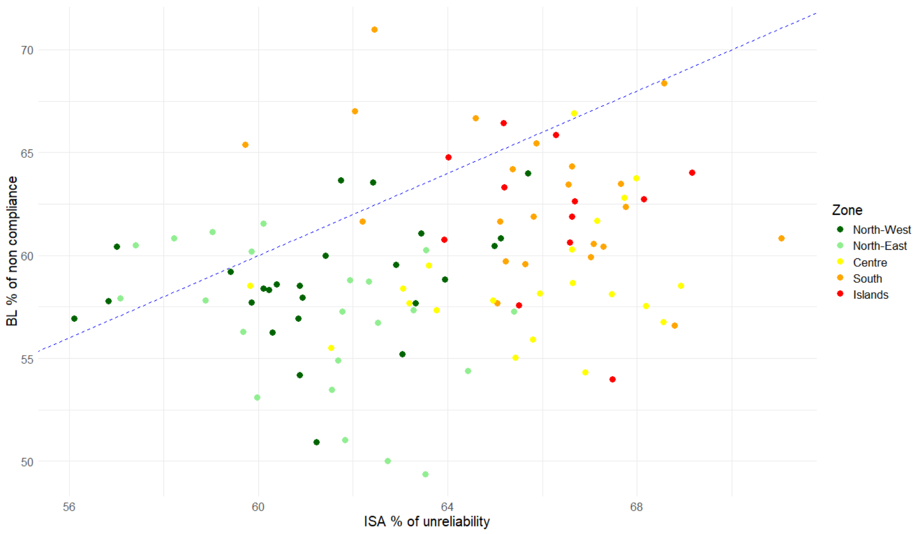

Figure 3: Percentage of potentially non-compliant firms according to the BL indicator vs. percentage of unreliability according to the ISA, at the provincial level. Different colours indicate provinces belonging to different macro-areas. The dashed line represents the bisector of the first quadrant.

lower-left area of the plot, being characterized by the lowest levels of non-compliance and unreliability, while those from the Center, South, and Islands are more frequently located in the upper-right section, exhibiting higher values for both indicators. Notably, a substantial number of points lie below the bisector, indicating that in many provinces, the ISA measure of unreliability slightly exceeds the BL-based estimate of non-compliance. Including financial statements from 2013 might have helped mitigate this discrepancy; however, these data are unfortunately not available in the AIDA database.

We remark that the precision of provincial estimates depends on the (sometimes small) number of sampled firms falling within each province. A minimum of 30 observations is generally considered the conventional threshold for stable estimation of mean-based indicators, in line with standard sampling practice. In our data, all provinces exceed this threshold, and most have more than 50 sampled firms. When excluding observations that cannot be classified as either compliant or non-compliant, the minimum number of observations per province is 39, with only five provinces having fewer than 50 sampled firms. We therefore adopted a more prudential threshold of 50 observations per province — a commonly used cut-off in applied work for segment-level analyses — and excluded these five provinces from previous analysis to ensure greater empirical reliability. Thus, the province with the fewest observations includes 55 firms and has an estimated standard error of 6.9%, whereas the province with the largest number of observations includes 4,291 firms and has a standard error of 0.7%.

Firm size	Compliant	Non-compliant	% of non-compliant	
Micro	11107	17948	61.8%	(61.2%-62.3%)
Small	4456	4791	51.8%	(50.8%-52.8%)
Medium	901	942	51.1%	(48.8%-53.4%)
Large	228	198	46.5%	(41.7%-51.2%)

Table 6: Number of potentially compliant and non-compliant firms, and percentage of non-compliant ones by firms' size, according to BL. Confidence intervals at the 95% level are provided in brackets.

3.3 Firm's size and tax evasion

In Italy, small firms exhibit a higher propensity to tax evasion compared to large ones. This disparity stems from several factors, including the relative invisibility of smaller firms, which are subject to less stringent oversight and fewer audits by tax authorities. Based on the report by CGIA-MESTRE (2019) for the year 2018, micro and small enterprises faced a 3% likelihood of being inspected, compared to 14% for medium-sized companies and 32% for large companies. Additionally, small businesses often handle a higher proportion of cash transactions, making it easier to under-report income (Mas-Montserrat et al., 2023). According to the MEF's report, about one-fifth of potential VAT goes uncollected (20.4% in 2019), amounting to €31.8 billion. A substantial portion of VAT evasion stems from small transactions, such as those in retail trade, which are easy to conceal from the tax authorities — for example, simply by not issuing a receipt. This phenomenon is largely attributed to so-called "consensual evasion", which occurs when a supplier and a customer agree to evade taxes together. In such cases, the company or professional avoids issuing an invoice or receipt, thereby evading IRPEF, while the final customer benefits from a discount equivalent to the VAT they would otherwise have had to pay. For medium and large enterprises, which primarily engage in transactions with other businesses, hiding invoices from the tax authorities is much more challenging. Larger companies, such as retail chains, multinationals, and banks, are inherently less likely to engage in this type of evasion. For them, issuing receipts is valuable as part of their internal accounting and reporting mechanisms. Another key factor is the substantial under-reporting of IRPEF by self-employed workers and small businesses. According to the MEF's estimates for 2019, approximately €32.5 billion of IRPEF revenue was evaded by these groups. This represents the most significant component of tax evasion, both in absolute and relative terms. On average, self-employed workers and small enterprises fail to pay 69% of the income tax they owe each year, nearly 2 percentage points of GDP.

Table 6 clearly shows an inverse relationship between the percentage of non-compliant firms and firms' size, with micro firms presenting a significantly larger propensity to evade taxes[14]. This finding reflects the stylized fact, just discussed, that smaller businesses tend

[14]We use the firms classification into micro, small, medium, or large based on criteria established by the European Commission Recommendation 2003/361/EC. A micro-firm is defined as a firm with fewer than

Macro-sector	BL % of non compliance		ISA % of unreliability
Agriculture	66.7%	(63.1%-70.3%)	74.2%
Commerce	58.1%	(56.9%-59.3%)	62.8%
Manufacturing	53.3%	(52.2%-54.4%)	59.7%
Professionals	56.9%	(53.6%-60.1%)	60.1%
Services	62.7%	(62.0%-63.3%)	64.4%

Table 7: Percentage of potentially non-compliant firms (BL) and percentage of unreliable tax returns (ISA) by economic macro-sector. Confidence intervals are provided in brackets.

to evade more, further lending support to the potential of the BL methodology to discern between firms that engage in tax evasion and those that comply with tax regulations.

3.4 Firm's activity sector and tax evasion

Tax evasion varies significantly across sectors of economic activity, reflecting differences in the nature of transactions, regulatory oversight, and the prevalence of cash-based operations. Sectors such as retail trade, hospitality, and personal services typically exhibit higher rates of evasion due to the frequent use of cash and the relative ease of under-reporting revenues by not issuing receipts or invoices. For example, small restaurants, bars, and hairdressers often operate in ways that make withholding income more feasible. In contrast, industries like manufacturing, finance, and telecommunications tend to have lower levels of evasion, as their operations are more heavily monitored and involve fewer direct cash transactions. Additionally, businesses in sectors where transactions occur predominantly between firms, such as wholesale trade or business-to-business services, face greater scrutiny due to the requirement of issuing invoices, making it more difficult to evade taxes.

Table 7 compares the percentage of potentially non-compliant firms across economic macro-sectors, as estimated using the BL methodology, with the percentage of unreliable tax returns reported by the ISA. Agriculture emerges as the macro-sector most prone to irregularities in financial declarations, followed by services and commerce. Professionals and manufacturing exhibit instead lower levels of irregularities, although the percentage of non-compliance among professionals is not significantly different from that in commerce. The Pearson correlation coefficient between the two measures is relatively high at 0.92, with an even stronger Spearman correlation of 1, reflecting the perfect agreement between the two rankings.

At a more detailed level of disaggregation, Figure 4 compares the percentage of potentially non-compliant firms identified by the BL indicator with the percentage of tax

10 employees, and either an annual turnover or a balance sheet total not exceeding 2 million euros. A small firm is a business with fewer than 50 employees, and either an annual turnover or a balance sheet total not exceeding 10 million euros. A medium-sized firm must have fewer than 250 employees, and either an annual turnover not exceeding 50 million euros or a balance sheet total not exceeding 43 million euros. A large enterprise is any firm that employs 250 or more people, or that exceeds both the turnover and balance sheet thresholds for medium-sized enterprises.

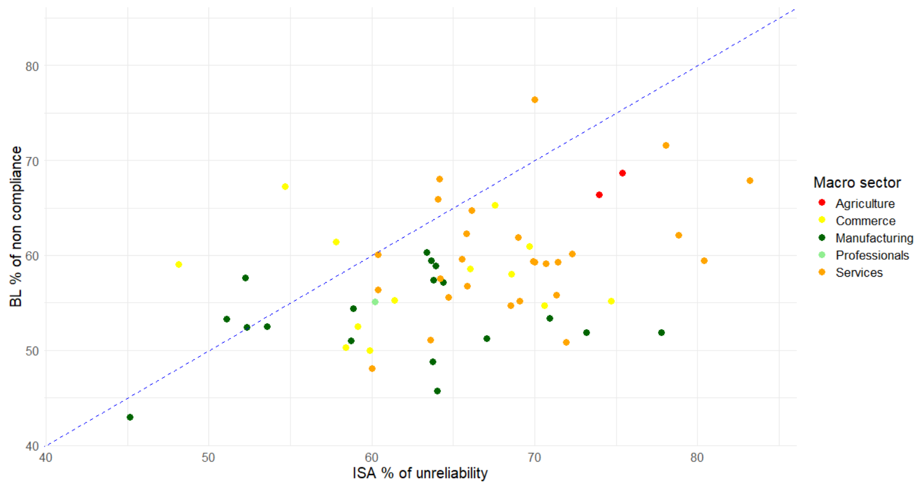

Figure 4: Percentage of potentially non-compliant firms according to the BL indicator vs. percentage of unreliability according to the ISA, by sector of economic activity. Different colours indicate sectors belonging to different macro-sectors. The dashed line represents the bisector of the first quadrant.

unreliability based on the ISA, across sectors of economic activity. Specifically, we refer to the 175 sectoral groups defined by the ISA methodology and use the conversion tables provided by the Italian Revenue Agency[15] to map firms into these groups according to their ATECO 2007 classification codes. To ensure statistical robustness, we restrict the analysis to sectoral groups that include at least 50 firms in our sample, thereby excluding smaller categories where percentages might be unstable or unreliable. The results reveal a clear positive relationship between the BL and ISA indicators, with a statistically significant Pearson correlation coefficient of 0.39. Each point in the plot represents a sectoral group and is colour-coded by macro-sector. As in previous analyses, sectors associated with Agriculture and Services exhibit the highest levels of non-compliance, clustering in the top-right corner of the plot. Conversely, the lower-left corner is predominantly occupied by Manufacturing sectors, along with a single sector related to Professional services.

4 Discussion and conclusions

In this paper, we investigated the potential of Benford's Law (BL) to detect tax-evading firms, with a focus on Italy. Using financial statement data from the AIDA database covering the years 2014–2022, we classified a firm as potentially non-compliant with tax regulations if its financial data deviated from BL in at least one year. This allowed us to construct a firm-level proxy for tax evasion. We then evaluated the validity of this approach by aggregating the firm-level indicator across regions, firm sizes, and sectors of economic

[15] The conversion tables are available at the following link:
https://www.agenziaentrate.gov.it/portale/documents/20143/5734779/Parte+generale+2024.pdf/3a5a2f34-7739-5ae0-7856-0965ded1879c

activity, and analysing whether the resulting distributions could replicate well-established stylized facts about tax evasion in Italy, as reflected in official aggregated data from the Italian Ministry of Economy and Finance and the Revenue Agency.

Our findings reveal a significant alignment between BL-based results and official indicators, suggesting that BL provides a simple yet powerful statistical benchmark that can be applied to publicly available financial data to flag anomalous patterns potentially associated with tax non-compliance. This effective method can help enhance current audit strategies and contribute to the development of more transparent, low-cost, and scalable tools for compliance monitoring.

Another possible use of the BL-based firm-level proxy for tax evasion lies in the field of economic research. Testing theoretical models or hypotheses on, for example, the distortions that tax evasion introduces into market functioning or on the effects it has on compliant firms' outcomes is particularly challenging, mainly due to the scarcity of reliable firm-level measures of tax compliance. In the same vein as Di Marzio et al. (2026), but at a more disaggregated level, the BL-based proxy could help address this issue by providing a consistent and low-cost indicator of potential firms' non-compliance. This would enable researchers to investigate how tax evasion affects market dynamics, resource allocation, productivity, and the competitive environment. Moreover, it opens the door to empirical studies that assess the effectiveness of policy interventions or audit strategies, especially in settings where administrative data are limited or unavailable.

A limitation of the proposed approach is that not all firms are required to file financial statements, particularly in the case of sole proprietorships and partnerships that fall below certain size thresholds. As a result, the BL-based methodology cannot be applied to this subset of businesses, which nonetheless represents a significant portion of the Italian productive structure, especially in sectors where tax evasion is more prevalent. This restricts the scope of the proxy to corporations and medium-to-large enterprises, for which balance sheet data are available in databases such as AIDA. To overcome this limitation, future research could explore the integration of alternative data sources, such as VAT declarations or information from digital invoicing systems, which could provide complementary indicators and might offer a way to generalize the detection strategy beyond balance-sheet-reporting firms.

Declarations

- **Funding** — This work has been funded by the PRIN 2022 under the Italian Ministry of University and Research (MUR) Prot. 2022YMLS4T - TEC - Tax Evasion and Corruption: theoretical models and empirical studies. A quantitative-based approach for the Italian case.

- **Conflict of interest** — All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

- **Ethics declaration** — The authors declare that they have complied with all ethical standards applicable to this research. No ethical approval was required for this study, as it did not involve human participants or animals.

References

F. Alali and S. Romero. Benford's Law: analyzing a decade of financial data. *Journal of Emerging Technologies in Accounting*, 10:1–39, 2013.

M. G. Allingham and A. Sandmo. Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1:323–338, 1972.

M. Almunia and D. Lopez-Rodriguez. Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy*, 10:1–38, 2018.

M. F. Arezzo and R. Cerqueti. A Benford's Law view of inspections' reasonability. *Physica A: Statistical Mechanics and its Applications*, 632:129294, 2023.

N. A. Aris, R. Othman, M. A. M. Bukhori, S. M. M. Arif, and M. A. A. Malek. Detecting accounting anomalies using Benford's Law evidence from the Malaysian public sector. *Management & Accounting Review*, 16:73–99, 2017.

M. Ausloos, C. Herteliu, and B. Ileanu. Breakdown of Benford's Law for birth data. *Physica A: Statistical Mechanics and its Applications*, 419:736–745, 2015.

M. Ausloos, R. Cerqueti, and T. A. Mir. Data science for assessing possible tax income manipulation: The case of Italy. *Chaos, Solitons & Fractals*, 104:238–256, 2017.

P. Bachas, R. N. F. Jaef, and A. Jensen. Size-dependent tax enforcement and compliance: Global evidence and aggregate implications. *Journal of Development Economics*, 140: 203–222, 2019.

I. N. Barros and M. L. Martins. Effects of taxes, redistribution actions and fiscal evasion on wealth inequality: An agent-based model approach. *Physica A: Statistical Mechanics and its Applications*, 679:130960, 2025.

F. Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78:551–72, 1938.

R. Benkraiem, S. Gaaya, and F. Lakhal. Tax avoidance, investor protection, and investment inefficiency: An international evidence. *Research in International Business and Finance*, 69:102258, 2024.

W. Boonzaaier, J. Harju, T. Matikka, and J. Pirttilä. How do small firms respond to tax schedule discontinuities? Evidence from South African tax registers. *International Tax and Public Finance*, 26:1104–1136, 2019.

R. Cerqueti and M. Maggi. Data validity and statistical conformity with Benford's Law. *Chaos, Solitons & Fractals*, 144:110740, 2021.

R. Cerqueti and M. Maggi. Classes of probability measures built on the properties of Benford's Law. *Advances in Statistical Analysis*, 109:197–216, 2025.

R. Cerqueti, M. Maggi, and J. Riccioni. Statistical methods for decision support systems in finance: how Benford's Law predicts financial risk. *Annals of Operations Research*, 342: 1445–1469, 2024.

CGIA-MESTRE. Fisco: Le grandi imprese evadono 16 volte in piú delle piccole. CGIA Report, Ufficio Studi CGIA Mestre, 2019. URL `https://www.cgiamestre.com/wp-content/uploads/2019/10/Accertamenti-fiscali-05.10.2019.pdf`.

CGIA-MESTRE. Il fisco adesso stana gli evasori: Nel 2022 recuperati 20 miliardi. CGIA Report, Ufficio Studi CGIA Mestre, 2023. URL `https://www.cgiamestre.com/wp-content/uploads/2023/05/Evasione-fiscale-20.05.2023.docx.pdf`.

R. Coppier, E. Michetti, and L. Scaccia. Industrial structure and evasion dynamics, is there any link? *Metroeconomica*, 73:960–986, 2022.

R. Coppier, E. Michetti, and L. Scaccia. Size-dependent enforcement, tax evasion and dimensional trap. *Computational Economics*, 65:585–611, 2023.

R. Coppier, E. Michetti, and L. Scaccia. Dimensional traps in evasion models and their effects on industrial structure. *Economic Modelling*, 132:106647, 2024.

I. Cunjak Mataković. The empirical analysis of financial reports of companies in Croatia: Benford distribution curve as a benchmark for first digits. *Croatian Review of Economic, Business and Social Statistics*, 5:90–100, 2019.

S. Das and H. Zhang. Rounding-up in reported EPS, behavioral thresholds, and earnings management. *Journal of Accounting and Economics*, 35:31–50, 2003.

D. Di Gioacchino and D. Fichera. Tax evasion and tax morale: A social network analysis. *European Journal of Political Economy*, 65:101922, 2020.

I. Di Marzio, S. Mocetti, and E. Rubolino. Market externalities of tax evasion. CESifo Working Paper No. 11896, CESifo Munich, 2026.

A. Diekmann. Not the first digit! Using Benford's Law to detect fraudulent scientific data. *Journal of Applied Statistics*, 34:321–329, 2007.

C. Durtschi, W. Hillison, and C. Pacini. The effective use of Benford's Law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting*, 5:17–34, 2004.

F. Gourio and N. Roys. Size-dependent regulations, firm size distribution, and reallocation. *Quantitative Economics*, 5:377–416, 2014.

T. Grammatikos and N. Papanikolaou. Applying Benford's Law to detect accounting data manipulation in the banking industry. *Journal of Financial Services Research*, 59: 115–142, 2021.

N. Guner, G. Ventura, and Y. Xu. Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11:721–744, 2008.

T. P. Hill. The significant-digit phenomenon. *The American Mathematical Monthly*, 102: 322–327, 1995a.

T. P. Hill. Base-invariance implies Benford's Law. *Proceedings of the American Mathematical Society*, 123:887–895, 1995b.

T. P. Hill. A statistical derivation of the significant-digit law. *Statistical Science*, 10:354–363, 1995c.

T. P. Hill and A. Berger. Benford online bibliography. `http://www.benfordonline.net/`, 2009.

ISTAT. *L'economia non osservata nei conti nazionali, anni 2018-2021.* https://www.istat.it/it/files//2023/10/Report-ECONOMIA-NON-OSSERVATA-2021.pdf, 2023.

I. Jianu and I. Jianu. Reliability of financial information from the perspective of Benford's Law. *Entropy*, 23:557, 2021.

A. E. Kossovsky. On the mistaken use of the chi-square test in Benford's Law. *Stats*, 4: 419–453, 2021.

J. J. López and J. Torres. Size-dependent policies, talent misallocation, and the return to skill. *Review of Economic Dynamics*, 38:59–93, 2020.

J. Martin, M. Parenti, and F. Toubal. Corporate tax avoidance and sales: Microevidence and aggregate implications. *Management Science*, to appear, 2026.

M. Mas-Montserrat, C. Colin, E. Ribault, and B. Brys. The design of presumptive tax regimes. OECD Taxation Working Papers 59, OECD Publishing, Paris, 2023. URL `https://doi.org/10.1787/141239bb-en`.

MEF. Relazione sull'economia non osservata e sull'evasione fiscale e contributiva. MEF Report, Ministero dell'Economia e delle Finanze, 2024. URL `https://www.mef.gov.it/export/sites/MEF/documenti-allegati/2024/Relazione-2024.pdf`.

T. A. Mir. The Benford Law behavior of the religious activity data. *Physica A: Statistical Mechanics and its Applications*, 408:1–9, 2014.

T. A. Mir, M. Ausloos, and R. Cerqueti. Benford's Law predicted digit distribution of aggregated income taxes: the surprising conformity of Italian cities and regions. *The European Physical Journal B*, 87:261, 2014.

C. C. Moul and J. V. C. Nye. Testing for strategic manipulation of economic and financial data. In *Benford's Law: Theory and Applications*. Princeton University Press, 2015.

S. Newcomb. Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics*, 4:39–40, 1881.

M. Nigrini. A taxpayer compliance application of Benford's Law. *The Journal of the American Taxation Association*, 18:72–91, 1996.

M. Nigrini. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. John Wiley & Sons, 2012.

J. Nye and C. Moul. The political economy of numbers: On the application of Benford's Law to international macroeconomic statistics. *The B.E. Journal of Macroeconomics*, 7, 2007.

K. Onji. The response of firms to eligibility thresholds: Evidence from the Japanese value-added tax. *Journal of Public Economics*, 93:766–775, 2009.

D. Parnes. Banks' off-balance sheet manipulations. *The Quarterly Review of Economics and Finance*, 86:314–331, 2022.

R. S. Pinkham. On the distribution of first significant digits. *The Annals of Mathematical Statistics*, 32:1223–1230, 1961.

K. V. Ramaswamy. Do size-dependent tax incentives discourage plant size expansion? Evidence from panel data in Indian manufacturing. *Margin: The Journal of Applied Economic Research*, 15:395–417, 2021.

B. Rauch, M. Göttsche, G. Brähler, and S. Enge. Measuring the quality of European statistics. In S. J. Miller, editor, *Benford's Law: Theory and Applications*, pages 235–243. Princeton University Press, Princeton, 2015.

M. Sambridge, H. Tkalčić, and A. Jackson. Benford's Law in the natural sciences. *Geophysical Research Letters*, 37:L22301, 2010.

F. Sandron and S. R. Hayford. Do populations conform to the law of anomalous numbers? *Population (English Edition, 2002-)*, 57:755–761, 2002.

S. Shikano and V. Mack. When does the second-digit Benford's Law-test signal an election fraud? Facts or misleading test results. *Jahrbücher für Nationalökonomie und Statistik*, 231:719–732, 2011.

A. M. Tran, M. D. Griffiths, and D. B. Winters. Small bank managers are prudent: A Benford's Law approach to analyzing loan loss allowances. *Journal of Economics and Business*, 125-126:106128, 2023.

C. Traxler and M. Spichtig. Social norms and the indirect evolution of conditional cooperation. *Journal of Economics*, 102:237–262, 2011.

H. Wickham, R. François, L. Henry, K. Müller, and D. Vaughan. *dplyr: A Grammar of Data Manipulation*, 2023. R package version 1.1.4, https://github.com/tidyverse/dplyr.