UNIVERSITY OF ECONOMICS HO CHI MINH CITY

CELG Research Awards

PORTFOLIO OPTIMIZATION IN THE VIETNAMESE STOCK MARKET: AN INTEGRATED PRINCIPAL COMPONENT ANALYSIS-DATA ENVELOPMENT ANALYSIS AND MEAN-VARIANCE ANALYSIS APPROACH

Student full name:	Vo Ho Kien Quoc
Student ID:	31211020288
Class:	DH47IVC03
Supervisor:	Nguyen Khanh Duy

Ho Chi Minh City, October 2024

Contents

Abstract

In the rapidly growing Vietnamese stock market, investors face the challenge of selecting optimal portfolios from an increasingly vast pool of companies. This study introduces an integrated approach combining Principal Component Analysis (PCA), Data Envelopment Analysis (DEA), and Mean-Variance Analysis (MVA) to enhance stock selection and portfolio optimization. The PCA-DEA method is employed to identify efficient firms by reducing dimensionality, while MVA is used to allocate portfolio weights, maximizing risk-adjusted returns. Using quarterly financial data from 1,586 listed companies across three Vietnamese stock exchanges, this research compares portfolio performance across different levels of information retention—100%, 90%, 80%, and 70%. The results demonstrate that portfolios retaining 80% and 90% of the data consistently outperform both the VNINDEX benchmark and portfolios with full or reduced information retention. The study underscores the importance of balancing information retention for optimal performance and provides a robust framework for investors seeking higher returns with managed risk in emerging markets like Vietnam.

1. Introduction

In financial markets, stocks remain one of the most popular investment choices, as investors seek to maximize returns by identifying high-performing companies. However, the inherent uncertainty and volatility of stock markets raise critical concerns about whether these high-performing firms will consistently deliver strong returns (Ho et al., 2009). This creates a need for robust portfolio optimization strategies that can manage both the potential for high returns and the associated risks.

Portfolio optimization involves a complex decision-making process where investors must balance multiple conflicting objectives, such as maximizing returns while minimizing risk (Jothimani et al., 2017). This process is typically divided into three key phases: asset selection, asset allocation, and asset management. Asset selection involves identifying a diverse set of investment options, asset allocation determines how much capital to allocate to each investment, and asset management focuses on monitoring and adjusting the portfolio to optimize performance over time.

Markowitz (1952) introduced the groundbreaking Mean-Variance (MV) framework to tackle this challenge by finding an optimal portfolio that minimizes risk for a given level of return or maximizes return for a given level of risk. This framework has since been refined by numerous researchers (Doerner et al., 2004; Cura, 2009; Golmakani & Fazel, 2011). Despite its popularity, the MV model assumes that asset returns follow a multivariate normal distribution, a condition often violated in real-world markets, particularly during periods of extreme volatility like the 1987 stock market crash (Aouni et al., 2014). Additionally, portfolios generated through the MV model tend to hold a large number of assets, which can increase transaction costs and complicate implementation, especially when additional constraints are applied (Singh et al., 2010).

The growing number of companies listed on stock exchanges in Vietnam adds to the complexity of stock selection. As of the end of August 2024, the Vietnam stock market consists of 1,586 listed companies across three exchanges: the Ho Chi Minh Stock Exchange (HOSE), the Hanoi Stock Exchange (HNX), and the Upcom Stock Exchange (ssc.gov.vn). Investors must sift through vast amounts of information, such as financial reports and economic indicators, to identify the best investment opportunities. The asset selection process is crucial because it directly influences the success of the asset allocation and management phases in portfolio construction.

To streamline the stock selection process, (Smith, 1990) proposed the use of DEA, a method introduced by Charnes et al. (1978) for measuring the efficiency of decision-making units (DMUs). However, DEA faces limitations, particularly when dealing with a high number of input and output variables relative to the number of DMUs, a phenomenon known as the "curse of dimensionality." To address this, the current study employs a hybrid approach combining PCA with DEA. This integrated PCA-DEA method reduces dimensionality before applying DEA, improving the model's ability to identify the most efficient stocks.

This research applies the PCA-DEA method to the Vietnamese stock market, aiming to optimize portfolio performance by selecting efficient stocks and applying Mean-Variance Analysis (MVA) to determine the optimal allocation. The following sections will review existing literature, present the methodology in detail, and discuss the empirical results based on quarterly stock data.

A brief review of previous works is provided in the next section. Section 2 discusses different methods for performance evaluation, including the PCA-DEA approach and MVA. The methodology adopted for the study is detailed in Section 3. Section 4 presents the results and discussion, while Section 5 concludes the study with key insights.

2. Background

2.1. Different methods for performance evaluation

Stock selection models utilize various methods to evaluate performance, addressing both quantifiable and non-quantifiable factors that interact in complex ways. Multivariate Statistical Analysis (MSA) offers a comprehensive approach for analyzing multiple variables simultaneously, providing insights into relationships and patterns within datasets. However, MSA requires large sample sizes and normal data distribution, which can limit its applicability in some contexts (Ho et al., 2009). Despite its broad applicability, methods lacking statistical testing may further limit the reliability of results (Utkin, 2007).

The Analytic Hierarchy Process (AHP) is a structured method that uses pairwise comparisons and expert judgments to derive weighted values for decision-making. While AHP is easy to apply and offers consistency checks, divergent expert opinions can yield unreliable outcomes, and it does not account for interrelationships between factors (Saaty, 2008; Ho et al., 2009).

Fuzzy Set Theory (FST) effectively handles uncertainty and simulates human decision-making through the construction of subordinate functions. Its strength lies in its ability to interchange qualitative and quantitative values; however, results can be influenced by the chosen membership function, potentially reducing objectivity (Ho et al., 2009; Huang et al., 2012). Additionally, FST does not capture relationships between variables, which may limit the depth of the analysis (Ho et al., 2009).

Grey Relation Analysis (GRA) examines the trend development between indicators to assess the strength of their relationships. GRA does not require large sample sizes or specific data distributions, making it flexible (Ho, 2006). However, its reliance on quantitative data and the subjective choice of the Grey Relation coefficient introduces potential variability in the results (Ho et al., 2009).

The Balanced Scorecard (BSC) integrates financial and non-financial performance metrics into a cohesive evaluation system. While it provides a comprehensive strategy for aligning organizational performance with strategic goals, BSC can be complex and time-consuming to implement, limiting its feasibility for some firms (Kaplan & Norton, 1998; Ho et al., 2009).

Financial Statement Analysis offers a concrete, objective evaluation based on firms' financial performance, yet it is limited by its inability to incorporate qualitative factors such as organizational potential or employee morale (Espahbodi, 1991; Ho et al., 2009).

Among these methods, DEA is particularly suited to this study due to its non-parametric approach and the fact that it requires minimal assumptions about the functional form of variables. DEA objectively handles multiple inputs and outputs without being influenced by scale, making it highly effective for comparing firms' performance. Moreover, DEA minimizes subjectivity by calculating efficiency values mathematically, making it an ideal tool for evaluating stock selection models (Chen, 2008; Ho et al., 2009). Although DEA has some limitations, such as sensitivity to sample size and the potential for an overly large efficiency frontier, its strengths in handling complex data make it the most suitable model for this research.

2.2. Data Envelopment Analysis

DEA is a non-parametric linear programming technique introduced by Charnes et al. (1978), extending the earlier work of Farrell (1957). DEA assesses the efficiency of DMUs by comparing the relationship between multiple inputs and outputs. The concept of efficiency, rooted in physical and engineering sciences, is measured through the CCR ratio definition, which generalizes the classical single-output to single-input ratio to multiple outputs and inputs without needing pre-assigned weights (Hwang & Chang, 2003).

The CCR ratio model, as applied in DEA, effectively combines multiple inputs and outputs into a single efficiency score, enabling the selection of efficient firms for investment. This approach provides an objective evaluation of overall efficiency and identifies sources of inefficiency (Charnes et al., 1978). DEA has since evolved to include various models: the BCC model (Banker et al., 1984) distinguishes between technical and scale inefficiencies, the Multiplicative models (Charnes et al., 1982) offer interpretations of the production process, and the Additive model (Charnes et al., 1985, 1987) connects DEA results to the economic concept of Pareto optimality (Koopmans, 1951).

DEA calculates relative efficiency as the ratio of the weighted average of outputs to the weighted average of inputs. A DMU is deemed efficient with a score of 1, while scores below 1 indicate inefficiency. This method provides a comprehensive framework for evaluating the efficiency of DMUs, making it a powerful tool for performance assessment and decision-making.

The ratio form of the CCR (input-oriented) model for n comparable DMUs is

$$max_{u,v}\, h_o(u,v) = \frac{\sum_r u_r y_{ro}}{\sum_i v_i x_{io}} \tag{1}$$

$$s.t.\ 0 \leq \frac{\sum_r u_r y_{rj}}{\sum_i v_i x_{ij}} \leq 1, for\ j = 0, 1, \dots, n$$

$$u_r \geq 0, for\ r = 1, 2, \dots, s$$

$$\mathrm{v_i} \geq 0, \mathrm{for\ i} = 1, 2, \dots, \mathrm{m}$$

where, x_{io} represents the i^{th} input of the o^{th} DMU;

v_i represents the weight of that input;

y_{io} represents the i^{th} output of the o^{th} DMU;

u_r represents the weight of that output;

x_{ij} and y_{rj} represent the the i^{th} input and r^{th} output, respectively of the j^{th} DMU;

h_o represents the relative efficiency of DMU_o with respect to other DMUs.

The ratio form inherently leads to an infinite number of optimal solutions, as any scaled version of an optimal solution (u^*, v^*) such as (βu^*, βv^*) with $\beta > 0$ remains optimal (Adler & Golany, 2007). The linear programming formulations of the CCR model and the constant returns-to-scale additive models for *n* comparable units are

$$max_{M,N}\, MY^o \tag{2}$$

$$s.t.\ NX^o = 1$$

$$NX - MY \geq\ 0$$

$$M, N \geq\ 0$$

$$min_{M,N} NX^o - MY^o \tag{3}$$

$$s.t.\ NX - MY \geq\ 0$$

$$M \geq \vec{1}$$

$$N\ \geq \vec{1}$$

(Contreras, 2020; Jothimani et al., 2017),

where, j outputs are denoted by Y;

i inputs are denoted by X;

$\vec{1}$ is vetor of ones;

Y^o represents the output column of DMU under consideration, DMU_o;

X^o represents the input column of DMU under consideration, DMU_o;

M represents the vector of output weights;

N represents the vector of the input weights.

Research over the years has consistently validated the effectiveness of Data Envelopment Analysis (DEA) as a strategy for stock selection and portfolio construction, often outperforming traditional methods. For example, a study assessing 27 U.S. online companies found that, particularly in the Internet industry, company effectiveness was more crucial than operating efficiency in delivering higher investor returns (C.-T. B. Ho et al., 2009). Another investigation in Taiwan applied DEA to portfolio construction, revealing that while the "size effect" was not a viable strategy, DEA-constructed portfolios consistently outperformed the market index (Chen, 2008). Additionally, earlier work on stock selection strategies in emerging markets confirmed that incorporating multiple characteristics, including analysts' earnings revisions, significantly enhanced performance (van der Hart et al., 2003). Collectively, these studies underscore DEA's potential as a robust tool for stock selection across diverse markets, benefiting both investors and managers.

DEA's effectiveness in stock selection is due to its ability to handle multiple input and output parameters. However, the accuracy of efficiency evaluations can be compromised when the ratio of DMUs to parameters is imbalanced. Golany and Roll (1989) recommended that the number of DMUs should be at least twice the number of parameters, while (Dyson et al., 2001) suggested it should exceed twice the product of inputs and outputs. Typically, this issue is addressed by reducing parameters, but this can significantly affect efficiency scores (Adler & Yazhemsky, 2010).

To address these challenges, researchers have explored several methods to reduce variables in DEA, such as PCA-DEA (Adler & Golany, 2001, 2002; Ueda & Hoshiai, 1997), partial correlation (Jenkins & Anderson, 2003), regression-based analysis (Ruggiero, 2005), and Efficiency Contribution Measure (ECM) (Pastor et al., 2002). Reviews by Adler and Yazhemsky (2010) and Jothimani et al. (2017) found PCA-DEA to be the most effective, as it minimizes information loss by retaining principal components and avoids the need for expert opinion or parameter elimination, with the added benefit of a short run-time.

2.3. Principle Component Analysis

PCA analyzes the variance structure of a dataset by creating linear combinations of the original variables, thereby reducing the data to a few principal components that typically account for 80-90% of the variance (Adler & Golany, 2001). When the majority of the variance is captured by the first few components, these components can effectively replace the original variables with minimal loss of information. Let the random vector $X = [X_1, X_2, \dots, X_p]$ (in this study case the original inputs or outputs chosen to be aggregated) have a covariance matrix V with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_3 \geq 0$ and corresponding normalized eigenvectors $l_1, l_2, \dots, l_p$.The linear combinations can then be expressed as, where the superscript t denotes the transpose operator:

$$X_{PC} = l_i^t X = l_{1i}^t X_1 + l_{2i}^t X_2 + \cdots + l_{pi}^t X_p \quad (4)$$

$$Var\left(X_{PC_i}\right) = l_i^t V l_i, i = 1,2, \dots, p \quad (5)$$

$$Correlation\left(X_{PC_i}, X_{PC_k}\right) = l_i^t V l_k, i = 1,2, \dots, p, k = 1,2, \dots, p \quad (6)$$

The principal components $X_{PC_1}, X_{PC_2}, \dots, X_{PC_p}$ are uncorrelated linear combinations of the original variables, ranked by the amount of variance they explain, in descending order. The entire set of principal components is the same size as the original set of variables. However, as certain principal components are discarded, the matrix L_x, which initially contains all eigenvectors l_i, reduces from dimensions $m \times m$ to $h \times m$, where h represents the number of retained principal components. Consequently, X_{PC} becomes a matrix of size $h \times n$, effectively capturing the essential variance with fewer dimensions.

2.4. The PCA-DEA fomulation

The concept of combining PCA and DEA methodologies was independently introduced by Adler and Golany (2001, 2002) and Ueda and Hoshiai (1997). Both approaches aimed to enhance the discriminatory power of DEA by incorporating PCA for data reduction. In these studies, it was suggested that variables could be grouped logically based on their roles in the production process, with each group then represented by its principal components (PCs). Alternatively, PCA can be applied to the entire set of input and output variables to condense the data into a few uncorrelated PCs that typically capture 80-90% of the variance, thereby preserving essential information while improving the efficiency of the DEA model. It is noteworthy that when the principal components explain 100% of the variance in the original data, the PCA-DEA formulation becomes equivalent to the traditional DEA model, maintaining its full discriminatory power (Adler & Yazhemsky, 2010).

Traditional methods of variable reduction, which involve removing certain variables, often result in a significant loss of information and can lead to underestimating efficiency scores (Jothimani et al., 2017). The PCA-DEA model, however, reduces this issue by discarding only the principal components that contribute the least to data variance, thus retaining the majority of the original information. The full detail of an input or output is maintained until its associated principal component is completely excluded. This method effectively minimizes information loss and enhances accuracy. Additionally, PCA-DEA offers a similar function to weight-restricted DEA but without the subjectivity associated with expert judgment.

The DEA models presented in equations 2 and 3 can be adapted to incorporate principal components (PCs), which can replace either all inputs or outputs. This adaptation allows the linear program in equation 4 to use both the original data and PCs, creating a generalized formulation that ensures the efficiency scores from the PCA-DEA model are consistent with those from the standard DEA model when all PCs are included (Adler & Yazhemsky, 2010; Jothimani et al., 2017).

In this PCA-DEA framework, the vectors of output and input weights, M_{PC} and N_{PC}, are derived using the PCA coefficients, such that $Y_{PC} = L_y Y$ and $X_{PC} = L_x X$, where L_y and L_x are the PCA coefficients for the output and input data, respectively. The resulting PCA-DEA formulations for CCR and constant returns-to-scale additive models are presented in equations 7 and 8 (Adler & Golany, 2001; Adler & Yazhemsky, 2010; Jothimani et al., 2017).

$$max_{M_{PC},N_{PC}} M_{PC}^t Y_{PC}^o \tag{7}$$

$$s.t.\ N_{PC}^t X_{PC}^o = 1$$

$$N_{PC}^t X_{PC} - M_{PC}^t Y_{PC} \geq 0$$

$$N_{PC}^t L_x \geq 0$$

$$M_{PC}^t \mathrm{L_y} \geq 0$$

$$M_{PC}\ and\ N_{PC}\ are\ free$$

$$min_{N_{L_y} M_{L_x}} N_{PC} X_{PC}^o - M_{PC} Y_{PC}^o \tag{8}$$

$$s.t.\ N_{PC} X_{PC} - M_{PC} Y_{PC} \geq 0$$

$$N_{PC} L_x \ \geq \vec{1}$$

$$M_{PC} L_y \ \geq \vec{1}$$

$$N_{PC_i} - N_{PC_{i+1}} \geq 0, for\ i = 1, \dots, m-1$$

$$M_{PC_i} - M_{PC_{i+1}} \geq 0, for\ i = 1, \dots, m-1$$

$$M_{PC}\ and\ N_{PC}\ are\ free$$

In PCA, PCs are prioritized in descending order of importance. Constraints $M_{PC_i} - M_{PC_{i+1}} \geq 0$ and $N_{PC} - N_{PC_{i+1}} \geq 0$ ensure that the weight of PC_1 is at least equal to or greater than PC_2, the weight of PC_2 to be at least equal to or greater than PC_3 and so on. In this way, PCA increases the discriminating ability of the DEA model (Adler & Golany, 2001).

2.5. Mean-Variance Analysis

MVA serves as a crucial tool in modern portfolio theory, providing a systematic approach for investors to balance risk and return (Markowitz, 1991). Risk is quantified through the variance or standard deviation of asset returns, while the expected return represents the anticipated performance of the investment. The primary goal of mean-variance analysis is to assist investors in distinguishing between various portfolios by evaluating their potential returns and associated risks.

This method involves two key mathematical concepts: portfolio return and portfolio volatility. Portfolio return is calculated as the weighted sum of individual asset returns, where the weights represent the proportion of each asset in the portfolio. Portfolio volatility, on the other hand, depends on both the variances of individual assets and the correlations between all asset pairs. The expected return of a portfolio can be expressed as:

$$E(R_p) = \sum_i w_i E(R_i)$$

$$s.t \sum_i w_i = 1$$

where R_p is the portfolio return, R_i is the return of asset i, and w_i is the weight of asset i in the portfolio. The weights are constrained such that the sum of the portfolio weights equals 1.

Portfolio risk, or variance, is calculated using the formula:

$$\sigma_p^2 = \sum_i w_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} w_i w_j \sigma_i \sigma_j p_{ij}$$

where σ_i is the standard deviation of asset i , and p_{ij} represents the correlation between the returns of assets i and j. This equation captures the interaction of both individual asset volatilities and the relationships between asset pairs.

A widely used performance metric in this context is the Sharpe ratio, which measures the excess return per unit of risk taken. The Sharpe ratio is calculated as:

$$Sharpe\ Ratio = \frac{E(R_p) - R_f}{\sigma_p}$$

where $E(R_p)$ is the expected return of the portfolio, R_f is the risk-free rate, and σ_p is the standard deviation (risk) of the portfolio. The objective of mean-variance optimization is to find the portfolio allocation that maximizes the Sharpe ratio, thereby delivering the highest possible return for a given level of risk (Sharpe, 1994). This framework provides a mathematical basis for constructing efficient portfolios, where risk is minimized for a given level of expected return or, alternatively, returns are maximized for a defined level of risk.

3. Methodology

3.1. Data description

As of the end of August 2024, the Vietnam stock market consists of a total of 1,586 listed companies[1] across three exchanges: HOSE, HNX, and the Upcom Stock Exchange. For each quarter, we select the 100 non-financial companies with the highest market capitalization. From this initial pool, firms with negative net income and those lacking complete data are excluded. The number of selected stocks for each quarter is presented in Table 1.

The financial data for this study is sourced from the Vietcap Securities Joint Stock Company (VCI) database, and historical price data is obtained from the Techcom Securities Joint Stock Company (TCBS) database. Additionally, the one-year government bond yield data is retrieved from the Investing platform[2].

[1] ssc.gov.vn
[2] investing.com

Table 1. Number of non-financial large-cap stocks with positive net income by quarter.

Year	Quarter	Number of Stocks
2020	2	71
	3	79
	4	80
2021	1	85
	2	83
	3	83
	4	81
2022	1	83
	2	87
	3	80
	4	77
2023	1	72
	2	79
	3	76
	4	75
2024	1	77
	2	78
	3	82

3.2. Defining DEA parameters

Selection of input and output parameters plays a vital role in evaluating the relative financial performance of the firms (Golany & Roll, 1989). In this study, we carefully select several key financial indicators to serve as input and output variables, based on relevant literature (Lawrence . & Joe, 1999; Lo & Lu, 2006; Serrano-Cinca et al., 2005). These variables are intended to capture both the financial efficiency and operational performance of the companies listed on the Vietnam stock market.

For the input variables, we include total assets, total equity, operating expenses, and cost of sales. Total assets are chosen because they represent the overall capacity of a firm to generate revenue, including current, fixed, and intangible assets. Intangible assets, such as goodwill resulting from mergers or acquisitions, can be challenging to quantify but are essential in understanding a company’s complete asset base. Total equity is selected as it reflects the ownership stake in the company, providing insights into how well the firm leverages its capital to generate returns. Operating expenses are included to capture the firm's cost structure. Since many firms, especially in emerging markets, may not provide detailed expense breakdowns, operating expenses offer a broad measure of the general costs associated with running the business. Furthermore, cost of sales is incorporated as it reflects the direct costs involved in producing goods or services, thus playing a significant role in determining how efficiently the company manages its production and supply chain to achieve profitability.

On the output side, we focus on net income and earnings per share (EPS). Net income is widely used as a measure of a company's ability to manage its resources and generate profits after accounting for all operating costs, taxes, and interest payments. It provides a clear perspective on overall profitability. EPS, in turn, gives insight into how much of the generated profit is distributed to shareholders, which is crucial for evaluating value creation from an investor's viewpoint.

By selecting these input and output variables, we aim to measure the financial efficiency of companies by assessing how well they utilize their resources—namely total assets, equity, operating expenses, and cost of sales—to generate revenue, profits, and shareholder value. This comprehensive approach enables us to rank and select potential stocks using the PCA-DEA model, ensuring that the most efficient companies are included in the portfolio selection process.

3.3. Framework for portfolio selection, allocation and management

In this study, the PCA-DEA approach is used to enhance the efficiency evaluation of firms, overcoming the limitations of traditional DEA models. While DEA can struggle with reduced discriminatory power in cases where the number of DMUs is limited, integrating PCA allows for dimensionality reduction while retaining essential information. This approach improves the model's ability to differentiate between efficient and inefficient firms, addressing the challenges of traditional DEA (Adler & Yazhemsky, 2010).

Additionally, MVA is incorporated to optimize the portfolio selection process. After identifying efficient firms using PCA-DEA, MVA is used to determine the optimal stock weightings that maximize the Sharpe ratio, balancing risk and return (Markowitz, 1991). The combination of PCA-DEA and MVA provides a comprehensive framework for selecting and optimizing potential stock portfolios.

The input and output data for the DEA model, used to identify the potential stock portfolio for each quarter, is sourced from the financial reports of the previous quarter. Additionally, data on stock prices and the 1-year government bond yield, used to determine the optimal weights for each stock in the portfolio, is also gathered from the previous quarter. At the end of each quarter, the potential portfolio is updated based on the newly released financial reports.

According to regulations, listed companies are required to submit their quarterly financial reports no later than 20 days after the quarter ends (thuvienphapluat.vn). Therefore, in this study, each quarter begins on the 21st of the first month of the quarter and ends on the 20th of the first month of the following quarter. This timeline allows for the analysis and potential portfolio adjustments based on the most up-to-date financial data.

In the pre-processing step, the data is standardized by dividing each value in the input and output datasets by their corresponding mean. This method ensures that all variables are on a similar scale, which helps to address scaling discrepancies and minimizes the likelihood of round-off errors in the computational processes (Sarkis, 2007).

The proposed research framework aims to evaluate and optimize stock portfolios using an integrated PCA-DEA-MVA approach. The steps are as follows:

1. *Apply PCA:* For each quarter (starting with Q2 2020), independently apply PCA to inputs and outputs to reduce dimensionality while retaining key variance.
2. *Full information retention (100%):* Retain 100% of the variance to capture all relevant information in the data.

3. *Evaluate efficiency:* Calculate the efficiency scores of firms using Equation 7. Firms with a score of 1 are deemed efficient.
4. *Select efficient firms:* Firms with efficiency scores of 1 are identified as potential stocks for the portfolio.
5. *Construct portfolio for Q2 2020:* Compile a list of efficient firms to create the Q2 2020 stock portfolio.
6. *Optimize portfolio with Mean-Variance Analysis:* Use mean-variance analysis to determine stock weights that maximize the Sharpe ratio, optimizing risk-adjusted returns.
7. *Monitor portfolio:* Track the portfolio's performance throughout the quarter.
8. Repeat steps 1-7 for each subsequent quarter.
9. Repeat the process for lower information retention levels (90%, 80%, and 70%) to analyze the impact on portfolio performance.

4. Results and discussion

4.1. Stocks selection

The efficient companies, identified by a score of 1, are selected for inclusion in the potential stock portfolio for each quarter. For the 80% information retention level, the results are presented in two tables: Table 2 summarizes efficiency scores, including minimum, maximum, average scores, and the number and percentage of efficient firms per quarter, while Table 3 lists the efficient firms by quarter. This analysis tracks stock performance and efficiency over time. Results for other information retention levels are provided in the appendix.

Table 1 presents an overview of the efficiency scores for each quarter, revealing fluctuations in firm performance. The maximum efficiency score remains consistently at 1, indicating that a few firms achieve optimal efficiency every quarter. However, the minimum efficiency scores show significant variation, ranging from 0.08 in Q4 2021 to

0.31 in Q1 2022. This wide disparity suggests that certain firms struggle with maintaining efficiency during specific periods, possibly due to sector-specific challenges or broader economic conditions, such as supply chain disruptions or regulatory changes.

The average efficiency scores across quarters fluctuate between 0.35 and 0.58. Lower average efficiency scores, such as 0.37 in Q4 2021, may indicate periods of economic instability where firms face challenges in optimizing resources. Conversely, higher average scores like 0.58 in Q1 2023 may reflect periods of economic recovery or stability, where firms are able to improve their performance. These trends suggest that external factors, such as market conditions or economic cycles, significantly impact firm efficiency across different periods.

The number of efficient firms, defined as those with an efficiency score of 1, also varies by quarter. For example, Q2 2023 has the highest number of efficient firms (8 firms, or 10.13%), while other quarters, such as Q3 2020 and Q4 2021, show only 3 efficient firms. This variation may be due to changing market conditions or economic factors that either enable or hinder firms' ability to optimize their operations. The relatively low percentage of efficient firms overall (ranging from 2.5% to 10.13%) indicates that only a small portion of companies consistently operate at optimal efficiency, making these firms valuable candidates for portfolio inclusion.

Table 2 provides a list of the firms that achieved an efficiency score of 1 in each quarter. Some firms, such as MWG (Mobile World Investment Corporation) and VCF (Vinacafé Bienhoa Joint Stock Company), appear consistently across multiple quarters. This consistency highlights their ability to maintain efficiency over time, suggesting strong management practices and adaptability to market changes. For instance, MWG's presence across several quarters may be attributed to its well-developed retail network and effective e-commerce strategies, allowing it to remain efficient even during challenging periods.

Table 2. Descriptive statistics of firms' efficiency for 80% information level by quarter.

Year	Quarter	Minimum Efficiency	Maximum Efficiency	Average Efficiency	Number of Efficient Firms	Efficient Firms (%)
2020	2	0.24	1.00	0.49	4	5.63%
	3	0.24	1.00	0.44	3	3.80%
	4	0.18	1.00	0.38	3	3.75%
2021	1	0.25	1.00	0.46	6	7.06%
	2	0.27	1.00	0.57	6	7.23%
	3	0.17	1.00	0.35	3	3.61%
	4	0.08	1.00	0.37	3	3.70%
2022	1	0.31	1.00	0.51	4	4.82%
	2	0.27	1.00	0.52	3	3.45%
	3	0.13	1.00	0.46	2	2.50%
	4	0.16	1.00	0.58	7	9.09%
2023	1	0.27	1.00	0.58	5	6.94%
	2	0.19	1.00	0.53	8	10.13%
	3	0.22	1.00	0.52	4	5.26%
	4	0.15	1.00	0.45	3	4.00%
2024	1	0.27	1.00	0.47	4	5.19%
	2	0.25	1.00	0.49	3	3.85%
	3	0.28	1.00	0.52	6	7.32%

Other firms, like DGC and HPG, also show up consistently in certain quarters, indicating that sectors such as chemicals and construction may have benefited from favorable demand or sector-specific trends during those times. However, some companies only appear in isolated quarters, which could reflect temporary advantages or challenges specific to that period. For example, economic recovery phases or temporary boosts in demand might lead to a firm achieving efficiency in one quarter but not in others.

There are several possible causes for these fluctuations in efficiency. Sector-specific factors, such as changes in regulatory environments, raw material costs, or consumer demand, can significantly impact firm performance. Additionally, broader economic cycles, such as the recovery seen in 2022 and 2023, may explain why some firms were able to regain efficiency after facing challenges in previous quarters. Firm-specific management practices also play a key role; companies with more adaptable leadership and strategies, like MWG and VCF, were able to sustain higher efficiency levels.

The implications for portfolio selection are clear. Firms that consistently appear as efficient in Table 2, such as MWG and VCF, represent stable investment opportunities due to their proven ability to maintain operational efficiency over time. Including these firms in a portfolio could provide a level of security, especially during volatile periods. Additionally, sector-specific firms like HPG and DGC may offer opportunities depending on broader sector trends, though investors should remain mindful of risks tied to specific industries.

Table 3. Potential stocks for 80% information level by quarter.

No.	2020			2021				2022				2023				2024		
	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
1	DBC	MWG	MWG	HPG	DBC	HPG	HPG	DGC	DGC	BSR	ACV	MCH	BMP	GMD	ACV	MCH	HVN	ACV
2	MWG	VCF	VCF	MCH	DHC	IPA	MWG	MSN	HPG	DGC	DGC	MWG	BSR	MWG	BSR	MWG	MWG	BMP
3	VCF	VNM	VNM	MWG	HPG	MWG	VNM	MWG	VCF		IDP	VCF	GAS	VCF	VCF	VCF	VCF	MCH
4	VNM			VCF	MWG			VCF			MSN	VEA	HPG	VNM		VNM		MWG
5				VEA	VCF						MWG	VNM	PDN					VCF
6				VNM	VEA						VCF		REE					VNM
7											VNM		VCF					
8													VEA					

Given the fluctuation in the number of efficient firms each quarter, it becomes evident that market timing is crucial. Some periods may see a higher proportion of efficient firms, while others experience fewer. This suggests that diversifying the portfolio across sectors and consistently performing firms may be essential to reducing risk, especially during periods of economic uncertainty or sector-specific downturns. By tracking these trends, investors can make more informed decisions and optimize their portfolio strategies.

4.2. Portfolios allocation and management

In this section, we analyze the performance of portfolios constructed using different levels of information retention (100%, 90%, 80%, and 70%) and compare them to the

VNINDEX benchmark. At the beginning of each new quarter, the potential stock portfolio is identified using PCA-DEA and reallocated using MVA. Both the stock selection and portfolio weights are updated quarterly based on new financial data. The chart illustrates the performance of these portfolios over time, with all portfolios starting at a value of 1000. Additionally, key performance metrics, including the Sharpe Ratio, annualized return, and volatility, are provided to give further insight into the relative efficiency of these portfolios.

The portfolio with 80% information retention demonstrates the best overall performance, reaching a final value of 7864.7 by the end of the period. It boasts an impressive annualized return of 46.45% and a Sharpe Ratio of 1.36, indicating a solid balance between risk and return. Although its volatility is somewhat elevated at 32.95%, this portfolio effectively navigates market fluctuations, exhibiting resilience and sustained growth over time. Its upward trajectory, despite occasional dips, highlights its ability to capitalize on opportunities in the market and maintain superior performance.

Similarly, the portfolio with 90% information retention also performs exceptionally well, finishing with a final value of 6930.4. This portfolio has the highest Sharpe Ratio of

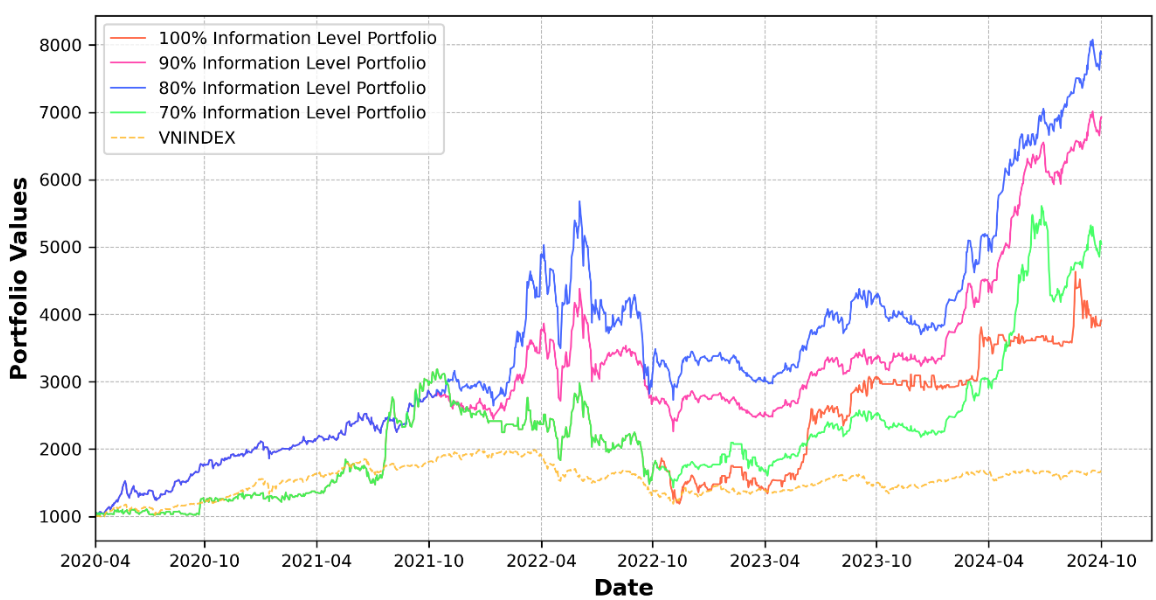

Figure 1. Portfolio values over time at different information levels.

1.44, reflecting its superior risk-adjusted returns, and its annualized return of 43.60% is notable. What makes this portfolio particularly attractive is its lower annualized volatility of 29.14%, which suggests that it offers a smoother ride for investors while still delivering substantial returns. The combination of relatively low risk and high return makes the 90% portfolio an ideal choice for investors seeking both stability and performance.

Table 4. Portfolios performance at different information levels.

Information Level:	100%	90%	80%	70%	VNINDEX
Sharpe Ratio	0.65	1.44	1.36	0.86	0.5
Annualized Return	30.71%	43.60%	46.45%	36.45%	11.65%
Annualized Volatility	44.29%	29.14%	32.95%	40.41%	19.85%
Final Portfolio Value	3909.9	6930.4	7864.7	5046.5	1286.4

In contrast, the portfolio that retains 100% of the information does not perform as well as the 80% and 90% portfolios. With a final value of 3909.9, it achieves an annualized return of 30.71% and a Sharpe Ratio of 0.65, both of which are considerably lower than the portfolios that retained less information. Additionally, its volatility is the highest among all portfolios at 44.29%, which indicates that the 100% portfolio is subject to greater swings in value. This underperformance suggests that retaining all available information introduces noise or less relevant data into the model, reducing its effectiveness in optimizing portfolio allocation. The results imply that using a more focused data set—such as the 80% or 90% information level portfolios—produces better outcomes by eliminating unnecessary variables that may obscure key signals in the data.

The portfolio with 70% information retention, while still outperforming the VNINDEX, also lags behind the 80% and 90% portfolios. It finishes with a final value of 5046.5 and an annualized return of 36.45%, coupled with a Sharpe Ratio of 0.86. Its relatively high volatility of 40.41% suggests that removing too much information from the data set reduces the portfolio's ability to consistently capture optimal performance. While this portfolio still offers positive returns and strong risk-adjusted metrics compared to the market index, it underperforms when compared to the portfolios with 80% and 90% information retention, likely due to the loss of critical information necessary for accurate stock selection and allocation.

When compared to the VNINDEX benchmark, the actively managed portfolios clearly outperform the market. The VNINDEX finishes with a final value of 1286.4, reflecting a modest annualized return of 11.65%. Its Sharpe Ratio of 0.5 and annualized volatility of 19.85% show that while the market is relatively stable, it offers significantly lower returns compared to the PCA-DEA and MVA-optimized portfolios. This stark contrast emphasizes the effectiveness of the PCA-DEA and MVA approach in generating superior risk-adjusted returns. Even the lowest-performing PCA-DEA portfolio (the 100% information level portfolio) still dramatically outperforms the VNINDEX, further underscoring the value of this combined methodology for portfolio management.

The key implication of these results is that reducing the amount of information retained through PCA to a manageable level, such as 80% or 90%, leads to better portfolio performance. The portfolios with these information levels offer the best trade-offs between risk and return, as demonstrated by their high Sharpe Ratios and superior final values. This suggests that retaining too much information (as in the 100% portfolio) can introduce noise, while discarding too much information (as in the 70% portfolio) can result in the loss of critical data. The findings highlight the importance of dimensionality reduction in financial modeling, as it helps to focus on the most relevant data without overfitting or losing valuable insights.

Furthermore, the superior performance of the PCA-DEA and MVA portfolios compared to the VNINDEX provides strong evidence that this approach can add significant value to investment strategies. Investors who apply this methodology can expect higher returns and better risk management than those who follow a market index-based strategy. The consistent outperformance of the 80% and 90% portfolios suggests that this combination of stock selection and allocation strategies can be highly effective in the Vietnamese stock market, providing a robust framework for optimizing investments.

5. Conclusion

This study demonstrates the effectiveness of combining Principal Component Analysis, Data Envelopment Analysis, and Mean-Variance Analysis in optimizing stock portfolio selection and allocation within the Vietnamese stock market. By applying PCA-DEA to identify efficient firms each quarter and reallocating portfolios using MVA, the research reveals that reducing dimensionality to a manageable level (80% and 90% information retention) leads to superior risk-adjusted returns. These portfolios consistently outperform both the VNINDEX and portfolios that retain too much or too little information, emphasizing the importance of striking the right balance in data retention.

The findings highlight that portfolios with 80% and 90% information retention offer the most favorable trade-offs between risk and return, as evidenced by their high Sharpe Ratios, strong annualized returns, and resilience in the face of market volatility. Conversely, retaining 100% of the data introduces noise and leads to underperformance, while excessive data reduction (70%) results in the loss of crucial information, reducing the model's predictive power. These results underline the necessity of reducing dimensionality in financial modeling to focus on the most relevant variables, improving the ability to make well-informed investment decisions.

Moreover, the study's consistent outperformance of the VNINDEX demonstrates the value of using an active management approach combining PCA-DEA and MVA. By focusing on efficiency in stock selection and optimizing portfolio allocation, investors can significantly enhance their returns and manage risk more effectively than simply following a market index. This approach provides a clear framework for portfolio optimization, particularly in emerging markets like Vietnam, where stock performance can be highly variable.

In conclusion, the integration of PCA, DEA, and MVA offers a powerful methodology for portfolio optimization, delivering superior results through efficient stock selection and risk management. Investors who adopt this approach can expect enhanced portfolio performance, even in volatile markets. This study provides valuable insights into the benefits of data-driven investment strategies, reinforcing the importance of careful data selection and robust portfolio allocation in achieving long-term financial success.

References

Adler, N., & Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. *European Journal of Operational Research*, *132*(2), 260–273. https://doi.org/10.1016/S0377-2217(00)00150-8

Adler, N., & Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the Operational Research Society*. https://doi.org/10.1057/palgrave.jors.2601400

Adler, N., & Golany, B. (2007). PCA-DEA. In Z. Joe & C. Wade D. (Eds.), *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis* (pp. 139–153). Springer US. https://doi.org/10.1007/978-0-387-71607-7_8

Adler, N., & Yazhemsky, E. (2010). Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction. *European Journal of Operational Research*, *202*(1), 273–284. https://doi.org/10.1016/j.ejor.2009.03.050

Aouni, B., Colapinto, C., & La Torre, D. (2014). Financial portfolio management through the goal programming model: Current state-of-the-art. *European Journal of Operational Research*, *234*(2), 536–545. https://doi.org/10.1016/j.ejor.2013.09.040

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, *30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Charnes, A., Cooper, W. W., Golany, B., Seiford, L., & Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, *30*(1), 91–107. https://doi.org/10.1016/0304-4076(85)90133-2

Charnes, A., Cooper, W. W., Learner, D. B., Phillips, F. Y., & Rousseau, J. (1987). *Explicit construction and analysis of some efficient empirical production functions in data envelopment analysis*.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Charnes, A., Cooper, W. W., Seiford, L., & Stutz, J. (1982). A multiplicative model for efficiency analysis. *Socio-Economic Planning Sciences*, *16*(5), 223–224. https://doi.org/10.1016/0038-0121(82)90029-5

Chen, H. (2008). Stock selection using data envelopment analysis. *Industrial Management & Data Systems*, *108*(9), 1255–1268. https://doi.org/10.1108/02635570810914928

Contreras, I. (2020). A review of the literature on DEA models under common set of weights. *Journal of Modelling in Management*, *15*(4), 1277–1300. https://doi.org/10.1108/JM2-02-2019-0043

Cura, T. (2009). Particle swarm optimization approach to portfolio optimization. *Nonlinear Analysis: Real World Applications*, *10*(4), 2396–2406. https://doi.org/10.1016/j.nonrwa.2008.04.023

Doerner, K., Gutjahr, W. J., Hartl, R. F., Strauss, C., & Stummer, C. (2004). Pareto Ant Colony Optimization: A Metaheuristic Approach to Multiobjective Portfolio Selection. *Annals of Operations Research*, *131*(1–4), 79–99. https://doi.org/10.1023/B:ANOR.0000039513.99038.c6

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259. https://doi.org/10.1016/S0377-2217(00)00149-1

Espahbodi, P. (1991). Identification of problem banks and binary choice models. *Journal of Banking & Finance*, *15*(1), 53–71. https://doi.org/10.1016/0378-4266(91)90037-M

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society: Series A (General)*, *120*(3), 253–281. https://doi.org/10.2307/2343100

Golany, B., & Roll, Y. (1989). An application procedure for DEA. *Omega*, *17*(3), 237–250. https://doi.org/10.1016/0305-0483(89)90029-7

Golmakani, H. R., & Fazel, M. (2011). Constrained Portfolio Selection using Particle Swarm Optimization. *Expert Systems with Applications*, *38*(7), 8327–8335. https://doi.org/10.1016/j.eswa.2011.01.020

Ho, C.-T. (2006). Measuring bank operations performance: An approach based on Grey Relation Analysis. *Journal of the Operational Research Society*, *57*(4), 337–349. https://doi.org/10.1057/palgrave.jors.2601985

Ho, C.-T. B., Wu, D. D., & Olson, D. L. (2009). A risk scoring model and application to measuring internet stock performance. *International Journal of Information Technology & Decision Making*, *08*(01), 133–149. https://doi.org/10.1142/S0219622009003302

Huang, C.-F., Chang, B., Cheng, D.-W., & Chang, C.-H. (2012). Feature Selection and Parameter Optimization of a Fuzzy-based Stock Selection Model Using Genetic Algorithms. *International Journal of Fuzzy Systems*, *14*.

Hwang, S.-N., & Chang, T.-Y. (2003). Using data envelopment analysis to measure hotel managerial efficiency change in Taiwan. *Tourism Management*, *24*(4), 357–369. https://doi.org/10.1016/S0261-5177(02)00112-7

Jenkins, L., & Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, *147*(1), 51–61. https://doi.org/10.1016/S0377-2217(02)00243-6

Jothimani, D., Shankar, R., & Yadav, S. S. (2017). A PCA-DEA framework for stock selection in Indian stock market. *Journal of Modelling in Management*, *12*(3), 386–403. https://doi.org/10.1108/JM2-09-2015-0073

Kaplan, R. S., & Norton, D. P. (1998). Putting the Balanced Scorecard to Work. In *The Economic Impact of Knowledge* (pp. 315–324). Elsevier. https://doi.org/10.1016/B978-0-7506-7009-8.50023-9

Koopmans, T. C. (1951). *Activity analysis of production and allocation*. https://krishikosh.egranth.ac.in/items/5228f77b-6513-46ed-9e5f-84621b170cc7

Lawrence ., S., & Joe, Z. (1999). Profitability and Marketability of the Top 55 U.S. Commercial Banks. *Management Science*. https://pubsonline.informs.org/doi/abs/10.1287/mnsc.45.9.1270

Lo, S. F., & Lu, W. M. (2006). Does size matter? Finding the profitability and marketability benchmark of financial holding companies. *Asia-Pacific Journal of Operational Research*. https://www.worldscientific.com/doi/abs/10.1142/S0217595906000930

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance*, *7*(1), 77–91.

Markowitz, H. M. (1991). Foundations of Portfolio Theory. *The Journal of Finance*, *46*(2), 469–477. https://doi.org/10.2307/2328831

Pastor, J. T., Ruiz, J. L., & Sirvent, I. (2002). A Statistical Test for Nested Radial Dea Models. *Operations Research*. https://doi.org/10.1287/opre.50.4.728.2866

Ruggiero, J. (2005). Impact assessment of input omission on dea. *International Journal of Information Technology & Decision Making*, *04*(03), 359–368. https://doi.org/10.1142/S021962200500160X

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, *1*(1), 83–98. https://doi.org/10.1504/IJSSCI.2008.017590

Sarkis, J. (2007). Preparing Your Data for DEA. In J. Zhu & W. D. Cook (Eds.), *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis* (pp. 305–320). Springer US. https://doi.org/10.1007/978-0-387-71607-7_17

Serrano-Cinca, C., Fuertes-Callén, Y., & Mar-Molinero, C. (2005). Measuring DEA efficiency in Internet companies. *Decision Support Systems*, *38*(4), 557–573. https://doi.org/10.1016/j.dss.2003.08.004

Sharpe, W. F. (1994). The Sharpe Ratio (Fall 1994). In P. L. Bernstein & F. J. Fabozzi (Eds.), *Streetwise: The Best of The Journal of Portfolio Management* (pp. 169–178). Princeton University Press. https://doi.org/10.1515/9781400829408-022

Singh, A. K., Sahu, R., & Bharadwaj, S. (2010). Portfolio evaluation using OWA-heuristic algorithm and data envelopment analysis. *The Journal of Risk Finance*, *11*(1), 75–88. https://doi.org/10.1108/15265941011012697

Smith, P. (1990). Data envelopment analysis applied to financial statements. *Omega*, *18*(2), 131–138. https://doi.org/10.1016/0305-0483(90)90060-M

Ueda, T., & Hoshiai, Y. (1997). Application of Principal Component Analysis for Parsimonious Summarization of Dea Inputs and/or Outputs. *Journal of the Operations Research Society of Japan*, *40*(4), 466–478. https://doi.org/10.15807/jorsj.40.466

Utkin, L. V. (2007). Risk analysis under partial prior information and nonmonotone utility functions. *International Journal of Information Technology & Decision Making*, *06*(04), 625–647. https://doi.org/10.1142/S0219622007002721

van der Hart, J., Slagter, E., & van Dijk, D. (2003). Stock selection strategies in emerging markets. *Journal of Empirical Finance*, *10*(1), 105–132. https://doi.org/10.1016/S0927-5398(02)00022-1

Appendix

Appendix 1. Descriptive Statistics of Firms' Efficiency for 70% Information Level by Quarter.

Year	Quarter	Minimum Efficiency	Maximum Efficiency	Average Efficiency	Number of Efficient Firms	Efficient Firms (%)
2020	2	0.1	1.00	0.33	1	1.41%
	3	0.03	1.00	0.3	1	1.27%
	4	0.06	1.00	0.24	1	1.25%
2021	1	0.07	1.00	0.32	1	1.18%
	2	0.08	1.00	0.45	1	1.20%
	3	0.07	1.00	0.24	1	1.20%
	4	0.05	1.00	0.35	1	1.23%
2022	1	0.06	1.00	0.34	1	1.20%
	2	0.27	1.00	0.52	3	3.45%
	3	0.13	1.00	0.46	2	2.50%
	4	0.16	1.00	0.58	7	9.09%
2023	1	0.07	1.00	0.37	1	1.39%
	2	0.06	1.00	0.45	7	8.86%
	3	0.22	1.00	0.52	4	5.26%
	4	0.15	1.00	0.45	3	4.00%
2024	1	0.27	1.00	0.47	4	5.19%
	2	0.13	1.00	0.41	2	2.56%
	3	0.17	1.00	0.49	5	6.10%

Appendix 2. Potential Stocks for 70% Information Level by Quarter.

No.	2020			2021				2022				2023				2024		
	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
1	VCF	VCF	VCF	VCF	HPG	IPA	HPG	VCF	DGC	BSR	ACV	VCF	BMP	GMD	ACV	MCH	HVN	ACV
2									HPG	DGC	DGC		BSR	MWG	BSR	MWG	VCF	BMP
3									VCF		IDP		GAS	VCF	VCF	VCF		MWG
4											MSN		PDN	VNM		VNM		VCF
5											MWG		REE					VNM
6											VCF		VCF					
7											VNM		VEA					

Appendix 3. Descriptive Statistics of Firms' Efficiency for 90% Information Level by Quarter.

Year	**Quarter**	**Minimum Efficiency**	**Maximum Efficiency**	**Average Efficiency**	**Number of Efficient Firms**	**Efficient Firms (%)**
2020	2	0.24	1.00	0.49	4	5.63%
	3	0.24	1.00	0.44	3	3.80%
	4	0.18	1.00	0.38	3	3.75%
2021	1	0.25	1.00	0.46	6	7.06%
	2	0.27	1.00	0.57	6	7.23%
	3	0.17	1.00	0.35	3	3.61%
	4	0.18	1.00	0.53	6	7.41%
2022	1	0.36	1.00	0.57	6	7.23%
	2	0.35	1.00	0.56	5	5.75%
	3	0.17	1.00	0.56	8	10.00%
	4	0.25	1.00	0.6	7	9.09%
2023	1	0.39	1.00	0.6	7	9.72%
	2	0.32	1.00	0.57	8	10.13%
	3	0.27	1.00	0.59	7	9.21%
	4	0.24	1.00	0.48	4	5.33%
2024	1	0.28	1.00	0.52	6	7.79%
	2	0.31	1.00	0.55	8	10.26%
	3	0.28	1.00	0.53	8	9.76%

Appendix 4. Potential Stocks for 90% Information Level by Quarter.

No.	**2020**			**2021**				**2022**				**2023**				**2024**		
	2	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**
1	DBC	MWG	MWG	HPG	DBC	HPG	HPG	ACV	DGC	ACV	ACV	DGC	BMP	ACV	ACV	ACV	ACV	ACV
2	MWG	VCF	VCF	MCH	DHC	IPA	MCH	DGC	DPM	BSR	DGC	MCH	BSR	GMD	BSR	HPG	FPT	BMP
3	VCF	VNM	VNM	MWG	HPG	MWG	MWG	MSN	GVR	DGC	IDP	MWG	GAS	GVR	HPG	MCH	HVN	MCH
4	VNM			VCF	MWG		VCF	MWG	HPG	GAS	MSN	PLX	HPG	MSN	VCF	MWG	MCH	MSN
5				VEA	VCF		VCS	VCF	VCF	GVR	MWG	VCF	PDN	MWG		VCF	MSN	MWG
6				VNM	VEA		VNM	VNM		HPG	VCF	VEA	REE	VCF		VNM	MWG	VCF
7										VCF	VNM	VNM	VCF	VNM			VCF	VEA
8										VHC			VEA				VNM	VNM

Appendix 5. Descriptive Statistics of Firms' Efficiency for 100% Information Level by Quarter.

Year	Quarter	Minimum Efficiency	Maximum Efficiency	Average Efficiency	Number of Efficient Firms	Efficient Firms (%)
2020	2	0.1	1.00	0.32	1	1.41%
	3	0.02	1.00	0.28	1	1.27%
	4	0.06	1.00	0.22	1	1.25%
2021	1	0.07	1.00	0.32	1	1.18%
	2	0.08	1.00	0.45	1	1.20%
	3	0.07	1.00	0.24	1	1.20%
	4	0.05	1.00	0.35	1	1.23%
2022	1	0.06	1.00	0.33	1	1.20%
	2	0.04	1.00	0.32	1	1.15%
	3	0.12	1.00	0.45	1	1.25%
	4	0.03	1.00	0.34	1	1.30%
2023	1	0.07	1.00	0.36	1	1.39%
	2	0.03	1.00	0.37	2	2.53%
	3	0.02	1.00	0.23	1	1.32%
	4	0.03	1.00	0.3	1	1.33%
2024	1	0.03	1.00	0.28	1	1.30%
	2	0.06	1.00	0.33	1	1.28%
	3	0.07	1.00	0.35	1	1.22%

Appendix 6. Potential Stocks for 100% Information Level by Quarter.

No.	2020			2021				2022				2023				2024		
	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3
1	VCF	VCF	VCF	VCF	HPG	IPA	HPG	VCF	DGC	DGC	DGC	VCF	PDN	GMD	VCF	VCF	VCF	VCF
2													VCF					