Can Benford's Law be used to detect financial fraud?

Ben Hull
Level 4 Project, MSci Natural Sciences
Supervisor: Professor I. Hughes.
Department of Physics, Durham University
Submitted: August 6, 2023

Benford's law (BL) describes the probability of a given digit occurring at a position (index) in a number [1]. Typically the law quantifies the probability of the digits $1, 2, ..., 9$ to occur in the first index. However, the law can be extended to the second index or any combination of indexes in a number, such as the first-second digit Benford's law [2]. The literature suggests that applying Benford's law to the reported earnings of firms, along with other financial statement variables, is a viable technique for identifying financial fraud (FF) [3, 4]. This report aims to investigate the use of BL in detecting data manipulations in price paid data (PPD) for properties in England and Wales and a confirmed case of FF involving the American International Group (AIG). Various methods used to assess conformity are presented and evaluated for their suitability in detecting nonconformity with BL. We analyse several test statistics by introducing artificial deviations into synthetic Benford sets and observing the effects on our test statistics, concluding that different metrics are sensitive to different types of deviations. We find that PPD is highly nonconforming with BL and that Benford tests are useful in quantifying data manipulations in this context. However, in the case of AIG, we find that testing annual reporting data for conformity with BL has its limitations. In particular, we find that annual reports do not conform with BL, whilst subsets of financial data found in these reports do and thus should be the focus of our analysis.

Contents

1. Introduction

1.1. Previous Literature

Newcomb gave the first mathematical statement of Benford's law (BL) after the observation that the first pages of logarithmic tables wear out faster than the last ones [1]. Over half a century later, Benford manually analysed twenty datasets, including the values of physical constants and population data, for conformity with Newcomb's findings[1] [2].

Subsequent work has popularised the use of BL in detecting anomalies in financial datasets. In 1988 Carslaw [3] analysed the frequency of second digits occurring in reported earnings and losses for New Zealand based firms, concluding that there is a tendency to overstate earnings and understate losses. This study gave evidence of goal-oriented behaviour in financial statement reporting. These results were reproduced for firms operating in the United States and the United Kingdom by Thomas [4] and Van Caneghem [5] respectively.

Tilden and Janes [6] showed that corporate accounting data conforms with BL under stable economic conditions. Mehta [7] applies BL, along with other forensic accounting techniques, to the financially fraudulent statements (FFS) of Toshiba during the 2008-2015 period, concluding

[1] Note that Benford's original paper used no goodness of fit parameters to determine the strength of this conformity. Instead, Benford speculated on the 'Probable Error' on the data collected.

that BL was a useful technique for detecting irregularities in reported accounting data in this instance.

1.2. Benford's Law

Contrary to one's intuition, the digits one to nine do not appear uniformly in the first index in many naturally occurring datasets[2]. For example, the Fibonacci numbers are more likely to start with one than any other digit, with nine the least likely digit to appear. Indeed the Fibonacci numbers are said to follow BL [9] which quantifies this property.

The term index is used to refer to a position in a number. At each index there will be an integer $0, 1, 2, ..., 9$. Indexes begin at the first significant digit in a number. For the integer 6301, the first index D_1 references the first position with a digit of 6, the second index D_2 references the second position, with a digit of 3 and similarly for the third and fourth indexes. In this case we would stay that $D_1 = 6, D_2 = 3, ...$ or $(D_1, D_2, D_3, D_4) = (6, 3, 0, 1)$. Note that there is no well-defined value for D_5 as the number 6301 has a length of four; that is 6301 is a four-digit number and there is no fifth index. By assumption D_1 is non-zero.

Benford-compliant sets conform with a logarithmic distribution, with each successive digit appearing with a lower frequency in the first index. Newcomb derived the following formula describing the probability of the integer $d_1 \in \{1, 2, ..., 9\}$ to occur in the first index D_1,

$$P[D_1 = d_1] = \log\left(1 + \frac{1}{d_1}\right). \tag{1}$$

In greater generality, the non-trivial positive integer $n = \sum_{i=0}^{m} 10^i a_i$ appears at the beginning of a number $N = \sum_{i=0}^{M} 10^i b_i$, with probability,

$$P[n] = \log\left(1 + \frac{1}{n}\right), \tag{2}$$

where $a_i, b_i \in \{0, 1, ..., 9\}$ and $a_m, b_M \neq 0$ with the decimal expansion of n having at most the same number of terms as N, that is, $m \leq M$. In Appendix A we prove that this statement is well defined.

A useful case of Equation 2 is the first-second digit law which states that the probability of $d_1 \in \{1, 2, ..., 9\}$ occurring at the first index D_1 and $d_2 \in \{0, 1, ..., 9\}$ occurring at the second index D_2 is,

$$P[(D_1, D_2) = (d_1, d_2)] = \log\left(1 + \frac{1}{10d_1 + d_2}\right). \tag{3}$$

Note that $10d_1 + d_2$ is the decimal expansion of d_1 in the first index and d_2 in the second. Summing over all possible values of d_1 gives the probability of d_2 appearing in the second index D_2,

[2] For instance, the Fibonacci numbers, which do not have a uniform distribution of first digits, appear in the growth patterns of sunflowers, pinecones and other plants and flowers [8]. Indeed the Fibonacci numbers are an example of a Benford set (see Figure 1).

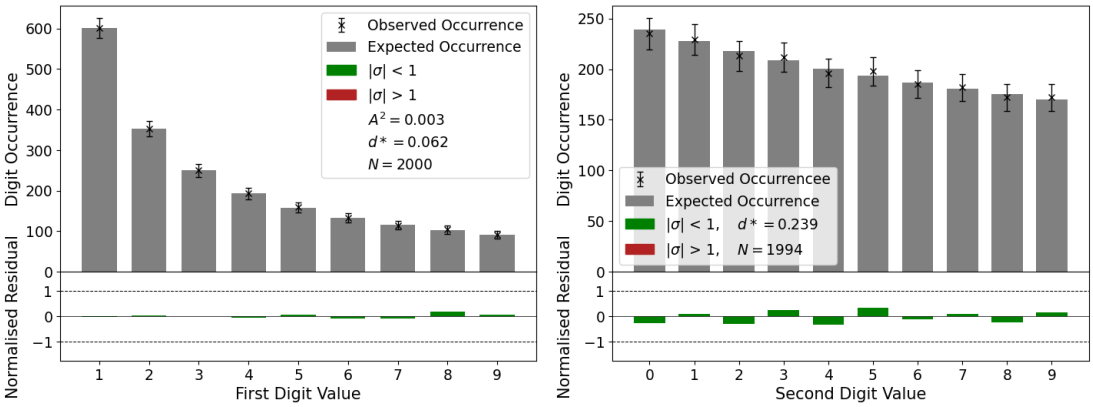


FIG. 1: An example of the application of Benford's first digit law (left) and the second digit law (right) to the first two thousand Fibonacci numbers. The observed occurrence (shown as error bars) is plotted against the expected occurrence due to Benford's law (shown as grey histograms bars). The error is calculated using the standard error on Poisson counts [10]. The normalised residuals show each digit's deviation from Benford's law, with green bars deviating by less than one standard error and red bars by more than one. Each figure has N data points; in this case, 2000. For the first digit test (left), one is the most common first digit with an expected occurrence of 30.1%. The expected occurrence of first digits decreases from one through to nine, which has an expected occurrence of 4.5%. Similar features hold for the second digit test (right); however, the variation of each digit's expected occurrence is far less pronounced. We see that the Fibonacci numbers comply with Benford's law in the first and second digit laws since all normalised residuals are within one standard error. We will discuss the metrics A^2 and d^* in Section 2.

$$P[D_2 = d_2] = \sum_{d_1=1}^{9} \log\left(1 + \frac{1}{10d_1 + d_2}\right). \tag{4}$$

Figure 1 shows an example of Benford's law applied to the first two thousand Fibonacci numbers. We can see that the occurrence of larger digits in the first and second indexes decreases, with this feature being most pronounced in the first digit test. Indeed this is a general feature of Benford's law with higher-order digit tests converging to a uniform distribution of $\{0, 1, ..., 9\}$; each of these integers will appear with near equal probability.

We see that all normalised residuals are within one standard error for both the first and second digit tests. Under the assumption that the normalised residuals should be normally distributed, we observe good conformity when 67% of normalised residuals are within one standard error. This is because we expect 67% of data points to be within one standard deviation of the mean for a normal distribution (here we take the mean to be zero and the standard deviation to be the standard error). Thus we expect no more than three normalised residuals to deviate by more than one standard error if the set we are considering conforms with BL. Therefore we conclude that the Fibonacci sequence conforms well with BL. The test statistics used in Figure 1 are discussed in Section 2 and will be used as another measure of conformity.

The error on the count of each digit is given by the standard error on Poisson counts. If a digit is observed in our dataset M times at some index, then the error in this digit count is $\sqrt{M}$

[10].

Whilst conformity with the Fibonacci sequence has been established, we still observe underlying Poisson noise in the data set, as evidenced by slight deviations from zero in the normalised residuals in Figure 1. Adding more terms to our dataset minimises the prominence of this Poisson noise. Determining when deviations from BL are due to underlying Poisson noise or whether they result from deliberate data manipulations is a central aim of this report.

1.3. Finite Range Benford's Law

If the range of values in our dataset is known, we can use a modified version of BL to analyse each digit's expected occurrence within this dynamical range. This has the advantage of considering the expected Benford distribution within the bounds of the dataset. However, we need to ensure that the data range is sufficiently large before applying these finite range techniques.

1.3.1. First Digit Finite Range Law

Sambridge, Tkalčić and Arroucau [11] formulated the first digit finite range version of Benford's Law (FRBL), that is, we consider datasets (or subsets of a dataset) in the range $U = [a\times 10^{\alpha}, b\times 10^{\beta}]$, $a, b \in [1, 10), \alpha, \beta \in \mathbb{N}$ with $\beta > \alpha$. This form of BL is useful when the upper and lower bounds of a dataset are known.

The probability of $x \in U$ having a first digit $d_1 \in \{1, 2, ..., 9\}$ is given by,

$$P[D_1 = d_1] = \frac{1}{\lambda_c}\left[(\beta - \alpha - 1)\log\left(1 + \frac{1}{d_1}\right) + \lambda_a + \lambda_b\right], \tag{5}$$

where,

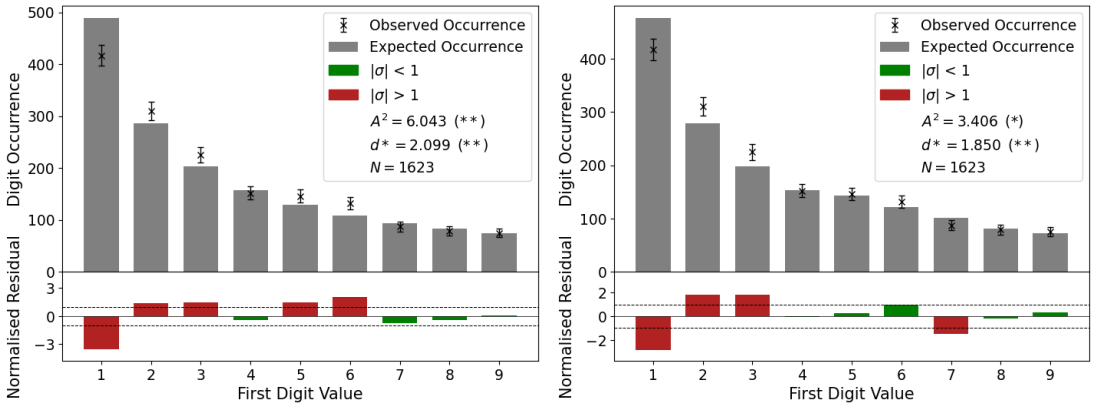
$$\lambda_c = (\beta - \alpha) + \log\left(\frac{b}{a}\right)$$

$$\lambda_a = \begin{cases} \log\left(1 + \frac{1}{d_1}\right) & : d_1 > a_1 \\ \log\left(\frac{1+d_1}{a}\right) & : d_1 = a_1 \\ 0 & : d_1 < a_1 \end{cases}$$

$$\lambda_b = \begin{cases} 0 & : d_1 > b_1 \\ \log\left(\frac{b}{d_1}\right) & : d_1 = b_1 \\ \log\left(1 + \frac{1}{d_1}\right) & : d_1 < b_1. \end{cases}$$

Here a_1 and b_1 are the first digits of a and b respectively. The terms λ_a and λ_b in Equation 5 correspond to boundary terms accounting for data in the ranges $[a \times 10^{\alpha}, 10^{\alpha+1}]$ and $[10^{\beta}, b \times 10^{\beta}]$, respectively. The first term in the bracket is BL applied to the range $[10^{\alpha+1}, 10^{\beta}]$. λ_c is a normalisation factor covering the entire finite range which ensures that all probabilities sum to one. A larger difference between α and β corresponds to a larger dynamical range which results in a better fit when applied to a Benford compliant set. Indeed, we recover Benford's law when $a, b = 1, 10$ or in the limit $\beta - \alpha \to \infty$.

Figure 2 shows an example of the first digit test and the first digit FRBL applied to consolidated financial data extracted from an annual report filed by the American International Group



(a) Benford's Law **(b)** Finite Range Benford Law

FIG. 2: The first digit test (subfigure (a)) and the first digit FRBL (subfigure (b)) applied to consolidated financial data extracted from an annual report filed by the American International Group in 2003. Note that for the first digit test, five of the nine normalised residuals are within one standard error and four for the FRBL. Notice that the data realises a poorer fit to the BL compared with the FRBL. We should be careful when applying the FRBL, ensuring that the dynamical range of the data is large enough to justify its use. The metrics d^* and A^2 will be discussed in Section 2.
(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

in 2003 [12]. We see that five residuals deviate by more than one standard error for the first digit test compared with four for the finite range law, which realises a better fit with BL. We should be careful when applying the FRBL ensuring that the dynamical range of the data is large enough.

1.3.2. Second Digit Finite Range Law

Similarly to Sambridge's FRBL, consider the range $U = [a \times 10^{\alpha}, b \times 10^{\beta}]$ with $\beta > \alpha$ with $a, b \in \mathbb{R}^{\geq 1}, \alpha, \beta \in \mathbb{N}$. Then the probability of $x \in U$ having a second digit $d_2 \in \{0, 1, ..., 9\}$ is given by,

$$P[D_2 = d_2] = \frac{1}{\lambda_c}\left[(\beta - \alpha - 1)\sum_{d_1=1}^{9}\log\left(1 + \frac{1}{10d_1 + d_2}\right) + \lambda_a + \lambda_b\right], \tag{6}$$

where,

$$\lambda_c = (\beta - \alpha) + \log\left(\frac{b}{a}\right),$$

$$\lambda_a = \begin{cases} \sum_{d_1=a_1}^{9}\log\left(1 + \frac{1}{10d_1+d_2}\right) & : d_2 > a_2, \\ \sum_{d_1=a_1+1}^{9}\log\left(1 + \frac{1}{10d_1+d_2}\right) & : d_2 < a_2, \\ \log\left(\frac{a_1+(d_2+1)10^{-1}}{a}\right) + \sum_{d_1=a_1+1}^{9}\log\left(1 + \frac{1}{10d_1+d_2}\right) & : d_2 = a_2. \end{cases}$$

$$\lambda_b = \begin{cases} \sum_{d_1=1}^{b_1-1}\log\left(1 + \frac{1}{10d_1+d_2}\right) & : d_2 > b_2, \\ \sum_{d_1=1}^{b_1}\log\left(1 + \frac{1}{10d_1+d_2}\right) & : d_2 < b_2, \\ \log\left(\frac{b}{b_1+10^{-1}d_2}\right) + \sum_{d_1=1}^{b_1-1}\log\left(1 + \frac{1}{10d_1+d_2}\right) & : d_2 = b_2. \end{cases}$$

Here a_1 and b_1 are the first digits of a and b respectively and a_2 and b_2 are the second digits of a and b respectively. This result is proven in Appendix B.

Similarly to the first digit FRBL, a greater dynamical range of data results in a better fit when applied to a Benford compliant set. We recover the second digit BL when $a, b = 1, 10$ or in the limit $\beta - \alpha \to \infty$.

2. Test Statistic Analysis

A well-defined method for assessing conformity with BL is required. Without such a formalism, the detection of anomalies in test sets is subjective and holds no underlying rigorous basis. This section defines two metrics for assessing conformity, the A^2 and d^* test statistics. We then apply each of these metrics to synthetically generated Benford compliant data with deviations introduced consistent with those expected in manipulated financial datasets. For example, we introduce rounding behaviour by pronouncing the lowest digit and suppressing the highest digit at a given index and observe the effects on the value of our test statistics. We also investigate the differences between the FRBL and the original BL by applying our test statistics over different finite ranges of data. We shall see that, in general, the application of the FRBL yields a better fit to the data on the basis of these test statistics over finite ranges.

Before we begin presenting and analysing these metrics, we first briefly discuss synthetic Benford set generation.

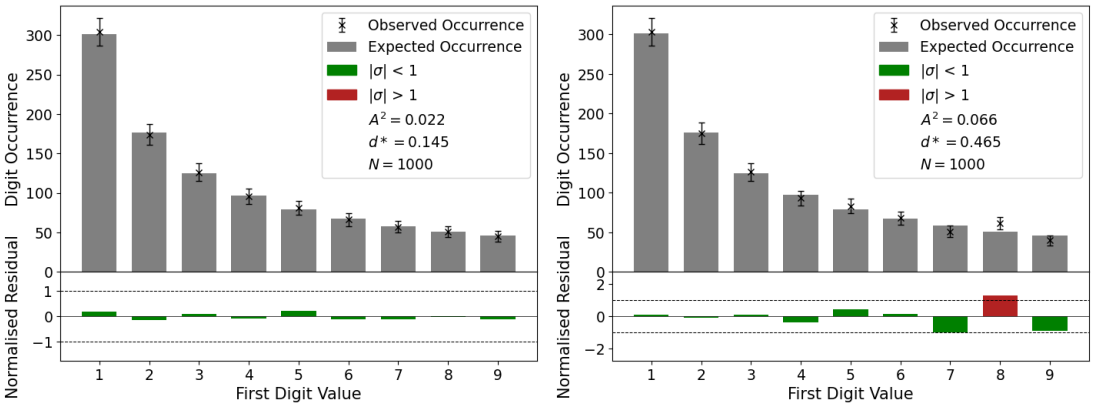
2.1. Synthetic Benford Sets

We define two techniques for generating Benford sets: geometric series and brute force methods. The former involves constructing a Benford set out of the terms in a geometric series, whereas the latter generates data using BL directly and relies on pseudo-random number generation. We shall see that geometric series produce precise Benford sets, whereas the degree of conformity of brute force based sets with BL varies significantly.

In our geometric series approach [13], we define $d := \log(b) - \log(a)$ for some $a, b \in \mathbb{N}$ with $b > a$ such that d is a positive integer. This condition ensures that the set is Benford[3]. If we define $r := 10^{\frac{d}{N}}$ with $N \in \mathbb{N}$ then the sequence $x_n = ar^n$ is Benford with $x_n \in [a, b]$. This defines a Benford set with N elements in the range $[a, b]$. A proof that this set is Benford is given in Appendix C. This gives us a precise fit with BL as evidenced by Figure 3.

Our brute force technique relies on the direct application of BL. We first calculate the cumulative probability distribution (CPD) of all possible four-digit sequences appearing at the beginning of a number using Equation 2. Note that we take first term in the CPD to be zero and the last term is one by definition. We select an element in the uniform distribution $\mathcal{U}(0, 1)$ and find the lowest upper and highest lower bound of this number in the CPD. The location of the lowest upper bound in the CPD allows us to randomly select four first digits for an element in our synthetic Benford set using BL. This number selection process and comparison with the CPD is repeated until we have the desired number of elements in our set. Note that random

[3] Indeed this condition on d is only required for the set to comply with BL; any value of $d \geq 1$ will generate a set compliant with the FRBL.



(a) Geometric **(b)** Brute force

FIG. 3: The first digit test applied to two synthetic Benford sets generated using (a) geometric series and (b) a brute force application of BL. Each set has 1000 elements in the range $[10^3, 10^7]$. Both figures conform well with Benford's law, as evidenced by the normalised residuals. In subfigure (a) all residuals are within one standard error, whilst in subfigure (b) all but one are within one standard error. In general, due to pseudo-random number generation, brute force based Benford sets do not conform with BL as well as geometric series based Benford sets. This is also shown by higher values of the test statistics d^* and A^2 in (b) compared with (a).

digits can be added to the end of each element to ensure that the synthetic Benford set is within a defined lower and upper bound such that the set is Benford (up to and including the fourth index).

Figure 3 shows an example of the first digit test applied to synthetic Benford sets generated using geometric and brute force techniques. Each method produces sets that conform with BL as shown by the normalised residuals, with the majority being within one standard error and the test statistics' low values. We note that geometric series produce a closer fit with BL compared with brute force sets. Geometric series utilises an analytic expression when generating Benford sets, which is proven to be Benford compliant. In particular, specifying a set of a given size with some upper and lower bound will always produce an identical Benford set under this algorithm. In contrast, the uniform number selection involved in the brute force method introduces deviations from BL into each set produced. This selection is also pseudo-random, which adds to the degree of this variation. This introduces significant variations in each set's conformity with BL based on consideration of the normalised residuals and test statistics.

Each type of Benford set has its specific uses and advantages. Geometric series gives an example of an analytic Benford set with an almost exact agreement with BL. We will use such sets as a basis to study our test statistics under several different manual deviations from BL that we expect in instances of FF. Brute force Benford sets give us insight into natural deviations from BL in the sets we analyse and will allow us to calculate p values for our test statistics.

2.2. d^* Analysis

The d test statistic is based on the Euclidean distance between observed and expected digit occurrences. For the first digit test the measure is defined as [14],

TABLE I: Calculated *p* values for d^* and A^2 for the first and second digit tests. α is the significance level; this is the probability of obtaining a result at least as extreme as the results actually observed, assuming the test set conforms with BL.

	First Digit Test		**Second Digit Test**	
Test Statistics	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
d^*	1.35	1.66	1.32	1.49
A^2	2.84	4.56	2.61	3.97

$$d := \sqrt{\sum_{i}^{9}(p_i - b_i)^2}, \tag{7}$$

where p_i is the observed proportion of digits in the first index having a value $i \in \{1, 2, ..., 9\}$ and b_i the corresponding expected probability according to BL. Morrow [15] introduces a similar definition for the d^* measure and computes the associated *p* values for the first digit test,

$$d^* = \sqrt{N\sum_{i=1}^{9}(p_i - b_i)^2}, \tag{8}$$

where N is the size of the set being analysed for conformity with BL. We can extend the definition of d^* to other digit tests as follows,

$$d^* := \sqrt{N\sum_{i}(p_i - b_i)^2}, \tag{9}$$

where p_i is the proportion of observations having $i = (d_n, ..., d_m, ..., d_2, d_1)$ as the $(D_n, ..., D_m, ..., D_2, D_1)$ indexes and b_i the corresponding expected proportion according to BL. The sum runs over all possible digit sequences in the indexes $(D_n, ..., D_m, ..., D_2, D_1)$.

This work uses Morrow's definition of the d^* statistic along with *p* values which are used to gauge conformity with BL. A *p* value is the probability of obtaining a result (in our case a value of our test statistics) at least as extreme as the results observed, assuming the null hypothesis is true [16]. In our case, the null hypothesis is that the set we are considering is Benford compliant.

Table I shows the *p* values for d^* when applying the first and second digit tests. These values were calculated by generating one thousand Benford sets in the range $[1, 10^6]$ each with ten thousand elements using brute force techniques. We then applied the first and second digit tests and computed the corresponding value of d^* for each set. Calculating the 95$^{\text{th}}$ and 99$^{\text{th}}$ percentiles for the first and second digit metric values gave us the 0.05 and 0.01 significance values, respectively. Notice that the second digit test has lower *p* values at both significance levels. This is to be expected as the distribution of second digits is closer to a uniform distribution than first digits. Since the lowest digit in the first index appears roughly 30% of the time compared with the lowest digit in the second index, which appears roughly 12% of the time, the absolute deviation from BL for the lowest digit in the first digit test is greater than in the second digit test. If the same proportion of digits deviate from BL in the lowest digit for both tests, this

will result in a higher absolute deviation of first digits compared with second digits. This higher absolute deviation contributes significantly to our test statistics resulting in higher *p* values for the first digit test.

One of the significant advantages of d^* compared to other metrics, such as Kolmogorov-Smirnov, is that d^* is a Benford specific formalism. Kolmogorov-Smirnov is a general statistic that assumes the null hypothesis of a continuous distribution however, in our case, we are working with discrete distributions. In particular, such a metric is conservative when testing discrete distributions; that is, the discrete case's *p* values are always less than, or equal to, the corresponding *p* values in the continuous case [17]. This means that the conformity of our test data with BL would be overestimated. In contrast, the d^* metric is designed for use with discrete data, and the *p* values have been explicitly calculated for use with BL. This makes the statistic extremely applicable to a Benford analysis.

Moreover, the *p* values associated with d^* apply to small sample sizes. Morrow calculates *p* values for the first digit test and compares values of d^* for various sample sizes constructed to deviate from BL at the 0.01 level and compares with the associated *p* values. Small sample sizes have simulated *p* values close to the expected *p* value, particularly for sample sizes with over eighty elements. In the limit that the sample size tends to infinity, the test statistic's asymptotics converges to the expected *p* value. Morrow's test statistic is valid for small and large sets; however, we should aim to analyse sets with over eighty elements when applying the d^* metric. A further discussion of *P* values is given in Appendix D.

2.3. A^2 Analysis

The Anderson-Darling or A^2 statistic is based on a comparison between the empirical probability distribution function of a set of data and the expected probability distribution function, which in this case given by BL [18].

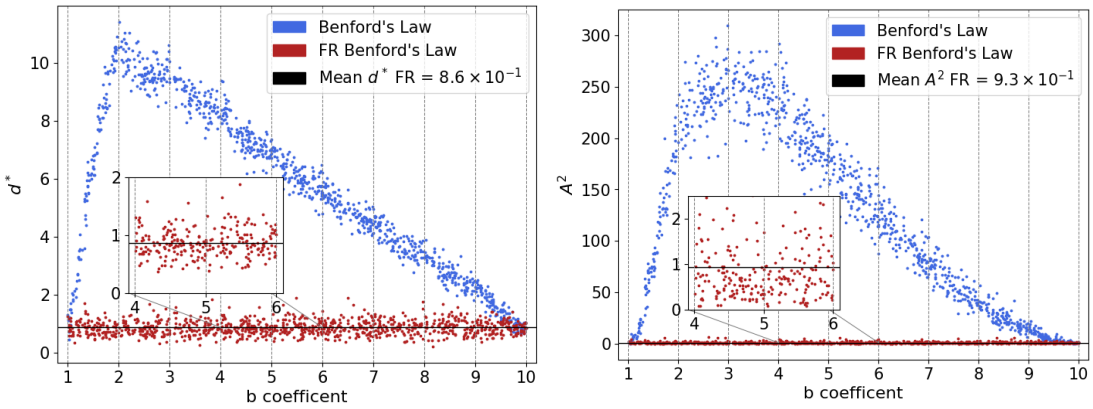
Let there be k cells in a discrete distribution labelled $1, 2, ..., k$ with the probability of a sampled data point belonging to cell j being p_j. In our case each cell j corresponds to an index (or combinations of indexes) in a number. Suppose N independent observations are made in an experiment. Let o_i and e_i be the number of observed and expected outcomes for cell i respectively. Define $S_j = \sum_{i=1}^{j} o_i$ and $T_j = \sum_{i=1}^{j} e_i$, then S_j/N and $H_j = T_j/N$ are the empirical and expected cumulative distribution functions respectively. Further take $Z_j = S_j - T_j$. Then the A^2 statistic for the discrete distributions is defined as,

$$A^2 = \frac{1}{N} \sum_{i=1}^{k-1} \frac{Z_j^2 p_j}{H_j(1 - H_j)}. \tag{10}$$

Note that we only sum over the first $k - 1$ terms rather than all k cells. This is because the last term in A^2 is of the form $0/0$, and is set equal to zero.

Table I shows the *p* values for A^2 for the first and second digit tests at various significance levels for a Benford distribution. These *p* values were calculated in the same way as for d^*. As with d^* the second digit test has lower *p* values.

A^2 is sensitive to deviations at the tails of distributions. Since we anticipate goal-oriented rounding behaviour, such as in the work of Carslaw [3], we require a test statistic with this property to detect such deviations. After roughly forty bins, A^2 will diverge and will no longer



(a) d^* **(b)** A^2

FIG. 4: The variation of the metrics d^* and A^2 when the first digit test is applied to brute force based Benford sets spanning different dynamical ranges, $[1, b \times 10^3]$, as a function of the coefficient $b \in [1, 10]$. Subfigures (a) and (b) show d^* and A^2 calculated for BL (shown in blue) and the FRBL (shown in red) in the first index over this finite range, respectively. Each plot has one thousand data points. We measure good conformity with BL at the endpoints $b = 1, 10$ with weaker conformity away from these values for both metrics. There are local maxima at $b = 2$ for d^* and $b \approx 3$ for A^2, the points with the poorest conformity with BL. All elements of the base sets in the range $[10^3, b \times 10^3]$ have one in the first index when $b < 2$. This introduces a disproportionate number of ones into the distribution of first digits, resulting in a poor fit with BL. When $b > 2$ other digits appear in the first index $(2, 3, 4, ..., 9)$ which improves conformity with BL. For d^* we observe a monotone increase between $b = 1$ and $b = 2$ followed by a monotone decrease between $b = 2$ and $b = 10$. This shows that the d^* metric is sensitive to the distribution's underlying shape rather than deviations in individual bins. Since A^2 is based on a cumulative distribution function, the shape of A^2 is less defined with maxima at around $b \approx 3$. Namely, we observe the greatest deviation from the expected cumulative distribution function at this value of b. Note that in both subfigures, we observe excellent conformity with the FRBL.

be a sensible statistic [18]. We must be careful not to apply this test statistic when dealing with distributions with many bins, such as in the first-second digit Benford test.

2.4. Finite Range Analysis

Figure 4 shows the variation of the d^* and A^2 statistics when the first digit test is applied to brute force based Benford sets in a finite dynamical range $[1, b \times 10^3]$ as a function of $b \in [1, 10]$. At all ranges, we observe excellent conformity with the FRBL for both metrics. This is expected as the FRBL applies to Benford sets spanning a finite dynamical range.

However, for BL (shown in blue), we observe an increase of our metrics, up to maxima at $b = 2$ for d^* and $b \approx 3$ for A^2, followed by a decrease up to $b = 10$. Increasing b between one and two introduces a disproportionate number of ones into the first index bringing the distribution away from BL. This coincides with an increase in the test statistics. Increasing b beyond two introduces higher digits into the first index, which brings the distribution closer to BL. This coincides with a decrease in the test statistics.

For d^*, the increase and decrease is monotone up to a defined maxima at $b = 2$. This

value of b is a turning point in the distribution of first digits; namely, as we increase b beyond this value, higher digits are introduced, bringing the distribution closer to BL. Since there is a corresponding turning point for d^*, this shows that d^* is sensitive to the overall shape of the distribution rather than individual deviations in each digit from BL. This metric can be utilised to measure conformity as a whole with BL but is unlikely to be sensitive to deviations of individual digits from BL.

In contrast, the A^2 statistic compares the empirical and expected cumulative probability distributions. This means that the statistic is sensitive to deviations that bring the empirical distribution function away from the expected distribution function. Hence, even at the turning point of the distribution of first digits, we do not observe a decrease in A^2. The large excess of ones in the first digit causes the empirical distribution function to deviate significantly from BL. Even though the shape of the distribution improves as b increases beyond two, as previously discussed, the empirical distribution function continues to deviate from expectation until $b \approx 3$. This shows that A^2 is sensitive to deviations in individual digits from BL rather than the overall shape of the distribution.

2.5. Variation of Statistics under Artificial Deviations

We model deviations in observed data from BL in an underlying synthetic Benford set, B (the base set). In doing so, we will analyse the effect such deviations have upon the d^* and A^2 test statistics. The base set used had one million elements in the range $[10^2, 10^6]$ and was generated using geometric series such that the base set is in excellent conformity with BL. These deviations are introduced in two ways, with each technique modelling features found in manipulated data.

The first technique pronounces a given digit, such as five, at a specified index in each element of the base set, at a percentage rate σ. For instance, we could pronounce the digit five at the second index at a rate of 25%, which is to say that 25% of the second digits in the base set are set to five.

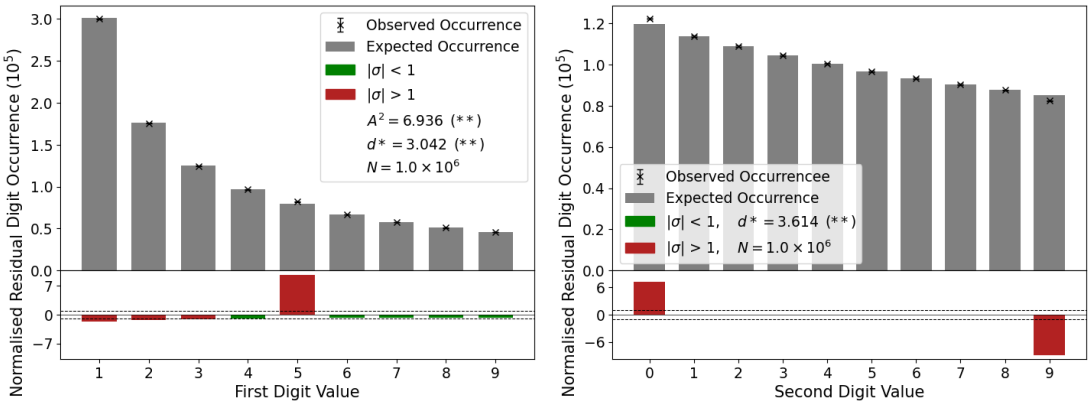
Our second method introduces rounding behaviour into the base set, more specifically upward rounding behaviour, at a specified index. A percentage σ of terms with nine in that index are rounded upward. This causes a deficit of nines in that index, favouring a surplus of zeros (or ones in the case of the first index).

Figure 5 shows an example of artificial deviations introduced in our base set, B. Subfigure (a) shows digit exaggeration in the first index and subfigure (b) shows rounding behaviour in the second index. Note that the values of σ responsible for the strength of the applied artificial deviations are not comparable between the two techniques.

2.5.1. Digit Exaggeration

To exaggerate a digit in our set B, pick an index D_i, a digit to pronounce at that index d_i and a rate at which to pronounce that digit $|\sigma| < 1$. For each element $x \in B$ select $y \in \mathcal{U}(0, 1)$. If $y \leq \sigma$, then set the value of the digit at D_i in x to d_i.

Figure 6 shows the variation of d^* and A^2 with respect to the parameter σ used to exaggerate a given digit in our Benford set. In this case, we have chosen to exaggerate the digit five in the



(a) Digit Exaggeration - First Index **(b)** Rounding Behaviour - Second Index

FIG. 5: Examples of artificial deviations introduced into a geometric Benford set with one million elements in the range $[10^2, 10^6]$. (a) shows the digit five exaggerated in the first index with a value of σ equal to 0.3. As expected, we observe an excess of fives in the distribution and a corresponding deficiency in the remaining digits. (b) shows the rounding behaviour introduced into the second index with a value of σ equal to 0.03. We see an excess of zeros and a deficiency of nines, as expected when elements in the Benford set have been rounded upward.

(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

first and second index. Each statistic is normalised with respect to their 0.05 p values given in Table I.

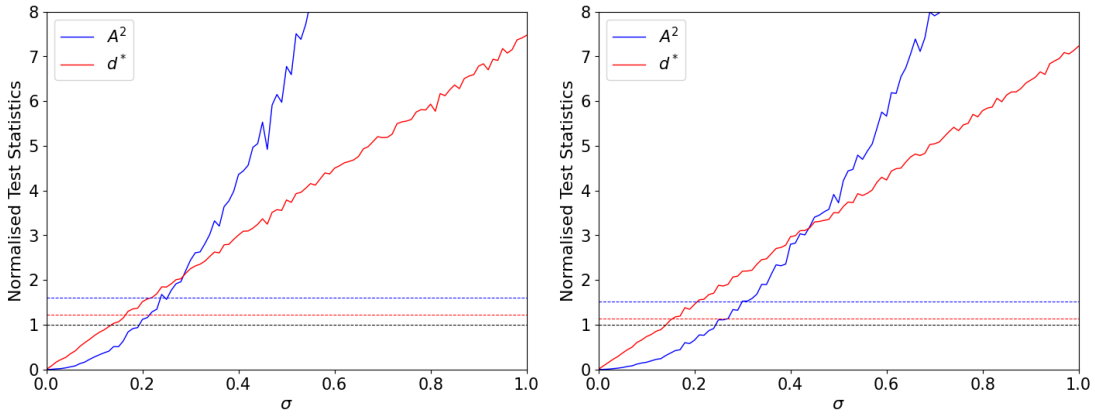
We firstly note that the metrics are self-consistent. That is, a local maxima/minima for d^* is a local maxima/minima for A^2 and vice versa. An increase in one metric is met with a corresponding increase in the other. This suggests that the metrics quantify conformity with BL consistently; however, each metric is sensitive to different deviations and digit tests.

d^* is a linear function of σ for both digit tests. Indeed we have a coefficient of determination [19], R^2, equal to 0.99 for the first index and 1.00 for the second index, indicating that d^* is linear in σ. This linearity suggests that d^* is not sensitive to significant deviations in individual digits compared with A^2. Indeed, as previously discussed, d^* describes the overall shape of the distribution.

At small values of σ and thus slight deviations in the digit five, d^* is greater than A^2. This suggests that slight deviations in individual digits will be identified by the d^* metric more frequently than A^2.

For A^2, we observe non-linear behaviour with an increasing gradient with increasing σ. Thus, A^2 will be effective in identifying significant deviations in individual digits. As discussed previously, for more minor deviations, d^* will be the more appropriate statistic.

We see similar behaviour in the d^* metric between the first and second digit tests. In particular, d^* is linear in each case reaching the 0.05 p value at a value of 0.17 for σ for both tests. Whereas, for A^2, we observe different features between each digit test. For the first digit test, we see a higher rate of increase in A^2 compared with the second digit test. This suggests that A^2 is more sensitive to deviations in individual digits in the first index than the second index and does not apply in the same manner between the two tests. d^* applies consistently between the two digit tests, which is an advantage of this metric.



(a) First Index **(b)** Second Index

FIG. 6: The dependence of the d^* and A^2 metrics, shown in red and blue respectively, on the parameter σ used to exaggerate a given digit in a Benford set. In this case, we have chosen to exaggerate the digit five. (a) shows the variation of the statistics when applying the first digit law and (b) the second digit law. The test statistics are normalised with respect to their 0.05 p values such that the 0.05 p value for each metric is equal to one. The dotted-lines show the corresponding value of the 0.01 p values for each test statistic. Across the first and second index, d^* is a linear function of σ. In contrast, A^2 is non-linear with an increasing gradient. d^* is of a similar form between the first and second index reaching the 0.01 p value at a value of 0.17 for σ in each case. A^2 is more sensitive to deviations in the digit five in the first index with a greater rate of increase compared with the second digit test.

2.5.2. Rounding Behaviour

To introduce rounding behaviour at an index D_i, pick a rate $|\sigma| < 1$ representing the prevalence of this rounding behaviour. For each element $x \in B$ with a nine in the index D_i choose $y \in \mathcal{U}(0, 1)$. If $y \leq \sigma$, then round up x such that x now has a zero (one, if we selected the first index) in the D_i index.

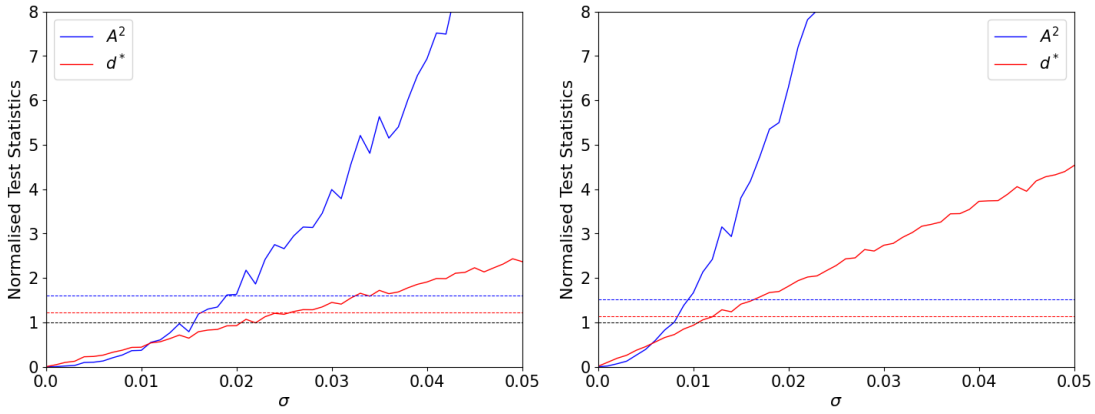
Figure 7 shows the variation of d^* and A^2 with respect to the parameter σ used to introduce rounding behaviour in the first and second indexes in our Benford set. Each statistic has been normalised with respect to their 0.05 p values as given in Table I.

As with digit exaggeration, the metrics are self-consistent with an increase/decrease in one metric met with an increase/decrease in the other. Again the metrics quantify conformity with BL consistently; however, each metric is sensitive to different features within a distribution of digits.

As before, d^* is roughly linear in σ for both digit tests. The coefficient of determination, R^2, is equal to 1.00 for both indexes, showing that d^* is linear in σ. We conclude that d^* is not sensitive to rounding compared with A^2 at large values of σ and that d^* is sensitive to the shape of our distribution.

Once again, A^2 is non-linear with an increasing gradient with increasing σ. Thus, A^2 will be effective in identifying significant rounding behaviour. At small values of $\sigma < 0.01$, that is a low rate of rounding, both metrics have similar values. This suggests that both metrics will have similar utility for low rates of rounding behaviour in detecting such deviations.

We see different behaviour for both metrics between digit tests. In particular, our metrics are



(a) First Index **(b)** Second Index

FIG. 7: The dependence of the d^* and A^2 metrics on the parameter σ used to introduce rounding behaviour into a Benford set. (a) shows the variation of the statistics when applying the first-digit law and (b) the second-digit law. The test statistics are normalised with respect to their 0.05 p values such that the 0.05 p value for each metric is equal to one. The dotted-lines show the corresponding value of the 0.01 p values for each test statistic. For both the first and second index, d^* is a linear function of σ whilst A^2 is non-linear with an increasing gradient. The rate of increase of both metrics is greater in the second index compared with the first, suggesting that we will have more success detecting rounding when utilising the second digit test. A^2 is far more sensitive to higher σ; that is, A^2 is far more sensitive to higher rates of rounding behaviour than d^*. However, for small σ, both statistics have similar normalised values suggesting a similar utility in detecting deviations for lower rates of rounding behaviour.

more sensitive to rounding behaviour in the second index. The gradient of d^* is greater in the second index, reaching the 0.05 p value at σ equal to 0.011 for the second digit test compared with 0.02 for the first digit test. A^2 had a greater increase in the gradient in the second index than the first, reaching the 0.05 p value at σ equal to 0.08 and 0.14, respectively. We will detect rounding behaviour in the second index far more readily using our test statistics compared with the first index.

3. Price Paid Data

In this section, we apply Benford's law to price paid data (PPD) for properties sold in England and Wales [20]. Such datasets form a solid foundation for a Benford analysis due to the psychological and monetary motivations when valuing properties. Indeed we shall see that PPD does not conform with BL. We quantify the nonconformity and investigate possible reasons why such a result is observed.

A notable factor that may influence house price values is stamp duty tax (SDT). SDT applies to property transactions in the United Kingdom and is payable to HM Revenues and Customs. The rates of SDT depend on the price of the property, i.e. the tax bracket of the property.

Of particular interest to us is the change in SDT in the year 2014. As of December 2014, tax is only payable on the portion of the property's value in each bracket. Previous to this, there was a set rate payable on total property price once a tax threshold was breached. Table II shows

TABLE II: Threshold values and SDT rates for residential properties valued under one million pounds sold in the United Kingdom [21]. Pre December 2014 the rate was payable on the total value of the property (the firm rate) whereas post December 2014 the rate was payable only on the portion of the value of the property in each tax bracket (the relaxed rate). Note that some of the tax brackets change between the two rates.

SDT Rate after Dec. 2014 (relaxed)		**SDT Rate after Dec. 2014 (firm)**	
Tax Bracket	**Tax Rate**	**Tax Bracket**	**Tax Rate**
£0−£125,000	0%	£0−£125,000	0%*
£125,001−£250,000	2%	£125,001−£250,000	2%*
£250,001−£925,000	5%	£250,001−£500,000	3%*
£925,001−£1.0m	10%	£500,000−£1.0m	4%*

* Payable on total property price once limit is breached.

a summary of these thresholds and the tax rates that apply. We will refer to the SDT rate after December 2014 as the relaxed rate and the SDT rate before December 2014 as the firm rate. In this investigation we focus on the tax threshold values at £125,001 and £250,001 since the majority of properties are priced around these values and each of these thresholds appear in both the firm and relaxed rates.

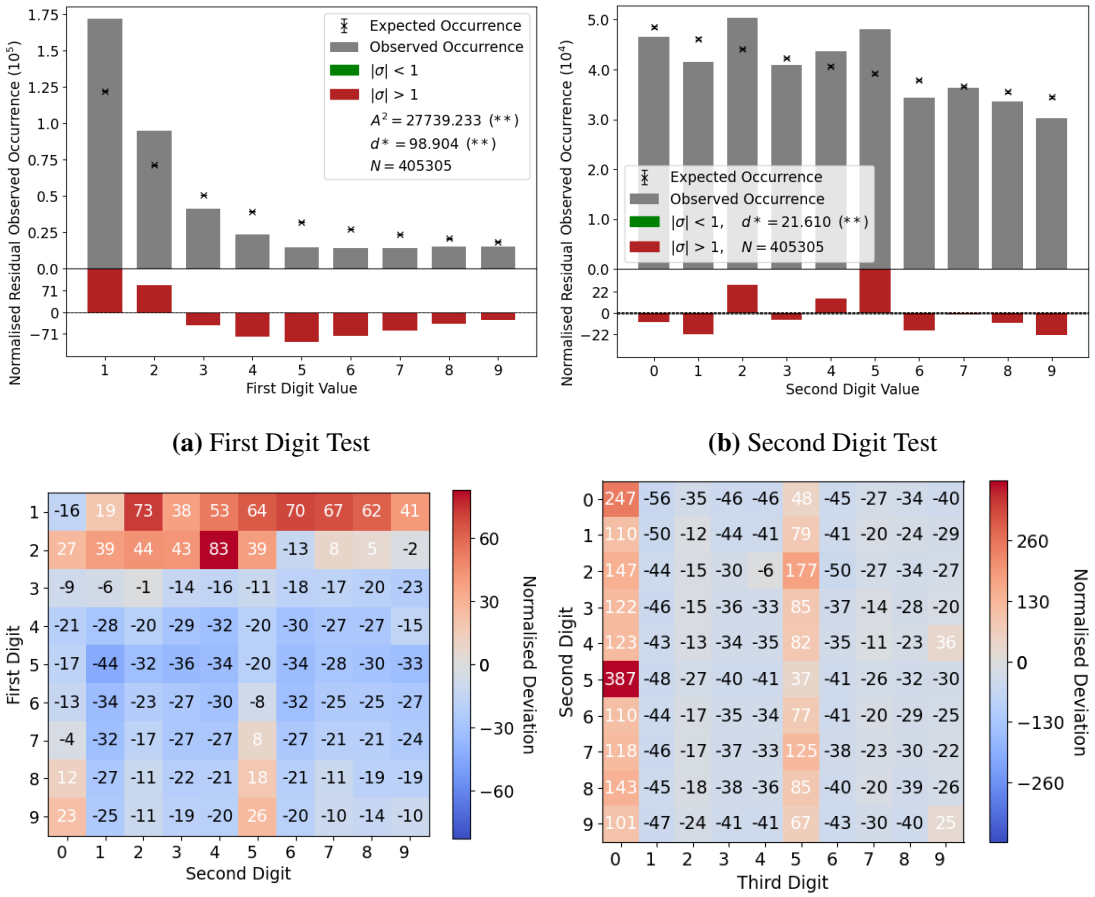
Before this change in SDT, there was a significant motivation to fix property values below these threshold values to avoid higher taxation. After the change in SDT, the tax increase paid only applies to the property's value above each threshold. Therefore there is less of an incentive to price properties below such values.

Moreover, there is a tendency to price houses at given reference points. Namely, we observed that property prices tended to set with zero or five in the second and third digit. A property is more likely to be priced at £250k or £255k than £252,500. This is similar to the cognitive reference points in commodity pricing described by Carslaw [3] in that there is a psychological tendency for properties to be valued at these price points.

3.1. Benford's Law Applied to PPD

Figure 8 shows the results of BL applied to PPD for 2013. Subfigures (a) and (b) show the first and second digit tests, respectively. Each of these tests suggests nonconformity with a Benford distribution at the 0.01 confidence level according to both the d^* and A^2 test statistics. This conclusion comes as no surprise as all the first and second digits deviate from expectation by more than one normalised error. For the first digit test, the distribution is skewed towards one and away from nine, whilst the second digit test shows that five is the most likely digit to appear in the second index.

Similarly, Figure 9 shows a Benford analysis for PPD in 2015. Again, subfigures (a) and (b) show the first and second digit tests, respectively. As before, we observe nonconformity at the 0.01 confidence level. All data points deviate from expectation by more than one normalised residual for both the first and second digit tests. The distribution of first digits is skewed towards one, and five deviates the most from the expected occurrence of second digits.



(a) First Digit Test

(b) Second Digit Test

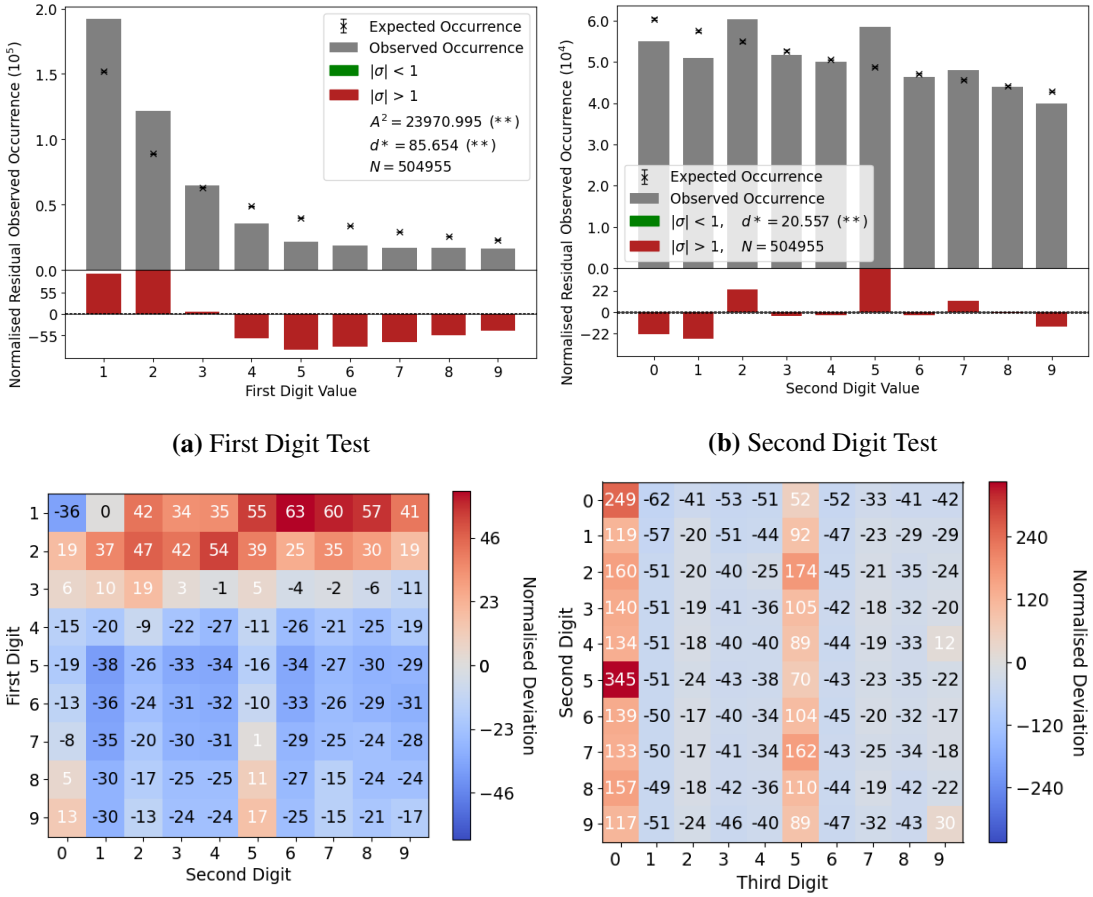
(c) First-Second Digit heatmap showing normalised deviation from BL

(d) Second-Third Digit heatmap showing normalised deviation from BL

FIG. 8: Subfigures (a) and (b) show the expected BL occurrence in the first and second indexes, respectively, and the observed occurrence for PPD in the year 2013. We see clear nonconformity with BL in both cases with all normalised residuals greater than one standard error and p values significant at the 0.01 level for the metrics d^* and A^2. Subfigures (c) and (d) are heatmaps showing the difference between the observed and expected Benford occurrence for the first-second and second-third index pairs, respectively. The heatmaps are normalised with respect to the standard error. Red colouration corresponds to a higher than expected occurrence and blue a lower than expected occurrence. In (a) we note a clear tendency for properties to be priced with one or two in the first index. Moreover, properties tend to be valued with five in the second digit, which can be seen in middle column corresponding to a second digit of five in (c) and the normalised residual at the digits five in (b). (d) shows that properties are priced most frequently with zero or five in the third index, as evidenced by the orange columns corresponding to third digits of zero and five. The most common second-third digit pairs are (0,0) and (5,0).

(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

From a purely traditional Benford analysis standpoint, it is difficult to distinguish the two distributions. Both are highly nonconforming and, upon casual inspection, appear to show the same trends. However, the heatmaps in subfigures (c) and (d) in Figure 8 and Figure 9 reveal subtle variations between the two.



(a) First Digit Test

(b) Second Digit Test

(c) First-Second Digit heatmap showing normalised deviation from BL

(d) Second-Third Digit heatmap showing normalised deviation from BL

FIG. 9: Subfigures (a) and (b) show the expected BL occurrence in the first and second indexes, respectively, and the observed occurrence for PPD in 2015. We see clear nonconformity with BL in both cases with all normalised residuals greater than one standard error and p values significant at the 0.01 level for the metrics d^* and A^2. Subfigures (c) and (d) are heatmaps showing the difference between the observed and expected Benford occurrence for the first-second and second-third index pairs, respectively. The heatmaps are normalised with respect to the standard error. Red colouration corresponds to a higher than expected occurrence, and blue a lower than expected occurrence. Similar to Figure 8, in (a) there is clear evidence suggesting houses are priced with one or two in the first index, as indicated by a high concentration of red values in the first two normalised residuals. Once again, houses tend to be valued with five in the second digit, which can be seen in the middle column corresponding to a second digit of five in (c) and the normalised residual at the digit five in (b). As in 2013, (d) shows that houses are frequently priced with zero or five in the third index, as evidenced by the red columns corresponding to third digits of zero and five. Again the most common second-third digit pairs are (0,0) and (5,0), emphasising the financial and psychological factors affecting property valuations.

(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

Subfigures (c) show the normalised deviation of the first and second digits from a Benford distribution. For PPD in the year 2013, as shown in Figure 8(c), we expect properties to be valued just below the tax threshold values shown in Table II. Indeed there is a spike for the

first and second digits $(D_1, D_2) = (1, 2), (2, 4)$ which correspond to valuations just below the threshold values of £125,001 and £250,001. We observe a similar yet less pronounced result in Figure 9(c). This is most likely due to the relaxed tax rates, as there is less motivation to fix prices below threshold values (see Table II).

Not all the first-second digit values $(1, 2)$ occur in the range £120k $-$ £125k. Indeed for the 2013 PPD, 52.4% of all data points in the range £120k $-$ £130k lie in the range £120k $-$ £125k compared with 48.5% for the 2015 PPD data. This suggests that there is more of a tendency to value properties under the threshold value for valuations made under firm tax rates compared to the relaxed rates.

We observe pricing at the reference points £120k and £125k in the range £120 $-$ £125. Indeed 17.7% and 18.4% of properties are priced at £120k in 2013 and 2015, respectively, in this price range. Moreover, 31.0% and 23.4% are priced at £125k in 2013 and 2015, respectively, showing a clear tendency for valuations to be made with zero or five in the third digit. This psychological bias is shown in Figure 8(d) and Figure 9(d), which show the normalised deviation of second and third digit occurrence from Benford's law. Indeed the most deviant second-third digit pairs are (5,0) and (0,0), showing the significance of this property pricing factor.

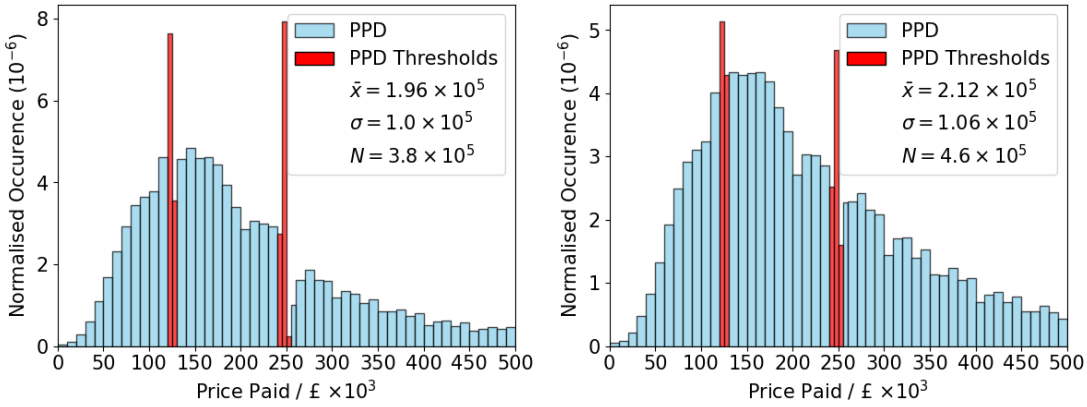
There is also a tendency to price houses just below psychological price points in agreement with the pricing behaviour described by Rosch [22]. In particular, in Figure 8(d) and Figure 9(d) we see spikes at the second-third digit pairs (4,9) and (9,9). Such price points correspond to prices set just below the psychological prices with five and zero in the second digit. This is done to give the impression of a significantly lower price than the property's actual price. That is a house priced at £249k appears to be significantly cheaper than if it were priced at £250k. This is a well-known technique used in sales and is used when marketing properties.

Figure 10 shows histograms of the distribution of PPD in the years 2013 and 2015. Subfigures (a) and (b) show data for 2013 and 2015, respectively. The bins were chosen beginning at one with a width of ten thousand, except around tax thresholds, in which instance they have a width of five thousand. Such a choice allows us to examine properties priced at £125k and £250k to avoid higher tax rates.

In 2013 there was a clear pattern of pricing properties to avoid higher taxation, particularly at the £250,001 threshold, which incurs a 3% tax rate on the entire property. Figure 8(c) shows the same features, as there is a higher count of the first-second digits (1,2) than expected, corresponding to the £125,001 threshold and an increased count of (2,4) corresponding to the £250,001 threshold. Similar yet less prominent features hold in 2015 primarily due to relaxed tax rates. Once again Figure 9(c) displays the same trends with spikes in the first-second digits (1,2) and (2,4) however such features are less appreciable when compared to Figure 8(c).

Although there is a clear trend in valuing properties below these threshold values, such valuations may be given based on psychological bias. That is, properties may be valued at £125k or £250k since there is an aforementioned bias toward pricing properties with a zero or five in the third index. Although there is a clear motivation to set prices at this value to maximise the profit made whilst selling the property and minimising tax paid, it is unclear whether this is done deliberately in every case or instead, whether this behaviour can be attributed to unconscious bias. It should also be noted that there is no mechanism for checking this within the scope of this investigation.

The results of first-digit analyses for both the 2013 and 2015 PPD can be explained using



(a) 2013 PPD showing threshold value drop-off **(b)** 2015 PPD showing threshold value drop-off

FIG. 10: Histograms showing the distribution of PPD data in the years 2013 and 2015. Bars shown in red represent prices near the tax threshold values £125,001 and £250,001. The legend shows the mean $\bar{x}$ and standard deviation σ of house prices as well as the number of house prices analysed N. There is clear pattern of houses been priced below these tax thresholds corresponding to spikes in our histograms just below these values and deficits immediately after. Indeed this trend is most pronounced in subfigure (a) under the firm tax rate in 2013, suggesting valuations are made with tax minimisation taken into consideration. Similar features hold in 2015 under the relaxed SDT rate however they are less pronounced, indicating that avoiding higher tax brackets was less of a concern under the relaxed rates of tax in 2015.

Figure 10. For example, in 2015, 62% of data points are in the range £100k − £299k. Indeed, inspecting Figure 10 reveals that most property prices are in the range £100k − £299k. Thus we would expect deviations from Benford's law due to this concentration of data points close to the mean (£212k). A similar argument holds for the 2013 PPD.

Neither PPD distribution conforms with BL and they appear on face value to display roughly the same trends. Although a traditional Benford analysis reveals this nonconformity, there is little insight indicating why this should be the case. Only upon closer inspection were the underlying causes of the features of each distribution realised. Whilst Benford's law effectively identifies data manipulations for PPD, further inquiry will be required to determine how and why these manipulations were made.

4. AIG Securities Fraud

On June 6th, 2005, the Securities and Exchange Commission (SEC) announced legal action against a former executive of the General Re Corporation (GEN RE) for their role in aiding American International Group (AIG) in committing FF [23]. AIG was accused of falsifying financial statements by engaging in several misleading transactions, "whose purpose was to paint a falsely rosy picture of AIG's financial results to analysts and investors" [24].

The most notable fraudulent transactions committed by AIG were two reinsurance agreements with GEN RE. The purpose of such transactions was to add $500 million to AIG's accounting and balance sheets across the final quarter of 2000 and the first quarter of 2001. These

transactions were intended to reduce AIG's losses in the third quarter of 2000 by introducing misleading accounting information.

These falsified loss reserve figures, which remained on AIG's financial statements and improperly inflated AIG's loss reserves by $500 million. This section applies a Benford analysis to two annual reports filed with the SEC by AIG in 2003 and 2004. In the 2004 report, AIG was forced to restate their financial figures to account for their fraudulent activity. We would therefore expect to find a change in the distribution of digits between 2003 (where fraud has not yet been accounted for) and 2004. In particular, we expect a higher reserve for loss to be reported in 2003 compared to 2004.

4.1. Benford's Law Applied to AIG's Annual Reports

We apply BL to two 10-K annual reports filed by AIG with the SEC [12]. We analyse sections of each report, including loss reserve data, looking for differences between the reports that could be indicative of FF. We expect worse conformity with BL in the 2003 report, in which fraud has not yet been accounted for, than in the restated 2004 report.

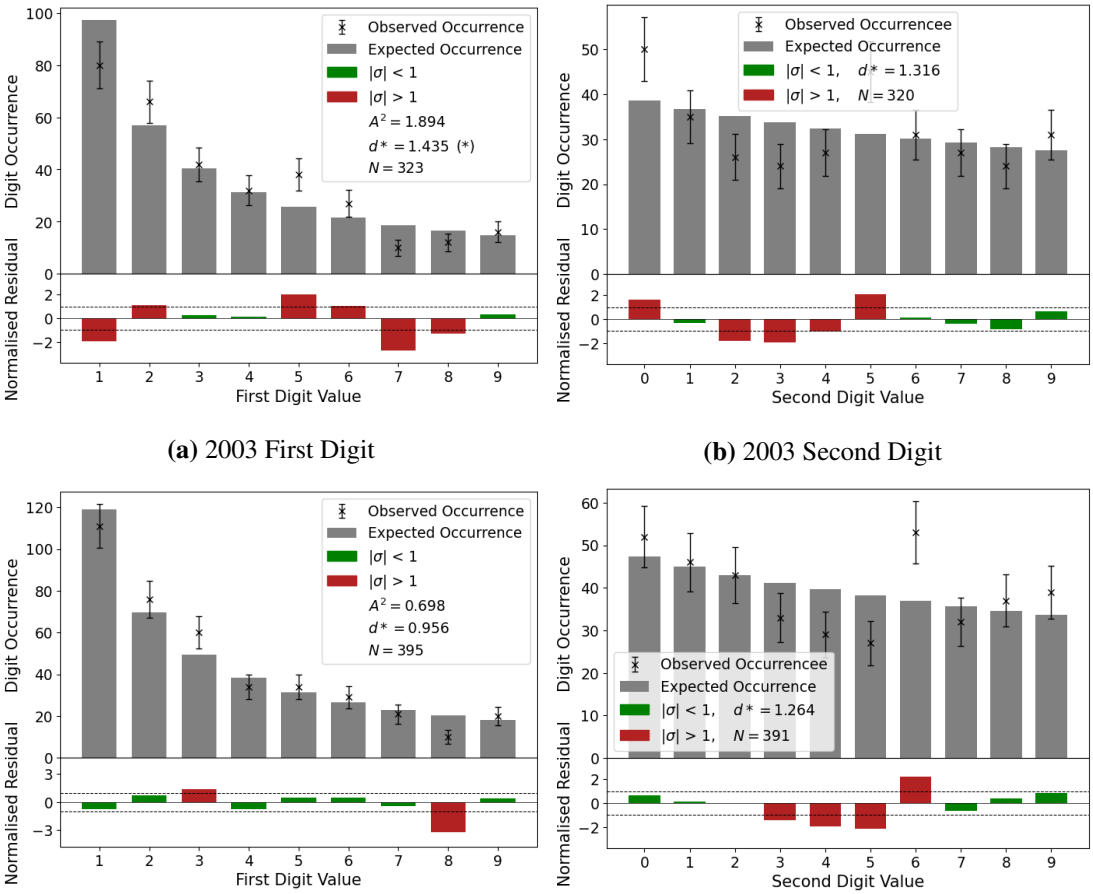
Software was written to extract financial data from identical sections of each annual report (see Appendix E for more details). These sections included selected consolidated financial data, revenues and income derived from operations and insurance investment operations, amongst other sections reporting financial data. We also only consider data relating to years appearing in both reports. For example, for the 2004 report, we ignore data relating to the year 2004 as this data will not be present in the 2003 report. This is to keep our analysis consistent and detect legitimate changes in reported data between the two reports.

Figure 11 shows the result of this analysis. The top row (subfigures (a) and (b)) shows the original data filed in 2003, and the bottom row (subfigures (c) and (d)) shows the restated data filed in 2004. We observe relatively poor conformity with BL in 2003 compared with 2004 in the first index. For the first digit test, in 2003, six out of nine residuals deviate from BL by more than one standard error compared with two in 2004. Moreover, our test statistics decrease between 2003 and 2004, from 1.89 to 0.69 for A^2 and 1.44 to 0.96 for d^*. Thus on an inspection of the first index, we conclude that conformity with BL improves between the two years. This suggests that, under the null hypothesis that non-fraudulent financial data conforms with BL, AIG engaged in fraudulent activity.

Different features hold for the second digit test; however, they are less pronounced. In 2003 and 2004, four out of ten normalised residuals deviate from BL by more than one standard error. Moreover, the values of our metrics do not change significantly between the two years. For example, in 2003 d^* is 1.32, compared with 1.26 in 2004. This suggests that there is no appreciable change in conformity with BL based on the second digit test. Hence we cannot conclude that fraud has taken place from the results of the second digit test.

Figure 11 has shown that our digit tests can lead us to conflicting conclusions. Indeed we do not have sufficient justification to conclude that fraud has occurred based on a Benford analysis. This shows us the importance of applying several different digit test to capture the distribution of digits over several indexes.

Figure 12 shows the application of BL to loss reserve data in annual reports filed by AIG in 2003 and 2004. The data relates to the years $1994 - 2003$. The top row (subfigures (a) and



(a) 2003 First Digit

(b) 2003 Second Digit

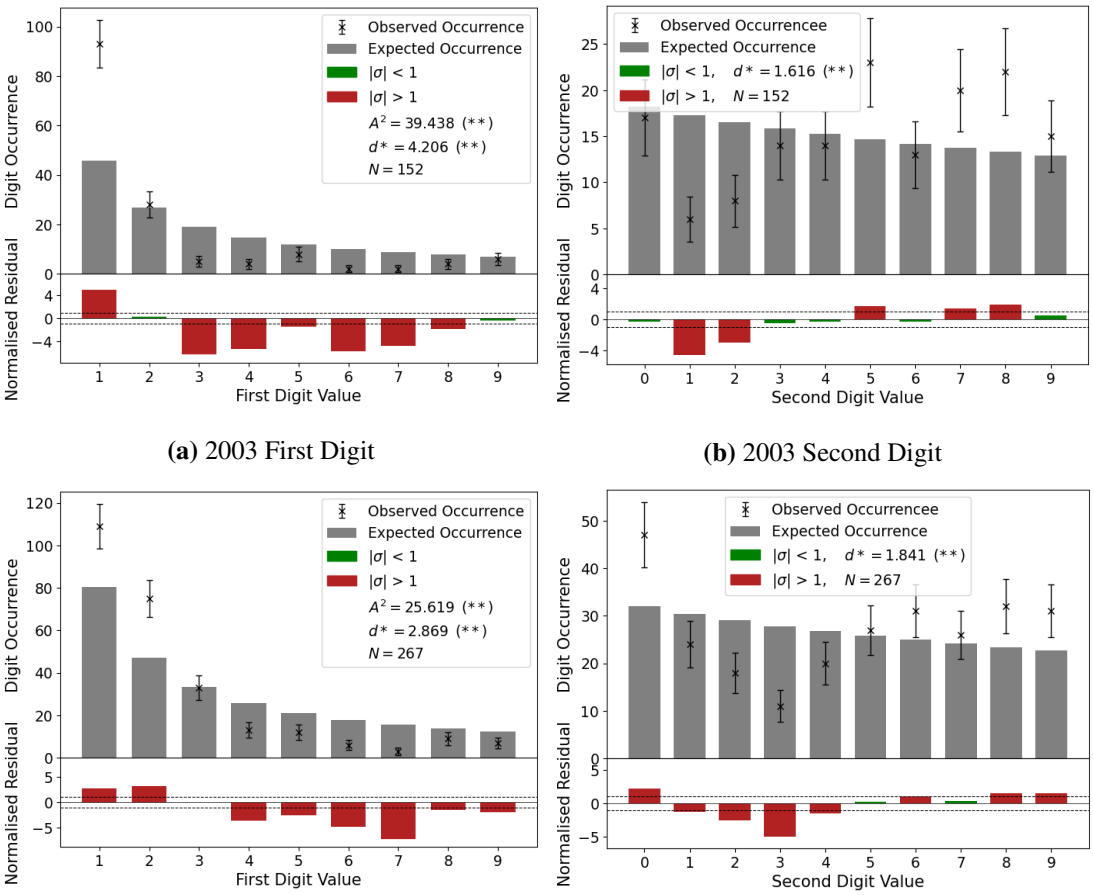
(c) 2004 First Digit (restated)

(d) 2004 Second Digit (restated)

FIG. 11: A Benford analysis of data extracted from identical sections of annual reports filed by AIG. The reports analysed were filed in 2003 and 2004 where we have extracted data relating to the years $2000 - 2003$. The top row shows the original data filed in 2003, and the bottom row the restated figures filed in 2004 whilst the lefthand column shows the first digit test and the righthand column the second digit test. For the first digit test, we see a relative improvement in conformity between 2003 and 2004. Six out of a possible nine residuals deviate by more than one standard error in 2003 compared with two in 2004. Furthermore, the values of our metrics decrease between the two years. However, we see no signs of improving conformity for the second digit test, with the same number of residuals deviating by more than one standard error and a similar value of our test statistics. Therefore we cannot conclude that fraud has occurred based on the tests performed.

(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

(b)) shows the original data filed in 2003, and the bottom row (subfigures (c) and (d)) shows the restated data filed in 2004. For the first digit test in 2003, we observe a relative improvement in conformity with BL between 2003 and 2004. In particular, our test statistics' values decrease from 39.44 to 25.62 for A^2 and 4.21 to 2.87 for d^*. However, our metrics are significant at the 0.01 level in both these years. Thus despite this improvement, we are not justified in assuming the distribution of first digits is Benford in 2004. Indeed we see a similar number of residuals deviating by more than one standard error, with seven in 2003 and eight in 2004. We conclude that both these distributions do not conform with BL in the first index.



(a) 2003 First Digit

(b) 2003 Second Digit

(c) 2004 First Digit (restated)

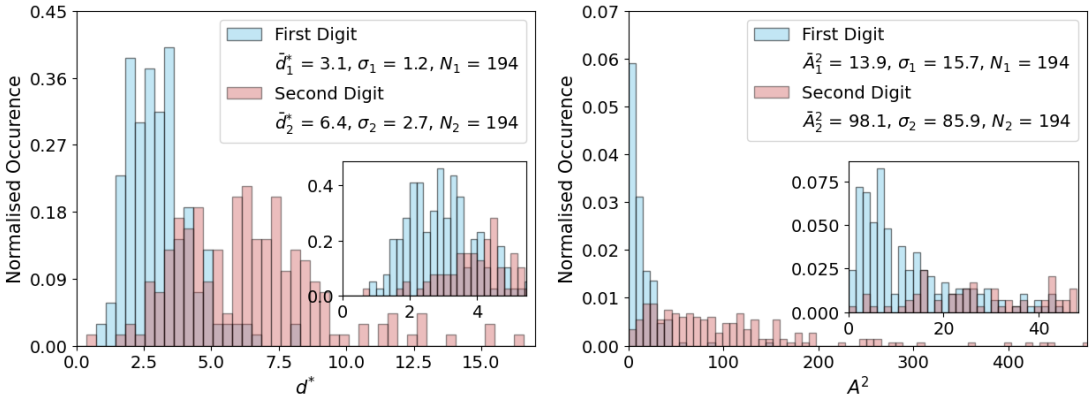
(d) 2004 Second Digit (restated)

FIG. 12: A Benford analysis of loss reserve data extracted from reports filed by AIG in 2003 and 2004. We have extracted data relating to the years $1994 - 2003$. The top row shows the original data filed in 2003, and the bottom row the restated figures filed in 2004 whilst the lefthand column shows the first digit test and the righthand column the second digit test. While we observe an improvement in conformity for the first digit test based on the test statistics, both years are significant at the 0.01 level, suggesting both years do not conform with BL. Moreover, a similar number of residuals deviate by more than one standard error: seven and eight 2003 and 2004 respectively. For the second digit test, we see a decrease in conformity based on our test statistics. There is no improvement based on residuals, with four deviating by more than one standard error in both years. Thus we have no justification to assert that fraud has occurred based on an analysis of the loss reserve data.

(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

In the second index, we observe worsening conformity with BL between 2003 and 2004. Three more residuals deviate by more than one standard error in 2004 compared with 2003. Our test statistics give conflicting results. d^* increases between 2003 and 2004 from 1.62 to 1.84 suggesting poorer conformity with BL in 2004. A^2 decreases from 4.90 to 3.91, suggesting a relative improvement in conformity in 2004. However, as with the first index, our metrics are significant at the 0.01 level in both these years. Therefore we once again conclude that these distributions do not conform with BL in the second index.

We have seen examples of a Benford analysis applied to two 10-K annual reports filed with



(a) d^* **(b)** A^2

FIG. 13: Histogram plots of d^*, in subfigure (a), and A^2, in subfigure (b), and (b), applied to our set of 194 annual reports. The first digit test is shown in blue, and the second digit test in red. The legend shows the mean values of our metrics in both cases alongside the standard deviation, σ_i, and the number of data-points in each set, N_i. We observe far higher values of our metrics when applying the second digit test showing that annual reports do not conform with BL in the second index compared with the first. This can be seen in the insets showing our metrics at lower values. We should also note that despite the first digit test showing better conformity, the majority of annual reports have test statistics above the 0.01 significance level. Therefore, it is unlikely that annual reports filed with SEC conform with BL in the first and second digits.

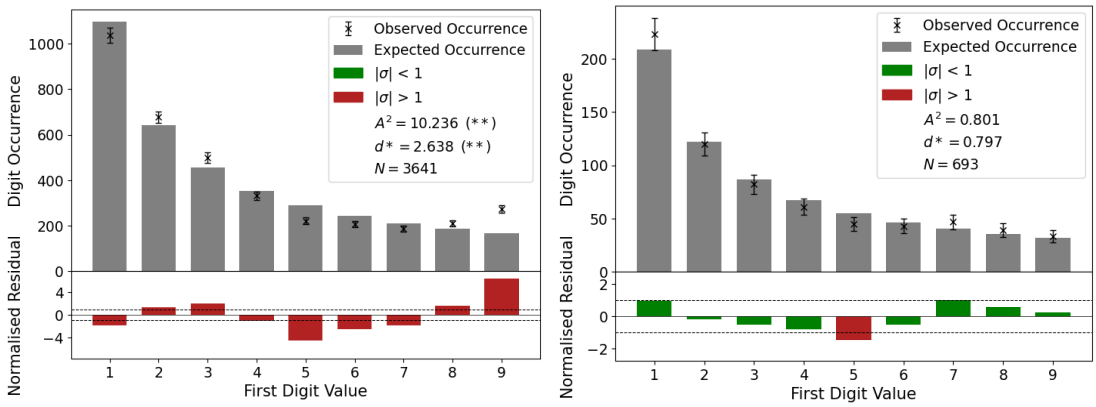
the SEC. Furthermore, we have compared identical sections of these reports and loss reserve data. To gain insight into our results, we now analyse a set of annual reports and assess conformity with BL. This will put the investigation into AIG into context and determine whether we expect data in annual reports to conform with BL.

4.2. General Annual Reports: Should we expect conformity with Benford's Law?

We now apply a Benford analysis to a variety of 10-K annual reports filed with the SEC in 2002 − 2006. In particular, we will calculate our test statistics for data extracted from 194 annual reports. Software was written to download, extract and sanitise data from a list of annual reports which were manually collected (see Appendix E for more details). We then computed our test statistics after an application of BL in the first and second indexes.

Figure 13 shows histogram plots of d^* and A^2 in subfigures (a) and (b) respectively applied to our set of annual reports. The first digit test is shown in blue, and the second digit test in red.

We see a drastic change in conformity with BL in the first and second index for both metrics. This is apparent after a visual inspection of both figures, with most second digit histogram bars appearing at higher values of our test statistics. In particular, for d^*, the mean value in the first digit test is 3.1 compared with 6.4 for the second digit test. This is a significantly higher mean value of the second digit test. We note similar more pronounced features for A^2, with a mean value for the first digit test of 13.9 compared with 98.1 for the second digit test. Therefore we expect annual reports to be far more nonconforming when applying the second digit test compared with the first digit test.



(a) Original

(b) Non-Financial Data Removed

FIG. 14: The first digit test applied to the Chateau Communities INC 2002 annual report. Subfigure (a) shows the entire report, whilst in subfigure (b), we have removed non-financial data relating to real estate. All normalised residuals deviate by more than one standard error for the original report, whilst only two deviate when irrelevant data has been removed. Moreover, our test statistics decrease dramatically from values at the 0.01 significance levels to values that suggest good conformity with BL. This suggests that annual reports may deviate from BL because they contain irrelevant non-financial data.

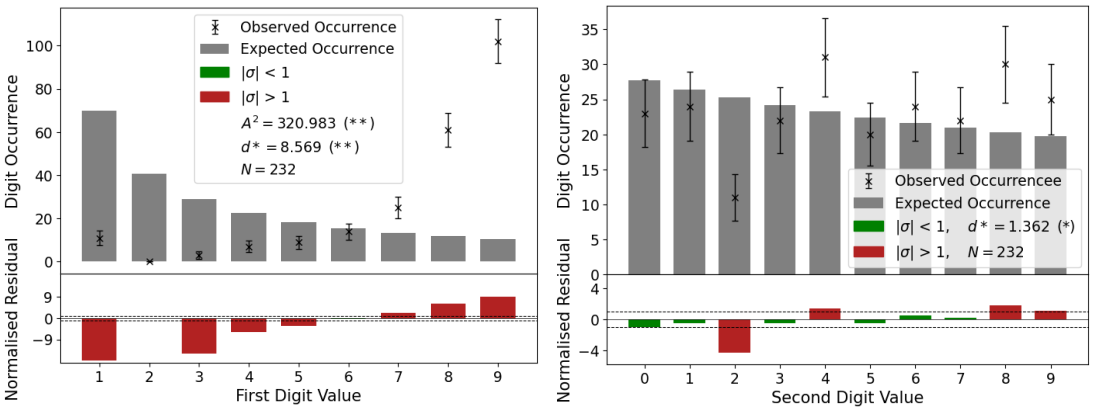
(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

However, consulting our p values once again, we note that even though the distribution of digits in the first index conforms with BL compared to the second index, our metrics' values suggest nonconformity with BL. Mean values in the first index of 3.1 and 13.9 for d^* and A^2 respectively are significantly higher than the 0.01 significance values of 1.66 and 4.56. Indeed only 14 and 40 of a possible 194 annual reports have values less than the 0.01 significance values for d^* and A^2, respectively. Therefore it is unlikely that annual reports will conform with BL in both indexes.

4.3. Irrelevant Data in Annual Reports

Several conjectures were made as to why we observe this disconformity. The most notable is that SEC reports often contain data that is not financial. We investigate the 10-K annual report filed by Chateau Communities INC in 2002. This dataset is of particular interest to us in the context of irrelevant data included in annual reports since this set contains large sections relating to real estate owned by the company. In particular, this includes entries such as the number of properties owned in a given area code, the number of properties occupied (as a percentage) and the average rent paid per site. To account for this non-financial data, we removed the report's real estate sections from our set and looked for changes in conformity with BL. We also analysed a subset of the real estate data separately, showing nonconformity with BL.

Figure 14 shows the first digit test applied to Chateau Communities INC's 2002 annual report. Subfigure (a) shows the data extracted from the entire report, whilst subfigure (b) shows the result with all non-financial data removed. All normalised residuals deviate by more than one standard deviation for the entire report, and the d^* and A^2 metrics are significant at the



(a) First Digit **(b)** Second Digit

FIG. 15: BL in the first index, shown in subfigure (a), and second index, shown in subfigure (b), applied to the percentage of properties occupied per area code owned by Chateau Communities INC in 2002. The first digit test is highly nonconforming with BL with all but one normalised residual deviating from BL by more than one standard error and high values of our metrics significant at the 0.01 level. We observe better conformity for the second digit test with only four normalised residuals deviating by more than one standard error. d^* is 1.36, significant at the 0.05 level, and A^2 is 4.40, significant at the 0.01 level. Note that for the first digit test the digit two has zero observed occurrences and thus does not have an associated error or residual.

(p values for hypothesis testing: * significant at the .05 level ** significant at the .01 level)

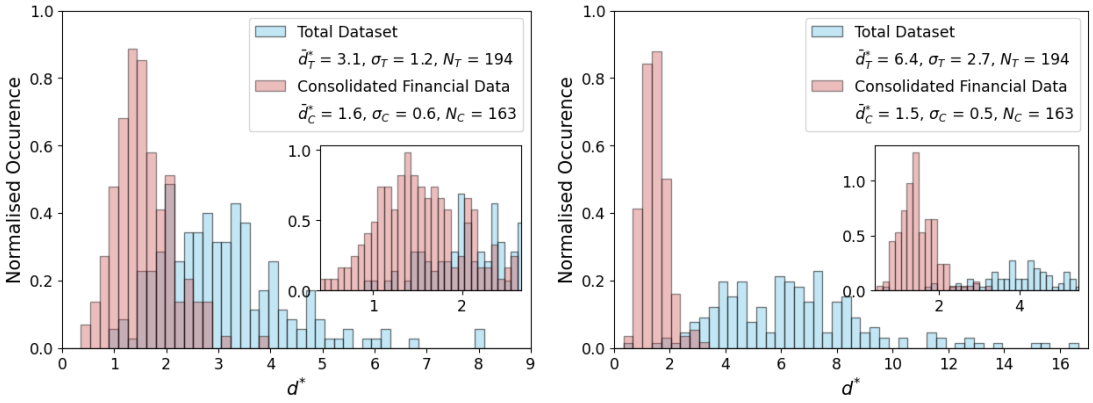
0.01 level. However, with irrelevant data removed from the report, only two residuals deviate by more than one standard error. We see a decrease in both test statistics, suggesting better conformity with BL. Therefore, with non-financial data removed from the report, we observe far better conformity with BL.

Figure 15 shows BL applied to a subset of real estate data found in Chateau Communities' 2002 annual report. The data relates to the percentage of properties occupied per area code. In the first index, we see poor conformity with BL, with a distribution almost opposite to what we would expect for a Benford compliant set. Eight normalised residuals deviate by more than one standard error, and our metrics are significant at the 0.01 level. The second index conforms better with BL; however, four residuals deviate by more than one standard error, and d^* and A^2 are significant at the 0.05 and 0.01 levels, respectively. This poor agreement with BL is due to the distribution of the percentage of occupied properties. Most area codes have roughly $80 - 100\%$ of properties occupied shown in the relative spikes at the digits $1, 8$ and 9.

4.4. Consolidated Financial Data

Having seen that annual reports deviate from BL, at least in part, since they contain irrelevant non-financial data, we will analyse consolidated financial data collected from our set of annual reports. Consolidated financial data relates to a company's financial affairs and its subsidiaries together as if they were one entity [25]. This will determine if annual reports filed with the SEC contain relevant Benford compliant financial data.

Figure 16 shows d^*, applied in the first index in subfigure (a) and the second in subfigure (b),



(a) First Digit **(b)** Second Digit

FIG. 16: Histogram plots of d^* computed for data extracted from the entire annual report, shown in blue, and consolidated financial data, shown in red, for our set of annual reports. Subfigure (a) shows the first digit test, and subfigure (b) shows the second digit test. We observe lower values of d^* for both digit tests when considering consolidated financial data compared with the entire dataset. When analysing consolidated data, the mean value of d^* decreases from 3.1 to 1.6 for the first digit test and decreases from 6.4 to 1.5 for the second digit test. Therefore, we can expect far better conformity with BL when considering consolidated financial data stated in annual reports compared with data extracted from the entire report.

to the entire dataset extracted from each annual report, shown in blue, and sections relating to consolidated financial data in each report, shown in red. We observe lower values of d^* when considering consolidated financial data for both digit tests. This is evidenced by the distribution of red bars, which are skewed to the left of blue bars. Moreover, for the first digit test, the mean value of d^* decreases from 3.1 to 1.6 whilst for the second digit test, we see a decrease from 6.4 to 1.5. Therefore, we conclude that for 10-K annual reports, we expect better conformity when considering subsets of the reports relating to consolidated financial data rather than the entire report.

We expect a more concentrated distribution of our d^* metric for consolidated financial data compared with the entire dataset. For the first digit test, the standard deviation of d^* decreases from 1.2 to 0.6, and for the second digit test, we see a decrease from 2.7 to 0.5. This again suggests better conformity with BL for consolidated financial data since the distribution of d^* is concentrated around lower values.

Note that we only have 163 data points for both digit tests for consolidated financial data, compared with 194 when considering the entire report. This is because some of the reports we consider do not have any sections relating to consolidated financial data. We observe similar behaviour for the A^2 metric.

In general 10-K annual reports filed with the SEC do not conform with BL. A significant factor in this is that such reports tend to contain irrelevant non-financial data. Figure 16 shows that by removing this data and considering only a subset of financial data, namely consolidated financial data, we observe far better conformity with BL. When identifying fraudulent activity based on BL it is more appropriate to consider subsets of financial data in annual reports rather

than all the data contained therein.

5. Conclusions and Future Considerations

5.1. Conclusions

This report gave an overview of the theoretical background behind BL, including various tests that are utilised to assess the conformity of data with this law. We also formalised the FRBL, which takes into consideration the dynamical range of our dataset. In doing so, we derived the second digit version of this law. However, during this investigation, finite range techniques were not helpful as most datasets considered were large and spanned several orders of magnitude, meaning any results derived from a finite range analysis gave similar results to BL.

Several methods of assessing conformity, namely the d^* and A^2 metrics, were introduced. We described generating synthetic Benford sets using a brute force application of the law and geometric series, discussing the differences between each technique, namely the natural deviation in brute forces sets and the analytical sets produced via geometric series. We also calculated the p values associated with each metric for the first and second digit tests by analysing a large collection of generated Benford sets. By introducing artificial deviations into a synthetic Benford set, we were able to study the changes in our metrics under two types of deviation expected in fraudulent datasets: rounding behaviour and digit exaggeration. In doing so, we found that d^* is sensitive to the shape of the distribution rather than individual deviations in digits compared with A^2. We arrived at the same conclusion when considering the variation of these metrics over different ranges of data.

We analysed PPD and found that BL was effective in identifying data manipulations. Using a combination of heatmap and histogram data visualisation, we were able to identify the financial and psychological factors behind property pricing. Whilst BL was effective in identifying these factors, a further investigation was required to verify these findings. In general, BL is a useful technique to identify data manipulations but will always demand a more rigorous examination of its findings, particularly in FF, which can have severe consequences for individuals and companies involved.

Finally, we examined a confirmed case of FF, namely securities fraud committed by AIG around the turn of the century. We analysed annual reporting data for AIG filed with the SEC to detect differences between the original and restated data, which accounted for AIG's fraudulent activity. Whilst on face value, there appeared to be an improvement in conformity between the two cases, a more general study of annual reports showed that they do not conform with BL. This is primarily due to the fact that they contain non-financial data. By considering a subset of financial data, namely consolidated financial data, within these reports, we observed far better conformity. Therefore the application of BL to detect FF based on annual reports should aim to analyse subsets of relevant financial data rather than the entire report.

5.2. Future Considerations

In the future, this investigation could be extended to analyse further financial datasets and cases of fraud. By applying the techniques presented in this work to different financial datasets, such as income data for various companies, we could determine if there is some probability of fraud

in income data. Moreover, an application of BL to other fraud cases could yield interesting results. Considering the results of this report, subsets of annual reports should be analysed rather than in the whole report in its entirety.

Another possible avenue of investigation is test statistic analysis. While this work introduces two types of artificial deviations into Benford sets, more complex deviations could be considered, for instance, combining different types of deviations. This would give a richer picture of our test statistics and determine which types of deviations from BL are likely to be detected.

It would also be beneficial to understand in which circumstances the FRBL is a more appropriate technique than BL. Throughout this report, we considered datasets with large dynamical ranges such that the FRBL was not appropriate. However, understanding in which cases we should apply the FRBL would be useful. Moreover, once the degree of conformity with BL has been established, it may be useful to apply the FRBL to subsets of data to identify further trends in the dataset. An example of this might be to use the FRBL to analyse data around tax threshold values for PPD. Moreover, calculation of the p values associated with the FRBL over various dynamical ranges of data would be required. This could be extended to analyse the variation of p values over different dynamical ranges of data.

References

[1] S. Newcomb, "Note on the frequency of use of the different digits in natural numbers," American Journal of mathematics, vol. 4, no. 1, pp. 39–40, 1881, doi: 10.2307/2369148.

[2] F. Benford, "The Law of Anomalous Numbers," Proceedings of the American Philosophical Society, vol. 78, no. 4, pp. 551–572, 1938, url: https://www.jstor.org/stable/984802.

[3] C. A. P. N. Carslaw, "Anomalies in Income Numbers: Evidence of Goal Oriented Behavior," The Accounting Review, vol. 63, no. 2, pp. 321–327, 1988, url: https://www.jstor.org/stable/248109.

[4] J. K. Thomas, "Unusual Patterns in Reported Earnings," The Accounting Review, vol. 64, no. 4, pp. 773–787, 1989, URL: https://www.jstor.org/stable/247861.

[5] T. Van Caneghem, "Earnings management induced by cognitive reference points," British Accounting Review, vol. 34, no. 2, pp. 167–178, 2002, doi: 10.1006/bare.2002.0190.

[6] C. Tilden and T. Janes, "Empirical evidence of financial statement manipulation during economic recessions," Journal of finance and accountancy, vol. 10, p. 1, 2012, url: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.470.5865rep=rep1type=pdf.

[7] A. Mehta and G. Bhavani, "Application of Forensic Tools to Detect fraud: The Case of Toshiba," Journal of Forensic and Investigative Accounting, vol. 9, no. 1, pp. 1188–1197, 2017, url: https://www.researchgate.net/publication/312623844.

[8] T. Omotehinwa and S. Ramon, "Fibonacci numbers and golden ratio in mathematics and science," International Journal of Computer and Information Technology, vol. 2, no. 4, pp. 630–638, 2013, url: https://www.researchgate.net/publication/334015286.

[9] S. Kunoff, "N! has the first digit property," The Fibonacci Quarterly, vol. 25, no. 4, pp. 365–367, 1987, url: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.388.6428rep=rep1type=pdf.

[10] I. Hughes and T. Hase, Measurements and their uncertainties: a practical guide to modern error analysis. Oxford University Press, 2010, p. 30.

[11] M. Sambridge, H. Tkal, and P. Arroucau, "Benford's Law of First Digits: From Mathematical Curiosity to Change Detector," Asia Pacific Mathematics Newsletter, vol. 1, no. 4, pp. 1–6, 2011, url: https://www.researchgate.net/publication/236176609.

[12] "SEC Filings," U.S. Securities and Exchange Commission, 2021, Accessed: 29/3/2021. [Online]. Available: https://sec.report/Document

[13] M. J. Nigrini, Benford's Law: Applications for forensic accounting, auditing, and fraud detection. John Wiley & Sons, 2012, vol. 586.

[14] W. K. Cho and B. J. Gaines, "Breaking the (Benford) law: Statistical fraud detection in campaign finance," American Statistician, 2007, doi: 10.1198/000313007X223496.

[15] J. Morrow, "Benford's law, families of distributions and a test basis," Centre for Economic Performance, London School of Economics and Political Science, 2014, url: http://eprints.lse.ac.uk/60364/1/dp1291.pdf.

[16] R. L. Wasserstein and N. A. Lazar, "The ASA Statement on p-Values: Context, Process, and Purpose," The American Statistician, vol. 70, no. 2, pp. 129–133, 2016, doi: 10.1080/00031305.2016.1154108.

[17] G. E. Noether, "Note on the kolmogorov statistic in the discrete case," Metrika: International Journal for Theoretical and Applied Statistics, vol. 7, no. 1, pp. 115–116, 1963, doi: 10.1007/BF02613966.

[18] V. Choulakian, R. Lockhart, and M. Stephens, "Cramér-von Mises statistics for discrete distributions," Canadian Journal of Statistics, 1994, doi: 10.2307/3315828.

[19] O. Renaud and M.-P. Victoria-Feser, "A robust coefficient of determination for regression," Journal of Statistical Planning and Inference, vol. 140, no. 7, pp. 1852–1862, 2010, doi: 10.1016/j.jspi.2010.01.008.

[20] HM Land Registry, "Archived Price Paid Data: 1995 to 2017," Available at: https://data.gov.uk/dataset/314f77b3-e702-4545-8bcb-9ef8262ea0fd/archived-price-paid-data-1995-to-2017, accessed: 06-01-20.

[21] The National Achieves. (2015) Stamp Duty Land Tax Act 2015. Accessed: 07-01-20. [Online]. Available: https://www.legislation.gov.uk/ukpga/2015/1/contents

[22] E. Rosch, "Cognitive reference points," Cognitive Psychology, vol. 7, no. 4, pp. 532 – 547, 1975, doi: 10.1016/0010-0285(75)90021-3.

[23] "SEC Charges GEN RE Executive for Aiding in AIG Securities Fraud," U.S. Securities and Exchange Commission, 2005, Accessed: 7/2/2021. [Online]. Available: https://www.sec.gov/news/press/2005-85.htm

[24] "SEC Complaint: American International Group, Inc." U.S. Securities and Exchange Commission, 2006, Accessed: 7/2/2021. [Online]. Available: https://www.sec.gov/litigation/complaints/comp19560.pdf

[25] R. J. Huefner and J. A. Largay, "Consolidated financial reporting: Accounting issues, financial reporting choices, and managerial implications," Journal of Managerial Issues, vol. 2, no. 1, pp. 26–40, 1990, url: https://www.jstor.org/stable/40603705.

Appendix A. Normalisation Of General Benford Formula

In this section we prove that Equation 2 is well defined. To reiterate, the integer n appears at the beginning of a number N with probability,

$$P[n] = \log\left(1 + \frac{1}{n}\right),$$

providing that the decimal expansion of n has at most the same number of terms as the decimal expansion of N.

We wish to prove that this formula is normalised such that all probabilities sum to one. We need to prove that,

$$\sum_{n \in A_\gamma} P_n = \sum_{n \in A_\gamma} \log\left(1 + \frac{1}{n}\right) = 1,$$

where $A_\gamma = \{m \in \mathbb{N} : 10^{\gamma-1} \leq m \leq 10^\gamma - 1\}$, with $\gamma \in \mathbb{N}$, is the set of integers of length γ appearing at the start of N. For example, for integers of length two appearing at the beginning of our number N, take $\gamma = 2$. Then,

$$A_2 = \{m \in \mathbb{N} : 10 \leq m \leq 99\} = \{10, 11, ..., 98, 99\},$$

which is the set of integers of length two allowed to appear at the beginning of N, since we do not consider numbers beginning with zero.

Proof: Take γ to be fixed. Then,

$$\begin{aligned}
\sum_{n \in A_\gamma} \log\left(1 + \frac{1}{n}\right) &= \sum_{j=10^{\gamma-1}}^{10^\gamma - 1} [\log(j+1) - \log(j)] \\
&= [\cancel{\log(10^{\gamma-1}+1)} - \log(10^{\gamma-1})] + [\cancel{\log(10^{\gamma-1}+2)} - \cancel{\log(10^{\gamma-1}+1)}] + \\
&... + [\log(10^\gamma) - \cancel{\log(10^\gamma - 1)}] \\
&= \log(10^\gamma) - \log(10^{\gamma-1}) \\
&= \log\left(\frac{10^\gamma}{10^{\gamma-1}}\right) \\
&= \log(10) \\
&= 1.
\end{aligned}$$

This ensures that probabilities sum to one and our equation is normalised properly.

Appendix B. Proof of Second Digit Finite Range Law

We work with a reciprocal density function to generate a Benford compliant set. The integral of the density function $P(x) = x^{-1}$ over the entire range gives the normalising constant λ_c:

$$\begin{aligned}\lambda_c &= \int_{a\times 10^{\alpha}}^{b\times 10^{\beta}} \frac{1}{x}\,\mathrm{d}x \\ &= (\beta-\alpha)\ln(10) + \ln\left(\frac{a}{b}\right).\end{aligned}$$

Define the disjoint sets $T_1 := [a\times 10^{\alpha}, 10^{\alpha+1}]$, $T_2 := [10^{\alpha+1}, 10^{\beta}]$ and $T_3 := [10^{\beta}, b\times 10^{\beta}]$ such that $U = T_1 \cup T_2 \cup T_3$. We consider each of these ranges separately.

$$\underline{T_2 := [10^{\alpha+1}, 10^{\beta}]}$$

Firstly consider the range $A^{\alpha+1} = [10^{\alpha+1}, 10^{\alpha+2}]$. The set $A^{\alpha+1}_{d_2}$, defined as the subset of $A^{\alpha+1}$ with $d_2 \in \{0,1,...,9\}$ as the second digit, is given by:

$$A^{\alpha+1}_{d_2} = \bigcup_{d_1=1}^{9} \left[(d_1 + 10^{-1}d_2)\times 10^{\alpha+1}, (d_1 + 10^{-1}(d_2+1))\times 10^{\alpha+1}\right].$$

We need to calculate the density function over $A^{\alpha+1}_{d_2}$, namely,

$$\begin{aligned}\int_{A^{\alpha+1}_{d_2}} \frac{1}{x}\,\mathrm{d}x &= \sum_{d_1=1}^{9} \int_{(d_1+10^{-1}d_2)\times 10^{\alpha+1}}^{(d_1+10^{-1}(d_2+1))\times 10^{\alpha+1}} \frac{1}{x}\,\mathrm{d}x \\ &= \sum_{d_1=1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right).\end{aligned}$$

Thus the integral of the density function over all of $T_2 = \bigcup_{i=1}^{\beta-\alpha-1} A^{\alpha+i}_{d_2}$ is given by,

$$\begin{aligned}\sum_{i=1}^{\beta-\alpha-1} \int_{A^{\alpha+i}_{d_2}} \frac{1}{x}\,\mathrm{d}x &= \sum_{i=1}^{\beta-\alpha-1}\sum_{d_1=1}^{9} \int_{(d_1+10^{-1}d_2)\times 10^{\alpha+1}}^{(d_1+10^{-1}(d_2+1))\times 10^{\alpha+1}} \frac{1}{x}\,\mathrm{d}x \\ &= \sum_{i=1}^{\beta-\alpha-1}\sum_{d_1=1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right) \\ &= (\beta-\alpha-1)\sum_{d_1=1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right).\end{aligned}$$

$$\underline{T_1 := [a\times 10^{\alpha}, 10^{\alpha+1}]}$$

The set of elements in T_1 with a second digit d_2, B_{d_2}, is given by:

$$\sigma_i := \bigcup_{d_1=i}^{9} \left[(d_1 + 10^{-1}d_2)\times 10^{\alpha+1}, (d_1 + 10^{-1}(d_2+1))\times 10^{\alpha+1}\right],$$

$$B_{d_2} = \begin{cases} \sigma_{a_1} & : d_2 > a_2, \\ \sigma_{a_1+1} & : d_2 < a_2, \\ [a\times 10^{\alpha}, (a_1 + (a_2+1)\times 10^{\alpha}] \cup \sigma_{a_1+1} & : d_2 = a_2. \end{cases}$$

We then take the integral over B_{d_2} of the density function to obtain the coefficient λ_a:

$$\lambda_a := \int_{B_{d_2}} \frac{1}{x}\,\mathrm{d}x = \begin{cases} \sum_{d_1=a_1}^{9} \int_{(d_1+10^{-1}d_2)10^{\alpha}}^{(d_1+10^{-1}(d_2+1))10^{\alpha}} \frac{1}{x}\,\mathrm{d}x & : d_2 > a_2, \\ \sum_{d_1=a_1+1}^{9} \int_{(d_1+10^{-1}d_2)10^{\alpha}}^{(d_1+10^{-1}(d_2+1))10^{\alpha}} \frac{1}{x}\,\mathrm{d}x & : d_2 < a_2, \\ \int_{a\times 10^{\alpha}}^{(a_1+10^{-1}(d_2+1))10^{\alpha}} \frac{1}{x}\,\mathrm{d}x + \sum_{d_1=a_1+1}^{9} \int_{(d_1+10^{-1}d_2)10^{\alpha}}^{(d_1+10^{-1}(d_2+1))10^{\alpha}} \frac{1}{x}\,\mathrm{d}x & : d_2 = a_2, \end{cases}$$

$$= \begin{cases} \sum_{d_1=a_1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right) & : d_2 > a_2, \\ \sum_{d_1=a_1+1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right) & : d_2 < a_2, \\ \ln\left(\frac{a_1+(d_2+1)10^{-1}}{a}\right) + \sum_{d_1=a_1+1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right) & : d_2 = a_2. \end{cases}$$

$$\underline{T_3 := [10^{\beta}, b\times 10^{\beta}]}$$

This section of the proof follows the same logic as the previous section (for T_1). The result gives the coefficient λ_b :

$$\lambda_b = \begin{cases} \sum_{d_1=1}^{b_1-1} \ln\left(1+\frac{1}{10d_1+d_2}\right) & : d_2 > b_2, \\ \sum_{d_1=1}^{b_1} \ln\left(1+\frac{1}{10d_1+d_2}\right) & : d_2 < b_2, \\ \ln\left(\frac{b}{b_1+10^{-1}d_2}\right) + \sum_{d_1=1}^{b_1-1} \ln\left(1+\frac{1}{10d_1+d_2}\right) & : d_2 = b_2. \end{cases}$$

Finally to calculate the probability of an element of U to have a second digit d_2 we add all the integrals over the density function for all the subsets in T_1, T_2 and T_3 with d_2 as a second digit together and divide by the integral of the density function over U. Thus we have,

$$\begin{aligned} P[D_2 = d_2] &= \frac{1}{\int_U \frac{1}{x}\,\mathrm{d}x}\left[\int_{A_{d_2}} \frac{1}{x}\,\mathrm{d}x + \int_{B_{d_2}} \frac{1}{x}\,\mathrm{d}x + \int_{C_{d_2}} \frac{1}{x}\,\mathrm{d}x\right] \\ &= \frac{1}{\lambda_c}\left[(\beta-\alpha-1)\sum_{d_1=1}^{9} \ln\left(1+\frac{1}{10d_1+d_2}\right) + \lambda_a + \lambda_b\right]. \end{aligned} \tag{11}$$

We now make a change of logarithmic basis to base 10. Note that this produces a factor of $\frac{1}{\log(e)}$ in front of all terms in Equation 11 (including, crucially, λ_c in the denominator) which then cancel giving the final expression in Equation 6 in terms of log to the base 10.

Appendix C. Proof that Geometric Series are Benford

Define $m_a := \lfloor \log(a) \rfloor$ and $m_b := \lfloor \log(b) \rfloor$ as the magnitudes of the lower-bound a and upper-bound b respectively. Then $m_b - m_a = \lfloor \log(b) - \log(a) \rfloor = \lfloor d \rfloor = d$, by our assumption that d is an integer. $\lfloor . \rfloor$ is the floor function; $\lfloor x \rfloor = \max\{m \in \mathbb{Z} | m \leq x\}$.

We break the range $[a, b]$ into two disjoint sets $T_1 = [10^{m_a+1}, 10^{m_b})$ and $T_2 = [a, 10^{m_a+1}) \cup [10^{m_b}, b]$ such that $[a, b] = T_1 \cup T_2$. At least one of T_i is non empty. This is encoded in the condition $d \in \mathbb{N}$. Take N to be sufficiently large such that for all $y \in \{z \in \mathbb{N} | m_a < z \leq m_b\}$, there exists $i \in \mathbb{N}$ such that $x_i = 10^y$. This means that for each set T_1 and T_2 we can consider a geometric series starting at the highest lower bound and ending at the lowest upper bound of the set. We will consider T_1 and T_2 separately and show that each subset is Benford. Since the union of two disjoint Benford sets is Benford this will complete the proof.

Before we continue with the proof we present a short lemma.

Lemma: $x_n = ar^n$ is Benford in $T^{'} = [10^{m+1}, 10^{m+2})$ for $m \in \mathbb{N}$.

If $T^{'}$ is empty then there is nothing to prove. Assume that $T^{'}$ is non-empty. The elements in $T^{'}$ with a first digit α are in $[\alpha \cdot 10^{m+1}, (\alpha+1) \cdot 10^{m+1})$. The number of terms l in our geometric series with first digit α satisfies $\alpha \cdot 10^{m+1} \cdot r^l = (\alpha + 1) \cdot 10^{m+1}$, or rather,

$$l = \frac{1}{\log(r)} \log\left(\frac{\alpha+1}{\alpha}\right).$$

The number of terms L in $T^{'}$ is given by, $10^{m+1} \cdot r^L = 10^{m+2}$ or,

$$L = \frac{1}{\log(r)}.$$

Thus the proportion of terms with a first digit α in $T^{'}$ is given by,

$$\frac{l}{L} = \frac{\frac{1}{\log(r)} \log\left(\frac{\alpha+1}{\alpha}\right)}{\frac{1}{\log(r)}} = \log\left(\frac{\alpha+1}{\alpha}\right).$$

Thus ar^n is Benford in $T^{'} = [10^{m+1}, 10^{m+2})$.

Case one, T_1: If T_1 is empty, that is a and b have the same magnitude, then there is nothing to prove. Assume that T_1 is non-empty.

Note that $T_1 = \bigcup_{\lambda=1}^{m_b - m_a - 1} [10^{m_a+\lambda}, 10^{m_a+1+\lambda}]$. Since the union of disjoint Benford sets are Benford, by the previous Lemma, ar^n is Benford in T_1.

Case 2, T_2: If T_2 is empty then we are done. Assume T_2 is non-empty. The total number of terms $L = L_1 + L_2$ in T_2 satisfies,

$$a \cdot r^{L_1} = 10^{m_a+1}, \quad 10^{m_b} \cdot r^{L_2} = b,$$

or,

$$L_1 = \frac{1}{\log(r)} \log\left(\frac{10^{m_a+1}}{a}\right), \quad L_2 = \frac{1}{\log(r)} \log\left(\frac{b}{10^{m_b}}\right).$$

Thus:

$$\begin{aligned}
L &= L_1 + L_2 \\
&= \frac{1}{\log(r)}\left[\log\left(\frac{10^{m_a+1}}{a}\right) + \log\left(\frac{b}{10^{m_b}}\right)\right] \\
&= \frac{1}{\log(r)}\left[\log(10^{m_a+1}) - \log(10^{m_b}) + \log(b) - \log(a)\right] \\
&= \frac{1}{\log(r)}\left[m_a + 1 - m_b + d\right] \\
&= \frac{1}{\log(r)}\left[1 - d + d\right] = \frac{1}{\log(r)}.
\end{aligned}$$

Now the elements of T_2 with a first digit β are members of,

$$B = \begin{cases} [\beta \cdot 10^{m_a}, (\beta+1)\cdot 10^{m_a}] & : a \leq \beta \cdot 10^{m_a} \\ [\beta \cdot 10^{m_b}, (\beta+1)\cdot 10^{m_b}] & : b \geq \beta \cdot 10^{m_b} \\ [a, (\beta+1)\cdot 10^{m_a}) \cup (\beta \cdot 10^{m_b}, b] & : a > \beta \cdot 10^{m_a}, b < (\beta+1)\cdot 10^{m_b} \end{cases}$$

The first two cases for B are identical to the proof of case 1. Thus take $a > \beta \cdot 10^{m_a}, b < (\beta+1)\cdot 10^{m_b}$. The number of elements $l = l_1 + l_2$ with a first digit of β satisfies,

$$a \cdot r^{l_1} = (\beta+1)\cdot 10^{m_a}, \quad \beta \cdot 10^{m_b} \cdot r^{l_2} = b.$$

That is,

$$l_1 = \frac{1}{\log(r)}\log\left(\frac{(\beta+1)10^{m_a}}{a}\right), \quad l_2 = \frac{1}{\log(r)}\log\left(\frac{b}{\beta \cdot 10^{m_b}}\right).$$

So,

$$\begin{aligned}
l &= l_1 + l_2 \\
&= \frac{1}{\log(r)}\left[\log\left(\frac{(\beta+1)10^{m_a}}{a}\right) + \log\left(\frac{b}{\beta \cdot 10^{m_b}}\right)\right] \\
&= \frac{1}{\log(r)}\left[\log\left(\frac{\beta+1}{\beta}\right) + \log\left(\frac{b \cdot 10^{m_a}}{a \cdot 10^{m_b}}\right)\right] \\
&= \frac{1}{\log(r)}\left[\log\left(\frac{\beta+1}{\beta}\right) + \log(10^{m_a}) - \log(10^{m_b}) + \log(b) - \log(a)\right] \\
&= \frac{1}{\log(r)}\log\left(\frac{\beta+1}{\beta}\right).
\end{aligned}$$

Dividing by the total number of terms in T_2 completes the proof.

Remark: Our assumption that N is sufficiently large such that for all $y \in \{z \in \mathbb{N} | m_a < z \leq m_b\}$, there exists $i \in \mathbb{N}$ such that $x_i = 10^y$ means that we can consider the regions T_1 and T_2 separately since there exists a term x_i on the boundaries of each set. If N was

small we would need to consider, for example, $T_1' = [10^{m_a+1} + \epsilon(N), 10^{m_b} - \delta(N)]$ with $x_j = 10^{m_a+1} + \epsilon(N) \in T_1'$ and $x_{j-1} \leq x_j \notin T_1'$ for some $j \in \mathbb{N}$. This adds extra complications to the proof for N small. However,

$$0 \leq \epsilon(N) \leq x_j - x_{j-1} = \left(1 - \frac{1}{r}\right) x_i.$$

Now $r := 10^{\frac{d}{N}}$ with d fixed. As $N \to +\infty$, $\frac{d}{N} \to 0$ such that $r \to 1$.

$$\lim_{N\to\infty} 0 = 0 \leq \lim_{N\to\infty} \epsilon(N) \leq \lim_{N\to\infty} \left(1 - \frac{1}{r}\right) x_i = \lim_{r\to 1} \left(1 - \frac{1}{r}\right) ar^i = 0.$$

A similar result holds for $\delta(N)$. Thus our assumption is justified in the limit $N \to +\infty$.

Appendix D. A Note On Significance Levels

P values are applied to our test statistics, which are a measure of the conformity of data with BL. For example, a *p* value of 0.05 corresponds to the value of our test statistic (either d^* or A^2), for which there is a 5% probability of collecting a result in which the test statistic is larger than this value, given that the set we are testing is Benford. In the event of this occurring, we would say that we can reject the result at the 0.05 significance level.

P values (and more generally test statistics) offer a method to quantify the incompatibility between a set of test data and a proposed statistical model for the data. In our context, this model is Benford's law. However, in isolation, a *p* value does not provide a reliable measure of the correctness of our hypothesis. A statement of the statistical significance of a result without context or further analysis provides little insight and certainly cannot form the basis of a rigorous conclusion. Therefore, in this investigation, data analysis did not end with a calculation of a *p* value. *P* values are taken as an indication as to what extent data may be incompatible with BL before further enquiry as to why this is the case is undertaken.

Appendix E. Software

Throughout the project, several tools were developed to assist with Benford analysis, figure production and data acquisition. This software was developed in Python and C. This toolset has multiple capabilities:

- Synthetic Benford set generation, including bruteforce and geometric series techniques. This allows for finer control over the degree of conformity with Benford's law.
- Introduce arbitrary deviations into Benford sets to observe changes in conformity. In particular, the introduction of rounding behaviour and digit pronunciation (exaggeration) for first and second digits.
- Traditional automated Benford tests, including the first, second and first-second digit tests. Each test includes plots of the data, either histograms or discrete heatmaps, and various test statistics assessing conformity.

- Finite range Benford's law including a generalised finite range implementation using synthetic Benford sets.
- Other general statistics and miscellaneous tools.

```
def main(mode):
    #Process mode of analysis
    try:
        str(mode)
    except:
        print("[Fatal Error] Cannot process mode", str(mode), ". Please enter a valid mode of analysis.")
        usage()
        exit()

    #Import data from argv[1]
    filename = sys.argv[1]
    data, lowerbound, upperbound = input_numbers(filename)
    print("[Debug] Starting First Digit Analysis")

    # Process the mode of analysis e.g. first digit analysis = 1
    if mode in ['1', '12', '12h', '12hn', '23', '23h', '23hn', '2']:
        data_raw, benford_raw, data_percent, benford_percent, z_statistic, von_mises_statistic, d_star_statistic = digit_test(data, mode)

        bins_to_plot = []

        if mode == '1':
            bins_to_plot = np.arange(9)
        elif mode == '2':
            bins_to_plot = np.arange(0, 10, 1)
        elif mode in ['12', '12h', '12hn']:
            bins_to_plot = np.arange(10, 100, 1)
        elif mode in ['23', '23hn', '23h']:
            for x in range(0, 100):
                bins_to_plot.append(x)

    elif mode == 'f1':
        data_raw, benford_raw, data_percent, benford_percent, z_statistic, von_mises_statistic, d_star_statistic, data_size = benford_finite_range(data, mode, lowerbound, upperbound)
        bins_to_plot = np.arange(9)

```

```
    elif mode == 'f2':
        data_raw, benford_raw, data_percent, benford_percent, z_statistic, von_mises_statistic, d_star_statistic, data_size = benford_finite_range(data, mode, lowerbound, upperbound)
        bins_to_plot = np.arange(10)

    # Compute sanitised dataset size
    data_size = 0
    for x in data_raw:
        data_size += x

    #Output results
    print("[Debug] Analysis complete. Outputing results.")
    print("\n\n###--- Analysis for", filename, "---###\n")
    output_digit_test(data_raw, benford_raw, z_statistic, mode)

    # Test statistic Output
    print("Cramer-von Mises test: {} {} {}".format(von_mises_statistic[0], von_mises_statistic[1],von_mises_statistic[2]))
    print("d* test: {}, {}{}".format(d_star_statistic[0], d_star_statistic[1], d_star_statistic[2]))
    #Legend significance levels
    print("\n * significant at the .05 level\n** significant at the .01 level\n")

    # Create plots of the data
    print("[Debug] Generating Plot of the data.")
    if mode in ['1','f1','f2','2','12']:
        plot_bar_chart(bins_to_plot, data_raw, benford_raw, data_size, von_mises_statistic[2], d_star_statistic[1], mode)

    elif mode in ['12h', '12hn', '23h', '23hn']:
        if 'hn' in mode:
            norm_residuals, null, null = compute_normalised_residuals(data_raw, benford_raw)
            plot_heat_map(norm_residuals, benford_percent, mode)
        else:
            plot_heat_map(data_raw, benford_raw, mode)

    print("[Debug] Output complete. Exiting.")
```

```
68      exit()
```

Listing 1: The main function of benford.py designed to import a list of numerical data, apply BL and produce figures of the results.

Listing 1 shows an example of code written for this project for benford.py a tool designed to automatically apply a Benford analysis to a set of data and produce figures of the results. The entire project is available at https://github.com/bluehood/benford_analysis.