ИССЛЕДОВАТЕЛЬСКАЯ СТАТЬЯ
https://doi.org/10.31063/AlterEconomics/2023.20-3.10
УДК 336.027
JEL G3; L2

Проверка достоверности финансовой отчетности европейских компаний законом Бенфорда[1]

Варвара В. НАЗАРОВА[1)], Ийя Ю. ЧУРАКОВА[2)], Дмитрий А. КУПРИЯНОВ[3)]
[1, 2, 3)] Национальный исследовательский университет «Высшая школа экономики», г. Санкт-Петербург, Российская Федерация

Для цитирования: Назарова, В. В., Чуракова, И. Ю., Куприянов, Д. А. (2023). Проверка достоверности финансовой отчетности европейских компаний законом Бенфорда. *AlterEconomics, 20*(3), 691–711. https://doi.org/10.31063/AlterEconomics/2023.20-3.10

Аннотация. Манипуляции финансовой отчетностью и достоверностью представленных в ней данных встречаются в экономике достаточно часто, и с 2001 г., со скандала с Enron, количество манипуляций продолжает расти. Компании стараются представить свои данные с наиболее выгодной позиции, увеличивая цену акции и привлекательность для инвесторов, используя порой неэтичные методы. Закон Бенфорда — это математический инструмент, открытый еще в 1938 г., а с 1995 г. примененный математиком М. Нигрини для анализа налоговых деклараций. В данном исследовании он используется как метод определения возможных фальсификаций бухгалтерской отчетности. Закон Бенфорда, или закон первой цифры, состоит в том, что вероятность появления в массивах данных чисел, начинающихся с единицы, в три раза больше, чем это предсказано нормальным распределением, и эту вероятность можно просчитать. Целью данного исследования является проверка данных европейских компаний на достоверность при помощи закона Бенфорда. В статье рассмотрены теоретические аспекты, связанные с законом Бенфорда, проанализированы важнейшие исследования по теме и рассмотрены методы, с помощью которых можно проверить данные отчетности на достоверность. Работа включает в себя анализ данных европейских компаний. На его основе создан алгоритм, благодаря которому удалось проверить данные отчетности и оценить степень их соответствия частотам Бенфорда. Данный подход актуален для проверки отчетности по отраслям, странам и отдельным компаниям. По результатам определены европейские страны и сферы деятельности, в которых наиболее часто выявляются фальсификации или умышленные корректировки финансовой отчетности.

Ключевые слова: закон Бенфорда, бухгалтерская отчетность, мошенничество, массив данных, фальсификация отчетности

RESEARCH ARTICLE

Assessing Financial Statement Reliability in European Companies Using Benford's Law

Varvara V. NAZAROVA[1] ✉, **Iya Yu. CHURAKOVA**[2], **Dmitriy A. KUPRIYANOV**[3]
[1, 2, 3] *HSE University, St. Petersburg, Russian Federation*

For citation: Nazarova, V. V., Churakova, I. Yu., & Kupriyanov, D. A. (2023). Assessing Financial Statement Reliability in European Companies Using Benford's Law. *AlterEconomics, 20*(3), 691–711.
https://doi.org/10.31063/AlterEconomics/2023.20-3.10

Abstract. The manipulation of financial reporting and data reliability in statements has become increasingly prevalent since the Enron scandal in 2001, with companies employing various strategies to enhance their stock prices and attract investors, occasionally resorting to unethical practices. Benford's Law, a mathematical tool dating back to 1938 and popularized by mathematician Nigrini in 1995 for analyzing tax declarations, is employed in this study as a means of identifying potential accounting irregularities. Benford's Law, also known as the "law of the first digit", asserts that numbers starting with one are three times more likely to appear in datasets than predicted by a normal distribution, with calculable probabilities. This study aims to validate European companies' data using Benford's Law. The paper explores the theoretical underpinnings of Benford's Law, reviews key research on the subject, and examines methods for assessing the reliability of reporting data. Additionally, the paper contains an analysis of data from European companies. Based on this analysis, an algorithm has been developed, enabling the verification of reporting data and the assessment of their compliance with Benford's frequencies. This approach proves valuable for examining financial reporting in various industries, countries, and individual companies. The findings help identify which European countries and sectors have the highest occurrences of financial statement falsification or intentional adjustments.

Keywords: Benford's law, financial statements, fraud, datasets, falsification of reporting

1. Введение

Представленные в финансовой отчетности компании данные напрямую влияют на цену ее акций и привлекательность компании для инвесторов. Несмотря на то, что такая отчетность проверяется аудиторскими фирмами в соответствии с международными стандартами, существует возможность манипуляции отчетностью, что ставит под угрозу не только инвесторов компании, но и ее менеджеров, аудиторскую фирму, а также снижает доверие к финансовому рынку в целом. Особенно же актуальной тема мошенничества с финансовой отчетностью стала после 2001 г., когда о преступлении Enron стало известно широкой общественности. Число подобных махинаций продолжает расти с каждым годом (Statista, 2020), и компании могут как завышать показатели финансовой отчетности, преувеличивая свою ценность для инвесторов или банков, так и занижать их, минимизируя налоговые платежи.

Закон Бенфорда является одним из методов, используемых при проверке данных на предмет их достоверности. Он определяет частоты значащих цифр в числах и предполагает, что поддельные числа имеют иную структуру, нежели действительные или случайные выборки. Согласно закону, первые цифры чисел подчиняются логарифмическому распределению и встречаются в числах не с одинаковой, а с разной вероятностью. Так, первая цифра любого числа (при соблюдении условий закона Бенфорда) равна единице в 30,1 % случаев, двойке — в 17,6 % случаев и т. д. до девяти. Это открытие было сделано астрономом Ньюкомбом в 1881 г., и сегодня этот закон используется во многих сферах, включая данные голосования на выборах, статистику больных COVID-19, данные о смертности населения и т д. В сфере аудита закон Бенфорда также широко применяется, и, хотя он не гарантирует 100 % вероятность искажения данных, это важный показатель, который может стать одним из тревожных сигналов и поводом к дальнейшей подробной проверке финансовой отчетности компании.

Метод проверки с использованием закона Бенфорда не зависит от вида финансовой отчетности, практически не зависит от специфики данных и может применяться при анализе большого массива данных, в т. ч. данных нескольких или нескольких тысяч отчетов одновременно. Это позволяет анализировать не только конкретную компанию или группу, но и целые отрасли или страны, как это было сделано в исследовании Хертели, в котором на мошенничество были проверены все румынские компании в сфере гостиничного бизнеса (Herteliu et al., 2021).

В прошлых исследованиях по теме основное внимание было сконцентрировано на методологии (ее улучшение и дополнение, к примеру, с помощью машинного обучения) или на проверке определенных компаний и их отчетностей. Продолжая идею Хертели, в текущем исследовании было решено проанализировать все европейские компании в целом и по различным категориям (страны, сферы), увеличив также время наблюдений (9 лет вместо одного года) (Herteliu et al., 2021).

Целью данного исследования является проверка данных европейских компаний на достоверность при помощи закона Бенфорда.

2. Теоретические аспекты закона Бенфорда

Первая работа, которая содержала указания на факт, что числа в природе распределены неравномерно, была опубликована в 1881 г. астрономом С. Ньюкомбом. Работая со справочниками логарифмов, он начал замечать, что в каждом из них сильнее остальных изношены страницы с логарифмами более младших чисел и практически целы страницы с логарифмами чисел, начинающихся на 8 или 9. Распределение для первых цифр чисел, предложенное Ньюкомбом, выглядело так, как представлено на рисунке 1.

Dig.	First Digit.	Second Digit.
0	. . .	0.1197
1	0.3010	0.1139
2	0.1761	0.1088
3	0.1249	0.1043
4	0.0969	0.1003
5	0.0792	0.0967
6	0.0669	0.0934
7	0.0580	0.0904
8	0.0512	0.0876
9	0.0458	0.0850

Рис. 1. *Распределение значений из работы Ньюкомба*
Источник: цитируется авторами по материалам статьи (Benford, 1938)
Fig. 1. *Distribution of Values proposed by Newcomb*

В 1938 г. физик Ф. Бенфорд опубликовал работу, в которой продолжил исследование феномена, открытого Ньюкомбом. В своем исследовании он проанализировал случайные выборки данных из разных сфер, в т. ч. данные о площади бассейна 300 рек, вес тысячи химических соединений, списки номеров домов разных улиц и т. д., и обнаружил, что в каждой таблице единица являлась первой цифрой у чисел с вероятностью 30,1 %, т. е. не с вероятностью 1 к 9, как логично было бы предположить, а с вероятностью 1 к 3. Двойка была первой цифрой у 17,6 % чисел, три — у 12,5 % и так по нисходящей до 9 (Benford, 1938).

Исследование Бенфорда имело более практическую направленность, нежели работа Ньюкомба, и если у последнего имелись противники, считавшие, что его теория неверна, то поспорить с Бенфордом уже было намного сложнее: огромные массивы данных из разных сфер, сходящиеся к одному распределению, не оставляли вопросов. Несмотря на это, доказательство закону Бенфорда по-прежнему предложено не было, как и хоть сколько-нибудь рационального объяснения.

Следующие 50 лет множество ученых публиковали свои работы, предлагая различные объяснения данному явлению. Наиболее важные исследования были собраны Рэйми в его статье «The First Digit Problem» (Raimi, 1976).

Первая категория работ, которую выделяет Рэйми, — работы, исследующие причины, по которым закон существует, область его применения и действия. Одной из таких работ является исследование В. Фурри и Г. Гурвица, в которой действие закона объясняется способом написания нами чисел (Furry & Hurwitz, 1944). Работа Пинхэма идейно продолжает работу Г. Гурвица и В. Фурри, представляя собой «теоретическую дискуссию», как называет ее сам автор, в которой анализируется область применения закона и предпринимается попытка объяснить причины его существования в природе (Pinkham, 1961). Кроме того, обозначается следующая закономерность: единственным распределением вероятности появления для первых значащих цифр, которое остается инвариантным относительно изменения масштаба базового распределения, является $\log_{10}\left(1+\frac{1}{n}\right)$, где n — первая значащая цифра числа, от 1 до 9. Здесь же было определено, что закону Бенфорда удовлетворяют любые односторонние устойчивые распределения, описываемыми четырьмя параметрами, два из которых определяют сдвиг. Благодаря работе Р. Пинхэма в дальнейшем удалось описать свойство закона Бенфорда, которое позволило применить его в новых сферах, таких как аудит.

В статье К. Адхикари и П. Саркара исследуется свойство наиболее значимой цифры (Adhikari & Sarkar, 1968). Доказано, что при возведении случайных чисел или их взаимно обратных чисел в степень все большей и большей степени они имеют в пределе логарифмическое распределение по значащей цифре. Это свойство проявляется в пределе и для произведений случайных чисел, поскольку число членов в произведении становится все больше и больше. Однако это свойство не проявляется в пределе для высших корней случайных чисел или их взаимно обратных корней. В действительности имеет место концентрация на некоторой определенной цифре. Делается вывод, что если случайная величина X имеет логарифмическое распределение старшей значащей цифры, то это же распределение имеют множества $1/X$ и CX, где C — любая константа. Таким образом, становится возможным сформулировать следующее свойство: если множества чисел $x_1, x_2, x_3 \ldots, x_n$ следуют закону Бенфорда, то множества чисел $1/X$ и c/X, где $c > 0$, также следуют ему.

В работе было открыто еще одно интересное свойство, которое в будущем стало востребовано авторами, использующими в бухгалтерском учете и проверке финансовой отчетности закон Бенфорда: если X — случайная величина, имеющая равномерное распределение в интервале $[1;10)$, а $Y = X^n$, где n — целое число, то верно следующее:

$$\lim_{n\to\infty} Prob\left(D_1, D_2 = d_1, d_2\right) = \log\left(1 + \frac{1}{d_1 d_2}\right),\ d_1 d_2 \in \{10, 11, \ldots, 99\}.$$

То есть если мы умножаем случайные величины, которые равномерно распределены, то по мере увеличения числа умножений множество будет стремиться стать набором Бенфорда. Количество умножений (или итераций), проведенное до того момента, когда множество будет практически не отклоняться от частот Бенфорда, зависит от диапазона выборки. К примеру, для выборки с распределением в интервале [7,5;9) потребуется намного больше итераций, чем выборке в интервале [4;9,5).

Следует отметить, что в литературе выделены также области, где закон Бенфорда не может применяться. В первую очередь это величины, размах вариации которых сравнительно мал (рост или IQ людей, например, не может иметь значительных различий в порядках чисел), искусственные массивы упорядоченных данных (например, номера домов на улице), цены на биткойны (но не объемы торгов), некоторые последовательности молекулярных весов в химии, а также данные, сгенерированные с помощью математических формул (Hasan, 2002; Sukanto, 2011). Ограничения статистических моделей в использовании закона Бенфорда для выявления различного рода искажений отчетности компаний также могут быть связаны с особенностями моделей, недостаточно чувствительных к отклонениям от выбранного закона распределения (Barabesi & Pratelli, 2020). Так как наборы чисел, не подчиняющихся закону Бенфорда, выявляются в основном эмпирически, то рекомендуется использовать соответствие данному закону лишь в качестве первых фильтров, дополняя другими статистическими методами, например, кластерным анализом (Зверев, Никифоров, 2018; Balashov et al., 2021).

Тем не менее внимание к использованию закона Бенфорда в финансах и аудите для выявления нарушений, аномалий различного рода, ошибок или мошенничества, которые могут присутствовать в наборе финансовых данных, растет.

3. Использование закона Бенфорда в аудите

В области аудита закон Бенфорда начал активно использоваться с 1980–х гг. Одной из наиболее значимых и фундаментальных стала работа М. Нигрини 1992 г. В связи с увеличивающейся проблемой уклонения от налогов он начал разработку новых методов проверки данных на достоверность. Он выделил четыре категории существующих на тот момент исследований в сфере аудита: обзорные исследования, аналитические исследования, а также исследования, использующие экспериментальную методологию и регрессионный анализ. Для аналитических исследований в простых моделях можно наблюдать четкие результаты определения того, какие факторы и в какой степени влияют на уровень уклонения от налогов. В опросах люди могут лгать и замалчивать какую-то информацию; в исследованиях, использующих экспериментальную методологию, может наблюдаться искажение данных по мере изменения условий эксперимента или выборки. Грамотно проведенное исследование в этой области совсем не гарантирует соответствия с практикой, т. к. имеет в основном теоретическую направленность. Проблемы с существующими исследованиями побудили М. Нигрини разработать новые методы анализа данных и, пытаясь ответить на вопрос «существует ли связь между используемыми числами (цифрами, составляющими числа) в налоговых декларациях и уклонением от уплаты налогов?», он создал модель, основанную на распределении частот в соответствии с законом Бенфорда, проанализировал данные с учетом коэффициента искажения. Результатом этой работы стал вывод, что группы с низким доходом чаще практикуют незапланированное уклонение от уплаты налогов, чем группы с высоким доходом, а также тот факт, что методология, основанная на значащих цифрах, может быть использована для выявления искажений в бухгалтерском учете. Помимо упомянутых результатов Нигрини определяет еще ряд областей, в которых может быть использован этот метод, а также описывает, как он может быть использован при обнаружении аномалий в больших объемах данных и применен к подмножествам более крупных наборов, чтобы сузить список подозрительных элементов и сделать процесс более точным (Nigrini, 1996; 2017; Nigrini & Mittermaier, 1997).

Спустя пять лет М. Нигрини публикует следующую работу, которая продолжает начатое в 1992 г. исследование, и предлагает шесть разработанных тестов для проверки, сценарии и области их применения, описывает каждый из них, а также указывает на ограничения их использования и дает указания для будущих исследователей. Автор указывает, что данные, которые следуют закону Бенфорда, скорее всего, будут иметь следующие характеристики:

— данные не содержат встроенных минимумов или максимумов (например, часовая ставка по зарплате или взносы на пенсионный счет);

— данные включают показатели, описывающие универсальные характеристики;

— показатели не являются индексами или номерами (к примеру, номер банковского счета, номер социального страхования, телефонный номер и т. д.);

— данные не являются тесно сгруппированными вокруг одного значения.

Также М. Нигрини в своей работе выделяет возможные сценарии использования закона Бенфорда. В их числе проверка: кредиторской задолженности, дубликатов платежей, информации из общей бухгалтерской книги, конвертированных данных в компьютерных системах, возвратов средств клиенту, эффективности обработки при большом количестве транзакций или при небольшой сумме, новых комбинаций отпускных цен и т. д.

Представленный ученым набор тестов для проверки состоял из 7 типов, каждый имел свои особенности и ограничения. В последующих научных публикациях эти тесты многократно использовались, проверялись и дополнялись. В базовом варианте существуют 7 тестов.

1. На первую значащую цифру. Анализируются частоты первых цифр в числах по закону Бенфорда.

2. На вторую значащую цифру. Анализируются частоты вторых цифр в числах по закону Бенфорда.

3. На первую и вторую значащие цифры по отдельности. Анализируются частоты первой значащей цифры в диапазоне от 1 до 9 и частоты второй значащей цифры в диапазоне от 0 до 9 по отдельности, затем составляется таблица соответствий и сравнивается с таблицей, согласующейся с законом Бенфорда.

4. На первые две комбинированные значащие цифры. Анализируются частоты не первой и второй цифры по отдельности, а вместе. Таким образом, таблица соответствия включает в себя значения от 10 до 99. Данный тест подходит для массивной выборки.

5. На первые три комбинированные значащие цифры. По аналогии с предыдущим тестом анализируются тройки цифр, таблица соответствия включает значения от 100 до 999. Для такого теста выборка должна быть еще больше, не менее 9 тысячи пунктов.

6. На дубликаты. Дополнительный тест, который опирается на методологию закона Бенфорда, но прямо не использует его. Анализируется частота появления дубликатов в выборке, затем данные сортируются по частоте повторов и проверяется плотность ряда из повторяющихся чисел. Тест на дубликаты можно использовать во внешнем аудите, при налоговых проверках и подобном.

7. На последние несколько цифр. Анализируются последние значащие цифры (одна, две или три), их частота сравнивается с частотами округления в большую или меньшую сторону по закону Бенфорда. Является менее популярным тестом ввиду неточности.

Дальнейшие исследования дополнили и развили предыдущую методологию. В работе У. Хиллисона было выяснено, в каких случаях закон может быть использован наиболее эффективно, а в каких аудиторам следует проявлять осторожность при его применении; какие виды мошенничества могут быть обнаружены с его помощью и потенциальные проблемы, которые возникнут, если будет недостаточно данных (Hillison et al., 2004).

Работа З. Кракара доказала использование закона Бенфорда и разработанной методологии при аудите информационных систем, в частности иностранных платежных систем (Krakar & Žgela, 2009). В исследование были включены данные хорватских банков за 2008 г., и было проанализировано в общей сложности 1 745 311 строк данных. В исследовании использовались три метода верификации: Хи-квадрат, Z-статистика и MAD. Авторами было обнаружено, что результаты теста MAD практически не зависят от размера выборки, следовательно, он может использоваться как с очень маленькими, так и с очень большими наборами данных, и, в отличие от хи-квадрат и Z-статистики, более понятен для аудиторов. Однако строгих и общепринятых предельных значений для MAD по-прежнему не существует, поэтому он наиболее эффективен при использовании в сочетании с первыми двумя тестами (Torres & Pericchi, 2011; Barney, 2016).

В 2013 г. Ф. Алали и С. Ромеро доказали возможность применения закона Бенфорда в сфере выявления признаков банкротства (Alali & Romero, 2013). Анализу подверглись данные американских коммерческих банков и их отчетность за последние 10 лет. Среди рассматриваемых обанкротившихся банков большинство демонстрировали показатели, несоответствующие закону Бенфорда, вследствие этого была предсказана функция выживания менее 50 % банков при несоответствии частот эталонному распределению. Кроме того, было замечено, что больше всего не соответствовали данные о дебиторской задолженности, средствах производства и имуществе компаний, и во многих случаях имело место положительное отклонение, что значит завышение показателей предприятия в отчетностях, в частности, стоимости активов.

Другое крупное исследование было проведено в 2017 г., где была затронута тема непрерывного аудита и то, как там может быть применен закон Бенфорда (Silva et al., 2017). Выборку исследования составили 210 899 контрактов, выданных шестьюдесятью управленческими подразделениями в двух штатах на северо-востоке Бразилии в 2010 г. Непрерывный аудит возможен благодаря алгоритмам, которые проверяют постоянно меняющийся набор данных в режиме реального времени. Отслеживая транзакции таким образом, организации могут снизить финансовые потери от рисков мошенничества. Авторам удалось определить отклонение от распределения частот во времени и увидеть, в какие конкретно моменты происходили максимальные пики, что позволило найти отдельные области, которые должны быть дополнительно обработаны аудиторами. Таким образом, ученые доказали, что закон Бенфорда может быть полезен и эффективно использован не только в обычном, но и в непрерывном аудите (Bhattacharya et al., 2011; Azevedo et al., 2021).

В 2017 г. М. Нигрини опубликовал еще одну статью, в которой проанализировал опубликованную литературу, исследующую закон Бенфорда и его применение, дополняя и критикуя их, а также представил очень хорошо структурированный обзор проблем и подводных камней, связанных с выборкой аудиторов, обобщив текущее состояние дел и добавив свои собственные идеи (Nigrini, 2017). В частности, он высказал мнение относительно тестов Хи-квадрат, Z-статистики и MAD, о которых будет оговорено далее в работе.

Самые последние статьи по теме развивают методологию закона Бенфорда, используя возможности нейронных сетей и машинного обучения, как в статье о проведенной проверке в 2017 г. в Испании (Badal-Valero, 2017). Это исследование основано на анализе базы данных из макро-дел (*macro-case*) об отмывании денег компанией и рядом ее поставщиков, некоторые из которых уже были признаны полицией мошенническими компаниями. Целью их работы было объединить закон Бенфорда и новые методы машинного обучения для создания мощного инструмента проверки. Они тестировали разные методы и виды нейронных сетей, в частности методы дерева решений, случайных лесов, экономичного обучения и SMOTE, которые широко используются в машинном обучении. После проверки каждой модели они пришли к выводу, что наиболее эффективным инструментом для их специфической выборки является модель SMOTE. Результатом исследования также стал список потенциальных поставщиков, которые изначально не были признаны мошенническими и к которым следует применить индивидуальную проверку. Таким образом, в то время как у автора имелись данные о компаниях, о которых было известно, что лишь 4 % из них мошенничали с отчетностью, обученные нейронные сети нашли еще ряд подобных компаний в выборке.

Кроме вышеупомянутых исследований, которые дополняют методологическую базу закона Бенфорда, было проведено огромное количество практических работ, в ходе которых были изучены счета в отчетах и после которых были обнаружены бухгалтерские махинации (Goncalves, 2021; Whitney, 1972; Wallace, 2002; Azevedo et al., 2021). Одно из таких исследований было проведено в Румынии (Herteliu et al., 2021). В нем авторы решили проверить, в какой степени данные о прибыли в румынском гостиничном бизнесе соответствуют закону Бенфорда. По мнению авторов, в гостиничном бизнесе в этой стране наблюдается высокий уровень уклонения от уплаты налогов, и одними из главных причин этого являются большая распространенность наличных при расчетах и коррупция, характерная именно для «сферы гостеприимства». Гипотеза заключалась в том, что данные о прибыли не будут соответствовать распределению Бенфорда, но результаты показали обратное — распределение почти не отклоняется от эталонных значений. Данные для анализа были взяты из открытых источников и отфильтрованы, т. к. многие компании на момент проведения работы не были активны, но числились в статистике. В общей сложности объем набора данных составил 28 350 компаний.

4. Методы расчета частот эмпирических распределений

В данной работе мы остановимся на тестах, приведенных выше: первом, втором и четвертом. Тесты на первую и вторую значащие цифры имеют распределение частот по Бенфорду, приведенное в таблице 1.

Формула для расчета вероятностей для первого теста (1):

$$P(D_1 = d_1) = \log_{10}\left(1+\frac{1}{d_1}\right), \text{ где } d_1 = (1,2,3,\ldots,9). \quad (1)$$

Это означает, что случайно выбранное число в выборке должно начинаться с цифры 1 в 30,1 % случаев (по формуле $\log_{10}(1 + 1/1)= 0{,}301$), с цифры 2 в 17,6 % случаев ($\log_{10}(1 + 1/2) = 0{,}176$) и так далее до 9.

Формула для расчета вероятностей второго теста (2):

$$P\left(D_2 = d_2\right) = \sum \log_{10}\left(1+\frac{1}{d_1 d_2}\right), \text{ где } d_2 = (1,2,3,\ldots,9). \quad (2)$$

По аналогии с предыдущим тестом это означает, что у случайно выбранного числа вторая цифра будет 0 в 11,97 % случаев, 1 в 11,39 % случаев и так далее до 9.

Для теста на первые две комбинированные цифры распределение частот приведено в таблице 2.

Таблица 1

Распределение вероятностей по Бенфорду для тестов на первую и вторую значащие цифры

Table 1

Benford Probability Distribution for First and Second Significant Digits

Тест на первую значащую цифру		Тест на вторую значащую цифру	
Разряд	**Вероятность по Бенфорду**	**Разряд**	**Вероятность по Бенфорду**
		0	11,97
1	30,1	1	11,39
2	17,61	2	10,88
3	12,49	3	10,43
4	9,69	4	10,03
5	7,92	5	9,67
6	6,7	6	9,34
7	5,8	7	9,04
8	5,12	8	8,76
9	4,58	9	8,5

Источник: составлено авторами на основе (Benford, 1938).

Таблица 2

Распределение вероятностей по Бенфорду для тестов на первые две комбинированные цифры

Table 2

Benford Probability Distribution for Tests on the First Two Joint Digits

Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду
10	4,139	26	1639	42	1,022	58	0,742	74	0,583	90	0,48

Окончание табл. 2 на след. стр.

Окончание табл. 2

Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду	Разряд	Вероятность по Бенфорду
11	3,779	27	1,579	43	0,998	59	0,73	75	0,575	91	0,475
12	3,476	28	1,524	44	0,976	60	0,718	76	0,568	92	0,470
13	3,218	29	1,472	45	0,955	61	0,706	77	0,56	93	0,464
14	2,996	30	1,424	46	0,934	62	0,695	78	0,553	94	0,460
15	2,803	31	1,379	47	0,914	63	0,684	79	0,546	95	0,455
16	2,633	32	1,336	48	0,895	64	0,673	80	0,540	96	0,450
17	2,482	33	1,296	49	0,877	65	0,663	81	0,533	97	0,445
18	2,348	34	1,259	50	0,86	66	0,653	82	0,526	98	0,441
19	2,228	35	1,223	51	0,843	67	0,643	83	0,520	99	0,436
20	2,119	36	1,190	52	0,827	68	0,634	84	0,514	—	0
21	2,020	37	1,158	53	0,812	69	0,625	85	0,508	—	0
22	1,931	38	1,128	54	0,797	70	0,616	86	0,502	—	0
23	1,848	39	1,100	55	0,783	71	0,607	87	0,496	—	0
24	1,773	40	1,072	56	0,769	72	0,599	88	0,491	—	0
25	1,703	41	0	57	0	73	0	89	0	—	0

Источник: составлено авторами на основе (Benford, 1938).

Формула для расчета вероятностей для комбинированных тестов (3):

$$P(D_1D_2 = d_1d_2) = \log_{10}\left(1+\frac{1}{d_1d_2}\right), \text{где } d_1\,d_2 = (10,11,12,13,\ldots,98,99). \qquad (3)$$

По аналогии с предыдущими тестами это означает, что в случайно выбранном числе две первые цифры будут являться парой чисел 1 и 0 в 4,14 % случаев, числами 1 и 1 в 3,78 % случаев, числами 1 и 2 в 3,48 % случаев и так далее до пары чисел 9 и 9, т. е. от 10 до 99.

Таким образом получили ожидаемые распределения по закону Бенфорда для каждого из трех тестов, и именно с ними будут сравниваться распределения частот, полученные из выборки. К примеру, для первого теста может получиться, что в выборке цифра 1 является первой цифрой у 50 % чисел, 2 — у 4 % чисел, 3 — у 15 % чисел и так далее до 9. Эти частоты сравниваются с ожидаемыми частотами по Бенфорду, и на этой основе заключается, соответствует ли распределение частот в нашей выборке частотам по Бенфорду. При проверке отклонения от ожидаемых частот можно использовать один из следующих тестов: Z-статистика; Хи-квадрат, тест Колмогорова–Смирнова или MAD (*Mean Absolute Deviation*).

Z-статистика является классическим тестом, и в первых работах по Бенфорду очень часто использовался именно он. В Z-статистике больший всплеск для меньшего набора данных может быть незначительным, а меньший всплеск для большего набора данных — значительным, поэтому тест страдает от так называемой проблемы «избыточной мощности». Она проявляется, когда выборка либо очень мала, либо очень массивна.

Тест Хи-квадрат является, по субъективной оценке, самым популярным из тестов на сегодняшний день, и, хотя также может страдать от «избыточной мощности», дает хорошие результаты. В отличие от MAD, имеет четкие критические значения и интуитивно очень понятен. Используется повсеместно, одно из последних исследований, где он применялся, — работа Хертели (Herteliu et al., 2021).

Тест Колмогорова–Смирнова основан на кумулятивной функции плотности. Он использует максимальную абсолютную разность между каждой из 90 (в данном случае) цифр.

MAD не страдает от «избыточной мощности», однако дать оценку результатам, полученным с помощью него, довольно проблематично. Четких критериев на этот счет попросту не существует. Сегодня тест MAD используют как дополнение для основных тестов и берут пороговые значения из вышеупомянутой статьи.

Учитывая методологию предыдущих исследований и специфику каждого теста, было принято решение использовать тест Хи-квадрат в качестве основного. Таким образом, при оценке отклонения от ожидаемых частот будут использованы следующие показатели:

Для первого теста $$x_{d_1}^2 = \sum_{d_1=1}^{9} \frac{\left(N_{d_1} - N_{e_{d_1}}\right)^2}{N_{e_{d_1}}},$$

Для второго теста $$x_{d_2}^2 = \sum_{d_2=0}^{9} \frac{\left(N_{d_2} - N_{e_{d_2}}\right)^2}{N_{e_{d_2}}},$$

Для третьего теста $$x_{d_1 d_2}^2 = \sum_{d_1 d_2=10}^{99} \frac{\left(N_{d_1 d_2} - N_{e_{d_1 d_2}}\right)^2}{N_{e_{d_1 d_2}}},$$

где N_d — фактические частоты первых цифр; N_{e_d} — ожидаемые частоты первых цифр по закону Бенфорда.

Полученные из формул значения будут сравниваться с критическими значениями для каждого из тестов, и если фактические значения будут превышать критические, то можно утверждать, что фактическое распределение не подчиняется закону Бенфорда (к примеру, фактическое значение для первого теста — 17,01; критическое значение — 15,51; 17,01 > 15,33 => распределение не подчиняется закону Бенфорда).

За нулевую принимается следующая гипотеза:

H_0 — эмпирическое распределение подчиняется закону Бенфорда.

При превышении критических значений в тесте нулевая гипотеза отвергается, принимается альтернативная гипотеза:

H_1 — эмпирическое распределение не подчиняется закону Бенфорда.

5. Уровень статистической значимости

После выбора метода оценки необходимо выбрать уровни значимости, при которых будут проводиться дальнейшие тесты. Уровень значимости показывает вероятность ошибки, т. е. вероятность отклонения нулевой гипотезы, когда она верна. Наиболее распространенный уровень значимости для подобных исследований при доверительной вероятности 95 % составляет 5 %. Для большей точности в тестах

Таблица 3

Критические значения теста Хи-квадрат для доверительной вероятности 95 % и 99 % и числа степеней свободы, соответствующих тестам Бенфорда

Table 3

Critical Values of the Chi-Square Test for 95 % and 99 % Confidence Levels and the Number of Degrees of Freedom Corresponding to Benford Tests

Тест	95 %	99 %	df
1 тест	15,5073	20,0902	8
2 тест	16,919	21,666	9
3 тест	112,022	122,942	89

Источник: составлено авторами.

также будет использоваться 1 %-ый уровень значимости (с 99 %-ой доверительной вероятностью). Критические значения для трех тестов приведены в таблице 3.

При проведении исследования важно также определить свойства, на которые указал Р. Бадал-Валеро в своей работе (Badal-Valero, 2017). Опираясь на работы Пинхэма и Хилла, он подмечает, что для распределения Бенфорда верны следующие закономерности: инвариантность по базе и по масштабу.

Инвариантность по масштабу означает, что при изменениях масштаба выборки, которая подчиняется закону Бенфорда, новая полученная таким путем выборка также будет подчиняться этому закону, и в этой ситуации закон Бенфорда будет выполняться для любых единиц измерения, т. е. даже если все значения выборки указаны в долларах, переводя их в рубли по определенному курсу, результаты исследования не изменятся, т. к. закон Бенфорда не чувствителен к системе измерения.

Инвариантность по базе можно определить как непрерывность распределения. Ранее в работе было упомянуто, что логарифмическое распределение является единственным непрерывным распределением, инвариантным по базе, поэтому для формулы расчета частот может использоваться не только основание 10, но и любое другое, например, двоичное основание. Из распределения частот Бенфорда можно также заметить, что чем дальше мы движемся по позициям цифр, тем сильнее они стремятся к вероятности 1/10, и в тесте на третью цифру вероятности начинают понемногу выравниваться, а в дальнейших тестах частоты могут различаться лишь на доли процентов. Поэтому тест на вторую цифру будет завершающим, т. к. дальнейшие тесты не имеют экономического смысла.

6. Массив данных, выборка и алгоритм проверки

Выборка, на базе которой проведено исследование, включает 7 557 европейских компаний за период с 2012 по 2021 гг., представленные в собрании Refinitiv Eikon database. Данные по компаниям были разделены по странам и секторам (табл. 4): Communication Services (CS), Consumer Discretionary (CD), Consumer Staples (CSt), Energy (EN), Financials (FI), Health Care (HC), Industrials (IN), Information Technology (IT), Materials (ML), Real Estate (RE), Utilities (UT). В качестве ключевого фактора исследования были приняты данные по прибыли до налогообложения (*Net Income Before Tax*). Распределение компаний по странам включало более 20 стран, лидерами по числу компаний стали страны, представленные в таблице 5. Данные компаний были подвергнуты фильтрации: был убран отрицательный знак перед показателем и исключены нулевые значения.

Таблица 4

Страны — лидеры по числу компаний в выборочной совокупности

Table 4

Top Countries by Number of Companies in the Sample

CS	CD	CSt	EN	FI	HC	IN	IT	ML	RE	UT
501	917	437	282	910	786	1449	952	536	568	219

Источник: составлено авторами по данным Refinitiv Eikon database

Таблица 5

Страны — лидеры по числу компаний в выборочной совокупности

Table 5

Top Countries by Number of Companies in the Sample

Страна	Число компаний
France	672
Poland	702
Germany	741
Sweden	926
United Kingdom	1322

Источник: составлено авторами по данным Refinitiv Eikon database.

Для каждого из трех тестов было создано отдельное распределение частот, и для одних и тех же данных распределение будет разным. К примеру, распределение для первого теста — вероятность появления первой значащей цифры (1, 2,… 9), для второго — вероятность появления второй значащей цифры (0, 1, …, 9), для третьего — вероятность появления комбинаций первых двух цифр (10, 11, … , 99). Разложив данные как вектор, появляется возможность провести тесты на первую, вторую и две комбинированные значащие цифры, а также посчитать отклонение для каждой значащей цифры, среднее отклонение и с помощью теста Хи-квадрат оценить отклонение от ожидаемых частот по Бенфорду при разных степенях значимости, таким образом подтвердив или отвергнув нулевую гипотезу.

7. Результаты исследования

Первым шагом была проверка *Net Income Before Tax* среди всех компаний за весь период. Фактически, результаты тестов показывают, подчиняются ли отчетности европейских компаний закону Бенфорда в целом.

Как можно видеть из рисунка 2 и таблицы 6, фактические частоты минимально отклоняются от Бенфорда, максимальное отклонение на цифре 6 — 6,9 %.

Таблица 6

Расчет критерия Хи-квадрат для отчетности европейских компаний

Table 6

Calculation of the Chi-Square Criterion for European Companies' Reports

Разряд	Вероятность по Бенфорду	Фактическая вероятность	Расхождение (%)	Значение теста Хи-квадрат
1	30,10	30,62	1,7	0,905
2	17,61	17,97	2,0	0,743
3	12,49	12,38	–0,9	0,111
4	9,69	9,46	–2,4	0,570

Окончание табл. 6 на след. стр.

Окончание табл. 6

Разряд	Вероятность по Бенфорду	Фактическая вероятность	Расхождение (%)	Значение теста Хи-квадрат
5	7,92	7,99	0,9	0,063
6	6,69	6,23	–6,9	3,232
7	5,80	5,66	–2,4	0,335
8	5,12	5,29	3,3	0,578
9	4,58	4,41	–3,7	0,621

Источник: составлено авторами по данным Refinitiv Eikon database.

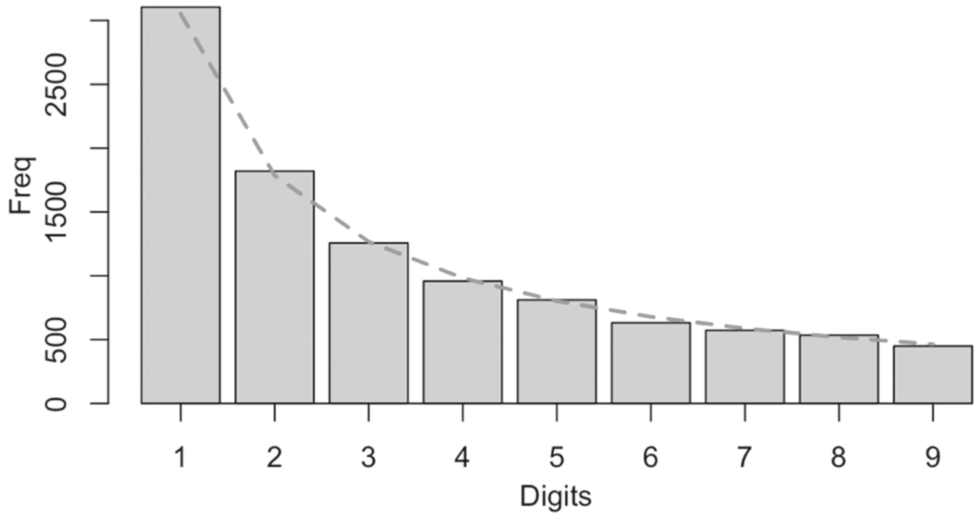

Рис. 2. *Сравнение фактических и ожидаемых частот для европейских компаний*
Источник: расчеты авторов
Fig. 2. *Comparison of Actual and Expected Frequencies for European Companies*

Сравнивая фактические значения критерия (табл. 7) и критические (табл. 3), можно увидеть, что значения Хи-квадрат не превышают критических ни в одном из тестов, таким образом, нулевая гипотеза не отвергается.

Анализ показал, что данные из таких сфер, как IT, энергетика, индустриальные, коммунальные предприятия, почти не отклоняются от распределения Бенфорда, а значения Хи-квадрат у этих секторов относительно низкие для каждого из трех тестов при любых уровнях значимости. Нулевая гипотеза не отвергается. Не соответствуют же распределению следующие секторы: сырье и материалы, недвижимость, а больше всего сферы финансов и здоровья (табл. 8). В тесте на первую зна-

Таблица 7

Фактические значения критерия Хи-квадрат для отчетности европейских компаний

Table 7

Actual Values of the Chi-Square Test for European Companies' Reports

Значения Хи-квадрат	Тест на первую цифру	Тест на вторую цифру	Тест на первые две комбинированные цифры
Фактические	7,159	7,383	98,493

Источник: составлено авторами по данным Refinitiv Eikon database

Таблица 8

Фактические значения критерия Хи-квадрат для отчетности европейских компаний в разрезе секторов

Table 8

Actual Values of the Chi-Square Test for European Companies' Reports, by Sector

Значение теста для сектора	Тест на первую цифру	Тест на вторую цифру	Тест на первые две комбинированные цифры
Information Technology	4,782	10,030	81,569
Energy	10,115	11,236	108,553
Industrials	8,634	14,201	85,870
Utilities	8,303	7,240	77,951
Financials	25,893***	9,148	141,202
Real Estate	18,367**	16759	110,914
Materials	17,754**	17,141**	113,001***
Healthcare	23,559***	24,425***	137,665***

** — коэффициенты, значимые на уровне 5 %.

*** — коэффициенты, значимые на уровне 1 %.

Источник: составлено авторами по данным Refinitiv Eikon Database.

чащую цифру компании из сфер недвижимости и сырья имеют показатели 18,367 и 17,754 соответственно, что превышает критическое значение для уровня значимости 0,05, но не превышает его при уровне значимости 0,01, поэтому можно утверждать, что распределение соответствует Бенфорду на уровне значимости 0,01, и нулевая гипотеза при данном уровне значимости не отвергается. Однако такого нельзя сказать про секторы финансов и здоровья. Показатель в сфере здоровья — 23,559, в то время как у финансов — 25,893, что не соответствует ни одному уровню значимости, из чего следует вывод: данные компаний не согласуются с законом Бенфорда, нулевая гипотеза отвергается.

Затем было решено детализировать финансовый сектор по странам и посмотреть, в каких из них прослеживается максимальное отклонение от распределения Бенфорда (табл. 9).

Анализ показал, что почти у всех стран полное соответствие с Бенфордом, у двух — Франции и Норвегии — показатели тестов были выше остальных в выборке, однако

Таблица 9

Фактические значения критерия Хи-квадрат для отчетности финансового сектора европейских компаний в разрезе стран

Table 9

Actual Values of the Chi-Square Test for the Reports of European Companies in the Financial Sector, by Country

Значение теста для сектора	Тест на первую цифру	Тест на вторую цифру	Тест на первые две комбинированные цифры
France Financials	13,56	15,77	102,21
Norway Financials	13,9	10,67	108
Switzerland Financials	32,72***	29,94***	130,87***

** — коэффициенты, значимые на уровне 5 %.

*** — коэффициенты, значимые на уровне 1 %

Источник: составлено авторами по данным Refinitiv Eikon database.

критических значений не превышали, в отличие от Швейцарии, которая не прошла ни один тест на соответствие ни при одном уровне значимости. Таким образом, лишь распределение для финансового сектора Швейцарии не согласуется с Бенфордом. Можно предположить, что имеет место манипуляция данными в отчетности швейцарских финансовых компаний или же непреднамеренное ее искажение. После изучения материалов по финансовому рынку данной страны было обнаружено несколько любопытных вещей, подтверждающих результаты анализа. Так, в 2019 г. консалтинговой компанией Commslab и исследовательским институтом Цюрихского университета был опубликован индекс репутации швейцарской экономики, который, продолжая свое падение, достиг самого низкого уровня с 2014 г. По словам исследователей, такое ухудшение было вызвано налоговой политикой государства, а также реальными случаями мошенничества, о которых стало известно общественности. Значительный ущерб репутации финансового сектора в 2018 г. был нанесен прежде всего скандалом с Райффайзенбанком, в результате которого бывший руководитель банка был привлечен к уголовной ответственности по подозрению в корпоративном мошенничестве. Кроме того, во время пандемии COVID-19 был зафиксирован рост случаев злоупотребления «*short-time working system*», при которой государство поддерживает компании средствами бюджета (по данным Государственного секретариата по экономическим вопросам Seco). В то же время упрощенные правила для получения компенсаций по этой системе подверглись критике со стороны Аудиторского бюро в 2020 г. Таким образом, результаты анализа по финансовому сектору Швейцарии подтвердились, практически же их можно интерпретировать следующим образом: инвестировать в финансовый сектор Швейцарии может быть довольно рискованно, отчетности же финансовых компаний Швейцарии могут показывать искаженные данные. Кроме того, данный результат говорит о том, что есть повод к проверке данных отчетности компаний, уже более подробной и основательной.

Кроме секторов, отдельно были рассмотрены данные по странам. У подавляющего большинства показатели согласовались с законом Бенфорда, самые высокие значения тестов были у четырех стран: Финляндии, Румынии, Франции и Швейцарии. Несмотря на относительно высокие показатели, Швейцария и Финляндия не превысили пороговых значений ни для одного из тестов, тогда как Румыния и Франция не прошли ни по одному из тестов. Из всех стран самое большое отклонение по тестам наблюдалось у Франции.

Таблица 10

Фактические значения критерия Хи-квадрат для отчетности европейских компаний в разрезе стран

Table 10

Actual Values of the Chi-Square Test for European Companies' Reports, by Country

Значение теста для сектора	Тест на первую цифру	Тест на вторую цифру	Тест на первые две комбинированные цифры
Finland	9,12	8,52	99,94
Romania	21,07***	23,95***	123,01***
France	35,18***	22,73***	132,2***
Switzerland	11,17	6,13	103,48

** — коэффициенты, значимые на уровне 5 %.

*** — коэффициенты, значимые на уровне 1 %.

Источник: составлено авторами по данным Refinitiv Eikon database.

Детализируя данные французских компаний, было обнаружено, что наибольшее несоответствие наблюдается у сектора потребительских товаров; значение Хи-квадрат превышает критические значения при всех уровнях значимости, наблюдается сильное отклонение на цифрах 3, 1, 9. Можно сделать вывод, что данные компаний этого сектора могут быть искажены и возможны определенные риски при принятии решений на основе финансовых отчетностей данных компаний. При изучении сторонних источников не было выявлено никаких определенных обстоятельств, способствующих такому результату, однако, как следствие возможных проблем, можно выделить закон, принятый в октябре 2018 г., «призванный усилить меры по борьбе с неисполнением налогоплательщиками своих налоговых и социальных обязанностей». По данному закону должна быть создана налоговая полиция, находящаяся в ведении мирового судьи и осуществляющая надзор за делами, требующего высокого уровня налоговой экспертизы и имеющих высокие бюджетные последствия. Кроме того, усиливается наказание за уклонение от налогов, а также отменяется требование о наличии заключения комиссии по налоговым правонарушениям (*verrou de Bercy*) перед подачей жалобы, т. е. упрощается процесс налоговой проверки. Как бы то ни было, результаты анализа говорят о том, что данные не соответствуют закону Бенфорда, и, как сказано выше, есть повод для дальнейшей подробной проверки данных компаний; инвесторам же следует проявлять осторожность при анализе отчетностей компаний из потребительского сектора Франции.

Что касается результатов по Румынии, то можно заметить, что в работе К. Хертелиу объектом исследования также являлись румынские компании, однако были рассмотрены данные лишь за 2008 г. (Herteliu et al., 2021). Результатом исследования стал вывод, что данные полностью согласуются с законом Бенфорда, хотя изначальное предположение автора было обратным. В текущем исследовании была взята более обширная выборка, и в результате догадка подтвердилась. Однако наибольшее отклонение было обнаружено в сфере здравоохранения, а не недвижимости. Данные не соответствуют Бенфорду при всех уровнях значимости, также можно сказать, что есть высокий риск искажения данных. При изучении сферы здравоохранения Румынии и сторонних источников было обнаружено, что в этой области высокий уровень коррупции (впрочем, как и в сфере недвижимости, о чем писал Хертелиу), что подтверждают многочисленные журналистские и уголовные расследования.

8. Заключение

Закон Бенфорда описывает распределение первых значащих цифр в данных, не имеющих никакой явной связи друг с другом. Первоначально это явление было обнаружено случайно, но сегодня оно имеет множество применений в различных областях, особенно при проверке фальсификации определенных данных, таких как финансовые отчеты. Закон Бенфорда является лишь одним из инструментов, используемых для выявления нарушений, который также может быть использован в области проверки данных в финансовой отчетности.

В ходе проведенного исследования были проанализированы данные европейских компаний за 2012–2021 гг. с детализацией по секторам и по странам. При анализе использовался закон Бенфорда и основанные на нем тесты, которые позволяют оценить достоверность рассматриваемых данных и дать оценку тому, насколько данные соответствуют или не соответствуют ожидаемому распределению. Проанализировав литературу по теме, для рассматриваемой выборки удалось подобрать наиболее эффективные методы оценки данных, после чего с помощью программного обеспечения были созданы три алгоритма, которые позволяют анализировать любые данные, оценивая их в реальном времени. Предельное число выборки составляет чуть больше 95 000 строк, что дает возможность использовать их в будущем с более массивной выборкой, результаты же исследования будут полезны будущим исследователям, решившимся взяться за анализ европейских компаний.

Несмотря на аудиторский контроль, компании по-прежнему могут манипулировать данными, и эта проблема становится все более острой с каждым годом. В данном контексте текущее исследование представляет интерес, т. к. помогает сделать выводы не о конкретных компаниях, а о секторах и странах, а также дает возможность детализировать данные вплоть до конкретных компаний.

Возможная применимость методов, основанных на законе Бенфорда, зависит от выборки исследования и количества данных, раскрываемых в бухгалтерской (финансовой) отчетности, что в случае углубленного анализа в области составления финансовой отчетности позволит вернуться к проблематике, проанализированной в настоящем исследовании.

По результатам исследования, авторы выявили и некоторые сложности в применении закона Бенфорда. Важно отметить, что корректировки, используемые для соблюдения закона Бенфорда, могут привести к систематической ошибке в конечном результате; это связано с масштабной инвариантностью выявленной статистической закономерности и происходит в основном при наличии больших наборов данных (Druică et al., 2018; Nigrini 2017). В частности, в цитируемых работах утверждается и статистически проверяется предположение, что большая выборка может быть умножена на соответствующую выбранную случайную величину или любой заданный скаляр без изменения соответствия данных закону распределения. Аналогичным образом можно удалить данные из выборки без радикального изменения значения p-теста при контроле достоверности закону Бенфорда. Наличие ограничений в рамках проверки закона Бенфорда заслуживает дополнительного анализа в будущих исследованиях в соответствии с рекомендациями вышеупомянутых литературных источников.

Список источников

Зверев, Е., Никифоров, А. (2018). Распределение Бенфорда: выявление нестандартных элементов в больших совокупностях финансовой информации. *Внутренний контроль в кредитной организации, 4* (40), 4–18.

Adhikari, K., Sarkar, P. (1968). Distribution of Most Significant Digit in Certain Functions Whose Arguments are Random Variables. *Sankhya-The Indian Journal of Statistics Series B, 30,* 47–58.

Alali, F., Romero, S. (2013). Characteristics of Failed U.S. Commercial Banks: An Exploratory Study. *Accounting & Finance, 53* (4), 1149–1174. https://doi.org/10.1111/j.1467-629X.2012.00491.x

Álvarez-Jareño, J. A., Badal-Valero, E., Pavía, J. M. (2017). Using Machine Learning for Financial Fraud Detection in the Accounts of Companies Investigated for Money Laundering. *Working Papers 2017/07.* Castellón, Spain: Universitat Jaume.

Azevedo, C. da S., Gonçalves, R. F., Gava, V. L., Spinola, M. de M. (2021). A Benford's Law Based Methodology for Fraud Detection in Social Welfare Programs: Bolsa Familia Analysis. *Physica A: Statistical Mechanics and its Applications, 567,* 125626. https://doi.org/10.1016/j.physa.2020.125626

Azevedo, C. da S., Gonçalves, R. F., Gava, V. L., Spinola, M. M. (2021). A Benford's Law-Based Method for Fraud Detection using R Library. *MethodsX, 8,* 101575, 1–10. https://doi.org/10.1016/j.mex.2021.101575

Balashov, V. S., Yan, Y., Zhu, X. (2021). Who Manipulates Data During Pandemics? Evidence From Newcomb-Benford Law. *Preprint from Research Square.* https://doi.org/10.21203/rs.3.rs-555372/v1

Barabesi, L., Pratelli, L. (2020). On the Generalized Benford Law. *Statistics & Probability Letters, 160,* 108702. https://doi.org/10.1016/j.spl.2020.108702

Barney, B. J., Schulzke, K. S. (2016). Moderating "Cry Wolf" Events with Excess MAD in Benford's Law Research and Practice. *Journal of Forensic Accounting, 1*(1), A66–A90.

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

Bhattacharya, S., Xu, D., Kumar, K. (2011). An ANN-based Auditor Decision Support System Using Benford's Law. *Decision Support Systems, 50*(3), 576–584. https://doi.org/10.1016/j.dss.2010.08.011

Druică, E., Oancea, B., Vâlsan, C. (2018). Benford's Law and the Limits of Digit Analysis. *International Journal of Accounting Information Systems, 31,* 75–82. https://doi.org/10.1016/j.accinf.2018.09.004

Furry, W., Hurwitz, H. (1945). Distribution of Numbers and Distribution of Significant Figures. *Nature, 155,* 52–53.

Hasan, B. (2002). Assessing data Authenticity with Benford's law. *Information Systems Control Journal, 6,* 41–45.

Herteliu, C. Jianu, I., Dragan, I. M., Apostu, S. A., Luchian, I. (2021). Testing Benford's Laws (Non)Conformity within Disclosed Companies' Financial Statements among Hospitality Industry in Romania. *Physica A: Statistical Mechanics and its Applications, 582.* 126221. https://doi.org/10.1016/j.physa.2021.126221

Hillison, W., Durtschi, C., Pacini, C. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting, V,* 17–34.

Krakar, Z., Žgela, M. (2009). Application of Benford's Law in Payment Systems Auditing. *Journal of Information and Organizational Sciences, 33*(1), 39–51.

Nigrini, M. (1996). A Taxpayer Compliance Application of Benford's Law. *The Journal of the American Taxation Association, 18*(1), 72–91.

Nigrini, M. (2017). Audit Sampling Using Benford's Law: A Review of the Literature with Some New Perspectives. *Journal of Emerging Technologies in Accounting, 14*(2), 29–46. https://doi.org/10.2308/jeta-51783

Nigrini, M., Mittermaier, L. (1997). The Use of Benford's Law as an Aid in Analytical Procedures. *Auditing: A Journal of Practice & Theory, 16*(2), 52–67.

Pinkham, R. S. (1961). On the Distribution of First Significant Digits. *The Annals of Mathematical Statistics, 32*(4), 1223–1230. https://doi.org/10.1214/aoms/1177704862

Raimi, R. A. (1976). The first digit problem. *The American Mathematical Monthly, 83*(7), 521–538. https://doi.org/10.2307/2319349

Silva, W. B. da, Travassos, S. K. de M., Costa, J. I. de F. (2017). Using the Newcomb-Benford Law as a Deviation Identification Method in Continuous Auditing Environments: A Proposal for Detecting Deviations over Time. *Revista Contabilidade & Finanças, 28*(73), 11–26. http://dx.doi.org/10.1590/1808-057x201702690

Torres, D., Pericchi, L. (2011). Quick Anomaly Detection by the Newcomb—Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela. *Statistical Science, 26*(4), 502–516. http://dx.doi.org/10.1214/09-STS296

Wallace, W. A. (2002). Assessing the quality of data used for benchmarking and decision-making. *The Journal of Government Financial Management, 51*(3), 16–22.

Whitney, R. E. (1972). Initial Digits for the Sequence of Primes. *The American Mathematical Monthly, 79*(2), 150–152. https://doi.org/10.2307/2316536

References

Adhikari, K., & Sarkar, P. (1968). Distribution of Most Significant Digit in Certain Functions Whose Arguments are Random Variables. *Sankhya-The Indian Journal of Statistics Series B, 30,* 47–58.

Alali, F., & Romero, S. (2013). Characteristics of Failed U.S. Commercial Banks: An Exploratory Study. *Accounting & Finance, 53* (4), 1149–1174. https://doi.org/10.1111/j.1467-629X.2012.00491.x

Álvarez-Jareño, J. A., Badal-Valero, E., & Pavía, J. M. (2017). Using Machine Learning for Financial Fraud Detection in the Accounts of Companies Investigated for Money Laundering. *Working Papers 2017/07.* Castellón, Spain: Universitat Jaume.

Azevedo, C. da S., Gonçalves, R. F., Gava, V. L., & Spinola, M. de M. (2021). A Benford's Law Based Methodology for Fraud Detection in Social Welfare Programs: Bolsa Familia Analysis. *Physica A: Statistical Mechanics and its Applications, 567,* 125626. https://doi.org/10.1016/j.physa.2020.125626

Azevedo, C. da S., Gonçalves, R. F., Gava, V. L., & Spinola, M. M. (2021). A Benford's Law-Based Method for Fraud Detection using R Library. *MethodsX, 8,* 101575, 1–10. https://doi.org/10.1016/j.mex.2021.101575

Balashov, V. S., Yan, Y., & Zhu, X. (2021). Who Manipulates Data During Pandemics? Evidence From Newcomb-Benford Law. *Preprint from Research Square.* https://doi.org/10.21203/rs.3.rs-555372/v1

Barabesi, L., & Pratelli, L. (2020). On the Generalized Benford Law. *Statistics & Probability Letters, 160,* 108702. https://doi.org/10.1016/j.spl.2020.108702

Barney, B. J., & Schulzke, K. S. (2016). Moderating "Cry Wolf" Events with Excess MAD in Benford's Law Research and Practice. *Journal of Forensic Accounting, 1* (1), A66–A90.

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

Bhattacharya, S., Xu, D., & Kumar, K. (2011). An ANN-based Auditor Decision Support System Using Benford's Law. *Decision Support Systems, 50* (3), 576–584. https://doi.org/10.1016/j.dss.2010.08.011

Druică, E., Oancea, B., & Vâlsan, C. (2018). Benford's Law and the Limits of Digit Analysis. *International Journal of Accounting Information Systems, 31,* 75–82. https://doi.org/10.1016/j.accinf.2018.09.004

Furry, W., & Hurwitz, H. (1945). Distribution of Numbers and Distribution of Significant Figures. *Nature, 155,* 52–53.

Hasan, B. (2002). Assessing data Authenticity with Benford's law. *Information Systems Control Journal, 6,* 41–45.

Herteliu, C. Jianu, I., Dragan, I. M., Apostu, S. A., & Luchian, I. (2021). Testing Benford's Laws (Non)Conformity within Disclosed Companies' Financial Statements among Hospitality Industry in Romania. *Physica A: Statistical Mechanics and its Applications, 582.* 126221. https://doi.org/10.1016/j.physa.2021.126221

Hillison, W., Durtschi, C., & Pacini, C. (2004). The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data. *Journal of Forensic Accounting, V,* 17–34.

Krakar, Z., & Žgela, M. (2009). Application of Benford's Law in Payment Systems Auditing. *Journal of Information and Organizational Sciences, 33* (1), 39–51.

Nigrini, M. (1996). A Taxpayer Compliance Application of Benford's Law. *The Journal of the American Taxation Association, 18* (1), 72–91.

Nigrini, M. (2017). Audit Sampling Using Benford's Law: A Review of the Literature with Some New Perspectives. *Journal of Emerging Technologies in Accounting, 14* (2), 29–46. https://doi.org/10.2308/jeta-51783

Nigrini, M., & Mittermaier, L. (1997). The Use of Benford's Law as an Aid in Analytical Procedures. Auditing: A Journal of Practice & Theory, 16 (2), 52–67.

Pinkham, R. S. (1961). On the Distribution of First Significant Digits. *The Annals of Mathematical Statistics, 32* (4), 1223–1230. https://doi.org/10.1214/aoms/1177704862

Raimi, R. A. (1976). The first digit problem. *The American Mathematical Monthly, 83* (7), 521–538. https://doi.org/10.2307/2319349

Silva, W. B. da, Travassos, S. K. de M., & Costa, J. I. de F. (2017). Using the Newcomb-Benford Law as a Deviation Identification Method in Continuous Auditing Environments: A Proposal for Detecting Deviations over Time. *Revista Contabilidade & Finanças, 28* (73), 11–26. http://dx.doi.org/10.1590/1808-057x201702690

Torres, D., & Pericchi, L. (2011). Quick Anomaly Detection by the Newcomb—Benford Law, with Applications to Electoral Processes Data from the USA, Puerto Rico and Venezuela. *Statistical Science, 26* (4), 502–516. http://dx.doi.org/10.1214/09-STS296

Wallace, W. A. (2002). Assessing the quality of data used for benchmarking and decision-making. *The Journal of Government Financial Management, 51* (3), 16–22.

Whitney, R. E. (1972). Initial Digits for the Sequence of Primes. *The American Mathematical Monthly, 79* (2), 150–152. https://doi.org/10.2307/2316536

Zverev, E., & Nikiforov, A. (2018). Raspredelenie Benforda: vyyavlenie nestandartnykh elementov v bol'shikh sovokupnostyakh finansovoy informatsii [Benford's Distribution: Identifying Irregular Elements in Large Sets of Financial Information]. *Vnutrenniy kontrol' v kreditnoy organizatsii [Internal Control in a Credit Institution], 4* (40), 4–18. (In Russ.)

Информация об авторах

Назарова Варвара Вадимовна — кандидат экономических наук, доцент департамента финансов СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики»; https://orcid.org/0000-0002-9127-1644 (Российская Федерация, 190121, г. Санкт-Петербург, ул. Союза Печатников, д. 16; e-mail: nvarvara@list.ru).

Чуракова Ийя Юрьевна — кандидат экономических наук, доцент департамента финансов СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики»; https://orcid.org/0000-0002-1791-607X (Российская Федерация, г. Санкт-Петербург, ул. Союза Печатников, д. 16; e-mail: iychurakova@hse.ru).

Куприянов Дмитрий Алексеевич — магистр ОП «Финансы» СПбШЭиМ, Национальный исследовательский университет «Высшая школа экономики», (Российская Федерация, г. Санкт-Петербург, ул. Союза Печатников, д. 16).

About the Authors

Varvara V. Nazarova — Cand. Sci. (Econ.), Associate Professor of Department of Finance, St. Petersburg School of Economics and Management, HSE University; https://orcid.org/0000-0002-9127-1644 (16, Soyuza Pechatnikov St., St. Petersburg, 190121, Russian Federation; e-mail: nvarvara@list.ru).

Iya Yu. Churakova — Cand. Sci. (Econ.), Associate Professor of Department of Finance, St. Petersburg School of Economics and Management, HSE University; https://orcid.org/0000-0002-1791-607X (16, Soyuza Pechatnikov St., St. Petersburg, 190121, Russian Federation; e-mail: iychurakova@hse.ru).

Dmitriy A. Kupriyanov — Master in Finance, St. Petersburg School of Economics and Management, HSE University (16, Soyuza Pechatnikov St., St. Petersburg, 190121, Russian Federation).

Дата поступления рукописи: 16.05.2023.

Прошла рецензирование: 07.06.2023.

Принято решение о публикации: 22.08.2023.

Received: 16 May 2023.

Reviewed: 07 Jun 2023.

Accepted: 22 Aug 2023.