Analysis of Linear Dependence between Financial Ratios and Stock Price Movements in Indonesian Bank using Principal Component Analysis

Leonardus Brain Fatolosja - 13524146
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13524146@std.stei.itb.ac.id, [2]leonardusbrain0@gmail.com

***Abstract*— This paper examines the correlation between financial ratios and stock price movements. The study focuses specifically on Indonesian bank, including BRI. Based on the financial ratios reported in their annual financial statements, this study investigates the linear dependence between financial ratios and daily stock price movements over a five-year period. The methodological approach employed is Principal Component Analysis (PCA).**

***Keywords*— Principal Component Analysis (PCA), linear algebra, financial ratios, stock price movements, BRI**

I. Introduction

Over the years, the Indonesian economy has experienced significant fluctuations, partly driven by COVID-19 Pandemic, which affected stock market from its initial outbreak through the recovery period. Along with the increasing trend in the number of stock market investors in Indonesia, several notable market events have occured in recent years. Many assume that some companies with poor financial ratios still experience increasing stock prices for reasons that remain unexplained.

According to conventional financial theory, Companies with strong financial ratios tend to have an increasing stock prices. For a while, a lot of people agree and believe at that theory. At the other hand, While holding a strong financial ratios, the banking sector in Indonesia having an abnormal price movements. Empirical observations in the Indonesian banking sector indicate that a number of banks with strong financial fundamentals have experienced volatile price movements, including a decreasing prices over a five-year period. This phenomenon challenges traditional financial theory and affect investors to reconsider the reliablity of financial ratios or fundamental analysis as sole indicators of stock performance.

These irregularities highlight the inherently complex and unpredictable nature of stock markets, underscoring the need for a more rigorous quantitative investigation into the underlying relationships between financial ratios and stock price movements. With Linear Algebra, the identification of dependence between financial ratios and stock prices becomes possible. By representing financial data in matrix form and applying Principal Component Analysis (PCA), high dimensional raw data, such as daily stock prices, can be transformed into informative lower dimensional representations.

These informative representations can be used to calculate the correlation among multiple variables. Furthermore, computational implementation enables automated process and allow for the analysis of large scale datasets. The use of large datasets improves the quality of both the results and the conclusions. Specifically, this paper will analysis Bank Rakyat Indonesia (BRI) as one of Indonesian Bank. The output from this paper will be a linear dependence between finantial ratios and stock price of BRI.

II. Theory

Before implementation of PCA and analysis the variable, involve financial ratios and stock price, its necessary to understand about PCA and fundamental matrix theory. In addition, core concepts of Linear Algebra must be reviewed in order to support the transformation of quantitative data into a valuable information that can be processed later. Last, essential and fundamental financial theory must be considered to ensure proper identification, selection, and collection of relevant datasets.

A. Matrix

Matrix is a set of elements represented as two - dimensional array. If a matrix has m row and n column, then A_{ij} denotes the element of matrix A in row i and column j. A vector is the same as matrix but with only one column or row.

1.) Basic Operations on Matrix
A matrix has some operations, including addition, substraction, scalar multiplication, and matrix multiplication.

Addition and Substraction
If two matrices have the same size (row and column), then the following equation applies:

$$(A \pm B)_{ij} = a_{ij} \pm b_{ij}$$

Scalar Multiplication
If c is a scalar, then the following equation applies:

$$(cA)_{ij} = ca_{ij}$$

Matrix Multiplication

If $A \in R^{m\times n}$ and $B \in R^{n\times p}$, then the following equation applies:

$$AB \in R^{m\times p}, \quad (AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

2.) Matrix Properties

The following properties apply to matrix:

- Associative : $(AB)C = A(BC)$
- Distributive : $A(B + C) = AB + AC$
- Not commutative : $AB \neq BA$
- Identity Matrix : $AI = IA = A$

3.) Transpose

A tranpose matrix is formed by swapping rows and columns in matrix.

$$(A^T)_{ij} = a_{ji}$$

Associative also apply to transpose matrix.

$$(AB)^T = B^T A^T$$

4.) Inverse

For a square matrix $A \in R^{n\times n}$, inverse A^{-1}satisfies:

$$A^{-1}A = AA^{-1} = I$$

Not every matrix has an inverse. Therefore, a matrix is called singular if it has not inverse.

5.) Orthogonal and Orthonormal Matrix

A matrix Q is orthogonal if:

$$Q^T Q = I$$

An Orthonormal matrix is an orthogonal matrix where all vectors inside it have magnitude of one.

B. Covariance Matrix

Suppose we have a random vector $x = (X_1, \dots, X_p)^\top$ with mean $\mu = E[x]$. The covariance matrix is

$$\Sigma = \mathrm{Cov}(x) = E[(x - \mu)(x - \mu)^\top].$$

The covariance between X_i *and* X_j with (i,j) entry is:

$$\Sigma_{ij} = \mathrm{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$$

From above, we can know that:

- Diagonal: (Σ_{ii}) is $\mathrm{Var}(X_i)$ or variance of each variable.
- Off-diagonal: Σ_{ij} tells whether two variables tend to move together/opposite/independent. (Correlation)

Sample Covariance Matrix

From sample, we can also create a covariance matrix. Let your dataset be n observation of p variables, arranged as matrix $X \in R^{n\times p}$ (rows = samples, columns = variables). Let $\bar{X}$ be the vector of column means and define the centered data:

$$X_c = X - 1\overline{x^\top}$$

A common estimator of the covariance matrix is the sample covariance matrix:

$$S = \frac{1}{n-1} X_c^\top X_c.$$

For a variables X and Y, The general form of a covariance matrix is given as follows:

$$S = \begin{bmatrix} \mathrm{Var}(X) & \mathrm{Cov}(X,Y) \\ \mathrm{Cov}(Y,X) & \mathrm{Var}(Y) \end{bmatrix}$$

If $\mathrm{Cov}(X,Y) > 0$, they tend to increase or decrease together. But if $\mathrm{Cov}(X,Y) < 0$, they move opposite meaning that one tends to rise when the other falls.

C. Eigenvalues and Eigenvectors

The primary function of eigen is to simplify the matrix by extracting the important value from it. It also means that it eliminate the noise of the matrix.

Let $A \in R^{nxn}$. A nonzero vector V is an **eigenvector** of A if applying A to V only scales it. So, there is no direction change of it.

$$A\,v = \lambda v$$

Where λ is the **eigenvalue** corresponding to V.

1.) Eigenvalues

In order to find eigenvalue of Matrix A, we must to rearrange previous equation.

$$(A - \lambda I)v = 0$$

For a nonzero solution V should be exist, the matrix $(A - \lambda I)$ must be singular, so:

$$det(A - \lambda I) = 0$$

From this equation, eigenvalue λ can be calculated.

2.) Eigenvectors

For each eigenvalue λ, eigenvectors are solution of:

$$(A - \lambda I)v = 0$$

The set of all solutions, including the zero vector, is called the eigenspaces.

3.) Diagonalization

If A has n linearly independent eigenvectors, stack them as columns of V and put eigenvalues in a diagonal matrix P. Then:

$$A = VPV^{-1}$$

This equation is useful for multiply matrix A with A. The reason is the process of multiplying A with itself becomes an easy operation because:

$$A^k = VP^kV^{-1}$$

D. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a linear dimensionality-reduction technique that replaces an original set of (possibly correlated) variables with a smaller set of new variables (principal components (PCs)) that are uncorrelated and ordered so that the first components retain the largest possible variance in the data.

To applying PCA, we must use concept of Covariance Matrix and Eigen (Eigenvalues and Eigenvectors) that has been explained before.

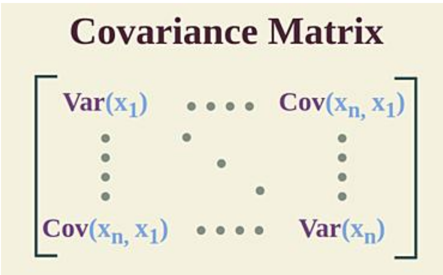
1.) Standardize Data

Before continue to applying PCA into some dataset, the data must be standardize for preventing variable dominance and ensuring equal weighting. For example, Net Interest Margin is expressed as a small percentage (0 – 1), while Total Assets may expressed as a big number (probably trillion rupiah). Without standardization, PCA would biasedly assign importance to variables with the largest absolute numerical ranges, mistaking high variance in scale for high informational significance. By applying the Z-score transformation as one of standardization method:

$$z = \frac{x - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation, the data is transformed so that each variables has a mean of 0 and a standard deviation of 1. This ensures that the PCA Algorithm threat every variables (financial ratios) as equally important at the beginning of analysis.

2.) Covariance Matrix
From part B, we learn the way to calculate covariance matrix. Once the covariance between every finacial ratios is calculated, they are arranged into a square matrix of size p X p (where p is the number of variables).



Picture 1.1 Covariance Matrix

(source: https://www.geeksforgeeks.org/maths/covariance-matrix/)

The diagonal of the matrix represents the variances of each feature and off-diagonal elements represent the correlation between different financial ratios.

3.) Principal Components and Eigen
The primary function of the Eigen concept is to simplify the matrix by extracting the most significant values from it. After the covariance matrix is established, we perform Eigen-decomposition to identify the new axes, known as Principal Components (PCs), where the data spreads out the most.

Eigenvectors represent the direction of the new axes. For a square covariance matrix A, an eigenvector X is a non-zero vector that satisfies the equation $AX = \lambda X$. This means that when the matrix acts on the vector, the direction remains unchanged because it only scales the vector. These eigenvectors define the stable directions of the data, which in this study represent the underlying relationships between various financial ratios and stock prices.

Eigenvalues are the scalars that represent the amount of variance captured by each eigenvector. Eigenvalues help rank these directions by importance. A higher eigenvalue indicates that the corresponding principal component explains a larger portion of the fluctuations in the banking data.

Once both Eigenvalues and Eigenvectors are calculated, they are ordered from highest to lowest. Principal Components 1 (PC1) is the direction associated with largest eigenvalue. This is also the direction of maximum variance. In this study, it's capturing the most dominant trend in the banking sector. Principal Components 2 (PC2) the next best direction that is orthogonal to PC1. Being orthogonal ensures that PC2 capture information that is uncorrelated with PC1, thus eliminating noise.

4.) Top Components and Transform Data
After the eigenvectors and eigenvalues are ranked, the final stage of PCA involves reducing the dimensionality of the dataset. This is achieved by selecting the top k components (those with the highest eigenvalues) which collectively account for a significant portion of the total variance (commonly targetting 95% or more). By discarding the components with low eigenvalues, we effectively filter out noise and less significant fluctuations, focusing only on the most robust patterns.

The original dataset is then transformed by projecting it onto these selected principal components. Mathematically, this transformation involves multiplying the standardized original data by the matrix of chosen eigenvectors. For example, a complex dataset with multiple financial ratios can be projected onto a 2D plane (PC1 and PC2) for visualization.

Specifically in this study, in this transformed space, the axes are no longer the original financial ratios like "NIM" or "LDR," but rather the Principal Components themselves. PC1 captures the maximum spread of the data, representing the most critical "signal" or trend in the banking sector's performance. PC2 captures the remaining variance while remaining strictly perpendicular (orthogonal) to PC1 making it's uncorrelated each other.

E. Financial Ratios in Banking

There are some financial ratios that will be a data for this analysis:

- Return Of Investment (ROI) : measures how efficiently a company uses its assets to generate profit. It is a core indicator of management performance.
- Price to Book Value (PBV) : a valuation ratio used to compare a company's current market price to its book value. It is particularly important to determine if a stock is overvalued or undervalued.

- Total Assets : represent the entire scale of the company, including loans, investments, and physical property. Larger assets often imply greater stability, though not necessarily higher growth.
- Market Capitalization : Market Capitalization represents the total market value of a company's outstanding shares. It is calculated by multiplying the current stock price by the total number of outstanding shares
- Return on Equity (ROE) : ROE is a powerful indicator of a company's profitability from the perspective of the shareholders. It measures how effectively the company uses the money invested by its owners to generate earnings.

III. Implementation

A. Planning

This paper is conducted to analyze the linear dependence between financial ratios and stock price movements of Indonesian banking companies, with Bank Rakyat Indonesia (BRI) as the primary case study. The planning stage establishes a systematic analytical roadmap that guides all subsequent implementation steps, from data preparation to result interpretation.

The analysis is structured into three major stages: data preparation, applying PCA algorithm, and dependence interpretation (result).

1.) Data Preparation

In the first stage, raw data are collected from publicly accessible financial disclosure platforms. Quarterly financial ratios of Bank Rakyat Indonesia (BBRI) are obtained from the official BBRI corporate website and Indonesian Stock Exchange (IDX) publications, while historical stock prices are obtained from publicly available market data providers.

These datasets are stored in structured CSV files to facilitate computational processing. This stage prepares all required data in a consistent and reproducible digital format and is described in detail in Section III-B.

2.) PCA Algorithm

In the second stage, the prepared dataset is transformed into a standardized multivariate matrix and processed using the Principal Component Analysis method described in Section II-D. PCA is applied to extract orthogonal principal components that summarize the covariance structure between financial ratios and stock price movement variables.

This stage produces quantitative outputs including eigenvalues, eigenvectors, explained variance ratios, and principal component scores. The detailed PCA procedure is presented in Section III-C.

3.) Result (Dependence Interpretation)

In the final stage, the extracted principal components are analyzed to identify dominant dependence patterns between financial ratios and stock price movements. The principal component loadings and score distributions are examined to reveal latent financial market coupling factors.

The resulting variance structures, loading matrices, and component score behaviors constitute the main analytical results of this study and are presented and discussed in Section III-D.

B. Data Preparation

The dataset used in this study consist of two primary data sources, including quarterly financial ratios of Bank Rakyat Indonesia (BRI) and historical monthly stock closing price of Bank Rakyat Indonesia (BRI).

1.) Financial Ratio

Quarterly financial ratio data were obtained form publicly available financial reports published on the official corporate website of Bank Rakyat Indonesia and the Indonesian Stock Exchange (IDX).

The financial ratios selected for this study are:

- Price to Book Value (PBV)
- Total Assets
- Return on Investment (ROI)
- Return on Equity (ROE)
- Market Capitalization

These variables were chosen to represent BRI’s performance. All quarter financial ratios were stored in a CSV file named ratios_quarterly.csv.

2.) Monthly Stock Price

Historical monthly closing prices of BBRI were obtained from publicly accessible financial market data providers that publish historical stock price records. Monthly closing price was selected as it represents the official market valuation at the end of each month. All monthly stock price data were stored in a CSV file named prices_monthly.csv, containing the monthly date and the corresponding closing price. This file serves as the direct input for computational processing.

However, because financial ratios are reported quarterly, direct use of monthly price data would result in inconsistent observation units. Therefore, a temporal aggregation process was performed to convert monthly stock prices into quarterly stock movement descriptors.

For each month t, the log return was first computed as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$$

Where P_t denotes the monthly closing price at month t.

Subsequently, for each financial quarter q, the monthly returns belonging to that quarter were aggregated to form quarterly stock movement variables:

- Quarterly Return

$$R_q = \sum_{t \in q} r_t$$

- Quarterly Volatility

$$\sigma_q = \sqrt{\frac{1}{|q|}\sum_{t\in q}\left(r_t - \bar{r_q}\right)^2}$$

- Quarterly Momentum

$$M_q = \frac{P_{\mathrm{end},q}}{P_{\mathrm{start},q}} - 1$$

- Quarterly Maximum Drawdown

$$MDDq = min(PqpeakPt - 1)$$

These returns are then aggregated by financial quarter to produce quarterly market movement variables, including quarterly return, volatility, momentum, and maximum drawdown. These variables summarize market dynamics within each quarter and enable consistent linear dependence analysis between market behavior and financial ratios.

The resulting quarterly stock movement descriptors are merged with the quarterly financial ratios to construct a unified multivariate dataset that serves as the input matrix for PCA.

The implementation of loading csv file and transform it so it can be processed by PCA is written in source code. Because the function take a lot of line, therefore we dont write it here.

```
# 1) Load ratios
ratios = load_ratios_flexible(args.ratios)

# 2) Load prices (optional) and compute quarterly movement features
prices = load_prices_monthly_optional(args.prices)
if prices is None:
    print("[INFO] Monthly price file not provided or not valid. Running PCA on ratios-only features.")
    dataset = ratios.copy()
    qprice = None
else:
    qprice = compute_quarterly_price_features(prices)

    # Merge (inner join to keep only quarters that exist in both)
    dataset = ratios.merge(qprice, on="Quarter", how="inner").sort_values("Quarter").reset_index(drop=True)
    if dataset.empty:
        raise ValueError(
            "Merge result is empty. Your ratios quarters do not overlap with price quarters.\n"
            "Check the date ranges and quarter labeling."
        )

    # Save derived quarterly price features for transparency
    qprice_out = qprice.copy()
    qprice_out["Quarter"] = qprice_out["Quarter"].astype(str)
    qprice_out.to_csv(os.path.join(args.outdir, "quarterly_price_features.csv"), index=False)
```

Then, a loading function is implemented in the main function to load csv file into the program. Now, the data is prepared to be processed by PCA Algorithm.

C. PCA Algorithm

In the program, First of all, we calculate the covariance matrix, eigenvalue, and eigenvectors in order to gain principal components.

```
# PCA
def run_pca(df: pd.DataFrame, feature_cols: List[str], n_components: Optional[int]):
    X = df[feature_cols].to_numpy(dtype=float)

    scaler = StandardScaler()
    Xs = scaler.fit_transform(X)

    pca = PCA(n_components=n_components)
    scores = pca.fit_transform(Xs)

    pcs = [f"PC{i+1}" for i in range(pca.n_components_)]

    loadings_df = pd.DataFrame(pca.components_.T, index=feature_cols, columns=pcs)
    evr_df = pd.DataFrame({"EVR": pca.explained_variance_ratio_}, index=pcs)
    evr_df["CumulativeEVR"] = evr_df["EVR"].cumsum()

    scores_df = pd.DataFrame(scores, columns=pcs)
    scores_df.insert(0, "Quarter", df["Quarter"].astype(str).values)

    return scaler, pca, evr_df, loadings_df, scores_df
```

Function run_pca contains a PCA Algorithm that will give some principal components from the dataset given. It will tells a correlation between quarter stock prices of BRI and quarter financial ratio of BRI from their financial reports.

D. Result

The program give five Principal Components value stored in pca_explained_variance.csv, and pca_loadings.csv. It also ordered by value from highest to lowest.

```
,EVR,CumulativeEVR
PC1,0.5212358538345199,0.5212358538345199
PC2,0.30231543394657867,0.8235512877810985
PC3,0.16512718287177636,0.9886784706528748
PC4,0.011065363170915644,0.9997438338237905
PC5,0.0002561661762094466,1.0
```

From pca_explained_variance, we know that PC1 and PC2 cover majority of the dataset. At the other hand, PC4 and PC5 have low impact due to low percentage.

```
,PC1,PC2,PC3,PC4,PC5
PBV,-0.06685755712602647,0.7983712351974056,-0.0944722495584277,-0.561838295133577,0.18315611387358666
TotalAsset,0.41365826726137,-0.20957293691543727,0.7609386297354748,-0.4154656337847048,0.18255610359319638
ROI,0.5706256375241744,-0.062442970356211276,-0.417369050141721,0.13870602374577728,0.6906887665321162
ROE,0.5778200322506287,-0.08563506635214835,-0.368247100109821,-0.3371423870809975,-0.6399372858607156
MarketCap,0.4061124193021832,0.5544817993279517,0.31975700943712354,0.6154849202663143,-0.21576933272168283
```

From pca_loadings.csv, the result tells that every variables have a correlation (linear dependence). ROI and ROE have a strong positive correlation. Total assets and market capitalization have a medium positive correlation. Finally, PBV has a little negative correlation.

IV. ANALYSIS

From the result we see on section III, the PCA results can be interpreted directly in terms of how much structure (shared linear variation) exists inside the combined dataset and which variables dominate that structure. The explained-variance report indicates that the first principal components, especially PC1 and PC2, capture most of the total variance, while the later components (PC4 and PC5) contribute only marginally. This means that the information contained in the original multivariate data can be represented well in a much lower dimensional space, and that the dominant relationships among variables are concentrated in the first two orthogonal directions. In other words, the dataset is not “randomly scattered” across all dimensions. instead, it has a strong low dimensional linear structure, which is exactly the type of structure PCA is designed to reveal.

The loading matrix further clarifies what that low dimensional structure represents. The strongest pattern observed is a clear positive alignment between ROI and ROE, meaning both variables contribute in the same direction with relatively large magnitude. Interpreted statistically, this indicates that ROI and ROE share substantial co-variation, so they are not independent descriptors of BRI’s performance in the observed period. Rather, they behave as a coupled profitability related signal. A second notable pattern is the medium positive alignment between Total Assets and Market Capitalization, which suggests that variations in firm scale and market value are also partially driven by a shared underlying dimension. In contrast, PBV contributes weakly and with a negative sign relative to those dominant directions, implying that its variation is not synchronized with the main profitability/size structure captured by the early principal components. This does not mean PBV is irrelevant, but it does indicate that PBV is not a primary driver of the dominant covariance structure during the sample window, and that its behavior may be influenced by other dynamics not captured strongly by the principal components explaining most variance.

Taken together, these findings support the core claim of this study: the variables included in the analysis exhibit meaningful linear dependence patterns, as shown by their grouped behavior under PCA. ROI–ROE emerges as the strongest coupled pair, Total Assets–Market Capitalization forms a second co-moving pair, and PBV behaves as a weaker, partially opposing valuation-related signal. Consequently, BRI's multivariate profile in this dataset can be summarized by a small number of latent linear factors rather than treated as many unrelated indicators, and the presence of such concentrated variance and aligned loadings provides empirical evidence that the financial-ratio variables are structured by shared linear relationships rather than independent fluctuations.

V. Conclusion

Based on the results of our analysis, it can be concluded that all examined variables exhibit linear dependence with stock price movements. However, the strength of these correlations varies across variables.

VI. Appendix

Github of the program can be accesed from the following link.
https://github.com/leains/algeo-makalah

Video can be accesed from the following link.
https://youtu.be/ljhZGB_lDHM?si=JMBu6-h_rfWl3wO4

VII. Acknowledgment

The author is deeply thank you to God, Mr. Arrival, and others that support author to make this paper.

References

[1] GeeksforGeeks, "Data Analysis: Principal Component Analysis (PCA)." [Online]. Available: https://www.geeksforgeeks.org/data-analysis/principal-component-analysis-pca/. Accessed: Dec. 24, 2025.

[2] GeeksforGeeks, "Covariance Matrix." [Online]. Available: http://geeksforgeeks.org/maths/covariance-matrix/. Accessed: Dec. 24, 2025.

[3] G. Strang, "6.1 Eigenvalues and Eigenvectors," Introduction to Linear Algebra (MIT), PDF lecture notes. [Online]. Available: https://math.mit.edu/~gs/linearalgebra/ila5/linearalgebra5_6-1.pdf. Accessed: Dec. 24, 2025.

[4] Stanford University, "Ch. 4: Linear Algebra," NBIO 228 Course Handout, PDF lecture notes. [Online]. Available: https://web.stanford.edu/class/nbio228-01/handouts/Ch4_Linear_Algebra.pdf. Accessed: Dec. 24, 2025.

[5] ReitData.io, "Data Analysis: The Relationship Between Correlation and PCA (Principal Component Analysis)," *Medium*. [Online]. Available: https://medium.com/@reitdata.io/data-analysis-the-relationship-between-correlation-and-pca-principal-component-analysis-3db5f1430a7b. Accessed: Dec. 24, 2025.

[6] Corporate Finance Institute, "Financial Ratios: Definition, Types, and Examples." Corporate Finance Institute, Oct. 23, 2025. [Online]. Available: https://corporatefinanceinstitute.com/resources/accounting/financial-ratios/. Accessed: Dec. 24, 2025.

DECLARATION

I hereby declare that this paper is my original work and has not been translated, adapted, or copied from any other source. All sources used in this study have been properly cited, and this paper is free from plagiarism.

Bandung, 24 Desember 2025

Leonardus Brain Fatolosja 13524146