Received 26 July 2024
Revised 6 November 2024
2 February 2025
23 June 2025
Accepted 19 September 2025

Evaluating the efficacy of modified Beneish and Dechow models for detecting earnings management in Polish companies

Marek Sylwestrzak and Julia Słomkowska
University of Warsaw, Warsaw, Poland

Abstract

Purpose – This article analyzes the role of several variants of the M-Score models and the F-Score model in the detection of earnings management in Poland. Previous research on earnings management concentrated only on the use of the original Beneish M-Score model.

Design/methodology/approach – The sample comprises 60 observations of nonfinancial Polish companies listed on the Warsaw Stock Exchange that received a monetary fine from the Polish Financial Supervision Authority (UKNF Board) for violation of International Financial Reporting Standards (IFRS) in the years 2010–2021. For each fraud observation, we selected the control observations based on industry (obtained from the SIC code), financial year and size of total assets.

Findings – The results indicate that the five-element M-Score model with a marginal value (−2.76) has the highest F-measure and the second-best accuracy in detecting earnings management than the F-Score model for the Polish sample. Moreover, the proprietary M-Score models usually had lower F-measure and accuracy than the original M-Score models, except for the model by Sylwestrzak (2022).

Research limitations/implications – Future research comparing the F-Score and M-Score models needs to include a larger sample of control companies due to the potential for choice-based sample bias.

Practical implications – The choice of an appropriate detection model and cut-off threshold may affect the conclusions regarding the analyzed sample.

Originality/value – The research provides empirical evidence of the appropriateness of modifying the M-Score model for country conditions and verifying the usefulness of the M-Score models modified by other authors.

Keywords Auditing, Earnings management, Financial reporting, Modified M-Score, Warsaw Stock Exchange

Paper type Research article

Introduction

The literature on detecting earnings management has evolved over the decades. The accounting scandals that occurred in the 2000s in the United States have also reshaped research into earnings management (DeFond, 2010). Earnings management refers to the manipulation of financial data reported in financial statements and raises investors' concerns about the quality of financial information available for listed companies. The Association of Certified Fraud Examiners (ACFE) states that financial statement fraud begins three to four years before its detection in a corporation (ACFE, 2022), and investors may suffer losses before detecting the manipulation. Financial statement fraud may threaten the integrity of the financial market, mislead investors about a company's actual financial condition and performance, and interfere with regulatory compliance.

Poland is one of the European Union countries that has experienced dynamic economic growth in recent years. According to the World Bank data, between 2010 and 2021, Poland's Gross Domestic Product (GDP) increased by 44.1%, while the GDP of EU countries grew by

Central European Management Journal
Vol. 34 No. 3, 2026
pp. 488-503
Emerald Publishing Limited
e-ISSN: 2658-2430
p-ISSN: 2658-0845
DOI 10.1108/CEMJ-07-2024-0225

19.1%. Furthermore, Poland's share in the EU's total GDP rose from 3.3 to 3.9%. During this period, the Warsaw Stock Exchange (WSE) also gained in importance. While the number of companies listed on the WSE's main market rose modestly from 400 in 2010 to 430 in 2021, the market capitalization of both domestic and foreign companies increased by PLN 516 bn, i.e. by over 39%, and the number of transactions per session rose by nearly 50%. Furthermore, data from the Federation of European Securities Exchanges indicate that the WSE is the largest stock exchange in Central and Eastern Europe in terms of both turnover and the number of listed companies. In this context, the ability to detect earnings management among companies constitutes a crucial factor, as such practices may undermine investor confidence in the market, potentially leading to capital outflows and economic destabilization. Given Poland's growing role in the EU economy and the increasing importance of the WSE in the region, it is particularly relevant to analyze earnings management practices in the Polish capital market.

The early detection of earnings management is a very important step to take to minimize future losses. Previous studies indicated that financial ratios are effective in financial statement fraud detection (Pazarskis, Drogalas, & Baltzi, 2017; Jan, 2018; Yao, Pan, Yang, Chen, & Li, 2019). Mathematical models use different financial ratios for fraud detection, and previous research confirms their abilities in earnings management detection. Our research approach incorporates two of the best-known mathematical models for fraud detection: the Beneish M-Score and the Dechow F-Score. The M-Score and the F-Score models exploit fundamental signals to predict future earnings, and the key component of false positive investors' costs is the forgone profit or the avoided loss due to not investing in a falsely flagged firm (Beneish & Vorst, 2022). However, Beneish, Lee, and Craig (2013) modified the total accruals ratio (TATA ratio) for the basic M-Score model, whereas the authors used the baseline model (Aghghaleh, Mohamed, & Rahmat, 2016; Husnurrosyidah & Fatihah, 2022). Some researchers (Paolone & Magazzino, 2014; Alfian & Triani, 2019; Svabova, Kramarova, Chutka, & Strakova, 2020; Papik & Papikova, 2020) preferred an alternative five-variable M-Score model to the eight-variable M-Score model for their analysis.

Previous studies on the Polish market examined the effectiveness of the M-Score model from 1999 and the five-element version (Comporek, 2020; Hołda, 2020; Wiszniowski, 2020; Sylwestrzak, 2022; Lesiak, 2024). Only Golec (2019) tested the modified M-Score introduced in 2013. However, instead of fraudulent firms, these studies considered companies that received an adverse or disclaimer opinion from the auditors (Golec, 2019; Sylwestrzak, 2022), did not include control companies (Comporek, 2020; Wiszniowski, 2020), tested the model on a very small number of observations (Hołda, 2020; Wiszniowski, 2020) or used only a single marginal value (Comporek, 2020; Hołda, 2020; Wiszniowski, 2020; Sylwestrzak, 2022). Only Sylwestrzak (2022) adapted the model for the Polish market and proposed the threshold values for indicators, but besides the mentioned limitation of the study, the evaluation of the modified M-Score model's usefulness was based on a sample of six companies, including three fraud ones. Given the dynamic development of the Polish capital market and the limitations of previous research, there remains a clear need for more comprehensive studies that verify the effectiveness of the modified M-Score model using robust methodology and representative data from the Polish context.

This study aims to evaluate the effectiveness of two M-Score and F-Score models, developed in the United States, in detecting earnings manipulation within the Polish market, using a properly constructed sample. The M-Score will be analyzed in three versions: the original model proposed in 1999, the modified version introduced in 2013 and the five-element version, also referred to as the Roxas model. As noted by authors studying bankruptcy prediction for the Polish market, researchers should not apply models from other markets uncritically (Mączyńska & Zawadzki, 2006; Iwanowicz, 2017; Kitowski, Kowal-Pawul, & Lichota, 2022). As a complementary objective, we aimed to investigate whether M-Score modifications proposed for non-US markets improve the identification of financial statement fraud in the Polish context. By comparing multiple versions of the M-Score to the F-Score, the study addresses a critical gap in the literature regarding the applicability and predictive performance of these models outside their original settings. It further evaluates whether model adaptations introduced by other authors yield superior classification metrics compared to the original M-Score model.

This article explores the potential of the M-Score and F-Score models as indicators of fraud in the Polish financial market. We analyzed the annual financial statements of companies listed on the main market of the Warsaw Stock Exchange (WSE) that received a monetary fine from the Polish Financial Supervision Authority (KNF Board) for violation of International Accounting Standards (IAS) or International Financial Reporting Standards (IFRS) and principles in their financial statements during the period 2010–2021. The contribution of this study is threefold. First, it verifies the usefulness of applying the most commonly used statistical models to detect the earnings management of listed companies in Poland. This serves as a response to previous research, where authors compared the accuracy of M-Score and F-Score models. Second, the study examines the effectiveness of the M-Score models that have changed over the years. Third, it verifies the usefulness of the modified M-Score models adapted by the authors to their markets.

The rest of the article is organized as follows: Section 2 presents the relevant literature on the Beneish and Dechow models. Section 3 describes the research methodology applied. Section 4 shows the analysis results. Section 5 presents a future research agenda derived from our findings in Section 4. Finally, Section 6 presents a discussion and points to possible further research directions.

Literature review

Beneish M-Score models

Discriminant models are one approach to classifying companies as manipulators or non-manipulators. The selection of indicators for the model is based on an analysis of the characteristic values observed in fraud and control firms. The advantages of discriminant models include their transparency and ease of interpretation, and their effectiveness has been demonstrated in the literature (Wuerges & Borba, 2014; Kanapickiene & Grundiene, 2015; Dong, Liao, & Zhang, 2018).

Beneish (1999) developed the M-Score model, which is one of the best-known methods for detecting accounting manipulations in the world. Beneish created the M-Score model (Formula 1) based on eight financial indicators, where the values for companies engaging in earnings manipulation will exhibit deviations from normal operations, which track changes in sales, incomes, accruals and margins (Beneish, 1999). Based on the estimation of relative costs of type I and type II errors, Beneish identified the three marginal values of the M-Score (−2.22, −1.78 and −1.49). In the literature, scholars mostly use the first two values. However, based on more recent data, Beneish *et al.* (2013) changed the TATA indicator on the ACCRUALS ratio and used only the (−1.78) marginal value of the M-Score model (Formula 2). The change in the accrual indicator in the model resulted from a change in the method of calculating accruals from a balance sheet approach to an income statement approach, which, as noted by the authors, was due to the evolution of the approach in the literature and the limited predictive power of the M-Score model, particularly in the middle quintiles of accruals results. On the other hand, some researchers (Buljubasic & Halilbegovic, 2017; Anning & Adusei, 2020; Nyakarimi, Kariuki, & Kariuki, 2020) prefer an alternative five-element M-Score model (Formula 3) to the eight-element M-Score model. The five-element M-Score model omits the SGAI, LEVI and TATA indicators and changes the marginal value of the M-Score to −2.76. However, some authors do not use the second change in the five-element M-Score model but use the standard marginal values. Numerous studies have found the five-variable M-Score model to be more accurate than the eight-variable model (Anning & Adusei, 2020; Lehenchuk, Mostenska, Tarasiuk, Polishchuk, & Gorodysky, 2021), but others did not confirm these results (Roxas, 2011; Buljubasic & Halilbegovic, 2017). The M-Score models can use an overall benchmark to determine whether the financial statement suggests earnings manipulation. By deconstructing the M-Score model into its constituent components, researchers can determine whether each calculation contains anomalous deviations or anomalies that require further investigation (Mantone, 2013).

$$M - Score = -4.84 + 0.920 * DSRI + 0.528 * GMI + 0.404 * AQI + 0.892 * SGI + 0.115 * DEPI - 0.172 * SGAI + 4.679 * TATA - 0.327 * LEVI \quad (1)$$

$$M - Score\,(modified) = -4.84 + 0.920 * DSRI + 0.528 * GMI + 0.404 * AQI + 0.892 * SGI + 0.115 * DEPI - 0.172 * SGAI + 4.679 * ACCRUALS - 0.327 * LEVI \quad (2)$$

$$M - Score\,(5 - element) = -6.065 + 0.823 * DSRI + 0.906 * GMI + 0.593 * AQI + 0.717 * SGI + 0.107 * DEPI \quad (3)$$

Building upon the above arguments, we formulated the following research question:

RQ1. Do eight-variable M-Score models have a higher F-measure metric than the five-element M-Score models?

Dechow F-Score model

The F-Score model is a financial fraud detection model that uses the scaled logistic probability technique developed by Dechow, Ge, Larson, and Sloan (2011). They developed it based on financial and nonfinancial variables that reflected the potential for overstating earnings in areas such as accrual quality, financial performance, nonfinancial measures, off-balance sheet activities and market-based measures. The final F-Score model relies on seven indicators from the financial statements. To calculate the F-Score value, one must first calculate the predicted value F (Formula 4). Next, the F value is converted into a probability value, which is divided by the unconditional probability of misstatement (Formula 5). The F-Score result shows how many times greater a certain company's probability of falsifying financial statements than a randomly selected company from the entire surveyed population. In other words, if the F-score is greater than 1, it indicates that there is a statistically high probability that the financial statements of a firm contain misstatements. However, an F-Score value greater than 1 can also indicate that the company changed its financial statements. Numerous authors have shown that the F-Score model performs with better accuracy than other fraud models (Hakami, Rahmat, Yaacob, & Saleh, 2020; Putra & Dinarjito, 2021; Saleh, Aladwan, Alsinglawi, Saleh, & Mahmoud, 2021; Hou, Lin, Li, Liu, & Zheng, 2023). However, the unconditional probability of misstatement impacts F-Score results, which were estimated for the US market sample.

$$F = -7.893 + 0.790 * RSST + 2.518 * CH\_REC + 1.191 * CH\_INV + 1.979 * SOFT\_ASSETS + 0.171 * CH\_CS - 0.932 * CH\_ROA + 1.029 * ISSUE \quad (4)$$

$$F - Score = \frac{\frac{e^F}{1+e^F}}{0.0037} \quad (5)$$

Table 1 reports the method of calculating the individual ratios of the M-Score and F-Score models.

Table 1. M-Score and F-Score models indicators

Ratio	Formula
DSRI	(Net receivables in year t/Sales in year t)/(Net receivables in year t-1/Sales in year t-1)
GMI	[(Sales in year t-1 − Cost of goods sold in year t-1)/Sales in year t-1]/[(Sales in year t − Cost of goods sold in year t)/Sales in year t]
AQI	[1 − (Current assets in year t + Property, plant & equipment in year t)/Total assets in year t]/[1 − (Current assets in year t-1 + Property, plant & equipment in year t-1)/Total assets in year t-1]
SGI	Sales in year t/Sales in year t-1
DEPI	[Depreciation in year t-1/(Depreciation in year t-1 + Property, plant & equipment in year t-1)]/[Depreciation in year t/(Depreciation in year t + Property, plant & equipment in year t)]
SGAI	(Sales, general and administrative cost in year t/Sales in year t)/(Sales, general and administrative cost in year t-1/Sales in year t-1)
LEVI	[(Current liabilities in year t + Total long term debt in year t)/Total assets in year t]/[(Current liabilities in year t-1 + Total long term debt in year t-1)/Total assets in year t-1]
TATA	[(Change in current assets − Change in cash) − (Change in current liabilities − Change in current maturities of long term debt − Change in income tax payable) - Depreciation and amortization in year t]/Total assets in year t]
ACCRUALS	(Net income in year t − Cash flow from operation in year t)/Total assets in year t
RSST	(ΔWC + ΔNCO + ΔFIN)/Average total assets WC = (Current assets − Cash and Short-term investment) − (Current liabilities − Debt in current liabilities) NCO = Total assets − Current assets − Investment and advances − (Total liabilities − Current liabilities - Long-term debt) FIN = (Short-term investment + Long-term investment) − (Long-term debt + Debt in current liabilities + Preferred stock)
CH_REC	(Accounts receivables in year t − Accounts receivables in year t-1)/Average total assets in year t
CH_INV	(Inventory in year t − Inventory in year t-1)/Average total assets in year t
SOFT_ ASSETS	(Total assets in year t - Property, plant & equipment in year t − Cash and cash equivalents in year t)/Total assets in year t
CH_CS	(Sales in year t − (Accounts receivables in year t − Accounts receivables t-1))/(Sales in year t-1 − (Accounts receivables in year t-1 − Accounts receivables in year t-2)) − 1
CH_ROA	(Net income in year t/Average total assets in year t) − (Net income in year t-1/Average total assets in year t-1)
ISSUE	Dummy variable equals 1 if additional securities were issued during the year of the manipulation and 0 if no securities were issued

Source(s): Sourced from Beneish (1999), Beneish *et al.* (2013) and Dechow *et al.* (2011)

Previous studies have shown that the F-Score model better detects financial statement fraud than the M-Score model (Aghghaleh *et al.*, 2016; Hakami *et al.*, 2020; Omeir, Vasiliauskaitė, & Soleimanizadeh, 2023; Chakrabarty, Moulton, Pugachev, & Wang, 2024). On the other hand, some research (Deniswara, Kesuma, & Louis, 2022; Husnurrosyidah & Fatihah, 2022; Hou *et al.*, 2023; Marais, Vermaak, & Shewell, 2023; Shu, He, Li, & Gong, 2022) found that the M-Score method is more effective than the F-Score method for detecting fraudulent financial statements. In turn, Wisdianti, Arum, and Wijaya (2022) and Xu, Shi, Xia, and Liu (2023) showed that both models achieved the same accuracy. Table 2 reports a comparison of the results of the M-Score and F-Score models in the analyzed studies.

Based on the arguments above, we asked the second research question:

RQ2. Does the F-Score model have a lower F-measure metric than both of the eight-variable M-Score models?

Table 2. Literature review

Author(s)	Country	No. of companies	M-Score (accuracy)	F-Score (accuracy)
Aghghaleh *et al.* (2016)	Malaysia	164	73.2%	76.2%
Chakrabarty *et al.* (2024)	USA	452	58.9%	59.3%
Deniswara *et al.* (2022)	Indonesia	28	nd	nd
Hakami *et al.* (2020)	Gulf Cooperation Council	365	83.5%	84.4%
Husnurrosyidah and Fatihah (2022)	Indonesia	47	68.1%	66.0%
Hou *et al.* (2023)	USA	10	70.0%	60.0%
Marais *et al.* (2023)	South Africa	274	84.8%	60.5%
Omeir *et al.* (2023)	Iran	12	83.0%	75.0%
Shu *et al.* (2022)	USA	7	85.7%	28.6%
Wisdianti *et al.* (2022)	Indonesia	8	25.0%	25.0%
Xu *et al.* (2023)	USA	20	nd	nd

Note(s): Nd – no data
Source(s): Authors' own elaboration

Modified M-Score models

Several authors (Ozcan, 2018; Erdogan & Erdogan, 2020; Shakouri, Taherabadi, Ghanbari, & Jamshidinavid, 2021; Sylwestrzak, 2022; Marais *et al.*, 2023) have adapted the Beneish model to the conditions of their own country, indicating a better fit of the model to national regulations. Formulas (6)–(11) present proprietary versions of M-Score models.

$$M - Score\,(Erdogan) = 3.642 + 2.958 * AQI + 2.980 * SGAI \tag{6}$$

$$\begin{aligned} M - Score\,(Marais1) = {} & -1.6897 + 0.0776 * DSRI - 0.1463 * GMI - 0.0692 * AQI \\ & - 0.2178 * SGI + 0.0307 * DEPI - 0.2042 * SGAI \\ & - 0.0603 * LEVI + 1.4912 * TATA \end{aligned} \tag{7}$$

$$\begin{aligned} M - Score\,(Marais2) = {} & -1.7944 + 0.1194 * DSRI - 0.1355 * GMI - 0.0722 * AQI \\ & - 0.1280 * SGI + 0.0347 * DEPI - 0.1875 * SGAI \\ & - 0.1538 * LEVI - 0.3350 * ACCRUALS \end{aligned} \tag{8}$$

$$\begin{aligned} M - Score\,(Ozcan) = {} & 14.583 - 1.162 * DSRI - 2.846 * GMI - 1.770 * AQI \\ & - 1.329 * SGI - 1.447 * DEPI + 1.011 * SGAI - 6.453 * LEVI \\ & - 19.313 * TATA \end{aligned} \tag{9}$$

$$\begin{aligned} M - Score\,(Shakouri) = {} & -20.42584 + 6.696443 * DSRI + 3.335502 * GMI \\ & + 2.701114 * AQI + 6.010995 * SGI + 0.849257 * DEPI \\ & - 1.315193 * SGAI - 2.924618 * LEVI + 32.82708 * TATA \end{aligned} \tag{10}$$

$$M - Score\,(Sylwestrzak) = 1.0181 - 0.4078 * GMI - 0.5693 * SGI - 0.8672 * DEPI - 4.0402 * ACCRUALS \tag{11}$$

In these proprietary M-Score models, the authors used different marginal values. Erdogan and Erdogan (2020) did not specify a marginal value, Marais *et al.* (2023) used the marginal value equal to -1.9653 (for M-Score with TATA ratio) and -1.9735 (for M-Score with ACCRUALS ratio) and Sylwestrzak (2022) used the marginal value equal to -0.0544. Ozcan (2018) and Shakouri *et al.* (2021) used a marginal value equal to -1.78. Therefore, we formulated the final research question as follows:

RQ3. Do modified M-Score models have a higher F-measure metric than eight-variable M-Score models?

Research methods

In the Polish legal system, the Polish Financial Supervision Authority (UKNF Board) is one of the bodies that ensures proper functioning, stability, security, transparency and trust in the financial market. The UKNF Board also protects the interests of market participants and imposes financial or legal sanctions for noncompliance with the IFRS guidelines.

The sample includes 60 observations of public companies listed on the WSE that received a monetary fine from the UKNF Board in the context of noncompliance with IAS or IFRS principles during the period 2010–2021. Fines imposed by the UKNF Board are related to situations of intentional omission or distortion of financial information by an entity and do not result from unintentional errors. The UKNF most often fines companies for:

(1) failing to provide liquidity risk disclosures and a lack of analysis of the maturity dates of financial assets held for liquidity risk management under IFRS 7;

(2) not including all subsidiaries of the parent company in consolidated financial statements, related to IAS 27;

(3) a lack of assessment of an entity's ability to continue as a going concern, related to IAS 1 and

(4) failure to disclose merger amounts under IFRS 3.

We had to use the list of companies fined by the UKNF Board because, in the Polish legal system, there is no legal act that refers to the definition of financial statement fraud. We matched the 60 fraudulent observations with 60 control observations, where we matched each fraudulent company with a corresponding control firm based on:

(1) Firm size: a control firm was considered similar if total assets were within $+/-30\%$ of total assets for the fraud firm in the fraud year;

(2) Financial year: annual reports for the control firm were available for the same period as the fraud firm and

(3) The industry: We reviewed the firm to identify a control firm within the same three-digit Standard Industrial Classification (SIC) as the fraud firm. If we found no match, we used the two-digit codes.

We collected the financial data (expressed in monetary units) from the annual financial statements of public companies.

We conducted the selection of control companies to match fraudulent companies in a 1:1 ratio based on a review of previous studies comparing both models (Aghghaleh *et al.*, 2016;

Hou *et al.*, 2023; Omeir *et al.*, 2023; Xu *et al.*, 2023). Applying the indicated criteria allows for the evaluation of whether control companies, which belong to the same industry and are similar in size, will be classified as non-manipulators. This approach enables an assessment of the model's specificity, not just its sensitivity, as in the studies by Chakrabarty *et al.* (2024), Deniswara *et al.* (2022), Shu *et al.* (2022) or Wisdianti *et al.* (2022). However, this contributes to sample selection bias, a point raised by Zmijewski (1984), Platt and Platt (2002) and Gruszczyński (2019). Nevertheless, this bias may affect the evaluation of individual cases but does not significantly impact the classification accuracy of the method (Zmijewski, 1984). Hakami *et al.* (2020) and Marais *et al.* (2023) used data from all public companies to verify the applicability of the M-Score and F-Score models. However, this affects the model's specificity, as companies from all sectors undergo analysis, differing in the nature of their business, the stage of their business cycle, their position in the industry and their size.

In the reviewed literature, the most commonly used metric for evaluating the performance of the M-Score and F-Score models was overall accuracy. However, as noted by Marais *et al.* (2023), accuracy is not appropriate due to the scarcity of earnings management cases. One may also assess the predictive performance of a model using sensitivity and specificity, which are presented to provide a clearer picture of classification performance.

In this study, we adopted the F-measure as the primary evaluation metric. The F-measure is the harmonic mean of precision and sensitivity, and thus symmetrically incorporates both indicators into a single metric. The classification performance metrics used in this study are calculated as follows:

(1) F-measure is the harmonic mean of sensitivity and precision, where precision is the number of true positive results divided by the total number of observations predicted as positive;

(2) Accuracy is the percentage of correctly classified fraud and control firms out of all firms;

(3) Sensitivity is the percentage of fraud companies correctly classified as manipulators out of all fraud firms and

(4) Specificity is the percentage of control firms correctly classified as non-manipulators out of all control firms.

Results

Descriptive statistics

Table 3 presents the results of the Mann–Whitney U-test. The test showed that there was a significant difference between the variables GMI, SGAI, LEVI and ACCRUALS from the M-Score model and RSST, CH_INV, SOFT_ASSETS, CH_CS and ISSUE from the F-Score model for the sample of fraud and control companies.

The low GMI ratio may indicate poor prospects for the firm, and manipulators may try to hide information about low performance, inflating revenues or understating costs. Similarly, a high SGAI value may signal a decline in sales and administrative efficiency, which could lead a company's management to commit financial statement fraud. A high value for the LEVI ratio indicates that the company has too many liabilities, and that is why incentives for fraud increase. A high level for ACCRUALS suggests that the net income increases significantly, but cash flow remains unchanged or increases slightly, which may indicate the manipulation of corporate earnings.

The RSST ratio is based on working capital accruals to include changes in long-term operating assets and liabilities, providing an opportunity for the company's management to manipulate financial statements to achieve the appropriate level of net income. The RSST ratio

Table 3. Summary statistics, by group

Ratio	Minimum Fraud	Control	Maximum Fraud	Control	Mean Fraud	Control	Std. deviation Fraud	Control
DSRI	0.061	0.011	35.960	2.048	2.679	0.950	6.076	0.406
GMI**	−19.618	−0.779	17.945	3.198	0.481	1.041	3.915	0.496
AQI	−0.525	−0.289	80.528	11.905	2.489	1.248	10.323	1.575
SGI	0.002	0.209	14.500	64.231	1.235	2.207	1.904	8.173
DEPI	0.000	0.046	6.821	5.095	1.106	1.142	0.956	0.766
SGAI**	0.049	0.023	29.910	2.358	2.596	0.967	5.738	0.321
LEVI***	0.206	0.069	3283.93	1.573	55.931	0.945	423.797	0.258
TATA	−0.808	−1.438	0.784	1.162	−0.083	−0.039	0.272	0.285
ACCRUALS***	−1.872	−1.614	0.274	0.332	−0.190	−0.064	0.360	0.227
RSST***	−3.962	−24.471	0.604	0.833	−0.144	−0.405	0.564	3.168
CH_REC	−0.737	−0.689	0.537	0.368	−0.011	−0.008	0.177	0.108
CH_INV**	−0.279	−0.087	0.433	0.389	−0.006	0.013	0.110	0.061
SOFT_ASSETS**	0.073	0.051	1.000	0.970	0.654	0.536	0.250	0.262
CH_CS*	−3.203	−2.573	3.565	2.997	−0.071	0.035	1.002	0.639
CH_ROA	−3.910	−0.206	1.353	0.244	−0.138	0.004	0.604	0.071
ISSUE**	0.000	0.000	1.000	1.000	0.517	0.300	0.504	0.462

Note(s): The variables are defined in Tables 1 and 2. ***, ** and * denote significance at the 1, 5 and 10% levels, respectively. We applied the Mann–Whitney U-test

Source(s): Authors' own elaboration

results for control companies are skewed by one outlier. Without considering outliers, the statistical values of the RSST ratio show significantly less variability compared to fraudulent companies. A large change in the CH_INV ratio can significantly affect gross profit and signal manipulation of the company's earnings. Fraud firms have a lower mean value of this indicator than control firms but significantly greater variability, as measured by the standard deviation, which leads to a higher coefficient of variation as well as higher minimum and maximum values. A high change in the SOFT_ASSETS ratio indicates that the management may be more likely to adjust assumptions about asset valuation other than cash and property, plant and equipment to meet short-term income targets. A high value of the CH_CS ratio suggests that the growth rate of sales does not relate to the growth rate of accounts receivable, which may indicate a manipulation of the company's business selling policy. The ISSUE ratio is linked with shares issuance, which suggests that the firm's management can convert stock options into shares and sell shares at times of high prices.

M-Score and F-Score models

Table 4 presents the classification results of the M-Score and F-Score models.

The basic M-Score model with a cut-off equal to (−2.22) classified correctly 26 fraud firms (43.3%) and 40 control firms (66.7%), while the basic M-Score model with a cut-off equal to (−1.78) had 20 (33.3%) fraud firms and 50 (83.3%) firms correctly classified, respectively. However, the change in the cut-off threshold decreased the F-measure by 4.7% points. The modified M-Score model correctly classified 20 (33.3%) fraud firms and 52 (86.7%) control firms. For the five-element M-Score model, as the marginal value increases, the model's sensitivity decreases (from 31.7 to 23.3%), but its specificity increases (from 61.7 to 90.0%). The five-element M-Score model with a cut-off equal to −2.76 achieved the highest F-measure, as it correctly classified 70 firms. This means that the eight-variable M-Score models have a lower value F-measure than the five-element M-Score models. The results from the F-Score model revealed that 20 fraud firms (33.3%) were classified as manipulators, while 51 control firms (85.0%) were classified as non-manipulators.

Table 5 presents the classification results of proprietary M-Score models.

Table 4. M-Score and F-Score model results

Model	M-Score (basic) cut-off = −2.22	M-Score (basic) cut-off = −1.78	M-Score (modified) cut-off = −1.78	M-Score (5-element) cut-off = −2.76	M-Score (5-element) cut-off = −2.22	M-Score (5-element) cut-off = −1.78	F-score
Accuracy	55.0%	58.3%	*60.0%*	58.3%	*60.0%*	56.7%	59.2%
Sensitivity	43.3%	33.3%	33.3%	*55.0%*	31.7%	23.3%	33.3%
Specificity	66.7%	83.3%	86.7%	61.7%	88.3%	*90.0%*	85.0%
F-measure	49.1%	44.4%	45.5%	*56.9%*	44.2%	35.0%	44.9%
Type I error	33.3%	16.7%	13.3%	38.3%	11.7%	*10.0%*	15.0%
Type II error	53.7%	66.7%	66.7%	*45.0%*	68.3%	76.7%	66.7%

Source(s): Authors' own elaboration

Table 5. The modified M-Score model results

Model	M-Score (Erdogan) cut-off = 11.59	M-Score (Marais1) cut-off = −1.9653	M-Score (Marais2) cut-off = −1.9735	M-Score (Ozcan) cut-off = −1.78	M-Score (Shakouri) cut-off = −1.78	M-Score (Sylwestrzak) cut-off = −0.0544
Accuracy	65.0%	54.2%	55.7%	44.2%	56.7%	*70.8%*
Sensitivity	35.0%	13.3%	17.7%	*71.7%*	35.0%	48.3%
Specificity	*95.0%*	*95.0%*	*95.0%*	16.7%	78.3%	93.3%
F-measure	50.0%	22.5%	27.4%	56.2%	44.7%	*62.4%*
Type I error	*5.0%*	*5.0%*	*5.0%*	83.3%	21.7%	6.7%
Type II error	65.0%	86.7%	82.3%	*28.3%*	65.0%	51.7%

Source(s): Authors' own elaboration

Only the Sylwestrzak M-Score version had higher accuracy and F-measure than both Beneish M-Score models, where the high accuracy resulted from the model's high specificity. Ozcan's M-Score version achieved the second-highest F-measure, primarily due to its strong classification performance for fraud firms. However, in terms of accuracy, it performed the weakest compared to the proprietary M-Score models. In turn, Erdogan's M-Score version had higher accuracy and F-measure value than both Beneish models. However, the authors did not provide the marginal M-Score value; therefore, we computed the probability cutoffs that would minimize the expected costs of misclassification with relative costs of Type I to Type II errors equal to 1:1. Therefore, for our sample, the marginal value equaled 11.59, as using one of the values suggested in the literature classified all observations as manipulators.

Discussion

RQ1. Do eight-variable M-Score models have a higher F-measure metric than the five-element M-Score models?

The five-element M-Score model with a marginal value (−2.76) proposed by Roxas (2011) had a better F-measure and also accuracy than the basic M-Score model, and these results are consistent with previous studies (Roxas, 2011; Anning & Adusei, 2020; Lehenchuk *et al.*, 2021). This is due to the significantly better classification of fraud firms than the basic M-Score model. However, using the marginal value (−2.22) worsens sensitivity but greatly improves specificity, which decreases F-measure. Moreover, using the marginal value (−1.78) increases the classification accuracy for control firms but reduces the classification accuracy for fraud companies to a greater extent, making the model less accurate than the basic M-Score model with the marginal value (−1.78). The changes proposed by Beneish *et al.* (2013) increased the accuracy and specificity of model fitting compared to the basic M-Score model. However, the F-measure was lower than that of the M-Score model with a marginal value (−2.22), due to a decline in the model's sensitivity. Moreover, we achieved similar F-measure and accuracy using a 5-element M-Score model with a marginal value (−2.22). The results confirmed the previous findings (Anning & Adusei, 2020; Nyakarimi *et al.*, 2020; Lehenchuk *et al.*, 2021) showing higher accuracy of the five-element M-Score model, but only in relation to the basic M-Score model. This result relates to the higher specificity of the modified M-Score model compared with the basic M-Score model. However, Kamal, Salleh, and Ahmad (2016) suggest that the basic M-Score model has higher accuracy in detecting financial statement fraud than the modified M-Score model.

RQ2. Does the F-Score model have a lower F-measure metric than both of the eight-variable M-Score models?

The F-measure for the F-Score model was slightly higher than that of the basic M-Score model with the marginal value (−1.78), but considerably lower than the model using the cut-off (−2.22). However, the accuracy of the F-Score model was higher than the basic M-Score model, as it significantly better classified the control firms, confirming the results of most authors (Aghghaleh *et al.*, 2016; Hakami *et al.*, 2020; Wisdianti *et al.*, 2022). On the other hand, we omitted the modified M-Score and the five-element M-Score model with a marginal value of (−2.22) from the literature comparison. Their F-measure and accuracy were higher than the F-Score results due to better classification of control firms, which confirms the results by Saleh *et al.* (2021) and Hou *et al.* (2023). Our study fills this gap and shows that the F-Score model had lower F-measure and accuracy than both indicated models. However, this needs to be the subject of further research in other markets. Furthermore, the unconditional probability of misstatement value requires estimation because, in the research, it serves as a reference value. Meanwhile, Dechow *et al.* (2011) showed that it refers to the number of fraud companies in the sample. The change in the unconditional probability of misstatement from 0.0037 to 0.0064, calculated based on the proportion of fraudulent companies to the number of nonfinancial public companies listed on the WSE during the study period, correctly classified 11 fraudulent companies (18.3%) and 58 control companies (96.7%). Compared to the unconditional probability of misstatement used by Dechow *et al.* (2011), the adjustment of this value to the conditions of the Polish market improved the model's specificity by 11.7% points but reduced its sensitivity by 15% points. In this context, the change in the unconditional probability of misstatement decreased the model's predictive accuracy.

RQ3. Do modified M-Score models have a higher F-measure metric than eight-variable M-Score models?

In most cases, the verification of the usefulness of the Beneish model, adapted by the selected authors, shows that its models had lower F-measure and accuracy than the modified M-Score or five-element M-Score models on the Polish sample. The exception to this was the modification of the M-Score model proposed by Sylwestrzak (2022). This results from the combination of high specificity and moderate sensitivity of the model compared to the modified versions of the M-Score proposed by other authors, as well as the eight-element and five-element versions of the M-Score model. However, this model was built on a sample of Polish companies that had received an adverse or disclaimer opinion from the auditors but had not received a monetary fine from the Polish Financial Supervision Authority for violation of IAS/IFRS principles related to their financial statements, and this may have affected the classification results of our sample. Furthermore, the proposed threshold values for the indicators in the modified M-Score model proposed by Sylwestrzak (2022) classified control firms much better – ranging from 68.3% for SGI to 91.7% for DEPI, but performed much worse for fraud firms – ranging from 23.3% for DEPI to 50.0% for SGI. In this case, one should adjust the critical values to better classify fraud companies.

The Mann–Whitney test results showed that more variables have significant differences between groups in the F-Score model compared to the M-Score model. Moreover, an additional logistic regression showed that, from the basic M-Score model, only two indicators (SGAI and LEVI) were significant, while, from the modified M-Score model, only the LEVI ratio was significant. Only the DSRI ratio was significant from the five-element M-Score model. These variables positively correlated with the likelihood of fraud occurring. This means that increasing the day sales outstanding ratio, sales, general and administrative costs and financial leverage increases the likelihood of financial data fraud. On the other hand, the regression of F-Score ratios showed that three indicators were significant: SOFT_ASSETS, CH_ROA and ISSUE. This means that increasing the value of soft assets, greater changes in the return on assets ratio and issuance of additional securities increase the likelihood of financial data fraud. However, the regression results indicate a weak fit of the M-Score and F-Score indicators to the Polish companies' financial data, as measured by the R-squared value, particularly for the five-element components of the M-Score and F-Score models. Table 6 presents the logistic regression results for the M-Score and F-Score models.

Conclusions

We investigated the ability of the two most popular fraud detection models, the M-Score and the F-Score, to identify manipulating firms in the Polish market. In the literature on earnings management detection, scholars apply various versions of the M-Score model. Therefore, we decided to verify which model type best fits the realities of the Polish market. The results showed that the five-element M-Score model with a marginal value of (−2.22) has the highest F-measure in detecting financial fraud. Moreover, this model has a better F-measure value and also accuracy than the F-Score model, due to the slightly higher specificity of the M-Score models. In turn, proprietary M-Score models had lower F-measure than the modified M-Score and F-Score models, except for the M-Score model proposed by

Table 6. Logistic regression results

Variables	M-Score	M-Score (modified)	M-Score (5-element)	F-score
Constant	−5.267*** (1.708)	−3.197** (1.276)	−0.027 (0.538)	−1.267** (0.521)
DSRI	0.388 (0.349)	0.561 (0.345)	0.515* (0.302)	
GMI	0.004 (0.090)	−0.020 (0.079)	−0.070 (0.078)	
AQI	0.119 (0.272)	0.027 (0.099)	0.025 (0.108)	
SGI	−0.004 (0.059)	0.010 (0.048)	−0.007 (0.040)	
DEPI	−0.133 (0.557)	−0.737 (0.519)	−0.542 (0.418)	
SGAI	1.219* (0.732)	0.999 (0.640)		
LEVI	3.401*** (1.133)	1.893** (0.745)		
TATA	2.387 (1.754)			
ACCRUALS		−1.513 (1.181)		
RSST				0.058 (0.151)
CH_REC				2.673 (1.934)
CH_INV				−3.273 (2.641)
SOFT_ASSETS				1.479* (0.807)
CH_CS				−0.291 (0.304)
CH_ROA				−2.248** (1.134)
ISSUE				0.831** (0.414)
Observations	120	120	120	120
Pseudo *R*-squared	21.6%	21.5%	7.9%	11.6%

Note(s): Standard deviations are in parentheses. Variables are defined in Tables 1 and 2, where ***, ** and * denote significance at the 1, 5 and 10% levels, respectively

Source(s): Authors' own elaboration

Sylwestrzak. On the other hand, the results obtained for Sylwestrzak's M-Score model confirm the conclusions of authors who have examined the use of foreign bankruptcy prediction models for the Polish market (Iwanowicz, 2017; Kitowski *et al.*, 2022; Mączyńska & Zawadzki, 2006). The authors state that due to legal differences, macroeconomic conditions and the specific nature of the accounting system, it is not appropriate to apply these models unconditionally. This is an important piece of information for users of financial statements, who should primarily rely on the results of models developed for the Polish market when assessing the likelihood of financial statement fraud. In addition to demonstrating high F-measure and accuracy, Sylwestrzak's M-Score model achieved notable specificity, thereby reducing the need for extensive testing to confirm that a given company is not engaging in manipulative practices. While the model did not correctly classify all observations, its results are more precise compared to the original M-Score or F-Score models, effectively narrowing the scope of financial data requiring detailed analysis. For the Polish market, key parameters were the gross margin index, sales, general and administrative expenses index, the accruals index measured by the income statement approach depreciation index. Analyzing the statistical values of these indicators enables investors to assess whether changes in these parameters are more characteristic of manipulative or non-manipulative companies.

This article demonstrates that the application of M-Score and F-Score models to detect earnings management should account for modifications tailored to the specific market, as such adaptations better reflect the unique characteristics of each country. The limitation of our study is the small number of observations included in the analysis, where we selected one control firm for one fraud company. An extension of our research would be to include all public companies in the analyzed period. Moreover, the M-Score models use an overall benchmark to determine whether the financial statement suggests earnings manipulation or an attempt to conceal embezzlement funds (Mantone, 2013).

References

Aghghaleh, S. F., Mohamed, Z. M., & Rahmat, M. M. (2016). Detecting financial statement frauds in Malaysia: Comparing the abilities of Beneish and Dechow models. *Asian Journal of Accounting & Governance*, *7*, 57–65. doi: 10.17576/AJAG-2016-07-05.

Alfian, F., & Triani, N. (2019). Fraudulent financial reporting detection using Beneish M-score model in public companies in 2012-2016. *Asia Pacific Fraud Journal*, *4*(1), 27–42. doi: 10.21532/apfj.001.19.04.01.03.

Anning, A., & Adusei, M. (2020). An analysis of financial statement manipulation among listed manufacturing and trading firms in Ghana. *Journal of African Business*, *23*(1), 1–15. doi: 10.1080/15228916.2020.1826856.

Association of Certified Fraud Examiners (2022). Report to the nations on occupational fraud and abuse. Available from: https://acfepublic.s3.us-west-2.amazonaws.com/2022+Report+to+the+Nations.pdf (accessed 10 June 2023).

Beneish, M. D. (1999). The detection of earnings manipulation. *Financial Analysts Journal*, *55*(5), 24–36. doi: 10.2469/faj.v55.n5.2296.

Beneish, M. D., & Vorst, P. (2022). The cost of fraud prediction errors. *The Accounting Review*, *97*(6), 91–121. doi: 10.2308/TAR-2020-0068.

Beneish, M. D., Lee, Ch., & Craig, D. (2013). Earnings manipulation and expected returns. *Financial Analysts Journal*, *69*(2), 57–82. doi: 10.2469/faj.v69.n2.1.

Buljubasic, E., & Halilbegovic, S. (2017). Detection of financial statement fraud using beneish model. In M. Obralic, & N. Tandir*International Conference on Economic and Social Studies* (pp. 252–262). doi: 10.14706/icesos178.

Chakrabarty, B., Moulton, P. C., Pugachev, L., & Wang, X. F. (2024). Catch me if you can: In search of accuracy, scope, and ease of fraud prediction. *Review of Accounting Studies*, *30*(2), 1268–1308.

doi: 10.1007/s11142-024-09854-4. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3352667 [accessed 15 May 2024].

Comporek, M. (2020). The effectiveness of the Beneish model in the detection of accounting violations – the example of companies sanctioned by the Polish Financial Supervision Authority. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, *64*(10), 18–30. doi: 10.15611/pn.2020.10.2.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. doi: 10.1111/j.1911-3846.2010.01041.x.

DeFond, M. L. (2010). Earnings quality research: Advances, challenges and future research. *Journal of Accounting and Economics*, *50*(2-3), 402–409. doi: 10.1016/j.jacceco.2010.10.004.

Deniswara, K., Kesuma, J. T., & Louis, V. (2022). Forensic accounting on Indonesian energy sector with Beneish M-score model analysis. In *ICEEG' 22 Proceedings of the 6th International Conference on E-Commerce, E-Business and E-Government*, New York. Association for Computing Machinery.

Dong, W., Liao, S., & Zhang, Z. (2018). Leveraging financial social media data for corporate fraud detection. *Journal of Management Information Systems*, *35*(2), 461–487. doi: 10.1080/07421222.2018.1451954.

Erdogan, M., & Erdogan, E. (2020). Financial statement manipulation: A beneish model application. In S. Grima, E. Boztepe, & P. Baldacchino (Eds.), *Contemporary Issues Audit Managerial Forensic Accounting*. Leeds: Emerald Publishing. doi: 10.1108/S1569-375920200000102014.

Golec, A. (2019). Ocena skuteczności modelu Beneisha w wykrywaniu manipulacji w sprawozdaniach finansowych. *Universitatis Lodziensis. Folia Oeconomica*, *2*(341), 161–182. doi: 10.18778/0208-6018.341.10.

Gruszczyński, M. (2019). On unbalanced sampling in bankruptcy prediction. *International Journal of Financial Studies*, *7*(2), 28. doi: 10.3390/ijfs7020028.

Hakami, T. A., Rahmat, M. M., Yaacob, M. H., & Saleh, N. M. (2020). Fraud detection gap between auditor and fraud detection models: Evidence from gulf cooperation council. *Asian Journal of Accounting & Governance*, *13*, 1–13. doi: 10.17576/AJAG-2020-13-01.

Hołda, A. (2020). Using the beneish M-score model: Evidence from non-financial companies listed on the Warsaw stock exchange. *Investment Management and Financial Innovations*, *17*(4), 389–401. doi: 10.21511/imfi.17(4).2020.33.

Hou, B., Lin, R., Li, X., Liu, Z., & Zheng, J. (2023). Using Beneish M-score and F-score to detect financial fraud. In X. Li, Y. Yuan, & J. Kent (Eds.). *Proceedings of the 6th International Conference on Economic Management and Green Development. Applied Economics and Policy Studies* (pp. 1383–1400) Singapore: Springer. doi: 10.1007/978-981-19-7826-5_131.

Husnurrosyidah, H., & Fatihah, I. (2022). Fraud detecting using Beneish M-score and F-score: Which is more effective?. *Equilibrium Jurnal Ekonomi Syariah*, *10*(1), 137–151. doi: 10.21043/equilibrium.v10i1.15351.

Iwanowicz, T. (2017). Ocena założenia o kontynuacji działalności przedsiębiorstwa z punktu widzenia biegłych rewidentów. *Zeszyty Teoretyczne Rachunkowości*, *95*, 9–30. doi: 10.5604/01.3001.0010.6564.

Jan, C. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: Evidence from Taiwan. *Sustainability*, *10*(2), 513. doi: 10.3390/su10020513.

Kamal, M. E. M., Salleh, M. F. M., & Ahmad, A. (2016). Detecting financial statement fraud by Malaysian public listed companies: The reliability of the Beneish M-score model. *Jurnal Pengurusan*, *46*, 23–32. doi: 10.17576/pengurusan-2016-46-03.

Kanapickiene, R., & Grundiene, Z. (2015). The model of fraud detection in financial statements by means of financial ratios. *Procedia-Social and Behavioral Sciences*, *213*, 321–327. doi: 10.1016/j.sbspro.2015.11.545.

Kitowski, J., Kowal-Pawul, A., & Lichota, W. (2022). „Identifying symptoms of bankruptcy risk based on bankruptcy prediction models—a case study of Poland. *Sustainability*, *14*(3), 1416. doi: 10.3390/su14031416.

Lehenchuk, S., Mostenska, T., Tarasiuk, H., Polishchuk, I., & Gorodysky, M. (2021). Financial statement fraud detection of Ukrainian corporations on the basis of Beneish model. In B. Alareeni, A. Hamdan, & I. Elgedawy (Eds.), *The Importance of New Technologies and Entrepreneurship in Business Development: In The Context of Economic Diversity in Developing Countries* (pp. 1341–1356). Cham: Springer. doi: 10.1007/978-3-030-69221-6_100.

Lesiak, M. (2024). Manipulacje danymi w sprawozdaniach finansowych spółek notowanych na GPW w Warszawie–ocena skuteczności modelu Beneisha. *Kwartalnik Nauk o Przedsiębiorstwie*, *72* (2), 67–78. doi: 10.33119/knop.2024.72.2.5.

Mączyńska, E., & Zawadzki, M. (2006). Dyskryminacyjne modele predykcji upadłości przedsiębiorstw. *Ekonomista*, *2*, 205–235.

Mantone, P. (2013). *Using analytics to detect possible fraud: Tools and techniques*. Hoboken: John Wiley & Sons.

Marais, A., Vermaak, C., & Shewell, P. (2023). Predicting financial statement manipulation in South Africa: A comparison of the Beneish and Dechow models. *Cogent Economics & Finance*, *11*(1), 2190215. doi: 10.1080/23322039.2023.2190215.

Nyakarimi, S. N., Kariuki, S. N., & Kariuki, P. W. (2020). Financial statements manipulations using Beneish model and probit regression model: A case of banking sector in Kenya. *European Online Journal of Natural and Social Sciences*, *9*(1), 253–264.

Omeir, A. K., Vasiliauskaitė, D., & Soleimanizadeh, E. (2023). Detection of financial statements fraud using Beneish and Dechow models. *Journal of Governance and Regulation*, *12*(3), 334–344. doi: 10.22495/jgrv12i3siart15.

Ozcan, A. (2018). The use of beneish model in forensic accounting: Evidence from Turkey. *Journal of Applied Economics and Business Research*, *8*(1), 57–67.

Paolone, F., & Magazzino, C. (2014). Earnings manipulation among the main industrial sectors. Evidence from Italy. *Economia Aziendale Online*, *5*(4), 253–261. doi: 10.4485/ea203-5498.005.0022.

Papik, M., & Papikova, L. (2020). Detection models for unintentional financial restatements. *Journal of Business Economics and Management*, *21*(1), 64–86. doi: 10.3846/jbem.2019.10179.

Pazarskis, M., Drogalas, G., & Baltzi, K. (2017). Detecting false financial statements: Evidence from Greece in the period of economic crisis. *Investment Management and Financial Innovations*, *14* (3), 102–112. doi: 10.21511/imfi.14(3).2017.10.

Platt, H. D., & Platt, M. B. (2002). Predicting corporate financial distress. Reflections on choice-based sample bias. *Journal of Economics and Finance*, *26*(2), 184–199. doi: 10.1007/BF02755985.

Putra, A. N., & Dinarjito, A. (2021). The effect of fraud pentagon and F-score model in detecting fraudulent financial reporting in Indonesia. *Jurnal Ilmiah Akuntansi dan Bisnis*, *16*(2), 247–263. doi: 10.24843/JIAB.2021.v16.i02.p05.

Roxas, M. (2011). Financial statement fraud detection using ratio and digital analysis. *Journal of Leadership, Accountability, and Ethics*, *8*(4), 56–66.

Saleh, M. M. A., Aladwan, M., Alsinglawi, O., Saleh, H., & Mahmoud, I. (2021). Predicting fraudulent financial statements using fraud detection models. *Academy of Strategic Management Journal*, *20*(3), 1–17.

Shakouri, M., Taherabadi, A., Ghanbari, M., & Jamshidinavid, B. (2021). Explaining the Beneish model and providing a comprehensive model of fraudulent financial reporting (FFR). *International Journal of Nonlinear Analysis and Applications*, *12*, 39–48. doi: 10.22075/ijnaa.2021.4793.

Shu, W., He, T., Li, X., & Gong, Y. (2022). Validity of fraud detection models on fraudulent US-listed Chinese companies. In E. Mallick, O. T. San, & B. Vilas (Eds.). *4th International Conference on Economic Management and Cultural Industry (ICEMCI 2022)* (pp. 1759–1781). Atlantis Press. doi: 10.2991/978-94-6463-098-5_199.

Svabova, L., Kramarova, K., Chutka, J., & Strakova, L. (2020). Detecting earnings manipulation and fraudulent financial reporting in Slovakia. *Oeconomia Copernicana*, *11*(3), 485–508. doi: 10.24136/oc.2020.020.

Sylwestrzak, M. (2022). Application of the beneish model on the Warsaw stock exchange. *Journal of Banking and Financial Economics*, *18*(2), 5–16. doi: 10.7172/2353-6845.jbfe.2022.2.1.

Wisdianti, D., Arum, E., & Wijaya, R. (2022). The analysis of fraud detection using the Beneish ratio index and the F score model method on the financial statements of the state-owned insurance companies registered in Indonesia. In A. F. Hindriana, K. R. K. Mahamud, S. Akhmaddhian, T. Supartono, N. Nugraha, & R. Rahim (Eds.). *Proceedings of the 2nd Universitas Kuningan International Conference on System, Engineering, and Technology*, Indonesia (pp. 575–586). European Alliance for Innovation. doi: 10.4108/eai.2-12-2021.2320344.

Wiszniowski, E. (2020). Accuracy of the Beneish model as applied to polish enterprises known to manipulate financial reports. In K. Soliman (Ed.), *Education Excellence and Innovation Management: A 2025 Vision to Sustain Economic Development during Global Challenges* (pp. 9218–9222). International Business Information Management Association (IBIMA).

Wuerges, A. F. E., & Borba, J. A. (2014). Accounting fraud: An estimation of detection probability. *Revista Brasileira De Gestão De Negócios*, *16*, 466–483. doi: 10.7819/rbgn.v16i52.1555.

Xu, H., Shi, W., Xia, J., & Liu, M. (2023). M-Score and F-score from the financial statement for company fraud prediction. *Advances in Economics, Management and Political Sciences*, *9*(1), 196–205. doi: 10.54254/2754-1169/19/20230138.

Yao, J., Pan, Y., Yang, S., Chen, Y., & Li, Y. (2019). Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: A multi-analytic approach. *Sustainability*, *11*(6), 1579. doi: 10.3390/su11061579.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, *22*, 59–82. doi: 10.2307/2490859.

Corresponding author
Marek Sylwestrzak can be contacted at: msylwestrzak@wne.uw.edu.pl