Testing the Mixing Property of the Newcomb-Benford Profile: Implications for the Audit Context

Edward J. Lusk[1] & Michael Halperin[2]

[1] The State University of New York (SUNY), Plattsburgh, NY, USA

[2] Director Lippincott Library, Wharton School, University of Pennsylvania, Philadelphia, PA, USA

Correspondence: Edward J. Lusk, The State University of New York (SUNY), Plattsburgh 101m Broad St. Plattsburgh, 12901, NY, USA. Tel: 518-564-4190 or 215-898-6803. E-mail: luskej@plattsburgh.edu or lusk@wharton.upenn.edu

Abstract

Introduction: Circa 1996 Theodore Hill offered a definitive proof that under certain conditions a data generating process is likely to produce observations that follow the Newcomb-Benford Log10 (N-B) first digit profile. The central feature of Hill's proof is the mixing property from which seems to follow base invariance for scale transformations. Further, it has been observed that small datasets are often not part of the N-B profile set. **Study Precise**: This suggests that, if indeed the mixing process underlies the generation of the N-B profile, that one should be able to take small Non-Conforming base-invariant datasets that are generated by uncorrupted processes and aggregate them to form datasets that conform to the N-B profile. **Results**: We demonstrate mixing convergence and find a systematic movement from Non-Conformity to Conformity at a transition point on the order of 250 data points. **Impact**: We suggest the practical importance of the Hill-Mixing result for the certification audit. We have all of these tests, datasets and results coded in a Decision Support System in VBA: Excel™ that is available from the authors free without restriction to its use.

Keywords: theodore hill, newcomb-benford conformity, CapitalCube, base-invariance

1. Introduction

The historically available information on the Digit Frequency Profile (DFP) starts in 1881 when Simon Newcomb (1835-1909), Professor of Mathematics, Astronomer awarded *The Gold Medal of London's Royal Astronomical Society* (1874), and extensively published Political Economist (See: *American National Biography Online*: Newcomb, Simon, 19th century) notices a curious pattern of *Wear & Tear* of his logarithmic tables—his Decision Support System of the 19th century. He offers (1881, p. 39):

That the ten digits do not occur with equal frequency must be evident to any one making much use of logarithmic tables, and noticing how much faster the first pages wear out than the last ones. The first significant figure is oftener I than any other digit, and the frequency diminishes up to 9.

Newcomb's "curiosity-note" gathers archive dust for some fifty years until Frank Benford (1938), an electrical engineer with *General Electric Inc*. USA with many patents to his credit, who curiously never directly cites Newcomb, makes and records the same observation, Benford (1938, p. 551):

It has been observed that the pages of a much used table of common logarithms show evidences of a selective use of the natural numbers. The pages containing the logarithms of the low numbers 1 and 2 are apt to be more stained and frayed by use than those of the higher numbers 8 and 9.

Newcomb and Benford both arrived at a simple mathematical formula to characterize the likely distribution of the nine first digits. To wit the (N-B Profile Log10):

$$Frequency[d_i] = LOG10\ (1 + 1/d_i)\ for\ i = 1, 2, ..., 9 \tag{1}$$

Following on the work of Nigrini (1996), this simple formula, if it is the underlying generating process for digital frequencies in the Big Data milieu, can be used to benchmark particular *Observed* digital frequency profiles for the purpose of generating *variance information* that can pique the interest of the auditor to the possible end of launching an extended procedures examination of the particular dataset under audit examination.

The important questions begged by the "non-intuitive" observation of Newcomb and Benford are:

Why should EQ(1) work as a general DFP-generating process estimator? And under what conditions can an auditor reasonably expect to use EQ(1) as a profiling-screen for selecting accounts for possible extended procedures investigation?

The first research to address the theoretic underpinning of the Log10 formula (EQ1), as a reasonable and appropriate surrogate for un-perturbed data generating processes only starts to appear some 25 years after Benford's paper. In the 1960s, various aspects of the theoretical context of the Newcomb-Benford curiosity are offered by: Pinkham (1961), Adhikari and Sarkar (1968), Duncan (1969), and Raimi (1969). However, Hill (1995 a, b, 1996) is usually credited with providing the conclusive theoretical support for the *Why* and *Conditional* questions posed above. *En bref,* Hill (1998) and Fewster (2009), show by convincing argumentation and illustration respectively that: *Where there are datasets formed from (i) many differentiable sources, or (ii) a kernel data-generating process with many reasonably "orthogonal" variations that data generated subject to these idiosyncratic conditions/constraints seems to follow the first digital pattern sketched out by the Log10 formula*. We shall term this as *Hill-Mixing*. Additionally, Hill (1995b) shows that *Mixing* generates datasets that have the base-invariance property. Finally, Nigrini and Mittermaier (1997) and Durtschi, Hillison and Pacini (2004) observe that small datasets are usually not found to *Conform* to the N-B first digit profile. These three results/observations: (i) Hill-Mixing, (ii) Base-Invariance, (iii) Inherent small sample size *Non-Conformity* form the contextual basis for our research, the objectives of which are to:

1) Investigate the *implication* of these three facets of the Newcomb-Benford Profile that: *There should be a systematic transition from Non-Conformity to Conformity as one moves from Small Base-Invariant datasets to their aggregation*. This will be validation testing of the Hill-Mixing logic.

2) Find, for our test data, the *Point of the Conformity transition*, and

3) Discuss the *implications* of this *Conformity* Transition point for the *Auditor executing a PCAOB certification audit*.

Consider now the following three elements of this research investigation that are needed to effect this Hill-Mixing testing:

1) An alternative to the Newcomb-Benford Profile: Log10 Profile that will facilitate profile-testing of dataset *Conformity*.

2) Datasets of financial information that are likely to be from un-perturbed or non-corrupted data generating processes. In this regard, we downloaded four datasets of firms the data of which was subjected to a PCAOB audit; therefore, this data that was accepted by the SEC as free from material reporting error.

3) Two additional *Non-Conformity & Conformity* testing protocols measures that have been reported in the literature. These will be used as a robustness check on the *Conformity* Transition Point where the Small *Non-Conforming* datasets transition to *Conforming* datasets.

2. An Alternative to the Newcomb-Benford Profile: Log10 Profile

Remarkable thought it is, we will need a protocol to make the decision if the first digit profile is likely to be generated from a process that conforms to the N-B profile, because neither Newcomb nor Benford suggested a testing protocol for deciding **if a particular dataset is *Conforming* or *Non-Conforming* in nature**. In this regard, recent research of Lusk and Halperin (2014a) has offered an alternative benchmark to the Log10 model which is (i) related to the Log10 model, and (ii) based in the empirical realizations collected by Benford (1938, Table 1, p. 553). Lusk and Halperin (2014a) report that the mean frequency profile reported by Benford should be refined. They offer:

"Benford reports a curious measure: In Table 1, p. 553 we find the Average (Arithmetic Mean) for the 20 point measures for the first digit set. To produce this measure, he took the simple average of the point frequency measures as reported in Table 1. Actually, this has no useful statistical meaning. The correct frequency average is the number of digits in a particular first digital Bin as a ratio to the sample total of 20,229. For example, in Table 1, p.553 Benford has the frequency proportion average for "1" as: 30.6%. However, the correct frequency percentage is the unit-vector product of the 20 frequencies with the respective sample sizes for the first digit. This develops for the 20,229 observations that 5,849.295 were "1s" and the proper ratio is 5,849.295/20,229 = 28.9%."

This suggests that for the Newcomb-Benford profile one could use the N-B Log10 generating function or could use the Lusk and Halperin (2014a) corrected means for the 20 samples that Benford reports. These two

alternatives are presented in Table 1 following:

Table 1. The Benford practical profile (BPP) and the N-B log10 profiles

First Digit Array	Corrected Means of Benford Datasets	N-B Log10	Difference (ColB less ColC)
Digit 1	0.28919	0.30103	-0.01184
Digit 2	0.19462	0.17609	0.018530
Digit 3	0.12665	0.12494	0.001710
Digit 4	0.09061	0.09691	-0.00630
Digit 5	0.07544	0.07918	-0.00374
Digit 6	0.06431	0.06695	-0.00264
Digit 7	0.05408	0.05799	-0.00391
Digit 8	0.05487	0.05115	0.00372
Digit 9	0.05052	0.04576	0.00476

Source: Lusk and Halperin (2014a).

These two benchmarks are effectively the same as a perusal of the differences reported in Col4 supports. However, the corrected means of the BPP are most relevant to forming a testing interval as they are not point realizations but rather empirical observations and so will have a reliable measure of variation. Lusk & Halperin (2014a) use this empirical variation to form a *screening interval* as presented in Table 2:

Table 2. The Benford practical profile (BPP) screening interval

First Digit Array	Corrected Means of Benford Datasets, n=20	Lower Benford Screening Window (BSW) Value	Upper Benford Screening Window (BSW) Value
Digit 1	0.289189	0.275377	0.303001
Digit 2	0.194622	0.179919	0.209324
Digit 3	0.126650	0.111340	0.141960
Digit 4	0.090612	0.074990	0.106235
Digit 5	0.075436	0.059684	0.091189
Digit 6	0.064314	0.048467	0.080161
Digit 7	0.054081	0.038147	0.070014
Digit 8	0.054872	0.038945	0.070798
Digit 9	0.050522	0.034558	0.066485

Source: Lusk and Halperin (2014a).

This screening interval can be used to determine if the realized first digit profile is in the screening interval (ColC to ColD). The simple test, programmed into the DSS: Hill-Mixing is:

If a particular digital frequency is in the closed Screening BPP-Interval (ColC to ColD) then that digital frequency is judged to be: *Conforming*; otherwise, i.e., the frequency is not in the Screening BPP-Interval, then that frequency would be labeled as: *Non-Conforming*. Lusk and Halperin (2014a) find that if 2/3 of the frequencies or more, that is 6, 7, 8 or 9 specific digits are outside the Screening BPP-Interval that the dataset is likely to be *Non-Conforming*; otherwise, if 1, 2, 3, 4, 5 are outside the Screening Interval then the dataset is likely to be *Conforming*.

3. The Datasets Used to Test the Mixing Transition

As introduced above, in testing the Hill-Mixing result, we have selected commercial data as reported in the financial statements of PCAOB audited firms where the two opinions on their financial statements were judged by the SEC as indicating that the financials were reasonable reflections of the results of operations; this being the case the firms were "listed" on their trading exchanges. "Listing" is a critical accrual criterion as this suggests that there was no evidence that the data generating processes of the firm were inappropriately modified, corrupted, or constrained so as not to be representative of generating processes that would be expected to produce data that would conform to the Newcomb-Benford profile. Here the research of Ley (1996); Nigrini and Mittermaier (1997); Durtschi, Hillison and Pacini (2004); Reddy and Sabastin (2012) and Lusk and Halperin (2014b) taken together suggest that *Corrupted data generating processes often do not produce data that follows*

the "natural" Newcomb-Benford profile. Another reason for selecting commercial data as reported in the firm's financial statements is that most of the variables should be highly correlated which is a feature of base-invariance. Therefore, base-invariance will be a second accrual condition, to wit: that there is strong factor association among the selected variables.

Specifically, we selected the two dimensions that are most often used in classifying organizations. The first has to do with the nature of the GAAP-USA Lens elected by the firm to record the results of operation over the accounting period. This impact-aspect of these elections is usually labeled: *Discretionary Accruals* or simply *Accruals*, made by management. These accruals are the way that management can create their conception of the organization as report in the financials using the variety of rules in the GAAP-Lens. For example, revenue recognition rules are surprisingly varied and management can opt to follow an "Aggressive" recognition recording perspective relative to recording/recognizing revenue in the short-term resulting in a relatively large revenue influx; alternatively management could adopt a relatively "Conservative" view of revenue recognition that would give a very different revenue profile for the organization. These discretionary GAAP-recording issues are usually argued in the MD&A section of the 10-K, (*See EDGAR; Section 7 & 7a of the filed 10-K*) and so have, usually, the approval of the certification auditors as well as the SEC. See the following early works where Discretionary Accruals are carefully discussed and illustrated: Dechow, Sloan, and Sweeney (1995); Frankel, Johnson, and Nelson (2002); Gul, Chen, and Tsui (2003); Hodgson and Praag (2006); Doyle, Ge, and McVay (2007); and Dechow, Hutton, Kim, and Sloan (2012).

The specific set of definitions used in the category triage are offered by the *CaptialCube Inc.* as follows (Note 1):

First Category: ***Conservative*** or ***Aggressive*** application of Generally Acceptable Accounting Principles (GAAP)*:

Conservative is defined as: "*Company's net income margin and percentage of accruals are both higher than peer median. Usually indicative of a company with "understated" income because of a conservative accounting policy.*"

Aggressive is defined as: "*Company's net income margin is higher than peer median while the percentage of accruals is lower than peer median. Usually indicative of a company with an aggressive accounting policy.*"

Second Category: ***High*** or ***Low*** Accounting Fundamentals.*

Fundamentals is defined as: The Fundamental Analysis score is calculated by comparing the company's performance relative to peer companies across multiple attributes like relative valuation, valuation drivers, operations diagnostic, etc. The Fundamental Analysis score is computed daily, and incorporates the latest company and peer data as of the previous day.

For the classification partition, we selected, on 26 Oct 2014, the *CapitalCube* reported data and cross-classified Fundamental with Aggressive and Conservative Accounting. Specifically, we then took the Highest and the Lowest Fundamentals score for each of the GAAP categories. This resulted in the following accruals: Low Fundamentals and Conservative Accounting (LFCA), n=43; High Fundamentals and Conservative Accounting, (HFCA) n=49; Low Fundamentals and Aggressive Accounting, (LFAA) n=53 and High Fundamentals and Aggressive Accounting, (HFAA) n=44. These four exclusive datasets will each be investigated for their *Mixing-Transition* profiles. For these four datasets, we selected at least one *Income Statement* variable and at least one *Balance Sheet* variable and then randomly added between two and five additional variables including also *Cash Flow from Operations*. This then generated the following four test sets that are presented in Table 3.

Table 3. The four accrual datasets and the measurement variables (in millions)

	Current Assets	Total Assets	Current Liabilities	Total Liabilities	**Gross Profit/Loss**	**COGS**	*Cash Flow Operations*
LFCA, n=43; EV:3.5(89%)		49.1		40.6	3.5		1.8
HFCA, n=49; EV:5.7(95%)	3.5	7.8	1.6	3.9	2.6		1.2
LFAA, n=53; EV:3.6(89%)		21.6		16.4	1.2		0.5
HFAA, n=44; EV:4.1(82%)	0.7	1.6		0.8	0.6	0.07	

Source: Accrual Data.

We then computed the usual Harman (1976) eigenvalues (EV) that pertain to the un-rotated factor matrix based

upon Pearson Correlation Coefficients. This is the test for a Base-Invariance effect. The values that are reported in the cells are the Mean values of these variables. As one can observe these four accrual sets are certainly reasonable characterizations of "typical" firms in the trading milieu. For example, these groups are "not in a Stressed Equity Position" as Total Assets on average are greater than Total Liabilities. Finally, for these variables over the four groups, the first eigenvalues are reported and the percentage of variance therein accounted for is reported. The second eigenvalues for all four groups had as their Range: (0.20 to 0.53) clearly arguing for base-invariance. For example, for the Low Fundamentals & Conservative Accounting partition (LFCA) we randomly selected: Total Assets; Total Liabilities; Gross Profit/Loss; and Cash Flow from Operations. The lead eigenvalue was 3.5 and accounted for 89% of the total variance. Given these datasets consider now the testing our expectation of the Hill-Mixing process.

4. The Hill Mixing Process: Testing the Aggregation of Base Invariant Datasets

Using the results of Nigrini and Mittermaier (1997) and Durtschi, Hillison, and Pacini (2004) that small datasets are likely to be *Non-Conforming* and, on the other side of the partition offered by Durtschi, Hillison and Pacini (2004) that, under the usual conditions, large dataset are likely to be *Conforming*, we first tested each of the individual variables of the four datasets for *Conformity*. Then we systematically but randomly aggregated the various individual datasets and then re-tested this aggregation for *Conformity*; we continued these aggregations until there was one final aggregated dataset that was tested for *Conformity*. Using the Bayes-conditional expectation, derived from the initial testing of the individual datasets, we then offer an inferential context for these aggregation results. Finally, given the inferential evidence that there was support for the Hill-Mixing effect, we conducted a robustness calibration of the final results. All of this information is reported following in Table 4.

Table 4. The aggregation and staged results of the accrued datasets

Datasets	LFCA, n= 43	HFCA, n = 49	LFAA, n = 53	HFAA, n = 44
NonConform%	2/4: 50%	6/6; 100%	3/4; 75%	4/5; 80%
First Digits	N= 171	N= 287	N= 211	N = 215
1	0.309942	0.331010	0.312796	0.306977
2	0.187135	0.184669	0.170616	0.162791
3	0.128655	0.132404	0.137441	0.125581
4	0.122807	0.108014	0.099526	0.111628
5	0.052632	0.045296	0.080569	0.079070
6	0.058480	0.062718	0.075829	0.046512
7	0.040936	0.055749	0.033175	0.055814
8	0.046784	0.045296	0.028436	0.055814
9	0.052632	0.034843	0.061611	0.055814
Results	**C:[3]**	**C:[3]**	**C:[4]**	**C:[4]**

Source: Summary of Accrual Data.

There are a variety of interesting results that may be gleaned from Table 4. We tested the conformity of the individual datasets for the four accrual groups. The *Non-Conformity* percentage of these tests, is noted as NonConform%. For example, for the Low Fundamentals and Conservative Accounting (LFCA) of the four datasets each of which had 43 observations at download (Total Assets, Total Liabilities, Net Profit/Loss & Cash Flow: Operations) two variable sets of the four or 50% tested to be *Non-Conforming* according to the BPP screening test and so is noted as: NonConform% 2/4: 50%. Along the last row is the same screening information for the First Digit Profile for the aggregation of all the data from the various variables in a particular data group. The number in () is the number of digits that were outside the BPP interval as presented in Table 2. For example, the LFCA group had a final aggregated sample size of N = 171 values greater than zero and for that aggregated dataset the First Digit Profile as noted in Col2 had (3) digits that were outside the BPP screening interval. *See the Appendix for the details of this computation*. Therefore, the LFCA was noted as a *Conforming* Dataset as: C(3). The principal result is that all four datasets after complete aggregation tested to be *Conforming*. The inferential test of this result is the simple Bernoulli test of the directional disposition in the binary space: C or NC. To fix the benchmark which is, of course, the Bayes conditional estimate, we took the dataset mix as downloaded from which the aggregate was formed. In this case we use the weighted average NonConform% over the four groups to determine the *incidence* of *Conformity* at download. In this case, the weighted average Conform% was: 23.04% (1-76.96%); this produced a p-value for finding four aggregate datasets that were Conforming in nature for the

usual Null of: p = 0.0028 to four places. This is certainly convincing evidence that:

Mixing is associated with forming aggregate datasets that move from, in large part, Non-Conforming (in weighted average of 76.96%) to Conforming in nature.

Another important result of this testing is the likely False Positive Investigation Error (FPIE) signal relative to electing extended procedures testing. Recall that the sample sizes at download were in the range (43 to 53) and had overall a *Non-Conformity* profile of 77%. However, as these datasets were downloaded from firms that were listed on trading exchanges their data was tested by the Audit LLPs and also scrutinized by the SEC. In this case, these datasets as reported are NOT likely to have warranted extended procedures testing. Therefore, if the auditor used the BPP to screen these datasets as downloaded, then the Auditor would make a FPIE, investigating when it is not likely to be warranted, about 77% of the time. This is of course due to the "small-sample size" tendency to *Non-Conformity*. Looking at the results of the aggregation for these datasets, there is no indication that extended procedures are in fact warranted–in fact, all of the four datasets tested to be *Conforming* and so the FPIE is likely to have been very low for the aggregation profile. *Implication*: *The sample size Range for these four aggregation-datasets was (171 to 287). And so as one moves away from low sample sizes of around 50 to sample sizes around 250, one can expect the Hill-Mixing result to correctly produce a Conformity result thus reducing the FPIE dramatically.*

To enrich this result, we examined the robustness of this important validation of Hill's Mixing proof as conditioned on base-invariance; we examined the *Conformity* of the four datasets in Table 3 using two other *Conformity* testing models and also for the Log10 benchmark where applicable. Consider now these results.

5. Robustness Testing of the Principal BPP Result

To effect this robustness testing, we used the following two models to make the decision if the observed First Digital profile (O_i) is inferentially different from the BPP (BPP_i) or in some cases the Log10 (Log_i) profile:

Model A: The χ^2 model where the test is: if $\{(O_i)$ and $[BPP_i]\}$ or $\{(O_i)$ and $(Log_i)\}$ are different in FD-profile using the standard marginal expectation projections over all 18 digits. In this regard, Lusk & Halperin (2014b) argue that if the overall χ^2 computed is > 15.507 which is the 95% inferential cut-off, that the dataset is *Non-Conforming* in nature. Also in this regard, the χ^2 sample size anomaly does not come into play as the sample sizes were projected using the upper limit suggested by Lusk & Halperin (2104b) of 440.

Model B: This model uses the Nigrini (1996) z-test results and tests at what sample size does the 6th z-value move past 1.96. Lusk and Halperin (2014c) find that if this Critical Sample Size (CSS) is greater than their suggested triage point of 1,825, then the dataset is likely to be *Conforming*; and, if the CSS ≤ 1,825 then the dataset is likely to be *Non-Conforming*.

Using these two models the robustness results are presented in Table 5.

Table 5. Robustness testing of the Hill-Mixing result

Datasets	Low Conservative (LC)	High Conservative (HC)	Low Aggressive (LA)	High Aggressive (HA)
Overall χ^2 BPP	0.008798	0.018972	0.016098	0.012581
Overall χ^2 Log10	0.006542	0.017087	0.013684	0.013108
z-Test BPP	20,600	1,850	6,100	7,100
z-Test Log10	23,100	6,100	7,850	4,850

Source: Accrual Data.

The Overall χ^2 values for the BPP or the Log10 profile relative to the (O_i) profile is remarkably low considering that the partitioning value is: 15.507. In this case, the χ^2 test results argue strongly for *Conformity* to the benchmarks. Further, the Critical Sample Size for the Nigrini z-test equations is, in all cases, greater than the triage point of 1,825 indicating that the datasets are likely to be *Conforming* in nature. These tests confirm the low FPIE result presented for the BPP testing. The last aspect that we wish to present is the implication of the validation of the Hill-Mixing result for the auditor.

6. Implications for the Auditor of the Hill-Mixing Results

The important implication of the empirical validation of the Hill-Mixing results is the need to condition the expectation of the auditor. Recall, the purpose of Digital Frequency testing is to identify candidates for possible discovery sample testing–it is, essentially, a screening protocol to ferret out, in an effective and efficient way,

those accounts that may warrant extended procedures testing in the audit. For example if *Non-Conformity* is detected this may be interpreted as indicating that extended procedures are warranted in the execution of the certification audit. However, the results reported above argue that first the auditor must consider the ***size of the datasets***, the Digital Frequency Profile (DFP) of which, has been developed. Small datasets, on the order of 50 observations, do not seem to provide sufficient observations to "fill in the digital" Bins so that their DFP can move towards the N-B or the BPP profile. In this case, the auditor invites, to be sure, the FPIE of selecting an account based upon its *Non-Conformity* profile where such *Non-Conformity* is driven by the small sample size rather than by a corrupted data generating process. With the results reported above the auditor can avoid this small-sample FPIE-anomaly by aggregating base-invariant, that is correlated, datasets to derive a sample of on the order of 250 observations, and then test such aggregations for *Non-Conformity*. This, we argue, will better approximate the FPIE consistent with the risk level of the audit as determined at the Analytic Procedures stages of the audit.

7. Summary, Conclusion and Outlook

7.1 Summary

The testing of the Hill-Mixing property has, for the datasets that we have collected, validated Hill's elegant proof (Note 2) of the Newcomb-Benford curiosity that leads to the first digit frequency profile popularized in the audit context by Nigrini (1996). We find that the observation of Nigrini and Mittermaier (1997) and Durtschi, Hillison, and Pacini (2004) that there is a small-dataset anomaly that by extension invites the FPIE in the audit context seems a valid concern. Finally, our results are conditioned on the fact that we, following Hill's results, used the base-invariance property to aggregate the small sample account dataset. Finally, we examined the robustness of these validation results by examining the DFP-nature of the finally aggregated datasets. In this regard, we used the DFT protocols that use the Overall χ^2 values for the BPP or the Log10 profile and the z-test calibration offered by Nigrini (1996).

7.2 Conclusion

The important recommendation that one may glean from these results is that aggregation of small correlated datasets of audit account variables, on the order of 50 observations, to form a single aggregate of at least 250 observations or so will move in the direction of avoiding testing based upon *Non-Conformity* where such an indication seems to result from the small sample size and so invites the FPIE.

7.3 Outlook

These results need to be replicated extensively to refine the FPIE small-sample size anomaly which is now initially set at 250 observations. In this regard, if auditors would contribute small un-corrupted datasets, on the order of 50 observations, drawn from the audit context, we would post these datasets on our *Commons* web-link for other researchers in this exciting Digital Frequency Profiling milieu. Additionally, one would extend this testing to non-base invariant aggregations. Finally, testing of the False Negative Investigation Error (FNIE), failing to investigate when it is likely needed, could enhance the DFP information for the audit context. Collecting such FNIE information will be very challenging as it assumes that there are corrupted data generating processes and that such data is available. In fact, in the usual audit context, such data is used to effect corrective actions and so the generating process is corrected and often the data from the corrupted process is lost. Possibly the only way to generate such data is to take *Non-Conforming* DFP reported in the literature and using these profiles run simulations of various sizes to examine the FNIE effect (Note 3).

References

Adhikari, A. K., & Sarkar, B. P. (1968). Distribution of most significant digit in certain functions whose arguments are random variables. *Sankhya-The Indian Journal of Statistics Series B, 30*, 47-58.

Benford, F. (1938). The law of anomalous numbers. *Proceedings of the American Philosophical Society, 78*, 551-572.

Dechow, P., Sloan, R., & Sweeney, A. (1995). Detecting earnings management. *The Accounting Review, 70*, 193-225.

Dechow, P., Hutton, A., Kim, J., & Sloan, R. (2012). Detecting earnings management: A new approach. *The Journal of Accounting Research, 70*, 275-334. http://dx.doi.org/10.1111/j.1475-679x.2012.00449.x

Doyle, J. T., Ge, W., & McVay, S. (2007). Accruals quality and internal control over financial reporting. *The Accounting Review, 82*, 1141-1170. http://dx.doi.org/10.2308/accr.2007.82.5.1141

Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's Law to assist in detecting fraud in accounting data. *Journal of Forensic Accounting, 5*, 17-34.

Duncan, R. (1969). Note on the initial digit. *Fibonacci Quarterly, 7*, 474-475.

Fewster, R. M. (2009). A simple explanation of Benford's Law. *American Statistician, 63*, 26-32. http://dx.doi.org/10.1198/tast.2009.0005

Frankel, R. M., Johnson, M. F., & Nelson, K. K. (2002). The relation between auditors' fees for nonaudit services and earnings management. *The Accounting Review, 77*, 71-105. http://dx.doi.org/10.2308/accr.2002.77.s-1.71

Gul, F. A., Chen, C. J., & Tsui, J. (2003). Discretionary accounting accruals, manages' incentives and audit fees. *Contemporary Accounting Research, 20*, 441-465. http://dx.doi.org/10.1506/686e-nf2j-73x6-g540

Harman, H. (1976). *Modern Factor Analysis*. Chicago, IL. USA: University of Chicago Press.

Hill, T. (1995a). The significant-digit phenomenon. *American Mathematical Monthly, 102*, 322-327. http://dx.doi.org/10.2307/2974952

Hill, T. (1995b). Base-invariance implies Benford's law. *Proceedings of the American Mathematical Society, 123*, 887-895. http://dx.doi.org/10.1090/S0002-9939-1995-1233974-8

Hill, T. (1996). A statistical derivation of the significant-digit law. *Statistical Science, 10*, 354-363.

Hill, T. (1998). The first digit phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data. *American Scientist, 86*, 358-363. http://dx.doi.org/10.1511/1998.4.358T. P.

Hodgson, A., & Praag, B. V. (2006). Information trading by corporate insiders based on accounting accruals: Forecasting economic performance. *Accounting and Finance, 46*, 819-842. http://dx.doi.org/10.1111/j.1467-629x.2006.00206.x

Ley, E. (1996). On the peculiar distribution of the U.S. stock indexes' digits. *American Statistician, 50*, 311-313. http://dx.doi.org/10.1080/00031305.1996.10473558#.Uq9HSdJDvTk

Lusk, E., & Halperin, M. (2014a). Using the Benford datasets and the Reddy & Sebastin results to form an audit alert screening heuristic: A Note. *IUP Journal of Accounting Research and Audit Practices, 8*, 56-69.

Lusk, E., & Halperin, M. (2014b). Detecting Newcomb-Benford digital frequency anomalies in the audit context: Suggested χ^2 Test Possibilities. *Journal of Accounting and Finance Research, 3*, 191-205. http://dx.doi.org/10.5430/afr.v3n2p191

Lusk, E., & Halperin, M. (2014c). Test of proportions screening for the Newcomb-Benford screen in the audit context: A likelihood triaging protocol. *Journal of Accounting and Finance Research, 4*, 166-180. http://dx.doi.org/10.5430/afr.v3n4p166

Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. *American Journal of Mathematics, 4*, 39-40. http://dx.doi.org/10.2307/2369148

Nigrini, M. (1996). A taxpayer compliance application of Benford's law. *Journal of American Taxation Association, 18*, 72-91.

Nigrini, M., & Mittermaier, L. (1997). The Use of Benford's Law as an aid in analytical procedures. *Auditing: A Journal of Practice & Theory, 16*, 52-67.

Pinkham, R. (1961). On the distribution of first significant digits. *Annals of Mathematical Statistics, 32*, 1223-1230. http://dx.doi.org/10.1214/aoms/1177704862

Raimi, R. (1969). The peculiar distribution of first digits. *Scientific American, 221*, 109-120. http://dx.doi.org/10.1038/scientificamerican1269-109

Reddy, Y. V., & Sebastin, A. (2012). Entropic analysis in financial forensics. *The IUP Journal of Accounting Research and Audit Practices, 11*, 42-57.

Ross, K. (2011). Benford's Law: A growth industry. *American Mathematical Monthly, 118*, 571-583.

http://dx.doi.org/10.4169/amer.math.monthly.118.07.571

Notes

Note 1. These CapitalCube definitions are found on: (http://www.capitalcube.com/blog/index.php/glossary/fundamental-analysis-score/) which we last accessed: 10 Feb 2015.

Note 2. The proof of the N–B observation as a general data generating process is challenging. In speaking of Hill's theoretical proof of the First Digit Phenomena, Ken Ross, former president of the Mathematical Association of America, (2011, p. 571) offers: "*The most successful seems to be Theodore Hill's very nice, but sophisticated, analysis in Hill (1996). I found his analysis challenging, and I am reasonably acquainted with probability*."

Note 3. In this regard, Lusk and Halperin (2014c) have offered a reasonably comprehensive list of 30 *Non-Conforming* datasets that are part of the Decision Support System that they used in developing the information for this research paper. This set of data could be a valuable source of simulation information.

Appendix A. Illustration of the Functioning of the Benford Practical Profile (BPP) Screening Interval

Digit	Lower Screening Boundary of the BPP	Gross Profit from Low Conservative Digital Profile	Upper Screening Boundary of the BPP
1	0.275377	0.34884	0.303001
2	0.179919	0.18605	0.209324
3	0.111340	0.06977	0.141960
4	0.074990	0.18605	0.106235
5	0.059684	0.04651	0.091189
6	0.048467	0.02326	0.080161
7	0.038147	0.02326	0.070014
8	0.038945	0.06977	0.070798
9	0.034558	0.04651	0.066485

Source: Accrual Data.

Above we illustrate how the BPP was used in evaluating the digital profiles relative to their *Conformity* or *Non-Conformity*. We selected the dataset from the Low Fundamentals & Conservative Accounting (LFCA) where the Gross Profit was reported. This was a small dataset n = 43 and its DFP is reported in Col3. The DSS: Hill-Mixing evaluates these values relative to the Lower (Col2) and Upper (Col4) screening boundaries of the BPP. The shaded cells of the Gross Profit profile are NOT in the BPP. As there are six not in the BPP this dataset is labeled as *Non-Conforming*. All of the various datasets were so analyzed using the DSS: Hill-Mixing.

Copyrights