NEW TRENDS IN ECONOMIC FORECASTING

NICOLAE MARIUS JULA[*]
NICOLETA JULA[**]

Abstract
Economic forecasting is a dynamic domain. New methods are developed and tested and the methodology needs to be updated according to economic reality. Classical approach in methodology must be completed with latest trends in econometric analysis and the forecasting methods have to benefit from the increasing computational power of the modern software. One of the main causes of false prediction is using altered data. In this paper, we will present the new concepts for data testing, adjusted for the Romanian economy, based on Benford's law.

Keywords: *Economic forecasting, econometric analysis, data testing, econometric software, Benford's law*

1.Introduction

For creating reliable econometric model, one must rely on existing data. There are many situations when the models and the obtained results are not useful because of the initial data. The altered data sets may create false signals and the conclusions based on these signals are not in accordance with the economic reality.

For example, we have analyzed in the papers signals of political inferences in economy in order to manipulate the voters for increasing the chances of reelections. These models were based also on data provided by authorities, data related to final results in parliamentary or presidential elections. Now, we are testing the data form parliamentary elections in Romania from December, the 9th, 2012.

There are numerous useful methods that can be conducted in data analysis in order to check data correctness and authenticity. One of contemporary and efficient method is application of so-called Benford's Law.

Why using Bendfor's Law? In 1972, Hal Varian suggested that the law could be used to detect possible fraud in lists of socio-economic data submitted in support of public planning decisions. Based on the plausible assumption that people who make up figures tend to distribute their digits fairly uniformly, a simple comparison of first-digit frequency distribution from the data with the expected distribution according to Benford's law ought to show up any anomalous results. Benford's law is used extensively in United States in legal status issues, election data, macroeconomic reported data and other scientific fraud detection algorithms.

2.Benford's Law – from random numbers to political fraud

2.1. Benford's law

Benford's law has its origins in the study of American astronomer Simon Newcomb, who observed that in logarithm tables (used at that time to perform calculations) the earlier pages (which contained numbers that started with 1) were much more worn than the other pages. Newcomb's published result is the first known instance of this observation and includes a distribution on the

[*] Lecturer, PhD, Faculty of Economic Sciences, “Nicolae Titulescu” University of Bucharest (mariusjula@univnt.ro).
[**] Professor, PhD, Faculty of Economic Sciences, “Nicolae Titulescu” University of Bucharest (nicoletajula@yahoo.com).

second digit, as well. Newcomb proposed a law that the probability of a single number N being the first digit of a number was equal to log(N + 1) − log(N).

The phenomenon was again noted in 1938 by the physicist Frank Benford[1], who tested it on data from 20 different domains and was credited for it. His data set included the surface areas of 335 rivers, the sizes of 3259 US populations, 104 physical constants, 1800 molecular weights, 5000 entries from a mathematical handbook, 308 numbers contained in an issue of Readers' Digest, the street addresses of the first 342 persons listed in American Men of Science and 418 death rates. The total number of observations used in the paper was 20,229.

The law stipulates that the distribution frequency of digits in data sources. A set of numbers would satisfy Benford's law if the leading digit d {1..9} occurs with probability:

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}(1+\frac{1}{d})$$

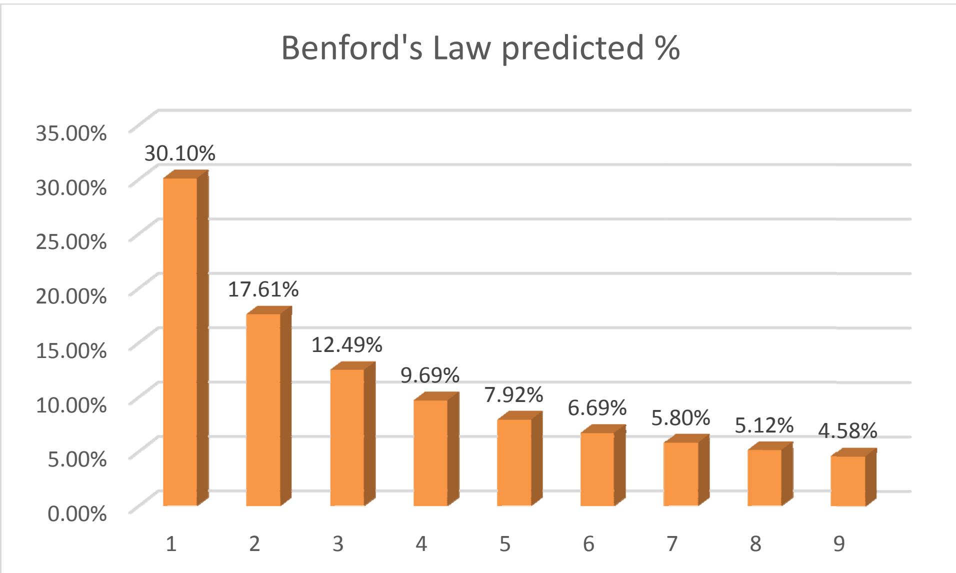

Figure 1 – Benford's distribution

The law is also valid for other bases besides decimal. Also, this law was extended to digits beyond the first. The general formula for probability that d {0..9} appears as the *n*-th (n>1) digit is:

$$\sum_{k=10^{n-2}}^{10^{n-1}-1} \log_{10}(1+\frac{1}{10k+d})$$

[1] Frank Benford, "The law of anomalous numbers", Proceedings of the American Philosophical Society 78 (4): 551–572, 1938.

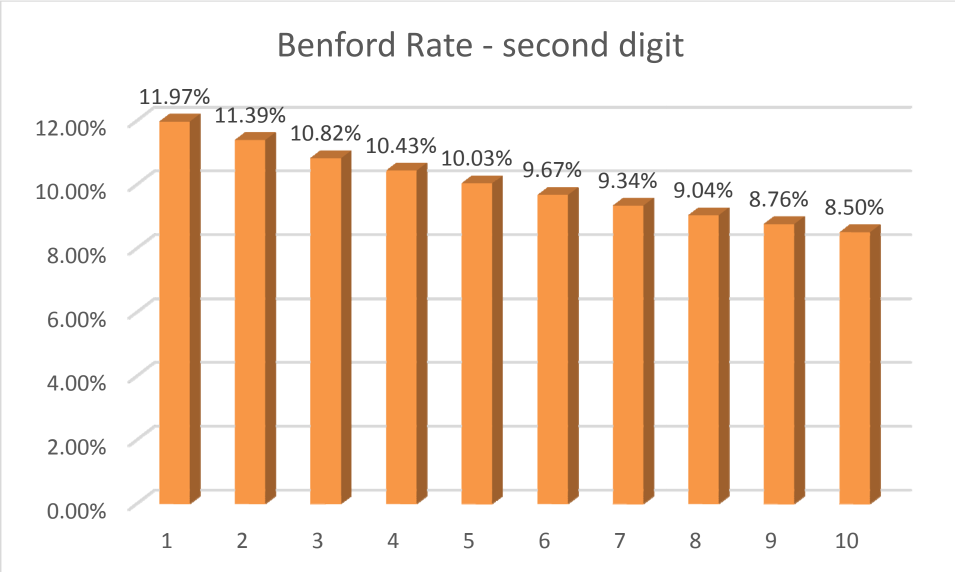

Figure 2 – Benford's law – second digit distribution

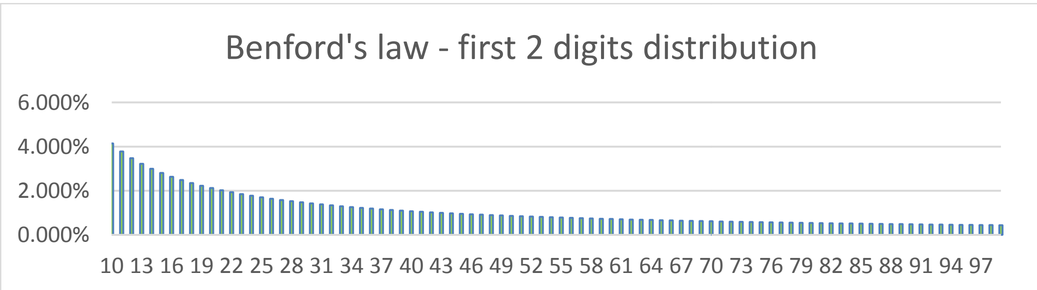

Figure 3 – Benford's law – first 2 digits distribution

2.2. Electoral data analysis

As Peter Klimeka et all[2] stated, free and fair elections are the cornerstone of every democratic society. A central characteristic of elections being free and fair is that each citizen's vote counts equal. However, Joseph Stalin believed that "it's not the people who vote that count; it's the people who count the votes." How can it be distinguished whether an election outcome represents the will of the people or the will of the counters?"

One can see elections as large-scale social experiments. A country is segmented into a usually large number of electoral units. This is also the case for Romania. Here we have electoral circumscriptions for each county, districts in Bucharest and electorate from abroad. The above-mentioned authors affirmed that "each unit may represent a standardized experiment, where each citizen articulates his/her political preference through a ballot. Although elections are one of the central pillars of a fully functioning democratic process, relatively little is known about how election fraud impacts and corrupts the results of these standardized experiments."

[2] Peter Klimeka, Yuri Yegorovb, Rudolf Hanela, Stefan Thurner (2012), Statistical detection of systematic election irregularities, Proceedings of the National Academy of Sciences of the United States of America, no. 41, vol 109, 16469–16473.

The specific literature acknowledge that there is an overabundance of ways of tampering with election outcomes (for instance, the redrawing of district boundaries known as gerrymandering or the barring of certain demographics from their right to vote, blocking access to voting locations). Some practices of manipulating voting results leave traces, which may be detected by statistical methods. Recently, Benford's law experienced a new start as a potential election fraud detection tool. In its original and naive formulation, Benford's law is the observation that, for many real world processes, the logarithm of the first significant digit is uniformly distributed. Deviations from this law may indicate that there are chances of data to be incorrect. For instance, suppose a significant number of reported vote counts in districts is completely made up and invented by someone preferring to pick numbers, which are multiples of 10. The digit 0 would then occur much more often as the last digit in the vote counts compared with uncorrupted numbers. Voting results from Russia[3], Germany[4], Argentina[5], and Nigeria[6] have been tested for the presence of election fraud using variations of this idea of digit-based analysis. There are also analysts who stipulate that the validity of Benford's law as a fraud detection method is subject to controversy. Peter Klimeka suggests that "the problem is that one needs to firmly establish a baseline of the expected distribution of digit occurrences for fair elections. Only then it can be asserted if actual numbers are over- or underrepresented and thus, suspicious. What is missing in this context is a theory that links specific fraud mechanisms to statistical anomalies[7]."

Walter Mebane[8] supports the idea of using Benford's law: "why should Benford's Law apply to vote count data?" He offers two mechanisms for why second digits of vote counts should follow a "Benford's Law-like distribution" which he refers to as the 2BL distribution. As stated above, Benford's Law does not apply to "simple random" data. Therefore, in order for Benford's Law to apply to vote count data vote count data cannot be generated simply randomly. Instead, due to the complexity inherent in the voting process, simple randomness should not be observed in voting outcomes. Thus, vote choice is not simply a "stochastic choice", but rather consists of a set of complex processes. Such processes are as follows. An individual voter first decides whether or not to vote and, secondly, who to vote for (or also which way to vote on a particular referendum). Finally, a voter must actually cast his or her ballot which can be done in a variety of ways: "election day voting in person, early voting, provisional ballots or mail-in ballots; on paper, with machine assistance or using some combination". In addition, there is always the potential for mistakes:

When all is said and done, most voters will look at each option on the ballot and have firm intentions either to select that option or not to select that option. Then for whatever reason—momentary confusion, bad eyesight, defective voting technology—a small proportion of those intended votes will not be cast or recorded correctly. A small proportion will be "mistakes" (Mebane).

The combination of the potential for mistakes and the set of complex processes produce vote counts that should "or will often" follow the 2BL distribution as laid out in Table 1. According to

---

3 Mebane WR, Kalinin K (2009) Comparative Election Fraud Detection. (The American Political Science Association, Toronto, ON, Canada).

4 Breunig C, Goerres A (2011) Searching for electoral irregularities in an established democracy: Applying Benford's Law tests to Bundestag elections in unified Germany. Elect Stud 30:534–545.

5 Cantu F, Saiegh SM (2011) Fraudulent democracy? An analysis of Argentina's infamous decade using supervised machine learning. Polit Anal 19:409–433.

6 Beber B, Scacco A (2012) What the numbers say: A digit-based test for election fraud. Polit Anal 20:211–234.

7 Deckert JD, Myagkov M, Ordeshook PC (2011) Benford's Law and the detection of election fraud. Polit Anal 19:245–268.

8 Mebane, Walter R., Jr. 2006b. "Election Forensics: The Second-digit Benford's Law Test and Recent American Presidential Elections." Earlier version presented at the Election Fraud Conference, Salt Lake City, Utah, September 29-30, 2006.

Mebane, "the kind of complexity that can produce counts with digits that follow Benford's Law refers to processes that are statistical mixtures (e.g., Janvresse and de la Rue 2004), which means that random portions of the data come from different statistical distributions" (Mebane).

Mebane uses simulations to show that when manipulations occur to 2BL distributed vote counts, this "will produce a significantly large value" of the test statistics. He shows that the test statistic is sensitive to departures from the 2BL distribution under a variety of scenarios: (1) when electoral manipulation occurs in a precinct for an already strong candidate; (2) when vote counts are manipulated in a close election (tie is expected according to the vote-generating process); and (3) when votes are manipulated in a precinct for a weak candidate. In addition, he shows that even small manipulations will produce significance (i.e. test statistic is sensitive to small manipulations). In other words, a massive amount of fraud does not have to occur for this test to detect fraud. However, "if the amount of manipulation is sufficiently small, the 2BL test will not signal that manipulation has occurred"

We have analyzed the data from parliamentary elections from 9th December, 2012. The selected data was from the official results published by Central Electoral Bureau (Biroul Electoral Central – BEC) – Final results – Per candidate statistics[9].

First analysis was made using the whole set of data, counting 8120 records. The data was recorded per electoral circumscriptions (41 counties, Bucharest with 6 districts and voters from abroad). We were looking into the column of "Obtained votes per candidate" and we are searching for differences from the Benford's law.

The tests included 3 phases:

- First digit distribution
- Second digit distribution
- First 2 digits distribution

For all these distributions, we were using also the chi-square test, which is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories. Here, we were testing if the observed distribution differs significantly from the Benford's law predicted distribution.

2.2.1. First digit distribution

Digit	Sample Frequency	Benford Rate	Sample Data Rate	Difference
1	2642	30.103%	32.585%	0.02482102
2	1498	17.609%	18.476%	0.00866454
3	1083	12.494%	13.357%	0.00863304
4	725	9.691%	8.942%	-0.00749215
5	583	7.918%	7.190%	-0.00727695
6	459	6.695%	5.661%	-0.01033603
7	420	5.799%	5.180%	-0.00619126
8	360	5.115%	4.440%	-0.00675193
9	338	4.576%	4.169%	-0.00407027

Table 1 – First digit test

[9] http://www.becparlamentare2012.ro/A-DOCUMENTE/Statistici/RezultateCandidati2012.xls

ChiTest: 99.99999999903%

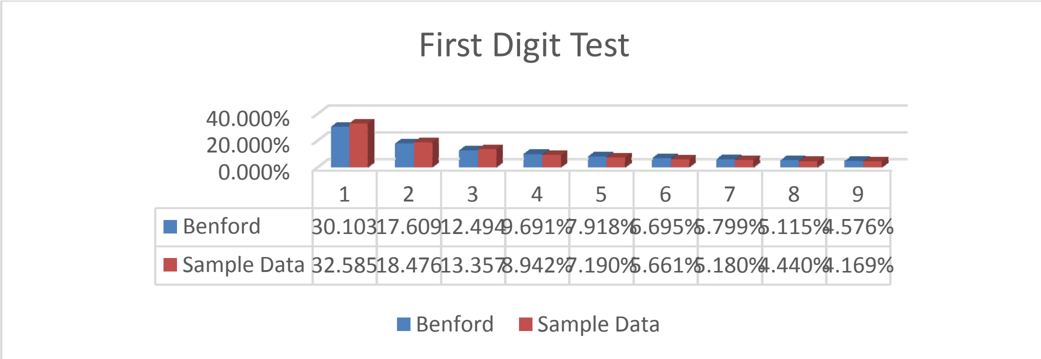

Figure 4 – First digit distribution

Conclusions: We observed a slight difference from the Benford's distribution, mainly the first 3 digits' frequency being above expected value and the others below. The Chi Square test indicates that the 2 distributions are almost identical, meaning that the test is not able to suggest a fraud in the data.

2.2.2. Second digit distribution

Digit	Sample Frequency	Benford Rate	Sample Data Rate	Difference
0	825	11.97%	11.84%	-0.0012986
1	846	11.39%	12.14%	0.0075047
2	782	10.82%	11.22%	0.0039912
3	728	10.43%	10.45%	0.0001326
4	702	10.03%	10.07%	0.0004218
5	686	9.67%	9.84%	0.0017559
6	625	9.34%	8.97%	-0.0036871
7	640	9.04%	9.18%	0.0014853
8	578	8.76%	8.29%	-0.0046313
9	557	8.50%	7.99%	-0.0050746

Table 2 – Second digit test

ChiTest: 100.00%

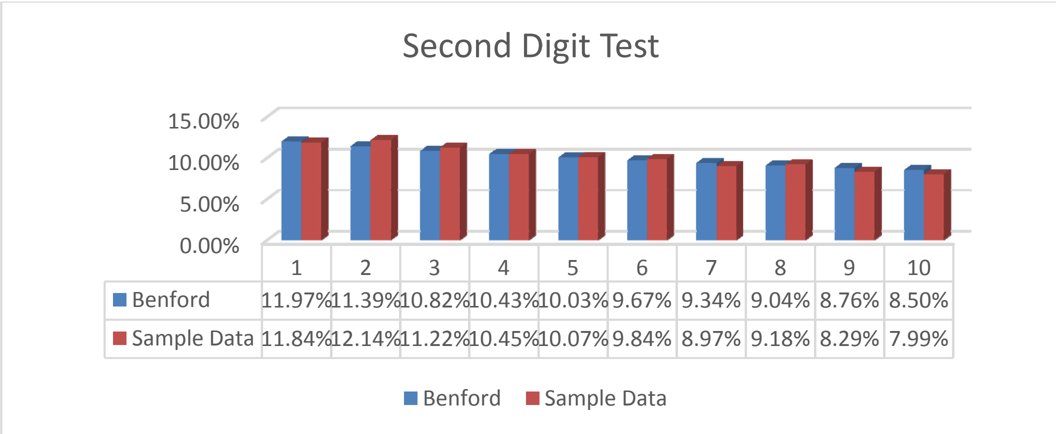

Figure 5 – Second digit distribution

Conclusions: The second digits 2 and 3 are above expected frequencies. The differences are insignificant according to Chi Square test.

2.2.3. First two digits distribution

Digits	Benford Rate	Sample Rate	Sample Frequency
10	4.139%	4.721%	329
11	3.779%	4.405%	307
12	3.476%	4.219%	294
13	3.218%	4.204%	293
14	2.996%	4.061%	283
15	2.803%	3.702%	258
16	2.633%	3.343%	233
17	2.482%	3.272%	228
18	2.348%	2.769%	193
19	2.228%	2.784%	194
20	2.119%	2.497%	174
…	…	…	…
96	0.450%	0.187%	13
97	0.445%	0.230%	16
98	0.441%	0.301%	21
99	0.436%	0.158%	11

Table 3 – First two digits test

ChiTest: 100.00%

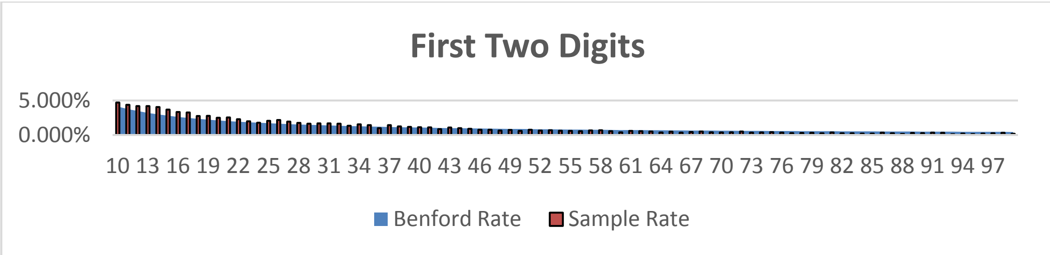

Figure 6 – First two digits distribution

Conclusions: The values starting with 1, 2 and 3 seem to have a slight value above expected distribution. All other starting digit pairs are under expected frequencies. This can suggest a possible data alteration (example: for a candidate with 24xx vote to 23xx or 25xx). The hypothesis is unstained by the Chi Square test.

3.Conclusions

Using Benford's law to discover data manipulation in the final results of the parliamentary elections from December, 2012, we can conclude that the data are validated. No important signs of abnormal distribution were detected.

As a further analysis, we recommend the analysis of the data using more fraud detection methods. This method is not exhaustive, as some economist suggest. This is not a perfect validation tool, is more like a test. If the data is in a category which supposed to obey Benford's law (like election data) and it fails, there is a signal of possible fraud. If the test is passed, doesn't mean the data is automatically validated, but more tests are always recommended to increase the confidence level.

References

- Battersby S.,"Statistics hint at fraud in Iranian election "(New Scientist: 24 June 2009).
- Beer TW, "Terminal digit preference: beware of Benford's law", J Clin Pathol 62 (2009): 192.
- Benford F., "The law of anomalous numbers". Proceedings of the American Philosophical Society 78 (1938): 551–572.
- Buttorff, G. Detecting fraud in America's gilded age. Working Paper No. 2, University of Iowa, (Cal Tech/MIT Voting Technology Project:2008).
- Cho WKT, Gaines BJ, " Breaking the (Benford) law: Statistical fraud detection in campaign finance." Amer Stat 61, (2007): 218–223.
- Deckert J., Mikhail Myagkov and Peter C. Ordeshook, "The Irrelevance of Benford's Law for Detecting Fraud in Elections", (Caltech/MIT Voting Technology Project: 2009) Working Paper No. 9.
- Diekmann A, "Not the First Digit! Using Benford's Law to detect fraudulent scientific data". J Appl Stat 34 (3) 2009:321-329, doi:10.1080/02664760601004940.
- Dümbgen L, Leuenberger C., "Explicit bounds for the approximation error in Benford's law." Elect. Comm. in Probab 13(2008): 99–112.
- Duncan R. L., "An Application of Uniform Distribution to the Fibonacci Numbers", The Fibonacci Quarterly, 5, (1967): 137–140.
- Durtschi C, Hillison W, Pacini C, "The effective use of Benford's Law to assist in detecting fraud in accounting data". J Forensic Accounting 5 (2004): 17–34.
- Fewster, R. M.,"A simple explanation of Benford's Law". The American Statistician 63 (2009): 26–32. doi:10.1198/tast.2009.0005.

- Hill T. P., "A Statistical Derivation of the Significant-Digit Law" (PDF). Statistical Science 10 (1995): 354–363.
- Hill T. P., "The first digit phenomenon" (PDF). American Scientist 86 (July–August 1998): 358.
- Hill T. P., "The Significant-Digit Phenomenon", The American Mathematical Monthly, Vol. 102, No. 4, (Apr., 1995): 322–327.
- Janvresse E. and Thierry de la Rue, "From Uniform Distributions to Benford's Law", Journal of Applied Probability, 41 (2004):1203-1210 doi:10.1239/jap/1101840566 MR2122815.
- Leemis LM, Schmeiser BW, Evans DL, " Survival distributions satisfying Benford's Law". Amer Stat 54, (2000):236–241.
- Mebane, Walter R., Jr., "Election Forensics: The Second-digit Benford's Law Test and Recent American Presidential Elections." Earlier version presented at the Election Fraud Conference, Salt Lake City, Utah, September (2006):29-30.
- Müller, Hans Christian." Greece Was Lying About Its Budget Numbers". (Forbes. September 12, 2011.).
- Newcomb S., "Note on the frequency of use of the different digits in natural numbers". American Journal of Mathematics (American Journal of Mathematics, Vol. 4, No. 1) 4 (1881): 39–40. doi:10.2307/2369148.
- Nigrini M , "A taxpayer compliance application of Benford's Law." Amer. Tax Assoc. 18, (1996): 72-91.
- Nigrini M.J., "I've Got Your Number". Journal of Accountancy, (May 1999).
- Raimi RA, "The first digit problem". American Mathematical Monthly, 83 (1976): 521–538.
- Sarkar P. B., "An Observation on the Significant Digits of Binomial Coefficients and Factorials", Sankhya B, 35, (1973): 363–364.
- Scott PD & Fasli, "Benford's Law: An empirical investigation and a novel explanation". CSM Technical Report 349 (2001).
- Smith S.W., "The Scientist and Engineer's Guide to Digital Signal Processing, chapter 34, Explaining Benford's Law". (2012).
- Stephens M.A, "Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables". Journal of the Royal Statistical Society, Series B (Methodological), 32, (1970):115–122.
- Suh I.S., Headrick T.C., "A comparative analysis of the bootstrap versus traditional statistical procedures applied to digital analysis based on Benford's Law", Journal of Forensic and Investigative Accounting 2(2) (2010):144-175.
- Suh IS, Headrick TC, Minaburo S, "An effective and efficient analytic technique: A bootstrap regression procedure and Benford's Law." J Forensic & Investigative Accounting 3 (3) (2011).
- Tolle C.R., Joanne L. Budzien, and Randall A. LaViolette "Do dynamical systems follow Benford?s law?", Chaos 10, 2, pp.331-336 (2000); doi:10.1063/1.166498.
- Varian, Hal. "Benford's law". The American Statistician 26: 65.
- Washington L. C., "Benford's Law for Fibonacci and Lucas Numbers", The Fibonacci Quarterly, 19.2, (1981): 175–177.