The potential of Benford's law to derive disaggregated tax evasion indices for Italy

Raffaella Coppier[1] Elisabetta Michetti[1] Luisa Scaccia[1]

[1]Dipartimento di Economia e Diritto
Università di Macerata

Rome, 27 June 2024

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Table of Contents

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | unimc UNIVERSITÀ DI MACERATA | DIPARTIMENTO DI ECONOMIA E DIRITTO

Table of Contents

Finanziato dall'Unione europea NextGenerationEU | Ministero dell'Università e della Ricerca | Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA | unimc UNIVERSITÀ DI MACERATA | DIPARTIMENTO DI ECONOMIA E DIRITTO

Benford's Law: from empirical observation . . .

- Empirical observation: in many real-life sets of numerical data, the leading digit is likely to be small

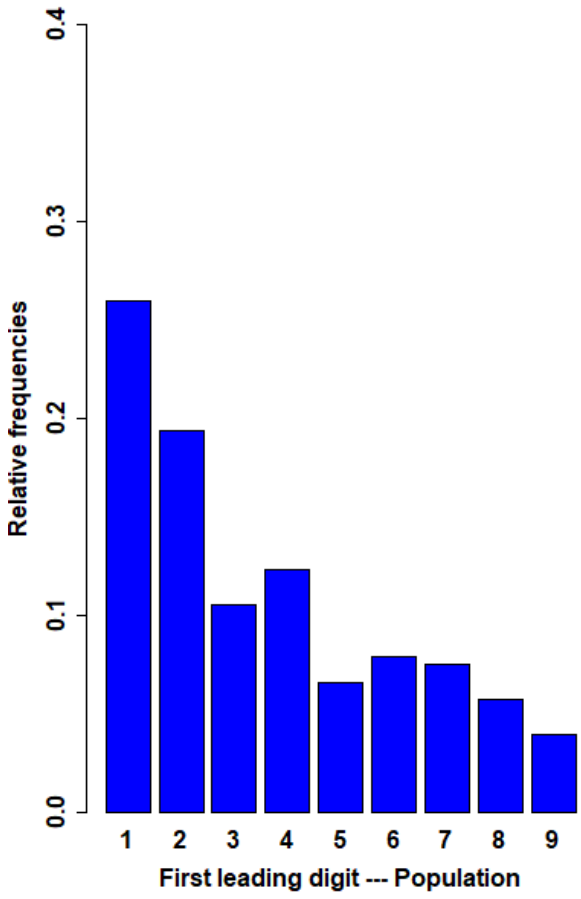
  $\longrightarrow$ The leading digit is the first from the left non-zero digit in a number. The leading digit in 23 is 2, in 0.31 is 3, in -1.5 is 1.

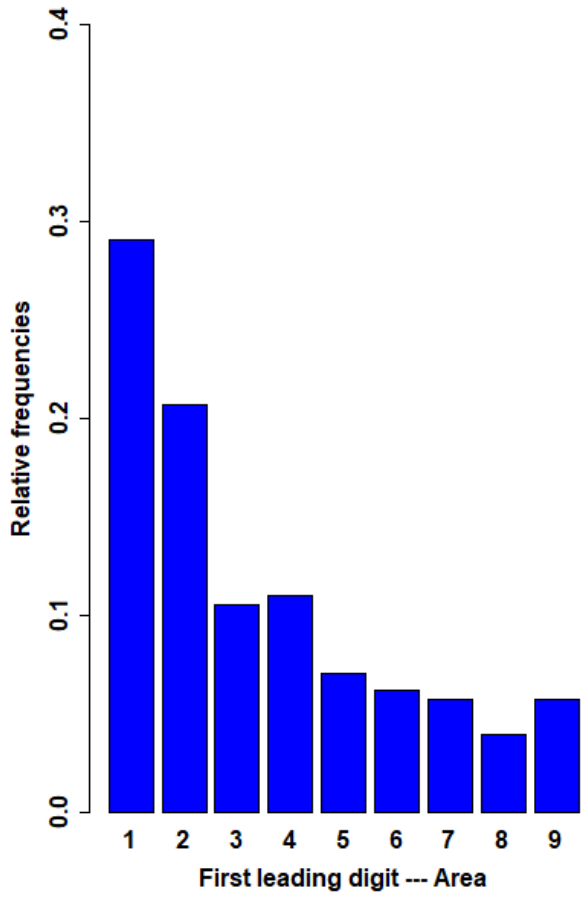
- Example

Country	Population	Area (km^2)	Per capita GDP
Afghanistan	**3**1,056,997	**6**47,500	**7**00
Albania	**3**,581,655	**2**8,748	**4**500
Algeria	**3**2,930,091	**2**,381,740	**6**000
⋮	⋮	⋮	⋮

Table: Countries of the world — some figures

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO





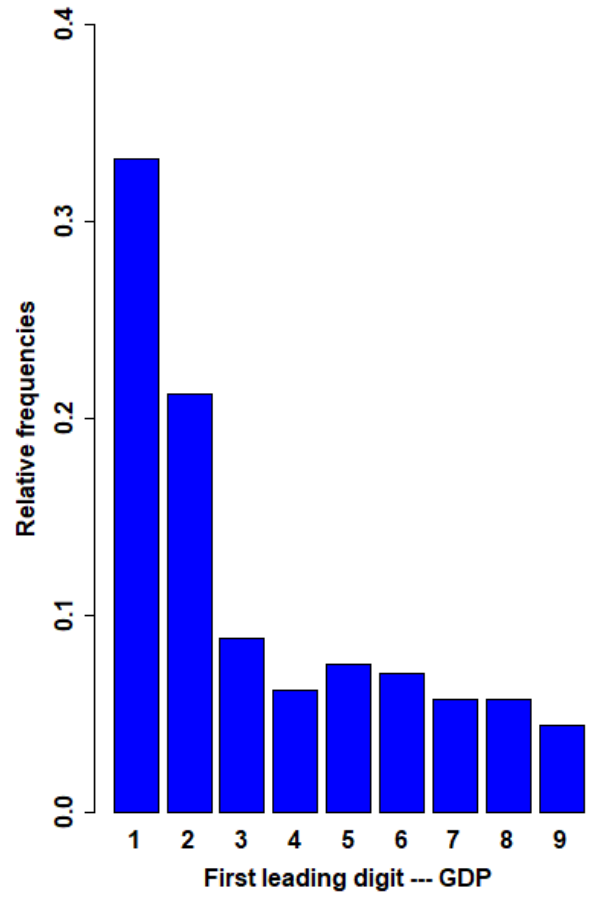


Figure: Distribution of first leading digit for Population, Area and GDP

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

. . . to a mathematical law!

- Benford's law, also known as the **Newcomb–Benford law**, the **law of anomalous numbers**, or the **first-digit law**, was first formalised by Newcomb (1881) and then made famous by Benford (1938).

Definition

A set of numbers is said to satisfy Benford's law if the leading digit d ($d \in \{1, \ldots, 9\}$) occurs with probability

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(1 + \frac{1}{d}\right)$$

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

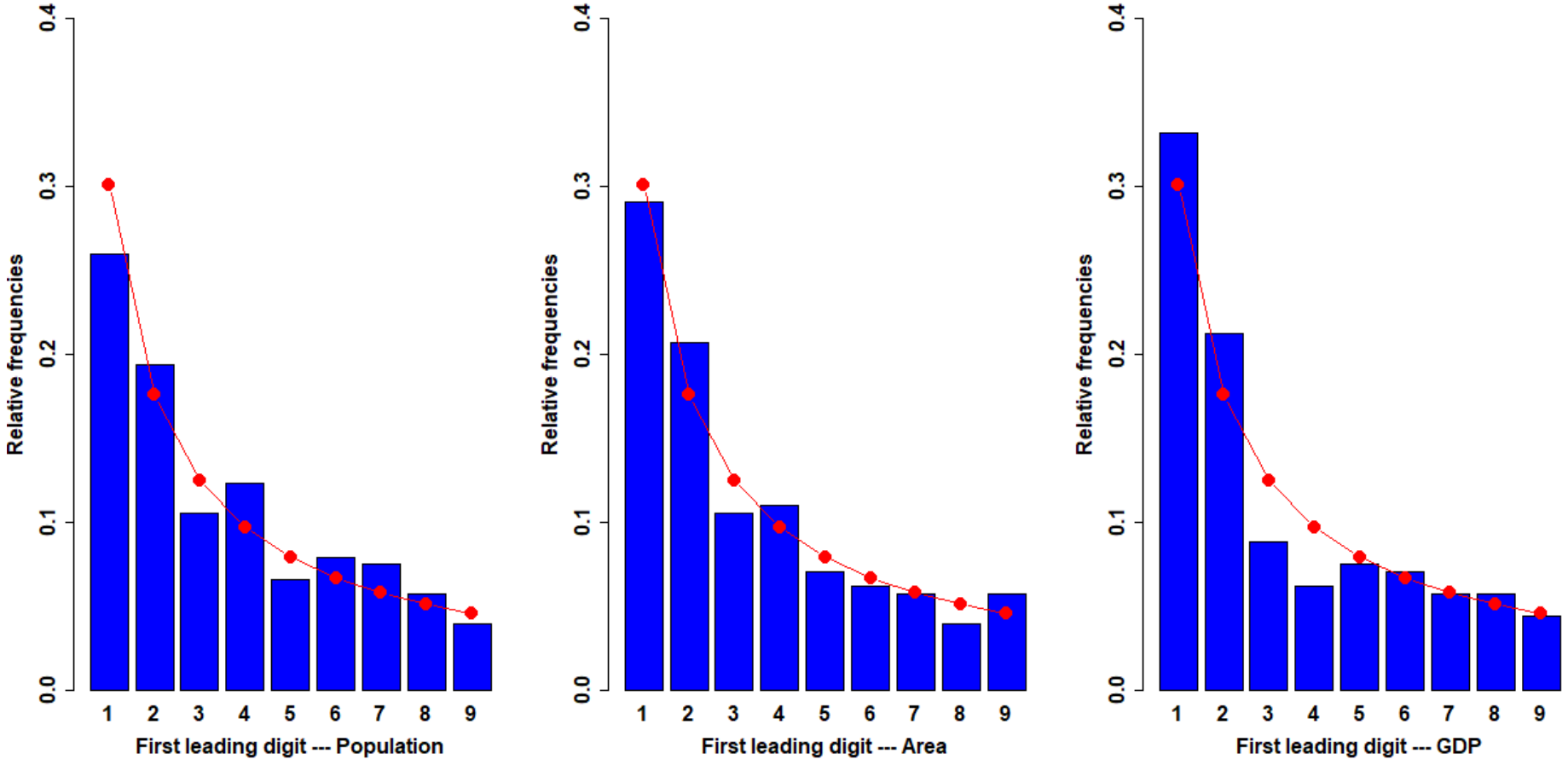


Figure: Distribution of first leading digit for Population, Area and GDP, and theoretical Benford's distribution (in red)

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Examples of data following the Benford's law

- **Population data**: populations of cities or countries, number of births, deaths, or diseases reported ...
- **Geophysical data**: mountains elevation, rivers length ...
- **Scientific data**: random numbers from some statistical distributions, gene lengths, physical constants, distance of stars ...
- **Financial and economic data**: stock prices, transaction volumes, invoice amounts, expense claims, taxable income, GDP ...
- **...**:

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Table of Contents

Finanziato dall'Unione europea NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

unimc UNIVERSITÀ DI MACERATA

DIPARTIMENTO DI ECONOMIA E DIRITTO

How can Benford's law help in fraud detection?

Remark

Many **real data** obeys Benford's law, but **fake data** do not!!!

$\longrightarrow$ Significant deviations from Benford's Law could indicate potential fraud. For example, if certain digits are overrepresented or underrepresented, it might suggest manipulation.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

How to measure deviations from Benford's law?

Consider a set of data and let

- n_d be the observed number of times that the digit d (for $d = 1, \ldots, 9$) appears as first digit within the data;
- p_d be the corresponding relative frequency, i.e. $p_d = n_d/n$, where $n = \sum_{d=1}^{9} n_d$;
- π_d be the theoretical proportion of times that the first digit is d according to the Benford's law

Then, different measures of non-compliance of the set of data with the Benford's law can be computed.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

- Chi-square statistic:

$$\chi^2 = \sum_{d=1}^{9} \frac{(n_d - n\pi_d)^2}{n\pi_d}$$

≤ 15.51 conformity
> 15.51 non conformity

- Mean Absolute Deviation (MAD):

$$MAD = \frac{1}{9} \sum_{d=1}^{9} |p_d - \pi_d|$$

≤ 0.006 close conformity
$0.006 - 0.012$ acceptable conformity
$0.012 - 0.015$ marg. acc. conformity
≥ 0.015 non conformity

- Sum of Squared Deviations (SSD):

$$SSD = \sum_{d=1}^{9} (p_d - \pi_d)^2$$

≤ 100 conformity
> 100 non conformity

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Some examples of applications

- Financial adviser Wesley Rhodes was convicted of defrauding investors. Statements that he sent them failed the first digit test and closer examination revealed that he had faked the data.
- Golbeck (2015) uncovered a Russian bot network on Twitter. She observed that for most users, the number of followers that their followers have adheres to Benford's law, but artificial accounts significantly veer from the pattern.
- Mebane (2010) used Benford's law to claim that the Iranian presidential election in 2009 was rigged.
- Rauch et.al. (2017) allege that Greece manipulated macroeconomic data in its application to join the eurozone, as they did not conform to Benford's law.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Table of Contents

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Some figures on tax evasion in Italy

- Tax and contribution evasion in 2021 amounted to 83.6 billion euros, of which approximately 73.2 billion were from lost tax revenues and 10.4 billion from lost contribution revenues.
- Out of the 73.2 billion of evaded taxes, about 30 billion are from IRPEF (Personal Income Tax) for self-employment and business, 18.1 from VAT, 8.5 from IRES (Corporate Income Tax) and 4.7 from IRAP (Regional Tax on Productive Activities).
- The principal tool available to tax authorities to enforce compliance is firms' auditing and fining in case of evasion. However, visiting or summoning the taxpayer to verify the information reported, the invoices, and other related documents, is quite a costly operation.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Firms' evasion propensity scores based on Benford's law

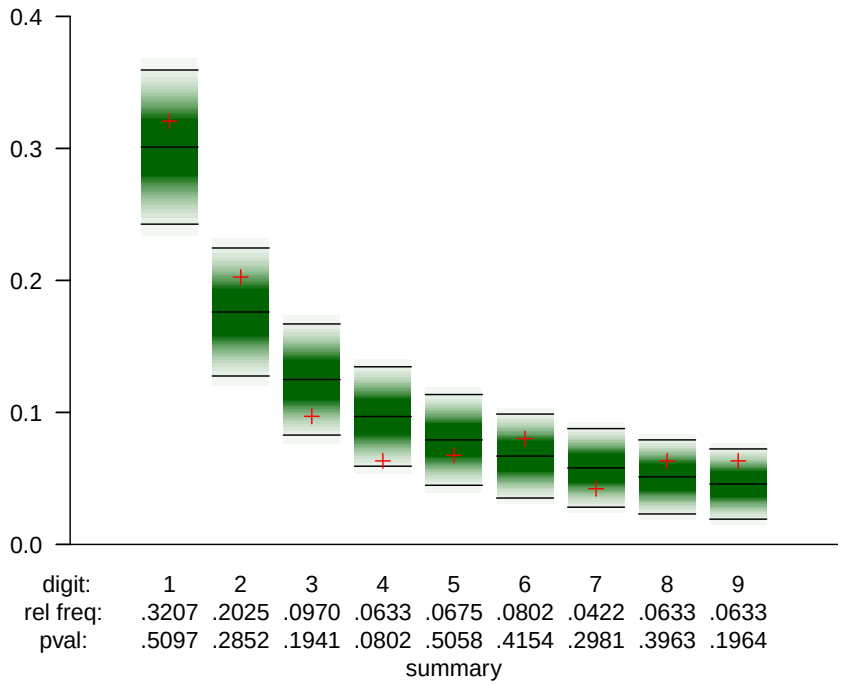
- We investigate the possibility of exploiting Benford's law on firms' balance sheet to construct an evasion propensity score at single firm level.
- We use the Bureau Van Dijk AIDA database of Italian firms' financial statements.
- Each financial statement (of industrial, commercial, and service companies) contains 243 line items (including sub-items and totals) of which 163 come from the balance sheet and 80 from the income statement.
- All duplicated and missing values are excluded from the analysis.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

A single firm example

As an illustrative example, we consider the firm SNEM, a S.P.A. of Business Process Outsourcing, located in Rome, and scoring three stars on the legality rating from AGCM

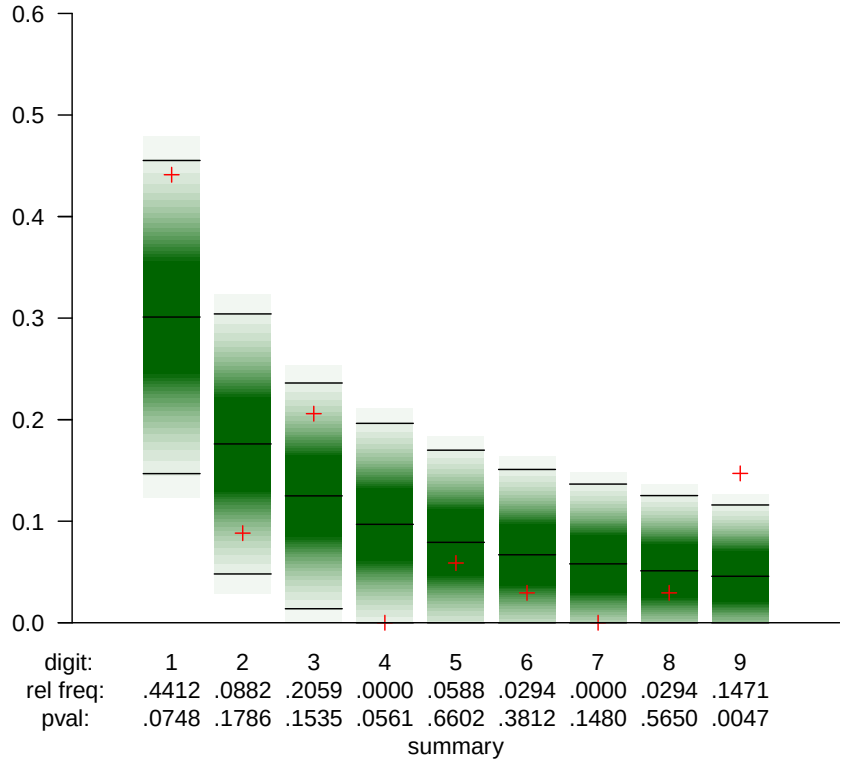
- $\chi^2 = 9.81$, p-value $= 0.279$
- $MAD = 0.0198$, p-value $= 0.146$
- $SSD = 0.0040$, p-value $= 0.295$



Another single firm example

As a second example, we consider the firm SCEB, an S.R.L.S. of accounting, auditing, bookkeeping and tax consultancy, located in Rome, recently accused of criminal organisation (Repubblica, February 29, 2024).

- $\chi^2 = 9.81$, p-value $= 0.279$
- $MAD = 0.0198$, p-value $= 0.146$
- $SSD = 0.0040$, p-value $= 0.295$



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

A control sample . . .

- As a **control sample**, we consider firms having obtained a three stars legality rating from AGCM (Agenzia Garante della Concorrenza e del Mercato, 2023).

$\longrightarrow$ The legality rating is a synthetic indicator of compliance with high standards of legality by companies that have applied for it and, more generally, of the level of attention paid to the proper management of their business.

$\longrightarrow$ This recognition takes the form of a score ranging from a minimum of one star to a maximum of three stars.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

. . . and a test sample

- As a **test sample**, we consider all firms with registered offices in Calabria and operating in the economic activity sector of Agriculture, forestry, and fishing.

$\longrightarrow$ Calabria is the region where the weight of the unobserved economy is the highest, accounting for 18.8% of the total added value (ISTAT, 2023).

$\longrightarrow$ The economic sector of Agriculture, forestry, and fishing is one of those where the weight of the unobserved economy is highest, accounting for 16.9% of the total added value (ISTAT, 2023).

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Results

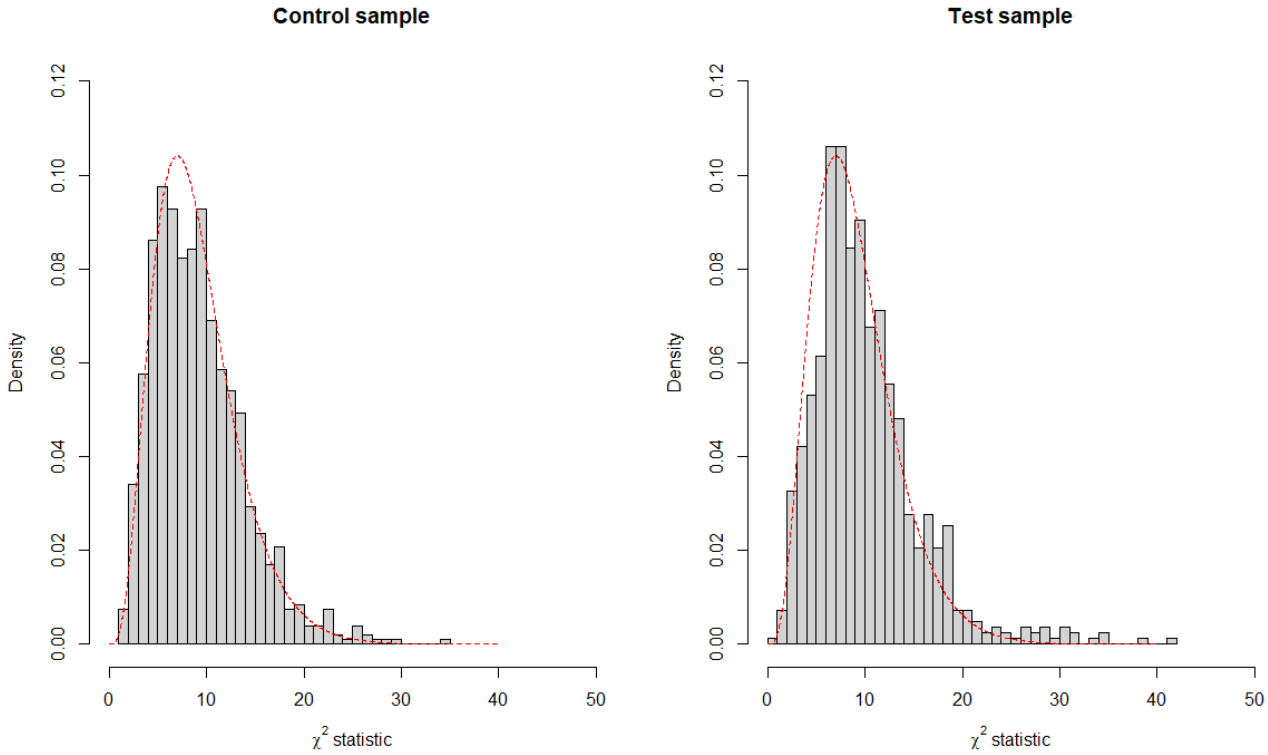


Figure: χ^2 statistic distribution for the control and test samples. Dashed red lines represents the distribution in case of compliance with Benfords' Law.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

	Number of firms	χ^2	MAD	SSD
Control sample	1057	0.091^{***}	0.090^{***}	0.092^{***}
Test sample	830	0.127^{***}	0.125^{***}	0.135^{***}
Difference	-227	0.037^{***}	0.035^{***}	0.043^{***}

Table: Estimated proportion of non compliant firms, according to the different measure of discrepancy

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Conclusions

- We explored the possibility of using Benford's law to detect financial statement manipulation, which would be helpful in:
  - targeting auditing and anti-fraud efforts
  - enhancing a better geographical, sectoral, and dimensional mapping of firm evasion
  - finding empirical support for theoretical evasion models
- Some promising results show that the Benford's law identifies a significantly large proportion of possibly manipulated financial statements in one of the Italian regions and economic sectors most affected by tax evasion.
- "Labelled" data on compliant and non compliant firms would be required to better determine the potential of Benford's law as a classifier in this field.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO

Bibliography

- Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, Vol. 78, pp. 551-572.
- Golbeck, J. (2015). Benford's Law Applies to Online Social Networks. PLoS ONE 10(8): e0135169
- Mebane, W.R. (2010). Fraud in the 2009 presidential election in Iran?. CHANCE 23, pp. 6–15.
- Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics 4, pp. 39-40.
- Rauch, B., Göttsche, M., Brähler, G. and Engel, S. (2017). Measuring the Quality of European Statistics, in Steven J. Miller (ed.), Benford's Law: Theory and Applications. Princeton, NJ.

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
unimc UNIVERSITÀ DI MACERATA
DIPARTIMENTO DI ECONOMIA E DIRITTO