Physica A 574 (2021) 125969

Contents lists available at ScienceDirect

Physica A

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

Benford's laws tests on S&P500 daily closing values and the corresponding daily log-returns both point to huge non-conformity

Marcel Ausloos [a,b,c,1], Valerio Ficcadenti [d,1,*], Gurjeet Dhesi [d,1], Muhammad Shakeel [e,1]

[a] School of Business, College of Social Sciences, Arts, and Humanities, University of Leicester, Brookfield, Leicester, LE2 1RQ, United Kingdom
[b] Department of Statistics and Econometrics, Bucharest University of Economic Studies, Calea Dorobantilor 15-17, 010552 Sector 1, Bucharest, Romania
[c] Group of Researchers for Applications of Physics in Economy and Sociology (GRAPES), Rue de la belle jardinière, 483, B-4031 Angleur, Liège, Belgium
[d] School of Business, London South Bank University, 103 Borough Road, SE1 0AA, London, United Kingdom
[e] Leicester Castle Business School, De Montfort University, Gateway House, LE1 9BH, Leicester, United Kingdom

ARTICLE INFO

Article history:
Received 3 November 2020
Received in revised form 13 March 2021
Available online 29 March 2021

Keywords:
S&P500
Benford's laws
Log-returns
Closing prices
Geometric Brownian motion

ABSTRACT

The so-called Benford's laws are of frequent use to detect anomalies and regularities in data sets, particularly in election results and financial statements. However, primary financial market indices have not been much studied, if studied at all, within such a perspective.

This paper presents features in the distributions of S&P500 daily closing values and the corresponding daily log-returns over a long time interval, [03/01/1950 - 22/08/2014], amounting to 16265 data points. We address the frequencies of the first, second, and first two significant digits and explore the conformance to Benford's laws of these distributions at five different (equal size) levels of disaggregation. The log-returns are studied for either positive or negative cases. The results for the S&P500 daily closing values are showing a remarkable lack of conformity, whatever the different levels of disaggregation. The causes of this non-conformity are discussed, pointing to the danger in taking Benford's laws for granted in large databases, whence drawing "definite conclusions". The agreements with Benford's laws are much better for the log-returns. Such a disparity in agreements finds an explanation in the data set itself: the index's inherent trends. To further validate this, daily returns have been simulated via the Geometric Brownian Motion and calibrating the simulations with the observed data averages and testing against Benford's laws when the log-returns distribution's standard deviation changes. One finds that the trend and the standard deviation of the distributions are relevant parameters in concluding about conformity with Benford's laws.

* Corresponding author.
E-mail address: ficcadv2@lsbu.ac.uk (V. Ficcadenti).

[1] Equally contributed to the study.

https://doi.org/10.1016/j.physa.2021.125969
0378-4371/

1. Introduction

Newcomb, in [1] noticed in 1881 that the first few pages of logarithmic table books are more thumbed than the latter ones. He inferred that numbers with smaller initial digits are more often looked for and used than numbers with larger initial digits. The author's observation was forgotten for about six decades [2].

In [3], Benford apparently and independently[2] went much ahead in detail in 1938 and tested the accuracy of his observation by analysing a large collection of (in total 20 000) numbers gathered from twenty diverse fields, thereby establishing the law as

$$P(d_1) \ = \ log_{10}(\frac{d_1+1}{d_1}) = log_{10}(1+\frac{1}{d_1}), \tag{1}$$

for $d_1 \ = \ 1, 2, 3, \ldots, 9$, where $P(d_1)$ is the probability of a number having the first non-zero digit d_1 and log_{10} is the logarithm in base 10. The first significant digit of a number is its left-most nonzero digit. According to Eq. (1), the smallest digit, 1, should appear as the first digit with the highest proportion (30.1%), whereas the largest digit, 9, should appear as the first digit with the least proportion (4.6%). Thus, N_{d_1}, the number of times the integer $d_1 = 1, 2, 3, \ldots, 9$ is observed to be occurring as the first digit in a data set, is given by the so called Benford's law for the first digit (BL1 hereafter):

$$N_{d_1} = N \ log_{10}(1+\frac{1}{d_1}), \quad d_1 = 1, 2, 3, \ldots, 9 \tag{2}$$

where N is the total number of considered data points.

One can show that the probability that $d_2 = 0, 1, 2, 3, \ldots, 9$ is encountered as the second digit is given by the Benford's law for the second digit (BL2 hereafter):

$$P_2(d_2) = \sum_{k=1}^{9} \log_{10}\left(1+\frac{1}{10\ k+d_2}\right) = log_{10}\left[\prod_{k=1}^{9}(\frac{10\ k+d_2+1}{10\ k+d_2})\right] \tag{3}$$

Moreover, one can extend BL1 to the first two digits, obtaining the so called BL12,

$$P_{12}(d_1\ d_2) = log_{10}(1+\frac{1}{d_1\ d_2}), \quad d_1\ d_2 = 10, 11, 12, \ldots, 98, 99 \tag{4}$$

Following a revival due to Nigrini [4,5], nowadays, these so-called Benford's laws [2,6,7] are of frequent use in order to detect anomalies and regularities in many data sets [e.g. see 8, were widely used survey data sets has been assessed]. In brief, can one trust the data?

Let us warn that Benford's laws (BLs) unique origin is not accepted by all theoreticians; in fact, it might not be unique. Moreover, some discussion rightly exists on whether Benford's laws should even be valid at all! One might also discuss how to test the validity (or not) of BLs as done in [9–11] and [12].

Usually, one considers that Benford's laws should be valid if there is no data manipulation or if human constraints are non-existent [13]. Yet, there are cases in which Benford's laws are either not hold, even though their occurrence should be expected [14], or on the contrary, are not expected to be observed, but are present [15–20]. Thus, testing BLs on various samples should bring some argument about discussing the controversies.

It seems strange that primary financial market indices have not been extensively analysed using Benford's laws. In Section 2, "State of the Art", we recall what is currently present in the literature concerning the use of Benford's laws for studying financial market indices.

In the present paper, we report our study of Benford's laws for first, second, and first two digits validity (called BL1, BL2, BL12), upon the S&P500's daily closing values and log-returns, over a long time interval: from 03/01/1950 till 22/08/2014. This amounts to 16 265 data points. The time series is downloaded from "Yahoo! Finance", an authoritative web site providing financial data.[3] In doing so, we are in line with studies like [20–23].

We discuss both daily closing values and daily log-returns. Moreover, we divide the whole time interval into five equal size subsets made of 3253 observations each. The interest in such disaggregation will be explained below.

We observe huge deviations of the S&P500 closing values through data histograms with respect to the BLs predictions (or expectations) in Section 3. The findings are in disagreement with [24–26]. We explain the causes of such a disagreement in Section 5. Concerning the log-returns, it results that the agreement with BLs is much better; we also explain why. The segmentation of the raw data into 5 time intervals is much serving the explanation.

Therefore, even though Benford's laws are mainly used to point out potential frauds in financial statements coming from companies [27] or countries [28], one may wonder (or expect) that investors can use such BLs for designing strategies or in building pertinent models based on volatility. Besides the findings related to data ranges and concerning the role of digits' frequencies at some position in the considered numbers, one may suggest further researches, pending that the considerations can be tied to other techniques based on the frequency of digits, like letters in a text [29]. This is also relevant to Bayesian approaches (or inputs) and Markov models in investor risk-taking aspects.

2 Benford does not cite Newcomb. In fact, neither papers have any bibliography.

3 https://finance.yahoo.com

Table 1
Statistical characteristics of the S&P500 daily closing values (CV) and corresponding log-returns (LR) distributions over the whole data set, [03/01/1950 - 22/08/2014]. The characteristics values are rounded to at most 5 significant digits.

S&P500	CV	LR
Minimum	16.66	−0.2290
Maximum	1992.4	0.10957
N. Points	16 265	16 264
Mean	451.45	$2.9403 \cdot 10^{-4}$
Std. Dev.	514.08	$9.7315 \cdot 10^{-3}$
Skewness	1.0637	−1.0311
Kurtosis	−0.32647	27.727

2. State of the art

Ley in 1996 has apparently been the first to examine "the peculiar distribution of the US stock indexes' digits" in [24]; from now one we use the quotation marks when we report phrases taken from the quoted papers. One has to wait for 2010 for Zhao and Wu's considerations on whether "Chinese stock indices agree with Benford's Law" [26]. In both cases, [24] and [26], Benford's law is claimed to be valid. Closely connected to our research, [25] checked whether financial markets like the S&P 500 case, from August 14, 1995, to October 17, 2007, thus 3067 data points, obeyed BL1 [25]. The authors also found some reasonable agreement, except that they claim the presence of anomalous times, like market crashes or special events.

Let us mention [30] and [31] where it is tested the "distribution of BIST-100 returns" along BLs. More recently, in 2018, [20] looked at whether BL1 could infer the reliability of financial reports in 6 developing countries. It was shown that "several visually anomalous data have to be a priori removed" to improve the agreement. Elsewhere, i.e. outside market indices studies, the authors of [32] studied LIBOR manipulation, performing an "Empirical Analysis of Financial Market Benchmarks Using Benford's Law". The authors point to "a concentration of notably high deviations from the Benford distribution". In [33], Alali and Romero studied a decade of financial data from a large sample of U.S. public companies along with a BL12 perspective. Alali and Romero broke down the data into six sub-periods, and found "different indicators of manipulation"; similar conclusions against Benford's law compliance are presented in [34], by the same authors.

In so reading, there is no need to say that more analysis can be welcome, and subsequent findings have to be discussed, as it follows here below.

3. Data and data analysis

We have access to the S&P500 daily closing values (CV) via the "Yahoo! Finance" web site. The downloaded data cover a period starting on 03/01/1950 and ending on 22/08/2014, see Fig. 1. This amounts to 16 265 data points. From such a set, one can easily obtain the 16 264 log-returns (LR); see Fig. 1 also. The main statistical characteristics of such a sample are reported in Table 1. Here, it is worth highlighting the huge difference in magnitude for the S&P500 closing values, ranging from ∼16 at the beginning of the time series to ∼2000 realized in 2014.

Since there is sometimes some discussions on the adequate size of the sample [12] and, for time series, about their stationarity, we have also divided the original sample into 5 equivalent size groups; thus, each set containing 3253 data points. The corresponding log-returns follow at once.

A BLs analysis is usually limited to the first and sometimes second digit. The second, third and fourth digit distributions are usually found to agree with BL2, BL3, and BL4; they can hardly be used for discussion. Sometimes, one finds a study of the first two - BL12 - (and first three digits, BL123). Thus, to prepare for a BLs analysis, one usually rounds up the data to 5 digits to avoid rounding the 4th significant digit if it occurs. We kept that rounding rule even though we only consider the first, second, and first two digits to test BL1, BL2 and BL12 on each S&P500 and log-returns sample. The statistical characteristics of such "adjusted values" are presented in Table 2, for the whole set and each subset. The notations seem to be obvious: CV_k and LR_k, with $k = I, \ldots, V$ refer to the subsets. For completeness, let us mention that the upper limits of such subsets are 3253, 6506, 9759, 13 012, and 16 265, respectively.[4]

One can observe much variety in the data reported in Table 2: for example, there are considerable negative log-returns due to a few crashes, whence the standard deviation can also be very high. Also, the skewness and kurtosis, either for the S&P500 raw data and for the log-returns, have different magnitude orders.

Thereafter, we can compare the number of first, second, and first two digits in such data sets (12, 5 for closing prices, 5 for log-returns and 2 for the global analysis). In the nomenclatures, we distinguish the 5 subsets by different symbols.

[4] It can be easily understood that do not take into account the first value of each log-return subset when dividing the whole set into 5 boxes, in order to have the same number of data points, i.e. 3252 for each subset. This is obviously far from a drastic "assumption"!

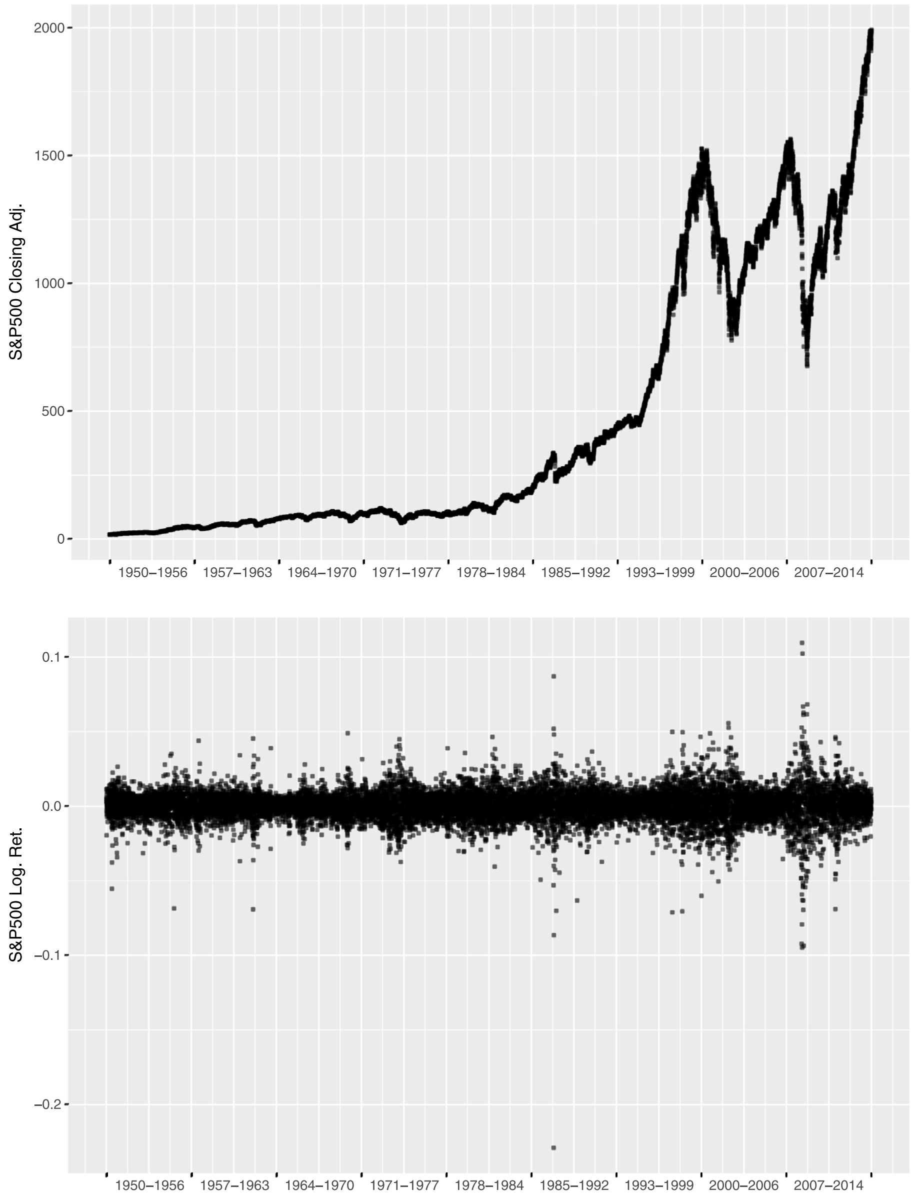

Fig. 1. (top) S&P500 daily closing value, (bottom) corresponding daily log-returns, between 03/01/1950 and 22/08/2014.

Two warnings first: (i) officially, a zero cannot be the first digit when studying BLs; (ii) decimal points separators are also ignored. Nevertheless, in our counting algorithms, we have kept 0 as a *bona fide* first (and second) digit in BLs tests on the log-returns. Indeed, in several (124) cases, these log-returns are strictly equal to 0 because there was no change in two successive S&P500 closing values. In such cases, the second digit is, of course, also 0. Keeping such a digit for the tests on log-returns allows one to observe the relative importance of such events; obviously $\leq\ \sim$1%. It is easily admitted that the importance is not great. However, necessarily, the number of observed events, N_k, with $k = I, \ldots, V$, thereafter differs in the previously imposed equal size intervals because the zeros are not homogeneously distributed across the 5 log-return subsets.

Here, we want at once to emphasize the following: some first digits, whence first two digits, values are missing in various subsets. For example, the missing first digits in each CV_k, can be found from Table 3; this is also clearly observed in the first digit figure for S&P500, Fig. 2, where one has stacked up the subset histograms.

This is not a trivial point; one understands (a posteriori) that this is due to the presence of different trends in the data; see the discussion in Section 5.

Fig. 2 presents the data for testing BL1 on the whole S&P500 daily closing values and on the corresponding log-returns. The division by colour provides information about the examined time intervals. Fig. 3 presents the corresponding BL2 data;

Table 2
Statistical characteristics of ("top") the S&P500 and ("bottom") the corresponding daily log-returns for the whole data set (16 265 and 16 264 data points, respectively), and for the five subsets (3253 and 3252 data points, respectively).

S&P500	*Min.*	*Max.*	*Total*	*Mean*	*Std.Dev.*	*Skewness*	*kurtosis*
CV	16.660	1992.4	7.34281[a]	451.45	514.08	1.0637	−0.32647
CV_I	16.660	72.640	0.13483[a]	41.447	15.828	0.076114	−1.2629
CV_{II}	62.070	120.24	0.29463[a]	90.571	12.092	−0.069025	−0.45554
CV_{III}	86.900	336.77	0.51063[a]	156.97	63.706	0.96070	−0.26424
CV_{IV}	263.82	1527.5	2.32312[a]	714.15	389.79	0.69803	−1.0539
CV_V	676.53	1992.4	4.07961[a]	1254.1	258.23	0.63624	0.30357
LR	−0.22900	0.10957	4.7821	2.9403[b]	97.315[b]	−1.0311	27.727
LR_I	−0.06909	0.04544	1.32420	4.0721[b]	74.371[b]	−0.71995	8.1493
LR_{II}	−0.03740	0.04900	0.35396	1.0884[b]	76.573[b]	0.22745	3.3794
LR_{III}	−0.22900	0.08709	1.13920	3.5031[b]	101.43[b]	−3.6266	84.952
LR_{IV}	−0.07113	0.04989	1.26110	3.8778[b]	96.784[b]	−0.35884	4.8634
LR_V	−0.09470	0.10957	0.72213	2.2206[b]	127.70[b]	−0.21277	9.1859

[a] $\equiv 10^6$
[b] $\equiv 10^{-4}$.

Table 3
Number of d_1 digits, in CV_k groups; observe that there are sometimes missing digits.

d_1	1	2	3	4	5	6	7	8	9
CV_I	234	898	252	812	689	299	69	0	0
CV_{II}	735	0	0	0	0	185	427	921	985
CV_{III}	2026	650	93	0	0	0	0	58	426
CV_{IV}	899	117	682	818	171	220	126	60	160
CV_V	2751	0	0	0	0	4	27	229	242

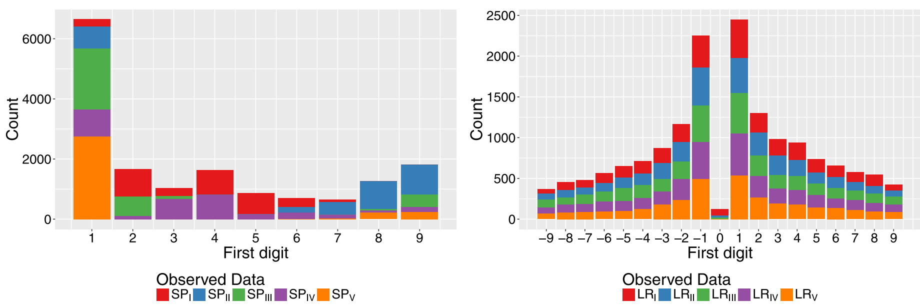

Fig. 2. Test of BL1 on (left) S&P500 closing value, (right) corresponding daily log-returns between 03/01/1950 and 22/08/2014.

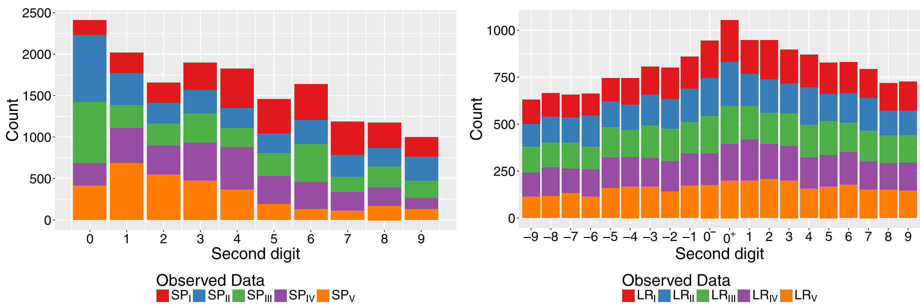

Fig. 3. Test of BL2 on (left) S&P500 closing value, (right) corresponding daily log-returns between 03/01/1950 and 22/08/2014.

Fig. 4 the data serving for a BL12 analysis. At once, visually, the S&P500 data looks hardly representable by a log function, like Eqs. (2)–(4). In contrast, the log-returns histograms have a more appealing form. Notice that we distinguish negative and positive log-returns and mention on each graph the occurrence of strictly zero and double zero values.

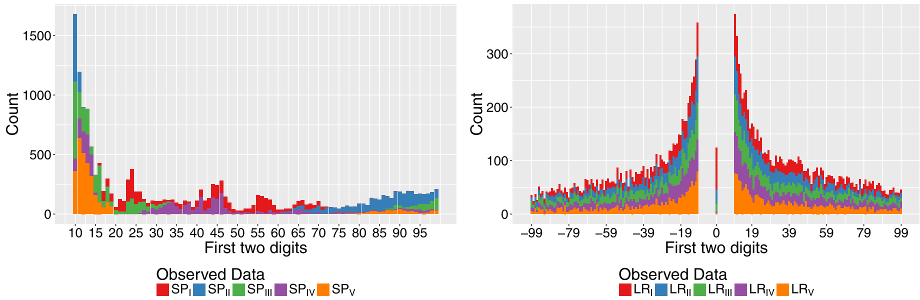

Fig. 4. Test of BL12 on (left) S&P500 closing value, (right) corresponding daily log-returns between 03/01/1950 and 22/08/2014.

Table 4
Results of χ^2 test of the daily closing values (CV) conformity with BL1, BL2, and BL12 for the S&P500 whole data set and for the 5 equal size subsets; the number N_k of observations (or data points) is indicated for each case: 16 265 for the whole time series and 3253 data points, for the five subsets, respectively. The number of degrees of freedom (dof) is easily derived from the number of bins. The corresponding $\chi_c^2(0.05)$ is given for an immediate comparison.

		BL1	BL2	BL12
	dof :	8	9	89
N_k	$\chi_c^2(0.05)$:	15.507	16.919	112.022
16 265	CV	3756.03	397.46	7084.40
3253	CV_I	2737.22	387.46	5030.895
3253	CV_{II}	10038.14	544.12	12840.86
3253	CV_{III}	2936.91	527.02	5561.862
3253	CV_{IV}	1476.73	302.02	3496.052
3253	CV_V	5187.90	813.99	7664.894

In Table 4, we report the χ^2 test of variables conformity with BL1, BL2, and BL12 for the S&P500 whole data set and for the subsets; the number N_k of observations (or data points) is indicated for each case: 16265 and, for the five subsets, 3253 data points, respectively. The number of degrees of freedom (dof), easily derived from the number of bins, is also indicated with the "critical" $\chi_c^2(0.05)$ value. One can hardly admit any conformity, given the large values compared to the "critical" $\chi_c^2(0.05)$ value. Even if a χ^2 test can be claimed as not being the most powerful test for BL conformance [10], the current results are so different from any reasonable expectation that the utilization of another test will be unlikely able to inverse the conclusions.

Let us turn our attention to the log-returns. As mentioned, there are 124 cases in which the log-returns are equal to 0 since the closing prices are identical two consecutive days; these cases occur unevenly in the different CV_k intervals: for completeness, let us mention their occurrence: 78, 26, 15, 3, and 2 times, respectively. This influences the number of observations N_k in each LR_k subgroup; see first column in Table 5. Therefore, $16264 - 124 = 16140$ cases are examined in the whole LR series. When dividing the LR series into 5 subsets, for coherence, the first value in the II, III, IV, V, subsets are disregarded since the first one (day) is missing in the LR_I case. Thus the number of LR observations on which to test BL1 amounts to $16260 - 124 = 16136$.

The number N of data points should be expected to be 3252 for the five LR subsets. However, N_k, the number of observations in the k-subset, varies in each subset, since one is not taking into account the number (124) of log-returns strictly equal to 0, and such a number is not uniformly distributed through the subsets. Moreover, notice that we distinguish (top of Table 5) the case of the absolute values of log-returns and those corresponding to either positive or negative log-return sign (two bottom sub-tables).

The χ^2 tests of variables conformity with BL1, BL2, and BL12 for the S&P500 corresponding daily log-returns (LR) for the whole data set and for the subsets are given in Table 5. BL1 is hardly obeyed, but the difference between the χ^2 values and the χ_c^2 is not so big as for the S&P500 closing prices sample. Some exceptional cases appear to obey BL1, all of the fall in the study of negative returns, LR_{II}^-, LR_{IV}^- and LR_V^-. The situation is almost perfect for BL2, for which only LR_{II}^+ is slightly disagreeing. In the case of BL12, only the latest subsets present some agreement, but the first subset and the whole sample series are surely not obeying BL12.

Our explanation follows in the conclusion section.

4. A benford's laws compliant price paths generator?

To stress the dependence from the distributional features of the data against the numerousness of the observations, we test the ability of the standard Geometric Brownian Motion (GBM), [35], in producing a Benford's laws compliant series of returns.

Table 5
Results of χ^2 conformity test with BL1, BL2, and BL12 for the S&P500 corresponding daily log-returns (LR) for the whole data set and for the subsets. The number of degrees of freedom (dof), easily derived from the number of bins, is also indicated. The number N of data points is equal to 16 264 for the whole set and should be expected to be 3252 for the five subsets; however, N_k, the "number of observations", varies for the various cases, since one is not taking into account the number (124) of log-return values strictly equal to 0. Moreover, notice that we distinguish (top of table) the case of the absolute values of log-returns and those corresponding to either positive or negative log-return sign (two "bottom sub-tables").

		BL1	BL2	BL12
	dof:	8	9	89
	$\chi_c^2(5\%)$	15.507	16.919	112.022
N_k		16 140		
16 136	LR	156.66	4.18	255.96
3174	LR_I	101.34	10.88	213.45
3226	LR_{II}	16.61	16.30	146.46
3237	LR_{III}	86.25	8.31	172.70
3249	LR_{IV}	33.99	4.49	101.39
3250	LR_V	19.11	5.42	102.23
N_k		8616		
8614	LR	115.06	4.73	198.02
1742	LR_I^+	74.97	5.41	168.53
1687	LR_{II}^+	28.58	21.38	135.47
1690	LR_{III}^+	33.30	5.87	108.90
1726	LR_{IV}^+	26.24	5.42	113.55
1769	LR_V^+	17.04	7.52	91.69
N_k		7524		
7522	LR	54.46	7.01	164.81
1432	LR_I^-	34.07	10.24	145.76
1539	LR_{II}^-	5.59	9.53	120.75
1547	LR_{III}^-	59.53	8.06	154.73
1523	LR_{IV}^-	15.37	5.30	83.38
1481	LR_V^-	12.56	13.98	123.31

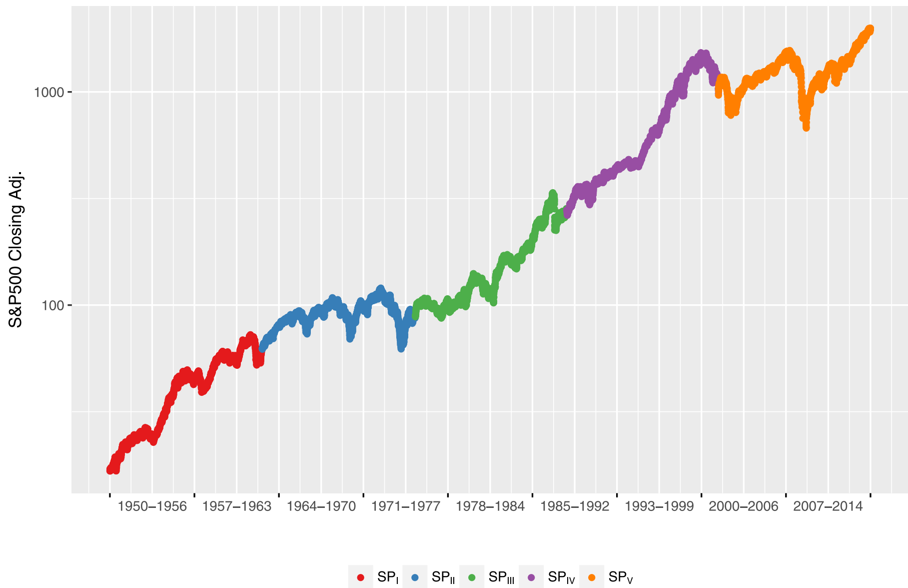

Fig. 5. Semilog plot of S&P500 closing values between 03/01/1950 and 22/08/2014 emphasizing studied subsets.

Along the standard GBM formulation, one has

$$\ln \frac{S_t}{S_0} = (\mu - \frac{\sigma^2}{2}) * t + \sigma * W_t \quad (5)$$

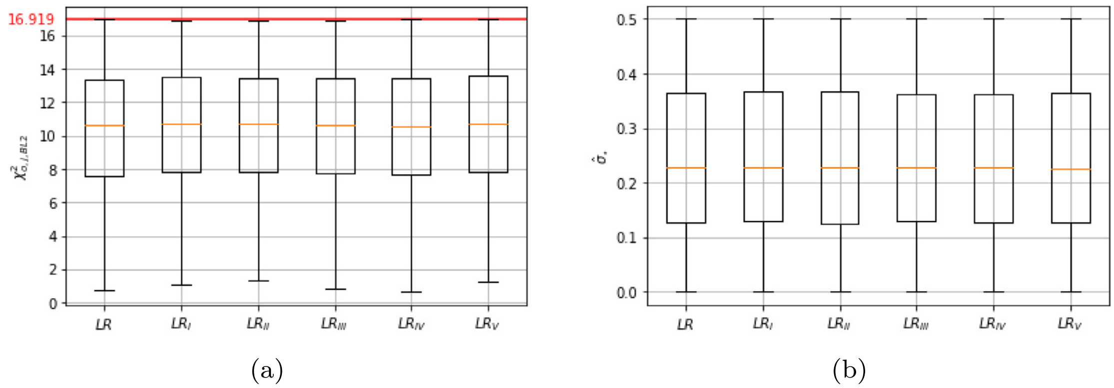

Fig. 6. These graphs contains the results coming form the application of criterion B. (a) is the boxplot of $\chi^2_{o,j,BL2}$ falling below the threshold level 16.919 (red line in the plot) for a 5% significance of the chi square test. So, it represents the chi square statistics coming from the simulated data and j is the pointer addressing the jth level of $\sigma_j = [0.0001 - 0.5]$ used to simulate the data. (b) contains the $\hat{\sigma}_\star$ corresponding to the simulated paths having passed the test. Namely, it contains the $\hat{\sigma}_\star$ corresponding to the statistics reported in (a).

where $\ln \frac{S_t}{S_0}$ is the log-returns (being S_0 and S_t the prices observed at inception and at time t respectively), μ is the mean and σ is the standard deviation of the log-returns, t is the time and W_t is the Wiener process or Brownian motion. Assuming log-normally distributed returns, calibrating the mean via the empirical observations, we aim at finding the level of σ that makes the returns simulated via the GBM as close as possible to the BL1, BL2, BL12 compliance, employing 2 criteria later described.

In so doing we have simulated the returns via the following relationship:

$$\overline{r_{\star,j}} = (\mu_\star - \frac{\sigma_j^2}{2}) * dt + \sigma_j * \sqrt{dt} * \overline{z_{\star,j}} \tag{6}$$

where dt has been set equal to 1 for convenience without harming the relationship, $\mu_\star$ is the average of the returns for the cases $\star = \{LR_I, LR_{II}, LR_{III}, LR_{IV}, LR_V\}$, σ_j is the jth standard deviation from the array ranging from 0.0001 to 0.5 with a step of 0.0001 (the range is set to embed the standard deviations reported in Table 2). $\overline{z_\star, j}$ is an array made of 5000 random extractions from a $N(0, 1)$. Thus, $\overline{r_{\star,j}}$ contains 5000 simulated returns with average $\mu_\star$ for each $\star$ and each j. Therefore, per each $\star$ of $r_\star$ we have a matrix with 5000 rows (simulated days) and 5000 columns (one per each σ_j).

From now on, $\hat{\sigma}_\star$ indicates the standard deviation which produced the most compliant BL price path for the respective $\mu_\star$. Therefore, for each column of $r_\star$, we calculate the chi-square statistic against the BL theoretical values for BL1, BL2 and BL12 [in doing so, we are in line with the usage of the test in comparing distributions, see 36,37]. We have determined the target levels $\hat{\sigma}_\star$ by using the following 2 criteria separately:

A. Minimum Euclidean distance between the chi-square distribution threshold levels at 5% significance (considering the respective degrees of freedom, see Table 5) and the observed chi-square levels. The relationship employed is:

$$d_{\star,j} = \sqrt{\sum_{\bullet \in BL} [\chi^2_{o,j,\bullet} - \chi^2_{c,\bullet}(5\%)]^2} \quad \forall j \vee \forall \star \tag{7}$$

Where, $BL = \{BL1, BL2, BL12\}$ representing respectively the stance for first, second and first two digits, $\chi^2_{o,j,\bullet}$ are the observed chi-square statistics and j is the pointer addressing the jth level of $\sigma_j = [0.0001 - 0.5]$; $\chi^2_{c,\bullet}(5\%)$ are the threshold taken for the case of 5% significance. For each vector, the $\min(d_{\star,j})$ is reported in Table 6.

B. Selection of the σ_js for which at least one among BL1, BL2 and BL12 passes the 5% chi-square test. The results are reported in Table 7, 8, Figs. 6(a) and 6(b).

For both criteria used, the results clearly prove that the standard GBM used for simulating returns makes it impossible to get joint compliance with BL1, BL2 and BL12 when starting from the mean calibrated on real data. Furthermore, even using 5000 simulated daily returns, one cannot reach satisfactory results.

A closer look at the results leads to additional comments. Under the criterion A (see Table 7), the returns simulated with the mean of LR_{III} leads to a $\hat{\sigma}_{LR_{III}} = 0.0865$ which is very close to the observed standard deviation for LR_{III}, namely 0.0101. However, the χ^2 tests fail for all the digits d_i apart from the second, as per the real data (see Table 5); in addition, the second digits presents remarkably low statistics. The other $\hat{\sigma}_\star$s are meaningless; namely, they give values rarely met in a financial market; the χ^2 statistics do not present a relationship with the sensibleness of the estimations.

The outcomes resulting from criterion B confirm that the second digits are the most BL compliant. Table 7 shows the number of cases for which the chi-square statistics pass the test with 5% significance; it happens in more than 70% of

Table 6
Results of χ^2 test of variables conformity with BL1, BL2, and BL12 for returns simulated according to Eq. (6) and with $\hat{\sigma}_\star$, which is the standard deviation estimated to satisfy the criterion A for each row. The number of degrees of freedom (*dof*), easily derived from the number of bins.

			BL1	BL2	BL12
dof :			8	9	89
$\chi_c^2(5\%)$			15.507	16.919	112.022
$\star$	$\hat{\sigma}_\star$	eucl. dist.	BL1	BL2	BL12
LR	0.2561	210.30	159.29	12.17	265.42
LR_I	0.2428	205.41	163.80	5.39	253.69
LR_{II}	0.2318	201.60	157.44	10.01	255.02
LR_{III}	0.0865	204.60	127.21	6.89	283.15
LR_{IV}	0.2081	195.18	152.88	13.01	250.63
LR_{IV}	0.2382	193.53	158.17	7.72	242.47

Table 7
Number of times that χ^2 test for verifying conformity with BL1, BL2, and BL12 for the simulated returns have returned acceptable results. Namely, with the returns simulated thanks to the relationship (6) and with $\sigma_\star \in [0.0001 - 0.5]$, we can see the number of times that the criterion B have been respected for each stance (row). Hence, the figures represent the frequencies of the chi-square statistics being below the threshold $\chi_c^2(5\%)$.

	BL1	BL2	BL12
dof :	8	9	89
$\chi_c^2(5\%)$	15.507	16.919	112.022
LR	0	3596	0
LR_I	0	3610	0
LR_{II}	0	3639	0
LR_{III}	0	3619	0
LR_{IV}	0	3640	0
LR_V	0	3576	0

Table 8
This is a statistical summary of the $\hat{\sigma}_\star$ resulting from the application of the criterion B. Namely, we report the mean, the standard deviation and the variation coefficient of the resulting $\hat{\sigma}_\star$. From Table 7 one can see that just the BL2 has passed cases, therefore, the statistics here reported concerns the $\hat{\sigma}_\star$ resulting from the passed tests for the second digit of the simulated data. The boxplots in Fig. 6(b) report additional information about $\hat{\sigma}_\star$ for each considered stance.

	μ	σ	μ/σ
LR	0.2417	0.1441	1.6778
LR_I	0.2423	0.1435	1.6886
LR_{II}	0.2419	0.1450	1.6688
LR_{III}	0.2427	0.1431	1.6967
LR_{IV}	0.2418	0.1436	1.6837
LR_V	0.2420	0.1442	1.6786

the cases for each stance, namely for the majority of the $\sigma_j \in [0.0001 - 0.5]$ plugged in Eq. (6). Fig. 6(a) hints about the distributions of the statistics whose frequencies are reported in Table 7. The $\hat{\sigma}_\star$ obtained when applying the criterion B are summarized in Table 8 and showed in Fig. 6(b). Most of them are pretty high, as testified by the mean and the standard deviation reported in the summary statistics and in the box.

The results show sensitivity to σ and to the presence of trends in the data. Besides, the d_i behaviours are different; therefore, per each digit studied against the respective Benford's law, dedicated consideration should be run before grasping conclusions on the data.

5. Conclusions

In light of increasing knowledge about applications of Benford's laws, we have analysed distributions' features of S&P500 daily closing values and the corresponding daily log-returns over a long time interval, that is, from the first days of January 1950 till almost the end of August 2014, amounting to 16 265 data points. We have addressed our considerations to the amount of first, second and first two significant digits. We have also explored the conformance to Benford's laws of these distributions, distinguishing five different (equal size) levels of disaggregation, in order to test some (non)stationarity (hidden) feature — if it might occur. Moreover, although this is not usual, we have distinguished negative log-returns from positive ones, plus their combination, since we have enough available data points.

The results for the S&P500 daily closing values (CV) are unexpectedly showing a huge lack of conformity, whatever the different levels of disaggregation. We have noticed that some first digits and first two digits values are missing in some subsets. The agreements with Benford's laws are much better for the log-returns (*LR*). Such a disparity in agreements finds an explanation in the data set itself, rather than in potential frauds!

This feature allows us to comment on some often forgotten criterion for testing the conformity of BLs [12]. Indeed, one should emphasize that BLs could only be usefully studied and observed if all digits – from 1 to 9 – are well represented as every first digit. A time-series or a set of data points should first be tested for its range, basically, the minimum and maximum values. The argument is here well sustained by observing the evolution of the S&P500 *CV* over time.

Fig. 5 provides a semi-log view of the S&P500 closing values; the five studied subsets are emphasized. This allows the understanding of the reasons why the distributions of digits are peculiar. The 1 as first digit is not present in some sectors due to the financial trends and steady states in the time series; for example, at the end of 1950, the S&P500 has no 1 as the first significant digit; the 1 came back steadily after $\sim$ 18 years, in September 1968. Another example showing why sometimes a BL analysis and anomaly deduction might be doubtful is found in sector CV_{II}: the index starts from $\simeq$ 62, reaches $\simeq$ 89, but never goes to any 50, or 20 or 200, a fortiori 300, etc. Thus, the index misses a few first digit values. The same observation is valid for the other sectors where the first digits are missing. In the present analysis of a financial market, this is due to the index's inherent trend. Such causes for non-conformity explains previously puzzling observations like in [15]. Related explanations do follow for cases of data containing crashes and long periods spent in growing and recovering [25].

Thus, besides a thorough analysis of a financial index, a case rarely examined over a so big set of observations, specific causes of this non-conformity are presented, pointing to the danger of taking Benford's laws for granted in huge databases, whence leading to "definite conclusions".

One often reads "the more, the better" as in [38] or [39] where it is claimed "the larger, the better" for applying Benford's laws and deducing frauds or not through lack of conformity or not. This is not true! A large set of data points is neither a sufficient nor necessary criterion for such a statistical conformity test [40]. Under this perspective, we have simulated 5000 daily returns using the averages of the real data presented in Section 3 with the Geometric Brownian Motion formulation, see Eq. (6). Also, for each mean, hence for each studied period, the standard deviations plugged in Eq. (6) range in [0.0001 - 0.5] (with steps of 0.0001). The BL compliance results are in line with the outcomes obtained with the observed data. This is an additional hint for the main point made in this research; in fact, one needs to consider the distributional features of the phenomenon under investigation instead of only focusing on the number of observations or the data's granularity. This type of comments is in line with [41], where the author has commented comparable exercise runs for studying fraud detection through BL in political elections.

Finally, recall that BLs are used to detect fraud mainly. Of course, some data sets can be hardly manipulated. We are all convinced that S&P 500 and other financial indices result from averages, thus apparently obeying the BL validity theoretical criteria, whence could not have fraudulent aspects. However, the present study suggests that one might use BLs at a more << microscopic level >>, that of company share price, as already appreciated by [27].

As already stated, one of the main findings of our research has been about the data range. Indeed to conform with the law, the data set must contain data in which each number 1 through 9 has an equal chance of being the leading digit; there should be equipartition [42,43]. However, this seems paradoxical. We show that the data transformation, from the raw index value to the log-return space, is crucial in observing that there is no data manipulation and obedience to BL. The trend value is avoided. Moreover, BL2 and BL12 are less sensitive to trend manipulation.

As so observed, one may imagine that BL2 and BL12 are of interest for investors since a change of the first digit is rare when share prices are higher than 10 (whatever the currency is, in fact). BL1 should be verified for steadily low prices, lower than 10. This would lead to an investment strategy similar to that considering the equivalence of digits in data series to letters in texts [44]. It would be interesting for financial analysts to reconsider a connection between Benford and Zipf law approaches.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S. Newcomb, Note on the frequency of use of the different digits in natural numbers, Amer. J. Math. 4 (1) (1881) 39–40.
[2] T.A. Mir, M. Ausloos, Benford's law: A "sleeping beauty" sleeping in the dirty pages of logarithmic tables, J. Assoc. Inf. Sci. Technol. 69 (3) (2018) 349–358.
[3] F. Benford, The law of anomalous numbers, Proc. Am. Phil. Soc. 78 (1) (1938) 551–572.
[4] M.J. Nigrini, A taxpayer compliance application of Benford's law, J. Am. Tax. Assoc. 18 (1) (1996) 72–91.
[5] M.J. Nigrini, L.J. Mittermaier, The use of Benford's law as an aid in analytical procedures, Auditing 16 (2) (1997) 52–67.
[6] A. Berger, T.P. Hill, An Introduction to Benford's Law, Princeton University Press, 2015.
[7] S.J. Miller, Benford's Law: Theory and Applications, Princeton University Press, 2015.
[8] M. Kaiser, Benford's law as an indicator of survey reliability – can we trust our data? J. Econ. Surv. 33 (5) (2019) 1602–1618.
[9] G. D'Amico, A semi-Markov approach to the stock valuation problem, Ann. Finance 9 (4) (2013) 589–610.
[10] M. Lesperance, W. Reed, M. Stephens, C. Tsao, B. Wilton, Assessing conformance with Benford's law: Goodness-of-fit tests and simultaneous confidence intervals, PLoS One 11 (3) (2016) e0151235.

[11] N. Omerzu, I. Kolar, Do the financial statements of listed companies on the Ljubljana Stock Exchange Pass the Benford's law test? Int. Bus. Res. 12 (1) (2019) 54–64.
[12] A. Berger, T.P. Hill, The mathematics of Benford's law: a primer, Stat. Methods Appl. 1 (2020) 1–17.
[13] U. Hassler, M. Hosseinkouchack, Testing the newcomb-Benford law: experimental evidence, Appl. Econom. Lett. 26 (21) (2019) 1762–1769.
[14] M. Ausloos, C. Herteliu, B. Ileanu, Breakdown of Benford's law for birth data, Physica A 419 (2015) 736–745.
[15] P. Clippe, M. Ausloos, Benford's law and Theil transform of financial data, Physica A 391 (24) (2012) 6556–6567.
[16] 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, Eur. Phys. J. B 87 (11) (2014) 261.
[17] R. Cerqueti, M. Ausloos, Evidence of economic regularities and disparities of Italian regions from aggregated tax income size data, Physica A 421 (2015) 187–207.
[18] T.A. Mir, The leading digit distribution of the worldwide illicit financial flows, Qual. Quant. 50 (1) (2016) 271–281.
[19] M. Ausloos, R. Cerqueti, T.A. Mir, Data science for assessing possible tax income manipulation: The case of Italy, Chaos Solitons Fractals 104 (2017) 238–256.
[20] J. Shi, M. Ausloos, T. Zhu, Benford's law first significant digit and distribution distances for testing the reliability of financial reports in developing countries, Physica A 492 (2018) 878–888.
[21] J.L. Juergens, L. Lindsey, Getting out early: an analysis of market making activity at the recommending analyst's firm, J. Finance 64 (5) (2009) 2327–2359.
[22] M.R. Haley, K-fold cross validation performance comparisons of six naive portfolio selection rules: how naive can you be and still have successful out-of-sample portfolio performance? Ann. Finance 13 (3) (2017) 341–353.
[23] J. Riccioni, R. Cerqueti, Regular paths in financial markets: Investigating the Benford's law, Chaos Solitons Fractals 107 (2018) 186–194.
[24] E. Ley, On the peculiar distribution of the US stock indexes' digits, Amer. Statist. 50 (4) (1996) 311–313.
[25] M. Corazza, A. Ellero, A. Zorzi, Checking financial markets via Benford's law: the S & P 500 case, in: Mathematical and Statistical Methods for Actuarial Sciences and Finance, Springer, 2010, pp. 93–102.
[26] S. Zhao, W. Wu, Does Chinese stock indices agree with Benford's law? in: 2010 International Conference on Management and Service Science, IEEE, 2010, pp. 1–3.
[27] A.D. Saville, Using Benford's law to detect data error and fraud: an examination of companies listed on the Johannesburg Stock Exchange: economics, S. Afr. J. Econ. Manag. Sci. 9 (3) (2006) 341–354.
[28] B. Rauch, M. Göttsche, G. Brähler, S. Engel, Fact and fiction in EU-governmental economic data, Ger. Econ. Rev. 12 (3) (2011) 243–255.
[29] E. Shulzinger, E. Bormashenko, On the universal quantitative pattern of the distribution of initial characters in general dictionaries: The exponential distribution is valid for various languages, J. Quant. Linguist. 24 (4) (2017) 273–288.
[30] M. Cinko, Testing distribution of BIST-100 returns by Benford law, J. Econ. Finance Account. 1 (3) (2014) 184–191.
[31] A. Aybars, L. Ataunal, An application of Benford's law to fundamental accounting figures reported by Borsa Istanbul (BIST) companies, J. Econ. Finance Account. 3 (3) (2016) 234–243.
[32] B. Rauch, M. Goettsche, F. El Mouaaouy, LIBOR manipulation–empirical analysis of financial market benchmarks using Benford's law, 2013, Available at SSRN 2363895.
[33] F.A. Alali, S. Romero, Benford's law: Analyzing a decade of financial data, J. Emerg. Technol. Account. 10 (1) (2013) 1–39.
[34] F. Alali, S. Romero, Characteristics of failed us commercial banks: an exploratory study, Account. Finance 53 (4) (2013) 1149–1174.
[35] A. Einstein, Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen, Ann. Phys. 4 (1905) 549–560.
[36] A.L. Christensen, R.M. Mohr, Nonprofit lobbying: Museums and collections capitalization, Financ. Account. Manag. 15 (2) (1999) 115–133.
[37] T. Dissanayake, S. Dellaportas, P.W.S. Yapa, The diffusion-adoption of accrual accounting in Sri Lankan local governments, Financ. Account. Manag. 36 (3) (2020) 261–277.
[38] G. Nairn, Beating fraudsters at their own game, 2011, www.geoffnairn.com/wordpress/?p=726. (Online; Accessed 23 September 2019).
[39] J.C. Collins, Using excel and Benford's law to detect fraud: Learn the formulas, functions, and techniques that enable efficient benford analysis of data sets, J. Account. 223 (4) (2017) 44.
[40] F. Heilig, E.J. Lusk, Testing the small size effect bias for Benford screening: The false negative signaling error, Bus. Manag. Dyn. 8 (6) (2018) 23.
[41] W.R. Mebane, Comment on Benford's law and the detection of election fraud, Polit. Anal. 19 (3) (2011) 269–272.
[42] É. Janvresse, T. De la Rue, From uniform distributions to Benford's law, J. Appl. Probab. 41 (4) (2004) 1203–1210.
[43] J.R. Iafrate, S.J. Miller, F.W. Strauch, Equipartitions and a distribution for numbers: A statistical model for Benford's law, Phys. Rev. E 91 (6) (2015) 062138.
[44] M. Ausloos, P. Bronlet, Strategy for investments from Zipf law(s), Physica A 324 (1–2) (2003) 30–37.