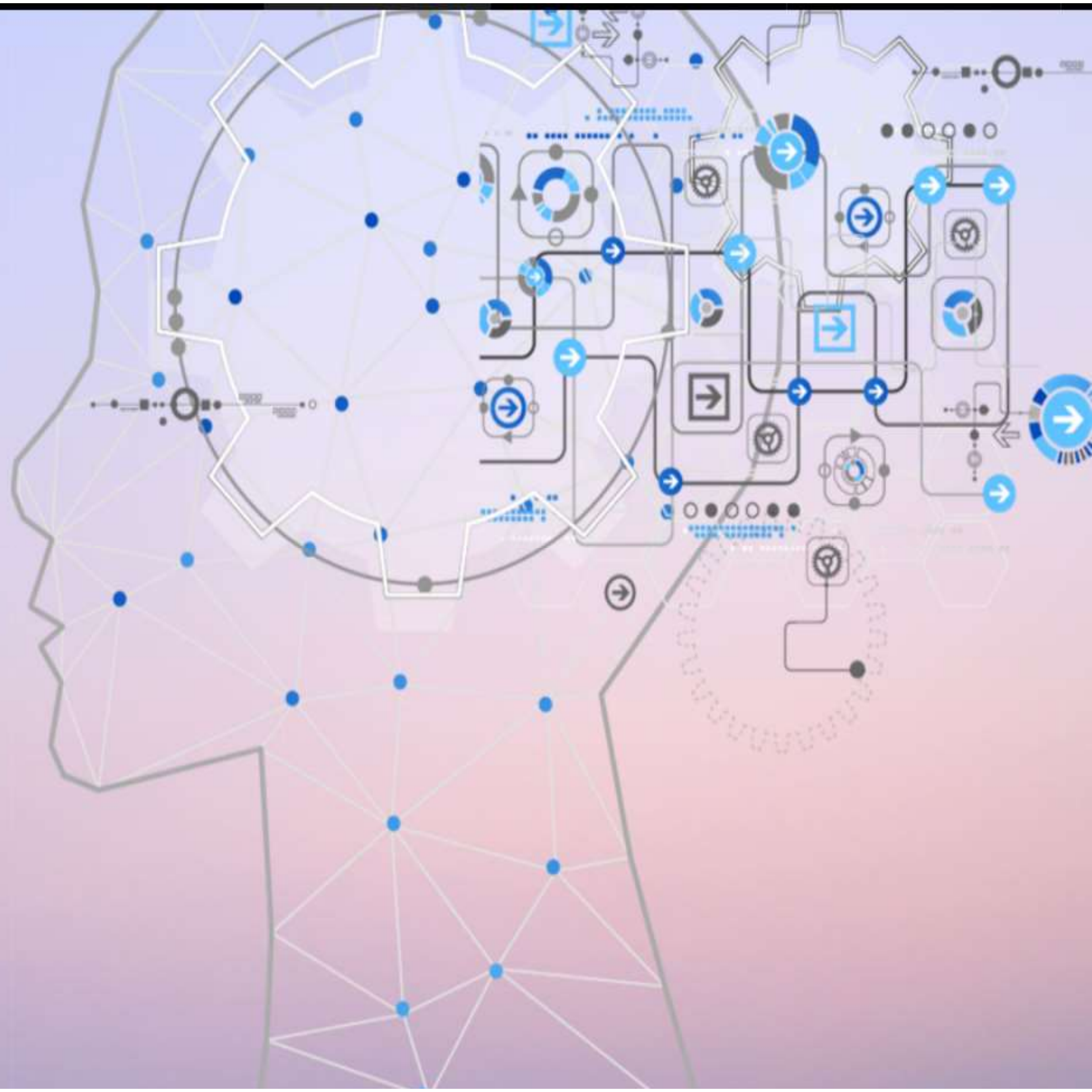
Fintech and Artificial Intelligence in Finance

FIN AI



The project is supported by

cost

EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

UNIVERSITY OF NAPLES FEDERICO II

DEPARTMENT OF POLITICAL SCIENCE
LABORATORY FOR STATISTICAL DATA ANALYSIS

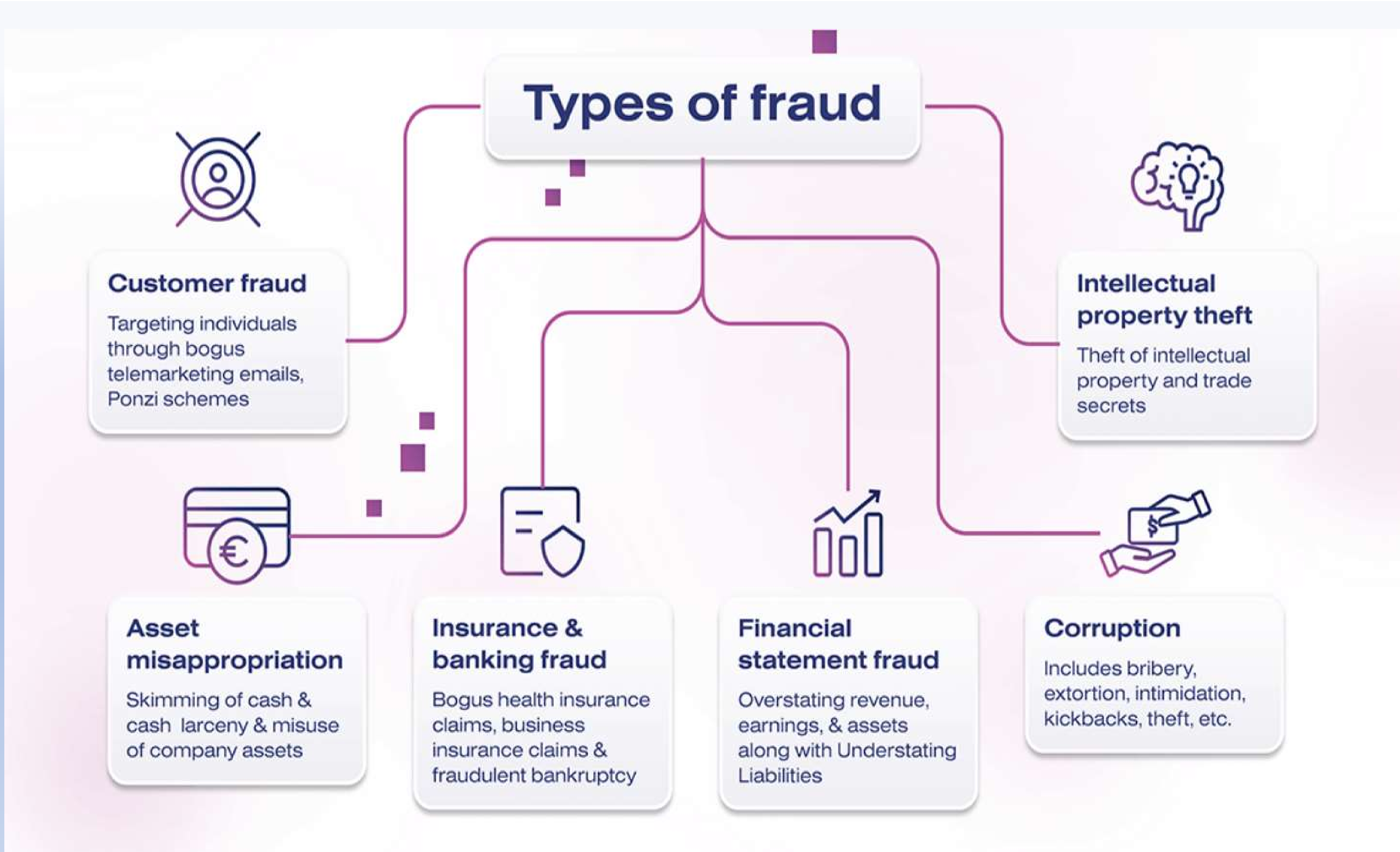
LabStat
Department of Political Science
University of Naples Federico II

The use of Financial Statements and Financial Ratios in fraud detection

Rezarta Shkurti Perri
University of Tirana, Albania

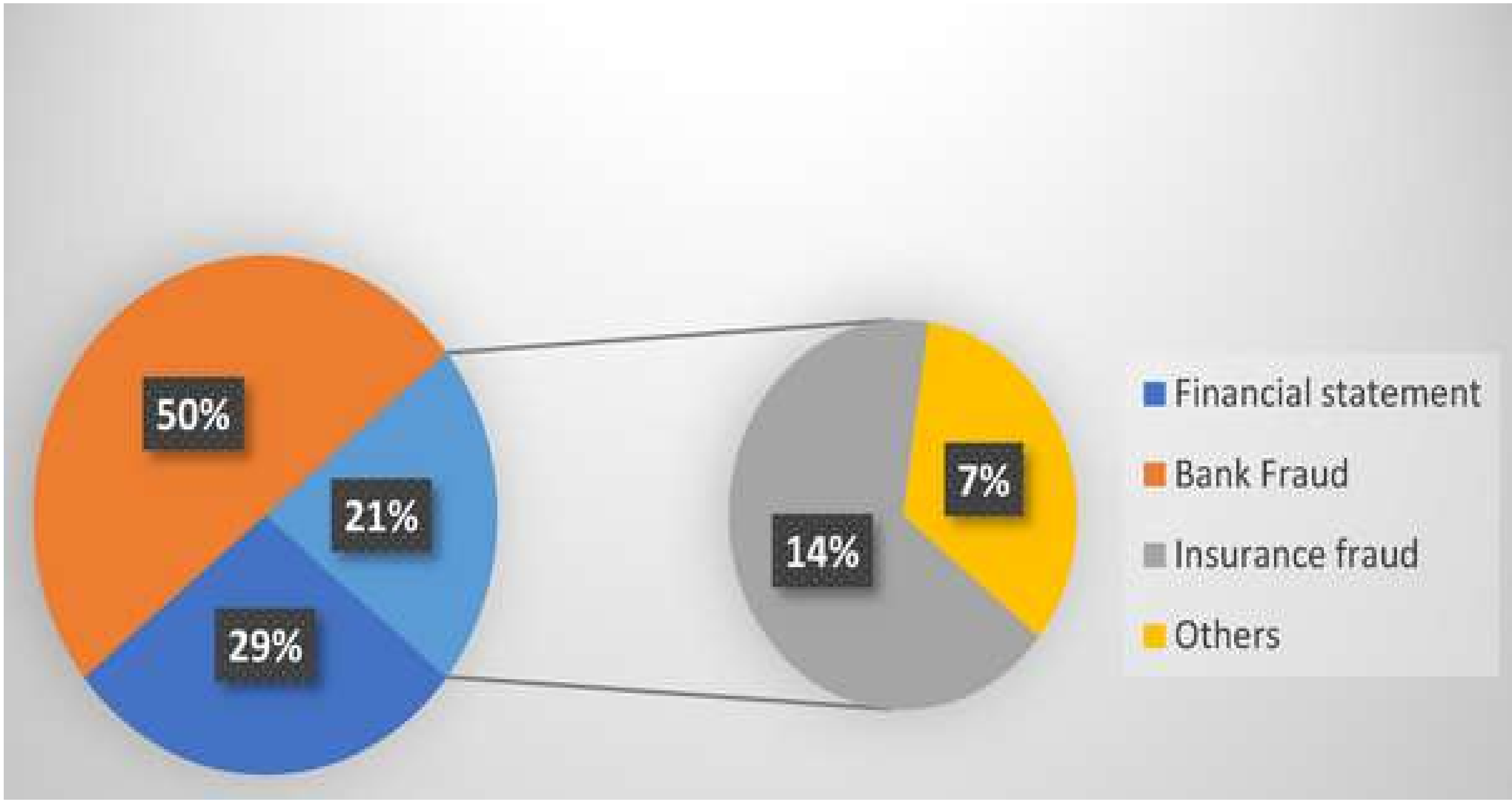
October 12 2023
12:30 CET
(Aula Pagano 2)

The use of Financial Statements information and Financial Ratios in fraud detection.


Types of fraud
Customer fraud
Targeting individuals through bogus telemarketing emails, Ponzi schemes
Intellectual property theft
Theft of intellectual property and trade secrets
Asset misappropriation
Skimming of cash & cash larceny & misuse of company assets
Insurance & banking fraud
Bogus health insurance claims, business insurance claims & fraudulent bankruptcy
Financial statement fraud
Overstating revenue, earnings, & assets along with Understating Liabilities
Corruption
Includes bribery, extortion, intimidation, kickbacks, theft, etc.

Fraud statistics

- Evaluation of financial statements fraud detection research: a multi-disciplinary analysis10.1057/s41310-019-00067-9 (Albizri, et.al, 2019)



Fraud Motivation

1. Financial difficulties and lower liquidity may be motivation for managers to engage in fraudulent activities.
2. Keep growing to meet targets.
3. Honor covenants, debt, and equity.
4. Achieve the threshold for bonuses.
5. Simply buy time until financial mistakes and losses can be properly corrected.

(Fanning & Cogger (1998), Kirkos et el. (2007), Ravisankar et al. (2011)

FIN AI

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Common Fraud Mechanisms

1. Inflated or deflated sales by inappropriately applying the revenue recognition standards;
2. Nonmatching sales with the appropriate cost of goods sold and this is reflected in ratios such as profit margins (on a gross, operating, or net basis);
3. Inventory reported at a lower cost than cost or market value hence impacting again the cost of goods sold but also the inventory to total assets ratio.

- **Whatever the mechanism, it is likely that it will show in the face of the financial statements**

Research context

Information from published financial statements and related financial ratios and indicators to detect phenomena as fraudulent financial statements, fraudulent transactions, and earnings management.

There is much interest in analyzing the published reports and statements with a specific focus on identifying the fraud cases, preferably before fraud jeopardizes the entities' activities.

But is that possible?

If Yes, How?

FIN AI

Fintech and Artificial Intelligence in Finance

cost

EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Research focus:
Tax Fraud as a subset of Financial Statements Fraud

Tax Fraud and Tax Audits

- Easier to obtain information if focused on tax fraud because of the Non-subjective Fraud/Nonfraud classification.
- **Tax Fraud** – Fraud with the primary intention of avoiding taxes.
- To mitigate the overall level of the tax risk, the tax auditors, engage in **tax audits**, whose focus is the examination of the tax returns to verify that financial information is being correctly reported.
- **Tax Audits** are usually time-consuming as they imply on-site controls and inspections for the audited companies.
- Tax auditors need to select entities for tax auditing.
- Hence the research problem of correctly detecting probable tax offenders among entities.

Tax Audit Process – How does it work?

Two stages of selecting companies that will be tax audited.

1. The Risk Module of the Tax Administration Information System selects up to 70 percent of entities that will be audited, based on indicators and red flags derived from its internal Business Intelligence algorithms.

2. Manual selection by Tax auditors of no more than 30% based on a "Risk Indicators Manual".

- Manual selection tends to be subjective.
- The list of the "Risk Indicators" that guide the manual selection process may be derived from reliable research thus avoiding subjectivism in selecting entities that will be tax audited.

Literature review

- Many studies have chosen the analysis of ratios as one of the methods to determine fraud (Feroz et al., 1991; Stice et al., 1991; Persons, 1995; Wells, 1997; Fanning & Cogger, 1998; Beneish, 1999; Spathis et al., 2002; Lenard & Alam, 2009; Ravisankar et al., 2011)
- Different scholars choose different financial ratios for fraud investigation and different research reveals different significant indicators.
- Explore **techniques** to detect manipulation and incentives to engage in earnings management in both dimensions, inflating earnings to achieve the maximization of shareholders' value and deflating earnings to minimize the taxable income.
- Find **variables** that indicate the probability of fraudulent activities.

Literature review for financial / tax fraud

Data:

- Annual reports
- Stock market information
- Financial statements
- Tax returns
- Directors and management disclosures and announcements
- Analysts forecasts.

Methods:

1. Traditional

a. Discriminant analysis

b. Logistic regression

c. multi-criteria decision aid (MCDA) technique

2. Modern

a. Neural networks

b. Machine learning

i. Supervised and unsupervised

Other tools:

1. Benford Law

2. Beneish manipulation index

3. The Dechow F Score Model.

FIN AI

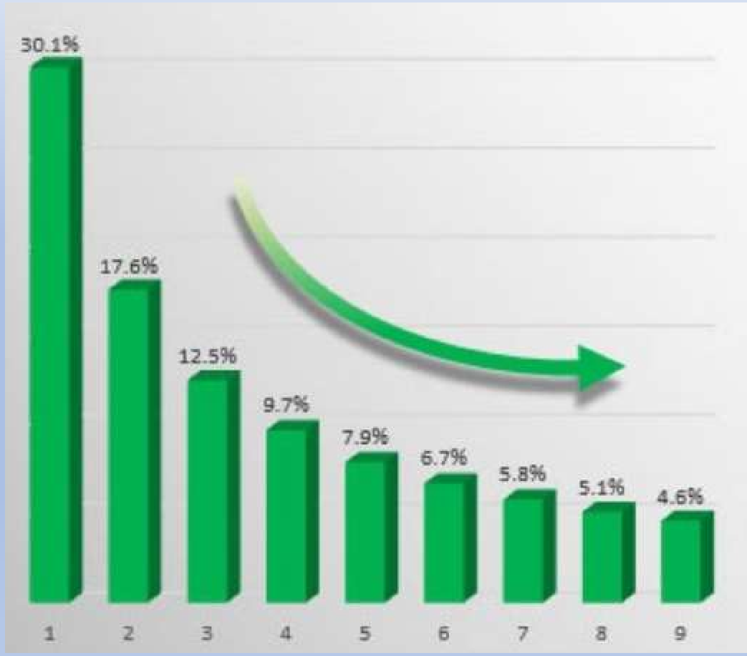
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Literature Review - Methods

- Use of computer-assisted audit techniques (CAATs) to investigate fraud and accounting irregularities. (Bay S, Kumaraswamy K, Markus G., Kumar R, Steier D M (2006) 'Large Scale Detection of Irregularities in Accounting Data' Center for Advanced Research, PricewaterhouseCoopers)
- The study proposes an expert system for identifying suspicious irregularities in detailed financial data and techniques for automatic analysis of company ledgers on a large scale, identifying irregularities. The system was named *Sherlock.*
- They showed that if possessing huge data sets with detailed financial information it is possible to use advanced technological tools to build an expert system that assists auditors in detecting irregularities and fraud.
- Advanced tools in detecting the entities with the probable higher tax audit risk, were also used in another study (Kallio M, Back B (2011) 'The Self-Organizing Map in Selecting Companies for Tax Audit' chapter from book IS Success Evaluation: Theory and Practice (pp.347-358). An unsupervised neural network method – the Self-Organizing Map (SOM) – to select entities for a tax audit.

Benford Law and its use in fraud and forensic

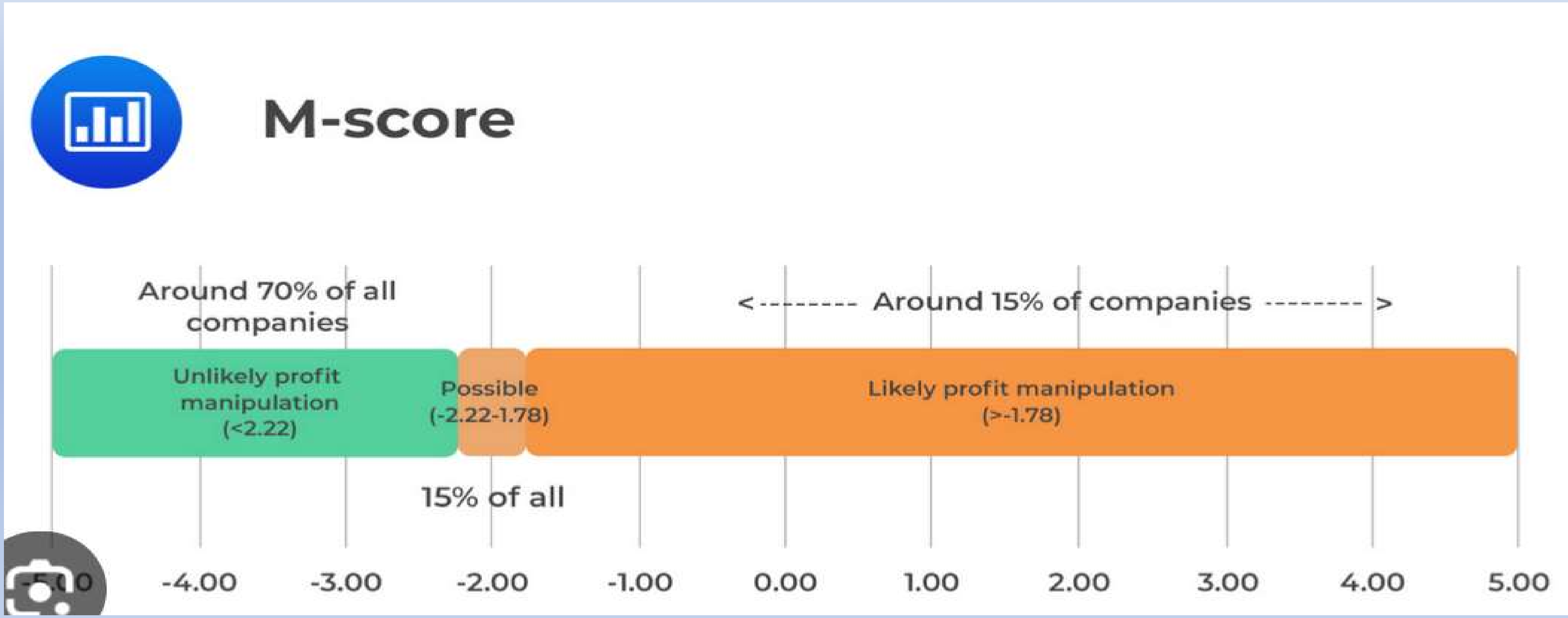
- Benford's Law is a rule that describes the distribution of first (leading) digits in economic or accounting data, but not only.
- *Briefly explained, Benford's Law maintains that the numeral 1 will be the leading digit in a genuine data set of numbers **30.1%** of the time; the numeral 2 will be the leading digit **17.6%** of the time; and each subsequent numeral, 3 through 9, will be the leading digit with decreasing frequency. This expected occurrence of leading digits represent what is known as the "Benford's Curve.*
- Several studies report that manipulated financial statements data deviate significantly from Benford's Law rule of numbers and therefore this law can be used as a tool for fraud detection.



Beneish Model for earnings manipulation and management - Beneish's M-Score

- Beneish's M-Score is a mathematical model that uses *eight financial ratios* weighted by coefficients to identify whether a company is likely to have manipulated its profits.
- Beneish surmises that companies are incentivized to manipulate profits if they have high sales growth, deteriorating gross margins, rising operating expenses, and rising leverage.
- They are also likely to manipulate profits by accelerating sales recognition, increasing cost deferrals, raising accruals, and reducing depreciation.
- *Application of Beneish M-Score Models and Data Mining to Detect Financial Fraud, Tarjoa, Nurul Herawatib, Elsevier Publishing, 2015.*

M Score classification chart



Modern Methods

- Machine learning and artificial intelligence algorithms (neural networks) perform better than linear regression or logistic regression when they are applied to the same data.
Wyrobek, 2020. "Application of machine learning models and artificial intelligence to analyze annual financial statements to identify companies with unfair corporate culture."

Purpose of research

- The purpose of this research is to reveal whether the information contained in the financial ratios and indicators could indicate tax level risk for entities, thus helping tax auditors during the selection procedures.
- To find if there is any link between financial information and tax risk, we focus on the information embedded in the financial ratios that can be computed from reported financial statements and the level of tax evasion that is subsequently discovered by tax auditors after they have conducted the tax audit.

FIN AI

COST EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Data Sample

- This study was performed by the beginning of 2023.
- Reports on tax audits performed during 2020 and 2021 and published in 2022.
- The audited financial statements of 2018 and 2019 for all regional tax directories of Albania. (The tax audits performed in a certain year cover two previous financial years)
- Confidential information. The data was made available upon a Non-Disclosure Agreement

Sample preparation in 3 steps.

<u>1402</u> audited economic entities audited in total for the concerning years.

1. A sample consisting of 15 percent of the total population, which is about **<u>210</u>** entities was randomly selected. *(The random selection option of MS Professional Excel was used as the list of audited entities was received in an Excel file format*).
2. After obtaining the sample, and a list of selected entities, we accessed (again under a confidential privileged access rights clause) the Information System of Tax Administration and downloaded the full financial statements and tax returns for the audited periods for all the companies in the sample.
3. Discard from the sample:
    a. microenterprises (with less than 5 employees)
    b. missing any piece of financial information

We were left with **<u>175</u>** and **<u>183</u>** entities with: a) complete financial statements for 2018 and 2019 and b) The Tax Audit Report which reports the **<u>level of tax evasion</u>**, if any found after the tax audit is performed on these entities.

Sample Administration

Various classifications of data:

- Entities grouped by type of activity: *construction & manufacturing*, *service & transport*, and *commerce*, and coded them by category with 1, 2, and 3 accordingly.
- By size as either the total assets, total revenues, or the number of employees. The size of the firm will later serve as a controlling variable during data analysis.
- Classification of the sample into two different categories based on the level of magnitude of tax evasion which will be *the dependent variable*. If the tax evasion was from 0 – 50.000 Albanian Lek (approximately 500 Euro) the entity is virtually risk-free and is coded with a 1; whereas if the level of the magnitude of tax evasion was found to be higher than 50.001 Albanian Lek (500 Euro), that entity is coded with a 2.

Variables included in the study.

- The dependent variable of the study is the magnitude of tax risk of tax-audited entities.
- From the literature review, several common financial ratios were consistently reported to successfully detect fraud and signaling manipulation – the independent variables.
  - Sales,
  - Accounts Receivables and liquidity ratios
  - Inventory
- Debt and profitability ratios
- Total 18 independent variables

No.	Category	Financial Ratio	Optimal Standard Value	Source of standardization of optimal vale	Codification
1	Liquidity Ratios	Current Ratio	150 %	Theoretical literature	0 if less than 150 1 if more than 15
2		Acid test (quick) Ratio	100 %	Theoretical literature	0 if less than 100 1 if more than 10
3		Liquid Ratio	50 %	Theoretical literature	0 if less than 50% 1 if more than 50
4	Debt Ratios	Non-Current Liabilities to Total Assets Ratio	30 %	Theoretical literature	0 if more than 30 1 if less than 30%
5		Total Debt Ratio	50 %	Theoretical literature	0 if more than 50 1 if less than 50%
6		Times interest earned	5 times to 1	Theoretical literature	0 if less than 5 1 if more than 5
7	Profitability Ratios	Net Profit Margin	Varies according to industries	INSTAT data about average profit margin according to each industry	0 if less than the a of industry 1 if more than a of industry
8		Profit before Taxes (EBT) Margin	Varies according to industries	INSTAT data about average profit margin according to each industry	0 if less than the a of industry 1 if more than a of industry
9		Return on Assets ROA	7 % adjusted with inflation (1.6% in 2018 and 2.6% in 2019)	Theoretical literature	0 if less than 7% adjustments 1 if more than 7% adjustments
10		Return on Equity ROE	Interest rate percentage (2.25% in 2018 and 2% in 2019)	Theoretical literature	0 if less than the i rate 1 if more than i rate

Other variables

No.	*Other financial ratios*
1	NCA Ratios = Fixed Assets to Total Assets
2	Inventory Ratio = Inventory to Total Assets
3	(Inventory + Accounts Receivable) / Total Assets
4	Total Personnel Expenses to Total Revenues
5	Operating Profit Margin
6	Administrative Expenses to Total Revenues
7	Difference between taxable income and reported income to total Revenues
8	Cash Flow from Operating Activities to Net Income

- These 8 indicators are not included in the univariate analysis as per lack of optimal or standard values in literature and/or because no comparable data in the official websites of INSTAT, etc.
- 18 ratios were calculated for 183 entities for 2 consecutive financial years.

Data analysis – traditional methods

1. Univariate analysis
    1. Refers to the average optimal values or range of indicators, hence, only 10 variables could be analyzed.
2. Binary logistic regression
    1. Whole data set
    2. For specific industry sectors individually

1 - Univariate analysis

No.	Category	Financial Ratio	2018 Significance	2019 Significance
1	Liquidity Ratios	Current Ratio	Not important (No statistically significant difference)	Not important (No statistically significant difference)
2		Acid test (quick) Ratio	Not important	Not important
3		**Liquid Ratio ***	Important Ratio (Statistically significant difference)	Important Ratio (Statistically significant difference)
4	Debt Ratios	Non-Current Liabilities to Total Assets Ratio	Not important	Not important
5		**Total Debt Ratio ***	Important Ratio (Statistically significant difference)	Important Ratio (Statistically significant difference)
6		Times's interest earned	Not important	Not important
7	Profitability Ratios	Net Profit Margin	Not important	Not important
8		Profit before Taxes (EBT) Margin	Not important	Not important
9		Return on Assets ROA	Not important	Not important
10		**Return on Equity ROE ***	Important Ratio (Statistically significant difference)	Important Ratio (Statistically significant difference)

2 - Binary logistic regression model

- *Variables with a correlation above 40% we eliminated from the model to avoid collinearity. Next, we excluded the outliers.*
- Model I regression based on the sample with all the entities, run separately for both years.

	2018				2019			
	B	S.E.	Wald	Exp(β)	B	S.E.	Wald	Exp(β)
Total Debt Ratio	-1.314***	.489	7.215	.269	.676	.535	1.598	1.965
NCA Ratio	1.340	.958	1.957	3.820	-2.126**	.846	6.314	.119
Inventory Ratio	.880	.744	1.401	2.412	-1.548*	.823	3.538	.213
Net profit margin	-1.548	1.473	1.104	.213	-1.530	1.283	1.422	.217
Revenue	.011***	.004	7.148	1.011	.000*	.000	3.707	1.000
Constant	1.500***	.490	9.362	4.484	2.151***	.647	11.040	8.596
N	175				183			

Regression Model I - 2018

- The regression model for both years, Model I's explaining variables are the Total debt ratio, Noncurrent liabilities to total assets (NCA) ratio, Inventory ratio, Net profit margin, and Total Revenues (as a controlling variable for size).
- For 2018, Model I is statistically significant, and Pseudo R^2 is 0.152.
- The percentage of correctly classifying the entities with high risk was 99.3%,
- The percentage of correctly classifying all entities with high and low risk was 84.6 %.
- In 2018, the Total Debt Ratio (p-value<0.01) and the Revenues (p-value<0.01) are statistically significant which confirms the importance of the total revenue level and the total debt level in determining the entities that have a higher probability of having higher tax risk.

FIN AI

cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Regression Model I - 2019

- For 2019 Model I is statistically significant (p-value<0.01), and Pseudo R^2 is 0.174.
- The percentage of correctly classifying all entities with high and low risk is 85.6%.
- Statistically significant variables for 2019 are the NCA ratio, Inventory ratios, and Revenues (whereas in the previous year, the NCA ratio and the Inventory ratio were not statistically significant).
- The 2019 Revenues are again, as in 2018, an important explaining variable (p-value<0.1).

An unusual finding

- The **net profit margin ratio**, is a statistically insignificant ratio.
- This may be very well explained by the formula used to calculate this ratio. We believe that this ratio, as calculated based on the reported income and not on taxable income is hiding the true "maneuvers" that companies are usually engaged in, to perform tax evasion. (*From the personal experience of the authors, we know the profit margin to be one of the first ratios that a tax auditor checks*). This fact is well known to everyone, including businesses who subsequently try to artificially adjust the taxable income and bring it to the "right amount".
- As the usual mechanism for this earnings management practice, companies may regard a portion of their expenses as non-tax deductible, thereby increasing taxable income and improving the value of this ratio, so that the tax auditor will not detect them.

Regression analysis for separate industry sectors

- We run logistic regression separately for the entities in the Construction & Manufacturing sector (**Model II),** entities in the Service & Transportation Sector **(Model III)** and entities in the Merchandising sector **(Model IV).**
- Model II Construction & Manufacturing Sector analyses 51 entities/observation for each year.
- The Model II for 2018 (the Construction & Manufacturing sector) was statistically significant ($p<0.1$), with a pseudo R^2 of 0.325. The total percentage of correctly classifying all entities with high and low risk is 88.2%. For 2018 the only statistically significant variable is the Inventory Ratio. We see that the Revenue variable has the same impact in this model as with Model I, but in this case, it is not statistically significant.

Model II

	2018				2019			
	B	S.E.	Wald	Exp(β)	B	S.E.	Wald	Exp(β)
Total Debt Ratio	.556	1.507	.136	1.744	1.132	1.600	.501	3.102
NCA Ratio	2.254	1.809	1.552	9.526	-2.406	1.699	2.007	.090
Inventory Ratio	5.611*	3.224	3.030	273.465	-4.104*	2.273	3.260	.017
Net profit margin	-1.217	2.714	.201	.296	-.003	2.149	.000	.997
Revenues	.000	.000	2.019	1.000	.000	.000	.730	1.000
Constant	-.581	1.372	.179	.559	2.556*	1.411	3.281	12.889

Model III

- Regarding sector 2, Services and Transport, Model III proved to be statistically not significant.
- The number of observations (entities) in this sector was respectively 31 in 2018 and 24 in 2019.

Model IV

- In the Merchandising sector there are 101 entities for the year 2018, Model IV is significant (p-value<0.01), and Pseudo R^2 is 0.295. The total percentage of correctly classifying all entities with high and low risk in the Merchandising sector is 84.2%.
- The Total Debt ratio (p-value <0.01) and the Revenues (p-value<0.05) are both statistically significant.
- Other ratios, the NCA ratio, Inventory Ratio, and the Net profit margin are not statistically significant.
- In the year 2019, Model IV for the Merchandising Sector included 108 entities and was statistically significant (p-value<0.05). The total percentage of correctly classifying all the entities both with high and low risk in this year was 85.2%. The ratios that we find statistically significant for the year 2019 are the NCA Ratio (p-value<0.05) and the Revenues (p-value<0.1).

Model IV

	2018				2019			
	B	S.E.	Wald	Exp(β)	B	S.E.	Wald	Exp(β)
Total Debt Ratio	-3.199**	1.150	7.740	.041	.517	.586	.780	1.678
NCA Ratio	1.332	1.391	.917	3.787	-2.486**	1.256	3.917	.083
Inventory Ratio	-.225	1.033	.047	.798	-1.055	1.057	.996	.348
Net profit margin	-.879	1.941	.205	.415	-5.406	4.635	1.361	.004
Revenues	.000**	.000	4.123	1.000	.000*	.000	3.067	1.000
Constant	2.976**	1.042	8.155	19.607	2.112**	.881	5.741	8.263

Main Findings:

- The dependent variables (financial ratios), behave differently in different sectors and in companies with different size (the Revenues is the control variable).
- In Construction and Manufacturing, inventory was a significant ratio which means that the higher the inventory levels of entities in this sector, the higher the tax fraud risk of that entity.
- Revenues and other variables are not significant while revenues were significant for the total set of firms (Model I).
- In the Merchandising sector (Model IV), the total debt ratio, the NCA Ratio, and the Revenues are statistically significant, (total compatibility with the results from Model I).
- Conclusion: for the Merchandising sector, specific risk indicators are high levels of total debt and high levels of investments in Fixed Assets.

Conclusions

- The independent variables are eighteen financial ratios and indexes, mainly from the reported financial statements and their filed tax returns. Eighteen financial ratios were calculated for each entity for two consecutive years, 2018 and 2019, and the respective Tax Audit Reports were analyzed to define the magnitude of the tax evasion.
- 1402 entities total population; sample = 183 entities.
- Both univariate and multivariate techniques are used to analyze the dataset.
- The univariate technique, reveals that:
  - out of the three liquidity ratios (current, acid test, and liquid ratio), only the liquid ratio holds statistical significance. Entities with a high liquid ratio were found to have a lower tax risk.
  - Among the solvency ratios, only the total debt ratio was found to be significant. Entities with a high debt ratio were found to have a higher tax audit risk.
  - In the profitability category, only the ROE ratio was important. We then proceeded to apply the univariate analysis to each sector separately, but the results were mixed, and no clear trend could be identified.

Conclusions

- Next, the multivariate analysis is applied.
- The results of Model I, showed that the Total debt ratio, Total Revenues, Non-current assets ratio, and Inventory Ratio were all statistically significant for both years and for all companies.
- It was unexpected to note that the net profit margin and liquidity ratios were not statistically significant in Model I, which may be attributed to companies using accounting techniques to avoid being caught by tax auditors.

Conclusions

- Model II (Construction and Manufacturing) revealed for 2018 the Inventory Ratio as the only statistically significant variable.
- Model III, (Services and Transportation Sector) yielded not statistically significant results for both years (probably due to fewer entities in this sector).
- Model IV (Merchandise) revealed as statistically significant ratios the total debt ratio, the Revenues, and the Non-Current Assets Ratio.
- Overall, based on results from both univariate and multivariate analysis we can conclude that the most common variables in these models that are linked to tax risk are the:
  - (1) liquid ratio, (2) ROE, (3) total debt ratio, (4) inventory ratio and (5) Non-Current assets ratio.
- Contrary to the widely spread beliefs that profit margin ratios do contain information on tax risk, we found no relation between these margins and the tax evasion magnitude probably due to many ways (and potentially manipulative as well) that this ratio is calculated and the low quality of data where it is based upon.

Contributions and limitations

- This research contributes by suggesting specific indicators that could be easily calculated based on reported published financial information and that could be helpful in the daily procedures of tax auditors minimizing their subjectivity in selecting entities that will be audited.
- A new, unique database with financial information, financial ratios, and indicators as well as audit report findings was prepared.
- Limitations:
  - limited time horizon, and application techniques.

Future perspectives

- Expand the scope of fraud investigation beyond mere tax fraud.
- Innovate methodology.
  - We intend to use Machine Learning open-source tools as a means to differentiate more accurately between fraud and non-fraud statements.
- An **unsupervised learning model** looks like the most suitable method given the scarcity of data due to the sensitivity of the matter and also nondisclosure of many fraud cases.
- It differs from a supervised learning model where all input information has to be labeled as good or bad.
- The unsupervised learning model is meant to detect anomalous behavior in cases where there is little transaction data or such data is not available at all. An unsupervised learning model continuously processes and analyzes new data and updates its models based on the findings. It learns to notice patterns and decide whether they're parts of legitimate or fraudulent operations. Deep learning in fraud detection is usually associated with unsupervised learning algorithms.

Thank you for your attention!

Questions and comments are welcome!