Should Benford's Law Be Used to Examine Financial Returns? Evidence from CRSP

Bruce M. Bird [*], Michael J. Sinkey,[†] Alexander Smith, [‡] and William J. Smith [§¶]

July 2016

Abstract

Benford's Law, also known as the first digit law or leading digit phenomenon, has been found to 'hold' for a number of different datasets. While some researchers find that Benford's Law may show promise as a fraud detection tool, we do not find strong evidence to indicate that Benford's Law should be used as a reliable "filter" for fraud in this context. Using a battery of statistical tests on plausibly non-fraudulent mutual fund data, we examine to extent to which using Benford's Law leads to Type I error (in which a distribution of returns violates Benford's Law, but no fraud has occurred.) To our knowledge, this is the first systematic comparison of rejection rates for Benford's Law among different test statistics. Depending on the statistical test that we use, our results indicate that, of the 52,407 funds in our sample, between 4,873 to 8,841 of the funds violate Benford's Law, a non-conformance rate of between 9.30 to 16.87 percent. Although Benford's Law appears to be a reasonably good model for the distribution of first digits in our data set, it does not hold "enough" one to be able to draw reasonable inferences of fraud for a *given* non-conforming mutual fund. Finally, we use two standard asset pricing models–the Fama/French three-factor model and the Carhart four-factor model–to examine whether Benford-conforming funds perform differently than non-conforming funds. We do not find any evidence to suggest that these funds have higher abnormal returns, nor do we find evidence to suggest that standard return tilts perform differently for non-conforming funds.

[*]Department of Accounting and Finance, University of West Georgia, 1601 N. Maple, Carrollton, GA 30118 e-mail: bbird@westga.edu

[†]Department of Economics, University of West Georgia, 1601 N. Maple, Carrollton, GA 30118 e-mail: wjsmith@westga.edu

[‡]Department of Economics, Georgia State University

[§]Department of Economics, University of West Georgia, 1601 N. Maple, Carrollton, GA 30118 e-mail: msinkey@westga.edu

[¶]We thank Joshua Boatright for excellent research assistance and Roberto Croce for helpful comments.

1 Introduction

It is often difficult to know whether or not a set of observed financial returns is genuine. This data authentication problem has given rise to a sizable literature devoted to developing tools to "filter" for potentially fraudulent returns. However, the process of discerning whether a set of returns is fraudulent is subject to Type I or Type II error (where Type I error would consist of erroneously flagging genuine returns as being fraudulent and Type II error would consist of failing to flag fraudulent returns as being fraudulent). In this article we focus on the degree to which Benford's Law is subject to Type I error. By assessing the magnitude to which using Benford's Law to identify fraudulent returns mistakenly flags genuine returns as being fraudulent.

Benford's Law, or the first-digit phenomenon, is a pervasive empirical distribution of the first digits of sets of numbers which states that the digit one is observed roughly 30% of the time and that each subsequent digit is observed far less frequently. It is sometimes referred to as the Newcomb-Benford Law to give credit to Simon Newcomb, a Canadian-American astronomer/mathematician who observed in the late 19th century that the first pages of books containing dedactic logarithms showed much more use and wear than did the last pages. In 1938, Frank Benford rediscovered Newcomb's work and expanded his observation by examining data from more than twenty different distinct domains, finding that, for a diverse set of data, distributions of first digits seemed to follow a logarithmic distribution.

Indeed, many real-world data sets seem to have first digit distributions that reasonably conform to Benford's Law and it commonly applies to numbers arising from "naturally-occurring" processes or transactions. Benford's Law has been shown to apply in a diverse set of different empirical settings, including the lengths of rivers, the street addresses of individuals, death rates, and population sizes. Because of its pervasiveness across seemingly disparate settings, Benford's Law is frequently proposed as a fraud detection tool, since those who commit fraud generally have a myriad of objectives and may be unaware of Benford's Law or unable to construct strings of numbers that adequately match the shape of its underlying probability distribution. Furthermore, it is well-known that individuals do poorly in tasks where they are asked to supply random numbers (see, for example, Camerer (2003) [6] and Cho and Gaines (2007) [9]). Because of this cognitive bias, Benford's Law may be suitable for detecting fraud, since individuals unaware of Benford's Law are unlikely to replicate it in settings when it is likely to hold.

The strategy of using Benford's Law as a "filter" for fraud originated in forensic accounting, where it has been used to detect fraud due to fictitious invoices (Nigrini 1996,[26], Nigrini (1997) [29] Nigrini (1999) [27]).[1] More recently, however, Benford Law has been used to examine, among other settings, survey data (Judge and Schechter (2009))[22], campaign finance data (Cho and Gaines (2007)) [9], regression coefficients in economics and finance (Günnel and Tödter (2009)) [17], the Libor rate [30], accounting data (Durtschi et al. (2004)[12] and Amiram et al. (2015) [2], international transactions (Suh et al. (2011)) [31], and international balance of payments data (Michalski and Stolz (2013))[25], among other setting s.[2]

The notion that return authenticity can be examined by using statistical properties of return digits is not new. By investigating the efficacy of Benford's Law at identifying potential fraud, we complement recent literature devoted to analyzing techniques that may successfully identify fraudulent hedge fund returns. More specifically, our research coincides with works which suggest that analyzing the distribution of return digits may be fruitful in identifying fraudulent returns. Recently, Bollen and Pool (2012)[5] examine whether the *last* digit of a return is uniformly distributed between 0-9; in their paper, they flag hedge funds with more extreme distributions than the uniform distribution as being potentially fraudulent. Similarly, Bollen and Pool (2009) [4] find a substantial positive discontinuity between the number of small positive and small negative returns that hedge fund managers report, but that this discontinuity is eliminated in the three months prior to an audit.[3]

In order for statistical tests of Benford's Law to be effective, α, the probability of Type I error must be small and β, the probability of Type II error must also be small. We examine the likelihood of Type I error with a data set which contains plausibly genuine returns: the mutual

---

[1] It should be noted that the nonconformance of a given distribution with a Benford distribution will typically necessitate the performance of additional auditing tests or procedures.

[2] In these settings, though, nonconformance of a given distribution with a Benford distribution creates an inference of fraud without additional tests or procedures.

[3] Jorion and Schwartz (2013) [21] demonstrate that this discontinuity can be explained rationally by incentive fees, and as such might not be direct evidence of manipulation.

funds from the CRSP survivorship-bias-free U.S. Mutual Funds database. Doing so allows us to specifically estimate whether the probability of Type I error is small using a large set of data which, theoretically, should be expected to follow Benford's Law. Additionally, CRSP data have distinct advantages over other financial data when examining Type I error. First, CRSP mutual funds are comprised of weighted compositions of returns from many different probability distributions, making them more likely to follow Benford's Law (Hill (1995) [18]). Second, CRSP data are available for long time periods at a monthly level, allowing us to conduct our tests with sufficient power. Third, we are able to observe a variety of potential characteristics which may limit adherence to Benford's Law, such as fund strategy, return size, and composition of holdings. Finally, CRSP data are survivorship-bias free, which serves as external validity of the authenticity of the returns, since it is possible that mutual funds which disappear or do not survive do so because they are fraudulent. We are able to include such funds in our analysis.

The basic premise for examining conformity of a dataset with Benford's Law is to recover an empirical distribution of first digits and to compare this distribution with the log distribution that underpins Benford's Law by using some type of non-parametric statistical test. However, there is no strong consensus in the literature about which test best accomplishes this task. For example, the chi-square test is well-known to perform poorly for small data sets (Michalski and Stolz (2013) [25], but the Kolmogorov-Smirnov test is known to perform poorly for large data sets (Nigrini (2012) [28] and Amiram (2015) [2]). Nonetheless, both tests are widely used and cited. We try to alleviate this test selection bias by using six of the most commonly-used tests in the literature to produce a range of estimates for Type I error: the chi-square test, Kolmogorov-Smirnov test, Kuiper test, m-test, d-test, and Binomial tests. We construct test-statistics for all six tests for over 50,000 different mutual funds by using the first digit of all within-fund monthly returns as the basis for our empirical distribution. We argue that this circumvents the possibility of our estimates for Type I error being biased by any one test. A byproduct of this approach is that, to our knowledge, we are able to produce the first systematic comparison of rejection rates for Benford's Law among different test statistics.

Our main findings are that estimates of Type I error might plausibly range from nine to sixteen percent in our overall sample, as we conduct each of our goodness-of-fit tests at the .01 level of significance. Our findings are not driven by any one test; Benford rejection rates are highly correlated across different subsamples, and each test rejects the null hypothesis of Benford-like data at least nine percent of the time. Our results are not being driven by abnormally high or low decile returns; if anything, abnormally high or low decile returns are less likely to violate Benford's Law compared to middle decile returns. Likewise, our results are not driven by fund objectives likely to violate Benford's Law, including money market funds and funds with large holdings of government bonds. These funds are likely to violate Benford's Law since they target stable returns, rather than returns that increase geometrically. While these funds do in fact violate Benford's Law much more frequently, we find that, even after removing such funds from our data, rejection rates far exceed one percent.

Our findings suggest that tests of conformance with Benford's Law excessively flag mutual funds in the CRSP data. Assuming such funds are unlikely to be fraudulent, however, it may be the case that funds flagged for non-conformance simply perform differently than funds which conform. We address this possibility in a few novel ways.

First, we compare the conformance of the distribution of first digits of mutual funds in our overall sample to funds to Benford's Law with the conformance of funds which have been formally liquidated or merged with other funds. These funds might be expected to violate Benford's Law far more frequently since fund managers are actively choosing to remove them from the marketplace. We find, though, that these funds violate Benford's Law only *slightly* more frequently than the overall sample.

We next complement our main findings by conducting an exercise to see if funds that violate Benford's Law have excess abnormal returns. We estimate abnormal returns (alphas) from two common asset pricing models–the Fama/French three-factor model and Carhart four-factor model–and compare the distributions of abnormal returns for funds that do not conform to Benford's Law to those that do. While excess returns from these models are commonly interpreted as fund managerial skill, excess returns more simply represent any unobserved components that affect returns, which may include skill. We argue that if fund managers were manipulating mutual fund returns in our sample, the associated returns of those funds would be less likely to correspond with known return determinants, and such funds would have higher abnormal returns when compared with the overall sample. Critically, those funds should also be less likely to conform to Benford's Law.

We do not find any evidence to suggest that funds which do not conform to Benford's Law perform abnormally better or worse compared to funds which do conform. We show that funds which do not conform to Benford's Law have indistinguishable excess returns when compared with our overall sample. Our results are robust across test in that there is no evidence that choosing a different statistical test for conformity would assist in flagging abnormally higher or lower returns. Our results are also robust across fund strategy. This finding is especially striking, since many strategies might be considered *ex-ante* as likely candidates for manipulation.

The rest of our paper is as follows. Section 2 provides a description of statistical properties associated with Benford's Law and the subsequent appropriateness of CRSP data. Section 3 outlines our statistical methodology and highlights the strengths and weaknesses of existing statistical tests. Section 4 describes the data, while Section 5 presents our main estimates of Type I error. Section 6 presents the asset pricing models used to estimate excess returns and formally compares the distributions of Benford-conforming funds to non-conforming funds. Section 7 concludes.

2 Applicability of CRSP Data to Benford's Law

2.1 Statistical Properties Associated with Benford-Like Data

It is known that Benford's Law does not hold for data drawn from symmetric distributions, such as the normal distribution or uniform distribution. However, mixtures of normal, uniform, chi-square, exponential, and other probability distributions tend to follow Benford's Law (Janvresse and Rue (2004)[20], Formann (2010)) [14]. According to Durtschi et al. (2004), [12] Benford's Law is likely to hold for quantitative data that is a mixture of two or more probability distributions. As an example, Benford's Law seems to hold for a large number of accounts payable and accounts receivable transactions, since these data points are mixtures of price and quantity distributions.

Finally, in a recent work, Michalski and Stolz (2013)[25] refer to three statistical properties which distributions whose first digits follow Benford's Law seem to possess. They find that the first digits of data-generating processes follow Benford's Law when those processes exhibit exponential growth (i.e., geometric sequences), are the results of random samples of random distributions, and exhibit scale invariance.

2.2 Appropriateness of CRSP Data

Our statistical approach is compromised if mutual fund return distributions are poor candidates for following Benford's Law, since data that are unlikely to follow it will be identified by the test regardless of whether or not they are purportedly fraudulent in nature (Type I error). However, we assert that mutual fund returns for most types of mutual funds are plausible statistical candidates for following Benford's Law. First, mutual funds with equity-based strategies are likely to experience exponential growth. While mutual fund holdings are non-random, they are mixtures of many random distributions. More importantly, it is unlikely that mutual fund managers are concerned with Benford-like properties of returns, which would mean that fund selections are made regardless of how the digit returns adhere to Benford's Law. Indeed, many mutual funds have hundreds of holdings from different sectors of the economy. Mutual funds with foreign holdings may hold mixtures of equities from many different countries with vastly different institutional features.

The CRSP database contains objective codes that clearly describe the mutual fund objectives that are announced to investors. This is important for studying Benford's Law because some fund objectives intuitively might not be expected to follow Benford's Law. As an example, the objectives of fixed-income and money market funds are to provide relatively constant rates of return. This, in turn, violates the assumption of Benford's Law that the underlying data exhibit exponential returns. Because mutual fund objectives are listed in the CRSP data, we can exclude any fund with objectives that are unlikely to produce returns consistent with Benford's Law, which mitigates the possibility of Type I error.

Our approach is also compromised if mutual funds returns are fraudulent, since in those cases the test would be accurately flagging returns. Our contention is that mutual funds listed in the

CRSP data are presumably non-fraudulent. We believe this is plausible for several reasons. Unlike hedge funds, all U.S. mutual funds must be registered with the SEC. Mutual funds are managed by a voted-upon board of directors, which mitigates the probability of any one individual manipulating returns. Next, mutual funds are subject to substantial disclosure requirements, in accordance with the Investment Company Act of 1940. During our sample period, the SEC required mutual funds to completely disclose holdings quarterly (before 1985 and after 2004) and bi-annually (from 1985-2004). The Investment Company Act also requires mutual funds to issue yearly audited financials.

We note that our claim that mutual funds are presumably non-fraudulent is substantially different from the notion that mutual fund companies are free from incentives to strategically target returns at key points during a year, such as "tax trading" or "window dressing." For example, mutual fund managers may more quickly dispose of their losing funds to minimize tax liability prior to the common October 31 year-end (Gibson et al. (2000) [15]). Managers may also purchase recent winners and sell losers before the applicable "disclosure dates" to publicize their new investments (Alexander et al. (2007) [1]). It is clear that mutual fund companies may have incentives to maximize inflows rather than risk-adjusted returns (Chevalier and Ellison (1997) [8]), but that those incentives do not necessarily constitute fraud. Although there is the possibility for perverse incentives for mutual fund managers, we believe that, due to disclosure requirements and periodic SEC requirements, our approach of using mutual fund data to investigate the possibility of Type I error is reasonable.

3 Statistical Approach

Numerous statistical techniques predicated on goodness-of-fit have been proposed to test for deviations from Benford's Law. These techniques consist of constructing a test-statistic that assesses goodness-of-fit between the observed data and an underlying probability distribution based on specified distance criteria. In order to reduce the likelihood that our findings related to Benford's Law are sensitive to any specific test, we implement and compare six commonly used goodness-of-fit tests that are currently used to examine whether observed data differ from Benford's Law. We do this for a few reasons. First, different test statistics will yield different rejection rates and there is no strong consensus on which test should be used. Second, to our knowledge, no systematic comparison of between-test rejection rates has yet occurred in the literature.

We define a deviation from Benford's Law as an observed discrete distribution that is statistically different from the distribution used as the underpinning for the law, i.e.,

$$Pr(X = k) = \log_{10}(1 + \frac{1}{k}), \tag{1}$$

where k is the first non-zero digit of the data string. While parametric methods are available to examine Benford's Law, we prefer to use methods that assess the overal shape of the empirical distribution.[4] The most common method to compare a discrete empirical distribution to a benchmark distribution is the Pearson's chi-squared test, which has been implemented in numerous studies of Benford's Law (Judge and Schechter (2009), Michalski and Stolz (2013) among others). This goodness-of-fit test compares observed frequencies to expected frequencies, and, for any sample of size N, compares the expected number of first digits to the actual number of first digits for all possible first digits k. The test statistic is then

$$\chi^2 = \sum_{i=1}^{9} \frac{(O_i - E_i)^2}{E_i}$$

where O_i is the observed frequency of the digits and E_i is the expected frequency. Because higher digits are observed infrequently according to Benford's Law, it is possible to have instances when the expected cell count is less than five, in which case the asymptotic chi-squared distribution may be inappropriate (Cochran (1952)). [10]. While this restriction has been shown to be restrictive (Conover (1999))[11], to overcome this potential problem with the chi-square test, we examine the robustness of our results by both including all funds in our samples and restricting our samples to funds with $N > 110$ returns, since $5 > \frac{110}{\log(1+\frac{1}{9})}$, which follows Michalski and Stolz (2013) [25]. As described in Cho and Gaines (2007), Morrow (2010), and Michalski and Stolz (2013), this test, while computationally straight-foward, tends to reject the null frequently for large N, which may

[4] One such test-statistic would be a difference in means test, where the empirical mean is compared to the expected value of a Benford-type distribution, as used in Günnel and Tödter (2009). [17] However, this test-statistic does not assess the overall shape of the empirical distribution and could be satisfied by any number of non-Benford distributions.

be undesirable.

Because of the limitations associated with the Pearson's chi-square test, we employ a battery of non-parametric tests based on evaluating discrepancies between the cumulative distribution function (CDF) associated with Benford's Law and the CDF associated with the observed distribution. We first employ the Kolmogorov-Smirnov test and a close derivation, the Kuiper test. While the Kolmogorov-Smirnov does not have a well-known distribution for discrete variables, we overcome this problem by following Jann (2008) [19] by using Monte Carlo simulations to determine the p-values for each test-statistic.[5] Letting N again be the number of observations, the K-S test statistic for any given CDF is as follows:

$$D_N = \sup_x |F_n(x) - F(x)|$$

where $F_n(x)$ is the observed CDF and $F(x)$ is the comparison CDF. We multiply an asymptotically-consistent version of this test-statistic by multiplying D_N by $\sqrt{N}$ to obtain the test-statistic for any sample size N. In summary, the K-S test finds the maximum difference between the observed CDF and the empirical CDF, and is particularly easy to implement for a discrete probability distribution.

We also implement the Kuiper test, a similar measure to the K-S test. While the K-S test concentrates on finding shifts in a probability distribution, the Kuiper test is more equipped to find spreads in the probability distribution. The Kuiper test statistic is:

$$V_N = \sup_x [F_n(x) - F(x)] + \sup_x [F(x) - F_n(x)]$$

where $F_n(x)$ and $F(x)$ have the same meaning as before. We modify this test-statistic as suggested by Stephens (1970) for smaller sample sizes, calculating instead $V_N^* = V_N \times (\sqrt{N} + .155 + \frac{.24}{\sqrt{N}})$. We use critical values as proposed in Giles (2006) [16] which are asymptotically consistent regardless of sample size.

Two other techniques have recently been proposed as methods for testing conformity to Ben-

[5] In particular, we simulate the exact p-values by computing both observed and expected frequencies for each fund in our data set and then by sampling from the null distribution with 10,000 replications. We investigate the robustness of our simulation approach by comparing the simulated p-values to the p-values suggested in Morrow (2010) and do not find any statistically different result.

ford's Law: the m test and the d test, first proposed in Cho/Gaines (2007), but also described in Morrow (2010) and Suh (2011). These tests are similar conceptually to least absolute deviation and quadratic loss measures and concentrate on measures of Euclidean distance. Both tests assess absolute distance criteria, but instead of measuring distances between the observed cdf and the proposed Benford-type cdf, these statistics measure the distance between the observed digit probabilities and the proposed discrete probabilities generated from Benford's Law as defined in (1). As can be seen in both test-statistics, large deviations between the test-statistic and any one node of the discrete probability distribution lead to an increased probability of rejection, making these statistics similar to the Pearson's chi-square measure. The m-statistic is as follows:

$$m = \max_x |Pr(X = k) - \log_{10}(1 + \frac{1}{k})|$$

while the d statistic is:

$$d = \left[\sum_{k=1}^{9} \left(Pr(X = k) - \log_{10}(1 + \frac{1}{k})\right)^2\right]^{\frac{1}{2}}$$

Both measures involve straight-forward calculations for a discrete probability distribution such as Benford's Law. Following Morrow (2010), we transform each test-statistic by multiplying by the test-statistic by $\sqrt{n}$ and subsequently use critical values at the .01 level of 1.212 for the m test and 1.569 for the d test.[6]

Finally we aggregate violations associated with a proportions test-statistic (z-test) by comparing the expected proportion of a digit based on Benford's Law to the actual proportion of the digit in the data, as suggested in Nigrini (1996).[29] For each digit, we calculate the proportions test statistic $z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ where $\hat{p}$ is the empirical proportion and p_0 is the proportion that is associated with Benford's Law. We note that this approach has limitations. First, it is probable that at least one digit would violate Benford's law regardless of whether the data is Benford-like or not; there is a 27 percent chance that at least one digit test will reject Benford's Law at a .05 significance level and a 39 percent chance at the .1 significance level. Additionally, as described in Durtschi

[6] Specifically, we construct $m^* = \sqrt{n} \max_x |Pr(X = k) - \log_{10}(1 + \frac{1}{k})|$ and $d^* = \sqrt{n} \left[\sum_{k=1}^{9} \left(Pr(X = k) - \log_{10}(1 + \frac{1}{k})\right)^2\right]^{\frac{1}{2}}$

et al. (2004) [12], for large data sets this type of test-statistic will reject Benford's Law extremely frequently, indicating excessive power. While we implement proportion test-statistics for our data for completeness, we agree with other researchers about the limitations of this approach.

4 Data and Empirical Strategy

We use data from the CRSP Survivor-Bias-Free US Mutual Fund Database from the University of Chicago's Center for Research in Security Prices. This data contains returns for over 52,000 unique mutual funds from 1960-2013. We use monthly returns as the data for the returns used to test Benford's Law in order to ensure sufficient statistical power to conduct our tests. For each fund in the database, we are able to observe the total return per share at month's end. We extract the first digit from this return in order to construct the empirical distribution for Benford's Law. We also observe a rich set of characteristics about the composition of each fund, including a breakdown of common stocks, preferred stocks, convertible bonds, corporate bonds, municipal bonds, government bonds, other securities, cash, fixed-income securities, asset-backed securities, mortgage-backed securities, and other equities. We also observe the expense ratio, whether or not the fund is a "dead" fund, whether or not the fund is open to new investors, and whether or not the fund is a retail fund.

The fund objective is of particular interest because it impacts the distribution of the first digits. Specifically, fund objectives which target stable rates of returns are not likely to have geometric returns, and, as such, their inclusion in the sample would bias our estimates of Type I error upward. To identify objectives, we use the CRSP objective code, which combines information on fund objectives from the Wiesenberger Fund Type Code (used prior to 1993), the Strategic Insight Objective Code (used from 1993-1998), and the Lipper Classification (used from 1998-on). CRSP codes are split into four general categories: equities, fixed income, mixed fixed equity and income, and other; of those categories, both equities and fixed income funds have a particularly rich set of sub-categories of objectives. In all, there are 271 different sub-categories of objectives, of which 241 are either equities or fixed-income. These categories are important, since we might expect wide groups of both fixed income and equity funds to violate Benford's Law (e.g., money market funds, TIPS, etc.). We examine each of these categories to ascertain whether the distribution of return first digits from the stated objective might reasonably be expected to follow Benford's Law, and note instances when assumptions that lead to conformance with Benford's Law are violated.

We construct the data for our sample as follows. First, we delete any month which does not have an observed percentage return (typically the first month of the time series, when no monthly return has yet been realized). We then find the first significant digit for each return, which is constructed from percentage changes in prices between returns. As an example, if a fund has a return of .0043 between months, then the first significant digit is '4,' which occurs with probability $\log_{10}(1+\frac{1}{4})$ in our data. We next tabulate the number of significant digits by digit for each fund and calculate each of the six test-statistics for each fund.

In order to construct a test-statistic for each fund we must construct an expected number of significant digits for each digit 1-9. This is based on the total number of returns by fund, N, and the expected frequency of a digit, given by $log(1+\frac{1}{k})$, where k is the digit. As an example, for a fund with 200 returns, we would calculate an expected number of ones as $\log_{10}(1+\frac{1}{1})\times 200 = 60.206$, an expected number of twos as $\log_{10}(1+\frac{1}{2})\times 200 = 35.218$, etc., such that the sum of expected digits equals 200. These expected frequencies serve as the expected distribution for the chi-square test statistic based on the sample size. The empirical cumulative distribution function is constructed by computing the observed percentages of digits for each digit 1-9. We then compare these amounts to the actual cumulative distribution, generated by $\sum_{i=1}^{k}\log_{10}(1+\frac{1}{k})$, where $k \in \{1, 2, \dots 9\}$ and $\sum_{i=1}^{9}\log_{10}(1+\frac{1}{k}) = 1$. This approach is the basis of the m, d, $K-S$, and Kuiper tests.

A consequence of our empirical strategy is that we only examine whether a fund violates Benford's Law at time T, i.e., we tabulate the first digits for all available returns of a fund. As a result, we do not perform "rolling" tests of whether a fund violates Benford's Law, (i.e., we do not evaluate whether or not a fund adheres to Benford's Law after each new return, only after we accumulate all of the returns for the fund.) We do this for a few reasons. First, imposing rolling tests might greatly increase the possibility of Type I error because we would be performing a number of different hypothesis tests *within-fund*. Next, any other cutoff that we might impose (i.e., evaluating whether a fund adheres to Benford's Law after one or two years of returns) lacks statistical justification since it would only represent part of the available empirical distribution of return digits.

5 Results

We present the results for our entire sample in Table 1. For each test, we assign a one if the test statistic indicates that the empirical distribution is statistically different from the Benford distribution at the $p < .01$ level, i.e., if the p-value associated with the test is less than .01. We then tabulate the frequency of tests for which the test statistic indicates a statistically significant difference.

Table 1 finds significant evidence that, for the entire sample, sets of returns seem to violate Benford's Law frequently. Depending upon the test used, the total sample violates Benford's Law from a range of 9.30% (for the M test) to 16.87% (for the Kolmogorov-Smirnov test). Because each test is conducted at the 99 percent confidence level, our results indicate that Benford's Law systematically over-rejects the null for mutual funds in our sample. Our results indicate that, of the 52,407 funds in our sample, between 4,873 to 8,841 of the funds in our sample have first-digit distributions that do not adhere to Benford's Law at the 99 percent level.

It is important to note that a vast majority of mutual funds in our data set adhere to Benford's Law, and it appears to be a reasonably good model for the distribution of first digits of our data set. However, Benford's Law does not seem to hold frequently "enough" for one to be able to draw a reasonable inference of fraud with regard to a given non-conforming mutual fund. Accordingly, our results suggest a significant possibility for Type I error.

Table 2 displays the percentages associated with violations of the $z-$test for the entire sample. This table is associated with the digit-by-digit proportions tests. In particular, for each fund, we check to see whether the proportion of ones, twos, etc., are as expected using the $z-$test described previously. We find that over 30% of the returns in our total sample exhibit at least one digit violation of Benford's Law and over ten percent exhibit two or more violations. For the entire sample, we find that 2.96 percent of funds in our sample have three or more digit-based violations at the 99 percent level, and that 7.95 percent of our funds have at least two digit-based violations.

This result is striking, given that, while the unconditional probability of receiving one or fewer violations of the proportions test statistics is 99.5 percent, our results are almost 16 times as large. These results are consistent with the rejection rates in Table 1.

Table 3 splits mutual funds by return decile. We do this because, should one wish to use Benford's Law to examine the potential for fraud, it is reasonable to think that funds with higher mean returns are more likely to be fraudulent, since they would, on average, be more attractive to investors and would attract more fund inflows. We split the funds into equally-spaced deciles and report the results for each decile. We find that mutual funds with average returns in the top decile seem to adhere to Benford's Law more frequently than those with returns in lower deciles. Specifically, rejection rates for the top decile of results vary between 1.23 and 7.56 percent, which is markedly lower than the full sample results in Table 1. This is likely due to the fact that fund objectives based on fixed income and money market strategies have lower returns on average in our sample than funds based on equity strategies, and we would not expect fund objectives based on such strategies to adhere to Benford's Law. Table 4 reports the results for the $z-$test by decile, and the results are consistent with the results from the other tests.

5.1 Fund Objectives

Mutual fund managers may have explicit incentives to construct stable returns or to provide relatively constant rates of interest, as in money market funds. In such cases we would not expect those fund returns to follow Benford's Law because those distributions would not have geometric returns. Inclusion of those fund objectives in our overall sample overstates the number of true violations of Benford's Law, since one does not expect those funds to follow Benford's Law. Other fund objectives, however, might be strong candidates for following Benford's Law. For example, Ley (1996) [23] finds that the distribution of digits of one-day returns of the S&P 500 Index reasonably adheres to Benford's Law. As such, mutual funds such as index funds, large-cap funds, or growth funds may reasonably be expected to follow Benford's Law, since they have a wide range of holdings from many different sectors. Foreign funds may be expected to follow Benford's Law for the same reason. Sector-specific funds might be less likely to follow Benford's Law, since such returns are exposed to sector-specific risk and it is less likely that these holdings represent random draws from random distributions.

Table 5 presents the results associated with fund objectives. We find that index funds seem to be the most likely fund type to adhere to Benford's Law, with violations from .59% to 7.37% of our sample, implying that we cannot reject the null hypothesis that Benford's Law fits for between 92.63%-99.41% of funds in our sample. We reject the null that a fund's returns follow Benford's Law for between 1.03%-9.09% of foreign holdings, between .85%-17.41% of mid-cap holdings, between .71%-24.73% of small-cap holdings, between 1.05%-11.36% for growth holdings, and between 3.96%-11.69% of sector-based holdings. We find that, within-test, the chi-square test is far more likely to reject the null than other tests, and that all tests are more likely to reject the null when compared with the m-test. In fact, these results would seem to indicate that the $m-$test might have substantially less Type I error than other tests, although we still find that the $m-$test rejects the null 3.96% of the time for sector-based holdings.

As expected, money market funds do not follow Benford's Law. We find that an overwhelming number of funds–depending upon the test used–between 81.58%-88.39% do not follow Benford's Law. This is consistent with our intuitive notion that stable rates of return should not exhibit the exponential growth required to give rise to Benford-like distributions of first digits. We find that mutual funds which hold higher percentages of government bonds also follow Benford's Law less frequently; specifically, between 11.11%-16.13% of such funds do not follow Benford's Law.

5.2 Do Funds That Are Delisted Follow Benford's Law?

We next examine sets of mutual funds that may be especially good candidates for violating Benford's Law for reasons unrelated to strategy. Dead funds are funds that are terminated by fund management groups for various reasons. Lunde et al. (1999) [24] find that both a fund's relative performance and the performance of the sector corresponding to fund strategy contribute to fund exits. They argue that funds are terminated for many reasons, but that the most common ones include consolidation of funds with similar objectives, insufficient funding, and poor performance relative to other funds in the group. While we have no reason to believe that dead mutual funds indicate fraud, one might believe that early terminations of a mutual fund might indicate fraud.

Table 6 produces results for dead funds. Within the CRSP data, a reason for why the fund is dead is available for 8,858 of 17,254 such funds. We concentrate on the 3,365 funds that have been formally liquidated and the 5,451 that have been merged into other funds. Because the reason for why a fund is dead is unknown for a large fraction of our sample, it is possible for a larger percentage of formally liquidated funds and merged funds to violate Benford's Law when compared with the complete sample.

The first row of the table presents results for all dead funds, while the second and third rows divide dead funds into two categories, formally liquidated funds and merged funds. For the entire sample, we find rejection rates that are higher than our overall sample, ranging from 9.89 to 17.73 percent of the sample. Funds that are formally delisted exhibit somewhat higher rejection rates such that between 11.15 to 21.63 percent of the sample violate Benford's Law. Merged funds also exhibit slightly higher rejection rates, as between 9.96 percent to 18.87 percent of these funds violate Benford's Law.

These findings highlight our concerns about using Benford's Law without adequately taking into account the possibility of Type I error. Taken literally, dead funds violate Benford's Law between ten to eighteen times more frequently than expected at the .01 level of significance. However, when compared with the main sample of funds, dead funds are only slightly more likely to violate Benford's Law. In our overall sample, funds violate Benford's Law between 9.30 to 16.87 percent of the time, which is not appreciably different in practice from our sample of dead funds. We conjecture that, absent a suitable control group, it might be plausible to conclude that dead funds in our sample have manipulated returns, and we assert that this conclusion would constitute Type I error. More generally, we contend that in many settings where Benford's Law is argued to be likely to hold, it does not "hold enough" for one to confidently assert that a rejection of conformity constitutes anything other than Type I error.

5.3 When do Different Test Statistics Reject Benford's Law?

Our results indicate that substantially more funds violate Benford's Law than would be expected at a .01 level of significance. Furthermore, the objectives and disclosure requirements of mutual funds make it difficult to reasonably interpret the results as evidence of fraud, since mutual fund objectives are generally well-known and strategies must be disclosed quarterly. In and of itself, this may not be worrisome. For example, the distributional properties of fund digits may not give rise to Benford's Law naturally, or, more probably, some subsets of funds within a category may violate at least one of the assumptions required for digit distributions to follow Benford's Law.

More worrisome, however, is the dispersion of rejection rates across test statistics. Regardless of the underlying distributional properties of first digits, a test that is selected should reject (or not reject) the null relatively similarly to any other test that is chosen. This is especially important because a researcher can choose any reasonable goodness-of-fit test to examine adherence to Benford's Law. This gives rise to potential selection bias in that researchers can select from a menu of tests with which to examine adherence to Benford's Law. Because of this, it is worth considering whether different goodness-of-fit tests exhibit similar behavior when rejecting the null hypothesis. If rejections are highly positively correlated with each other, then, in practice, it may not matter which test is used. Accordingly, in Table 7 we report a correlation table that contains the rejections of five of the different test statistics that we use.[7]. Table 6 shows that rejections are moderately to strongly positively correlated, ranging from correlation coefficients of .4831 to .7217. The m test has the lowest rejection rates and is least likely to be correlated with the other test statistics, while the Kuiper and d tests are somewhat more likely to reject the null when the other tests reject.

We also test to see whether funds with larger numbers of returns have stronger correlations between tests. As described in Morrow (2010), the m and d tests may not perform as well for samples with fewer than 80 returns. Furthermore, as described in Michalski and Stolz (2013), the chi-square test may be inappropriate for samples with small expected counts, so we also examine the correlations between tests for larger samples. We do not find any substantial difference between correlations between the total sample and our restricted sample of funds with larger numbers of

[7]We exclude the binomial violations test because it is not plausible to directly compare the number of binomial violations to the rejections associated with other test statistics

returns; there is somewhat more variation in the correlations, but rejections are still positively correlated, with correlations that range from .4127 to .7404. Again, the m test has the lowest rejection rates and is least likely to be correlated with the other test statistics.[8]

Taken together, Tables 1, 3, 5, 6, and 7 highlight the sensitivity of rejection rates to test statistic selection. Were one to use the $m-$test or Kuiper test as the sole test-statistic to assess the goodness-of-fit of the empirical distribution to the Benford's distribution, one would be far less likely to reject the null hypothesis of Benford-like data than if one uses the chi-squared test or Kolmogorov-Smirnov test. While rejection rates are correlated between tests (i.e., all tests systematically reject the null more frequently for money market funds compared to index funds, and all tests systematically reject the null more frequently for middle-decile mean returns compared to higher-decile mean returns), there are vast within-category differences in rejection rates between test-statistics.

6 Do Benford-Conforming Mutual Funds Perform Differently than Non-Conforming Funds?

Our primary analysis examines the extent to which series of mutual fund returns violate Benford's Law. In this section we ask whether violating Benford's Law has any practical significance when predicting mutual fund returns. To this end, we examine whether there are any systematic differences between return determinants among funds that follow Benford's Law and funds that do not follow Benford's Law. We also examine whether there are systematic differences in abnormal returns among funds that follow Benford's Law. In principle, adherence of mutual fund returns to Benford's Law should not impact the magnitude of abnormal returns, nor should it impact how much known determinants affect returns. Nonetheless, manipulation of fund returns could possibly bias stated average returns upward, which would generate larger abnormal returns, since lower returns would not attract new investors.

Whether or not a fund adheres to Benford's Law should not affect the magnitude of fund

[8]We note that the correlation matrix does not account for fund objective. However, funds that would be expected to violate Benford's Law should theoretically violate Benford's Law regardless of the test statistic that is computed.

returns. Because of this, we choose not to model fund performance as a function of Benford's Law, as such models would be misspecified. To address this issue, we instead choose to estimate two standard models of mutual fund performance–the Fama-French three-factor model [13] and the Carhart four-factor model [7]–for every mutual fund in the CRSP database. These models are among the most common models that "investors might reasonably employ when evaluating the performance of mutual funds" (Barber et al. 2016, p.14) [3] and should capture many of the factors that determine mutual fund performance. We argue that estimating these models for each mutual fund allows us to net out other factors known to impact returns. We then examine the differences in fitted distributions of abnormal returns between funds that follow Benford's Law and those that do not.

In particular, for each mutual fund i, we estimate the following cross-sectional models:

(FF three-factor)

$$r_{i,t} = \alpha_{i,T} + \beta_{i,t}(r_{m,t} - r_{f,t}) + s_{i,t}SMB_t + h_{i,t}HML_t + e_{i,t} \tag{2}$$

where $r_{i,t}$ is the excess return for mutual fund i for month t over the risk-free rate of return, $r_{m,t}$ is the market rate of return, which includes all NYSE, AMEX, and NASDAQ funds and $r_{f,t}$ is the risk-free rate of return, measured by the one-month Treasury bill return. SMB_t represents the differential return on portfolios constructed based on size (as defined by market equity) segmented by book to market (value, neutral, or growth).[9] HML_t represents the differential returns on portfolios based on value vs. growth at month t. [10] and $e_{i,t}$ is an error term. We also estimate:

(Carhart four-factor)

$$r_{i,t} = \alpha_{i,T} + \beta_{i,t}(r_{m,t} - r_{f,t}) + s_{i,t}SMB_t + h_{i,t}HML_t + p_{i,t}MOM_t + e_{i,t} \tag{3}$$

which is the Fama/French three factor model with an additional term for momentum in stock returns, MOM_t, defined as $\frac{1}{2}(\text{small high} + \text{big high}) - \frac{1}{2}(\text{small low} + \text{big low})$.

[9] In particular, SMB, as defined by Fama and French, is $\frac{1}{3}(\text{small value} + \text{small neutral} + \text{small growth}) - \frac{1}{3}(\text{big value} + \text{big neutral} + \text{big growth})$.

[10] HML_t, as defined by Fama and French, is $\frac{1}{2}(\text{small value} + \text{big value}) - \frac{1}{2}(\text{small growth} + \text{big growth})$

The parameter of interest in each model is $\alpha_{i,T}$, which is the monthly abnormal return of mutual fund i averaged across T periods (sometimes described as "skill"). While alpha is often considered to be a measure of fund manager performance, it also captures any unobserved factors that may impact returns. If, for example, a manager's strategy led to a violation of Benford's Law, both the strategy and the violation would be captured in alpha.

Empirically, we continue to use monthly mutual fund returns and fund strategies from the CRSP mutual fund database. We match our returns to whether or not a fund violated Benford's Law as delineated in Section 4, and use violations associated with the Chi-square, $K-S$, Kuiper, $m-$, and $d-$tests. We augment our data by gathering data on the aggregate market return, SMB, HML, and MOM from portfolios formed by Fama and French.[11]

We begin by noting that the estimates we produce for alpha are averaged across all available time periods for a mutual fund. As with our analysis of Benford's Law, we produce estimates for all available returns of a fund, not for specific intervals of within-fund returns. Because of this, we do not capture whether a particular fund experiences a positive alpha at any point in its history, nor do we observe whether a particular fund violates Benford's Law for some sequence of its returns, but not other sequences. We only estimate alpha after observing all available returns, and we compare those estimates to violations of Benford's Law based on the entire distribution of first digits.

We proceed in two parts. First, we examine whether there are systematic differences between abnormal returns of funds that follow Benford's Law, and those that do not, by estimating the distribution of α for all of the funds in our sample. Because we do not know the exact distributions of alpha and are interested in where the distributions may differ, we then visually compare the density estimates of the overall distribution to the density estimates of the subsets of funds that do not follow Benford's Law. We perform an analogous procedure for the distributions of alpha based on fund strategy. Since adherence to Benford's Law is impacted by fund strategy, it may be the case that any differences that we observe between the overall distribution of alphas and the distributions of alphas associated with funds that violate Benford's Law may simply capture

[11] These factors are available on Kenneth French's online data library: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`

differences in fund strategies. [12]

Next, we examine the distributions of known return determinants. In principle, adherence to Benford's Law should be orthogonal to the other factors (HML, SMB, MOM, and $(r_{m,t} - r_{f,t})$ that we consider. The returns of funds that do not adhere to Benford's Law should not be impacted differently by known factors when compared with those that do adhere. To address this possibility, we examine the distributions of each of the tilts for all funds $(\beta_{i,t}, s_{i,t}, h_{i,t}$ and $p_{i,t})$.

We visually compare the distributions of abnormal returns for a few reasons. First, we are particularly interested in distributions that violate Benford's Law that also differ from the overall sample. If, for example, funds with distributions that violate Benford's Law have appreciably higher alphas, one might flag those differences as *prima facie* evidence of potentially fraudulent returns, but one would probably not flag funds with lower alphas as being potentially fraudulent. Next, we note that goodness-of-fit tests may be attractive, such as the Kolmogorov-Smirnov or Cramer-Von-Mises tests, which compare the baseline distributions of the estimated parameters to distributions of funds that violate Benford's Law. However, one must have a hypothesized distribution for the baseline in order to credibly implement such tests. Because we are estimating both the baseline distribution and the distributions of alphas for funds that violate Benford's Law, we do not have a known baseline distribution for comparison.

6.1 Results

We first produce kernel density estimates of both the three-factor alpha and four-factor alpha for our entire sample in Figure 1.[13] The dashed line represents a zero baseline, and the figure shows that both the mean and median alphas are slightly below zero, although the vast majority of alphas are distributed close to zero. This indicates that a majority of funds do not have abnormal returns after accounting for momentum, value, size, and market return. These estimates are consistent with the literature that suggests the relative inability of fund managers to outperform the market.

Figure 2 depicts estimates for the densities of three-factor alpha *conditional* on violating Ben-

[12]We examine this possibility later by splitting our sample both by Benford's Law adherence and fund strategy.

[13]Our estimates are produced using an Epachinikov kernel with bandwith of .0155, but are robust to other bandwidths selected.

ford's Law. Each line represents a kernel density estimate of the distribution of alpha for the subset of funds that violates Benford's Law according to a specific statistical test. Again, we find that alphas appear to be normally distributed; that their mean and median is slightly below zero, but that most funds have abnormal returns close to zero. What is striking is that this is true regardless of which test for Benford's Law is implemented. Simply put, Benford's Law does not seem to have any explanatory power regarding the distribution of returns.

Figure 3 depicts estimates for the densities of four-factor alphas conditional on violating Benford's Law. Again, each line represents a kernel density estimate of the distribution of alphas for the subset of funds that violate Benford's Law. Both the qualitative and quantitative features of the estimates mirror Figure 2, indicating that adding an additional determinant of returns does not appreciably impact the distribution of returns.

Taken together, Figures 2 and 3 do not provide supporting evidence that funds which violate Benford's Law have higher alphas relative to those that do not. There is no practical difference in abnormal returns between funds whose returns violate Benford's Law and those funds whose returns conform to the law. These results corroborate our findings in Tables 3 and 4 that funds which violate Benford's Law are more likely to have average historical returns in the middle deciles. Additionally, many of the funds which violate Benford's Law within our sample are fixed income or money market funds, which have lower returns relative to the market in our sample period. We would expect these funds to have lower abnormal returns relative to the other funds in the sample irrespective of whether or not the distributions of their digits follow Benford's Law.

Figure 4 depicts estimates of the distributions of four-factor alphas for four sets of equity funds: index funds, mid-cap, growth, and foreign funds. (We only report the distributions for four-factor alphas as the results for three-factor alphas are largely similar). Funds in these categories violate Benford's Law less frequently compared to money market or government funds. Additionally, the holdings of these funds may be more closely priced by three-factor and four-factor models since they primarily hold domestic equities.

The top left panel of Figure 4 provides density estimates of the distributions of alphas for index funds, conditional on whether or not they violate Benford's Law. We find that the distribution of abnormal returns is tightly centered around zero, which is reasonable considering the returns of these funds are designed to follow the market return. We do not find any difference between funds that conform to Benford's Law and funds that do not. This is unsurprising in the context of Table 5, which finds that index funds are highly likely to conform to Benford's Law.

The top right panel of Figure 4 provides density estimates of the distributions of alphas for mid-cap funds. These funds are more actively managed than index funds We again do not find significant evidence to suggest funds which violate Benford's Law perform appreciably worse or better than those that do not. If anything, the overall sample of alphas is slightly more dispersed when compared to different samples of funds that violate Benford's Law.

The bottom two panels of Figure 4 provide evidence for growth and foreign funds. These funds are likely to be more actively managed and may involve more risk-taking by their managers. In principle, these funds might be higher-likelihood candidates for managerial tactics designed to increase fund flows. For both panels, we again do not find any evidence to suggest that funds which do not conform to Benford's Law have systematically lower abnormal returns when compared with the overall sample of funds; this result is robust across test. Again, each distribution is centered around zero, and there is little difference between the distributions of the different statistical tests and the distributions of the full sample.

Finally, in Figure 5 we report estimates of the densities of the other parameters in the return model– ($\beta_{i,t}, s_{i,t}, h_{i,t}$ and $p_{i,t}$). We find that, for each distribution of parameter estimates, funds that follow Benford's Law are essentially the same as those that do not. This is intuitive, as rejections for Benford's Law are determined independently of the four factors used to predict returns. Still, we provide these estimates to rule out other potential channels for Benford's Law to affect returns.

To summarize, Figures 2-4 examine distributions of abnormal returns for the entire sample of mutual funds. We examine abnormal returns in the context of two common models of returns and across a number of different subsamples that should plausibly have the most active fund management. Moreover, we exclude samples of funds whose properties make them more likely to violate Benford's Law. Since it is well-established that investors tend to chase funds that report high returns, fund managers have strong incentives to produce higher returns. However, regardless of sub-sample or return model, we do not find evidence to suggest that Benford's Law effectively "flags" funds with abnormally high returns.

7 Conclusions

Benford's Law has been used to detect fraud in a number of different empirical settings. A common motivation in such research is the notion that the data studied are potentially fraudulent and that Benford's Law has promise in sifting between data likely to be genuine and data likely not to be. In this paper, we perform an exhaustive empirical exercise that compares six of the most common goodness-of-fit tests for Benford's Law for 52,000 separate mutual funds.

We find seemingly robust evidence that statistical goodness-of-fit tests used to compare underlying empirical distributions of first digits to Benford-like distributions of first digits may overreject the null hypothesis that the data are not fraudulent. Moreover, we find that this phenomenon is persistent across return decile, fund strategy, funds which are liquidated, and funds which are merged into other funds. We find estimates that, regardless of method, Benford's Law is seemingly rejected anywhere from five to sixteen percent of the time in the overall sample, and that this result is relatively robust across return deciles, fund objectives, and test-statistic used.

We extend our analysis to a thought exercise in which we estimate two plausible models (the Fama/French three-factor and the Carhart four-factor model) of returns and net out any factors known to impact returns. We find that adherence to Benford's Law does not impact the distribution of parameter estimates associated with known return factors. We find no evidence that funds whose return digits violate Benford's Law perform abnormally worse or abnormally better than funds which do not. This result is robust across test and across fund strategy. Because one would expect funds which violate Benford's Law to perform better if their returns were being manipulated, we take this result as an indicator of Type I error.

We also find evidence of sizable differences in rejection rates depending on which Benford's test is used. Rejection rates are relatively strongly positively correlated, but differences in rejection rates are nonetheless substantial within sample. These differences have substantial practical consequences for the researcher, as demonstrated in our sub-sample analysis of abnormal returns for mid-cap funds. Even within our data, different tests flagged different returns in a way where one would possibly conclude a potential for fraud from one test, but not from the other. Simply put, researchers should consider the possibility that rejections of Benford's Law are not due to fraudulent data, but may instead be due to Type I error.

References

[1] Gordon Alexander, Gjergji Cici, and Scott Gibson. Does Motivation Matter When Assessing Trade Performance? An Analysis of Mutual Funds. *Review of Financial Studies*, 20, 2007.

[2] Dan Amiram, Zahn Bozanic, and Ethan Rouen. Financial Statement Errors: Evidence from the Distributional Properties of Financial Statement Numbers. *Review of Accounting Studies*, 20:1540–1593, 2015.

[3] Brad Barber, Xing Haung, and Terrence Odean. Which Risk Factors Matter to Investors? Evidence from Mutual Fund Flows. 2016. unpublished manuscript.

[4] Nicolas Bollen and Veronica Pool. Do Hedge Fund Managers Misreport Retruns? Evidence from the Pooled Distribution. *Journal of Finance*, 64(5):2257–2288, 2009.

[5] Nicolas Bollen and Veronica Pool. Suspicious Patterns in Hedge Fund Returns and the Risk of Fraud. *Review of Financial Studies*, 25(9):2673–2702, 2012.

[6] Colin Camerer. *Behavioral Game Theory*. 2003.

[7] Mark Carhart. On Persistence in Mutual Fund Performance. *Journal of Finance*, 52(1):57–82, 1997.

[8] Judith Chevalier and Glenn Ellison. Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy*, 105:1167–1200, 1997.

[9] Wendy K. Tam Cho and Brian J. Gaines. Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance. *The American Statistician*, 61:218–223, 2007.

[10] W. G. Cochran. The Chi-Squared Test of Goodness of Fit. *Annals of Mathematical Statistics*, 23:315–345, 1952.

[11] W.J. Conover. *Practical Nonparametric Statistics, Third Edition*. Wiley, New York, New York USA, 1999.

[12] Cindy Durtschi, William Hillison, and Carl Pacini. The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting*, 5:17–34, 2004.

[13] Eugene Fama and Kenneth French. Common Risk Factors in the Returns on Bonds and Stocks. *Journal of Financial Economics*, 33:3–53, 1993.

[14] Anton Formann. The Newcomb-Benford Law in Its Relation to Some Common Distributions. *PLOS one*, 5(5), 2010.

[15] Scott Gibson, Assem Safieddine, and Sheridan Titman. Tax-Motivated Trading and Price Pressure: An Analysis of Mutual Fund Holdings. *Journal of Financial and Quantitative Analysis*, 35:369–386, 2000.

[16] David E. Giles. Benford's Law and Naturally Occurring Prices in Certain eBay Auctions. *Applied Economics Letters*, 14(3):157–161, 2007.

[17] Stefan Günnel and Karl-Heinz Tödter. Does Benford's Law Hold in Economic Research and Forecasting? *Empirica*, 36(3):273–292, 2009.

[18] Theodore Hill. Base-Invariance Implies Benford's Law. *Proceedings of the American Mathematical Society*, 123(3):887–895, 1995.

[19] Ben Jann. Multinomial Goodness-Of-Fit: Large-Sample Tests with Survey Design Correction and Exact Tests for Small Samples. *The Stata Journal*, 8(2):147–169, 2008.

[20] Elise Janvresse and Thierry de la Rue. From Uniform Distributions to Benford's Law. *Journal of Applied Probability*, 41:1203–1210, 2004.

[21] Philippe Jorion and Christopher Schwarz. Are Hedge Fund Managers Systematically Misreporting? Or Not? *Journal of Financial Economics*, 111:311–327, 2014.

[22] George Judge and Laura Schechter. Detecting Problems in Survey Data Using Benford's Law. *Journal of Human Resources*, 47(1):1–24, 2009.

[23] Eduardo Ley. On the Peculiar Distribution of the U. S. Stock Indices' Digits. *The American Statistician*, 50:311–313, 1996.

[24] Asger Lunde, Allan Timmerman, and David Blake. The Hazards of Mutual Fund Underperformance: A Cox Regression Analysis. *Journal of Empirical Finance*, 6:121=152, 1999.

[25] Tomasz Michalski and Gilles Stoltz. Do Countries Falsify Economic Data Strategically? Some Evidence That They Might. *Review of Economics and Statistics*, 95(3):591–616, 2013.

[26] Mark Nigrini. Taxpayer Compliance Application of Benford's Law. *Journal of the American Taxation Association*, 18:72–92, 1996.

[27] Mark Nigrini. I've Got Your Number. *Journal of Accountancy*, pages 79–83, 1999.

[28] Mark Nigrini. *Benford's Law: Applications for Forensic Accounting, Auditing, and Fraud Detection*. Wiley, 2012.

[29] Mark J. Nigrini and Linda J. Mittermaier. The Use of Benford's Law as an Aid in Analytical Procedures. *Auditing: A Journal of Practice and Theory*, 16(2):52–67, 1997.

[30] Sofia Villas-Boas Rosa Abrantes-Metz and George Judge. Tracking the LIBOR Rate. *Applied Economics Letters*, 18(10):893–899, 2011.

[31] Ik Seon Suh, Todd C. Headrick, and Sandra Minaburo. An Effective and Efficient Analytic Technique: A Bootstrap Regression Procedure and Benford's Law. *Journal of Forensic and Investigative Accounting*, 3:25–45, 2011.

Table 1: m, d, χ^2, K-S, and Kuiper test results for Benford's Law: Full Sample

	m	d	χ^2	K-S	Kuiper
Full Sample	9.30%	15.46%	15.47%	16.87%	11.51%

Each cell reports the percentage of times that Benford's Law is violated at the 99 percent confidence level for a selected sample of mutual funds. The total sample contains 52,407 funds.

Table 2: $z-$test results for Benford's Law: Full Sample, Deciles, and Large Samples

Number of Violations	0	1	2	3	4	5+
Full Sample	69.55%	22.50%	4.99 %	1.72 %	.70%	.54%

Each cell reports the percentage of digit violations associated with Benford's Law at the 99 percent confidence level for a selected sample of mutual funds. The total sample contains 52,407 funds.

Table 3: m, d, χ^2, K-S, and Kuiper test results for Benford's Law: Individual Deciles, and Large Samples

	m	d	χ^2	K-S	Kuiper
90th+ percentile	1.22%	5.52%	5.29%	7.13%	7.56%
80th+ percentile	2.10%	13.93%	11.53%	15.96%	8.13%
70th+ percentile	2.90%	15.15%	11.13%	15.19%	7.82%
60th+ percentile	6.54%	14.90%	10.07%	13.99%	7.25%
50th+ percentile	11.16%	16.50%	11.56%	15.00%	9.16%
40th+ percentile	11.79%	18.09%	14.27%	13.87%	11.26%
30th+ percentile	13.05%	18.76%	17.72%	16.43%	13.66%
20th+ percentile	24.58%	29.73%	29.66%	26.99%	25.99%
10th+ percentile	17.59%	19.33%	20.20%	19.86%	19.79%
lowest percentile	2.02%	2.69%	3.32%	4.62%	4.47%

Each cell reports the percentage of times that Benford's Law is violated at the 99 percent confidence level for a selected sample of mutual funds. The total sample contains 52,407 funds. Deciles are equally spaced and contain roughly 5,240 returns each. Because mutual funds vary in the number of returns, we use Monte Carlo approximation to compute p-values associated with the K-S test. The Kuiper statistic is modified by the number of returns for the individual mutual fund as suggested in Stephens (1970).

Table 4: $z-$test results for Benford's Law: Deciles

Number of Violations	0	1	2	3	4	5+
90th+ percentile	77.35%	20.21%	2.08%	.36%	0%	0%
80th+ percentile	70.73%	22.84%	5.38%	.90%	.13%	.02%
70th+ percentile	70.83%	22.69%	5.48%	.86%	.15%	0%
60th+ percentile	70.67%	24.56%	4.31%	.44%	.02%	0%
50th+ percentile	69.93%	24.10%	5.19%	.63%	.11%	.02%
40th+ percentile	68.09%	24.31%	5.61%	1.55%	.29%	.15%
30th+ percentile	65.88%	23.56%	6.70%	2.92%	.55%	.39%
20th+ percentile	57.13%	22.51%	8.74%	5.65%	3.30%	2.63%
10th+ percentile	64.97%	21.68%	4.96%	3.87%	2.44%	2.08%
lowest percentile	79.96%	18.53%	1.45%	.06%	0%	0%

Each cell reports the percentage of times that Benford's Law is violated at the 99 percent confidence level for a selected sample of mutual funds. The total sample contains 52,407 funds. Deciles are equally spaced and contain roughly 5,240 returns each.

Table 5: m, d, χ^2, K-S, and Kuiper test results for Benford's Law: Selected Fund Objectives

	m	d	χ^2	K-S	Kuiper
Equities					
Domestic (small-cap)	.71%	19.52%	24.73%	15.44%	12.65%
Domestic (mid-cap)	.85%	15.53%	17.41%	11.8%	6.84%
Domestic (index)	.59%	4.72%	7.37%	5.6%	2.06%
Domestic (growth)	1.05%	9.06%	11.36%	8.5%	4.82%
Domestic (sector-based)	3.96%	11.69%	9.29%	9.21%	6.67%
Foreign	1.03%	9.09%	7.96%	6.35%	3.92%
Fixed Income					
Foreign	6.88%	8.32%	9.75%	7.39%	4.62%
Government	12.47%	16.13%	14.68%	15.07%	11.11%
Money-Market	82.99%	86.42%	81.58%	88.39%	87.56%

Each cell reports the rejection rates for different subsets of mutual funds depending on objective. Our data contain 2,850 domestic small-cap funds, 2,137 domestic mid-cap funds, 341 domestic index funds, 7,299 growth funds, 2,549 sector-based funds, 7,023 foreign funds, and 2,651 money market funds.

Table 6: m, d, χ^2, K-S, and Kuiper test results for Benford's Law: Dead Funds

	m	d	χ^2	K-S	Kuiper
Dead Funds (full sample)	9.89%	15.72%	16.44%	17.73%	11.26%
Delisted Funds	11.15%	16.23%	21.45%	21.63%	14.16%
Merged Funds	9.96%	15.97%	16.81%	18.87%	10.07%

Each cell reports the rejection rates for different subsets of mutual funds depending on whether the fund was delisted or was merged. Our data contain 3,365 funds formally defined as delisted and 5,451 funds that were merged into other funds. The full sample of dead funds contains 17,254 funds.

Table 7: Correlation Matrix Between Rejection Rates: m, d, χ^2, K-S, and Kuiper Tests

Full Sample	K-S	m	d	χ^2	Kuiper
K-S	1				
m	.4831	1			
d	.6175	.6598	1		
χ^2	.6111	.545	.7186	1	
Kuiper	.6584	.5847	.6974	.7217	1
$n > 110$					
K-S	1				
m	.4127	1			
d	.5843	.5728	1		
χ^2	.5901	.4578	.6819	1	
Kuiper	.6150	.5032	.6753	.7404	1

Each cell reports the Pearson's correlation coefficient between the rejections of Benford's Law for two different goodness-of-fit measures. The total sample contains 52,407 funds. The second panel restricts the sample to funds with more than 110 returns, and consists of 15,883 funds.

Figure 1: Distributions of Fitted Alphas, Three-Factor and Four-Factor Models

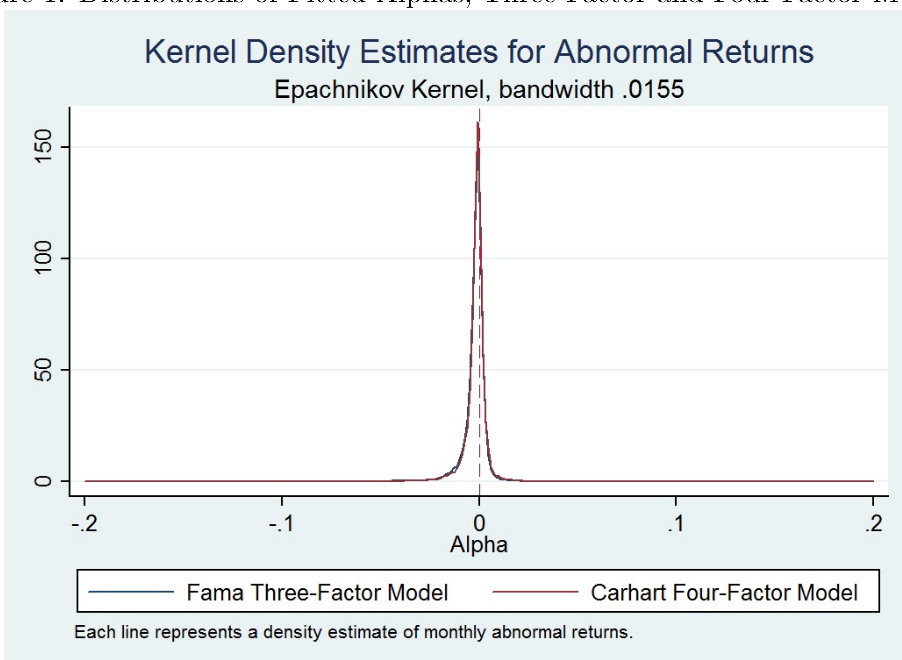

Figure 2: Distributions of Fitted Alphas When Fund Violates Benford's Law, Three-Factor Model

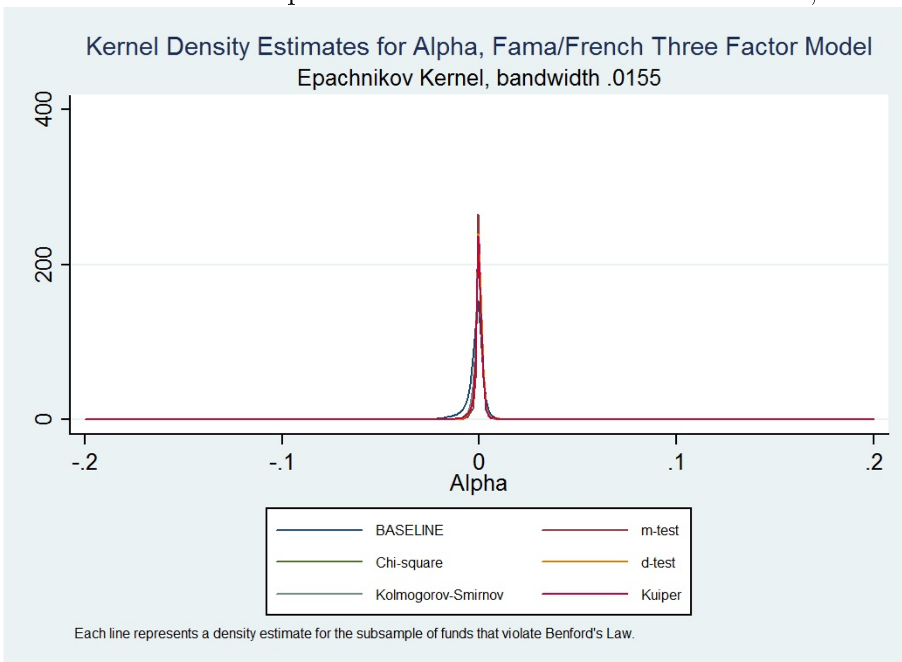

Figure 3: Distributions of Fitted Alphas When Fund Violates Benford's Law, Four-Factor Model

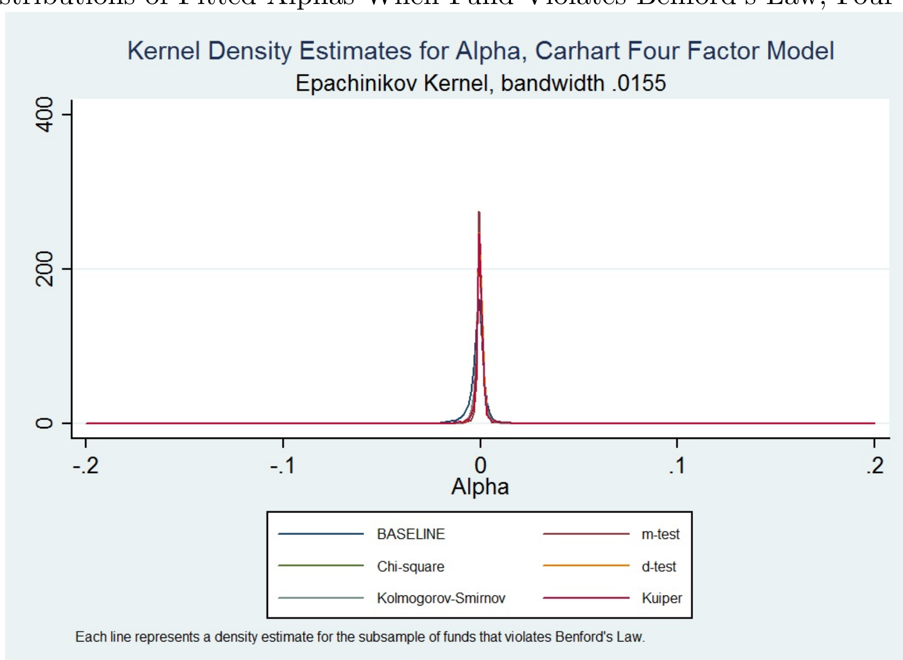

Figure 4: Distributions of Fitted Alphas When Fund Violates Benford's Law, Style Subsamples

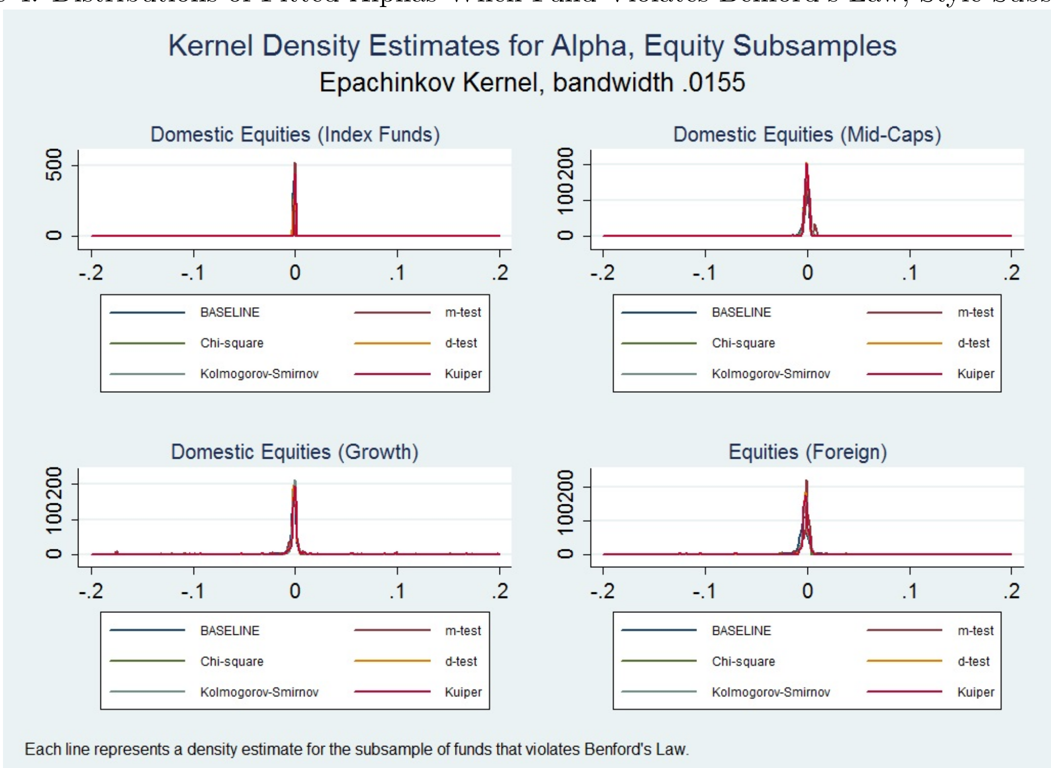

Figure 5: Distributions of Four Factor Estimates When Fund Violates Benford's Law