

## Financial statements fraud identifiers

Robert Zenzerović & Josip Šajrih

To cite this article: Robert Zenzerović & Josip Šajrih (2023) Financial statements fraud identifiers, Economic Research-Ekonomiska Istraživanja, 36:3, 2218916, DOI: [10.1080/1331677X.2023.2218916](https://doi.org/10.1080/1331677X.2023.2218916)

To link to this article: <https://doi.org/10.1080/1331677X.2023.2218916>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 13 Jun 2023.



Submit your article to this journal [↗](#)



Article views: 236



View related articles [↗](#)



View Crossmark data [↗](#)

## Financial statements fraud identifiers

Robert Zenzerović<sup>a</sup> and Josip Šajrih<sup>b</sup>

<sup>a</sup>Faculty of Economics and Tourism “Dr. Mijo Mirković”, Juraj Dobrila University of Pula, Pula, Republic of Croatia; <sup>b</sup>Basellandschaftliche Kantonalbank, Pratteln, Switzerland

### ABSTRACT

Contemporary research among fraud professionals indicates that organizations lose 5% of revenues from fraud every year which makes the research in this area and the derivation of fraud detection models very important. The purpose of the article is to develop a new accounting tool that will help companies and investors in prompt fraud detection and prevention which can finally result in the preservation of financial stability as well as more efficient capital allocation. In this context the main objective of the research is to test the significance of some financial statements' relations that has not been used in the previous research using the dataset from SEC AAERs presented and included in Bao et al.'s research as well as to combine them with existing ones and consequently develop new financial statement fraud detection model. Another objective consists of presenting some of the most significant and contemporary research in the field of financial statement fraud detection models and comparing their quality using the ROC analysis. Research results were generated by using the SMOTE algorithm and logistic regression analysis on the dataset of 146,045 cases for a period from 1982 to 2014 and point out five independent variables used by Bao et al. The financial statement fraud detection model comprised of change in free cash flow, percentage of soft assets, sale of common and preferred stock, change in cash sales, and change in receivables shows a sufficient level of discriminant power with 67% area under ROC curve. The model derived could be used as a starting point for fraud detection preventing the significant losses the company and stakeholders could face.

### ARTICLE HISTORY

Received 26 July 2022  
Accepted 23 May 2023

### KEYWORDS

Financial statement fraud;  
fraud detection models;  
logistic regression analysis;  
SMOTE algorithm; ROC  
analysis

### JEL CODES

M41; M42

## 1. Introduction

Fraud has always been a global issue. Throughout the history, human creativity to find easier ways has always been active. From the shepherds to the hackers, from the Hammurabi code of laws to the EU Directive 2019/1937 of the European Parliament and of the Council of 23 October 2019 on the protection of persons who report

**CONTACT** Robert Zenzerović  [robert.zenzerovic@unipu.hr](mailto:robert.zenzerovic@unipu.hr)

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

breaches of Union law, there are always significant discrepancies between fraud practices and methods of preventing and detecting various types of frauds. According to contemporary research among professionals dealing with fraud, organizations lose 5% of revenue from fraud every year (ACFE, 2022). The Association of Certified Fraud Examiners (ACFE), the organization which performs the longest consecutive research of fraud, finds that corruption is the most common fraud scheme in every global region. Asset misappropriation schemes and financial statements fraud schemes are another two types of schemes examined in research where the first one results in the lowest median loss of 100,000 USD, while the financial statements fraud generated a constant highest median loss which amounted to 593,000 USD in 2022 report (ACFE, 2022). Because only between 9 and 17% of the total loss are fully compensated (depending on region) and the organizations with the fewest employees had the highest median loss, authors think that researching fraud will probably be among the most challenging area in the accounting field.

The focus of this article is on financial statements' fraud. An overview of some of the most significant and contemporary research in the field of financial statements fraud detection models is presented which is one of the objectives of this article. Another and more important objective of the article consists in testing the significance of some financial statements positions' relations that has not been used in the previous research using the dataset from SEC AAERs presented and included in Bao et al.'s (2020) study. The logistic regression technique has been used and the financial statements fraud detection model has been developed. In this way, the next purpose of the research has been fulfilled—the development of a new accounting tool (financial statements fraud detection model) that will help companies and investors, creditors, financial analysts, and other stakeholders in prompt fraud detection and prevention what can result in preservation of financial stability as well as more efficient capital allocation.

Fraud prevention is crucial if a company strives to eliminate the possibility of fraud in the long run (Dimitrijević et al., 2020, p. 370). Losses related to fraud could cause the company going concern questionable, particularly the small and medium ones that are faced with relatively more significant losses than bigger ones which can be related to weaker systems of internal control and lower consciousness of fraud prevention.

## 2. Previous research

Financial statements fraud is traditionally the most expensive type of fraud whose costs are very difficult to precisely calculate. This is the reason why some research, like ACFE, estimates the losses from financial statements fraud as a value of falsification or differences between real and falsified publicly disclosed values. In their last ten Reports to the nations on occupational fraud that include 20 years period, ACFE reported financial statements fraud as the most expensive one reaching a median loss of 4.25 million USD in 2002 with significant variation during the period till 2012 when they reached one million USD (ACFE, 2022). From 2014 to 2022, losses from fraudulent financial reporting started to decrease and finally reached 593,000 USD.

This trend can be explained by stronger systems of corporate governance and systems of regulation that have been established in the previous decade followed by the Sarbanes-Oxley Act and new legislation that has been brought after in other parts of the world.

Fraudulent financial reporting got into the focus of researchers along with numerous financial scandals which included some of the biggest companies in developed countries (e.g., Enron in the US, Parmalat in Italy, Toshiba in Japan, Agrokor in Croatia, and others) and much smaller one's years earlier. The development of financial markets and the fundamental analysis, which has a starting point in financial statements, arise the need to develop diagnostic and prognostic tools that will be easily applicable and have high classification accuracy, reducing the probability of mistakes. That is why this article presents some of the most known and relevant research that developed financial statements fraud detection models and compares their discriminant power using the ROC analysis. Some of the relevant research does not include correct financial statements in their samples so the models could not be tested for type II error. This is the reason why only the models comparable to the one derived in this article are briefly elaborated.

Further in the article, some significant research is presented. Green and Choi (Green & Choi, 1997) applied the neural network technique to the sample of 192 falsified and 3,173 correct financial statements using publicly available data. Their model correctly classified 100% of falsified and 7.1 of correct financial statements i.e., companies that disclosed these financial statements with the area under the curve of 0.472 or 47.2%.

Another author that is considered to be among the most known is Messod Beneish. He used probit analysis on the sample of 149 falsified and 3,389 correct financial statements collected from publicly available sources. Beneish's model (1999) correctly classified 54.2% of falsified financial reports and 45.5% of correct financial statements with the area under the curve reaching 0.492.

The third significant research has been done by Dechow et al. (2011) who used the logistic regression analysis, technique used in this research. The sample size included 57 companies with falsified financial statements and 1,244 correct reports. The database has been compiled by detailed examination of firms that have been subject to enforcement actions by the Securities and Exchange Commission (SEC) for allegedly misstating their financial statements. Since 1982, the SEC has issued Accounting and Auditing Enforcement Releases (AAERs) during or after an investigation against a company, an auditor, or an officer for alleged accounting and/or auditing misconduct. These releases provide varying degrees of detail on the nature of the misconduct, the individuals and entities involved, and the effect on the financial statements (Dechow et al., 2011). The logistic regression model correctly classified 70% of falsified financial reports and 84.9% of correct financial statements with the area under the curve reaching 0.762.

Cecchini et al. (2010) performed research on 132 falsified and 3,187 correct financial statements using support vector machines and kernel methodology. Data has been collected as well from SEC AAERs source. Their model correctly classified 80% of falsified financial reports and 90.6% of correct financial statements with the area under the curve reaching 0.878.

Finally, the last models examined in this article are the ones derived from the research of Bao et al. (2019). Research has been performed using the data collected from SEC AAERs source on which the machine learning approach was applied. They derived four different models and the first one included 28 raw financial data as input variables. The second model included 14 financial ratios while the third model included input variables from the previous two models. The last model comprised all 294 raw financial data. The quality of the models measured by area under the ROC curve indicates that the first model with 28 raw financial data shows the highest quality reaching 0.725 area. The third, fourth, and second model has reached 0.696, 0.692 and 0.659 area under the curve.

### 3. Financial statements fraud

According to International Standard on Auditing 240—The auditor’s responsibilities relating to fraud in an audit of financial statements, fraud is defined as an intentional act by one or more individuals among management, those charged with governance, employees, or third parties, involving the use of deception to obtain an unjust or illegal advantage (IAASB, 2020, p. 168). Fraudulent financial reporting involves intentional misstatements including omissions of amounts or disclosures in financial statements to deceive financial statements users. It can be caused by the efforts of management to manage earnings to deceive financial statement users by influencing their perceptions as to the entity’s performance and profitability (IAASB, 2020, p. 177). Financial statements fraud is performed by management and/or those charged for governance. They try to omit quantitative or qualitative information from financial statements as well as disclose falsified ones. The focus of this article is on quantitative information published in financial statements. Financial statements fraud mostly includes net worth (net assets) or net income overstatements and/or understatement where fraud perpetrators use various techniques that could be summarized into groups of time differences, fictitious revenues, concealed or overstated liabilities and expenses, improper asset valuation, improper disclosures, and understated revenues. According to the latest research by ACFE (2022) financial statements fraud appears in 9% of the cases but is the type of fraud that relates to most stakeholders. Financial statements fraud can disrupt trust in financial reporting and corporate governance and consequently lead to less efficient allocation of capital. This is the reason why the authors focused on developing a new model that can help investment society and other stakeholders to detect fraudulent financial reporting.

#### 3.1. Research questions

The main objective of the research is to test the significance of some financial statements positions’ relations that has not been used in the research using the dataset from SEC AAERs presented and included in Bao et al.’s (2020) study and consequently develop new financial statements fraud detection model. Another objective consists of a presentation of some of the most significant and contemporary research in the field of financial statements fraud detection models and comparing their

quality using the ROC analysis. To achieve the research objectives, the authors set the following research questions:

RQ 1: Could the new financial statements fraud detection model include some additional financial statements positions?

RQ 2: What is the level of new financial statements fraud detection model discriminant power in comparison with other contemporary comparable models?

Authors have explored the first research question using SMOTE algorithm and logistic regression analysis on the big dataset for a long period described below, while the answer to the second research question has been given using the ROC analysis.

### **3.2. Research methodology**

As it was mentioned before, the objective of this research article is to analyse available research results and consequently test the significance of financial statement positions' relations that reflect some of the areas in which financial statements are falsified and which have not been used in research yet. For research purposes, a financial database used by Bao et al. has been used (<https://github.com/JarFraud/FraudDetection>). Data has been collected for all publicly traded U.S. firms over the period 1991–2008. The sample started in 1991 because there is a significant shift in US firms' fraudulent behaviour as well as the nature of SEC enforcement starting around that time. The sample ends in 2008 because the regulators reduced the enforcement of accounting fraud starting from around 2009, increasing the possibility that many accounting fraud cases remain undetected for the post-2008 period (Bao et al., 2020, p. 203). Accounting data were collected from COMPUSTAT, a fundamental annual database for fiscal years 1991 to 2014. An AAER database includes all the AAERs announced over the period between 17 May 1982, and 31 December 2014. Data used in this research were downloaded from a public 'GitHub' directory called 'JarFraud/FraudDetection'. The dataset consists of 28 raw accounting data variables and 14 financial ratio variables. The total number of proven fraudulent cases in the dataset is 964 and 145,081 valid ones, resulting in a total number of recordings of 146,045.

We have extended the research of Bao et al. (2020) by selecting some of the financial data as input variables they disclosed according to the existing accounting theories in the field of financial statements fraud. Ten input variables from Bao et al. selected are change in free cash flow, change in return on assets, percentage of soft assets (where soft assets represent the percentage of assets that are neither cash nor plant, property, and equipment), long-term debt issuance, sale of common and preferred stock, change in cash sales, change in inventory, change in receivables, book-to-market value and working capital accruals (calculation of variables are available in Dechow et al. (2011, p. 60). We added seven additional input variables according to the fact that they can represent the fictitious revenues, accumulating the inventories and timing differences. These variables are the change in debt in short-term liabilities, the relation between the change in revenues and the change in the cost of goods sold, the relation between the change in revenues and the change in receivables, the relation between the change in inventories and change in average assets, the relation

**Table 1.** Descriptive statistics – fraudulent financial statements.

	Change in free cash flow	Percentage of soft assets	Sale of common stock	Change in cash sales	Change in receivables	Sale of common stock adj.
Count	964	964	964	964	964	964
Mean	– 0.101974	0.634767	105,666,387.00	0.373692	0.037340	0.328175
Standard deviation	0.320505	0.221423	1,017,366,345.00	1,188,827.00	0.095761	4,730,640.00
Minimum	– 2,980,601.00	0.004515	0.002000	– 6,484,197.00	– 0.304659	– 0.163153
25%	– 0.162038	0.490044	2,254,250.00	0.017222	– 0.003117	– 0.152680
50%	– 0.046642	0.683939	14,602,000.00	0.165750	0.017337	– 0.095265
75%	0.011440	0.807266	37,085,500.00	0.415991	0.060124	0.009281
Maximum	1,314,497.00	0.993599	30,563,000,000.00	9,600,000.00	0.382809	141,951,367.00

Source: Research results.

**Table 2.** Fraudulent financial statements non-fraudulent financial statements

	Change in free cash flow	Percentage of soft assets	Sale of common stock	Change in cash sales	Change in receivables	Sale of common stock adj.
Count	145.081	145.081	145.081	145.081	145.081	145.081
Mean	– 0.008224	0.498893	34.620.567,00	0.184024	0.013764	– 0.002181
Standard deviation	0.539070	0.273664	199.134.125,00	1.306.438,00	0.081917	0.925951
Minimum	– 2.980.601,00	0.004515	– 516.713.000,00	– 6.484.197,00	– 0.304659	– 2.565.820,00
25%	– 0.116445	0.269911	0.808000	– 0.049796	– 0.009019	– 0.159405
50%	– 0.021232	0.521593	9.534.000,00	0.099919	0.009036	– 0.118830
75%	0.056262	0.725019	27.657.300,00	0.214009	0.032684	– 0.034559
Maximum	4.061.810,00	0.993599	27.619.000.000,00	9.600.000,00	0.382809	128.262.096,00

Source: Research results

between EBITDA and revenues, net debt/EBITDA and the relation between free cash flow to revenues.

The first step of the analysis was to describe and clean the data. Several approaches have been taken. As the goal was to search for a relationship by using logistic regression analysis, all AAER companies' financial statements were treated as 1—fraudulent or falsified and all correct statements as 0. Tables 1 and 2 show the descriptive statistics of the chosen variables.

Logistic regression analysis (or logit regression) includes the estimation of the parameters of a logistic model (the coefficients in the linear combination). Formally, in binary logistic regression, there is a single binary dependent variable, coded by an indicator variable, where the two values are labelled '0' and '1', while the independent variables can each be a binary variable (two classes, coded by an indicator variable) or a continuous variable (any real value). The corresponding probability of the value labelled '1' can vary between 0 (certainly the value '0') and 1 (certainly the value '1'). The first step was to show a correlation between the financial data variables in the database. After that the variables with low multicollinearity were chosen for further analysis, as high multicollinearity can decrease model predictivity. As a threshold for high multicollinearity a limit of +15% was used. In the second step, data was cleaned using several approaches. Missing data were replaced by variable averages, zeros, or normalized by using the z-score normalization.

After the process of data cleaning, data have been balanced using a Synthetic Minority Oversampling Technique—SMOTE algorithm. The data in the dataset was



unbalanced as there were a large number of nonfraudulent examples and a minority of AAERs or proven fraudulent examples. The most widely used approach to synthesizing new examples is called the Synthetic Minority Oversampling technique, or SMOTE for short. This technique was described by Nitesh Chawla et al. (2002). SMOTE works by selecting examples that are close to the feature space, drawing a line between the examples in the feature space, and drawing a new sample at a point along that line. This procedure can be used to create as many synthetic examples for the minority class as are required. As described in the paper, it suggests first using random undersampling to trim the number of examples in the majority class, then using SMOTE to oversample the minority class to balance the class distribution. The approach is effective because new synthetic examples from the minority class are created that are plausible, that is, they are relatively close in feature space to existing examples from the minority class. A general downside of the approach is that synthetic examples are created without considering the majority class, possibly resulting in ambiguous examples if there is a strong overlap for the classes. By using the SMOTE algorithm, data were balanced 50:50, which allowed us to perform a logistic regression accurately and logically.

After the data has been balanced, logistic regression analysis was performed to determine the  $p$ -values of selected variables. If the selected variables  $p$ -value was over 0.05 the variable has been discarded. Train and test data have been divided at 70:30.

Upon receiving desired variables with acceptable  $p$ -values a receiving operating characteristic curve (ROC) was constructed and the area under the curve (AUC) was calculated to comprehend the discriminatory strength of the model. If the AUC is over 0.5 the model shows statistical importance and discriminatory strength exist.

ROC analysis is a useful tool for evaluating the performance of diagnostic tests and more generally for evaluating the accuracy of a statistical model (e.g., logistic regression, linear discriminant analysis) that classifies subjects into 1 of 2 categories (Zou et al., 2007, p. 654). ROC curve has been applied in many other areas including psychology, atmospheric sciences, biosciences, experimental psychology, finance, geosciences, sociology, machine learning, and data mining (Gonçalves et al., 2014, p. 3).

Financial statements fraud detection models are classification models that produce continuous outputs which are compared with thresholds to predict the probability of fraudulent financial reporting. The model's qualities are often estimated by their discriminant power where the models with higher power are considered to be better than others. This could be correct if the classification error costs are the same which is not the case in fraudulent financial reporting prediction. Type I error appears when the models classify the financial statements as correct while in reality they are falsified. On the other side, type II error is the error of classifying the correct financial statements as falsified. ROC analysis is a better model's quality estimator because it does not assume the equality of classification error costs which is the case with financial statements fraud detection models. Discriminant power represented by the area under the curve will be the indicator of the model's quality.

The final step in estimating the financial statements fraud detection model's discriminant power i.e., their quality is comparing the proportion under the model's ROC curve with theoretical critical values.<sup>1</sup> Different authors that perform research



**Table 3.** Model discriminant power estimation.

Proportion of area under curve	Model discriminant power
0.50 – 0.60	Insufficient
0.60 – 0.70 (0.50 – 0.75)	Sufficient
0.70 – 0.80 (0.75 – 0.92)	Good
0.80 – 0.90 (0.92 – 0.97)	Very good
0.90 – 1.00 (0.97 – 1.00)	Excellent

Source: Rozga (2009).

using ROC analysis use different critical values in estimating the model's discriminant power (Rozga, 2009; Simon, 2000). Table 3 shows the intervals of proportions under the ROC curve with an estimation of model discriminant power suggested by two groups of authors.<sup>2</sup>

In further steps, a confusion matrix was constructed to show the number of right predictions against actual ones and several type I and type II errors. The final part of the research consists of showing the prediction ability of the model and the F1 score.

In statistical analysis of binary classification, the F1 score is a measure of a test's accuracy. It is calculated from the precision and recall of the test, where the precision is the number of true positive results divided by the number of all positive results, including those not identified correctly, and the recall is the number of true positive results divided by the number of all samples that should have been identified as positive. Precision is also known as a positive predictive value, and recall is also known as sensitivity in diagnostic binary classification. The F1 score is the harmonic mean of the precision and recall.

The final part includes model development—a usage of calculated weights for favorable variables implemented in a prediction logistic function. All analysis was performed in the Python programming language by using the next modules: pandas, imblearn, sklearn, seaborn, numpy, matplotlib, statsmodel, and openpyxl.

### 3.3. Research results

Conduction of various testing showed that the following independent variables have acceptable  $p$ -values in the logistic regression model against a binary dependent variable—fraudulent financial statements: change in free cash flow, percentage of soft assets, sale of common and preferred stock, change in cash sales and change in receivables. The additional seven input variables that the author suggested have been discarded as not significant according to the unfavourable values generated in research. Consequently, the new financial statements fraud detection model does not include those additional financial statements positions and the answer to the first research question is negative. Despite that authors have developed a new model using the logistic regression with existing datasets and financial statement fraud identifiers and continue the research to determine its discriminant power and compare it with other models.

To improve the results, different approaches were taken to test the ability of the data to predict correctly. Logistic regression was tested on raw data where missing values were treated as 0, median, and column average. The best and most reliable result was replacing the missing values with the average of the data set. This was

**Table 4.** Logistic regression analysis results (*p*-values, variable weights, and standard error).

Independent variable	Variable label	<i>p</i> -value - Missing values = average	Coefficient (linear function weights)	Standard error
Change in free cash flow	x1	0.0000	−0.3108	0.090
Percentage of soft assets	x2	0.0000	1.9306	0.065
Sale of common and preferred stock (z-score normalized)	x3	0.0000	0.0473	0.134
Change in cash sales	x4	0.0001	0.0708	0.009
Change in receivables	x5	0.0000	1.6850	0.021
Constant (Intercept)	b0	0.0000	−6.1939	0.339

Source: Research results.

**Table 5.** Area under the curve (AUC) and classifier accuracy.

Data set	Area under the curve (AUC)	Classifier accuracy
Missing values = 0	0.50	0.99
Missing values = median	0.50	0.99
Missing values = average	0.67	0.99

Source: Research results.

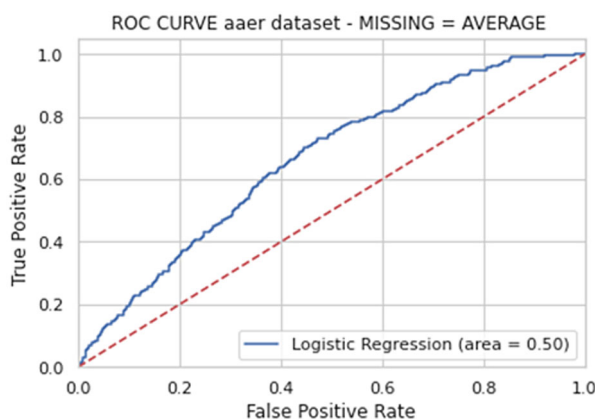
used in further research. Other results that were produced with median and missing = 0 were discarded and not used in the final model development as they were unsatisfactory. Table 4 shows *p*-values, variable linear function weights, and standard error of the favourable variables which were chosen for model development. Variable sale of common and preferred stock was twice adjusted because the variable average was high. Therefore, a z-score normalization was applied to normalize the data. The model showed better results with the variable change in free cash flow even though it has a negative coefficient. For that reason, it was kept in the model. The authors even took the approach of not using the constant in the derived model and also removing the variable change in free cash flows. This approach was also discarded as the results received were less favourable than the ones shown in the paper.

Table 5 shows the area under the curve and the classifier accuracy—values of all three above-mentioned data sets. Every data set consists of independent variables mentioned in Table 4 and the same binary dependent variable ‘fraudulent financial statements’.

In Table 5, only the dataset where missing values of the selected independent variables were replaced by variable averages produced a favourable result, a discriminatory strength, or sufficient classification ability of this model. Figure 1 shows the last model (whose missing variables are changed by average) discriminant power graphically by ROC curve. Results of other datasets were discarded as the discriminatory strength of the model was weak.

In the following step, a confusion matrix was created for the results of the dataset which has shown an adequate relationship to the dependent variable. Table 6 shows these results.

As the data set consists mostly of statements that are not fraudulent, the model better predicted correctly cases that are not fraudulent. It predicted one fraudulent case. Type I errors occurred in 279 cases or 0.608% and type II errors in two cases or 0.0043%. It is important to mention that the majority of the model accuracy of 0.66 out of 0.67 in prediction is a result of one variable and that is the sale of common



Source: Research results

**Figure 1.** Receiver operating characteristic curve.

Source: Research results.

**Table 6.** Confusion matrix of the results.

Actual/Predicted	Falsified	Correct
Falsified	45,532	2
Correct	279	1

Source: Research results.

**Table 7.** Model precision, recall, and F1 score.

	Precision	Recall	F1 score
Correct	0.99	1.00	1.00
Falsified	0.33	0.00	0.01
Accuracy			0.99
Macro average	0.66	0.50	0.50
Weighted average	0.99	0.99	0.99

Source: Research results.

and preferred stock, which indicates that higher sales in common and preferred stock are a strong indicator that the financial statements might be fraudulent. In the final model testing a probability of 74% was calculated on one fraudulent case. In that case, the sale of common and preferred stock was extremely out of dataset bounds even after normalization through z-score normalization.

Table 7 shows a research result by calculating the precision, recall, and F1 score of a dataset, indicating the stronger relationship and discriminatory strength between the dependent and independent variables.

Table 7 shows that even though there exists a relationship between the dependent and independent variables and discriminatory strength is present, the model does not predict truly fraudulent statements very accurately. Model precision is around 0.66, but model recall is 0.50. F1 score combines the precision and recall of a classifier by taking their harmonic mean. Therefore, it can be said that the model is overall above 50% of the time accurate because even though the model precision is 0.66 and 66% of positive results belong to the positive class, only 50% of them are positive class predictions from the positive examples in the dataset.

**Table 8.** Discriminant power of the models' estimation using the area under the curve.

Model	Area under curve (AUC) in %	Discriminant power
Green and Choi	47.2	Insufficient
Beneish	49.2	Insufficient
Dechow et al.	76.2	Good
Checchini et l.	87.8	Very good
Bao et al.	65.9 – 72.5	Sufficient - Good
Zenzerović and Šajrih	67	Sufficient

Source: Research results.

In the final part of this research, a model was derived. The final model is a logistic regression function where a constant and weights of favourable variables are included. In Equation 1, a final model is shown by using the variables and labels from Table 3. Equation 1. Financial statements fraud detection model:

$$p^{\wedge} = \frac{1}{1 + e^{-(-6.139 + (-0.3108 * x1) + (1.9306 * x2) + (0.0473 * x3) + (0.0708 * x4) + (1.6850 * x5))}} \quad (1)$$

where

$p^{\wedge}$  = probability of fraud

$x1$  = Subject Change in free cash flow

$x2$  = Subject Percentage of soft assets

$x3$  = Subject Sale of common and preferred stock

$x4$  = Subject Change in cash sales

$x5$  = Subject Change in receivables

To answer the second research question, the models' discriminant power should be compared and estimated. Table 8 shows the comparisons of models using the area under the curve.

Models' discriminant power varies significantly. Considering the models' discriminant power from Table 3, the Checchini et al.'s model shows the best result with very good discriminant strength. It is followed by Dechow et al. with good discriminant power as well as one of the Bao et al. models, while another Bao et al.'s model as well as our model shows sufficient discriminant power. Beneish and Green and Choi's models do not show sufficient discriminant power. Despite the various levels of models' discriminant powers, they should be used with caution and could be used in combination to draw a first impression of the probability that the financial statements are fraudulent.

#### 4. Discussion

The financial statements fraud detection model derived shows sufficient discriminant power which makes it applicable as a starting point for fraud detection. Compared with other models it shows sufficient discriminant power which is consistent with models derived from Bao et al. whose dataset has been used. The best discriminant power among the model compared has been achieved by Checchini et al. who analysed the financial statement fraud identifiers used by Beneish, Summers and Sweeney (1998), and Dechow et al. The model could be applicable for prompt fraud detection preventing the significant losses the company and investors, creditors, and other

stakeholders could face. It could be used as a fraud prediction model as well as a tool for improving the quality of financial reporting. The model is easy to use and apply but should be taken with caution when estimating the complexity that financial statements fraud includes. Five independent variables included in the model covers the various segments of financial statements fraud. Change in cash sales and change in receivables are variables that relate to revenues falsification while the percentage of soft assets indicates that companies with falsified financial statements use discretion in estimating the value of an asset that is not cash or plant, property, and equipment or are not willing to adjust it to fair value when they are obliged to. On the other side, the companies for which the fair value measurement is not mandatory can use it as a mean for financial statements manipulation, particularly in cases when they are not applying fair value consistently. Change in free cash flow can indicate that companies with falsified financial statements had more significant changes than other ones which can be explained by the fact that they are restraining themselves from capital expenditures before and in misstatement year and/or generating lower cash flows from operations. The last, but statistically most significant independent variable is the sale of common and preferred stock which indicate that misstated firms are raising capital probably because they are not having access to sufficient sources of financing and/or they are misleading investors with preferred stock issuance with fixed dividend.

Sales of treasury stocks are stocks that were outstanding and bought back from the stockholders and kept on the books. Selling that kind of assets could be portrayed as operating or financing activity in the cash flow statement or as an operating sales revenue or financial revenue in the profit and loss statement which is misleading to stakeholders according to the fact that it should be classified within financial activities. The sale of common and preferred stock usually means selling more than two-thirds of outstanding common and preferred stock. Companies tend to resort to this kind of activity usually at the end of their life cycle indicating a change in the future through an acquisition, merger, or bankruptcy with preceding. Although it is possible to show that kind of business activity as inflows from operating activity instead of cash flow from financing what is appropriate, the aim could be to boost up the cash flow from operating activities as well as operating revenues what could also influence the profit and loss account where higher revenues and a larger operating income are falsified what mislead of the stockholders.

The results show that there exists a relationship between fraudulent financial statements and raw accounting data. The dataset used in this research included data from all sectors and types of business activities. But every business activity is unique by itself and has specific operations which are hard to derive so the next area of research could include deriving the model for a particular activity.

One of the biggest problems represents a small number of actually proven fraudulent statements against the ones that are valid or quasi-valid. This represents a challenge in the research as similarities between proven fraudulent financial statements are on the verge of randomness by looking only at raw accounting data. Pure accounting data can only give a hint on what to look for and where. This opens a new possibility for including and analysing nonfinancial data and identifying

significant ones that contribute to detecting financial statements fraud. The future of financial statements fraud detection most likely lies in nonfinancial data that could be derived from notes to financial statements, nonfinancial publicly disclosed information as well as from other informal publicly available sources.

## Notes

1. Simple model's quality assessment could be done by comparing their classification results with critical value. Critical value in this case is theoretical probability increased by 25%.
2. The authors differ according to the proportions of area under ROC curve and appropriate discrimination power estimation. First group proportion of area is shown outside while other group proportion is shown inside the brackets.

## References

- Association of Certified Fraud Examiners: Occupational Fraud. (2022). A report to the nations, <https://legacy.acfe.com/report-to-the-nations/2022/>
- Bao, Y., Bin, K., Bin, L., Julia, Y., & Jie, Z. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research*, 58(1), 199–235. <https://doi.org/10.1111/1475-679X.12292>
- Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, 55(5), 24–36. <https://doi.org/10.2469/faj.v55.n5.2296>
- Cecchini, M., Aytug, H., Koehler, J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science*, 56(7), 1146–1160. <https://doi.org/10.1287/mnsc.1100.1174>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Database. <https://github.com/JarFraud/FraudDetection>
- Dechow, P. M., Weili, G., Chad, R. L., & Sloan, R. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, 28(1), 17–82. <https://doi.org/10.1111/j.1911-3846.2010.01041.x>
- Dimitrijević, D., Karapavlović, N., & Milutinović, S. (2020). Fraud prevention measures in Serbian small and medium-sized enterprises: Existence and effectiveness. *Ekonomika Preduzeća*, 68(5-6), 369–382. [https://www.ses.org.rs/uploads/dimitrijevic\\_et\\_al\\_201109\\_72136\\_855.pdf](https://www.ses.org.rs/uploads/dimitrijevic_et_al_201109_72136_855.pdf)
- Gonçalves, L., Subtil, A., Rosário Oliveira, M., & de Zea Bermudez, P. (2014). ROC curve estimation: An overview. *Statistical Journal*, 12(1), 1–20. <https://www.ine.pt/revstat/pdf/rs140101.pdf>
- Green, P., & Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing: J. Practice Theory*, 16(1), 14–29.
- IAASB. (2020). Handbook of international quality control, auditing, review, other assurance, and related services pronouncements, <https://www.ifac.org/system/files/publications/files/IAASB-2020-Handbook-Volume-1.pdf>
- Rozga, A. (2009). *Statistika za ekonomiste*. Ekonomski fakultet u Splitu, Split.
- Simon, S. (2000). [http://www.dia.fi.upm.es/~jaferran/teaching/sistemasayudadecision/Stats\\_%20ROC%20curve%20%28August%2018,%201999%29.pdf](http://www.dia.fi.upm.es/~jaferran/teaching/sistemasayudadecision/Stats_%20ROC%20curve%20%28August%2018,%201999%29.pdf)
- Song, M., Naoto, O., & Akinobu, S. (2016). Predicting accounting fraud: evidence from Japan. *The Japanese Accounting Review*, 6(2016), 17–63. <https://doi.org/10.11640/tjar.6.2016.01>
- Summers, S. L., & Sweeney, J. T. (1998). Fraudulent misstated financial statements and insider trading: An empirical analysis. *Accounting Review*, 73(1), 131–146.
- Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, 115(5), 654–657. <https://www.ahajournals.org/doi/10.1161/circulationaha.105.594929>