Regular paths in financial markets:

investigating the Benford’s Law

Jessica Riccioni & Roy Cerqueti[[1]](#footnote-1)

University of Macerata - Department of Economics and Law

Via Crescimbeni 20, Macerata I-62100, Italy

**Abstract**

This paper aims at verifying whenever the Benford’s Law is valid in the context of global stock markets, which are here viewed as complex systems. In so doing, we pursue the scope of assessing the presence of data regularities and interpret obtained discrepancies.

Specifically, we check the reliability of Benford’s Law for all the indexes listed on the stock exchanges of several countries, with a particular reference to prices and volumes of stocks.

To pursue our scope, we adopt comparison criteria grounded on statistical theory, like the Chi-squared test for both the distributions of the first and the second meaningful digits.

Evidence of violations is provided and some insights taken from the historical facts and economic shocks are carried out.

**Keywords:** Data science; Benford law; Stock markets; Finance.

**1. Introduction**

The evolution of prices and volumes in the international stock markets has a remarkable informative content. In fact, financial markets and their components are highly correlated with the surrounding socio-political environment and a shock of noneconomic nature might have an effect on the economic system.

This paper moves from this scientific ground. We aim at analyzing the context of global stock markets with a specific focus on all the indexes composing them. In particular, we deal with the nature of complex system of the markets, and discuss whenever the available time series associated to prices and volumes of the stocks exhibit a sort of “regularity”.

Indeed, datasets exhibit often deviations from a pure uniform law, even if they are not constructed in an “ad-hoc” manner. Think about countries’ GDPs, with a great percentage of low levels and a few data with high values.

In this respect, the so-called Benford’s Law (BL), firstly empirically observed by Newcomb (1881) and lately formalized by Benford (1938), has been proved to hold in a wide set of contexts.

BL is an evident regularity of the distribution of the digits of a number of large datasets. It states that the frequency of the first and second digits of the values of a set of data decreases with the value of the digit, and achieves its maximum when the digit is “one”.

More in details, Benford (1938) provided an extensive study of 20,149 data from 20 different areas, and calculated the frequency of each first digit from 1 to 9 and of each second digit from 0 to 9.

The outcome of this effort has led to the statement of an exact rule given by a logarithmic relationship.

For the first digit, BL says that:

(1)

where is the probability that a number has the first digit equals to , with .

As for the probability of the second digit, BL can be written as follows:

. (2)

where .

The theoretical frequencies are shown in Table 1.

Table 1. Relative frequencies of the first and second significant digit

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1st digit |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Frequency |  | 0.301 | 0.176 | 0.125 | 0.097 | 0.079 | 0.067 | 0.058 | 0.051 | 0.044 |
| 2nd digit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Frequency | 0.119 | 0.114 | 0.109 | 0.104 | 0.100 | 0.097 | 0.093 | 0.090 | 0.088 | 0.085 |

Generalizations of formulas (1) and (2) have been also provided, leading to the distributions of the digits subsequent to the second and obtaining frequencies more equally distributed. From the fifth digit the distribution no longer follows the BL but becomes a uniform distribution.

Not all dataset are eligible to be controlled for agreement with the BL. In particular, data must be not chosen under an “ad hoc” criterion and the cardinality of the dataset must be large. Such requirements are fulfilled in our specific context.

We check the validity of the BL for extensive daily series of prices and volumes of 4166 stocks, listed in the main international financial markets. Several clusterings of such series are performed at the level of the related geographic areas (Italy, Europe, USA, Asia and Pacific area and a generic “other relevant markets”) and of the individual stock markets.

The beginning of the reference period varies with the stock, on the basis of the free availability of the data on the yahoo.com/finance website. The ending date is for all November 14th, 2014.

Both first and second digits distributions are explored.

The analysis is performed by the application of a Chi-squared procedure, to check whenever the empirical data obeys with the theoretical frequencies of BL.

The analysis reveals the presence of deviations of the data digits from the statements of the BL. In accord to the literature (see the next Section), they might be caused by shocks of different nature which are able to influence the dynamics of the international stock markets components. A discussion of such deviations and also of the regularities is provided, along with a related socio-economic-political contextualization.

The rest of the paper is organized as follows: next Section contains a critical review of the main contributions of the literature. Section 3 is devoted to a detailed description of the data used and of the methodological instruments employed to perform the analysis. Section 4 outlines and comments the obtained results, and proposes a reading of the deviations from the statements of the BL through the occurred events. Last Section offers some conclusive remarks and gives some suggestions for future research lines.

**2. Relevant literature**

Here are briefly described the cases in which this law has been applied with success over the years.

In the context of social science, Cho and Gaines (2007) took care of fraud in funding for charity campaigns. See also Mir (2012, 2014) and Ausloos et al (2012, 2015), who tested the consistency of social data and assessed the presence of some sort of manipulation.

In the political environment Mebane (2011) used the law for verify the trustworthiness of the election results. In this respect, the law was also used by the Iranian minister Boudewijn F. Roukema to ensure the accuracy of the electoral results in the presidential election of 2009, as a consequence of the complaint about electoral fraud made by Moussavi (opposition leader) against the election of President Ahmadinejad. City after city, the minister analysed the relative frequencies of the first significant digit of the number of votes and compared the frequencies obtained with the theoretical frequencies of BL. As a result Roukema remarked that while all the frequencies of the other digits were very similar to those theoretical, the number 7 appeared too many times, hence supporting the truthfulness of the charges of Moussavi. Decided to settle the matter, the minister remarked that the anomaly concerned three out of the six larger areas of Iran. In these areas the winner Ahmadinejad had won a share of the vote much higher than in other areas.

Also Varian (1972) suggested the possibility of using BL to identify any falsification in the collection of data used to support political decisions. The author proposed to compare the relative frequency of the first digit of the numbers used to the support political decisions with the theoretical frequencies of BL, so as to highlight any abnormal results. He applied the law in a set of land planning for 777 tracts in San Francisco Bay. The law was respected.

Nigrini (1996, 1999, 2012) was able to show the benefits of the possible application of BL to detect the falsification and the frauds in accounting and activity auditing (which of course may not be intentional, such as arithmetic errors and, more generally, errors in calculation or misapplication of the applicable accounting standards; or it could be outright intentional frauds, such as the alteration of records or documents, the lack of enforcement of accounting standards, the omission of some results or finally recording non existent transactions). This system is now used in the majority of American States to support the identification of tax fraud.

In the economic field BL can also be applied to check the efficiency of financial indexes, and in general financial quantities. Ley (1996) calculated if the distribution of the relative frequencies of the first significant digit of daily returns of two American stock indexes (ie the S & P for the period from 1926 to 1993 and the Dow Jones for the period from 1900 to 1993), using the BL: the result obtained that both stock indexes respected this law and it was a demonstration of the current index efficiency.

Realdon (2008) and Ausloos et al. (2016) dealt with the assessment of the validity of the BL on the sovereign credit default swap markets. Carrera (2015) analysed exchange rates to check whether there exists the possibility of manipulation for policy management. In support of macroeconomic data, Nye and Moul (2007) studied the GDP data of the Penn World Tables. Günnel and Tödter (2009) applied the BL on the forecast of GDP growth and inflation of German consumer prices. Ranch (2011) examined the abnormal data of national and financial accounts of the EU countries from 1999 to 2009. Holz (2014) evaluated the quality of the Chinese GDP advancing suspicions on the statistics published by the National Bureau of Statistics of China. Michalski and Stoltz (2013) examined the balance of payments data for 103 countries between 1989 and 2007. Mir (2016) studied the illicit financial outflows from developing countries. Finally, Mir et al. (2014) and Ausloos et al. (2017) have instead applied the BL to the aggregate of income taxes of municipalities and Italian regions for the period between 2007 and 2011.

In this work we have tried to interpret the efficacy and functionality of financial markets worldwide. We checked the reliability of BL for all stock indexes listed on the stock exchanges of the various countries but not to apply it on daily returns but on prices and average volumes of stocks comprising indexes.

We are in the line traced by other scholars dealing with the assessment of BL in financial markets (see e.g. De Ceuster et al., 1998; Corazza et al., 2010 and Karavardar, 2014). However, this is the first paper dealing with a so global context, with all the international stock markets and both the key stock variables (volumes and prices).

It will be interesting to see how this application of BL analyses and comments consistent results; in fact, the cases in which the Law will be verified are various, and we provide a brief discussion on the technical and historical reasons leading to the failure of this law.

**3. Data analysis and methodology**

**3.1 Data**

All processed data are freely available on the web (see e.g. <https://finance.yahoo.com>). The dataset is of scientific reliability, as also witnessed by authoritative financial studies who had referred to such a dataset for collecting empirical data (see e.g. Antweiler and Frank, 2004; Juergens and Lindsey, 2009).

We use all the indexes listed on the stock exchanges of the various countries whose time series are available, with a particular reference to daily adjusted closing price and volume of stocks that compose the financial indexes.

The period under investigation is the available one, and depends on the individual stocks. Indeed, after a careful analysis, we have noticed that listing days of an asset in a Stock Exchange are different. Thus, the starting points of the time series vary. Some stocks have data available from 1962 (e.g. The Walt Disney Company, The Coca-Cola Company, both listed in the 65 Composite US Index) while other ones have only recent data, from 2014 (e.g. Cerved Information Solutions SPA, listed in the FTSE Italia MIDCAP and Emas Offshore Regs, listed in the OSE All Share of the Norwegian Stock Exchange). The period under investigation ends for all the stocks in November 14th, 2014.

The outcome of the data collection procedure is thousands of time series which reflect the recent overall activity of the stock markets.

The database have been divided into five macro areas: the Italy, the Europe, the Usa, the Asia and the Pacific Area and finally the area called “Other relevant markets” composed by indexes not included in the previous groups but still relevant.

For meeting the requirement of the Benford Law –which represents the object of our analysis- we have restricted the analysis to the individual indexes with more than 100 components. A further assessment of the validity of the BL has been implemented also on the overall sample collecting all the indexes, in order to give an interpretation of the global equity markets. Data of a few components are not available, and they are not included in the analysis.

The macro areas with relative stocks available have been grouped as follows: Italy(476 stocks available), Europe (451 stocks available), Usa (2759 stocks available); Asia and Pacific area (364 stocks available), other relevant markets (116 stocks available).

We now enter the details on the considered dataset.

For Italy, we have a number of relative indices. Specifically, we have FTSE ITALIA MICRO CAP (3 components); FTSE ITALIA SMALL CAP (126 components); FTSE ITALIA MID CAP (60 components including 2 not available); FTSE ITALIA STAR (69 components); FTSE MIB (40 components including 1 not available); FTSE ITALIA ALL SHARE (183 components including 2 not available).

For Europe, the considered indices are BEL-20 (20 components, but only 3 available); OMX COPENHAGEN (2 components); CAC 4O (40 components including 1 not available); DAX (30 components); OSE ALL SHARE (163 components including 6 not available); STOCKHOLM GENERAL (23 components including 1 not available); SWISS MARKET (20 components including 6 not available); FTSE 100 (94 components including 1 not available); EUROSTOXX 50 (49 components including 6 not available); TECDAX (30 components). For what concerns the OSE ALL SHARE, an additional explanation is needed. Indeed, there are two components of such an index which have never been traded, so their average volume is zero. Such stocks are then removed in the analysis of the first significant digit of the volume, where the total amount of considered stocks is then 155 instead of 157.

For Usa, the considered indexes are DOW JONES COMPOSITE AVARAGE (65 components); DOW JONES INDUSTRIAL AVARAGE (30 components); DOW JONES TRANSPORTATION AVARAGE (20 components); DOW JONES UTILITY AVARAGE (15 components); NASDAQ COMPOSITE (2558 components including 36 not available); NASDAQ 100 (107 components).

For Asia and Pacific area we have the following indices: JAKARTA COMPOSITE (299 components including 63 not available); NXZ 50 GROSS INDEX (50 components including 1 not available); STRAITS TIMES INDEX (30 components including 1 not available); HANG SENG INDEX (50 components).

Lastly, we have also included in the analysis an area called “other relevant markets”, on the basis of the association of the indices to countries not belonging to the classes defined above. In particular, we have MERVAL BUENOSAIRES (13 components); BOVESPA (70 components including 1 not available); IPC (35 components including 1 not available).

Overall, the studied stocks are 4166, so that the considered time series (prices and volumes) are 8332.

**3.2 Methodology**

The analysis of all the components of each index has been made. The possible not compliance with BL has been contextualized in the socio-economic environment, to gain insights on the roots of violation on the basis of the main events occurred in the considered period.

First of all, we have computed the mean values over time of prices and volumes for all the stocks. BL has been verified both on the distribution of the first and of the second *significant* digits over the samples of mean prices and mean volumes.

To clarify the term *significant*, a brief example is provided: the first digit of 4.1562 is 4 and the second digit is 1; the first digit of 315 is 3 and the second digit is 1; the first significant digit of 0.0154 is 1 and the second digit is 5.

The empirical frequencies of the digits have been compared with the theoretical distribution given by the BL. To pursue our scope, we have adopted a comparison criterion grounded on statistical theory which is one of the most used in the literature for testing the BL: the Pearson’s test. Such a test is of parametric type and serves for checking if the empirical frequencies are statistically different from theoretical frequencies of the BL.

In our framework, this test can be written as

(3)

where:

* is the frequency actually detected for character ;
* is the theoretical frequency of character , according to the BL.

Table 2 contains the p-values corresponding to various significance levels for 8 and 9 degrees of freedom. Indeed, 8 degrees of freedom are needed for verifying the conformity of the first meaningful digit (n = 9 and n-1 = 8), so that . Differently, 9 degrees of freedom are the ones employed for verifying the conformity of the second meaningful digit (n = 10 and n-1 = 9); in this case ;.

Table 2. Distribution of p-values for 8 and 9 degrees of freedom, according to the significance levels.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Significance levels** | | | | | | | | | |
|  | **1,00** | **0,99** | **0,95** | **0,90** | **0,25** | **0,10** | **0,05** | **0,025** | **0,01** | **0,005** |
| **Degrees of freedom** | **p-values** | | | | | | | | | |
| **8** | 1,34 | 1,65 | 2,73 | 3,49 | 10,2 | 13,4 | 15,5 | 17,54 | 20,1 | 21,96 |
| **9** | 1,74 | 2,09 | 3,33 | 4,17 | 11,4 | 14,7 | 16,9 | 19,02 | 21,7 | 23,59 |

**4. Results**

This section contains the analysis and the discussion of the results obtained from the empirical experiment. In particular, we aim at interpreting the financial markets through the assessment of the validity of the BL, mainly when such a law is violated.

The study has been performed at different levels. First, we have dealt with the indexes with more than 100 components, in order to have consistency in the respect of the cardinality of the dataset. In this respect, we recall that a large dataset is a requirement for checking the validity of the BL. Second, we have carried out the analysis of the indexes aggregated in terms of geographical areas of the relative markets. Third, we have analysed all the indexes together to have a general picture of the world situation.

In the tables below (Tables 3 and 4) one can find the results of the test for the indexes with more than 100 components and for the geographic areas.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st digit volume | 1st digit price | 2nd digit volume | 2nd digit price |
| FTSE ITALIA SMALL CAP | 8,89418683 | 12,2156522 | 6,72392148 | 10,7901592 |
| FTSE ITALIA ALL SHARE | 17,7539262 | 4,28808661 | 4,95303644 | 7,41873505 |
| OSE ALL SHARE | 9,46096051 | 9,25376369 | 2,44966909 | 8,02013759 |
| NASDAQ COMPOSITE | 14,50280609 | 93,43717036 | 7,430743972 | 12,38956914 |
| NASDAQ 100 | 8,02891215 | 19,109046 | 4,78071238 | 13,0184766 |
| JAKARTA COMPOSITE | 8,788330727 | 15,27951537 | 5,742951162 | 12,88689165 |

Table 3. test performed for individual indexes with more than 100 components.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1st DIGIT VOLUME | 1st DIGIT PRICE | 2nd DIGIT VOLUME | 2nd DIGIT PRICE |
| Italy | 34,89306957 | 5,768783508 | 11,94656804 | 14,57381575 |
| Europe 1\* | 10,26652222 | 8,021512551 | 8,698752939 | 5,995530794 |
| Europe 2\* | 31,01624831 | 9,923093492 | 10,66715576 | 11,86938179 |
| Usa | 11,8702497 | 113,3609736 | 5,650480794 | 17,58469152 |
| Asia-Pacific area | 9,450189145 | 13,44280063 | 8,902870173 | 7,029834036 |
| Other relevant markets | 9,964310315 | 4,448626433 | 8,110803132 | 8,217477778 |
| World 1\* | 5,624102229 | 104,9908995 | 3,999015609 | 11,28583735 |
| World 2\* | 10,14766471 | 94,67223226 | 7,169797966 | 10,12147404 |

Table 4. test performed for indexes divided by geographic areas. Europe 1\* and World 1\* denote the cases in which Italy is considered only by the FTSE MIB, while Europe 2\* and World 2\* are associated to Italy considered by all indexes.

Both first and second significant digits are considered for prices and volumes.

There are some cases in which the test fails, and the null hypothesis of accordance with the BL is not verified. In the set of non conformity cases, we mention the 1st digit of volume of FTSE ITALIA ALL SHARE, Europe 2\* and Italy; the 1st digit of adjusted closing price of NASDAQ COMPOSITE, NASDAQ 100, Usa, Asia-Pacific area, World 1\* and World 2\*; the 2nd digit of adjusted closing price of Italy and Usa.

Differently, there are some cases in which the critical values of the test are below confidence levels of at least 50% up to 95%. In these cases we argue that the null hypothesis is accepted and conclude that data fit with BL.

Among the cases of conformity, we point attention to the 1st digit of volume of WORLD 1\*; the 1st digit of adjusted closing price of FTSE ITALIA ALL SHARE, Italy and Other relevant markets; the 2nd digit of volume of FTSE ITALIA SMALL CAP, FTSE ITALIA ALL SHARE, OSE ALL SHARE, NASDAQ 100, JAKARTA COMPOSITE, Usa, World 1\* and World 2\*; the 2nd digit of adjusted closing price of FTSE ITALIA ALL SHARE, OSE ALL SHARE, Europe 1\* and Asia-Pacific area.

It is worth to discuss the obtained findings.

For what concerns FTSE ITALIA ALL SHARE, which contains all the stocks listed in the Italian Stock Exchange Borsa Italiana, the distribution of the first digit of the prices is very similar to the theoretical distribution provided by Benford. In fact, the test gives the very low critical value of 4.288, and the validity of the BL is obtained for a confidence level slightly lower than 80%.

Such a conformity is confirmed also by a pure visual inspection of the interpolating curves of the distributions (see Fig. 1).

Fig. 1: The comparison between the real distributions of the 1st digit of adjusted price and the frequency distribution of Benford for the FTSE ITALIA ALL SHARE.

Also for the analysis of the World joining all the indexes together (considering Italy only through the FTSE MIB), a confidence level of 90% is obtained for the first digit of the volume.

From a visual examination of this chart, it can be note an excellent compliance between the two distributions (see Fig. 2).

Fig. 2: The comparison between the real distributions of the 1st digit of volume and the frequency distribution of Benford for World including FTSE MIB.

Now we focus our attention to the particular case of the OSE ALL SHARE. This index includes all the stocks with high capitalization listed in the Norwegian Stock Exchange.

By the calculation of the test for the second meaningful digit of the volume the obtained confidence level is 95% (the critical value is 2.45). This represents an extraordinary result.

In some situations, there is an excellent compliance of the BL with the empirical distribution of the digits even if the visual appeal of the concordance of the corresponding plots is questionable. Such a discrepancy depends on the cardinality of the considered dataset, since a small deviation might be graphically relevant in presence of a small number of data. This said, we point out that the scientifically consistent method for assessing the validity of the BL is the statistical test. (see Fig. 3)

In this respect, notice that the OSE ALL SHARE includes only 157 stocks, while in the analysis of the overall World the included stocks are 3727.

Fig. 3: The comparison between the real distributions of the 2nd digit of volume and the frequency distribution of Benford for the OSE ALL SHARE.

We now discuss some cases of not conformity, where BL fails.

The first case concerns Italy.

For the first digit of the volume, BL is violated (see Fig. 4).

Fig. 4: Graphical comparison between the real distributions of the 1st digit of volume and the frequency distribution of Benford for Italy.

Several interpretations can be given to the not conformity of first digit of the volume with the BL for Italy. Such motivations do not include an incompleteness problem of the dataset, since the database of Italy is the most complete one with very few data missing.

We rather present an historical perspective over the last 15 years, bearing in mind that the considered period for the Italian dataset begins in January 2000. Indeed, when the BL is not valid, there may be events that have "manipulated" the considered dataset.

Surely, the transition between Lira and Euro at the beginning of this century has created an instability which have caused a collapse in the Italian market until 2003, when a slow recovery started, till the market boom in 2007. Unfortunately, the subprime mortgage crisis in 2008 in the United States was around the corner, and it had a great impact in the Italian markets. These events are, in our view, those which has led to a distortion of a “normal random course” for Italian market, hence leading to the noncompliance with the BL.

Another case of not conformity concerns the first digit of the volumes of the macro area called Europe.

Looking at the graphical appeal, one can better grasp the failure of the BL (see Fig. 5).

Fig. 5: The comparison between the empirical distributions of the 1st digit of volume and the frequency distribution of Benford for the Europe considering all of Italian indexes.

An important technical aspect which could have affected the results of the test is the presence of a large amount of missing data. In fact, only few countries are represented in the analysis of Europe, because the time series of many indexes are unavailable (ATX, ISEQ 20 PRICE, BEX 35, PX INDEX, RTS INDEX and OMX Helsinki).

The introduction of the Euro and the American crisis in 2008 might have played a relevant role in the obtained nonconformity outcome. In fact, the negative trends of the European stocks and the increase of the European sovereign debt with the generalized fear of possible default of some countries, may have led to a distortions between empirical distribution and the BL.

Figures 6-8 illustrate the cases of Nasdaq Composite and of the USA in general, for the first and the second significant digit of the price.

Fig. 6: The comparison between the real distributions of the 1st digit of adjusted closing price and the frequency distribution of Benford for the NASDAQ COMPOSITE.

Fig. 7: The comparison between the real distributions of the 1st digit of adjusted closing price and the frequency distribution of Benford for the USA.

Fig. 8: The comparison between the real distributions of the 2nd digit of adjusted closing price and the frequency distribution of Benford for the USA.

The discussion of the USA cases is now provided.

The Nasdaq Composite is made up of 2558 stocks, hence being the most consistent index of the set of the investigated ones. Also the American area is the most relevant one, with 2759 available stocks. BL fails in all these cases.

As for European case, there is an incompleteness problem. Indeed, it has not been possible to find the data of many indexes (the entire S&P and the Russell are missing). This technical reason may be behind the obtained results.

Under a different point of view, the events which seem to be plausibly responsible for the non-validity of BL are of different types. We mention the Black Monday, which represents one of the greatest financial collapse in size globally, all in a single day: October 19th, 1987. It is also important to recall the Twin Towers attacks, when the terrorist attacks at the core of the USA has precipitated the world markets in a state of shock. The invasion of Iraq by the Americans in March 2003 is also a very relevant event. Indeed, USA markets have suffered a lot during the Second Gulf War. That period has experienced a negative peak price ever recorded before. Last but not least, the serious subprime mortgage crisis, with the bursting of the real estate bubble (due to the heavy exposure of the largest investment banks in the area of subprime mortgages). This crisis turned into a financial crisis in 2008 and took a global character which affected every country in the world. Such economic drawback is commonly defined as the worst global financial crisis since 1929.

**5. Conclusion**

The application of the BL has highlighted the complexity of the financial markets and has emphasized the sensitivity to the events around them. Several discrepancies have been observed and commented. The motivations behind them can be found in technical characteristics of the dataset but also in the breakthrough events which provide basically a reshuffling and shake of the ordinary course of finance.

The analysis has been carried out over first and second relevant digits, for prices and volumes and for all the international markets with a large enough number of data. The peculiar attention to geographical areas and to the overall sample of the World markets brings further consistency to the obtained results.

**References**

Antweiler W., Frank M. Z. (2004), Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance* 59, 1259–1294.

Ausloos M., Castellano R., Cerqueti R., (2016). “Regularities and Discrepancies of Credit Default Swaps: a Data Science approach through BL”, *Chaos, Solitons and Fractals* 90, 8-17.

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

Ausloos M., Herteliu C., Ileanu B., (2015). “Breakdown of BL for birth data” *Physica A* 419, 736-745.

Benford F., (1938). “The Law of Anomalous Numbers.” *Proceedings of the American Philosophical Society* 78(4), 551-72.

Carrera C., (2015). “Tracking exchange rate management in Latin America” *Review of Financial Economics*, 25(C), 35-41.

Corazza M., Ellero A., Zorzi A., (2010). “Checking financial markets via Benford's law: the S&P 500 case”. In: Mathematical and statistical methods for actuarial sciences and finance, M. Corazza and C. Pizzi, Eds. (Springer-Verlag Italia), 93-102.

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

El Sehity, T., Hoelzl E., Kirchler E., (2005). “Price Developments after a Nominal Shock: BL and Psychological Pricing after the Euro Introduction.” *International Journal of Research in Marketing* 22(4), 471-80.

Günnel S., Tödter K. H., (2009). “Does BL Hold in Economic Research and Forecasting?” *Empirica* 36(3), 273-292.

Holz C. A., (2014). “The quality of China’s GDP statistics” *China Economic Review* 30, 309-338.

Juergens J. L., Lindsey L. (2009), Getting Out Early: An Analysis of Market Making Activity at the Recommending Analyst's Firm. *The Journal of Finance* 64, 2327–2359.

Karavardar, A., (2014). “Benford's Law and an Analysis in Istanbul Stock Exchange (BIST)” *International Journal of Business and Management* 9, 160-172.

Ley E., (1996). "On the Peculiar Distribution of the U.S. Stock Indices Digits." *The American Statistician* 50, 311-313.

Mebane Jr., Walter R., (2011). "Comment on" BL and the Detection of Election Fraud."” *Political Analysis* 19(3), 269-272.

Michalski T., Stoltz G., (2013). “Do Countries Falsify Economic Data Strategically? Some Evidence That They Might.” *Review of Economics and Statistics* 95(2), 591-616.

Mir T. A., (2016). “The leading digit distribution of the worldwide illicit financial flows” *Quality & Quantity* 50(1), 271-281.

Mir T.A., (2012). “The law of the leading digits and the world religions.” *Physica A* 391(3), 792–798.

Mir T.A., (2014). “The Benford law behavior of the religious activity data.” *Physica A* 408, 1–9.

Mir T. A., Ausloos M., Cerqueti R., (2014). “BL predicted digit distribution of aggregated income taxes: The surprising conformity of Italian cities and regions.” *The European Physical Journal B* 87(261).

Newcomb S., (1881). “Note on the Frequence of Use of the Different Digits in Natural Numbers.” *American Journal of Mathematics* 4(1/4), 39-40.

Nigrini M.J., (1996). “A Taxpayer Compliance Application of BL.” *Journal of the American Taxation Association* 18(1), 72-91.

Nigrini M. J., “BL: Applications for Forensic Accounting, Auditing, and Fraud Detection.” Hoboken, N.J.: John Wiley & Sons, 2012.

Nigrini M. J., (1999). "I've Got Your Number." *Journal of Accountancy* 187, 79-83.

Nye J., Moul C., (2007). “The Political Economy of Numbers: On the Application of BL to International Macroeconomic Statistics” [*The B.E. Journal of Macroeconomics*](http://econpapers.repec.org/article/bpjbejmac/) 7(1), 1-14.

Realdon M. (2008). Credit default swap rates and stock prices. *Applied Financial Economics Letters*, 4(4), 241–248.

Tam Cho, W.K., Gaines B.J., (2007). “Breaking the (Benford) Law. Statistical Fraud Detection in Campaign Finance” *The American Statistician* 61(3), 218-223.

Varian H.R., (1972). “BL” *The American Statistician* 23, 65-66.

1. Corresponding author. University of Macerata - Department of Economics and Law, via Crescimbeni 20, Macerata I-62100, Italy. Tel.: +39 0733 2583246. Fax: +39 0733 2583205. Email: roy.cerqueti@unimc.it. [↑](#footnote-ref-1)