



Data science for assessing possible tax income manipulation: The case of Italy



Marcel Ausloos^{a,b}, Roy Cerqueti^{c,*}, Tariq A. Mir^d

^a Institute of Accounting and Finance, School of Management, University of Leicester, University Road, Leicester LE1 7RH, UK

^b GRAPES, rue de la Belle Jardinière, B-4031 Liège, Federation Wallonie-Bruxelles, Belgium

^c University of Macerata, Department of Economics and Law, via Crescimbeni 20, Macerata I-62100, Italy

^d Nuclear Research Laboratory, Astrophysical Sciences Division, Bhabha Atomic Research Center, Srinagar 190 006, Jammu and Kashmir, India

ARTICLE INFO

Article history:

Received 5 July 2017

Accepted 15 August 2017

JEL classification:

H71

C82

Keywords:

Data science

Benford law

Aggregated income tax

Data manipulation

Italy

ABSTRACT

This paper explores a real-world fundamental theme under a data science perspective. It specifically discusses whether fraud or manipulation can be observed in and from municipality income tax size distributions, through their aggregation from citizen fiscal reports. The study case pertains to official data obtained from the Italian Ministry of Economics and Finance over the period 2007–2011. All Italian (20) regions are considered. The considered data science approach concretizes in the adoption of the Benford first digit law as quantitative tool. Marked disparities are found, - for several regions, leading to unexpected “conclusions”. The most eye browsing regions are not the expected ones according to classical imagination about Italy financial shadow matters.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

This paper deals with the relevant theme of identifying the existence of anomalies in tax incomes. We specifically focus on the case of Italian regions. The problem is faced under a data science perspective, which is suitable for the scope of the study. Indeed, data science represents nowadays a major area in the research frontier for processing large sets of data (see e.g. [17] and references therein contained).

The relevance of the study lies in the evidence that assessing the errors in financial statements is a major task of auditors, regulators, or analysts not only in financial markets, but also in macroeconomic and public affairs, like on governmental economic data. Reports of accurate financial statement data are crucial, even essential, to the management of public budgets. Thus, it is mandatory to observe whether misestimations, mistakes, biases, or even manipulations have occurred or are occurring [60]. On the other hand, academic researchers must propose ways for detecting errors or anomalies. Many methods have been proposed and steps taken in creating and validating techniques to assess different con-

structs of errors [23]. However, despite substantial progress, in this safety area, available methods present deficiencies that limit their usefulness, - sometimes due to unclear hypotheses underlying the method. Most likely, this will continue for ever, since it is well known that the imagination of crooks leads to further more sophisticated manipulation, while reaction of “policy makers” is impaired by legal processes. Yet, controls are challenged by intelligent people, lacking classical ethics.

Without suggesting opprobrium on all Italian citizens because of supposed to be tax evasion by a few, before debating individual cases, it is often admitted that Italy is one of the top countries losing to tax evasion (after the USA and Brazil) through a GDP ranking (<http://investorplace.com/investorpolitics/10-worst-countries-for-tax-evasion/#.Vvkqf3AR54c>), or through the amount of tax loss as a result of shadow economy. Income manipulation might thus be rightfully tested on such a (country) case, - as somewhat in line with recently discussed pertinent topics, from points of view related to ethics and organized crime, by e.g. [4,15,16], or [46].

Of course, not only citizens falsify their financial data, but also firms and even governments [2,41,43]. For example, questions have been raised about the data submitted by Greece to the Eurostat to meet the strict deficit criteria set by the European Union (EU), - see [62], or about the macroeconomic data of China (Holz [37]).

* Corresponding author.

E-mail addresses: ma683@le.ac.uk, marcel.ausloos@ulg.ac.be (M. Ausloos), roy.cerqueti@unimc.it (R. Cerqueti), taarik.mir@gmail.com (T.A. Mir).

Table 1

Frequency (“Freq.”) of the first digit (d) in a set of data; d values ranging from 1 to 9, - according to BL1, Eq. (1).

d	1	2	3	4	5	6	7	8	9
Freq.	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

Managers can engage in more or less corporate tax avoidance than shareholders would otherwise prefer [6]. On the other hand, firms have incentives to manipulate earnings in order to convince investors, e.g. to report a rounded to an upper value number when they have profits (i.e., USD 40 million) and to report a number such as USD 39.95 million, when they have losses, - as discussed by [65], having observed such unusual patterns in reported earnings. This rounding approach points to a “moderate manipulation” of the data. However, its relevance is considered not to be negligible for investors.

At the “lower level”, that of citizens, it is also known that seemingly small rounding manipulations can influence financial statement users’ perception of credit quality [34]. At another level, that where the citizen is immersed in a crowd, and expects to be protected by some shadow due to a bigger cheater, it is interesting to raise the question whether a collective effect can be seen. This can be done through examining income tax contributions at local levels. This accounting level is the core of our investigation and report.

A review of statistical methods of fraud detection has been provided by Bolton and Hand [13,14], while accountants’ perceptions regarding fraud detection and prevention methods have been recently discussed [12]; see also [68] for a quick summary or [39] for a specific discussion of a couple of techniques.

In this context, the data science approach proposed in the present paper is based on the Benford law.

This Benford law, originally for the first digit (BL1) distribution of data sets, follows a logarithmic law:

$$P(d) = \log_{10} \left(1 + \frac{1}{d} \right), \quad d = 1, 2, \dots, 9, \quad (1)$$

where $P(d)$ is the probability that the first digit is equal to d in the data set; \log_{10} being the logarithm in base 10.

This “law” stems from observations by [47] and later independently by [11] that the distribution of the 1st digit is more concentrated on smaller values: the digit “1” has the highest frequency, “9” the lowest frequency. In Table 1, the frequency of the first digit, as given by BL1, is recalled for the reader convenience. Thereafter, mathematics can suggest empirical law for the 2nd, 3rd, & digit distribution. However, since the latter becomes quickly rather uniform, it becomes hard (but it is done) to use such high level digits for testing the validity of reported data. Thus, let us concentrate our aim below to the first digit, i.e. on the validity (or not) of BL1 in a specific case, serving as a paradigm for other big data investigations.

In general, Benford law has to be recommended because it contains many advantages like not being affected by scale invariance, and is admitted to be of help when there is no supporting document to prove the authenticity of the “transactions” [67]. Nowadays, this so called “law” provides a convenient basis for digital analysis of sequences of numbers of similar nature. For example, an analysis based on Benford law has been used in a wide variety of ways to identify instances of employee theft and tax evasion [34] and also deviation of the exchange rates from regular paths [18] or of the Libor rates [1].

In fact, following [65], a “manipulation expectation” can be obtained using the [11] law, as was later pointed out by [48–50].

Since [51], it is admitted that BL1 can be used to detect fraud in accounting data reporting individual incomes. The presentation

and demonstration of the Newcomb–Benford law (1881–1938), as a powerful methodology in the audit field, were further emphasized by [29,36,58,61], among others, and also recently in [21], Nigrini and Miller [5,8,22,45,57]. In fact, BL1 is also applied outside the financial audit realm; e.g. see [31] for image forensics, [7] for birth rate anomalies, [59] for maternal mortality rates, or elsewhere in the natural sciences [63], and on religious activities [44]. We notice that the literature is huge: see [3], for approximately the last decade, [25], or [10], which contains a rather exhaustive list of references.

Limiting ourselves at the public government financial realm data tampering, let us mention, among others, the application of Benford law to selected balances in the Comprehensive Annual Financial Reports of the fifty states of the United States [35,38]; or the “political economy of numbers” international macroeconomic statistics [52]. A study on the analysis of the digit distribution of 134281 contracts issued by 20 management units in two states, in Brazil, also found significant deviations from Benford’s law [24]; see also [33]. Fraud detection (and prevention methods) in the Malaysian Public Sector have been discussed in [53].

At a lower scale, - ours if it has to be recalled, using Benford’s law has permitted to uncover deficiencies in the data reported by local governments, like municipalities and states in several countries, or example, USA or Brazil. The digit distributions of the financial statements of 3 municipalities, Valejo City, Orange County and Jefferson County, have been shown to have significant departures from that expected on the basis of BL1 [35].

In Italy, tax collection is a fundamental source of revenues for local governments, enabling the efficient delivery of services [54]. On the other hand, tax evasion is known to be widespread across Italy ([15,30], Marino and Zizza [20,32]).

Obviously, any financial distress of municipalities, resulting from income tax evasion, has severe repercussions on the lives of the taxpayers and municipal employees [9]. This is annoying for the collectivity, thus it seems important to have some better oversight of the quality of financial statements and accountability in view of the demand (and use) of funds, say returning from the Italian government. Moreover, these concerns on data quality, on one hand, and the admittedly poor auditing procedures being used in Italy, on the other hand, have resurfaced vigorously following the bankruptcy of a number of local government bodies during recent financial crisis [54], - in fact, to be fair, more generally, across several industrialized countries.

Within this review of specific accounting features relevant to our research, it might be finally interesting to point to the reader a very recent and specific (by “chance”) Italian case, i.e. the detection of anomalies in receivables and payables in Italian universities, by [19]. This shows that intermediary levels of financial data scales may contain intriguing features, whence further suggesting to raise questions on the detection of manipulations, through deviations from Benford law, as here the case of tax incomes in e.g. regions.

Thus, we have considered the aggregated values of the income tax reports of each of the 20 IT regions, - over a recent quinquennium: [2007–2011] for which the data is available. As suggested by [42] we calibrate our analysis with a χ^2 test.

Our point of view is politico-economic unique: the accounting reliability of the citizen contributions to the IT GDP, - even though questions are numerous. Hopefully, within this framework, one can also (i) enlarge the knowledge of BL1 application range, (ii) contribute to a better application of BL1 in accounting, and even (iii) indicate that one can reach socio-economic-political conclusions.

The paper is organized as follows: Section 2 is about the methodology. Section 3 contains the description of the data. The

Download English Version:

<https://daneshyari.com/en/article/5499479>

Download Persian Version:

<https://daneshyari.com/article/5499479>

[Daneshyari.com](https://daneshyari.com)