INDCOMP: A Shiny App for Open Data Repository of the Performance of an Indonesian Company Listed at the Indonesia Stock Exchange

Prana Ugiana Gio
Department of Mathematics, Universitas Sumatera Utara, Indonesia
prana@usu.ac.id (corresponding author)

Herman Mawengkang
Department of Mathematics, Universitas Sumatera Utara, Indonesia
hmawengkang@yahoo.co.id

Muhammad Zarlis
Computer Science Department, Bina Nusantara University, Indonesia
muhammad.zarlis@binus.edu

Saib Suwilo
Department of Mathematics, Universitas Sumatera Utara, Indonesia
saib@usu.ac.id

Received: 13 June 2024 | Revised: 3 July 2024 | Accepted: 5 July 2024

 DOI: https://doi.org/10.48084/etasr.8131

ABSTRACT

Investors, practitioners, and stock researchers highly need data related to financial performance to predict a company's financial health condition, which is used as a basis to consider investing in it. The Indonesia Stock Exchange (IDX) website provides reports on the company's financial performance. Unfortunately, the company's financial data found on the IDX website are in PDF format, and researchers must download them one by one, which takes a long time. This study presents a website-based application, named Indonesia Company Performance (INDCOMP), built using the R programming language and involving various R packages and frameworks to assist investors, practitioners, and stock researchers in studying the financial performance of companies. This application can help users quickly access the financial performance data of various companies, present financial performance data in data tables, and perform data visualizations as well as statistical analyses.

Keywords-financial; R shiny; INDCOMP; computational; statistics

I. INTRODUCTION

The number of investors in Indonesia at the end of the second semester of 2022 increased by 37.68% year-on-year (yoy) to 10,311,152 [1]. Meanwhile, the value of share ownership at the same time increased by 18.77% to Rp. 5,921.60 trillion. Local ownership increased by Rp. 522.53 trillion (19.36%) yoy. Foreign ownership increased by Rp. 413.34 trillion (18.08%) yoy. Based on investor type, in the second semester of 2022, local and foreign share ownership was dominated by corporate investors amounting to Rp. 1,503.77 trillion (46.67%) and Rp. 689.58 trillion (25.55%) yoy [2]. Information on the financial performance of a company is very important for investors to decide to invest in it [3]. By analyzing the company's financial performance, investors can determine its financial health [4]. The Indonesia Stock Exchange (IDX) website hosts financial performance reports of listed companies. In addition to sector-based classification, a company can be categorized by board level, namely main, development, acceleration, special monitoring, and new economy. Researchers who need financial performance data from various companies should download them one at a time from the IDX website, then open them, and finally import them into an Excel spreadsheet. This process takes a long time if the number of companies being studied is large, and importing financial data into Excel is also a time-consuming task [5].

In [6,7] financial data from Nikkei NEEDS were analyzed using a database system, including records of 1,500 companies listed in the first sector of the Tokyo Stock Market. This approach deployed a double-log model with a normal error distribution, which is an exploratory data analysis method. Data visualization informed the selection of total assets and the number of employees as explanatory factors to describe the relationship between sales data and other variables. This study also utilized massive amounts of financial data from companies around the world. The findings pointed to a skew-t error distribution in a double-log model as the best fit. The model took into account the companies' overall sales volume in relation to their staff count and asset value. Some double-log models with skew-symmetric distributions were evaluated following a cross-validation approach, and the findings were derived by comparing their Akaike Information Criterion [8, 9].

Shiny is an R package that can be used to build interactive web-based applications [10] and has been employed by many applications. ShinyItemAnalysis provides various psychometric methods, sample data, model equations, parameter estimation, and interpretation of results [11]. EpiExploreR can be utilized to analyze spatiotemporal data on animal diseases [12]. SAFA-HsG can be implemented for nucleotide frequency analysis [13]. In [14], a Shiny application was presented to analyze and visualize the thesis and dissertation data from the Brazilian Digital Library. In [15], CIPR was based on Shiny to assess gene expression profiles from unknown cell clusters against mouse or human data or data provided by users. In [16], a Shiny application, called MEPHAS, was presented to perform statistical data analysis and predictions for clinicians and pharmacy researchers. In [17], a Shiny application was introduced for learning modules related to the prediction of ecological data. In [18], a Shiny application was proposed to calculate the number of samples in the case of a longitudinal study. In [19], a Shiny application, called VisProDom, was presented to help users visualize protein domains and transcript structures simultaneously using either built-in or uploaded genome data.

Due to the limitations of the IDX website, namely the inefficiency in the financial performance data collection process, and the lack of a public repository, this study aims to build a web-based application named INDCOMP (The Indonesia Company Performance). This application aims to provide various financial performance data for companies, such as Return On Assets (ROA), Return On Equity (ROE), Net Profit Margin (NPM), Debt-to-Asset Ratio (DAR), and Debt-to-Equity Ratio (DER). Investors and researchers can deploy these data to analyze the health of a company's financial performance or investment. INDCOMP users can also perform several statistical analyses directly utilizing the provided data.

II. INDCOMP R PACKAGE

The R programming language and some of its packages were put into service to build INDCOMP in the RStudio environment. The R programming language is both free and open-source. Linear and nonlinear models, time series analysis, clustering, visualization, classification, and classical statistical tests are only a few of the many statistical methods it offers [20, 21]. Table I shows the R packages employed in the development of INDCOMP.

TABLE I. R-PACKAGES FOR INDCOMP

Features	R package	Description
Web-based application	shiny	Web application framework for R [10].
	shinyAce	Ace editor bindings to enable a rich text editing environment within Shiny [22].
	ggthemes	Extra themes, scales, and geometries for ggplot2 [23].
Data manipulation	plyr	Tools for splitting, applying, and combining data [24].
	dplyr	A fast, consistent tool for working with data frames both in memory and out of memory [24].
Spreadsheet	DT	Data objects in R can be rendered as HTML tables using the DataTables JavaScript library (typically via R markdown or Shiny) [25].
	rhandsontable	An R interface to the handsontable JavaScript library, which is a minimalist Excel-like data grid editor [26].
Visualization	ggplot2	Create elegant data visualizations using the grammar of graphics [27].
	plotly	Create interactive web graphics via plotly.js [28].
	corrplot	Visualization of a correlation matrix.
	Ggally	Some functions include a pairwise plot matrix, a two-group pairwise plot matrix, a parallel coordinates plot, a survival plot, and several functions to plot networks.
Statistical analysis	GLDEX	Fitting single and mixture of generalized lambda distributions [29].
Print output	openxlsx	Read, write, and edit xlsx files [30].

The development of INDCOMP was based on six primary components. Interactive and visually appealing user interfaces can be built implementing packages, such as Shiny, ShinyAce, and ggthemes, which are part of the web-based application's architecture. The plyr and dplyr packages are used for data organizing and manipulation, offering powerful methods to effectively transform and summarize data. The DT and rhandsontable packages support INDCOMP's spreadsheet functions, enabling dynamic and editable table presentations. INDCOMP relies on ggplot2, plotly, corrplot, and Ggally to facilitate visualization. Data analysis is made more thorough and informative with the help of these tools, which allow the construction of various static and interactive displays. The GLDEX package is utilized for data modeling, which employs the Generalized Lambda Distribution (GLD) technique to provide accurate and flexible data modeling. The openxlsx package is employed to export the results to Excel files, making the data and findings easier to share and access.

III. GENERALIZED LAMBDA DISTRIBUTION

Aligning a probability distribution with empirical data is a crucial component of every data analysis system [31]. An important obstacle in statistical data analysis is determining a statistical probability distribution that matches the actual data accurately [32]. Karl Pearson initiated fundamental research on the process of adjusting statistical distributions to data [33]. The field of fitting statistical distributions to data has undergone significant evolution, with applications extending to numerous disciplines, method developments, and improvements in computer approaches. Statistical modeling is widely used in various sectors, including science, technology, medicine, engineering, and other areas of human knowledge [34-46]. The primary and widely deployed model is the probability distribution, which establishes a connection between the values of essential variables and the probability of their occurrence. A continuous probability distribution refers to the probability distribution of an observed variable that can assume any value within a certain interval [33, 47].

GLD is well-known for its exceptional flexibility, as it can accurately represent distributions with different shapes [48, 49]. Due to recent progress, GLD has become widely applicable in almost all cases [34] and has been effectively utilized in several fields, such as construction, atmospheric data analysis, quality control, and medicine [49]. GLD is a flexible probability distribution that can approach various probability distributions, such as normal, Student's t, Weibull distribution, and many others [50]. Authors in [51] introduced GLD with four parameters, which is defined using the quantile function as follows [33, 51]:

$$Q(y) = Q(y;\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \lambda_1 + \frac{y^{\lambda_3}-(1-y)^{\lambda_4}}{\lambda_2} \quad (1)$$

Note that the y values are in the interval $0 \le y \le 1$. The parameters λ_1 and λ_2 represent the location and scale parameters, respectively, while λ_3 and λ_4 represent the skewness and kurtosis of $GLD(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$. GLD has been successfully used in various disciplines. GLD has also been employed to explain the distribution of income in a population [52], financial data [48, 49], and stocks [53]. The following is a method of moments approach for GLD. Consider data $x_1, x_2, \dots, x_n$ and the i^{th} moment α_i defined as [54]:

$$\hat{\alpha}_1 = \frac{\sum_{i=1}^{n} x_i}{n},$$

$$\hat{\alpha}_2 = \frac{\sum_{i=1}^{n}(x_i-\hat{\alpha}_1)^2}{n},$$

$$\hat{\alpha}_3 = \frac{\sum_{i=1}^{n}(x_i-\hat{\alpha}_1)^3}{n(\hat{\alpha}_2)^{1.5}},$$

$$\hat{\alpha}_4 = \frac{\sum_{i=1}^{n}(x_i-\hat{\alpha}_1)^4}{n(\hat{\alpha}_2)^2} \quad (2)$$

Putting $a = \frac{1}{\lambda_2}$ and $b = \lambda_1 - \frac{1}{\lambda_1\lambda_2} + \frac{1}{\lambda_2\lambda_4}$ with $Y = \frac{(X-b)}{a}$, and using $E(X^k) = \int_0^1 \left(Q^{-1}(u)\right)^k$ and binomial expansion gives:

$$s_k = E(Y^k),$$

$$s_k = \int_0^1 \left(\frac{u^{\lambda_3}}{\lambda_3} - \frac{(1-u)^{\lambda_4}}{\lambda_4}\right) du,$$

$$s_k = \int_0^1 \sum_{j=0}^{k} \binom{k}{j} (-1)^j \left(\frac{u^{\lambda_3(k-j)}}{\lambda_3^{k-j}} - \frac{(1-u)^{\lambda_4 j}}{\lambda_4^{j}}\right) du,$$

$$s_k = \sum_{j=0}^{k} \binom{k}{j} \frac{(-1)^j}{\lambda_3^{k-j}\lambda_4^{j}}\, \beta(\lambda_3(k-j)+1, \lambda_4 j+1) \quad (3)$$

where $\beta(*)$ denotes the β function. Note that both arguments must be positive values, which means that $\min(\lambda_3,\lambda_4) > -\frac{1}{k}$ if the distribution has finite k^{th} moments. The k^{th} center moment (except for the first moment) from $Q(y)$ is denoted by μ_k and given as:

$$\mu_1 = \frac{1}{\lambda_2}(s_1) - \frac{1}{\lambda_2\lambda_3} + \frac{1}{\lambda_2\lambda_4},$$

$$\mu_2 = \frac{1}{\lambda_2}(s_2 - s_1^2),$$

$$\mu_3 = \frac{1}{\lambda_2^3}(s_3 - 3s_1s_2 + 2s_1^3),$$

$$\mu_4 = \frac{1}{\lambda_2^4}(s_4 - 4s_1s_3 + 6_1^2 s_2 - 3s_1^4) \quad (4)$$

In theory, α_3 and α_4 are given by:

$$\alpha_3 = \frac{s_3-3s_1s_2+2s_1^3}{(s_2-s_1)^{\frac{3}{2}}},$$

$$\alpha_4 = \frac{s_4-4s_1s_3+6s_1^2s_2-3s_1^4}{(s_2-s_1)^2} \quad (5)$$

λ_3 and λ_4 can be found by minimizing (6), where $\hat{\alpha}_3$ and $\hat{\alpha}_4$ are sample values using sample moments.

$$\sqrt{(\hat{\alpha}_3-\alpha_3)^2 + (\hat{\alpha}_4-\alpha_4)^2} \quad (6)$$

This equation is used in the optimization process to determine the initial values in such a way that the expression is minimized, thereby obtaining the optimal values for λ_3 and λ_4 Once λ_3 and λ_4 are determined, λ_1 and λ_2 can be calculated utlizing:

$$\lambda_2 = \frac{\sqrt{s_2-s_1^2}}{\hat{\alpha}_2},$$

$$\lambda_1 = \hat{\alpha}_1 + \frac{1}{\lambda_2}\left(\frac{1}{\lambda_3+1} - \frac{1}{\lambda_4+1}\right) \quad (7)$$

IV. DATASET

The Indonesian company financial performance dataset presents data from various company financial performance variables. The data source comes from annual reports that can be accessed publicly and freely on the IDX website. Data for 451 companies were collected from 2020 to 2022 and classified by sector and board level. There are 11 classifications of company sectors, namely healthcare, basic materials, financials, transportation and logistics, technology, consumer non-cyclical, industrials, energy, consumer cyclical, and infrastructures, with two board levels, namely main and development. Table II presents the variables, data types, and data sources in the dataset. The dataset contains 24 variables, with data sources coming from the annual financial report. The annual financial report data for each company can be downloaded for free on the IDX website in PDF format. The data were then extracted into an Excel spreadsheet. Table III provides a description and calculation formula for each financial performance variable.

TABLE II. DATASET

Section	Variables	Data type	Data source
Income statement	Total Revenue	Ratio	Annual Report
	Net Income Common Stockholders	Ratio	Annual Report
	Basic EPS	Ratio	Annual Report
	Basic Average Shares	Ratio	Annual Report
Balance sheet	Total Assets	Ratio	Annual Report
	Total Liabilities Net Minority Interest	Ratio	Annual Report
	Total Equity Gross Minority Interest	Ratio	Annual Report
	Share Issued	Ratio	Annual Report
	Common Stock Equity	Ratio	Annual Report
	Net Tangible Assets	Ratio	Annual Report
	Tangible Book Value	Ratio	Annual Report
Cash flow	Cash Flows from Used in Operating Activities Direct	Ratio	Annual Report
	Investing Cash Flow	Ratio	Annual Report
	Financing Cash Flow	Ratio	Annual Report
	Capital Expenditure	Ratio	Annual Report
	Free Cash Flow	Ratio	Annual Report
Financial performance ratios	Return on Assets (ROA)	Percentage (%)	Annual Report
	Return on Equity (ROE)	Percentage (%)	Annual Report
	Net Profit Margin (NPM)	Percentage (%)	Annual Report
	Debt to Asset Ratio (DAR)	Percentage (%)	Annual Report
	Debt to Equity Ratio (DER)	Percentage (%)	Annual Report
	Stock Price	Ratio	Annual Report
	Adjusted Stock Price	Ratio	Annual Report
	Volume	Ratio	Annual Report

TABLE III. DESCRIPTION AND FORMULA FOR EACH FINANCIAL PERFORMANCE RATIO VARIABLE

Index	Description	Formula
ROA	It is one of the most basic metrics in measuring business success [55]. ROA measures a company's ability to generate profits using the assets it owns [56-58]. A higher ROA indicates that a company is better at utilizing its assets to achieve profits. ROA can be used to decide which company is the best to invest in by comparing the ROA of similar companies.	Net Income Common Stockholders / Total Assets × 100
ROE	Many investors consider ROE for investing in shares [56]. ROE measures a company's ability to generate profits using funds invested by shareholders. ROE can also be used to compare which company is better in terms of generating profits. If ROE increases, share prices also tend to increase, and vice versa [57, 59].	Net Income Common Stockholders / Total Equity Gross Minority Interest * 100
NPM	NPM is the first candidate to measure overall profitability [57]. Net profit, net margin, or net profit of a company is the most comprehensive measure of profitability. Net income explains how much money a company keeps after paying all its costs and expenses. Companies can also lose money, and in this case, it is called net loss [56].	Net Income Common Stockholders / Total Revenue * 100
DAR	DAR is the ratio between the company's debt and the assets owned by the company. DAR measures the ability of a company's assets to cover its debts [58].	Total Liabilities Net Minority Interest / Total Assets * 100
DER	DER is one of the most basic measures of a company's debt burden, explaining how a company's debt compares to the amount of money it earns from stock investors. The higher the DER, the greater the debt burden compared to shares [56].	Total Liabilities Net Minority Interest / Total Equity Gross Minority Interest * 100

V. RESULTS

Figure 1 illustrates the dataset within the INDCOMP program. The dataset comprises financial performance data from companies listed on the IDX, classified into 11 sectors: consumer cyclical, basic materials, healthcare, financials, transportation & logistics, technology, industrials, consumer non-cyclical, infrastructures, and energy. In addition, there are two levels of boards: main and development. INDCOMP presents a collection of 24 financial indicators, such as Total Revenue, Net Income Common Stockholders, Basic EPS, Basic Average Shares, Total Assets, Total Liabilities, Net Minority Interest, Total Equity Gross Minority Interest, Share Issued, Common Stock Equity, Net Tangible Assets, Tangible Book Value, Cash Flows from Used in Operating Activities Direct, Investing Cash Flow, Financing Cash Flow, Capital Expenditure, Free Cash Flow, Return on Assets (ROA), Return on Equity (ROE), Net Profit Margin (NPM), Debt to Asset Ratio (DAR), Debt to Equity Ratio (DER), Stock Price, Adjusted Stock Price, and Volume. INDCOMP users can apply filters or make selections based on sector, year, and board level to perform an analysis.

The filtering functionality enables users to customize the dataset according to their specific requirements and areas of interest, making it a potent instrument for performing in-depth financial analysis. The INDCOMP dataset uses red to represent negative values and green to represent positive values. Color coding facilitates the rapid identification and comprehension of crucial financial indicators, thus improving the usability and legibility of the data. The abbreviation NA denotes that the value is unavailable or not provided. It is crucial for users to comprehend the dataset and to make well-informed judgments based on the accessible data. Figure 2 portrays the functionality that allows the visualization of the correlation matrix between the financial performance characteristics of companies. Comprehending the connections among these variables is vital for investors, analysts, and decision-makers to assess the overall well-being and effectiveness of companies. INDCOMP provides users with the option to use Pearson or Spearman correlation methods, based on the characteristics of the data and the goals of their study. This option enables improved flexibility and precision in capturing the fundamental patterns and interconnections within the financial dataset. Moreover, INDCOMP streamlines the process of choosing variables for correlation calculations. Users can easily select from a wide range of financial factors in the dataset, making the analysis process more efficient and quick. In addition, the platform provides a variety of customization options, including the ability to modify font size and the number of decimal places displayed in the correlation matrix. These customization capabilities allow users to personalize the visualization according to their tastes and analytical needs, guaranteeing a smooth and effective analysis experience.

INDCOMP helps to understand the intricate connections among financial factors by offering flexibility and user-friendly features. Users can analyze the relationships between various indicators, detect possible patterns or trends, and acquire useful insights into the factors that influence financial performance. Users can use these advanced analytical capabilities to make well-informed judgments, develop strategic plans, and optimize investment strategies with certainty. The correlation matrix feature in INDCOMP is a potent tool for performing thorough financial research and revealing useful insights into the interrelatedness of financial variables. INDCOMP enables users to make informed decisions and navigate the ever-changing financial markets by providing tools for risk assessment, investment evaluation, and performance analysis.

Dataset

No	Perusahaan	Sektor	Papan	Tahun	Tanggal Pencatatan	Tahun Pencatatan	Total Revenue	Net Income Common Stockholders	Basic EPS	Basic Average Shares	Net Income from Continuing & Discontinued Operation	Total Assets
1	EMTK 2022	Technology	Utama	2022	12 Januari 2010	2010	15524642337.0	5444009955.0	89.39	60899140	5444009955.0	44469025417.0
2	ATIC 2022	Technology	Utama	2022	8 Juli 2015	2015	7809976522.0	68786548.0	31.7	2315361	68786548.0	4217064303.0
3	MLPT 2022	Technology	Utama	2022	8 Juli 2013	2013	3442223000.0	559004000.0	298.0	1875000	541965000.0	2720784000.0
4	MTDL	Technol	Utama	2022	9 April	1990	209882	580496	47.0	122768	580496	858289

Fig. 1. Dataset.

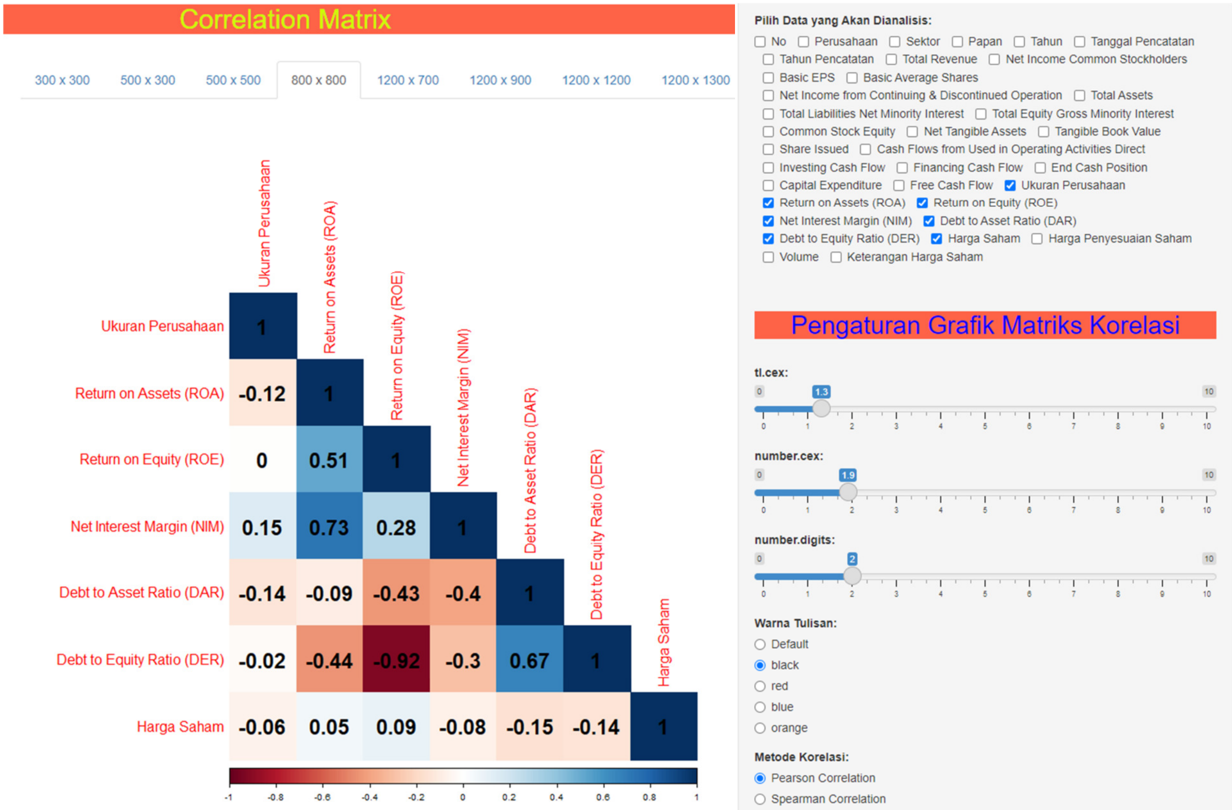


Fig. 2. Correlation matrix features.

Figure 2 demonstrates that the Pearson correlation coefficient between the Net Interest Margin (NIM) and ROA is 0.73, demonstrating a robust positive link between them. This indicates that as NIM increases, there is a strong probability that ROA will also increase and vice versa. The positive correlation exhibits a symbiotic connection between the effectiveness of asset management and the net profit margin of a corporation. On the contrary, the Pearson correlation coefficient between DER and ROE is -0.92, manifesting a highly robust negative connection between them. As the DER grows, indicating a higher level of debt relative to equity, the ROE tends to drop and vice versa. The negative correlation reveals a reciprocal connection between a company's financial structure and the return on investment for shareholders.

In the correlation matrix, boxes with values closer to -1 or 1 are tinted darker, while boxes with values closer to 0 are shaded lighter. This color-coding method enables users to quickly discern the magnitude and direction of correlations between different pairs of variables. As a result, users can easily identify important trends in the data and make well-informed judgments in their financial analysis. Moreover, the correlation matrix functionality allows users to simultaneously examine correlations between several financial performance parameters. By comprehensively displaying these links, users can acquire more profound insights into the interconnections among many facets of a company's financial performance. With a comprehensive understanding, users can identify possible areas for improvement, evaluate risks with greater efficiency, and develop plans to maximize overall financial results. The correlation matrix is useful for investors, analysts, and decision-makers, as it enables them to make data-driven decisions and navigate the complexity of financial markets with greater confidence and knowledge.

Figure 3 depicts a visualization of the scatterplot along with the data distribution. For example, the scatterplot between ROA and NIM has a positive direction and is known to have a correlation value of 0.726. The ROA distribution tends to be symmetrical, whereas the NIM distribution tends to be skewed to the right. The scatterplot between ROE and DER has a negative direction and is known to have a correlation value of -0.925.

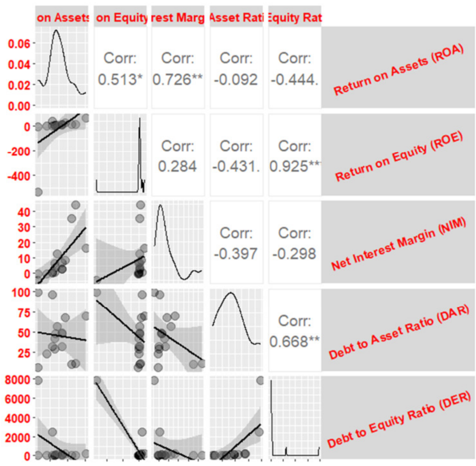
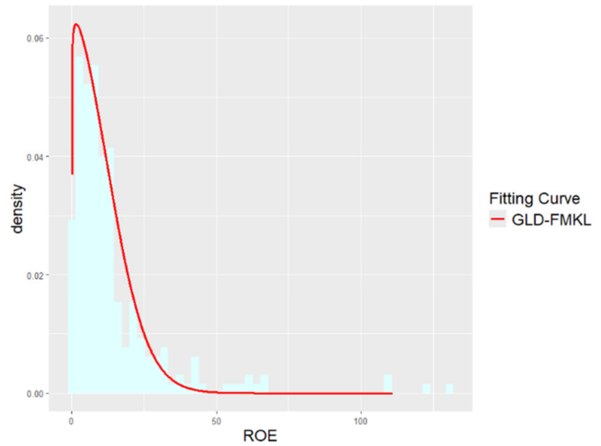


Fig. 3. Scatterplot and data distribution.

Figures 4 and 5 illustrate the application of GLD in matching the ROA and ROE data distribution. In addition to GLD, normal, Weibull, lognormal, and gamma distribution approaches are used in data fitting. The distribution of ROA and ROE tends to be skewed to the right.

(a)



(b)

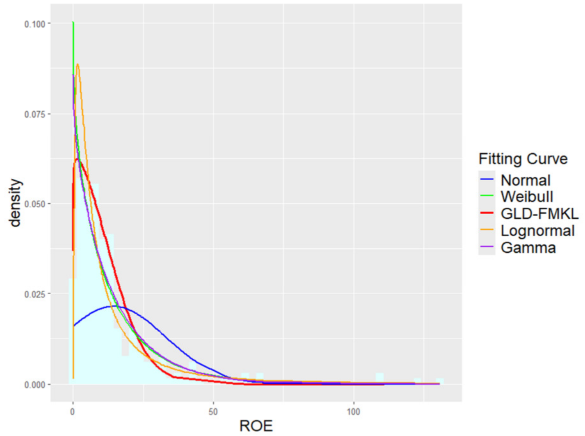


Fig. 4. Fitting ROE toward GLD (a) and All distribution (b).

Table III portrays the results of the significance tests for fitting the distribution models to the data. The models implemented include the normal, Weibull, lognormal, and gamma distributions along with the GLD. Figure 4(a) shows the application of GLD to ROE, with the histogram demonstrating a right-skewed data distribution. Figure 4(b) displays the utilization of the normal, Weibull, lognormal, gamma, and GLD models to fit the data. The findings of the significance test indicate that only the GLD model fits the data well, as evidenced by a p-value greater than 0.05. ROA and ROE data were utilized for the simulation of model fitting. Figure 5(a) discloses the application of GLD to ROA, with the histogram exhibiting a noticeable right skew in the data distribution. Figure 5(b) presents the use of the normal, Weibull, lognormal, Gamma, and GLD models to accurately match the data.

The findings in Table III provide useful insights into the efficacy of different distribution models in accurately fitting the ROA and ROE data. The GLD model is a suitable choice for calculating ROA, as it demonstrates a favorable match with a p-value of 0.227. This suggests that the GLD model effectively represents the range of variability in ROA for various observations. In contrast, the normal distribution model, which has a considerably smaller p-value, indicates that it is not suitable for accurately reflecting the ROA data. This emphasizes the drawbacks of assuming a normal distribution for this metric. However, alternative distribution models, such as Weibull, Lognormal, and Gamma have favorable fits, indicating their potential utility in simulating ROA. Regarding ROE, the GLD model demonstrates its effectiveness once again, as observed by a substantial p-value of 0.824, showing strong alignment with the data. This supports the idea that the GLD model effectively depicts the distribution of ROE, precisely capturing its variability across several observations.

(a)

(b)

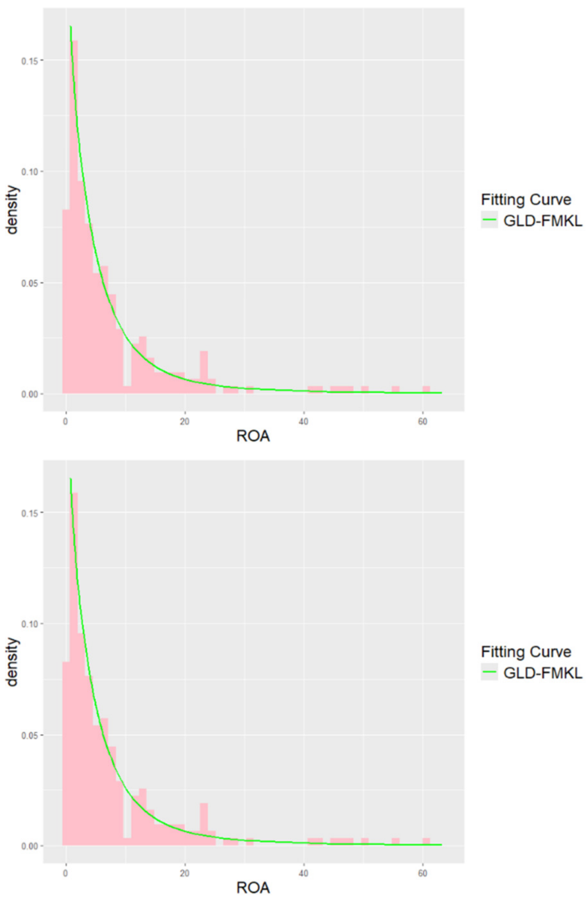


Fig. 5. Fitting ROA toward GLD (a) and All Distribution (a).

TABLE IV. COMPARISON OF GOODNES OF FIT RESULTS

Variable	Distribution	p-values	Conclusion
ROA	GLD	p = 0.227	p > 0.05 (Fit)
	Normal	p = 6.0228e-11	p < 0.05 (Not Fit)
	Weibull	p = 0.5598	p > 0.05 (Fit)
	Lognormal	p = 0.4665	p > 0.05 (Fit)
	Gamma	p = 0.2796	p > 0.05 (Fit)
ROE	GLD	p = 0.824	p > 0.05 (Fit)
	Normal	p = 1.7508e-12	p < 0.05 (Not Fit)
	Weibull	p = 0.0334	p < 0.05 (Not Fit)
	Lognormal	p = 0.0339	p < 0.05 (Not Fit)
	Gamma	p = 0.0149	p < 0.05 (Not Fit)

The findings of the significance test reveal that the Weibull, lognormal, gamma, and GLD models effectively fit the ROE data since all p-values are greater than 0.05. However, the conventional model is not a good fit, as indicated by a p-value less than 0.05. According to the findings, the GLD model is preferable to accurately fit both ROE and ROA data, as it exhibits a stronger match compared to alternative distribution models. The asymmetrical structure of the data distributions implies that conventional distribution models may not adequately capture the inherent patterns in the financial performance measurements. Therefore, the GLD model provides a more precise representation of the data and improves the effectiveness of financial analysis and modeling efforts.

The GLD model's supremacy implies that it is well-suited for capturing the underlying intricacies in the distributions of these variables. However, the inconsistency of various distributions underscores the importance of meticulous deliberation and verification when selecting a model for financial modeling and decision-making objectives. These observations can provide valuable information for strategic planning, risk assessment, and investment decisions, improving the efficiency of financial analysis in many situations.

VI. CONCLUSION

INDCOMP is an innovative online tool created to provide extensive financial performance information on companies listed on the Indonesia Stock Exchange (IDX). INDCOMP seeks to provide a wide range of data and analytical tools to assist investors, practitioners, and researchers in assessing and quantifying the health of these enterprises. This platform is notable for its practicality, versatility, and advanced capabilities, many of which are not accessible on the IDX website. An important benefit of INDCOMP is its intuitive interface, which enables users to easily retrieve and personalize financial performance statistics. Users can apply filters to the data, allowing them to narrow the information based on particular sectors, years, and individual companies. By offering a high degree of personalization, customers can acquire precise data without having to search for extraneous information. INDCOMP offers customizable data analysis tools that cater to the needs of investors seeking insight into the financial stability of a given sector. Furthermore, INDCOMP surpasses simple data presentation by integrating diverse visualization methods.

These technologies facilitate the creation of charts, graphs, and other visual aids, enhancing the ease and intuitiveness of data interpretation. For example, users can visually represent patterns in revenue, profit margins, or debt levels as they change over time. These visualizations are essential for identifying patterns, making well-informed decisions, and presenting the data clearly and understandably. This feature greatly improves the user experience by making complex financial data easily accessible and comprehensible to users with different degrees of knowledge. INDCOMP provides comprehensive statistical analysis capabilities in addition to data display. Within the platform, users can perform a variety of statistical tests and analyses. This capability is very beneficial for academics and analysts who need to perform a thorough data analysis. INDCOMP offers superior statistical features compared to the IDX website, allowing users to perform comprehensive studies without the need for additional software. As an illustration, users can calculate financial ratios and perform regression analysis or hypothesis testing, all within the same platform. One notable characteristic of INDCOMP is its ability to replicate different probability distributions for financial performance data.

Of all the options, GLD has demonstrated the greatest precision in fitting the data. When compared with normal, lognormal, Weibull, and gamma distributions in simulations, GLD consistently demonstrated the most accurate match for the financial performance data of companies listed on IDX. Precision in data fitting is essential to generate reliable projections and well-informed investment choices. INDCOMP is an indispensable tool for investors to make well-informed decisions. Through the examination of past financial data and the representation of patterns, investors can detect possible investment prospects and assess the risks associated with various companies. With the ability to perform statistical studies, investors can review the financial stability and performance of organizations with more precision. The use of an all-encompassing method for data analysis has the potential to result in more astute investment choices, enhancing investment results.

ACKNOWLEDGMENT

This research was fully funded by the Universitas Sumatera Utara. The authors express their sincere appreciation to the external reviewer, Rezzy Eko Caraka, who is a Professor at Chaoyang University of Technology, Taiwan, and a Researcher at BRIN Indonesia. His invaluable assistance was instrumental in ensuring the accuracy and thoroughness of this study. His deep knowledge, astute observations, and comprehensive evaluations significantly enhanced the excellence, reliability, and effectiveness of this approach and the analysis of its findings. The authors appreciate his unwavering attention and strong commitment to achieving academic excellence, which has significantly contributed to improving the overall credibility of the study project.

DATA AVAILABILITY

The source code, materials, and data findings of this study are openly available in full access by the corresponding author.

REFERENCES

[1] "Indonesia Central Securities Depository (KSEI)." https://www.ksei.co.id/?setLocale=en-US.

[2] "Statistik Pasar Modal Semester II-2022," Financial Services Authority (OJK), Jakarta, Indonesia, 2022. [Online]. Available: https://www.ojk.go.id/id/kanal/pasar-modal/data-dan-statistik/statistik-pasar-modal/Documents/Statistik%20Pasar%20Modal%20Semester%20II-2022.pdf.

[3] N. Yalcin, A. Bayrakdaroglu, and C. Kahraman, "Application of fuzzy multi-criteria decision making methods for financial performance evaluation of Turkish manufacturing industries," *Expert Systems with Applications*, vol. 39, no. 1, pp. 350–364, Jan. 2012, https://doi.org/10.1016/j.eswa.2011.07.024.

[4] K. Kara, G. C. Yalçın, A. Çetinkaya, V. Simic, and D. Pamucar, "A single-valued neutrosophic CIMAS-CRITIC-RBNAR decision support model for the financial performance analysis: A study of technology companies," *Socio-Economic Planning Sciences*, vol. 92, Apr. 2024, Art. no. 101851, https://doi.org/10.1016/j.seps.2024.101851.

[5] "Indonesia Stock Exchange (IDX)." https://www.idx.co.id/.

[6] M. Jimichi and S. Maeda, "Visualization and Statistical Modeling of Financial Data with R," presented at the R User Conference, 2014.

[7] M. Jimichi, "Building of Financial Database Servers," Technical Report, 2010.

[8] M. Jimichi, D. Miyamoto, C. Saka, and S. Nagata, "Visualization and statistical modeling of financial big data: double-log modeling with skew-symmetric error distributions," *Japanese Journal of Statistics and Data Science*, vol. 1, no. 2, pp. 347–371, Dec. 2018, https://doi.org/10.1007/s42081-018-0019-1.

[9] M. Jimichi, Y. Kawasaki, D. Miyamoto, C. Saka, and S. Nagata, "Statistical Modeling of Financial Data with Skew-Symmetric Error Distributions," *Symmetry*, vol. 15, no. 9, Sep. 2023, Art. no. 1772, https://doi.org/10.3390/sym15091772.

[10] W. Chang *et al.*, "shiny: Web Application Framework for R." [Online]. Available: https://cran.r-project.org/web/packages/shiny/index.html.

[11] P. Martinková and A. Drabinová, "ShinyItemAnalysis for Teaching Psychometrics and to Enforce Routine Analysis of Educational Tests," *The R Journal*, vol. 10, no. 2, 2019, Art. no. 503, https://doi.org/10.32614/RJ-2018-074.

[12] L. Savini, L. Candeloro, S. Perticara, and A. Conte, "EpiExploreR: A Shiny Web Application for the Analysis of Animal Disease Data," *Microorganisms*, vol. 7, no. 12, Dec. 2019, Art. no. 680, https://doi.org/10.3390/microorganisms7120680.

[13] B. Sivaprakasam and P. Sadagopan, "Development of an Interactive Web Application 'Shiny App for Frequency Analysis on Homo sapiens Genome (SAFA-HsG),'" *Interdisciplinary Sciences: Computational Life Sciences*, vol. 11, no. 4, pp. 723–729, Dec. 2019, https://doi.org/10.1007/s12539-019-00340-z.

[14] L. de F. Ramalho and W. R. de C. Segundo, "R-Shiny as an Interface for Data Visualization and Data Analysis on the Brazilian Digital Library of Theses and Dissertations (BDTD)," *Publications*, vol. 8, no. 2, Jun. 2020, Art. no. 24, https://doi.org/10.3390/publications8020024.

[15] H. A. Ekiz, C. J. Conley, W. Z. Stephens, and R. M. O'Connell, "CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments," *BMC Bioinformatics*, vol. 21, no. 1, May 2020, Art. no. 191, https://doi.org/10.1186/s12859-020-3538-2.

[16] Y. Zhou, S. Leung, S. Mizutani, T. Takagi, and Y. S. Tian, "MEPHAS: an interactive graphical user interface for medical and pharmaceutical statistical analysis with R and Shiny," *BMC Bioinformatics*, vol. 21, no. 1, May 2020, Art. no. 183, https://doi.org/10.1186/s12859-020-3494-x.

[17] T. N. Moore, R. Q. Thomas, W. M. Woelmer, and C. C. Carey, "Integrating Ecological Forecasting into Undergraduate Ecology Curricula with an R Shiny Application-Based Teaching Module," *Forecasting*, vol. 4, no. 3, pp. 604–633, Sep. 2022, https://doi.org/10.3390/forecast4030033.

[18] S. Iddi and M. C Donohue, "Power and Sample Size for Longitudinal Models in R - The longpower Package and Shiny App," *The R Journal*, vol. 14, no. 1, pp. 264–282, Jul. 2022, https://doi.org/10.32614/RJ-2022-022.

[19] H. Wang *et al.*, "VisProDom: an interactive Shiny/R application for displaying protein domains with transcriptional features," *BMC Genomics*, vol. 23, no. 1, Jun. 2022, Art. no. 472, https://doi.org/10.1186/s12864-022-08713-3.

[20] "R: The R Project for Statistical Computing." https://www.r-project.org/.

[21] "RStudio IDE." [Online]. Available: https://www.posit.co/.

[22] V. Nijs, F. Fang, Trestle Technology, Llc, and J. Allen, "shinyAce: Ace Editor Bindings for Shiny." [Online]. Available: https://cran.R-project.org/package=shinyAce.

[23] J. B. Arnoid, "ggthemes: Extra Themes, Scales and Geoms for ggplot2." [Online]. Available: https://cran.r-project.org/web/packages/ggthemes/.

[24] H. Wickham, "plyr: Tools for Splitting, Applying and Combining Data." [Online]. Available: https://cran.r-project.org/web/packages/plyr/.

[25] Y. Xie *et al.*, "DT: A Wrapper of the JavaScript Library 'DataTables.'" [Online]. Available: https://cran.r-project.org/web/packages/DT/.

[26] J. Owen *et al.*, "rhandsontable: Interface to the 'Handsontable.js' Library." [Online]. Available: https://cran.r-project.org/web/packages/rhandsontable/.

[27] H. Wickham *et al.*, "ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics." [Online]. Available: https://cran.r-project.org/web/packages/ggplot2/.

[28] C. Sievert *et al.*, "plotly: Create Interactive Web Graphics via 'plotly.js.'" [Online]. Available: https://cran.r-project.org/web/packages/plotly/.

[29] S. Su *et al.*, "GLDEX: Fitting Single and Mixture of Generalised Lambda Distributions." [Online]. Available: https://cran.r-project.org/web/packages/GLDEX/.

[30] P. Schauberger, A. Walker, L. Braglia, J. Sturm, J. M. Garbuszus, and J. M. Barbone, "openxlsx: Read, Write and Edit xlsx Files." [Online]. Available: https://cran.r-project.org/web/packages/openxlsx/.

[31] B. Fournier, N. Rupin, M. Bigerelle, D. Najjar, A. Iost, and R. Wilcox, "Estimating the parameters of a generalized lambda distribution," *Computational Statistics & Data Analysis*, vol. 51, no. 6, pp. 2813–2835, Mar. 2007, https://doi.org/10.1016/j.csda.2006.09.043.

[32] S. Su, "A Discretized Approach to Flexibly Fit Generalized Lambda Distributions to Data," *Journal of Modern Applied Statistical Methods*, vol. 4, no. 2, Nov. 2005, https://doi.org/10.22237/jmasm/1130803560.

[33] Z. A. Karian and E. J. Dudewicz, *Handbook of Fitting Statistical Distributions With R*, 1st edition. Boca Raton, FL, USA: Chapman & Hall, 2010.

[34] A. H. Syafrina, A. Norzaida, and O. N. Shazwani, "Stochastic Modeling of Rainfall Series in Kelantan Using an Advanced Weather Generator," *Engineering, Technology & Applied Science Research*, vol. 8, no. 1, pp. 2537–2541, Feb. 2018, https://doi.org/10.48084/etasr.1709.

[35] P. Balamurugan, T. Yuvaraj, and P. Muthukannan, "Optimal Allocation of DSTATCOM in Distribution Network Using Whale Optimization Algorithm," *Engineering, Technology & Applied Science Research*, vol. 8, no. 5, pp. 3445–3449, Oct. 2018, https://doi.org/10.48084/etasr.2302.

[36] R. E. Caraka, M. Noh, R. C. Chen, Y. Lee, P. U. Gio, and B. Pardamean, "Connecting Climate and Communicable Disease to Penta Helix Using Hierarchical Likelihood Structural Equation Modelling," *Symmetry*, vol. 13, no. 4, Apr. 2021, Art. no. 657, https://doi.org/10.3390/sym13040657.

[37] R. E. Caraka *et al.*, "Cluster Around Latent Variable for Vulnerability Towards Natural Hazards, Non-Natural Hazards, Social Hazards in West Papua," *IEEE Access*, vol. 9, pp. 1972–1986, 2021, https://doi.org/10.1109/ACCESS.2020.3038883.

[38] R. E. Caraka *et al.*, "Micro, Small, and Medium Enterprises' Business Vulnerability Cluster in Indonesia: An Analysis Using Optimized Fuzzy Geodemographic Clustering," *Sustainability*, vol. 13, no. 14, Jan. 2021, Art. no. 7807, https://doi.org/10.3390/su13147807.

[39] R. E. Caraka, R. C. Chen, S. W. Huang, S. Y. Chiou, P. U. Gio, and B. Pardamean, "Big data ordination towards intensive care event count cases using fast computing GLLVMS," *BMC Medical Research Methodology*, vol. 22, no. 1, Mar. 2022, Art. no. 77, https://doi.org/10.1186/s12874-022-01538-4.

[40] R. E. Caraka *et al.*, "Connectivity, sport events, and tourism development of Mandalika's special economic zone: A perspective from big data cognitive analytics," *Cogent Business & Management*, vol. 10, no. 1, Dec. 2023, Art. no. 2183565, https://doi.org/10.1080/23311975.2023.2183565.

[41] R. E. Caraka *et al.*, "Albatross analytics a hands-on into practice: statistical and data science application," *Journal of Big Data*, vol. 9, no. 1, May 2022, Art. no. 70, https://doi.org/10.1186/s40537-022-00626-y.

[42] F. H. Putra and R. Kurniawan, "Clustering for Disaster Areas Endemic Dengue Hemorrhagic Fever Based on Factors Had Caused in East Java Using Fuzzy Geographically Weighted Clustering - Particle Swarm Optimization," *Jurnal Aplikasi Statistika & Komputasi Statistik*, vol. 8, no. 1, pp. 27–27, Jun. 2016, https://doi.org/10.34123/jurnalasks.v8i1.10.

[43] B. I. Nasution, R. Kurniawan, T. H. Siagian, and A. Fudholi, "Revisiting social vulnerability analysis in Indonesia: An optimized spatial fuzzy clustering approach," *International Journal of Disaster Risk Reduction*, vol. 51, Dec. 2020, Art. no. 101801, https://doi.org/10.1016/j.ijdrr.2020.101801.

[44] P. A. Kaban, R. Kurniawan, R. E. Caraka, B. Pardamean, B. Yuniarto, and Sukim, "Biclustering Method to Capture the Spatial Pattern and to Identify the Causes of Social Vulnerability in Indonesia: A New Recommendation for Disaster Mitigation Policy," *Procedia Computer Science*, vol. 157, pp. 31–37, Jan. 2019, https://doi.org/10.1016/j.procs.2019.08.138.

[45] N. Masseran, "Markov Chain model for the stochastic behaviors of wind-direction data," *Energy Conversion and Management*, vol. 92, pp. 266–274, Mar. 2015, https://doi.org/10.1016/j.enconman.2014.12.045.

[46] N. Masseran and M. A. M. Safari, "Modeling the transition behaviors of PM10 pollution index," *Environmental Monitoring and Assessment*, vol. 192, no. 7, Jul. 2020, Art. no. 441, https://doi.org/10.1007/s10661-020-08376-1.

[47] M. M. Hamasha, "A Mathematical Approximation of the Left-sided Truncated Normal Distribution Using the Cadwell Approximation Model," *Engineering, Technology & Applied Science Research*, vol. 7, no. 1, pp. 1382–1386, Feb. 2017, https://doi.org/10.48084/etasr.988.

[48] Y. Chalabi, D. J. Scott, and D. Wuertz, "Flexible distribution modeling with the generalized lambda distribution," MPRA Paper, Nov. 2012. [Online]. Available: https://mpra.ub.uni-muenchen.de/43333/.

[49] Y. Chalabi, D. J. Scott, and D. Wurtz, "The Generalized Lambda Distribution as an Alternative to Model Financial Returns," ETH Econohysics Working and White Papers Series.

[50] D. S. Dedduwakumara, L. A. Prendergast, and R. G. Staudte, "A simple and efficient method for finding the closest generalized lambda distribution to a specific model," *Cogent Mathematics & Statistics*, vol. 6, no. 1, Jan. 2019, Art. no. 1602929, https://doi.org/10.1080/25742558.2019.1602929.

[51] J. S. Ramberg and B. W. Schmeiser, "An approximate method for generating asymmetric random variables," *Communications of the ACM*, vol. 17, no. 2, pp. 78–82, Oct. 1974, https://doi.org/10.1145/360827.360840.

[52] A. Tarsitano, "Fitting The Generalized Lambda Distribution to Income Data," in *COMPSTAT 2004 Symposium*, 2004, pp. 1861–1867.

[53] P. Owusu Junior and C. H. Korkpoe, "Generalised Lambda Distributions by Method of Moments and Maximum Likelihood using the JSE-ASI Returns," *Asian Journal of Finance & Accounting*, vol. 9, no. 1, Apr. 2017, Art. no. 224, https://doi.org/10.5296/ajfa.v9i1.10913.

[54] S. Su, "Numerical maximum log likelihood estimation for generalized lambda distributions," *Computational Statistics & Data Analysis*, vol. 51, no. 8, pp. 3983–3998, May 2007, https://doi.org/10.1016/j.csda.2006.06.008.

[55] M. A. Tracy, Ratio Analysis Fundamentals: How 17 Financial Ratios Can Allow You to Analyse Any Business on the Planet, 1st ed. CreateSpace Independent Publishing Platform, 2012.

[56] M. Krantz, *Fundamental Analysis For Dummies*, 2nd ed. Wiley, 2016.

[57] P. Walton and W. Aerts, *Global Financial Accounting and Reporting: Principles and Analysis*. London, UK: CENGAGE Learning, 2006.

[58] R. Palat, *Fundamental Analysis for Investors*, 4th ed. Vision Books, 2016.

[59] S. Kurniati, "Stock returns and financial performance as mediation variables in the influence of good corporate governance on corporate value," *Corporate Governance: The International Journal of Business in Society*, vol. 19, no. 6, pp. 1289–1309, Dec. 2019, https://doi.org/10.1108/CG-10-2018-0308.