Journal Pre-proof

A Benford's Law view of inspections' reasonability

Maria Felice Arezzo, Roy Cerqueti

PII: S0378-4371(23)00849-X
DOI: https://doi.org/10.1016/j.physa.2023.129294
Reference: PHYSA 129294

To appear in: *Physica A*

Received date : 22 July 2023
Revised date : 28 September 2023

Please cite this article as: M.F. Arezzo and R. Cerqueti, A Benford's Law view of inspections' reasonability, *Physica A* (2023), doi: https://doi.org/10.1016/j.physa.2023.129294.

A Benford's Law view of inspections' reasonability

Maria Felice Arezzo[1*] and Roy Cerqueti[2,3]

[1*]MEMOTEF, Sapienza University of Rome, Via del Castro Laurenziano, 9, Rome, 00161, Italy.
[2]Dep. of Social and Economic Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, Rome, 00185, Italy.
[3]GRANEM, University of Angers, SFR CONFLUENCES, Angers, F-49000, France.

*Corresponding author(s). E-mail(s): mariafelice.arezzo@uniroma1.it;
Contributing authors: roy.cerqueti@uniroma1.it;

Abstract

Beyond offering room for methodological research, the assessment of data regularities has relevant implications for applications. In this respect, Benford's Law represents one of the key instruments to detect the presence of possible data manipulation. This paper contributes to this debate by dealing with the analysis of the labor inspections' reliability through an econophysic approach. Specifically, we check the validity of the Benford's Law for a large set of Italian firms' balance sheets and income statements, in both cases of firms inspected and not inspected by the National Institute of Social Security and for all the Italian regions. In so doing, we provide a panoramic view of the plausability of the inspection activities at a regional and financial item level.

Keywords: Benford law, Undeclared work, Labour inspections, Administrative data

1 Introduction

At the EU Commission level, the undeclared work is defined as *any paid activities that are lawful as regards their nature but not declared to the public authorities*[1]. More specifically, we talk about a special form of evasion when we refer to undeclared work,

[1]https://ec.europa.eu/social/main.jsp?catId=1298&langId=en

in that taxes due on gross wages and social security contributions are not paid in this case.

Empirical evidence and standard economic theory point out the detrimental effects deriving directly from this harming behavior [17, 36]. In particular, undeclared work leads to unfair fiscal treatment of the taxpayers, distortion of the competition (those who do not pay taxes have an advantage because they reduce their costs, increasing level of the tax rate for recovering from undeclared incomes and fund public expenditure, high unemployment level and low productivity.

As a consequence, it is of paramount importance to have tools that can help public authorities in contrasting undeclared work. In this respect, labor inspections play a crucial role. However, inspections are costly for the policymakers. Therefore, the issue related to the analysis of their efficiency deserves a deep exploration.

We contribute to the debate on the topic by employing the Benford's law (BL, hereafter) as a methodology for evaluating the labor inspections efficacy. In so doing, we also provide a guide for policymakers on the firms' characteristics to be monitored with care.

BL describes the frequency distribution of leading digits in many natural and human phenomena. Such a data regularity property has been discovered in [29] and better formalized by [7]. In particular, BL states that the leading first digit d in a set of numbers from many real-life data is not uniformly distributed, as one might naively expect, but it follows a logarithm-type law given by:

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

where p_d is the probability that a number from the dataset has its first digit equal to d. There are other versions of the BL beyond the one expressed in equation (1), by exploring the $h - th$ digit with $h > 1$ or combinations of digits [32]. The employment of such generalizations is required when the first digits are not enough to respond to the research question – and this is not the case, in our paper.

BL has drawn a lot of attention in the scientific world especially since the nineties of the last century [6]. One reason is its validity in an incredibly large number of situations from various scientific contexts like finance and economics (see e.g. [2, 4, 12, 19, 23, 26, 34]), hydrology and geological sciences (see e.g. [3, 13, 35]), electoral studies (see e.g. [11, 25, 33]), internet search patterns ([20]).

The interested reader can refer to [38] for the story of BL along with its theoretical advancements from the original formulation and a list of its applications. More recently, [27] elaborates on the worthiness of this law, quoting also some historical facts about its introduction.

BL can be viewed a "natural property" of some datasets. The violation of the BL points to the presence of distortions and possible manipulations of the data. For this reason, such a statistical law is often exploited in the context of fraud detection. This is particularly true in the context of accounting, where BL is an acknowledged instrument of fraud detection (see e.g. [16, 18, 30–32]). Indeed, the violation of the BL can mirror misreporting or cheating behaviors when dealing with datasets composed by balance sheet and income statements entries (on this, see also [21]).

This paper moves from this premise. We test the compliance to BL of firms' income statements in the context of labour inspections. More in detail, we consider a high-quality dataset of Italian firms' performance indicators composed of two subsamples. One of them contains companies audited in 2005 by the National Institute of Social Security while the other one relies to randomly selected firms. Firms are clustered at Italian regions level, to have a clear view of the different regional realities. In this framework, we pursue two different targets. On one side, we aim at identifying the items breaking more evidently the compliance with the BL; on the other side, we make an *ex-post* assessment of inspections' plausibility.

The latter statement needs further clarification. Labour inspectors visit several firms each year, either based on their knowledge of the economic and production fabric of the territory they are in charge of, or because they receive whistle-blowing. If the inspections are plausible, then the sample of inspected firms is far less compliant to BL than any random sample drawn from a population of active firms. In other words, for inspections to be plausible, the conformity of the non-random sample of inspected firms should be much lower than that of any other random sample.

In the described framework, a prominent role is played by the statistical methodology used to assess if a variable complies to BL. The conformity of a given random variable with BL is statistically proved when the value of the considered statistical distance is below a threshold. We adopt the critical values of [32] and [21], who derive valid thresholds for the Mean Absolute Deviation (MAD) and the Sum of Squared Deviations (SSD) to assess the compliance of a random variable to BL (see also the methodology developed in [9] for the formalization of these critical values).

Results are quite interesting, and offer a panoramic view of the regional realities in Italy as well as of the financial variables more prone to violate BL. Specifically, while in the random sample the conformity to BL is maintained for the items in the top part of the income statement (such as revenues and costs) and degrades for those in the bottom part (EBIT, EBITDA and profits) in the sample of inspected firms this pattern is scattered. As the accounting literature points out [24, 28, 37], items in the top part of income statements are impossible to manipulate unless there is a will of frauds; in the bottom part, companies have more room for manoeuvre and, while complying with the law, items can be legitimately changed in order to pay less tax. This findings are consistent with the fraudulent behaviour that firms with undeclared workers must adopt in order to cover up their misconduct.

The remainder of the paper is as follows. In section 2 we describe the data used in the analysis along with the methodology; in section 3 we report the results and finally we draw our conclusions in 4.

2 Data and methodology

This section contains the description of the unique and high-quality dataset at the focus of the study and the methodological devices employed for the analysis. The data are available to the authors thanks to a research agreement between the Italian National Institute of Statistics (ISTAT), the National Institute of Social Security (INPS) and the National Institute for Public Policy Analysis (INAPP).

2.1 Description of the dataset

In this work we make use of two samples of data drawn from the population of Italian building and construction companies (NACE section: F) operating in 2005. The size of the population is 529,757.

Both samples come from an administrative source, that is data gathered by public authorities with a typical aim of control/knowledge of the statistical units. This kind of data are a very precious sources of information because they allow to study phenomena that would remain otherwise unknown.

The first sample consists of the 14,777 building and construction companies (bcc henceforth) audited in 2005 by the National Institute of Social Security inspectors and the second is a random sample of 75,382 bcc. The two samples overlap by 2,539 units.

The first sample contains information on firms' main economic characteristics as retrieved from the so-called "Studi di Settore" (SS in the following) of the Italian Revenue Agency[2] It is worth noting that this sample is non-random because inspectors choose to visit firms that are *suspicious* – i.e., more prone to non-compliance – for some reasons.

The second sample (the random one) contains the same information from SS.

The SS archive contains an exhaustive list of information on corporate organization, firm structure, management and governance. We considered the following variables: bcc location region, revenues (i.e. the money a business generates through its normal operations), turnover (i.e. the total value of the sale of services or goods during a financial year), added value (the capacity of a company to generate wealth), cost of sales (all costs used to create a product or service), total costs (global amount of costs), cost of raw material (cost of substances used in the manufacturing of goods), cost of services (includes all the direct costs involved in producing the services), EBITDA (Earning before interest, taxes, depreciation and amortization is a measure of core corporate profitability) EBIT (Earnings before interests and taxes), and profit.

To better understand the relevance of these financial variables to the labor part of the company, it is worth recalling that a company's balance sheet and the items presented in it are interrelated and provide an economic representation of the company itself. This presentation must be internally consistent. If the use of labour units is concealed, it is necessary to alter several accounting entries in order to try to give the financial statements an apparent consistency. For example, a construction company that is able to carry out many contracts by using undeclared workers cannot show the actual turnover achieved in its balance sheet, as this would force it to show consistent labour costs. It must therefore alter its turnover and revenues in order to justify low labour costs. This need to manipulate various balance sheet items, which companies using undeclared labour have and which is not necessary for regular companies, is the basis for the application of BL. In this sense, we assume that irregular firms will be less compliant with BL *also* for those balance sheet items for which the legislator leaves less room for manoeuvre (such as turnover and costs) and, if inspections are

[2]The SS is an administrative data source that can be explored also through the ISTAT data warehouse. It collects a set of data on income statement features that firms with at most 5 million euros of income are obliged to communicate to the Revenue Agency on a yearly basis.

plausible, they will be concentrated on those firms that are most likely to be irregular and therefore least compliant with BL.

In Table 1 we report some descriptive statistics of the two samples that give insights on their differences, being the inspected firms typically larger.

Table 1 Descriptive statistics of the two samples (inspected and random sample). Values are in euros.

	Sample of inspected firms (n=14,777)		Random sample (n=75,382)	
Variable	Mean	SE	Mean	SE
Revenues	533052.86	6644.90	232584.06	1892.13
Turnover	538371.64	6549.35	235592.74	1861.09
Added value	475302.52	7533.23	212320.48	2109.90
Cost of sales	173887.43	3357.47	84363.41	982.78
Costs	267868.11	4194.17	121271.70	1234.27
Cost of raw material	190977.36	3183.17	87536.79	959.78
Cost of services	93980.68	2310.04	40767.89	652.22
EBITDA	126797.77	1746.88	64771.84	466.25
EBIT	112586.46	1637.27	58489.23	434.25
Profit	33369.33	569.23	28048.51	175.15

From the economic data in the two samples, we computed the corresponding first-digit data-sets. Then, we computed the empirical first-digit Benford distribution for each of the 20 Italian regions. As measures to assess the difformity between the empirical and the theoretical Benford distribution, we used the *Mean Absolute Deviation* (MAD, hereafter) and the *Sum of Squared Deviations* (SSD), as described in detail in sub-section 2.2.

2.2 Methodology

Given a dataset, the empirical probability that the first digits of one of its elements is d will be denoted by $\tilde{p}_d$, for each $d = 1, 2, \ldots, 9$.

In order to check the validity of the BL, we employ two concepts of distance between the empirical probabilities $\tilde{p}$'s and the theoretical probabilities p's in Eq. (1) coming from the BL.

The Minimum Absolute Deviation (MAD) is defined by

$$MAD = \frac{1}{9}\sum_{d=1}^{9} |p_d - \tilde{p}_d|. \tag{2}$$

MAD is evaluated as the most reliable test for checking the validity of the BL, with a value below 0.006 as close conformity, between 0.006 and 0.012 as acceptable conformity, marginally acceptable conformity for values between 0.012 and 0.015 and nonconformity otherwise (see [32])

The Sum of Squared Deviations (SSD) is defined as follows:

$$SSD = \sum_{d=1}^{9} (p_d - \tilde{p}_d)^2. \tag{3}$$

When SSD is above 100, then we have noncompliance with the Benford's law (see [21]).

There are important reasons for preferring the employment of MAD and SSD to other well-known measures of conformity, such as χ^2. Indeed, there is a large strand of the literature on BL pointing out that χ^2 is not appropriate for its excessive power; specifically, χ^2 tends to reject Benfordness even when the deviations from the theoretical BL are negligible – mainly in presence of large samples [15, 22, 32]. In this respect, several scholars favour the employment of MAD and SSD as suitable statistical tests, in that they do not depend on the sample size [8, 9, 14, 32].

To improve the readability of the results, we built some measures to evaluate the plausibility of region's inspections. First of all, using the thresholds identified in [32] and [21] (on this, see also [9]), for each region we determined which income statement items do (1) or do not (0) comply with BL. This task was done for both samples (inspected and random). We then compiled the results into a contingency table such as:

		Random sample	
		Compliant	Non-compliant
Inspected	Compliant	a	b
firms	Non-compliant	c	d

where a is the number of features for which both samples are compliant (match), b is the number of attributes such that the sample of inspected firms is compliant but the random sample is not (mismatch), c is the number of attributes where it happens the other way around (mismatch) and d is the number of attributes where both samples do not comply (match).

Since we want to understand how different the two samples are, we computed the metrics Q of Yule and of VARI [1, 10] as follows:

$$\begin{aligned} D_{YuleQ} &= \frac{2 \cdot b \cdot c}{a \cdot d + b \cdot c} \\ D_{VARI} &= \frac{b + c}{4 \cdot (a + b + c + d)} \end{aligned} \tag{4}$$

besides the two distances, we also computed a measure to help understand the regions that have a more efficient inspection activities. The measure is $I_{BL} = c/(c + b)$ and the closer it is to the upper bound (i.e. 1) the most effective are the inspections according to BL. In fact, it means that the sample of inspected firms has a number of non-compliant features higher than the total mismatches.

3 Results and discussion

As a premise, we point out that the BL is a device with a global meaning over a collection of data, without having the possibility of identifying the discrepancies of individual firms. In this respect, the BL allows to point to a possible data manipulation problem, without entering the specific context of the identification of the manipulators.

Fig. 1 depicts four heat-maps of the MADs and SSDs measures for the two samples at use. The shade of grey is proportional to the value of MAD (graphs in top panel) and SSD (bottom panels): the higher the value the darkest the grey. Left panel are the inspected firms while right panel are the random samples.

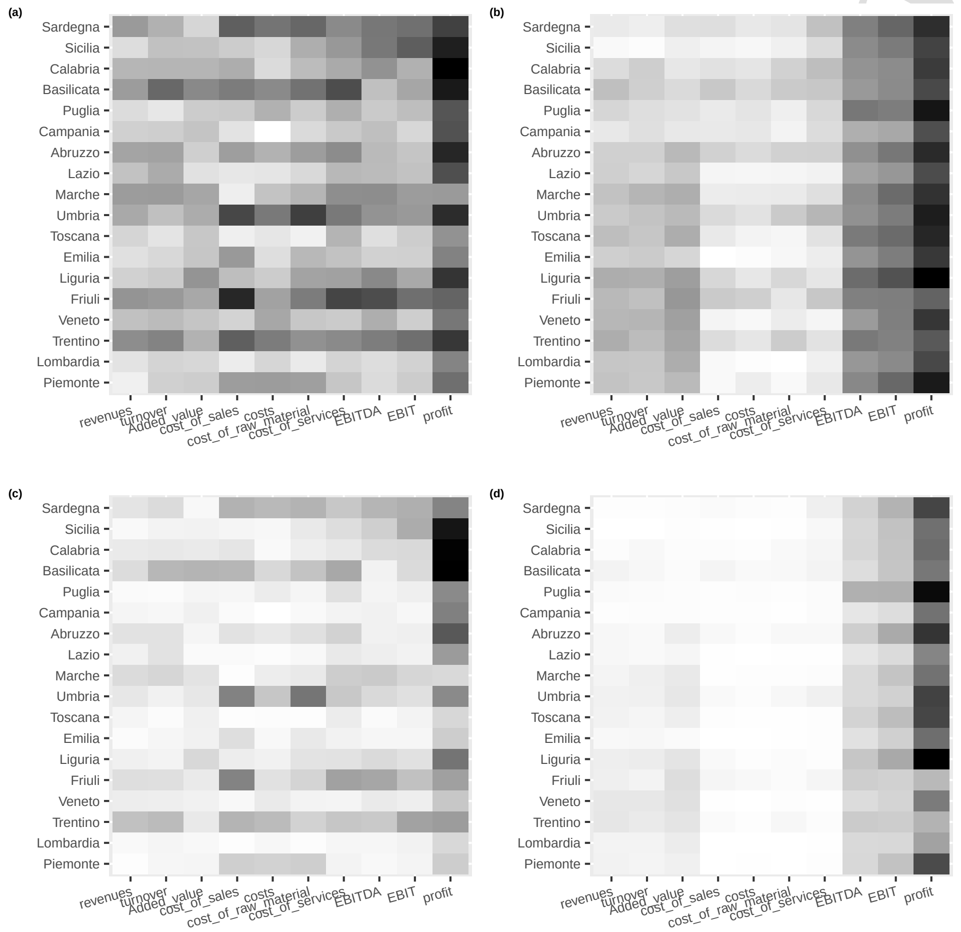

Fig. 1 Heatmap of the conformity measures. Panel: (a) MAD inspected firms; (b) MAD random sample; (c) SSD inspected firms; (d) SSD random sample. Darker grey means less conforming.

The $x--$ and $y--$ axes contains the elements of the sets of the financial variables and regions, respectively. Regions are arranged from the South (top) to the North (bottom). In doing so, the heatmap offers a clear view of the reality of Northern-Central-Southern Italy. As for the financial variables, the ordering reflects that in the income statement to allow a clearer visualization of a possible fraudulent behavior. In fact, as recalled in the introduction, items in the top part of income statements (like revenues, turnover and different types of costs) are manipulate if there is a will of frauds while in the bottom part, companies have more room for manoeuvre within the rule of law.

We can read results both in rows and columns. When reading by row, we can visually assess which regions made better choices in selecting the firms to inspect. In fact, we should compare the greys in the inspected sample (left) with those in the random sample (right).

If we consider the sample of inspected firms, then the darker the grey in the heatmap, the more plausible the inspections. Indeed, the dark zones are associated to low compliance with the BL, hence validating the effectiveness of the inspections.

Heatmaps on the left side of Fig. 1 appear to be darker than those on the right side. This means that inspected firms (left side) have financial parameters less compliant with the BL than those in the random sample (right side). This is a crucial information on the overall inspection process, in that it explains that such inspections are in general worthy and effective – taking place in cases of non compliance with the BL. At a regional level, this outcome can be observed in the vast majority of the cases.

Relevant insights can be obtained when reading the heatmaps by column – hence, pointing to the considered financial variables. When reading by column, we identify the items in the financial statement that do not follow BL. For example, the variables regarding all types of costs (costs of sales, costs of raw materials and so on) comply quite well with BL and this is probably due to a higher level of difficulty in the manipulation of data that are based on other firms' invoices. A completely different picture appears when looking at the EBIT, EBITDA and profit: their calculation has a greater margin of arbitrariness because the law allows companies to choose how to compute certain items and they often make their choices in a way that reduces the tax burden.

By combining regions and variables, we observe that MAD offers a debatable efficiency of the inspections on some variables (revenues, turnover, added value) for the regions in the Northern part of Italy (see the lower-left quadrant, with Liguria, Friuli, Veneto, Trentino, Lombardia and Piemonte), with a heatmap that is darker in the sample case than in the inspected one.

The results obtained with the first digit, clearly point to a response to our research question either under a regional point of view as well as at the level of the specific variables that have been considered. Therefore we did not proceeded with any analysis at a higher digit order.

The results of the metrics Q of Yule and of VARI are reported in Table 2.

Firstly, with a correlation coefficient of 0.77 between the two metrics, we can see that the results are consistent. The regions with the greatest distance between random and inspected samples are Emilia, Marche, Umbria, Liguria, Piemonte, Abruzzo,

Table 2 Distances between samples for all regions.

	D_{yuleQ}	D_{vari}	I_{BL}
Piemonte	1.06	0.12	0.50
Lombardia	0.00	0.06	0.00
Trentino	2.00	0.12	0.67
Veneto	0.00	0.12	0.00
Friuli	2.00	0.12	0.83
Liguria	1.00	0.12	0.33
Emilia	0.25	0.06	0.33
Toscana	0.00	0.08	0.00
Umbria	0.67	0.10	0.60
Marche	0.33	0.08	0.75
Lazio	0.00	0.04	0.00
Abruzzo	1.80	0.19	0.67
Campania	0.00	0.06	0.00
Puglia	0.00	0.06	0.00
Basilicata	2.00	0.21	0.80
Calabria	0.00	0.04	0.00
Sicilia	0.00	0.02	1.00
Sardegna	0.00	0.12	1.00

Trentino, Friuli and Basilicata. These are the same as those with a higher value of the index in the third column of Table 2, which, as mentioned above, measures the plausibility of inspections in terms of BL.

Another very interesting result is obtained by comparing the index I_{BL} with the regional irregularity rate of 2005 calculated by ISTAT. The correlation coefficient with the irregularity rate[3] is 0.59. It means that there is a strong correlation between the plausibility of the inspections and the irregularity rate. This result emphasizes that BL is a useful instrument for inspection authorities. It should be stressed, though, that in the computation of the correlation we excluded the southern regions, where the mafia is well known to have a strong presence in construction companies. In these cases, for obvious reasons, inspection activity is extremely delicate and the choices made by inspectors may not be completely free.

4 Conclusions

In this work, we explored the possibility of using BL to assess the plausibility of inspection activity. The results clearly demonstrate its usefulness in this regards. To reach this conclusion, we compared some items in the income statement in a sample of inspected companies and in a random sample. The implicit assumption is that since the companies inspected are those most likely to be irregular, they must be less compliant with the BL than those in the random sample. The intuition behind our work is that since each company is an *unicum*, the accounting records must reflect this unity. If part of the workforce is hidden (as in irregular enterprises), it is necessary to alter the accounts in order to give them an apparent coherence. In doing so, however, the BL is easily broken. The results clearly point in this direction.

[3] Data on regional irregularity rate is available upon request or it can be found on the ISTAT website.

Furthermore, since it is possible to measure the degree of compliance with the BL for each Italian region and for each income statement item, it is possible to understand where there is room for improvement in inspection activity and which items need to be given more attention in order to understand the irregularity of the company. In this regard, all the items related to costs are difficult to manipulate unless there is a fraudulent intent, as well as revenues, turnover and added value. These are those to look at when searching for help in improving inspection activities.

We point out that the results of the present paper are often in agreement with the outcomes of [4], even if the considered empirical instances and the reference periods are quite different. For example, Sardegna is confirmed to have a low conformity with the BL for the special case of inspected firms, while Liguria is still not conform with Benford in the random sample.

Acknowledgements

The authors, acknowledge the financial support from the La Sapienza University of Rome under the grant number RP120172B8134439.

Data availability

Data will be made available on request.

References

[1] A. Agresti, Categorical data analysis. New York: Wiley. 2013

[2] M. Ausloos, R. Castellano, R. Cerqueti, Regularities and discrepancies of credit default swaps: a data science approach through Benford's law. Chaos, Solitons and Fractals, 90, (2016) 8-17.

[3] M. Ausloos, R. Cerqueti, C. Lupi, Long-range properties and data validity for hydrogeological time series: The case of the Paglia river. Physica A: Statistical Mechanics and its Applications, 470, (2017), 39-50.

[4] M. Ausloos, R. Cerqueti, T.A. Mir, Data science for assessing possible tax income manipulation: The case of Italy. Chaos, Solitons and Fractals, 104, (2017), 238-256.

[5] M. Ausloos, C. Herteliu, B. Ileanu, Breakdown of Benford's law for birth data. Physica A: Statistical Mechanics and its Applications, 419, B. (2015), 736-745.

[6] L. Barabesi, A. Cerioli, D. Perrotta, Forum on Benford's law and statistical methods for the detection of frauds. Statistical Methods & Applications, 30, (2021), 767-778.

[7] F. Benford, The Law of Anomalous Numbers. Proceedings of the American Philosophical Society, 78, 551-572, 1938.

[8] R. Cerqueti, C. Lupi. Severe testing of Benford's law, TEST: An Official Journal of the Spanish Society of Statistics and Operations Research, 32(2), (2023), 677-694.

[9] R. Cerqueti, M. Maggi, Data validity and statistical conformity with Benford's Law. Chaos, Solitons and Fractals, 144, (2021), 110740.

[10] S.-S. Choi, C. Sung-Hyuk, C.C. Tappert, A Survey of Binary Similarity and Distance Measures. Journal on Systemics, Cybernetics and Informatics 8 (2010), 43-48.

[11] J. Deckert, M. Myagkov, P.C. Ordeshook, Benford's Law and the detection of election fraud. Political Analysis, 19(3), (2011), 245-268.

[12] M. J. De Ceuster, G. Dhaene, T. Schatteman, On the hypothesis of psychological barriers in stock markets and Benford's Law. Journal of Empirical Finance, 5(3), (1998), 263-279.

[13] J. Díaz, J. Gallart, M. Ruiz, On the ability of the Benford's Law to detect earthquakes and discriminate seismic signals. Seismological Research Letters, 86(1), (2015), 192-201.

[14] P. D. Drake, M. J. Nigrini, Computer assisted analytical procedures using Benford's Law, Journal of Accounting Education, 18(2), (2000), 127-146.

[15] E. Druica, B. Oancea, V. Călin, (2018), Benford's law and the limits of digit analysis, International Journal of Accounting Information Systems, 31(C), (2018), 75-82.

[16] C. Durtschi, W. Hillison, C. Pacini, The effective use of Benford's law to assist in detecting fraud in accounting data. Journal of Forensic Accounting, 5(1), (2004), 17-34.

[17] European Commission, Stepping up the fight against undeclared work, COM(2007) 628 final. Brussels: European Commission, 2007.

[18] T. Grammatikos, N.I. Papanikolaou, Applying Benford's law to detect accounting data manipulation in the banking industry. Journal of Financial Services Research, 59, (2021), 115-142.

[19] C. Herteliu, I. Jianu, I. M. Dragan, S. Apostu, I. Luchian, Testing Benford's Laws (non)conformity within disclosed companies' financial statements among hospitality industry in Romania, Physica A, 582, (2021), 126221.

[20] B.V. Ileanu, M. Ausloos, C. Herteliu, M.P. Cristescu. Intriguing behavior when testing the impact of quotation marks usage in Google search results. Quality & Quantity 53, 2507–2519 (2019).

[21] A. E. Kossovsky, Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications. Singapore: World Scientific, 2014.

[22] A.E. Kossovsky. On the mistaken use of the chi-square test in Benford's law, Stats,4(2), (2921) 419–453.

[23] E. Ley, On the peculiar distribution of the US stock indexes' digits. The American Statistician, 50(4), (1996), 311-313.

[24] K. Malikov, S. Manson, J. Coakley, Earnings management using classification shifting of revenues, The British Accounting Review, 50 (3), (2018), 291–305.

[25] W.R. Mebane, Comment on Benford's Law and the detection of election fraud. Political Analysis, 19(3), (2011), 269-272.

[26] T.A. Mir, M. Ausloos, 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(11), (2014), 261.

[27] T.A. Mir, M. Ausloos, Benford's law: A "sleeping beauty" sleeping in the dirty pages of logarithmic tables. Journal of the Association for Information Science and Technology, 69(3), (2018), 349-358.

[28] K. H. M. Naser, Creative Financial Accounting. Its nature and use. Prentice-Hall, Hemel Hempstead eds, 1993.

[29] S. Newcomb, Note on the Frequency of Use of the Different Digits in Natural Numbers. American Journal of Mathematics, 4(1/4), (1881), 39-40.

[30] M.J. Nigrini, A taxpayer compliance application of Benford's law. The Journal of the American Taxation Association, 18(1), (1996), 72-91.

[31] M.J. Nigrini, I've got your number. Journal of Accountancy, 187(5), (1999), 79-83.

[32] M.J. Nigrini, Benford's Law: Applications for forensic accounting, auditing, and fraud detection, John Wiley & Sons, 2012.

[33] L. Pericchi, D. Torres, Quick Anomaly Detection by the Newcomb—Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela. Statistical Science, 26(4), (2011), 502-516.

[34] J. Riccioni, R. Cerqueti, Regular paths in financial markets: Investigating the Benford's law. Chaos, Solitons and Fractals, 107, (2018), 186-194.

[35] M. Sambridge, H. Tkalcic, A. Jackson, Benford's law in the natural sciences. Geophysical Research Letters, 37(22), (2010), L22301.

[36] Small Business Council, Small business in the informal economy: making the transition to the formal economy. London: Small Business Council, 2004

[37] T. Sutton, Corporate financial accounting and Reporting, Prentice Hall, Harlow, 1998.

[38] J. Torres, S. Fernandez, A. Gamero, A. Sola, How do numbers begin? (The first digit law). European Journal of Physics, 28(3), (2007), L17.

Highlights

- Labor inspections plausibility is discussed;
- The Benford law is used for detecting financial anomalies;
- Compliance to Benford law of Italian building and construction companies is presented;
- Regional differences and similarities are discussed.

Credit authors statement

September 27, 2023

Maria Felice Arezzo: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing.
Roy Cerqueti: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing.

Declaration of interest statement

September 27, 2023

The authors declare no conflict of interest.