University of Victoria
Department of Economics

Econometrics Working Paper EWP0505

ISSN 1485-6441

BENFORD'S LAW AND NATURALLY OCCURRING PRICES IN CERTAIN ebaY AUCTIONS*

David E. Giles

Department of Economics, University of Victoria
Victoria, B.C., Canada V8W 2Y2

May, 2005

Abstract

We show that certain the winning bids for certain ebaY auctions obey Benford's Law. One implication of this is that it is unlikely that these bids are subject to collusion among bidders, or "shilling" on the part of sellers. Parenthetically, we also show that numbers from the naturally occurring Fibonacci and Lucas sequences also obey Benford's Law.

Keywords: Benford's Law; auction prices; football; Kuiper test; Watson test

JEL Classifications: C12; C14; C16; D44

Author Contact:

David Giles, Dept. of Economics, University of Victoria, P.O. Box 1700, STN CSC, Victoria, B.C., Canada V8W 2Y2; e-mail: dgiles@uvic.ca; FAX: (250) 721-6214

1. Introduction

Consider the daily return on a stock on the NYSE. What is the probability that the first significant digit in the return is a "1", or a "9", for example? Contrary to what one might think, the answer is *not* (1/9) in either case. Newcomb (1881) appears to have been the first to notice that many naturally occurring decimal numbers have a first significant digit that is "1" with probability approximately 31%, while the probability of this digit being "9" is less than 5%. His examination of this phenomenon was motivated by observing that the first few pages of books of logarithm tables were generally dirtier (as a result of being thumbed more often) than were later pages. His results were not widely appreciated, and were unwittingly "re-discovered" by Benford (1938).

Benford provided numerous examples of data that obey a particular non-uniform distribution with respect to first significant digits, and the phenomenon became known as "Benford's Law". His examples included the areas of lakes, the lengths of rivers, and the molecular weights of compounds. Subsequently, a number of studies (*e.g.*, Pinkham, 1961; Raimi, 1969, 1976; Ley, 1996; and de Ceuster *et al*., 1998) have confirmed that Benford's Law holds for a very large range of phenomena, including certain economic data. The purpose of this note is to examine whether Benford's Law holds in the context of certain ebaY auction prices. The importance of the results that we obtain is that they can be used as a test of collusion in that market place

In the next section we discuss Benford's Law and illustrate its application to two important naturally occurring data-sets. The main results of the paper are given in section 3, where we consider an analysis of ebaY auction prices; and section 4 concludes.

2. Benford's Law and Natural Phenomena

Let d denote the first significant digit in a random decimal number, N. So, for example, if N = 194.23 then d = 1; and if N = 0.00476 then d = 4. Benford's Law dictates that

$$\Pr.[d = k] = \log_{10}[1 + (1/k)] \quad ; \quad k = 1, \ldots.., 9.$$

This result can be established *via* combinatorial arguments (*e.g.*, Cohen, 1976), and straightforward statistical derivations of this result are also available (*e.g.*, Hill, 1995). Further motivation is provided by De Ceuster *et al*. (1998), for example.

Studies that have drawn attention to the relevance of Benford's Law to economic data include those of Ley and Varian (1994); Ley (1996), who investigated one-day returns on the Dow-Jones and Standard and Poor indices for U.S. stocks; and Pietronero *et al*. (2001), who considered prices on the Madrid, Vienna and Zurich stock exchanges. One application of Benford's Law that is relevant to the present paper arises in the context of auditing tax data. Fraudulent activity tends to result in figures that are "too uniform", relative to Benford's Law, as was discussed by Nigrini (1996). This application of Benford's Law is surveyed by Geyer and Williamson (2004).

To lay the basis for our subsequent analysis, let us consider two specific examples of Benford's Law. As this law apparently occurs widely in the context of natural phenomena, one might speculate that it will be satisfied by the well-known Fibonacci series, {1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, ……}, where the terms satisfy the recurrence relationship: $F(n+1) = F(n) + F(n-1)$, and where the irrational number

$$\phi = Limit_{n\to\infty}[F(n+1)/F(n)] \approx 1.61803....$$

is the co-called 'golden section' that occurs widely in nature. Figure 1 compares the empirical frequency distribution of first digits for the first 1,476 Fibonacci numbers with the corresponding distribution predicted by Benford's Law.[1] Recently, Canessa (2003) has proposed a general statistical thermodynamic theory that explains, *inter alia*, why Fibonacci sequences should obey Benford's Law.[2] The visual similarity between the two frequency distributions in Figure 1 is apparent, and a formal statistical test of their equality (and hence of Canessa's theory) is discussed in the next section.

The series of Lucas (1878) numbers is {1, 3, 4, 7, 11, 18, 29, 47, 76, 123, …. }, where the terms follow the same recurrence relationship as for the Fibonacci sequence. Interestingly,

$$Limit_{n\to\infty}[L(n+1)/L(n)] = \phi \ ,$$

and in fact this is true for any sequence of integers of the form {1, x, (1+x), [(1+x) +x], …….}. There are numerous relationships between Fibonacci and Lucas numbers. While the Lucas numbers are less well known than those of Fibonacci's sequence, they occur extensively in nature, such as in lunar eclipse cycles, and in the properties of electrical circuits, for example.[4] Figure 1 also compares the empirical frequency distribution of first significant digits for the first 1,475 Lucas numbers with the distribution predicted by Benford's Law.[5] The extent to which the Lucas sequence accords with this law is even more striking than for the Fibonacci sequence, but this result does not appear to have been observed previously in the literature.

3. Analysis of Some ebaY Auction Prices

The question we wish to address is whether or not the closing prices of successful auctions on the ebaY internet auction site are naturally occurring numbers. If they are not, then this may be the result of collusion among bidders and/or interference ("shilling") on the part of the seller. In our investigation we have considered all of the successful auctions for tickets for professional football games in the "event tickets" category on ebaY, for the period 25 November to 3 December, 2004.[6] Auctions that ended with the "Buy-it-Now" option, and all Dutch auctions, were excluded from the data. This resulted in 1,161 successful auction completions.

By not considering college football games we avoided the distortions that might arise in those cases where there is a "two-tier" ticketing system. By this we mean the situation where there are "student only" tickets and "regular" tickets. Information was recorded for each winning bid. All prices are in U.S. dollars as this is the official ebaY currency.[7] Figure 2 compares the empirical distributions of the first significant digits of prices per lot and prices per ticket with the distribution predicted by Benford's Law. Although these results are less striking than those in Figure 1, they still suggest that the auction prices accord with Benford's Law. Of course, a formal test of the equivalence of the distributions is required, and we now turn to this issue.

Typically, chi-square goodness-of-fit tests have been used in association with Benford's Law. Various other nonparametric tests, such as "integrated deviations" tests of the Cramér-von Mises type, or "supremum tests" of the Kolmogorov-Smirnov type, could be considered here. These tests involve comparing the empirical distribution function of the data with the hypothesized distribution, and they rely on the Glivenko-Cantelli Theorem for their justification.[8] However, none of these tests are actually appropriate here. It is important to note that our data have a special characteristic that must be taken into account when considering this testing problem – they follow a circular distribution. This is clear if we consider a pair of prices such as \$99.99 and \$100, say, whose first significant digits are "9" and "1" respectively. The values "9" and "1" are as far apart as is possible within the set of first significant digits, yet the associated prices are as close as is possible in an ebaY auction (where bids are in dollars and cents). Recently, Giles (2005) has recognized the relevance of distributional circularity in the dating of business cycles.

The phenomenon of circularity is common with many types of statistical data, and there is an extensive literature dealing with testing for the underlying distribution in such cases (*e.g.*, Fisher, 1993). One well-known example that meets our needs here is Kuiper's (1959) V_N test, which is a modified Kolmogorov-Smirnov test. The crucial point is that such tests are invariant to the location of the origin and the scale on the circle, whereas conventional such tests are not.[9] Suppose that the empirical distribution function for a sample of size N is $F_N(x)$, and that the null population distribution function is $F_0(x)$. Then, Kuiper's basic test statistic is given by:

$$V_N = D_N^+ + D_N^- \quad ,$$

where the two components are the one-sided Kolmogorov-Smirnov statistics:

$$D_N^+ = \sup_{-\infty<x<\infty}. [F_N(x) - F_0(x)]$$
$$D_N^- = \sup_{-\infty<x<\infty}. [F_0(x) - F_N(x)] \quad .$$

Another important property of Kuiper's test is that the null distribution of the test statistic is invariant to the hypothesized distribution, regardless of the sample size. Critical values for the null distribution of $V_N^* = V_N[N^{1/2} + 0.155 + 0.24N^{-1/2}]$ are tabulated for by Stephens (1970).[10] These critical values are provided together with our test results in Table 1. As can be seen there, the inferences that were drawn visually from Figures 1 and 2 are very strongly supported by the formal tests – we cannot reject the hypothesis that the data for the Fibonacci and Lucas sequences, and for the ebaY auction prices obey Benford's Law.

4. Conclusions

In this paper we have tested Benford's Law against a small set of winning prices for ebaY auctions for pro-football tickets, and shown that this law cannot be rejected. Had there been collusion among bidders, or "shilling" on the part of sellers in these auctions, it is unlikely that this positive result would have been achieved. This demonstrates the usefulness of Benford's Law in a different "auditing" context than has been considered previously, and suggests that further research along these lines would be valuable. It also sets the scene for work in progress that investigates psychological barriers in ebaY auctions.

Peripherally, we have also confirmed Canessa's (2003) result that numbers from the Fibonacci sequence obey Benford's Law, and we have shown that this also holds for numbers from the important and closely related Lucas sequence.

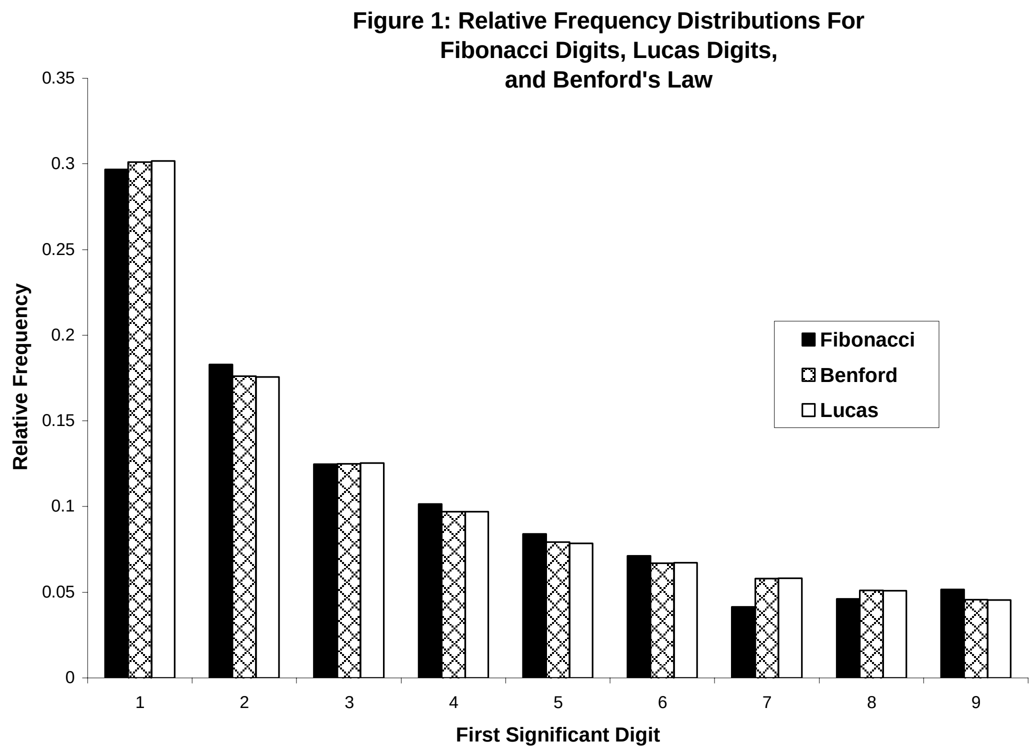
Figure 1: Relative Frequency Distributions For
Fibonacci Digits, Lucas Digits,
and Benford's Law
Relative Frequency
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
First Significant Digit
Fibonacci
Benford
Lucas

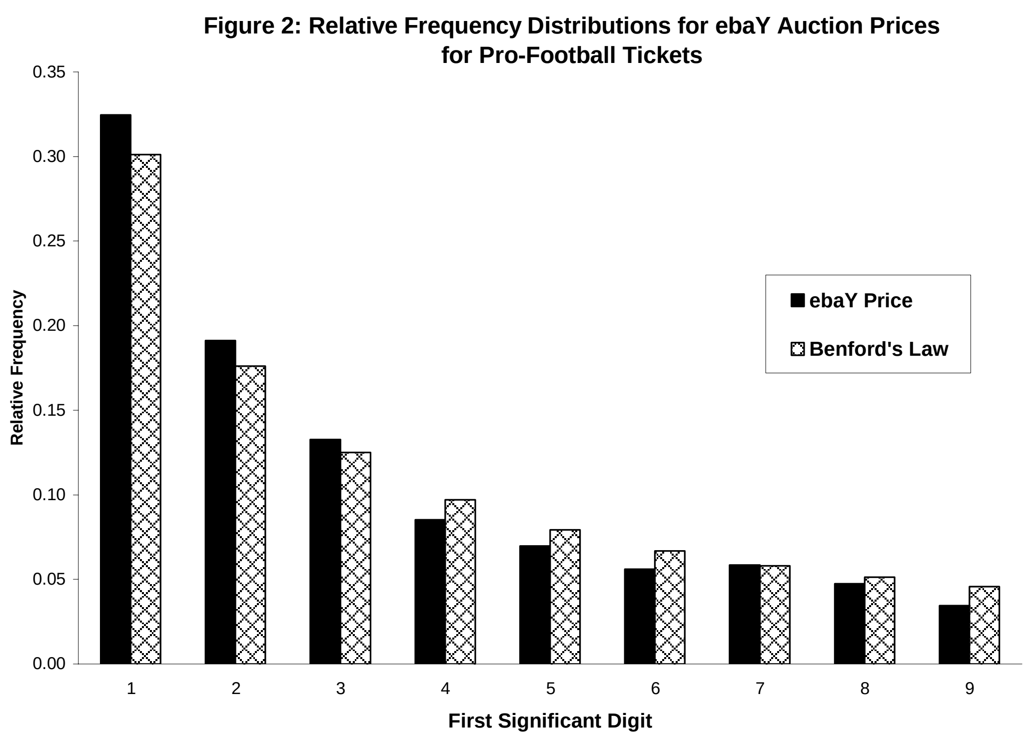
Figure 2: Relative Frequency Distributions for ebaY Auction Prices
for Pro-Football Tickets
Relative Frequency
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
6
7
8
9
First Significant Digit
ebaY Price
Benford's Law

Table 1

Kuiper Test Results*

	Fibonacci Numbers	Lucas Numbers	ebaY Prices
N	1,476	1,475	1,161
$\mathbf{V_N}$	0.0217	0.0007	0.0465
$\mathbf{V^*_N}$	0.8370	0.0270	1.5919

* 10%, 5% and 1% critical values for V^*_N are 1.620, 1.747 and 2.001, regardless of the sample size.

References

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

Canessa, E., 2003. Theory of analogous force in number sets. Physica A 328, 44-52.

Cohen, D., 1976. An explanation of the first digit phenomenon. Journal of Combinatorial Theory 20, 367-370.

De Ceuster, M. K. J., G. Dhaene and T. Schatteman, 1998. On the hypothesis of psychological barriers in stock markets and Benford's Law. Journal of Empirical Finance 5, 263-279.

Durbin, J., 1973. Distribution Theory for Tests Based on the Sample Distribution Function. SIAM, Philadelphia PA.

Fisher, N.I., 1993. Statistical Analysis of Circular Data. Cambridge University Press, Cambridge.

Geyer, C. L. and P. P. Williamson, 2004. Detecting fraud in data sets using Benford's Law. Communications in Statistics B 33, 229-246.

Giles, D. E., 2005. Testing for a Santa Claus effect in business cycles. Economics Letters, in press.

Gottwald, G. A. and M. Nicol, 2002. On the nature of Benford's Law. Physica A 303, 387-396.

Hill, T. P., 1995. A statistical derivation of the first digit phenomenon. Statistical Science 10, 354-363,

Kuiper, N. H., 1959. Alternative proof of a theorem of Birnhaum and Pyke. Annals of Mathematical Statistics 30, 251-252.

Ley, E., 1996. On the peculiar distribution of the U.S. stock indexes' digits. American Statistician 50, 311-313.

Ley, E. and H. Varian, 1994. Are there psychological barriers in the Dow-Jones index? Applied Financial Economics 4, 217-224.

Lucas, E., 1878. The theory of simply periodic numerical functions. American Journal of Mathematics 1, 184-240, 289-321.

McMinn, D., 2003. Fibonacci – Lucas numbers and moon-sun cycles. http://www.davidmcminn.com/pages/fiblucas.htm

Newcomb, S., 1881. Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics 4, 39-40.

Nigrini, M., 1996. A taxpayer compliance application of Benford's Law. Journal of the American Taxation Association 18, 72-91.

Pietronero, L., E. Tosatti, V. Tosatti, and A. Vespignani, 2001. Explaining the uneven distribution of numbers in nature: the laws of Benford and Zipf. Physica A 293, 297-304.

Pinkham, R. S., 1961. On the distribution of first significant digits. Annals of Mathematical Statistics 32, 1223-1230.

Raimi, R. A., 1969. The peculiar distribution of first digits. Scientific American, 109-119.

Raimi, R. A., 1976.The first digit problem. American Mathematical Monthly 83, 521-538.

Scott, L. H., undated. Electrical equivalence of discrete cascaded T sections, electrical transmission lines, and Fibonacci and Lucas numbers. http://www.lhsorg.org/

Stephens, M. A., 1965. The goodness-of-fit statistic, V_N : Distribution and significance points. Biometrika 52, 309-321.

Footnotes

* I am very grateful to Sophie Hesling for her excellent research assistance.

1. This number of observations is dictated by machine precision. The last number in the sequence is $F(1{,}476) = 0.1306989E+309$.

2. Pietronero *et al*. (2001) provide an explanation for Benford's Law in terms of multiplicative dynamical processes, and derive a generalization in terms of a power law. Gottwald and Nicol (2002) prove that systems satisfying Benford's Law do not have to be random or chaotic – they may be deterministic, and are more related to the (near) multiplicative nature of the process than to its chaoticity.

3. For example, for $n > 1$, every Lucas number is the sum of two Fibonacci numbers, and for all n, $L(n) = [F(n+2) - F(n-2)]$ and $[L(n-1)L(n+1) + F(n-1)F(n+1)] = 6[F(n)]^2$. Moreover, $Limit_{n\to\infty}[L(n)] = \phi^n$, and this relationship is approximated well even when 'n' is quite small. For example, for n = 8, 9, 10 we have L = 47, 76, 123; and ϕ^n = 46.971, 75.999 and 123.046.

4. See McMinn (2003), and Scott (undated). Also, in a sequence of 'n' equal resistors, connected alternately in series and in parallel, then $Limit_{n\to\infty}$ [Total resistance] = ϕ. If the wiring is alternately in parallel and in sequence, $Limit_{n\to\infty}$ [Total resistance] = $1/\phi$.

5. The last number in the sequence is $L(1{,}475) = 0.111630E+309$.

6. More specifically, the time-span is from 00:00:01 PST on 25 November, to 24:00:00 PST on 3 December, 2004.

7. Strictly, it is possible for other currencies to be used, but there were no such examples in our population. In any case, Benford's Law is scale and base invariant, so the choice of (a common) currency does not affect our results.

8. For a detailed technical discussion, see Durbin (1973), for example. The Glivenko-Cantelli Theorem implies that, under quite mild conditions, the empirical distribution function converges uniformly and with probability one to the population distribution function.

9. Note that if the data are circular, as is the case with the digits "1" through "9", the distribution of the test statistic should not depend on whether we begin counting at "1" and end at "9" (say), or alternatively begin counting at "6" and end at "5" (say).

10. See Stephens (197, p.118). The purpose of transforming from V_N to V^*_N is to create a statistic whose null distribution is independent of the sample size. This is especially helpful for large "N".