Benford's Law in Scientific Research

Werner Hürlimann

Abstract— As departures from Benford's law have been observed in many scientific data sets, there is a theoretical need to understand such discrepancies. We argue that the use of parametric extensions to Benford's law is appropriate and demonstrate this for several first significant digit distributions taken from theoretical scientific laws or extracted from real-world data sets.

2010 Mathematics Subject Classification: Primary 62E15, Secondary 62P05, 62P10, 62P25, 62P35.

Keywords— first significant digit, generalized Benford law, Pareto Benford law, two-sided power Benford law, Pareto Benford law, mean absolute deviation, weighted least squares.

1 Introduction

Newcomb [28] observed that in logarithmic tables lower significant digits occur more frequently than upper ones. His "law of probability of the occurrence of numbers" remained unnoticed until Benford [3] verified it empirically. Soon after, this was reported twice in Nature by Goudsmit and Furry [11], Furry and Hurwitz [10], and even formulated as a "harmonic law of Statistics" by Furlan [9]. In this respect, there is an increasing trend to study "the Mathematics of Harmony" and integrate it into general science (e.g. Stakhov [40]). A particular instance is the newspaper mention by Szpiro [39] of a generalized Benford law for prime numbers as formulated in Luque and Lacasa [26].

At present, Benford's law is such omnipresent in the sciences that there is no need to motivate it again. Completing the bibliography [17], the number of publications listed in Berger and Hill [4] and Beebe [2] approaches the first thousand and they appeared in nearly 350 journals in almost all contemporary research fields (e.g. Beebe [1]).

Departures from Benford's law have been detected in many data sets in the natural sciences. As several studies suggest, there is a theoretical need to explain such discrepancies. In this respect, we ask whether the use of parametric extensions to Benford's law is appropriate in general science. Our approach to answer this question follows multiple paths.

We recall in Section 2 the existence of parametric first digit distributions (FDD) that extend and embed Benford's law. Essentially three useful parametric FDD models are considered, namely the generalized Benford (GB), the two-sided power Benford (TSPB), and the Pareto Benford (PB). They are all realized by at least one probabilistic model. Although parametric extensions of Benford's law exist in mathematics, especially number theory, they have not been fully recognized at an early stage, as argued at the beginning of Section 3. Then, based on several theoretically sound first digit distributions from physics and actuarial sciences, Stigler's alternative first digit law, and eight published first digit data sets from the geophysical, earth sciences, social sciences, biomedicine and finance, we demonstrate improved goodness-of-fit of the parametric extensions of Benford's law and comment the obtained results.

In particular, a new theoretical law in the field of actuarial sciences, called Bernegger-Lloyd law, that almost obeys Benford's law, is displayed. We demonstrate that Stigler's law and Benford's law differ fundamentally from both a probabilistic and statistical viewpoint. In the social sciences, we show that twitter users by followers count are almost perfectly Pareto Benford, a result that improves the analysis in [21]. A data set from biomedicine about magneto-encephalograms (MEG data) is statistically rejected as being not conform to both Benford's and Stigler's law, the last one against a previous belief. On the other hand, we improve the analysis of Rodriguez [34] concerning the data set from Ley [24] (Dow Jones index from finance). While the FDD does not conform to Benford's law it does with Stigler's law, and additionally the Pareto Benford extension is also selected. Section 4 concludes the present exposé with some characteristic statements about the use of parametric extensions of Benford's law in theory and applications.

Werner Hürlimann works as Independent Researcher in Pure Mathematics and its Applications, Switzerland. E-mail: whurlimann@bluewin.ch

2 Embedding Benford's law into parametric FDD extensions

Since Hill [13] a lot of attention has been put on the theory of random variables in probability spaces in order to better understand the probability distributions that follow or closely approximate Benford's law. Along this path one might mention papers from Leemis et al. [23], Engel and Leuenberger [8], and others, to recent contributions by Wójcik [44], [45]. Relying on [15], [18], our focus concentrates again on those parametric FDD's that extend and embed Benford's law. Three parametric extensions of Benford's archetype are considered.

Generalized Benford law (GB)

Based on a Bayesian probabilistic model, the author in [21], Proposition 3.1, derives in a new way the so-called *generalized Benford law* (GB) with FDD defined for $\gamma \in (-\infty, \infty)$ by

$$GB(d;\gamma) = \left\{ \begin{array}{ll} \dfrac{d^{-\gamma} - (1+d)^{-\gamma}}{1-10^{-\gamma}}, \quad if \quad \gamma \neq 0, & \\ \log(1+d^{-1}), \quad if \quad \gamma = 0, & \end{array} \quad d = 1,...,9 \right\}.$$

In the limiting case $\gamma \to 0$ the two formulas are consistent and generate Benford's law. One sees that the GB generates

ISSN 2229-5518

monotone decreasing probabilities if, and only if, one has $\gamma \geq -1$, where the limiting case $\gamma = -1$ is the discrete uniform distribution. Therefore, GB is defined as monotone decreasing FDD for all $\gamma \in [-1,\infty)$ and includes Benford's law as special case $\gamma = 0$. Moreover, this FDD is tilted toward a uniform distribution for $\gamma \in [-1,0)$ and is more tilted than Benford's law for $\gamma \in (0,\infty)$. The author [21], Proposition 3.2, also derives the exponential Benford law (EB), which coincides with the GB in the special case $\gamma \in [0,\infty)$.

Two-sided power Benford law (TSPB)

The author [15], Theorem 2.1, derives this one-parameter extension of Benford's law from a two-sided power random variable. For $c \geq 1$ the corresponding FDD is given by

$$TSPB(d;c) = \frac{1}{2}\left\{ \begin{array}{l} [\log(1+d)]^c - [\log(d)]^c \\ -[1-\log(1+d)]^c + [1-\log(d)]^c \end{array} \right\}, \quad d = 1,...,9 .$$

The special cases $c = 1,2$ coincide with Benford's law, which henceforth embed into the TSPB.

Pareto Benford law (PB)

The author [18], Theorem 3.1, obtains this two-parameter extension of Benford's law from a special case of the double Pareto distribution studied in Reed [33]. For parameters $\alpha, \beta > 0$ the FDD takes the form

$$PB(d;\alpha,\beta) = \frac{\alpha}{\alpha+\beta}\left\{[\log(1+d)]^{\beta} - [\log(d)]^{\beta}\right\}$$
$$+\frac{\beta}{\alpha+\beta} \cdot \sum_{k=1}^{\infty}\left\{[k+\log(d)]^{-\alpha} - [k+\log(1+d)]^{-\alpha}\right\}, \quad d \in \{1,...,9\}.$$

One notes that Benford's law is embedded into the PB. Indeed, setting $\beta = 1$ and letting α goes to infinity, one sees that the PB converges to Benford's law.

Another one-parameter extension of Benford's law that encompasses Stigler's law (see Section 3) and the uniform FDD has been constructed by Rodriguez [34] and revisited in [21]. It is remarkable that all the considered parametric extensions of Benford's law are related to power law distributions. Is there a mathematical reason for this phenomenon? Are there necessary conditions for embedding Benford's law into parametric FDD extensions?

3 The Parametric FDD Extensions in Scientific Research

The usefulness of parametric FDD's of Benford type in mathematics is self-evident. However, this does not seem to have been fully recognized at an early stage. For example, the asymptotic distribution of the first digit of m-th integer powers is GB distributed with parameter $\gamma = 1/m$ as the number of digits goes to infinity (e.g. [16], Theorem 1, see also Ross [35]). This result can be extended to other number sequences (see [19], [20]). Note that sequences of integer powers have already been studied by Furlan [9], Section III.13 for squares, pp. 172-191, Section III.15 for cubes, pp. 207-217, and Section III.16 for general powers, pp. 219-229. On the other hand, the first digit sequences of prime numbers follow a GB with size-dependent parameter that converges necessarily to the uniform distribution, as shown by Luque and Lacasa [26].

In view of these mathematical results about the GB distribution, one may ask whether the use of parametric FDD extensions to Benford's law is appropriate in general science. For this, it is natural to consider data sets from physics and earth's science, where Benford's law is known to fit it empirically quite well. To underpin our study with some more theoretical background, we begin with three prominent theoretical laws from statistical and quantum mechanics, namely the Boltzmann-Gibbs (BG), the Fermi-Dirac (FD) and the Bose-Einstein (BE) distributions, which have been studied from Benford's point of view in Shao and Ma [37].

Boltzmann-Gibbs (BG)

This classical macroscopic law of statistical mechanics applies to particles in a system without quantum effects. The density for the energy x of the system is expressed by the exponential law $f_{BG}(x) = \beta \cdot \exp(-\beta x), x \geq 0$, where $\beta = 1/kT$, with k Boltzmann's constant and T the thermodynamic temperature. Its FDD is denoted $BG(d;\beta)$.

Fermi-Dirac (FD)

This microscopic law of quantum mechanics applies to a system in thermodynamic equilibrium that consists of many identical particles described by anti-symmetrical wave function with half-integer spin. The normalized probability density of the system energy reads $f_{FD}(x) = (\beta/\ln 2) \cdot \{\exp(\beta x)+1\}^{-1}$. The corresponding FDD is denoted $FD(d;\beta)$.

Bose-Einstein (BE)

The Bose-Einstein law is the statistics associated to a system of particles described by symmetrical wave function with integral spin. The density of the system energy, which cannot be normalized, is proportional to $f_{BE}(x) = \{\exp(\beta x)-1\}^{-1}, x \geq 0$. The FDD is denoted $BE(d;\beta)$.

It is known that these theoretical distributions are closely related to Benford's law in the following manner. First, the first two FDD's satisfy the invariant properties $BG(d;\beta) = BG(d;10\beta)$ and $FD(d;\beta) = FD(d;10\beta)$, which lead to periodic functions on the β-logarithmic scale (Figure 1 in Engel and Leuenberger [8], and Figures 1 and 2 in Shao and Ma [37]). These invariant properties imply that both FDD's fluctuate around Benford's law in the mean in a mathematical precise sense as follows. The transformed functions $BG^*(d;\alpha)=BG(d;10^\alpha)$ and $FD^*(d;\alpha)=FD(d;10^\alpha)$ are 1-periodic with respect to the parameter α and a Fourier series calculation (under the assumption that $\alpha=\log(\beta)$ is uniformly distributed) shows that $\int_0^1 BG^*(d;\alpha)d\alpha=B(d)$ and $\int_0^1 FD^*(d;\alpha)d\alpha=B(d)$, i.e. in the average Benford's law is recovered (consult the original papers for this). Despite of these nice features, the conformity of these theoretical laws to Benford's law is not overwhelming and can be improved, as argued below in TABLE 4. On the other hand, the third Bose-Einstein FDD is exactly Benford distributed (see Shao and Ma [37], Section 3.3, and Hill [14]).

The above BG, FD and BE laws have also been used in actuarial sciences. They have been extended by Bernegger [5] to a class of two-parameter distributions over finite and infinite ranges. Extracted from his analysis, the following special case is worthwhile to be mentioned.

Bernegger-Lloyd (BL)

Stefan Bernegger, a Swiss actuary physicist, works at the interface between actuarial sciences and physics (actuar-physics). This researcher relates the above three distributions from physics to a topic in reinsurance called exposure curves. In his paper, the Boltzmann-Gibbs distribution is confounded with the Maxwell-Boltzmann distribution, and, as in Shao and Ma [38], should be better called Maxwell-Boltzmann-Gibbs (in honor of the three pioneers of statistical mechanics). After some detailed analysis, Bernegger derives in his equation (3.6) a two-parameter loss distribution that belongs to an exposure curve over an interval $[0,1]$. In terms of parameters $b\ge 0,\ g\ge 1$, the loss distribution reads

$$F(x)=\begin{cases}1, & x=1,\\ 0, & x<1\wedge(g=1\vee b=0),\\ 1-\{1+(g-1)x\}^{-1}, & x<1\wedge b=1\wedge g>1,\\ 1-b^x, & x<1\wedge bg=1\wedge g>1,\\ 1-(1-b)\{(g-1)b^{1-x}+(1-gb)\}, & \\ x<1\wedge b>0\wedge b\ne 1\wedge bg\ne 1\wedge g>1 & \end{cases} \tag{1}$$

The parametric restriction $bg=1\wedge g>1$ belongs to the Boltzmann-Gibbs (BG), $bg<1$ to the Fermi-Dirac (FD) and $bg>1$ to the Bose-Einstein (BE). Bernegger shows in Section 4.3 that some exposure curves used in non-proportional property reinsurance fit (1). In particular, he shows this for the subclass of the BG-FD-BE distribution (1) parameterized by the exposures curves that belong to the one-parameter choice

$$\begin{aligned}b&=b(c)=\exp\{3.1-0.15c(1+c)\},\\ g&=g(c)=\exp\{(0.78+0.12c)c\},\quad c\in[0,5]\end{aligned}. \tag{2}$$

The special case $c=5$ coincides very well with a curve used by Lloyd to rate industrial risks. This distribution is called hereafter Bernegger-Lloyd (BL). Its behaviour is similar to the Bose-Einstein law. Henceforth, it is expected to be closely approximated by Benford's law, a result that is statistically confirmed later on.

Tsallis q-exponential (q-Exp)

Inspired by a probabilistic description of multifractal geometries, Tsallis [42] introduced a generalization of the standard Boltzmann-Gibbs entropy and derived from it the Tsallis q-exponential distribution from the principle of maximum entropy (see e.g. Tsallis [43]). Generalizing the BG law, the probability density for the energy x is expressed by the q-exponential law

$$\begin{aligned}&f_{q-Exp}(x)=\beta(2-q)\cdot\{1-(1-q)\beta x\}^{1/(1-q)},\\ &x\ge 0,\ 1\le q<2\end{aligned}. \tag{3}$$

If $q\to 1$ one recovers the BG law. As shown in Shao and Ma [38], Section III, the explained FDD mean-value property shared by the BG and FD laws, also holds for the q-exponential FDD. That is, the latter fluctuates around Benford's law in the mean. Moreover, the amplitude of fluctuation diminishes as q increases from 1 to 2, and in the limit as $q\to 2$ it conforms exactly to Benford's law. Therefore, it seems interesting to analyze to what extent the q-exponential FDD conforms to parametric extensions of Benford's law in case the parameter q is not too far away from 1.

Stigler's law (SL)

Stigler [41], a future Nobel laureate in economics, proposed an alternative to Benford's law, which is less skewed towards the lower digits. Stigler's law (SL) is determined by the formula

$$\begin{aligned}&SL(d)=\frac{1}{9}\left(1+\frac{10}{9}\ln(10)+d\ln(d)-(1+d)\ln(1+d)\right),\\ &d=1,\ldots,9\end{aligned}. \tag{4}$$

This FDD resembles a bit the generalized Benford law for $\gamma\in(0,1]$ but is in fact very different. First, as explained in Lee et al. [22], the two laws are based on different probabilistic assumptions. While Stigler assumes that the largest entries in statistical tables equally likely begin with $d=1,\ldots,9$, Benford assumes that smaller numbers with corresponding smaller first significant digits occur more often as bounds for statistical tables. The earlier explanation by Raimi [32], Section 5, is also instructive. A purely statistical discrepancy will be clarified later. Stigler's law is closely related to the "random upper bound model" RUBM considered by Cáceres et al. [6] in a biomedicine context. In fact, the latter authors formulate a randomized version of Stigler's law. Performing simulations up to a size of 10'000, they allow Stigler's upper bound to vary randomly and obtain (through numerical simulations) a parameter-free FDD close to Stigler's law. In fact, as the number of simulations grows to infinity the RUBM converges to Stigler's law, as shown in [21]. Therefore, in view of the analytical formula (4), there is no need to perform RUBM simulations. Previous nice derivations of Stigler's law are found in Rodriguez [34] and Lee et al. [22].

The FDD's of the above theoretical distributions are summarized in TABLE 1 (for simplicity, the scale parameters of the BG, FD and qExp are equal to 1, the q parameter of qExp is equal to 1.1). Of course, other theoretical FDD's exist in the literature, and one might analyze them by the same method. To us, it is similarly and even more important to consider real-world data sets and look at their first digits, and, if necessary, even second and third digits. Among the many possibilities, a small sample of nine data sets has been retained for illustration. The first four real-world FDD's in TABLE 1 are taken from Sambridge et al. [36] and concern the geophysical and earth sciences. The next three are taken from the social sciences: the population data of countries around the world, the twitter users by followers count, and the people killed by terrorism (1970-2013), all three are found in Long [25]. The last two illustrates for biomedicine: the global infectious decease data is from Sambridge et al. [36] and the MEG data (magnetoencephalograms from a healthy male) is from Cáceres et al. [6].

TABLE 1
Theoretical FDD's and real-world first digit data sets

FDD's / data sets	First Digit								
	1	2	3	4	5	6	7	8	9
Theoretical FDD's									
Boltzmann-Gibbs (BG)	0.3297	0.1743	0.1127	0.0860	0.0726	0.0643	0.0582	0.0533	0.0490
Fermi-Dirac (FD)	0.3436	0.1841	0.1114	0.0805	0.0667	0.0595	0.0549	0.0512	0.0480
Bernegger-Lloyd (BL)	0.3008	0.1770	0.1252	0.0967	0.0788	0.0665	0.0576	0.0509	0.0466
q-exponential (q-Exp)	0.3196	0.1797	0.1194	0.0900	0.0737	0.0635	0.0564	0.0510	0.0467
Stigler's law (SL)	0.2413	0.1832	0.1455	0.1174	0.0950	0.0764	0.0605	0.0465	0.0342
Real-world FDD's									
Earth's gravity	0.3296	0.1660	0.1120	0.0850	0.0750	0.0670	0.0594	0.0557	0.0503
Geomagnetic field	0.2890	0.1770	0.1330	0.0940	0.0810	0.0690	0.0610	0.0510	0.0450
Seismic wavespeeds	0.3004	0.1760	0.1330	0.0980	0.0790	0.0640	0.0560	0.0489	0.0447
S-A seismogram	0.2839	0.1569	0.1249	0.0960	0.0897	0.0737	0.0652	0.0604	0.0493
Population of countries	0.2741	0.1629	0.1230	0.1061	0.0934	0.0684	0.0653	0.0531	0.0537
Twitter users by followers	0.3262	0.1666	0.1181	0.0926	0.0763	0.0655	0.0577	0.0514	0.0456
Terrorism deaths	0.5193	0.1814	0.0990	0.0640	0.0469	0.0325	0.0247	0.0187	0.0135
Global infectious decease	0.3371	0.1671	0.1321	0.1070	0.0730	0.0540	0.0456	0.0507	0.0334
MEG	0.2580	0.2130	0.1610	0.1150	0.0820	0.0560	0.0450	0.0380	0.0320

At the time being there is no simple exact mathematical test to decide whether a given FDD conforms to Benford's law or another related parametric extension to it (cf. Morrow [27]). Despite this theoretical lack, a lot of experience has been accumulated to assess conformity to Benford's law, which might be extended to parametric FDD's. In this respect, the mean absolute deviation (MAD) test developed by Nigrini [29], Table 7.1, suffices for our purpose (see TABLE 2). Recall the definition of the MAD statistics. Given two FDD's, which may depend on parameters or not, say $F_1(d)$ and $F_2(d)$, $d = 1,...,9$, the MAD measure is defined and denoted by

$$MAD = \frac{1}{9} \cdot \sum_{d=1}^{9} \left| F_1(d) - F_2(d) \right| . \quad (4)$$

The MAD statistics are calculated in TABLE 3 and re-used in TABLE 4 as follows. First, the minimum MAD estimators of the parametric extensions are computed and their minimum values are reported in TABLE 3. Taking into account the (extended) critical values in TABLE 2 the conformity to Benford's law and other FDD's is then assessed in TABLE 4.

TABLE 2
MAD critical values and conformity to FDD

MAD critical values	FDD Conformity	Abbreviation
$MAD \le 6 \cdot 10^{-3}$	Close conformity	C
$6 \cdot 10^{-3} < MAD \le 12 \cdot 10^{-3}$	Acceptable	AC
$12 \cdot 10^{-3} < MAD \le 15 \cdot 10^{-3}$	Marginal	MC
$MAD > 15 \cdot 10^{-3}$	Nonconformity	NC

TABLE 3
MAD goodness-of-fit for some parametric FDD's

	10^3 MAD				10^3 WLS			
Theoretical FDD's	**B**	**GB**	**TSPB**	**PB**	**B**	**GB**	**TSPB**	**PB**
Boltzmann-Gibbs (BG)	7.591	5.637	5.465	**3.280**	6.128	5.729	5.754	**2.079**
Fermi-Dirac (FD)	11.756	5.745	10.682	**5.135**	13.676	7.453	13.653	**4.117**
Bernegger-Lloyd (BL)	0.421	0.418	**0.389**	0.757	**0.0282**	0.0291	0.0338	0.0798
q-exponential (q-Exp)	5.131	2.437	4.703	**2.083**	2.583	1.295	2.517	**0.740**
Stigler's law (SL)	16.86	12.90	12.44	**9.38**	27.76	25.55	16.34	**15.49**
Real-world FDD's	**B**	**GB**	**TSPB**	**PB**	**B**	**GB**	**TSPB**	**PB**
Earth's gravity	8.694	7.544	4.143	**2.951**	7.204	9.241	2.969	**1.683**
Geomagnetic field	3.522	2.507	3.207	**1.959**	1.367	**1.098**	1.193	**0.764**
Seismic wavespeeds	2.034	2.013	**1.837**	1.994	0.856	0.880	**0.599**	0.869
S-A seismogram	8.282	4.423	8.128	**2.283**	7.995	3.160	7.549	**1.008**
Population of countries	9.347	4.119	8.996	**2.913**	9.257	2.382	9.067	**1.891**
Twitter users by followers	5.648	4.415	3.399	**0.244**	3.321	3.106	2.714	**0.018**
Terrorism deaths	49.684	**2.004**	48.594	7.833	268.31	**1.122**	238.59	15.961
Global infectious decease	11.849	7.012	11.812	**6.923**	15.218	**7.380**	14.957	9.012
MEG	20.86	20.31	**15.55**	18.28	39.99	42.57	**22.46**	43.14

Besides decision upon conformity, we use additionally the probability weighted least squares (WLS) measure used earlier by Leemis et al. [23] (chi-square divided by sample size) to decide upon the preferred FDD choices. Again, this measure can be used for both theoretical FDD's or/and FDD's derived from sample data. Indeed, suppose $F_1(d)$ must be chosen to approximate $F_2(d)$ and suppose both have been derived from a sample of same size N. Then, by definition of the WLS measure, one has

$$WLS = \frac{1}{N} \cdot \sum_{d=1}^{9} \frac{(N \cdot F_1(d) - N \cdot F_2(d))^2}{N \cdot F_1(d)} = \sum_{d=1}^{9} \frac{(F_1(d) - F_2(d))^2}{F_1(d)} . \quad (5)$$

ISSN 2229-5518

Theoretically, if the FDD's are known with certainty, the WLS measure does not depend on the sample size. It can therefore be used as a rule of thumb to choose the best fit to a given FDD among various alternatives.

Some interesting observations can be made. Among the theoretical FDD's, the Bernegger-Lloyd conforms very well to Benford's law, and it is the only one, which is actually chosen with first priority by the WLS selection criterion. This is no surprise in view of the fact that it behaves similarly to the BE law. The qExp generalization of the BG generates smaller MAD values and improved close conformity in comparison with the BG. For three of them, namely BG, FD and qExp, the WLS selection criterion chooses first the PB and then the GB. Special comments on Stigler's law follow afterwards. Concerning the real-world FDD's there is close conformity for the geomagnetic field, the seismic wave speeds and the twitter users FDD's. First choices are the PB, TSPB and PB respectively (GB, PB and TSP are second choices). The earth's gravity, the seismogram and population FDD's all perform equally well. For all three the PB is the first choice, followed by the TSPB (earth's gravity) and the GB (other two data sets). The fit of the infectious decease FDD is acceptable conform with respect to all parametric extensions, the first WLS choice being the GB followed by the PB. For all these FDD's (except the BL and the terrorism data) the PB is always selected first or second according to the WLS measure. This can be expected in view of the extra free parameter. A bit surprising is, however, the almost perfect fit of the twitter data with the PB. The terrorism data seems a priori far away from Benford's law (big difference for the digit one). It is a surprise that it conforms to the GB and is acceptable to the PB (the latter failing to be selected by merely 0.961 units of the WLS measure). In general, compared to the TSPB the GB seems to play a more important role, at least for the theoretical FDD's. This confirms partially the prominent role GB seems to play in number theory.

It remains to discuss Stigler's law and the MEG data. First, TABLE 4 shows that the MEG data reveals no conformity with any of Benford's extensions. Second, a separate calculation of the MAD and WLS measures between Stigler's law and the MEG FDD yields $MAD = 13.78 \cdot 10^{-3}, WLS = 20.6 \cdot 10^{-3}$. In contrast to Cáceres et al. [6], which claim that the RUBM/Stigler law "can likely explain the observed behaviour of MEG data", the WLS measure does not select Stigler's law to explain the MEG data. Third, Stigler's law is not in conformity with Benford's law $(MAD = 16.86 \cdot 10^{-3})$ in accordance with the statistical discrepancy stated previously in the text. This law is acceptable conform to the PB law $(MAD = 9.38 \cdot 10^{-3})$ but is slightly rejected by the WLS selection criterion $(WLS = 15.49 \cdot 10^{-3})$. Although the used critical values have not been theoretically well funded so far, they appear reasonable to us. According to Lee et al. [22], footnote 2, p.83, there exist only few empirical data sets following Stigler's law. An exception seems to be the stock market data of Ley [24] (Dow Jones index), which fits better Stigler's law than Benford's law, as shown by Rodriguez [34]. Improving this statement, TABLE 2 rejects Benford's law $(MAD = 16.5 \cdot 10^{-3}, WLS = 27.08 \cdot 10^{-3})$ and confirms Stigler's law $(MAD = 3.4 \cdot 10^{-3}, WLS = 2.05 \cdot 10^{-3})$. The parametric extensions of Benford's law are not rejected but only the PB is selected by the WLS criterion $(MAD = 8.97 \cdot 10^{-3}, WLS = 11.6 \cdot 10^{-3})$. Although Stigler's law and the MEG data are not selected by the WLS criterion for any of the considered parametric extensions of Benford's law, there exist different probabilistic models of FDD's for which they will be selected. For this, the reader is refereed to the recent paper [15], which proposes a fine structure index for Benford's law.

TABLE 4

MAD test and WLS choices ($WLS < 15 \cdot 10^{-3}$)

	generalized MAD test				WLS criterion	
Theoretical FDD's	**B**	**BG**	**TSPB**	**PB**	**Choice 1**	**Choice 2**
Boltzmann-Gibbs (BG)	AC	C	C	C	PB	GB
Fermi-Dirac (FD)	AC	C	AC	C	PB	GB
Bernegger-Lloyd (BL)	C	C	C	C	B	GB
q-exponential (q-Exp)	C	C	C	C	PB	GB
Stigler's law (SL)	NC	MC	MC	AC	no choice	no choice
Real-world FDD's	**B**	**BG**	**TSPB**	**PB**	**Choice 1**	**Choice 2**
Earth's gravity	AC	AC	C	C	PB	TSPB
Geomagnetic field	C	C	C	C	PB	GB
Seismic wavespeeds	C	C	C	C	TSPB	PB
S-A seismogram	AC	C	AC	C	PB	GB
Population of countries	AC	C	AC	C	PB	GB
Twitter users by followers	C	C	C	C	PB	TSPB
Terrorism deaths	NC	C	NC	AC	GB	no choice
Global infectious decease	AC	AC	AC	AC	GB	PB
MEG	NC	NC	NC	NC	no choice	no choice

Finally, it is worthwhile to mention some other results related to the role GB might play for some real-world data sets. For example, Pietronero et al. [31] mention that the California earthquake magnitude distribution follows the Gutenberg-Richter law [12], a power law which corresponds to a GB with approximate $\gamma = -1$. Similarly, Nigrini and Miller [30] find that the sizes of lakes and wetlands conform to a GB via power law behaviour.

4 Conclusion

The interest of parametric extensions of Benford's law in scientific research has many facets. First, parametric extensions of Benford's law often yield a better data fit than Benford's law. Second, by embedding Benford's law in a parametric FDD a simple statistical procedure to validate Benford's law is obtained. If Benford's model is sufficiently close to its embedding it is automatically revealed. The Bernegger-Lloyd distribution is a typical example for this phenomenon. Third, the use of parametric FDD's suggests a potential for improved applications in known and new fields of scientific research (e.g. fraud detection pioneered by Nigrini, prediction of earthquakes by Sambridge et al. [36], etc.). It is our hope that this brief presentation helps justify the high flexibility of the three

ISSN 2229-5518

parametric Benford model extensions to cope with data sets of Benford type (at least those with decreasing FDD's). Although there exists FDD's which exactly follow Benford's law, for example the Bose-Einstein law in physics and the Bernegger-Lloyd distribution in actuarial sciences, not all distributions with monotonically decreasing first significant digits will follow Benford's law. Given that parametric extensions of Benford's law exist in mathematics (GB in number theory) and geophysics (GB for Gutenberg-Richter law), and Galileo Galilei's quote that "the book of nature is written in the language of mathematics", one may ask researchers from any scientific discipline to reveal further FDD's that truly follow parametric extensions of Benford's law.

REFERENCES

[1] N.H.F. Beebe, "Newcomb, Benford, Pareto, Heaps, and Zipf. Are arbitrary numbers random?", preprint, 20 Feb. 2003, available at http://www.math.utah.edu/~beebe/talks/2013/benfords-law/.

[2] N.H.F. Beebe, "A bibliography of publications about Benford's law, Heap's law, and Zipf's law", version 1.79, 11 Jun. 2015, available at ftp://ftp.math.utah.edu/pub/tex/bib/benfords-law.pdf.

[3] F. Benford, "The law of anomalous numbers", *Proc. Amer. Phil. Soc.*, vol. 78, pp. 551-572. 1938.

[4] A. Berger, T. Hill and E. Rogers, "Benford Online Bibliography", 2009-15, available at http://www.benfordonline.net/.

[5] S. Bernegger, "The Swiss Re exposure curves and the MBBEFD distribution class", *ASTIN Bulletin*, vol. 27, no. 1, pp. 99-111, 1997.

[6] H. Cáceres, J.L. Pollo García, C.M. Martínez Ortiz and L. García Dominguez, "First digit distribution in some biological data sets. Possible explanations for departures from Benford's law", *Electron. J. Biomed.*, vol. 1, pp. 27-35, 2008.

[7] J.R. van Dorp and S. Kotz. "The standard two-sided power distribution and its properties: with applications in financial engineering", *The American Statistician*, vol. 56, no. 2, pp. 90-99, 2002.

[8] H.-A. Engel and C. Leuenberger, "Benford's law for exponential random variables", *Statistics and Probability Letters*, vol. 63, no. 4, pp. 361-365, 2003.

[9] L.V. Furlan, *Das Harmoniegesetz der Statistik: Eine Untersuchung über die metrische Interdependenz der sozialen Erscheinungen*. Basel, Verlag für Recht und Gesellschaft A-G, xiii:504, 1946. (Reviewed in J. Amer. Statist. Assoc., vol. 43, no. 242, pp. 325–328).

[10] W.H. Furry and H. Hurwitz, "Distribution of numbers and distribution of significant digits", *Nature*, vol. 155, no. 3924, pp. 52-53, 1945.

[11] S.A. Goudsmit and W.H. Furry, "Significant figures of numbers in statistical tables", Nature, vol. 154, no. 3921, pp. 800-801, 1944.

[12] B. Gutenberg and C.F. Richter, "Frequency of earthquakes in California", *Bull. Seismol. Soc. Amer.*, vol. 34, pp. 185-188, 1944.

[13] T.P. Hill, "A statistical derivation of the significant-digit law", *Statist. Sci.*, vol. 10, pp. 354-363, 1995.

[14] T.P. Hill, "Base-invariance implies Benford's law", *Proc. Amer. Math. Soc.*, vol. 123, pp. 887-895, 1995.

[15] W. Hürlimann, "A generalized Benford law and its application", *Adv. Appl. Statist.*, vol. 3, no. 3, pp. 217-228, 2003.

[16] W. Hürlimann, "Integer powers and Benford's law", *Int. J. Pure Appl. Math.*, vol. 11, no. 1, 39-46, 2004.

[17] W. Hürlimann, "Benford's law from 1881 to 2006: a bibliography", available at : www.arxiv.org/abs/math.ST/0607168.

[18] W. Hürlimann, "Generalizing Benford's law using power laws: application to integer sequences", *Int. J. Math. Math. Sci.*, Article ID 970284, 2009

[19] W. Hürlimann, "A first digit theorem for square-free integer powers", *Pure Math. Sci.*, vol. 3, no. 3, pp. 129-139, 2014.

[20] W. Hürlimann, "A first digit theorem for powers of perfect powers", *Commun. in Math. and Appl.*, vol. 5, no. 3, pp. 91-99, 2014.

[21] W. Hürlimann, "On the uniform random upper bound family of first significant digit distributions", *Journal of Informetrics*, vol. 9, no. 2, pp. 349-358, 2015.

[22] J. Lee, W.K.T. Cho and G.G. Judge, "Stigler's approach to recovering the distribution of first significant digits in natural data sets", *Statist. Probab. Lett.*, vol. 80, no. 2, pp. 82-88, 2010.

[23] L.M. Leemis, B.W. Schmeier and D.L. Evans, "Survival distributions satisfying Benford's law", *The American Statistician*, vol. 54, no. 3, pp. 1-6, 2000.

[24] E. Ley, "On the peculiar distribution of the US stock indexes' digits", *The American Statistician*, vol. 50, no. 4, pp. 311-313, 1996.

[25] J. Long, "Testing Benford's law", online website, available at http://testingbenfordslaw.com/ (last accessed Dec. 2014).

[26] B. Luque and L. Lacasa, "The first-digit frequencies of prime numbers and Riemann zeta zeros", *Proc. Royal Soc. , Ser. A*, vol. 465, pp. 2197-2216, 2009.

[27] J. Morrow, "Benford's law, families of distributions and a test basis", *CEP Discussion Paper*, No. 1291, London School of Economics and Political Science, London, 2014.

[28] S. Newcomb, "Note on the frequency of use of the different digits in natural numbers", *Amer. J. Math.*, vol. 4, pp. 39-40, 1881.

[29] M.J. Nigrini, *Benford's Law. Applications for forensic accounting, auditing, and fraud detection*. J. Wiley & Sons, Hoboken, New Jersey, 2012.

[30] M.J. Nigrini and S.J. Miller, "Benford's law applied to hydrological data – results and relevance to other geophysical data", *Mathematical Geology*, vol. 39, no. 5, 469-490, 2007.

[31] L. Pietronero, E. Tossati, V. Tossati and A. Vespignani, "Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf", *Physica A* , vol. 293, pp. 297-304, 2001.

[32] R.A. Raimi, "The first digit problem", *The Amer. Math. Monthly*, vol. 83, no. 7, pp. 521-538, 1976.

[33] W.J. Reed, "The Pareto, Zipf and other power laws", *Economics Letters*, vol. 74, pp. 15-19, 2001.

[34] Rodriguez, "First significant digit patterns from mixtures of uniform digits", *The American Statistician*, vol. 58, no. 1, pp. 64-71, 2004.

[35] K.A. Ross, "First digits of squares and cubes", *Math. Magazine*, vol. 85, pp. 36-42, 2012.

[36] M. Sambridge, H. Tkalčić and A. Jackson, "Benford's law in the natural sciences", *Geophysical Research Letters*, vol. 37, L22301, 2010.

[37] L. Shao and B.-Q. Ma, "The significant digit law in statistical physics", *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 16, pp. 3109-3116, 2010.

[38] L. Shao, and B.-Q. Ma, "First-digit law in nonextensive Statistics", *Phys. Rev.* E82, 2010.

[39] G. Szpiro, "Neues aus dem Reich der Primzahlen", *Neue Zürcher Zeitung*, 27 May 2009.

[40] A. Stakhov, *The Mathematics of Harmony. From Euclid to Contemporary Mathematics and Computer Science*. Series on Knots and Everything, Vol. 22. World Scientific Publishing Co. Pte. Ltd., Singapore, 2009.

[41] G.J. Stigler, "The distribution of leading digits in statistical tables", Working paper, 1945. Univ. of Chicago, Regenstein Library Special Collections, George J. Stigler Archives.

[42] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics", *Journal of Statistical Physics*, vol. 52, pp. 479–487, 1988.

[43] C. Tsallis, *Introduction to Nonextensive Statistics Mechanics – Approaching a Complex World*. Springer, New-York, 2010.

[44] M.R. Wójcik, "How fast increasing powers of a continuous random variable converge to Benford's law", *Statist. Probab. Lett.*, vol. 83, no. 12, pp. 2688-2692, 2013.

[45] M.R. Wójcik, "A characterization of Benford's law through generalized scale-invariance", *Math. Social Sciences*, vol. 71, pp. 1-5, 2014.