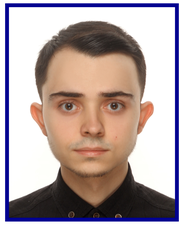
#13

BICA VPS 2023

Improving the Methodology for Integrated Testing of Journal Entries by Benford's Law.

Pavel Y. Leonov*[1], Viktor M. Sushkov[1], Sofia A. Boiko[1], Margarita A. Stepanenkova[1] [1]National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Kashira Hwy, 31, Moscow 115409, Russia. pyleonov@mephi.ru

Pavel Y. Leonov



Viktor M. Sushkov

SUMMARY

The study focuses on improving a methodology for applying Benford's Law tests in detecting distortions within accounting practices. Primary, advanced, and associated tests are conducted to assess the natural character of journal entries of a construction company. Additionally, machine learning techniques such as K-means clustering, random forest, and elliptic envelope are used to analyze the test results and identify highly suspicious transactions within the dataset. The outcomes indicate that the selected transactions flagged by the tests are indeed suspicious.

INTRODUCTION

Currently, fraudulent activities in accounting that result in distorted financial statements have become a pervasive issue and occupy a prominent position in global statistics on economic crimes.

One of the most effective ways to detect fraudulent activities and suspicious transactions is to check the conformity of journal entries to the Benford's Law. By combining machine learning techniques with Benford's Law tests, it becomes feasible to identify the most suspicious transactions and reduce the size of the sample to be analyzed manually.

APPROACH

We built a clustering model in order to identify a group of suspicious transactions. The model was applied to a diverse range of companies spanning various industries. As an illustrative example, we present the results of the analysis conducted on a construction company, which is considered one of the industries with a high susceptibility to fraud.

METHODS

Benford's Law states that if a data set is randomly generated under the influence of external factors, the leading digits form a discrete exponential distribution. The probability of a digit being the first is measured using the following formula:

$$P(d_1) = log_{10}\left(1 + \frac{1}{d_1}\right);\ d_1 = 1, 2, \dots, 9$$

Based on the Benford distribution, M. Nigrini developed 8 statistical tests to check the natural character of data sets and classified them into primary, advanced and associated. Generally, the tests of each group are applied sequentially.

Conducting the tests also involves calculating statistical characteristics to assess conformity to Benford's law. We used Z-statistic, Mean Absolute Deviation, Bhattacharya coefficient, Kullback-Leibler divergence, Kolmogorov-Smirnov statistic, Chi-square statistic, and Mantissas Test.

RESULTS

The input for the study was a set of journal entries from the general ledger for a 4-year period, containing 144 thousand records. The initial data for building the model were the following statistics calculated for each transaction amount:

- X1: Z-statistics for the first digit test;
- X2: Z-statistics for the second digit test;
- X3: Z-statistics for the first-two digits test;
- X4: Frequency of transaction amounts under the first two digits;
- X5: Z-statistics for the second-order test;
- X6: Frequency of all transaction amounts;
- X7: Z-statistics for the last-two digits test.

Clustering of the data was carried out using K-means. We chose the number of clusters equal to 2.

The average values of features in clusters are presented in Figure 1.

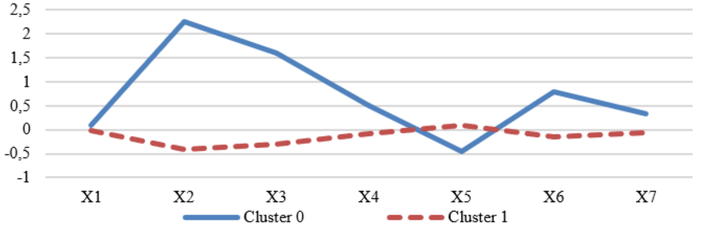


Fig. 1. Average values of features in clusters

ANALYSIS

72% of suspicious cluster operations have abnormally round amounts. Especially suspicious are the round amounts contained in contracts for the supply of raw materials.

DISCUSSION

It can be concluded that distortions are probably caused by rounding the transaction amounts down. Thus, considering the cost of goods or services is below the "fair market" price in the documents, the company reduced the amount of taxable income.

CONCLUSIONS

1. The application of mathematical methods has proven to significantly enhance the productivity of organizations that conduct activities in the field of financial control. The proposed approach makes it possible to free professionals from routine transaction checks, thereby allowing them to narrow their focus onto operations necessitating additional procedures. The results have demonstrated high effectiveness, as the operations identified through the tests were found to be suspicious, with significant monetary values.
2. By combining Benford's Law with machine learning techniques, this research offers a new perspective on how to analyze and interpret data in the context of fraud detection. However, further research is needed to optimize machine learning algorithms for processing large arrays of journal entries and improving the accuracy of identifying suspicious operations. Additionally, it is worth further investigating and assessing the practical implementation and scalability of this approach in various organizational contexts