On the Peculiar Distribution of the U.S. Stock Indeces' Digits

Eduardo Ley
Resources for the Future, Washington DC

Version: November 29, 1994

Abstract. Recent research has focused on studying the patterns in the digits of closely followed stock market indeces (see, *e.g.*, Ley and Varian (1994) and Koedijk and Stork (1994)). In this paper, we find that the series of one-day returns on the Dow-Jones Industrial Average Index (DJIA) and the Standard and Poor's Index (S&P) reasonably agrees with Benford's law and, therefore, belongs to the family of *anomalous* or *outlaw* numbers.

Keywords. Dow-Jones Index, Standard and Poor's, Benford's Law

JEL Classification System. G14

Address. Eduardo Ley, Resources for the Future, 1616 P Street NW, Washington DC 20036. Email: ley@rff.org.

1. Benford's Law

Benford (1938) presented a general *law of anomalous numbers:* the probability that a random decimal begins with the digit i, $f(i)$, is given by

$$f(i) = \log\left(\frac{i+1}{i}\right) \qquad \text{for } i = 1, \ldots, 9;$$

where $\log(x)$ represents logarithm of x to the base 10. Benford arrived at this law —that he believed was a general law of nature related to the general logarithmic character of natural phenomena— after observing that tables of logarithms in libraries tend to be dirtier at the beginning than at the end. This couldn't mean anything but that people have more occasion to look up numbers beginning with 1 than with 2, and so on up to 9. He investigated 20 tables of numbers of such different nature as the surface and area of 353 rivers, the street addresses of the first 342 persons listed in *American Men of Science*, the specific heat and molecular weight of thousands of chemical compounds, and the numbers of consecutive front page news items of a newspaper.

Benford found that the first digit '1' appeared with frequency .306 while the first digit '9' appeared with a frequency of only .047 —if every number appeared equally often all the observed frequencies should approximate $\frac{1}{9} = .111$.[1] Benford's law turns out to be the only distribution of first digits which is scale invariant in the sense that multiplication by a constant doesn't change the distribution of first digits, Pinkham (1961). Some tables of numbers are in better agreement with Benford's law than others. Systematic tables —*e.g.*, a list of square roots— or closely knit tables —*e.g.*, specific heats of chemical compounds— do

I thank Mark Steel and Hal Varian for their comments and suggestions.

[1] Note that, by definition, 0 is not a possible outcome. We're looking at the first digit regardless of decimal-point position, which does not affect the place of an entry in a logarithm table—*i.e.*, $\log(x/10^k) = \log(x) - k$.

not correspond well with Benford's law. On the other hand, tables of *anomalous numbers* —*i.e.*, "those outlaw numbers that are without known relationship rather than to those that individually follow an orderly course" —do agree with Benford's law quite well. See Raimi (1969) for a fascinating discussion.

Recent research has focused on studying the patterns in the digits of closely followed stock market indeces (see, *e.g.*, Ley and Varian (1994) and Koedijk and Stork (1994)). Here, we shall use Benford's law to see how *anomalous* are the series of one-day returns on the two most important U.S. stock indences: the Dow-Jones Industrial Average Index (DJIA) and the Standard and Poor's Index (S&P).[2]

2. Statistical Analysis

Let p_t be the closing value of the stock index at time t. The one-day return on the index, r_t, is defined as

$$r_t = \frac{\ln p_{t+1} - \ln p_t}{d_t} \times 100,$$

where d_t is the number of days between trading days t and $t+1$—*e.g.*, if t corresponds to a Friday and Monday is the next trading day, then $d_t = 3$. Since $p_{t+1} = p_t \exp\{\ln(p_{t+1}/p_t)\}$ we have that r_t is the continuous-time rate of return for the period between t and $t+1$. Since we want to have periods of equal length, we divide by the number of days between trading days —*i.e.*, whenever we have a holiday or weekend the computed r_t is an *average* rate of return, which then gets assigned to all days between t and $t+1$.[3]

[2] For the DJIA we will be using data from January 1900 to June 1993; for the S&P's and we will be using data from January 1926 to June 1993. (The data are available via anonymous ftp from econ.lsa.umich.edu.)

[3] For the DJIA (1900–1993), we have that 81.15% of the cases have $d_t = 1$, 10.80% have

Looking only at the first significant digit of r_t we then obtain a vector $x = (x_1, x_2, \ldots, x_9)$ where x_i is the number of times that r_t's first significant digit is $i \in \{1, 2, \ldots, 9\}$. We can think of X as being generated by a Multinomial distribution with parameter θ (9×1), thus

$$f(x|\theta) = \frac{(\sum_{j=1}^{9} x_j)!}{\prod_{j=1}^{9} x_j!} \prod_{j=1}^{9} \theta_j^{x}.$$

On the basis of prior ignorance, we could assume that each digit was equally likely and postulate a prior density for θ with mean $(\frac{1}{9}, \frac{1}{9}, \ldots, \frac{1}{9})$. The natural conjugate prior is a Dirichlet density with parameter $(\alpha, \alpha, \ldots, \alpha)$; thus

$$g(\theta) = \frac{\Gamma(9\alpha)}{\Gamma(\alpha)^9} \prod_{j=1}^{9} \theta_j^{\alpha - 1}.$$

The bigger α, the bigger the confidence on our prior knowledge —*i.e.*, the smaller the dispersion around θ's postulated prior mean.

The posterior distribution of θ is given by a Dirichlet with parameter $(\alpha + x_1, \alpha + x_2, \ldots, \alpha + x_3)$ —see, *e.g.*, DeGroot (1971) for details— that is we have that

$$h(\theta|x) = \frac{\Gamma(9\alpha + \sum_{j=1}^{9} x_j)}{\prod_{j=1}^{9} \Gamma(\alpha + x_j)} \prod_{j=1}^{9} \theta_j^{\alpha + x_j - 1}.$$

$d_t = 2$, 7.20% have $d_t = 3$, and less than 1% have $d_t \geq 4$. For the S&P's (1926–1993), we have that 78.19% of the cases have $d_t = 1$, 7.67% have $d_t = 2$, 12.06% have $d_t = 3$, and 2.09% have $d_t \geq 4$. Two things are worth noting. First, until June of 1952 the NYSE operated on Saturdays. Second, until January 1928 the S&P's was computed on a weekly basis only.

2.1. Posterior Results

The data information is very strong and for reasonable values of the single hyper-parameter, α, —*i.e.*, those leading to reasonably large prior standard deviations— the results will be rather insensitive to the postulated value of α. Therefore, we only show posterior results for $\alpha = 1$.[4] As can be seen from table 1 or figure 1, the posterior mean —which, given the large number of observations corresponds to the observed frequencies— gets very close to Benford's theoretical frequencies.

Table 1. Prior and Posterior Mean *vs* Benford's Law

i	Prior	Posterior		Benford's Law
		DJIA(1900–93)	S&P's (1926–93)	
1	11.11%	28.94%	29.17%	30.10%
2	11.11%	16.78%	16.96%	17.61%
3	11.11%	12.38%	13.42%	12.49%
4	11.11%	9.99%	9.87%	9.69%
5	11.11%	8.48%	7.76%	7.92%
6	11.11%	7.23%	7.13%	6.69%
7	11.11%	6.15%	5.60%	5.80%
8	11.11%	5.32%	5.36%	5.12%
9	11.11%	4.72%	4.73%	4.58%
$\sum \lvert\hat{\theta}_i - f(i)\rvert$		4.22%	3.88%	
$\chi^2(8)$		71.98	43.46	
N		33,804	24,126	

As it can be seen in Figure 1, the series of one-day returns on the DJIA

[4] With $\alpha_1 = 1$ for all i, the prior standard deviation of θ_i is 0.0994 for all i. For the DJIA, the posterior standard deviations of θ_i are 0.0025, 0.0020, 0.0018, 0.0016, 0.0015 0.0014, 0.0013, 0.0012, and 0.0012 for $i = 1, \ldots, 9$. For the S&P's, they are: 0.0029 0.0024, 0.0022, 0.0019, 0.0017, 0.0017, 0.0015, 0.0015, and 0.0014. Making the α_i's smaller results in bigger prior standard deviations giving more weight to the data information.

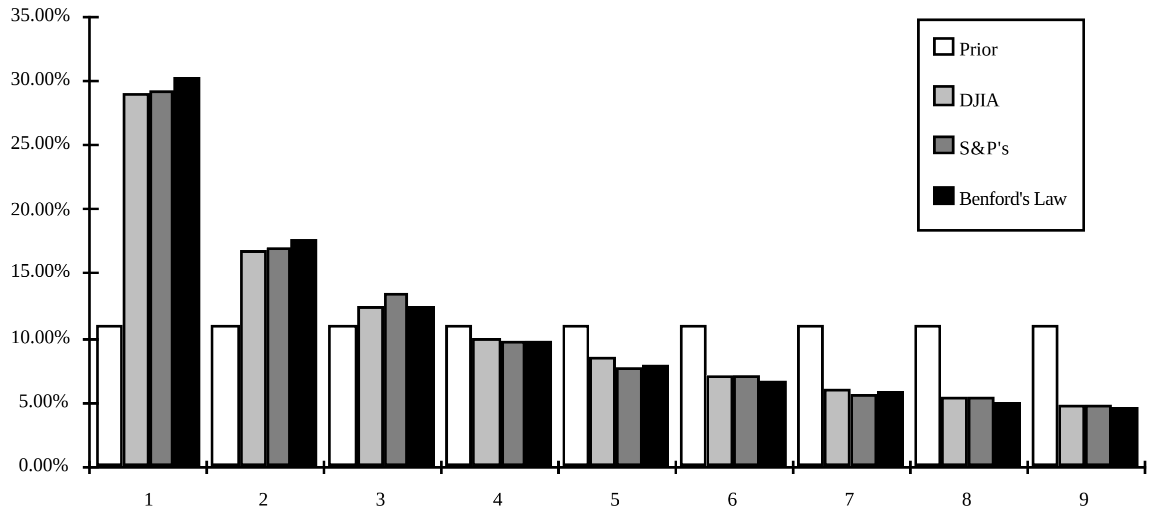

Figure 1. Prior and Posterior Means *vs* Benford's Law

and S&P's indeces, can both be reasonably classified to belong to the family of "outlawed numbers without known relationship." The observed frequencies roughly agree with the theoretical frequencies given by Benford's law. The sum of absolute deviations from the theoretical frequencies are 3.88 for the S&P's and 4.22 for the DJIA, which would have ranked these series 3^{rd} and 4^{th} in goodness of fit among the 22 series consisting of themselves and Benford's 20 original series, between "Pressure Lost, Air Flow," and "H.P. Lost in Air Flow."

However, if we performed the usual likelihood-ratio test or the chi-square test of goodness of fit, we would reject the null because of the huge power that any of these tests would have, given the large number of observations —*i.e.*, the classical acceptance region shrinks with sample size, given a significance level. The chi-square statistic, $\chi^2(8) = N \times \sum_{i=1}^{9} (\hat{\theta}_i - f(i))^2 / f(i)$ where the hat indicates observed frequencies, is shown on table 1. If one takes models as mere approximations to reality, not as perfect data-generating mechanisms, then this can only be viewed as a weakness of the Neyman-Pearson theory —see, *e.g.*, Berger (1985), page 20, or Leamer (1983). In particular, if we had presented table 1 for only the last ten years of data, 1983–1993, the chi-square statistics would have been 12.93 (DJIA) and 13.12 (S&P's) *vs* a value of 15.51 for the 95^{th} quantile of a chi-square distribution with 8 d.f.; so the Benford hypothesis would not be rejected on the basis of this test.

2.2. Small Changes and Big Changes

We could interpret Benford's Law in the present context as merely stating that small movements in the DJIA are more likely than big ones. That is, if 0.1 is more likely a value for $|r|$ than 0.2, and, in turn, 0.2 is more likely than 0.3, ..., then we could observe something similar to Benford's Law. If small changes happen more often than big changes, one could ask whether the same holds uniformly within changes of the same *magnitude*. For example, does Benford Law hold when we only look at numbers that fall within the interval $[10^k, 10^{k+1})$ for different values of k? Table 2 contains the relative frequencies of first digits for three different intervals of $|r|$.

As we can see in table 2, the relative frequencies are closer to Benford's Law when the (absolute) movement in the DJIA is between 0.1% and 1%–which happens about $\frac{2}{3}$ of the time. The other columns do not resemble Benford's Law at all. Roughly, a particular first digit is more probable the closer the implied movement is to 0.1%. On aggregate, over all groups, however, the observed frequencies are remarkably close to Benford's Law.

3. Concluding Remarks

Recent research has focused on studying the patterns in the digits of closely followed stock market indeces (see, *e.g.*, Ley and Varian (1994) and Koedijk and

Table 2. Relative Frequency of First Digits by Intervals

	DJIA(1900–1993)			S&P's (1926–1993)		
i	$\|r\| < .1$	$.1 \leq \|r\| < 1$	$\|r\| \geq 1$	$\|r\| < .1$	$.1 \leq \|r\| < 1$	$\|r\| \geq 1$
1	10.05%	23.29%	77.34%	8.05%	26.13%	74.91%
2	10.95%	18.88%	14.47%	9.88%	19.53%	15.48%
3	12.09%	14.33%	4.02%	15.67%	14.42%	5.36%
4	10.27%	11.68%	2.06%	12.16%	10.78%	2.17%
5	11.55%	9.30%	1.09%	10.30%	8.41%	0.89%
6	11.88%	7.46%	0.54%	12.48%	6.83%	0.54%
7	11.57%	6.00%	0.18%	9.05%	5.62%	0.36%
8	10.69%	5.00%	0.22%	11.84%	4.40%	0.21%
9	10.94%	4.06%	0.08%	10.56%	3.87%	0.09%
N	6,162	22,598	5,044	5,008	15,758	3,360
$\%N$	18.23%	66.85%	14.92%	20.76%	65.32%	13.93%

Stork (1994)). We find that the series of one-day returns on the DJIA follows Benford's *Law of Anomalous Numbers*. Therefore, they can be classified into the group of "outlaw numbers without known relationship;" (Benford (1938)). The analysis presented here suggests that small changes are more likely than big ones; at the same time, the closer the daily changes are (in absolute value) to 0.1%, the more probable they are too.

References

DeGroot, M. (1971): *Optimal Statistical Decisions*. London: McGraw-Hill.

Benford, F. (1938): "The Law of Anomalous Numbers," *Proceedings of the Americal Philosophical Society*, **78**:4, 551–572.

Berger, J.O. (1985): *Statistical Decision Theory and Bayesian Analysis*. Berlin: Springer-Berlag.

Koedijk, K.G. and P.A. Stork (1994): "Should we care? Psychological barriers in stock markets.," *Economics Letters*, **44**:4, 427–432.

Leamer, E. (1983): "Model Choice and Specification Analysis," in *Handbook of Econometrics*, ed. Z. Griliches and M. Intriligator. Amsterdam: North-Holland.

Ley, E. and H.R. Varian (1994): "Are there Psychological Barriers in the Dow-Jones Index?," *Applied Financial Economics*, **4**, 217–224.

Pierce, P.S. (1986): *The Dow Jones Averages, 1885–1985. Cenntenial Edition*. Homewood, Illinois: Dow Jones-Irwin.

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*, **221**:6, 109–120.