Benford's Law and the Risk of Financial Fraud

What is Benford's Law, and can it be applied to detect financial fraud? Controversies surrounding the integrity of LIBOR setting and reported sovereign economic data serve as examples that Benford fraud detection is sometimes misleading. This article suggests best practices for Benford's Law analysis.

By J. M. Pimbley

What the world calls "Benford's Law" is a marvelous mélange of mathematics, data, philosophy, empiricism, theory, mysticism and fraud. Even with all these qualifiers, one can easily describe this law and then ask the simple questions that have challenged investigators for more than a century.

Take a large collection of positive numbers, which may be integers or real numbers or both. Focus only on the first (non-zero) digit of each number. Count how many numbers of the large collection have each of the nine possibilities (1-9) as the first digit. For typical number collections – which we'll generally call "datasets" – the first digit is not equally distributed among the values 1-9. Instead, a first digit of "1" occurs roughly with frequency 30%, while the frequency of first digit "9" is just 4.6%.

That's the easy description. The simple questions include the following: Is this really true? Why is it true? Why *should* it be true? Does it ever fail?

In terms of written history, Simon Newcomb (1881) published the first known discussion of and solution to this "distribution of digits" problem. The next development was Frank Benford's analysis (1938) of this same topic. Benford had not been aware of the earlier Newcomb discovery. Newcomb and Benford both provided intriguing data, and reached the same conclusion regarding the mathematical form of the digit distribution; however, both also failed to convince subsequent investigators of the validity of their explanations.

In 1976, Ralph Raimi provided an excellent history and summary of Benford's Law research. Theodore Hill published a quasi-proof of the Law in 1995,[1] while Mark Nigrini (1999) has applied the Law in numerous accounting contexts.

Even more recently, Rachel Fewster (2009) wrote an enjoyable history and some valuable thoughts on applicability of Benford, but our favorite discussion, due to his unique insights, is Steven Smith's (2007) explanation of this Law.

Our primary interest in this topic is the alleged utility of Benford's Law in detecting fraud in financial and accounting contexts. We develop a straightforward understanding of Benford more generally, and we apply this understanding to the potential detection of financial fraud. Litigated disputes, in which Benford's Law is not nearly as appropriate or helpful as some parties believe, do exist.

Fraud Detection with Benford's Law

The subtitle to Nigrini's 1999 article reads: "How a mathematical phenomenon can help CPAs uncover fraud and other irregularities." Nigrini's assertion is that the auditor or accountant can determine the distribution of first digits in datasets arising from accounts payable data; estimates in the general ledger; customer refunds; and numerous other potentially misstated financial information. Significant deviation of the observed first-digit distribution from Benford's Law implies possible fraud.

In Nigrini's words, "because human choices are not random, invented numbers are unlikely to follow Benford's Law." He also explains that a dominant "behavioral feature" of fraudulent values is the dishonest person's desire to keep numbers (such as costs or expenses) beneath a specific threshold value, presumably to escape the notice of auditors or managers.

One could imagine a similar desire to falsely *inflate* values, such as returns on equity or economic growth rates, above a threshold value. In 2005, Nigrini also applied his Benford review to Enron's accounting.

Reporting and prices in the financial world offer abundant possibilities for the application of Benford's Law to search for fraud. Investigation of alleged fraud in the setting of LIBOR values is a timely example.[2] Since alleged manipulation of index values pertaining to the interest rate swap, foreign exchange, commodity and other markets is similar in many respects to LIBOR allegations, Benford's Law may provide a forensic tool in many upcoming investigations.

Another current financial controversy is the sovereign reporting of economic data.[3] Sovereign bond investors, the *International Monetary Fund* (IMF), credit rating agencies, economists, voters, taxpayers and others would have reason to apply tests such as Benford's Law to scrutinize the veracity of reported information.

The review procedure is straightforward. Simply assemble the data of interest (e.g., annual changes in gross domestic product, inflation rates and government spending), run a short algorithm to determine the distribution of first digits and compare the resulting distribution to that of Benford's Law.[4]

Explanation of Benford's Law

To re-state the Benford problem, first let's find or generate a large dataset of positive numbers. The first non-zero digit of each number will take one of the values k with k=1,2,3,⋯,9. If the dataset conforms with Benford's Law, the distribution of first digits p_k will be[5]

$$p_k \; = \; \log\left(\frac{k+1}{k}\right), \;\; k = 1, 2, 3, \cdots, 9 \qquad (1)$$

In this expression, "log" is the base-ten logarithm. This Benford distribution is not uniform and, therefore, not intuitive. One's unthinking, though reasonable, impression would be that a "9" is as likely to appear as a "1" as a leading digit for a collection of apparently random numbers. But the p_9 and p_1 values from equation (1) are 4.6% and 30.1%, respectively. Thus, a "1" is more than six times as likely to occur as a "9" in the leading digit of a Benford collection of numbers.

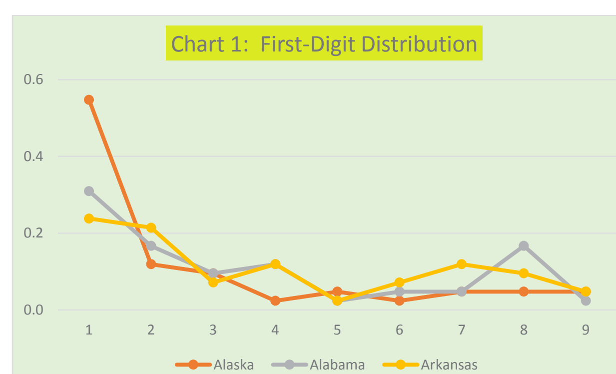
Previous authors – such as Benford (1938), Raimi (1976), Nigrini (1999), Fewster (2009), Smith (2007) and many others we do not cite[6] – provide datasets to show varying degrees of conformance with Benford's Law. There are dataset examples as well that do not comport with Benford. We provide our own "new contribution" here.

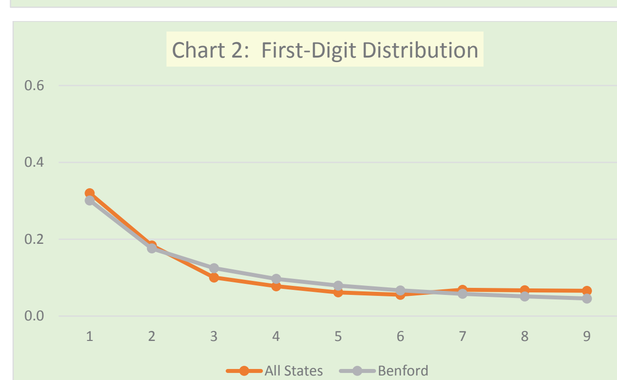
From the *Federal Housing Finance Agency* (FHFA) website, we downloaded the average residential mortgage loan amount by state in the U.S. for every year from 1969 through 2010.[7] Given this span of years, there are 42 data points per state, and more than 2,000 data points upon aggregation of all the states. Table I below shows the observed first-digit distribution p_k for all aggregated states and also for the first three states:

Table I: First-Digit Distribution for Average Loan Amount by State for 1969-2010

Digit	All States	Alaska	Alabama	Arkansas
1	0.320	0.548	0.310	0.238
2	0.184	0.119	0.167	0.214
3	0.101	0.095	0.095	0.071
4	0.078	0.024	0.119	0.119
5	0.062	0.048	0.024	0.024
6	0.056	0.024	0.048	0.071
7	0.068	0.048	0.048	0.119
8	0.067	0.048	0.167	0.095
9	0.066	0.048	0.024	0.048

Charts 1 and 2 below plot the distributions of Table I and the Benford distribution.





The distributions for the individual states of Chart 1 differ from each other and do not approximate the Benford Distribution. Chart 2 shows, however, that the aggregated data for all U.S. states is much closer to Benford.[8] This observation that aggregating multiple datasets produces better agreement with Benford is unanimous among investigators and traces back to Benford (1938).

When Should the Data Follow Benford's Law?

In a separate study, we found in Pimbley (2014) many example datasets that do not produce Benford Distributions. The "uniform distribution" – including $U(0,1)$ to $U(0,10^m)$; the "flat-top algebraic" distribution; the $m>1$ "pure algebraic" distribution; and the "positive normal" distribution – are all clearly non-Benford.

The "positive exponential" distribution is somewhat closer to Benford in some cases. The "log-normal" distribution converges precisely to Benford, as far as one can judge with numerical calculations. The x^{-1} "pure algebraic" distribution is exactly Benford. An indisputable conclusion is that one may not assume the first digits of an empirical dataset should obey Benford's Law.

We do not believe any study prior to Pimbley (2014) considered all of these example datasets to show Benford versus non-Benford behavior.[9] But virtually all prior research addresses the topic of how to determine which datasets will conform to Benford and which will not. Two clear principles are that the dataset should span several orders of magnitude (e.g., Fewster (2009)) and that *mixtures* of datasets are more likely to produce the Benford distribution for digits (e.g., Benford (1938), Raimi (1976) and Hill (1995)) than individual datasets.

The empirical requirement to span several orders of magnitude is understandable, by reference, to example. One of the Pimbley (2014) numerical cases is the normal distribution centered at x=15 with standard deviation $\sigma=1$. This distribution is almost entirely contained between 10 and 20.

The probability of the first digit being "1" is essentially 100%. To have any non-zero probability of finding a "9" as first digit, the distribution must extend upward to 90 and/or downward below 10. Thus, a dataset distribution must at least span one order of magnitude. An example of this sort provides intuition but not a solution.

Smith (2007) provided a creative and convincing condition that a probability density function (PDF) $f(x)$ must satisfy to be consistent with a Benford distribution.[10] The derivation of first-digit probabilities p_k with k=1,2,3,⋯,9 is relatively straightforward to express (if not solve) for arbitrary $f(x)$, as follows:

$$p_k = \sum_{n=-\infty}^{+\infty} \int_{k\cdot 10^n}^{(k+1)10^n} f(x)\,dx \qquad (2)$$

The limits of integration for each n specify the range in x pertaining to each digit k. Since the widths of these infinite number of "windows" are all equal under a logarithmic transformation, Smith (2007) applied the transformation $Y=\log X$ to work with the PDF $g(y)$ in this transformed variable y. Thus, we can write p_k as

$$p_k = \int_{-\infty}^{+\infty} dy\, g(y)\, W_k(y) \qquad (3)$$

in which the $W_k(y)$ is an infinite sum of "window functions" equal to one in the intervals $n+\log k < y < n+\log(k+1)$, and equal to zero otherwise.

At this point, Smith (2007) adds a "scale variable" to measure how the p_k change when all "x-space" data values are multiplied by a common factor. In the logarithmic "y-space," this scaling is additive. This implied additional property of scale invariance has a long history in the scrutiny of Benford's Law (see Raimi (1976)). Denoting the arbitrary logarithmic shift as $\boldsymbol{\alpha}$,

$$p_k(\alpha) = \int_{-\infty}^{+\infty} dy\, g(y-\alpha)\, W_k(y) \qquad (4)$$

Taking the Fourier transform of equation (4) gives the transformed $\tilde{p}_k(\lambda)$ $\left(= \int_{-\infty}^{+\infty} d\alpha\, p_k(\alpha) e^{-i\lambda\alpha}\right)$ as a product (by the convolution theorem) of two individual transforms representing g and W_k. The transform for W_k gives an infinite sum of Dirac delta functions at evenly spaced frequencies $\lambda=2n\pi$ for $n\epsilon(-\infty,+\infty)$. Still following Smith (2007), the criterion for satisfaction of Benford's Law is that the transformed $\tilde{g}^*(\lambda)$ be small or zero at all non-zero "window frequencies" $2n\pi$ $(n\neq 0)$. Considering the magnitude of relevant terms, we write this "small" condition somewhat subjectively as

$$\left| \int_0^{\infty} dx\, f(x)\, e^{\pm i 2n\pi \log x} \right| < 0.1 \qquad \text{for all } n \neq 0 \qquad (5)$$

after transforming back to the "x-space."

Inequality (5) is not the friendliest expression we've ever encountered, but it's tractable in several cases of interest. For example, recall the PDF for the uniform distribution $U(0,1)$: $f(x)=1$ when $0<x<1$ and equals zero for all other values of x. Plugging this functional form into the left-hand side of (5), we get $|1\pm i2n\pi|^{-1}$.

This result clearly does not satisfy the inequality of (5) and, therefore, the distribution $U(0,1)$ fails our "Benford Test." Had we employed $U(0,10^m)$ rather than $U(0,1)$, we would have obtained precisely the same outcome. Failure of the Benford test, of course, is what we expect, given the direct calculation of p_k for the uniform distribution in Pimbley (2014).[11]

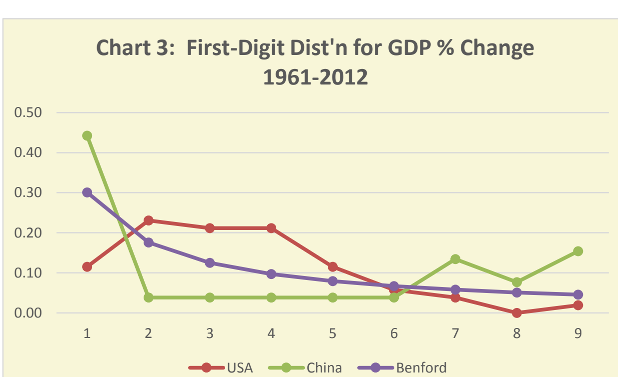
Errors in the Application of Benford's Law to Fraud Detection

The challenge in applying Benford's Law to the detection of fraud is that legitimate, naturally occurring datasets *often* obey this law, but there is no guarantee that they must as we found in the last section.

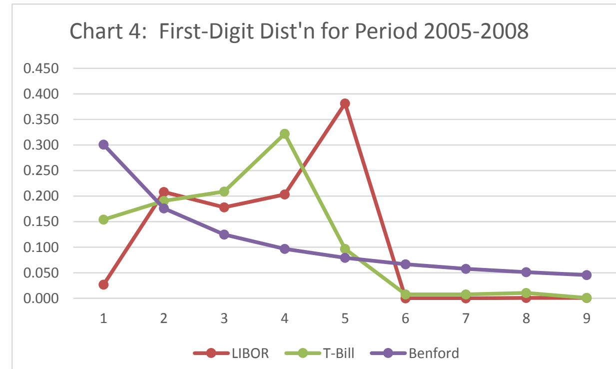
In the appendix, we discuss the conditions required for an empirical dataset to satisfy Benford's Law. The best conclusion we or anybody find is that the dataset must have a range that extends over several orders of magnitude or possibly consist of a quasi-random mixture of smaller groups of data.[12]

Financial analysts, therefore, are likely to err when they apply Benford's Law to a dataset that does *not* span several orders of magnitude. Consider the news article we cited earlier arguing that China's gross domestic product (GDP) data violates Benford's Law.[13] One's immediate analysis should find that the (absolute value) annual percentage change in GDP is not expected to span "several orders of magnitude." A maximum likely reported range is 0.1% to 20% or so. Though this range exceeds two orders of magnitude, it's likely that the true effective range of a sovereign's GDP growth data would be confined.

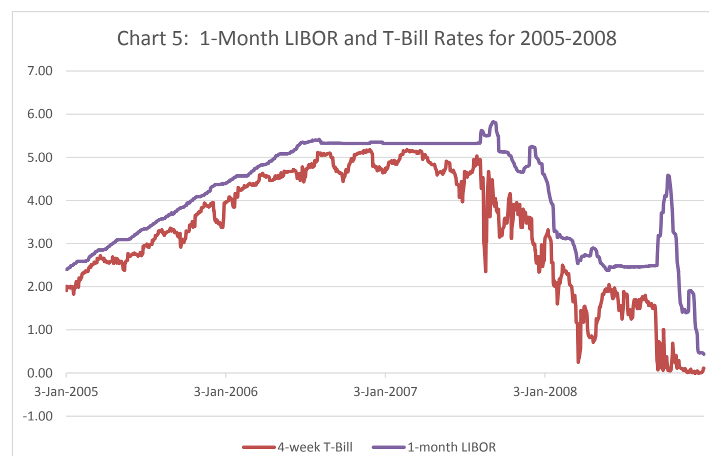
Chart 3 (below) plots first-digits distributions for 50 years of GDP growth data for China and the United States relative to the Benford distribution.[14] Neither country's GDP growth comports with the Benford distribution. It is far more likely that GDP growth is simply not a statistic likely to match the Benford conditions than that the GDP data of the two countries are fraudulent.



Consider next the allegations that Benford's Law implies fraudulent LIBOR settings of recent years.[16] Chart 4 shows the first-digit distributions for (one-month) LIBOR and Treasury bill ("T-Bill") for this interest rate data from early 2005 to late 2008.[17]



Neither LIBOR nor T-Bill data match the Benford distribution. Plotting the actual LIBOR and T-Bill values for this period, it's clear the numerical values do not span "several orders of magnitude" (see Chart 5, below). Thus, it's not surprising that Chart 4 shows no agreement with Benford.[18]



Ironically, a simple glance at Chart 5 might persuade a skeptical analyst that LIBOR settings are contrived. Both the increasing LIBOR period of early 2005 to mid-2006 and the stable period thereafter are "too smooth." But we cannot cite Benford's Law as the basis for this skepticism. Benford's Law simply does not apply to datasets without broad variation.

We should note that both the China GDP study and the LIBOR study employed the Benford's Law distribution for the *second* digit rather than the *first* digit. The authors evidently recognized the arguments we make here that Benford's Law should not apply to the GDP and LIBOR datasets. Yet it is a different error to assume a Benford *second*-digit law will hold when the Benford *first*-digit law does not.

As is clear in Newcomb (1881) and Hill (1995), the analytical foundation is the same for both first and second digits. That is, this is simply a mathematical syllogism. A Benford-compliant dataset (see the appendix) satisfies conditions such that we can derive the expected distributions of first and second digits to get what the world calls “Benford’s Law” for the two distributions. There is no Benford variant that deliberately derives a second-digit distribution, for example, under conditions for which it is clear the dataset does not conform to the first-digit Benford distribution. If the first-digit law does not apply, then neither does the second.[19]

Best Practices in the Application of Benford’s Law to Fraud Detection

The primary difficulty in applying Benford’s Law to the detection of fraud is that many datasets *do not* naturally satisfy Benford’s Law. While some datasets *do* largely follow the Benford behavior, there is no “bright line” test to distinguish the two types.

As a result, we identify three best practices in this analysis. First, review the candidate dataset to gain comfort that the numerical values span “several orders of magnitude.” Second, always include an additional, comparable dataset for the Benford review. Third, realize that a Benford result implying potential fraud is merely a flag for review and not a standalone indicator of fraud.

In this article – in Table I and Charts 1, 3 and 4 – we have directly subjected more than one dataset to comparison to the Benford distribution of first digits. When China’s GDP data diverged from Benford in Chart 3, for example, the analyst recognizes immediately that similar U.S. data diverges as well. The “Benford failure” of China GDP data would only have been striking and worth pursuing if the comparable dataset (U.S. GDP data in this case) had conformed closely to the Benford distribution.

Now consider a case in which the first-digit distribution of a financial dataset does deviate significantly from the Benford distribution, the dataset numerical values are broadly varying, and comparable datasets do match Benford. This is the scenario in which Benford identifies potential fraud or other modification of data.

The Benford result cannot tell us *which* values are “wrong” or provide any diagnostic information beyond the distribution of the first digits. Further, in a scientific or litigation sense, it is not possible to prove that a non-fraudulent dataset *would* follow a Benford distribution. Hence, the role of the Benford test is simply to flag specific datasets for scrutiny rather than to allege data irregularity with failure of the Benford review as one supporting statement.

Application of Benford’s Law will be fruitful in some analyses. But financial risk managers and investigators should always apply generalized common sense, curiosity, skepticism, models and diverse automated procedures to the review of data integrity.

Application of Benford’s Law will be fruitful in some analyses. But financial risk managers and investigators should always apply generalized common sense, curiosity, skepticism, models and diverse automated procedures to the review of data integrity.

Closing Thoughts

This article discussed the origin and meaning of Benford’s Law and its application to the detection of financial fraud. There are many datasets that *do not* obey Benford’s Law and there is no strict ability to determine which datasets naturally follow the Benford distribution and which do not.

Consequently, analytical errors and misapplications occur. We identified past studies alleging fraud in sovereign economic data and LIBOR settings in this discussion. While it is entirely possible that fraud does exist in these contexts, Benford’s Law is often the wrong tool to detect such fraud.

Our new research in the mathematics underlying Benford’s Law includes the development of new dataset examples, the analytical solutions for new probability density functions and the explanation of the Benford distribution in terms of a logarithmic transformation with “bin shifting.” These new results, combined with our review of recent fraud allegations, produced our suggested best practices for application of Benford’s Law.

Joe Pimbley (PhD) is Principal of Maxwell Consulting, a consulting firm he founded in 2010, and a member of Risk Professional’s editorial board. His recent and current engagements include financial risk management advisory, underwriting for structured and other financial instruments, and litigation testimony and consultation. In a prominent engagement from 2009 to 2010, Joe served as a lead investigator for the Examiner appointed by the Lehman bankruptcy court to resolve numerous issuers pertaining to history’s largest bankruptcy.

APPENDIX: Why do Many Datasets Obey Benford's Law?

There exists a relatively simple explanation for the appearance of the Benford distribution of digits in sufficiently large and broad datasets. This explanation is entirely consistent with the Hill (1995) theorem regarding a limiting distribution (reminiscent of the Central Limit Theorem). Earlier authors – Newcomb (1881), Benford (1938), Raimi (1976) and others – had also recognized the importance of a mixture of distributions. Let's put this history aside momentarily and consider a different angle.

Reading the first digit of each number in a large dataset is a sequence of mathematical operations, which occur as follows:

(i) take the base-ten logarithm;

(ii) shift this step (i) value by the unique integer value that produces a result in the (logarithm) range *(0,1)*;

(iii) raise 10 to the power of this step (ii) value (which is the inverse transformation to step (i)); and

(iv) read the integer part of this step (iii) value, which will be a number in the range 1, 2, …, 9.

As an example, consider the number 4,371.7. While we can see immediately that the first digit is "4," we must instead apply the mathematical sequence we just defined (or its equivalent) for a coded algorithm.

The base-ten logarithm of step (i) is 3.6407 (to five significant digits). To do the translation to the range *(0,1)* of step (ii), we must subtract the integer "3" to get 0.6407. Raising 10 to the power of this value of step (iii) gives 4.3717. Finally, reading the integer part of 4.3717 (the "int" operation in many computer languages) gives the result of "4" for this step (iv).

Of these four steps, it is the second "shifting step" that is critical to the Benford distribution. The shifting step takes the values in every integer range – such as *(3,4), (11,12), (-6,-5), et cetera* – and combines them all into the range *(0,1)*. If the ultimate, aggregated collection of values shifted into *(0,1)* is uniform in this range, then the transformation of step (iii) to the x-range $x \in (1,10)$ will have a PDF $f(x) \sim x^{-1}$. As we determined earlier, it is this specific PDF that produces the Benford digit distribution.

We consider it reasonable and plausible that the "shifting step" will have the tendency to produce uniform distributions within the logarithm range *(0,1)* when the logarithm of the dataset has elements in many of the integer "bins." That is, when the original dataset spans several orders of magnitude, then several "bins" will be populated and the sum of contents of several "bins" may produce an approximation of a uniform distribution in *(0,1)*.

These statements are mere conjecture. Yet we propose the *ansatz* that naturally occurring datasets will tend to produce uniform distributions under the transformation and shifting of steps (i) and (ii) above when the number of "populated bins" is sufficiently large.

Given the *ansatz*, we then claim that the existence of the Benford Distribution is due only to the choice of transformation in step (i) (and its inverse in step (iii)). We get the Benford distribution (for many but not all datasets) when the step (i) transformation is the base-ten logarithm. We would find an entirely different distribution of first digits if we chose a transformation other than the base-ten logarithm.

REFERENCES

S. Newcomb. "Note on the Frequency of Use of the Different Digits in Natural Numbers," *Am. J. Math.*4(1), 39-40, 1881.

R. M. Abrantes-Metz, M. Kraten, A.D. Metz and G.S. Seow. "LIBOR manipulation?" *J. Banking & Finance* 36, 136-150, 2012.

R. Abrantes-Metz, S. B. Villas-Boas and G. G. Judge. "Tracking the LIBOR Rate," *CUDARE Working Papers*, 2013.

F. Benford. "The Law of Anomalous Numbers," Proc. *Amer. Phil. Soc.*78, 551-72, 1938.

M. H. DeGroot. *Probability and Statistics, 2nd Edition*, Addison-Wesley, 1989.

R. M. Fewster. "A Simple Explanation of Benford's Law," *Amer. Stat.* 63(1), 26-32, 2009.

A. K. Formann. "The Newcomb-Benford Law in its Relation to Some Common Distributions," *PLoS One*.5 (5), May 2010.

T. P. Hill. "A Statistical Derivation of the Significant-Digit Law," *Stat. Sci.* 10 (4), 354-63, 1995.

M. J. Nigrini. "I've Got Your Number," *J. Accountancy*, May 1999.

M. J. Nigrini. "An assessment of the change in the incidence of earnings management around the Enron-Andersen episode," *Review Accounting Finance* 4, 92-110, 2005.

J. M. Pimbley. "Benford's Law as a Logarithmic Transformation," *Maxwell Consulting Archives*, 2014.

R. Raimi. "The First Digit Problem," *Amer. Math. Monthly* 83(7), 521-38, 1976.

S. W. Smith. "Explaining Benford's Law," Chapter 34 in T*he Scientist and Engineer's Guide to Digital Signal Processing*, available at http://www.dspguide.com/.

FOOTNOTES

1. See Hill (1995). In this paper, Hill described his Theorem 3 as "help[ing] explain and predict the appearance of the logarithmic distribution in significant digits of tabulated data."

2. See, for example, Abrantes-Metz et al (2012).

3. See, for example, "China Data Suspected Says 75-Year-Old Theory: Cutting Research," *Bloomberg News*, January 10, 2013. Also, see "Don't Lie to Me, Argentina," *The Economist*, February 25, 2012.

4. It is reasonable to apply the Pearson Chi-Square Test to determine whether an empirical distribution of first digits matches the Benford's Law distribution. See, for example, Abrantes-Metz et al (2013). Alternatively, one could reason that a person or group perpetrating fraud would adjust or manufacture data with a nearly uniform distribution of first digits. In this case, it could be reasonable for the analyst to test the hypothesis that the first-digit distribution is uniform. Failure to reject the uniform distribution hypothesis would then be the fraud indicator.

5. As Newcomb (1881) demonstrated, it is little additional work to determine the distributions of the trailing digits as well as the leading digit.

6. Raimi (1976), Hill (1995) and Fewster (2009) provide good citations to others' empirical data collections. As Nigrini (1999) notes, there exist more than 150 academic papers on Benford's Law, published from the 1940's through the end of the twentieth century. An online record of more than 600 such papers exists at http://benfordonline.net/.

7. See the website http://www.fhfa.gov/.

8. Though "much closer to Benford," as we say, Chart 2 does not truly show a good fit. The Q statistic of the Pearson Chi-Square Test for the Chart 2 data is 71. Based on the chi-square distribution with eight degrees of freedom, one rejects the null hypothesis (agreement with Benford) with 99% confidence for this Q value, since it is greater than 20.09. See, for example, DeGroot(1989).

9. Formann (2010) provided numerical calculations for the U(0,1), positive exponential and log-normal distributions; he also provided some others we do not include.

10. For positive real values x, the expression f(x)dx with $0<x<\infty$ is the probability that x resides in the domain (x,x+dx). The PDF f(x) is non-negative and must satisfy the normalization $\int_0^\infty f(x)dx = 1$.

11. Pimbley (2014) applies this Benford test of inequality (5) to several more complex probability density functions, and finds agreement with all other known results.

12. It is possible to create a special case in which a dataset spans just one order of magnitude and satisfies Benford's Law exactly, but we exclude such contrived cases.

13. "China Data Suspected Says 75-Year-Old Theory: Cutting Research," *Bloomberg News*, January 10, 2013.

14. We downloaded GDP data from the World Bank database.

15. Of course, this finding is not a validation that the GDP data is not fraudulent. Benford's Law simply does not apply.

16. See Abrantes-Metz et al (2012) and Abrantes-Metz et al (2013).

17. We downloaded LIBOR and Treasury bill data from the FRED database.

18. We also created the analogues of Charts 4 and 5 for the LIBOR and T-Bill rates during the earlier and longer period 1986-2004. We find the same conclusions: both LIBOR and T-Bill values do not follow the Benford Distribution, and the ranges of the values of both data series are limited. We employed 3-month interest rate date rather than 1-month data due to the absence of 4-week T-Bill data for a portion of this long period in our data download.

19. For clarity, let us note that an unusual distribution of second digits may well be indicative of fraud. In the news report of the GDP study, the authors note that the sovereign may boost numerical values higher to reach the next first digit. (As cited earlier, see "China Data Suspected Says 75-Year-Old Theory: Cutting Research," *Bloomberg News*, January 10, 2013.) In that case, the second digit may be zero with higher frequency relative to un-adjusted numbers. Our point here is that the Benford Law for the second-digit distribution cannot be the logical comparison to detect this form of data adjustment.

20. This prescription for "several orders of magnitude" is admittedly vague. One might reason that a dataset with both positive and negative values will automatically satisfy the criterion, since the inclusion of zero within the range of the data strictly implies an infinity of orders of magnitude. In practice, the "infinity" will be limited by the precision of the data. The appendix discussion adds some clarification, but this prescription does remain vague nonetheless. Our recommendation is simply that the analyst should determine that the dataset spans at least one order of magnitude. With just one order of magnitude, though, there will not be a Benford distribution, unless the data essentially has a PDF ~x^(-1). As the appendix notes, the dataset will need the vague "several orders of magnitude" to achieve near-Benford Distribution.

21. We believe that Theorem 3 in Hill (1995) likely suffices to prove these statements for the special case of the base-ten logarithm transformation under Hill's stated assumptions and restrictions.