Risk Analysis and Portfolio Management Using Principal Component Analysis: A Study on Indonesia's IDX30 Stock Index

Nathaniel Jonathan Rusli - 13523013[1,2]
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
[1]13523013@std.stei.itb.ac.id, [2]omgitsnathaniels@gmail.com

***Abstract*—This paper explores the application of Principal Component Analysis (PCA) in risk analysis and portfolio management within the IDX30 stock index. PCA effectively identifies risk factors and quantifies their impact on individual stocks, offering a detailed understanding of systematic and idiosyncratic risks. By analyzing stock loadings and constructing portfolios based on principal components, PCA demonstrates its value in creating diversified portfolios and managing risk exposure. The study evaluates various portfolio strategies, including those based on top principal components, cumulative variance thresholds, and custom weighting schemes. Results reveal that while PCA is a critical tool for assessing risks and uncovering diversification opportunities, it does not guarantee superior performance. The paper emphasizes the importance of complementing PCA insights with economic and strategic considerations to achieve tailored investment goals, balancing risk and return.**

***Keywords*—IDX30 stock index, Portfolio management, Principal component analysis (PCA), Risk analysis**

I. Introduction

In modern finance, the quest for effective risk analysis and optimal portfolio management remains a sustaining endeavor. Analyzing large datasets, such as stock indices like Indonesia's IDX30, presents challenges due to high asset correlations, which can obscure critical insights. Whereas Markowitz's mean-variance theory paves the foundation of modern portfolio theory, its practical application is hindered by its high sensitivity to input assumptions, often leading to suboptimal diversification. The modern paradigm of risk-based allocation strategy, which constructs a portfolio based solely on the variance-covariance of assets, such as minimum variance, most diversified portfolio, risk parity, and diversified risk parity, neglects to effectively handle datasets with a large number of highly correlated assets [1].

Principal Component Analysis (PCA) offers a robust mathematical approach to tackle such challenges. By transforming high-dimensional data into a smaller set of uncorrelated components, PCA captures the most significant sources of variance within the dataset, highlighting vital factors for smarter investment decisions [2]. The IDX30 index comprises the top 30 stocks on the Indonesia Stock Exchange, selected based on relatively large market capitalization, high liquidity, and strong fundamentals [3]. Its composition reflects various sectors, making it an ideal candidate for risk analysis and portfolio management using PCA.

This paper aims to explore the application of PCA on the IDX30 stock index to uncover key risk factors and optimize stock portfolio construction. By leveraging PCA's capability to reduce dimensionality and isolate critical sources of variance, the study provides actionable insights for risk analysis and portfolio management. The results are expected to contribute to more effective investment strategies, particularly in emerging markets.

II. Literature Review

A. Variance and Covariance

Variance and covariance are two fundamental concepts in statistics that measure different aspects of data distribution and relationships between variables. Both concepts will be extensively explored throughout this study, as they form the mathematical foundation for techniques such as PCA.

1) Variance: Quantifies how much a set of data points deviate from their mean value. It provides a measure of spread or dispersion within a single dataset. In the context of PCA, higher variance indicates that a component contains more information about the dataset. Variance (σ) is calculated using the formula:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \qquad (1)$$

where N is the number of data points, x_i represents each data point, and μ is the mean of the data points.

2) Covariance: Measures how two random variables change together, indicating the direction of the linear relationship between the variables. The covariance of two variables X and Y is calculated using the formula:

$$Cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) \tag{2}$$

where N is the number of the paired data points, x_i and y_i are individual sample points, $\bar{x}$ and $\bar{y}$ are the means of X and Y, respectively. Covariance can be held in a matrix called the covariance matrix [4].

B. Eigenvalues and Eigenvectors

1) Definition: Eigenvalues are scalar values that indicate how much an eigenvector is stretched or compressed during a transformation. Formally, if A is a square matrix and v is an eigenvector, then the relationship can be expressed as follows:

$$Av = \lambda v \tag{3}$$

Eigenvectors are non-zero vectors that, when transformed by a matrix, result in a vector that is a scalar multiple of the original vector, meaning the direction of the eigenvector remains unchanged (or reversed if the eigenvalue is negative) after the transformation [5], [6].

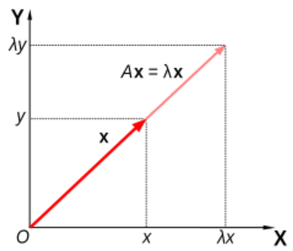

Fig. 1. Eigenvalues and eigenvectors visualization [Source: https://en.wikipedia.org/wiki/File:Eigenvalue_equation.svg, Accessed: Dec. 29, 2024.]

2) Eigenvalues and Eigenvectors in PCA: Eigenvalues and eigenvectors are often utilized in the domain of machine learning, particularly in techniques like PCA, where they play pivotal roles in achieving dimensionality reduction.

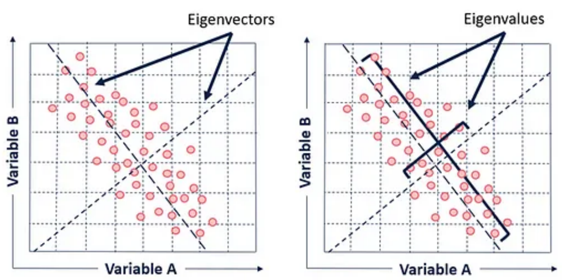

Fig. 2. Visualization of eigenvectors and eigenvalues as PC [Source: https://medium.com/intro-to-artificial-intelligence/principal-component-analysis-pca-cd282196b7d5, Accessed: Dec. 29, 2024.]

Eigenvalues quantify the amount of variance captured by each principal component (PC). Larger eigenvalues indicate a greater portion of the data's variance of the corresponding eigenvector. Eigenvectors represent the directions of these principal components in the transformed space. Each eigenvector defines a new axis, aligning with the directions of maximum variance. Thus, eigenvectors can be visually interpreted as vectors that point in the direction of maximum variance. By sorting eigenvalues in a descending order, the most informative eigenvectors can be prioritized, allowing the reduction of noisy dimensions in the dataset [6].

3) Calculation: Eigenvalues and the corresponding eigenvectors of a matrix A can be found by following these simple steps. In order to find the eigenvalues, compute the characteristic polynomial which is obtained from the determinant equation:

$$det(A - \lambda I) = 0 \tag{4}$$

Eigenvectors then can be calculated by substituting each eigenvalue (λ) back into the equation:

$$(A - \lambda I)v = 0 \tag{5}$$

To note, not all matrices A have corresponding eigenvalues and eigenvectors, as these exist only if A is a square matrix and satisfies the characteristic equation as shown in (4) [5].

C. Singular Value Decomposition (SVD)

1) Definition: Singular Value Decomposition (SVD) decomposes a matrix A into three other matrices:

$$A = U\,S\,V^T \tag{6}$$

where U contains left singular vectors (u_i), S or Σ is a diagonal matrix of singular values (σ_i), and V^T contains right singular vectors (v_i).

The matrix AA^T and A^TA are very notable since these matrices are symmetrical, square, positive semidefinite, have the same positive eigenvalues, and both have the same rank r as A. Eigenvectors for AA^T are named as u_i and A^TA as v_i. Left singular vectors (u_i) can then be concatenated into U and right singular vectors (v_i) into V, forming orthogonal matrices with each eigenvector being orthonormal. Since eigenvectors can be differently ordered to produce U and V, the eigenvectors are ordered such that vectors higher eigenvalues come before those with smaller values. In comparison to eigendecomposition, SVD works on both square and non-square matrices [7], [8].

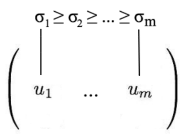

Fig. 4. Standardization of the eigenvectors order of U and V [Source: https://jonathan-hui.medium.com/machine-learning-singular-value-decomposition-svd-principal-component-analysis-pca-1d45e885e49, Accessed: Dec. 29, 2024.]

The SVD decomposition can also be recognized as a series of outer products of u_i and v_i:

$$A = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T \quad (7)$$

2) SVD Connection to PCA: The reduced form of SVD decomposition as shown in (7) is crucial to understand the components of A, offering a powerful method to decompose an m x n matrix of entangled data into r components. Since u_i and v_i are unit vectors, we can ignore terms $u_i\sigma_i v_i^T$ with very small singular values (σ_i). If data is highly correlated, many singular values (σ_i) can be expected to be small and ignored [8].

Just like PCA, SVD can be used for dimensionality reduction by selecting a few singular values and their corresponding vectors, effectively capturing the most significant features (with highest variance) of the data while reducing noise [8]. But in practical applications, the right singular vectors (V) correspond to the principal components derived in PCA.

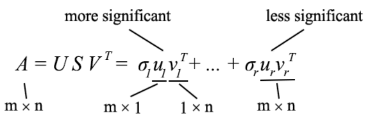

Fig. 5. SVD decomposition and its reduced form [Source:https://jonathan-hui.medium.com/machine-learning-singular-value-decomposition-svd-principal-component-analysis-pca-1d45e885e49, Accessed: Dec. 29, 2024.]

D. Principal Component Analysis (PCA)

1) Definition: Principal Component Analysis (PCA) is a statistical technique primarily used for dimensionality reduction while preserving as much variance (information) as possible derived from the dataset. This method is particularly useful when dealing with high-dimensional data, allowing for easier analysis and visualization.

PCA projects m-dimensional input features onto a k-dimensional space defined by k principal components. representing latent factors. PCA reorients the axes of a dataset to align with directions of maximum variance. The new coordinate system lines up with the eigenvectors, making correlations and patterns more apparent [6].

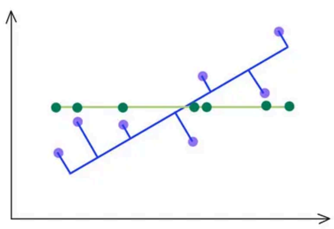

Fig. 6. Principle components reorientation for maximum variance [Source:https://jonathan-hui.medium.com/machine-learning-singular-value-decomposition-svd-principal-component-analysis-pca-1d45e885e49, Accessed: Dec. 29, 2024.]

2) Computation Methods: PCA can be computed through various methods. Traditionally, it is performed through Eigenvalue Decomposition (EVD) on the sample variance matrix, providing a method based on EVD. Alternatively, SVD is another approach that has demonstrated advantages. SVD is particularly efficient for high-dimensional datasets, offering greater computational speed, numerical speed, and stability [9]. Mature machine learning libraries, such as Scikit-learn, also implement the SVD method for their PCA computation, as outlined in their documentation [10].

Utilizing the SVD method, PCA can be computed as follows:

a) Center the Data: Subtract the mean of each feature from the dataset to ensure the data is zero-centered:

$$Z = X - \mu \quad (8)$$

where Z is the centered data matrix, X is the original data matrix, and μ is the mean vector of features.

b) Standardization (Optional): If the features have varying units or scales, standardize them by dividing each feature by their standard deviations (σ):

$$Z_{standardized} = \frac{Z}{\sigma} \quad (9)$$

c) Compute SVD: Apply SVD to the centered (and possibly standardized) data matrix Z:

$$Z = U S V^T \quad (10)$$

d) Calculate Variance Explained (Optional): Divide the square of each singular value (derived from the diagonal of S) by the sum of all squared singular values to determine the proportion of variance explained by each principal component:

$$Variance\ Explained\ by\ PC_i = \frac{\sigma_i^2}{\Sigma \sigma_j^2} \quad (11)$$

e) Extract Principal Components: Choose the top k singular values and corresponding columns of V^T or rows of V (principal components) based on explained variance.

f) Transform Data: Project the original data onto the reduced principal component:

$$Z_{transformed} = Z V_k^T \quad (12)$$

where V_k^T consists of the first k columns of V^T, corresponding to the top k principal components.

E. Risk Analysis Using PCA

Stocks are commonly influenced by three main types of risks: systematic, sectoral, and idiosyncratic risks. Systematic risk refers to market-wide factors, such as economic recessions, geopolitical events, or even macroeconomics changes like fluctuations in interest rates. Systematic risk affects all stocks in varying degrees and cannot be diversified away. Sectoral risk is specific to particular industries or sectors, such as changes in commodity prices affecting energy stocks. Idiosyncratic risk, on the other hand, is unique to individual companies, such as management decisions, product success or failures, or even lawsuits. Unlike systematic risk, idiosyncratic risk can typically be reduced and diversified away.

PCA offers valuable insights into which stocks are most exposed to systematic risks by analyzing each stock's contribution to dominant principal components. Stocks with high loading values in the first principal component (PC1) tend to have a significant impact on overall market risk. On the other hand, subsequent principal components highlight diversification opportunities by identifying assets with weaker correlations to the broader market.

Highly correlated datasets often require only one or two PCs to capture a large portion of the total variation. Conversely, if the data is less correlated or only correlated within certain subgroups, more PCs may be necessary. Examining the variance explained by successive PCs provides a way to estimate the effective dimensionality of the dataset.

The explained variance of each component also allows investors to assess how much of the total market risk is attributed to systematic versus idiosyncratic factors. For example, a sharp decline in explained variance after PC1 indicates that most of the risk is market-driven, while a gradual decline suggests a combination of systematic and specific (sectoral or idiosyncratic) risks.

Whereas this reduction in dimensionality is helpful, it does not provide an economic interpretation of the PCs. Specifically, it remains unclear whether the PCs are influenced by all variables in the dataset or just a subset, requiring further investigation to assign meaningful economic interpretations to the components [11].

F. Portfolio Management Using PCA

PCA offers a powerful framework for optimizing portfolio management and constructing eigenportfolios. One key application of PCA in portfolio optimization is risk-return optimization. By allocating portfolio weights based on the variance contribution of principal components, investors can enhance risk-adjusted returns. For instance, assigning lower weights to components with high exposure to systematic risk, such as PC1, and prioritizing diversifying components like PC2 and PC3, allows for a greater stability and more favorable risk-return profile.

Another critical application of PCA is weight optimization. Using the loadings of each stock in principal components, investors can determine factor-based weights that align with expected return and desired risk exposure. Stocks exhibiting higher sensitivity to beneficial principal components, such as those driving long term growth or sectoral outperformance, can be assigned greater weights. Additionally, traditional techniques, such as mean-variance optimization (MVO) and Sharpe ratio calculation can be combined with PCA by incorporating expected returns derived from principal components.

It is clear that PCA supports diversification by identifying uncorrelated components within the dataset. Picking stocks that load significantly on different principal components ensures reduced exposure to specific risks associated with correlated assets. This enhances portfolio resilience and mitigates the impact of adverse events affecting particular sectors or market trends. However, PCA is less of a suitable method for determining how many individual stocks to retain in a portfolio. This limitation arises from PCA's reliance on the covariance matrix [12].

To tackle challenges associated with noisy data in the covariance matrix and build a more robust eigenportfolios, techniques such as spectral cut-off and spectral selection have been introduced. Spectral cut-off involves discarding smaller eigenvalues, retaining only significant eigenvalues and their corresponding eigenvectors for portfolio construction. Spectral selection, on the other hand, offers a more precise approach by retaining eigenvalues and eigenvectors that hold economic or strategic importance. For instance, investors might prioritize components associated with systematic or sectoral risks, discarding those linked to idiosyncratic noise [13].

G. IDX30 Stock Index

The IDX30 stock index, a key benchmark on the Indonesia Stock Exchange (IDX), comprises the top 30 stocks selected based on market capitalization, liquidity, and fundamental performance from the LQ45 stock index. It serves as a vital indicator towards the Indonesian market's health, reflecting the performances of its most prominent companies. With representation across diverse sectors such as finance, consumer goods, energy, and telecommunications, the IDX30 offers a comprehensive view of Indonesia's economic landscape [3].

TOP 10 CONSTITUENTS

NO.	CODE	COMPANY NAME	FF MC (IDR)	INDEX WEIGHT	SECTOR
1.	BBRI	Bank Rakyat Indonesia (Persero) Tbk.	258.63 T	16.01%	Financials
2.	BMRI	Bank Mandiri (Persero) Tbk.	245.35 T	15.19%	Financials
3.	BBCA	Bank Central Asia Tbk.	244.81 T	15.16%	Financials
4.	TLKM	Telkom Indonesia (Persero) Tbk.	187.55 T	11.61%	Infrastructures
5.	ASII	Astra International Tbk.	93.55 T	5.79%	Industrials
6.	BBNI	Bank Negara Indonesia (Persero) Tbk.	84.57 T	5.24%	Financials
7.	GOTO	GoTo Gojek Tokopedia Tbk.	75.03 T	4.65%	Technology
8.	AMRT	Sumber Alfaria Trijaya Tbk.	49.94 T	3.09%	Consumer Non-Cyclicals
9.	UNTR	United Tractors Tbk.	32.37 T	2.00%	Industrials
10.	ADRO	Adaro Energy Indonesia Tbk.	30.91 T	1.91%	Energy
		Total 10	1,302.70 T	80.66%	

Fig. 7. Sector weights and point index of IDX30 per November 2024 [Source: IDX Index Fact Sheet IDX30 November 2024, Accessed: Dec. 29, 2024]

As an index in an emerging market, the IDX30 reflects unique economic dynamics. These characteristics pose both challenges, such as difficulty in uncovering independent risk factors, and opportunities, such as testing advanced techniques like PCA to uncover vital risk factors. The IDX30's manageable size and relevance to investors in both emerging and developed markets further justify its selection as the dataset for this study.

III. Implementation

A. Programming Language and Supporting Tools

For the implementation of risk analysis and portfolio management using PCA, Python is used as the programming language, leveraging its extensive ecosystem of supporting libraries. These libraries streamline the process of data collection, processing, analysis, and visualization:

```
import yfinance as yf
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

The yfinance library is used for fetching historical stock price data directly from Yahoo Finance, simplifying the process of downloading and organizing market data for multiple stocks. The numpy library is used for more efficient numerical computations, matrix operations, and statistical processing, including the computation of SVD. The pandas library is used for data manipulation, cleaning, and organization of stock price and return data. The matplotlib library is used to create comprehensive visualizations of key results, enhancing the interpretability of data.

B. Data Collection and Preprocessing

The PCA process will be applied to the daily returns of IDX30 stocks. The historical price data spans from November 2024 to the end of December 2024, aligning with the IDX30 stock composition update in November 2024. However, the timeframe of the historical price data can vary depending on the goals of the analysis. For instance, the November to December 2024 data emphasizes the most recent market dynamics and stock relationships, providing insights into the current risk and diversification landscape. On the other hand, a longer backtracking period, such as January 2024 to December 2024, may capture broader market trends and performance over time, offering a more comprehensive perspective on risk and portfolio behavior.

The fetched historical price data is then converted into daily returns and cleaned by removing any incomplete price data to ensure consistency across all stocks and dates:

```
idx30_tickers = [
    "BBCA.JK", "BBRI.JK", "BMRI.JK",
    "TLKM.JK", "ASII.JK", "BBNI.JK",
    "ICBP.JK", "AMRT.JK", "UNTR.JK",
    "GOTO.JK", "BRPT.JK", "CPIN.JK",
    "ADRO.JK", "UNVR.JK", "INDF.JK",
    "KLBF.JK", "MBMA.JK", "PGEO.JK",
    "MDKA.JK", "PGAS.JK", "INKP.JK",
    "INCO.JK", "ANTM.JK", "ARTO.JK",
    "PTBA.JK", "MEDC.JK", "MAPI.JK",
    "SMGR.JK", "AKRA.JK", "ACES.JK"
]

data = yf.download(idx30_tickers,
start="2024-11-1", end="2024-12-30")
prices = data["Close"].dropna()
returns = prices.pct_change().dropna()
```

C. Principal Component Analysis Implementation

As the central focus of this analysis, PCA is implemented through a function that utilizes the SVD method. This function not only computes and returns the principal components, but also outputs the explained variance ratio, singular values, and transformed data, providing supporting data for further analyses:

```
def svdPCA(X, n_components=None):
    X_mean = np.mean(X, axis=0)
    X_centered = X - X_mean
    U, S, Vt = np.linalg.svd(X_centered,
full_matrices=False)

    if n_components is None:
        n_components = min(X.shape)

    components = Vt[:n_components]
    for i in range(components.shape[0]):
        max_abs_idx =
np.argmax(np.abs(components[i]))
        if components[i, max_abs_idx] <
0:
            components[i] *= -1
            U[:, i] *= -1

    singular_values = S[:n_components]
    explained_variance =
(S[:n_components] ** 2) / (X.shape[0] -
1)
    total_var = (S ** 2).sum() /
(X.shape[0] - 1)
    explained_variance_ratio =
explained_variance / total_var
    transformed_X = np.dot(X_centered,
components.T)

    return transformed_X, components,
explained_variance_ratio, singular_values
```

D. Principal Components Risk Analysis

To analyze risks through principal components, PCA must first be applied to the historical stock returns data. The historical stock returns serve as an essential input for PCA, as they capture day-to-day variations in stock prices, reflecting the underlying market dynamics and volatility. By utilizing the all relevant informations derived from the PCA computation, further analyses, such as principal components stock loadings calculation, can be performed:

```
# Perform PCA on historical returns data
transformed_returns, components,
explained_variance_ratio, singular_values
= svdPCA(returns.values)
```

E. Principal Components Portfolio Management

The principal components portfolio management is done by constructing and analyzing portfolios using weights derived from principal components to evaluate risk-return dynamics effectively. These varying portfolio compositions include the PC1 portfolio (reflects systematic risk), Top 3 PC, Top 5 PC, and Top 10 PC portfolios (combines multiple components to incorporate sector-specific and idiosyncratic risks), and the 95% Variance portfolio (heavily diversifies by including all components that explain 95% of the variance), Additionally, Equal Weight portfolio, allocating equal weights to all stocks, is included as a benchmark and Custom Weight portfolio, allocating real IDX30 stocks weight, is included to reflect real-world considerations. Portfolio returns are calculated by applying respective weights to stock returns and cumulative returns are derived to envision long-term portfolio performance:

```
# Number of components explaining 95%
variance
num_components =
np.argmax(cumulative_variance >= 0.95) +
1

# Portfolio Weights Construction
num_components_95 =
np.argmax(np.cumsum(explained_variance_ra
tio) >= 0.95) + 1
portfolio_weights = {
    'PC1': components[0] /
np.sum(np.abs(components[0])),
    'Top_3_PC': np.sum(components[:3],
axis=0) /
np.sum(np.abs(np.sum(components[:3],
axis=0))),
    'Top_5_PC': np.sum(components[:5],
axis=0) /
np.sum(np.abs(np.sum(components[:5],
axis=0))),
    'Top_10_PC': np.sum(components[:10],
axis=0) /
np.sum(np.abs(np.sum(components[:10],
axis=0))),
    '95_Variance':
np.sum(components[:num_components_95],
axis=0) /
np.sum(np.abs(np.sum(components[:num_comp
onents_95], axis=0))),
    'Equal_Weight':
np.ones(len(idx30_tickers)) /
len(idx30_tickers),
    'Custom_Weighted':
np.array([custom_portfolio_weights[ticker
] for ticker in idx30_tickers])
}

# Portfolio Returns Calculation
portfolio_returns = {}
for name, weights in
portfolio_weights.items():
    portfolio_returns[name] =
np.dot(returns.values, weights)

# Convert Portfolio Returns to Cumulative
Returns
cumulative_returns = {name: np.cumprod(1
+ ret) for name, ret in
portfolio_returns.items()}
cumulative_returns = {name: cum_ret - 1
for name, cum_ret in
cumulative_returns.items()}
```

IV. Results and Discussion

A. The IDX30 Stock Index Characteristics

Using the historical data following the most recent update of the IDX30 composition, the analysis of the historical data following the latest IDX30 composition reallocation reveals that 16 principal components (PCs) are required to retain 95% of the variance in the dataset. This relatively high number of PCs indicates that the IDX30 stock index exhibits a complex structure with significant variability distributed across multiple dimensions.

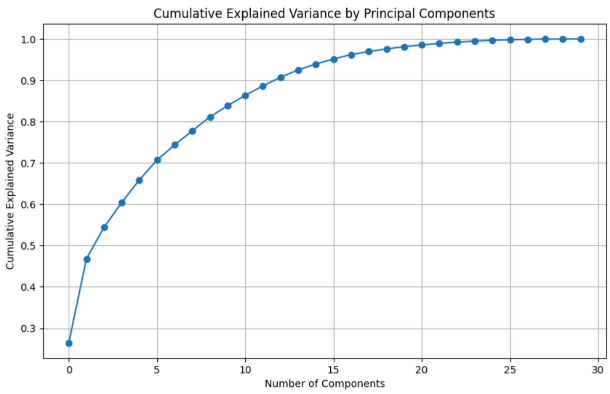

Fig. 8. Cumulative explained variance of the IDX30 returns data [Source: Author's calculation using PCA on IDX30 historical returns]

The necessity for 16 PCs to explain 95% of the variance can be attributed to the multi-sectoral nature of the IDX30. Stocks in different sectors tend to exhibit unique patterns of returns driven by distinct economic factors, such as interest rates, commodity prices, or consumer demand. For example, the financial sector may respond to monetary policy shifts, whereas the energy sector is more influenced by global oil prices. This heterogeneity increases the dimensionality of the dataset, as no single or small group of PCs can capture the diverse factors affecting all sectors. The implication is that risk analysis and portfolio management strategies must account for this complexity by considering contributions from multiple PCs, rather than focusing solely on the dominant ones

B. Risk Analysis

By analyzing the stock loadings of each principal component (PC), the contributions of the IDX30 stocks towards various factors can be understood.

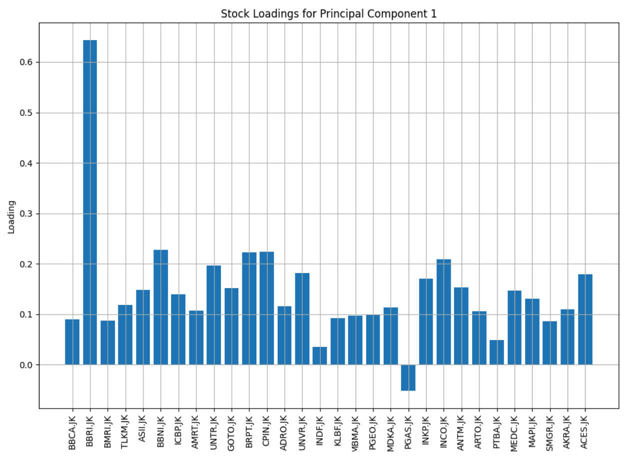

Fig. 9. IDX30 stock loadings for the first principal component (PC1) [Source: Author's calculation using stocks coefficients of the corresponding principal component]

Using the PC1 stock loadings data illustrated in Fig. 9, the stocks are categorized based on their loadings levels, as detailed in Table I. Column (a) represents the stock loading categories, whereas column (b) lists the corresponding tickers within each category.

TABLE I
LIST OF STOCK TICKERS GROUPED BY THEIR FIRST PRINCIPAL COMPONENT LOADINGS

Stock Loadings[a]	*Stock Tickers*[b]
High-loading (> 20%)	BBRI.JK, BBNI.JK, BRPT.JK, CPIN.JK, INCO.JK
Moderate-loading (10-20%)	TLKM.JK, ASII.JK, ICBP.JK, AMRT.JK, GOTO.KJ, ADRO.JK, UNVR.JK, MDKA.JK, INKP.JK, ANTM.JK, ARTO.JK, MEDC.JK, MAPI.JK, AKRA.JK, ACES.JK
Low-loading (< 10%)	BBCA.JK, BMRI.JK, INDF.JK, KLBF.JK, MBMA.JK, PGEO.JK, PGAS.JK, PTBA.JK, SMGR.JK

The first principal component (PC1) captures systematic market-wide risk, reflected in the significant loadings of most stocks. Stocks with high PC1 loadings are more exposed to market-driven risks, making them key contributors to portfolio variance and performance, especially during economic turbulence. Investors exposed to these high-loading stocks should be aware of their higher influence on the portfolio's total risk.

In contrast, stocks with moderate or low loadings are less sensitive to market-wide factors, often reflecting idiosyncratic or sector-specific risks. These lower-loading stocks can enhance portfolio diversification. A balanced portfolio combining both high and low PC1 loadings leverages their complementary characteristics for stability and diversification. To note, a negative stock loadings signifies that the stock moves inversely to the direction of the principal component.

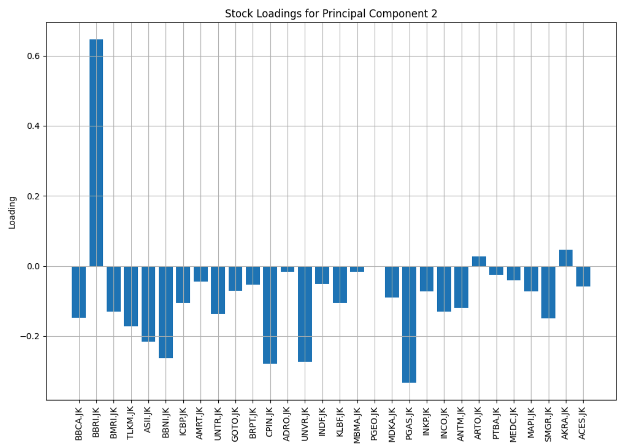

Fig. 10. IDX30 stock loadings for principal component 2 (PC2) [Source: Author's calculation using stocks coefficients of the corresponding principal component]

Examining PC2 and subsequent components (e.g., PC3 and beyond) reveals increasingly nuanced and orthogonal aspects of variance in the IDX30. While PC1 highlights overall market trends, subsequent components capture sectoral contrasts or thematic behaviors, reflecting interactions among specific industry groups. Subsequent components may offer granular insights into the dataset's structure and unique dimensions of risk and opportunity.

For instance, as illustrated in Fig. 10, PC2 loadings vary significantly, with BBRI.JK emerging as a dominant influence. Figuring out the interpretation or the factors related to PC2 may require further extensive analysis, but it could still offer insights into sectoral dynamics or external influences.

C. Portfolio Management

Utilizing principal component stock loadings, portfolio stock weights are calculated. The stock weights in a portfolio can vary significantly depending on the method of portfolio construction as illustrated in Fig. 11 and Fig. 12.

For instance, portfolios based on the first principal component (PC1) tend to assign higher weights to stocks with greater exposure to market-wide systematic risks, as indicated by their loadings on PC1. Conversely, portfolios constructed using multiple principal components, such as the 95% Variance portfolio, covering 95% of the variance, distribute weights more broadly, incorporating both systematic and idiosyncratic factors.

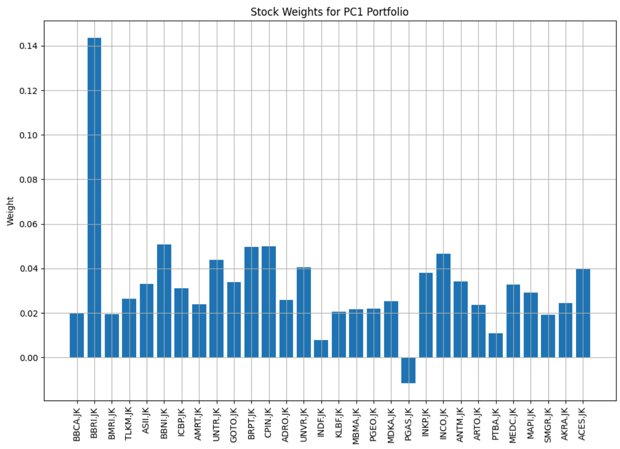

Fig. 11. Stock weights of the PC1 Portfolio [Source: Author's calculation using stock loadings of each principal component]

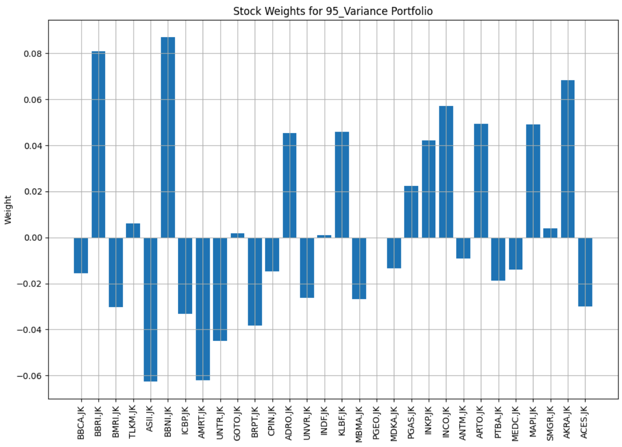

Fig. 12. Stock weights of the 95 Variance Portfolio [Source: Author's calculation using stock loadings of each principal component]

Furthermore, comparing portfolio performances allows for a clear evaluation about the impact of each portfolio weighting strategy. This analysis is essential for identifying the effectiveness of diversification and risk mitigation techniques, helping investors choose a portfolio that aligns with their objectives.

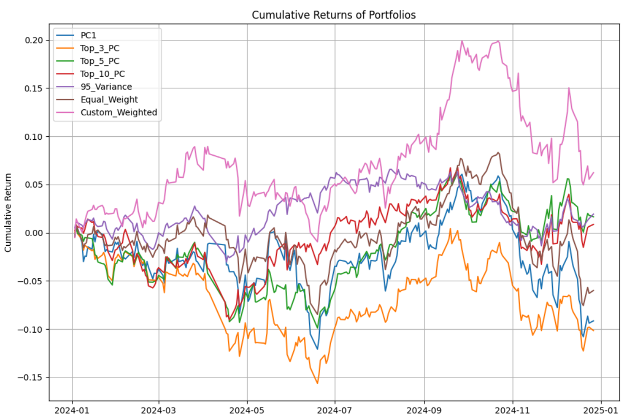

Fig. 13. Cumulative return of constructed portfolios [Source: Author's calculation using each portfolio's stock weight and historical data]

To enhance the accuracy of backtesting, this analysis employs a data span from January to December 2024, allowing for a comprehensive evaluation of cumulative returns among the constructed portfolios. The use of this extended time frame ensures that the performance trends are reflective of varying market conditions over the year.

To evaluate portfolio performance and cumulative returns effectively, standard deviations are calculated. Risk-adjusted returns are then derived by dividing the cumulative returns by their respective standard deviations. All performance metrics for the portfolios are presented in Table II, with column (a) displaying cumulative returns, column (b) detailing standard deviations, and column (c) outlining risk-adjusted returns.

TABLE II
PERFORMANCE OF EACH CONSTRUCTED PORTFOLIO

Portfolio	*Cumulative Return*[a]	*Standard Deviation*[b]	*Risk-Adjusted Return*[c]
PC1	-9%	0.0115	-7.92
Top 3 PC	-10%	0.0101	-10.08
Top 5 PC	2%	0.0089	1.86
Top 10 PC	1%	0.0086	1.00
95% Variance	2%	0.0059	3.27
Equal Weight	-6%	0.0094	-6.39
Custom Weight	6%	0.0105	5.93

As illustrated in Fig. 13 and Table II, portfolio performance varies significantly, offering valuable insights. The PC1 portfolio, heavily influenced by systematic market-wide risks, underperformed with a cumulative return of -9% and a risk-adjusted return of -7.92, reflecting its vulnerability to market downturns due to concentrated exposure to broader trends. Similarly, the Top 3 PC portfolio yielded a cumulative return of -10% and a risk-adjusted return of -10.08, showing that relying solely on the first few principal components can result in overexposure to dominant risk factors and does not guarantee better performance.

In contrast, portfolios incorporating a broader range of components, such as the Top 5 PC and 95% Variance portfolios, achieved positive cumulative returns of 2% each, with risk-adjusted returns of 1.86 and 3.27, respectively. These results highlight the advantage of diversifying across additional principal components, effectively mitigating systematic risks and capturing more nuanced aspects of variance.

The Custom Weight portfolio, designed with predefined stock weightings to favor specific investment priorities, outperformed all others with a cumulative return of 6% and a risk-adjusted return of 5.93, showcasing the benefits of strategic weighting tailored to growth opportunities or undervalued stocks.

Although PCA does not ensure superior performance, it remains a critical tool for risk management and diversification, as evidenced by the decreasing portfolio standard deviation with increasing diversification (see column (b) of Table II). Investors should use PCA to identify and manage risks but complement it with custom weighting strategies to enhance returns. For instance, more aggressive investors may prefer the Custom Weight portfolio for its superior gain potential, while risk-averse investors may choose the 95% Variance portfolio, which balances modest returns with lower volatility.

V. Conclusion

PCA proves to be a powerful tool for uncovering risk factors and uncovering the stocks most impacted among them. Its ability to analyze asset weights within a portfolio provides valuable insights into risk distribution and portfolio composition. By leveraging PCA, investors can effectively construct diversified portfolios with reduced risk exposure. However, whereas PCA excels in identifying risks and diversification opportunities, it does not guarantee superior performance. To achieve optimal results, PCA should be paired with economic or strategic considerations, ensuring that portfolio decisions align with broader investment objectives and market conditions.

VI. Acknowledgment

The author expresses heartfelt gratitude to God Almighty for granting the strength, perseverance, and opportunity to successfully complete this paper. The author also wishes to express profound gratitude to Ir. Rila Mandala, M.Eng., Ph.D., the lecturer for the IF2123 Linear Algebra and Geometry course, for his unwavering guidance and inspiration throughout his time teaching the students.

VII. Appendix

The complete source code used for the analysis, including data preprocessing, PCA implementation, and portfolio management computations, is available on GitHub. Access the code repository here: GitHub Repository Link.

References

[1] L. Yang. "An Application of Principal Component Analysis to Stock Portfolio Management," Master of Commerce in Finance, Department of Economics and Finance University of Canterbury. Accessed: Dec. 28, 2024. [Online]. Available: https://ir.canterbury.ac.nz/items/1bced3e5-abd7-4277-9c97-65c90c215298.

[2] J.R, "Principal Component Analysis for Portfolio Optimization," *MarketBulls,* Jun. 19, 2024. https://market-bulls.com/principal-component-analysis-portfolio-optimization/ (accessed Dec. 29, 2024).

[3] Indonesia Stock Exchange, "IDX Index Fact Sheet IDX30," Nov. 2024.https://www.idx.co.id/id/data-pasar/laporan-statistik/fact-sheet-index/ (accessed: Dec. 29, 2024).

[4] verifiedmetrics, "Covariance vs. Variance: Top differences you should know | Verified Metrics," *www.verifiedmetrics.com*. https://www.verifiedmetrics.com/blog/covariance-vs-variance-top-differences-you-should-know (accessed Dec. 29, 2024).

[5] R. Munir. "Nilai Eigen dan Vektor Eigen (Bagian 1)", 2023. https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2023-2024/Algeo-19-Nilai-Eigen-dan-Vektor-Eigen-Bagian1-2023.pdf (accessed: Dec. 29, 2024).

[6] Dataheadhunters, "Dissecting Eigenvectors: Their Role in Dimensionality Reduction," *Dataheadhunters.com*, Jan. 07, 2024. https://dataheadhunters.com/academy/dissecting-eigenvectors-their-role-in-dimensionality-reduction/ (accessed Dec. 29, 2024).

[7] R. Munir. "Singular Value Decomposition (SVD) (Bagian 1)", 2023.https://informatika.stei.itb.ac.id/~rinaldi.munir/AljabarGeometri/2023-2024/Algeo-21-Singular-value-decomposition-Bagian1-2023.pdf (accessed: Dec. 29, 2024).

[8] J. Hui, "Machine Learning — Singular Value Decomposition (SVD) & Principal Component Analysis (PCA)," *Medium*, Feb. 08,2020.https://jonathan-hui.medium.com/machine-learning-singular-value-decomposition-svd-principal-component-analysis-pca-1d45e885e491 (accessed Dec. 29, 2024).

[9] K. He, "SVD in Machine Learning: PCA," *Medium*, May 18, 2020. https://towardsdatascience.com/svd-in-machine-learning-pca-f25cf9b837ae (accessed Dec. 29, 2024).

[10] scikit-learn, "PCA," *scikit-learn*, 2024. https://scikit-learn.org/1.5/modules/generated/sklearn.decomposition.PCA.html (accessed Dec. 29, 2024).

[11] Mico Loretan, "Generating market risk scenarios using principal components analysis: Methodological and practical considerations," *Research Gate*, Jan. 01, 1997. https://www.researchgate.net/publication/253200764_Generating_market_risk_scenarios_using_principal_components_analysis_Methodological_and_practical_considerations (accessed Dec. 30, 2024).

[12] G. Pasini, "PRINCIPAL COMPONENT ANALYSIS FOR STOCK PORTFOLIO MANAGEMENT," *International Journal of Pure and Apllied Mathematics*, vol. 115, no. 1, Jun. 2017, doi: https://doi.org/10.12732/ijpam.v115i1.12.

[13] D. Guo, P. Boyle, C. Weng, and T. Wirjanto, "Eigen Portfolio Selection: A Robust Approach to Sharpe Ratio Maximization," Nov. 2017. Accessed: Dec. 30, 2024. [Online]. Available: https://business.unl.edu/academic-programs/departments/finance/actuarial-science/seminar-series/documents/WengPaper.pdf.

Statement

I hereby declare that this paper is my own work, not a paraphrase or translation of someone else's paper, and not plagiarism.

Bandung, 1 January 2025

Nathaniel Jonathan Rusli
13523013