ARTICLE IN PRESS

Physica A ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries

Jing Shi [a], Marcel Ausloos [a,*], Tingting Zhu [b]

[a] School of Business, Ken Edwards Building, University of Leicester, Leicester, LE1 7RH, United Kingdom
[b] Faculty of Business and Law, Hugh Aston Building, De Montfort University, Leicester, LE1 9BH, United Kingdom

ARTICLE INFO

Article history:
Received 1 October 2017
Available online xxxx

Keywords:
Benford's law (BL)
Outliers
Distributions distance to BL

ABSTRACT

We discuss a common suspicion about reported financial data, in 10 industrial sectors of the 6 so called "main developing countries" over the time interval [2000–2014]. These data are examined through Benford's law first significant digit and through distribution distances tests.

It is shown that several visually anomalous data have to be *a priori* removed. Thereafter, the distributions much better follow the first digit significant law, indicating the usefulness of a Benford's law test from the research starting line. The same holds true for distance tests. A few outliers are pointed out.

1. Introduction

The Benford's law is a phenomenological law about the probability distribution of the first significant digits (henceforth FSD) in a data set [1]. Newcomb (in 1881) and later Benford (in 1938) [2,3] observed that the occurrence of the first significant digit in many data sets is not uniform, but tends to follow a logarithmic distribution such that the smaller digits appear as a first significant digit more frequently than the larger ones, according to

$$N_d = N\ log_{10}(1 + \frac{1}{d}), \quad d \equiv 1, 2, 3, \ldots, 9 \tag{1}$$

where N is the total number of considered 1st digits, in short, the number of data points, and N_d is the number of the so observed integer d ($= 1, 2, 3, \ldots, 9$) being the starting one (1st) in the data set list [4].

The law, Eq. (1), nowadays called Benford's law (BL), is widely applied in the investigation of data manipulation by researchers in finance and economy [5–7]. The BL can be used not only to identify falsely created data, e.g. in corporations financial statements as shown by Nigrini [8], but also to verify the (non)reliability of macroeconomic data [9]. Furthermore, because of its special features, BL has been employed as a quality test criterion for various employed data. An extensive bibliography, from 1881 up to 2006, on Benford's law papers including theories, applications, generalizations and warnings can be found in a Hürlimann unpublished work but available on internet [10]; see also books by Kossovsky for more recent reviews [11,12].

In particular, the Benford's law test (henceforth BL test) has been applied to many financial numerical series. Concentrating on stock market indexes, let us mention (in a chronological order) Ley (in 1996) [13] with a Bayesian approach observed

* Correspondence to: GRAPES, 483 rue de la belle jardinière, B-4031 Liège, Belgium.
E-mail addresses: 248299825@qq.com (J. Shi), ma683@le.ac.uk, marcel.ausloos@ulg.ac.be (M. Ausloos), tingting.zhu@dmu.ac.uk (T. Zhu).

https://doi.org/10.1016/j.physa.2017.11.017

the distribution of U.S. stock indexes' digits: the series of one-day returns, for the Dow Jones Industrial Average (DJIA) and the Standard and Poor's Index (S&P), were found to follow Benford's Law. Ley interestingly concluded that distributions that follow Benford's Law are distributions where small changes are more likely to occur than large changes. The FTSE[1] 100 was examined by De Ceuster et al. (in 1998) [14].

After about a 10 years lapse with no investigation of the sort, Krakar and Žgela (in 2009) studied the Zagreb Stock Exchange and found out that the closing prices did not follow BL [15].

Corazza, et al. (in 2010) checked the S&P 500 [16], while Zhao and Wu (also in 2010) wondered whether the Shanghai Stock Exchange Composite index and the Shenzhen Stock Exchange Component index agree with Benford's Law, finding that Benford's law reasonably holds for these two (main) Chinese stock indexes [17].

Žgela (in 2011) analysis of DAX[2] percentage changes over 10 years [2001–2011] led to the conclusion that DAX values were not in accordance with the FSD Law [18]. More recently, the Istanbul Stock Exchange (BIST) attracted some attention: Karavardar (in 2014) [19] found no disagreement with BL for the monthly returns (over 26 years), while Cinko (2014) found no disagreement for the daily returns (over 23 years) [20].

Under the efficient market hypothesis, the stock price index should be completely controlled by the market, thus would not be manipulated by human (or government) intervention. Under this situation, we could use the BL test to detect if faults exist in the price indexes.

For completeness, let us mention opinions wondering whether BL has to be ever applied, e.g. as discussed by Mebane [5] or Durtschi et al. [21]. An argument on the non-reliability of detecting fraud by BL is based on the knowledge of BL by potential swindlers, aware of the necessity of conforming to the BL. In fact, such a possibility might not be surprising, or is even sustained, in view of findings on income taxes regularity in some Italian provinces by Mir et al. [7,22]. Such an argument suggests to consider alternative methods complementing the BL classical analysis [23–25].

We propose here a data aggregation method (see also Mir et al. [7]) which could prove to be useful, especially in view of the fact that it would be particularly difficult to fabricate data conforming to BL by various unconnected agents along this scheme.

We have conducted the BL test in order to evaluate whether the collected Financial Times Security Exchange (henceforth FTSE) global price index of some financial data possibly contains error values. We have examined the case of monthly log-returns for six (top—from a Gross Domestic Product (GDP) point of view) emerging countries, over 10 different industrial sectors—over 15 recent years, as reported by FTSE. There is no study to our knowledge, about data from emerging (or developing) countries, on the scale endeavored here below, but we mention that the GDP growth rate was studied for Germany by Gunnel and Tödter, in 2009, [26].

Therefore, the present paper goes as follows: after quickly outlining the BL test, in Section 2.1, we complement the methodology introducing a "complementary distance study" Section 2.2. In Section 3, after explaining the countries selection (Section 3.1), the raw data acquisition is recalled, i.e. monthly log-returns, next aggregated over 6 relevant emerging countries for a 15 year time interval (Section 3.2). The data of interest is displayed through histograms in Section 3.3.

We have performed the χ^2 tests in order to compare the BL to the observed distributions. The results appear unsatisfactory. However, a visual review indicates anomalous values in the raw data. We have removed them and redone χ^2 tests with respect to BL test values; a comparison of results before and after removal of anomalous data is presented in Section 3.4.

A discussion, practically oriented on BL and on the distribution distance to BL is found in Section 4.1. In addition, since we also investigate a complementary quantity, the distance between the 10 industry price indexes FSD distributions and BL is discussed in Section 4.2. Some synthesis allows for a discussion in Section 5. Section 6 serves for a conclusion.

2. Methodology

2.1. Benford's law

Benford's law, Eq. (1), is known as the first digit law or the law of the leading digits. According to Eq. (1), in a given data set the probability of occurrence of a certain digit as the first (1st) significant digit decreases logarithmically as the value of the digit increases from 1 to 9. Thus, digit 1 should appear as the first significant digit about 30.10% times, and similarly 9 should appear about 4.58% times.

Benford's law, Eq. (1), holds for data sets in which the occurrence of numbers is free from any restriction; significant deviations from the Benford distribution may indicate fraudulent or corrupted data [27].

Deviation from the expected Benford's law distribution is calculated from the χ^2 statistic according to

$$\chi^2_{1BL} = \sum_{i=1}^{9} \frac{(e_{i1} - b_{i1B})^2}{b_{i1B}}, \tag{2}$$

where e_{i1} is the observed frequency of each FSD in the price index data; b_{i1B} is the frequency expected from Benford's law. The 10%, 5%, and 1% critical values for χ^2, with 8 degrees of freedom, are 13.36, 15.51, and 20.09.

[1] Financial Times Security Exchange.

[2] Deutscher Aktien IndeX.

Please cite this article in press as: J. Shi, et al., Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries, Physica A (2017), https://doi.org/10.1016/j.physa.2017.11.017.