STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–13.
Published online in International Academic Press (www.IAPress.org)

Dynamic Portfolio Optimisation in Morocco's Stock Market through Machine-Learning Selection and Complex Mean-Variance allocation

Achraf BOUHMADY [1,*], Hamza Kadiri [2,3], Khalid Belkhoutout [2], Nadia Raissi [1]

[1] *LMSA, Department of Mathematics, Faculty of Science, Mohammed V University in Rabat, Morocco*
[2] *LERJP, Department of Economics, Faculty of Law, Economic and Social Sciences of Souissi, Mohammed V University in Rabat, Morocco*
[3] *Laboratory of Management, Finance, Marketing, and International Development MFMDI, SupMTI, Rabat, Morocco*

Abstract Emerging markets, such as Morocco's stock exchange, encounter structural challenges, including liquidity constraints, sectoral concentration, and low macroeconomic sensitivity, which render traditional portfolio optimization methods inadequate. Consequently, market forecasts, whether based on chartist or fundamental analysis, become less relevant. This study proposes a hybrid framework that integrates machine learning (ML) techniques for stock selection with a novel Mean-Variance Complex-Based (MVCB) optimization method to enhance performance within the Moroccan All Shares Index (MASI). Five ML models, including Ridge Regression, Stepwise Regression, Random Forest, Generalized Boosted Regression, and XGBoost, are employed to predict returns based on fundamental and technical indicators, with XGBoost demonstrating superior predictive accuracy. The MVCB method utilizes complex returns derived from the Hilbert Transform, effectively capturing dynamic market correlations and phase-amplitude relationships to optimize portfolio weights in the presence of volatility. Backtesting results indicate that the MVCB portfolio surpasses traditional mean-variance (MV) and market benchmarks, achieving a 10.48% annualized gross return with a volatility of 3.52% and a Sharpe ratio of 2.48, compared to 1.12 for the MASI. Additionally, sector diversification and a reduction in left-tail risk (19.3%) contribute to mitigating correlation breakdowns during crises. By synergizing predictive ML with adaptive optimization, this framework addresses the inherent instability of emerging markets, providing a robust and scalable solution for enhancing risk-adjusted returns.

Keywords Modeling, Forecasting, Hybrid models, Machine learning, Deep Learning, Portfolio optimization.

DOI: 10.19139/soic-2310-5070-2529

1. Introduction

In recent decades, portfolio optimization has become an important area in finance, helping investors balance risk and return in different market situations [1]. However, traditional methods used in emerging markets, which often have low liquidity, low sensitivity to economic changes, and a focus on specific sectors, can be less effective [2]. Common diversification strategies usually aim to reduce volatility by spreading investments across various asset types, regions, and company sizes [3]. Yet, major shocks like the 2008 financial crisis and the COVID-19 pandemic

*Correspondence to: Achraf BOUHMADY (Email: bouhmadiachraf@gmail.com). Department of Mathematics, Faculty of Science, Mohammed V University in Rabat. B.P. 8007 Nations-Unies Avenue des Nations Unies Agdal, Rabat, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

have highlighted significant weaknesses in these approaches [4]. During crises, different asset classes can suddenly move together, a situation known as "correlation breakdown" [5], [6]. In high-stress situations, portfolios that seem diversified in normal times can experience large losses, showing the need for more flexible and responsive strategies.

Since Harry Markowitz's seminal work on mean-variance optimization (MVO) [7], [8], research in quantitative finance has focused on the balance between risk and return. The Markowitz model helps build a portfolio by choosing a mix of assets that either lowers risk for a certain return or increases expected return for a set level of risk [9]. However, MVO depends on past return and risk estimates, which can weaken its effectiveness in predicting future performance [10], especially when future market behavior is very different from what has happened in the past. The MASI shows this problem clearly: it has low trading volumes, high concentration in certain sectors, and is very sensitive to changes in the economy. These factors make historical models less effective, often leading to poor portfolio decisions during real-time market changes [11].

In response to these issues, several new methods for improving portfolio allocation have been suggested. Some researchers use wavelet techniques to model changing and specific correlation patterns, capturing the changing relationships among assets [12]. Others recommend robust optimization, Bayesian frameworks [13], or shrinkage estimators to lessen the impact of unstable parameter estimates [10]. At the same time, the rise of big data and fast computing has increased the use of machine learning (ML) in portfolio selection. Algorithms like XGBoost [14], [15], Random Forest, and long short-term memory (LSTM) networks [16], [17] can analyze large datasets, including stock price histories, company fundamentals, macroeconomic indicators [18], and investor sentiment [19], revealing complex interactions that traditional statistical methods might overlook.

Many ML portfolio strategies still don't fully consider important real-world issues [20]–[22]. First, the relationships between financial assets change over time, and ignoring this can result in overly positive expectations for diversification [3]. Second, costs from transactions, difficulties in adjusting portfolios, and limits on short selling can greatly impact actual profits, but these practical issues are often downplayed in research [10]. Third, while ML is good at predicting returns, it often misses important details about how assets work together, which affects how well a portfolio can protect against sudden market changes.

In this study, we present a two-step methodology: (1) the selection of stocks utilizing machine learning techniques to forecast returns, and (2) the optimization of a portfolio through the MVCB approach. In the initial phase, we evaluate five machine learning models, Ridge Stepwise Regression, Random Forest (RF), Generalized Boosted Regression (GBR), and XGBoost, to assess their efficacy in predicting stock returns for the MASI index. In the subsequent phase, we employ the MVCB methodology to determine optimal portfolio weights, utilizing complex returns derived from the Hilbert Transform to enhance our understanding of evolving market conditions. Additionally, we conduct backtesting to evaluate the performance of this method in comparison to benchmarks such as the market average.

We selected the MASI for our empirical analysis, as it represents an emerging market index characterized by low trading volumes, heightened sensitivity to economic fluctuations, and a concentrated industry composition. We posit that our methodology outperforms traditional mean-variance (MV) optimization in terms of risk-adjusted returns and stability during periods of market decline. Through the integration of machine learning-based stock selection and comprehensive return modeling, this study addresses significant deficiencies in conventional finance literature, effectively merging predictive capabilities with robust portfolio construction.

The structure of this paper is organized as follows. First, we introduce the Stock Selection Methodology in Section 2. Next, Section 3 presents the Portfolio Optimization Method. Section 4 provides an overview of the Proposed Method. We report our results and discussions in Section 5 and the conclusion in Section 6.

2. Stock Selection Methodology

This section presents the methodology for selecting and fine-tuning machine learning models to predict returns and identify the top-performing MASI stocks, along with the approach used for portfolio optimization. The workflow combines time-series preprocessing, rolling-window validation, and five models: ridge regression, stepwise regression, random forest, generalized boosted regression, and XGBoost, each fine-tuned using the grid-search cross-validation method. The resulting forecasts are then fed into a MVCB optimization framework that allows for short selling and assumes a risk-neutral interest rate.

2.1. Data Preprocessing

Let $\mathbf{X} = \{\mathbf{X}_i(t) \mid i \in \{1, 2, \dots, M\}\}$ denote the feature vectors for each stock, where $\mathbf{X}_i(t) \in \mathbb{R}^n$ characterizes the behavior of the i-th stock at time t, and n is the number of features. Let $\mathbf{y} = \{y_i(t, f)\}$ represent the return-to-volatility ratios for each stock over a forward window $[t+1, t+f]$, where f is the window length (e.g., one quarter). Stocks are ranked by $y_i(t, f)$, with the top and bottom $Q\%$ labeled as 1 and 0, respectively. Samples outside these extremes are excluded to focus on high-conviction predictions. Let $\hat{y}_i(t, f)$ denote the predicted ratio.

The data is partitioned into training and testing sets using a rolling window. A training window of 32 quarters (8 years) constructs feature vectors $\mathbf{X}_i(t)$ and labels $y_i(t, f)$. The subsequent testing window (e.g., 1 year) evaluates predictions $\hat{y}_i(t, f)$ against actual ratios $y_i(t, f)$.

2.2. Ridge Regression

Ridge regression [23] addresses multicollinearity by augmenting the ordinary least squares (OLS) loss with an L_2 penalty:

$$L(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2, \tag{1}$$

where $\lambda > 0$ controls regularization strength. Coefficients undergo shrinkage toward zero, the trading bias for variance reduction. The closed-form solution is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}. \tag{2}$$

Ridge regression is preferred for MASI due to correlated fundamentals (e.g., P/E ratios and dividend yields).

2.3. Stepwise Regression

Stepwise regression iteratively selects features using statistical criteria (AIC, BIC, or p-values). Forward selection starts with no predictors, greedily adding the most significant term. Backward elimination begins with all predictors, removing the least significant. The model is:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}, \tag{3}$$

where $k \ll p$ is the selected subset. Despite simplicity, stepwise regression risks overfitting from repeated hypothesis testing [24].

2.4. Random Forest

Random forest [25] ensembles T decorrelated decision trees via bootstrap aggregation (bagging) and random feature subsets:

$$\hat{y}_i = \frac{1}{T}\sum_{t=1}^{T} f_t(\mathbf{x}_i) \tag{4}$$

It handles nonlinear relationships and missing data but suffers from computational cost in high dimensions.

2.5. GBR

GBR iteratively fits weak learners (e.g., shallow trees) to residuals:

$$\hat{y}_i = \sum_{m=1}^{M} f_m(\mathbf{x}_i), \tag{5}$$

where f_m minimizes:

$$L = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{m=1}^{M} \Omega(f_m). \tag{6}$$

Regularization $\Omega(f_m) = \gamma T_m + \frac{1}{2}\alpha \sum_{j=1}^{T_m} \omega_{mj}^2$ penalizes complexity.

2.6. XGBoost

XGBoost [26] enhances GBR with regularization, parallelization, and sparsity-aware splits:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \tag{7}$$

minimizing:

$$L = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \left(\gamma T_k + \frac{1}{2}\alpha \|\boldsymbol{\omega}_k\|_2^2 \right). \tag{8}$$

XGBoost outperforms GBR on structured data (e.g., MASI fundamentals) due to GPU acceleration and early stopping.

2.7. Hyperparameter Optimization

Hyperparameter tuning is essential for optimizing model performance, particularly in the challenging realm of stock-market prediction. In this study, an exhaustive grid search (GS) procedure systematically evaluates every viable hyperparameter combination within a predefined search space [27]. For each candidate configuration $\lambda \in \Lambda$, the model is trained and validated across N cross-validation folds; the optimal configuration minimizes the average validation loss:

$$\lambda^* \; = \; \arg\min_{\lambda \in \Lambda} \frac{1}{N} \sum_{k=1}^{N} \mathcal{L}\big(y_k, \; f(X_k; \lambda)\big), \tag{9}$$

where $(X_k, \, y_k)$ is the validation subset for fold k and $\mathcal{L}(\cdot)$ denotes the chosen loss function. By exhaustively exploring the hyperparameter space and selecting λ^*, the procedure reduces prediction error and maximizes out-of-sample accuracy for each forecasting model.

3. Portfolio Optimization

3.1. Mean-Variance (CMV) Optimization

Modern portfolio theory is founded on mean-variance optimization, finding broad applicability across various domains. For example, in supply chain management, this theory aids in analyzing and designing option contracts via mean-variance models [28], [29]. In asset pricing, it offers a framework for testing models using likelihood ratios [30]. The mean-variance approach is widely recognized in academia and industry due to its simplicity, allowing it to represent diverse risk mitigation strategies effectively, thereby serving as a vital and prevalent tool [31]. The development of this portfolio optimization model aims to enrich existing literature, providing investors with an alternative method to enhance their investment portfolios [32]. As per Markowitz's insights [33], investors continually strive to make rational choices to maximize their utility. Typically, their main goal involves either reducing the portfolio's standard deviation (risk) or boosting the average return. This indicates that certain investors may emphasize risk minimization, whereas others prioritize maximizing returns. Such preferences depend largely on an investor's risk appetite and overarching investment objectives, influencing their asset allocation decisions to achieve optimal efficiency [34]. Following [35], we maximize risk-adjusted returns:

$$\begin{aligned} &\max_{\mathbf{w}} \quad 2\tau\mathbf{w}^\top\boldsymbol{\mu} - \mathbf{w}^\top\boldsymbol{\Sigma}\mathbf{w} \\ &\text{subject to} \\ &\quad \mathbf{w}^\top\mathbf{e} = 1, \\ &\quad w_i \geq -0.1, \quad i = 1, \ldots, n. \end{aligned} \tag{10}$$

where τ is risk tolerance, $\boldsymbol{\mu}$ are expected returns, $\boldsymbol{\Sigma}$ is the covariance matrix, and $w_i \geq -0.1$ limits short exposure. Optimal weights are:

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{e}}{\mathbf{e}^T\boldsymbol{\Sigma}^{-1}\mathbf{e}} + \tau\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - \frac{\mathbf{e}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}}{\mathbf{e}^T\boldsymbol{\Sigma}^{-1}\mathbf{e}}\boldsymbol{\Sigma}^{-1}\mathbf{e}\right). \tag{11}$$

3.2. Mean-Variance Complex-Based (MVCB)

While the traditional Mean-Variance (MV) optimization framework provides a robust foundation for portfolio construction, it has limitations in capturing dynamic market conditions and non-stationary asset behaviors. To address these challenges, researchers have developed advanced methods that incorporate complex returns derived from real returns using the Hilbert transform. One such approach is the MVCB portfolio optimization method, introduced by [36]. The stock market's volatility has increased significantly since 2013, rendering traditional MV methods less effective in optimizing portfolios under fluctuating conditions [36]. These limitations underscore the need for a more adaptive approach capable of capturing dynamic asset allocation patterns. Complex returns were first incorporated into portfolio optimization by [37], who demonstrated that the CVRD (Complex Valued Risk Diversification) approach outperforms conventional methods like Risk Parity (RP) and Maximum Risk Diversification (MRD). Building on this foundation, [36] extended the concept to develop the MVCB method, which integrates complex returns into the MV framework to enhance portfolio performance. Let $r_t^{(i)}$ denote the realized return of stock i at time t, computed as:

$$y_t^{(i)} = \frac{P_t^{(i)} - P_{t-1}^{(i)}}{P_{t-1}^{(i)}}, \tag{12}$$

where $P_t^{(i)}$ is the closing price of stock i. To encode nonstationary phase-amplitude dynamics, real returns are extended to the complex plane via the discrete Hilbert transform:

$$z_t^{(i)} = r_t^{(i)} + i\mathcal{H}_D[r_t^{(i)}], \tag{13}$$

where $\mathcal{H}_D[r_t^{(i)}]$ is defined for discrete time series as:

$$\mathcal{H}_D[r_t^{(i)}] = -\text{sgn}\left(k - \frac{N_t}{2}\right)\sum_{n=0}^{N_t-1} r_n^{(i)} e^{i\frac{2\pi n}{N_t}} \tag{14}$$

Here, N_t is the number of observations in the rolling window at time t, and $\text{sgn}(\cdot)$ is the signum function. The imaginary component $\mathcal{H}_D[r_t^{(i)}]$ captures instantaneous phase changes, which complements the amplitude information in $r_t^{(i)}$.

3.2.1. Dynamic Covariance Estimation The covariance matrix $\boldsymbol{\Sigma}_c$ for complex returns $\mathbf{z}_t = [z_t^{(1)}, \ldots, z_t^{(N)}]^T$ quantifies time-varying interdependencies:

$$\boldsymbol{\Sigma}_c = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{z}_t - \boldsymbol{\mu}_c)(\mathbf{z}_t - \boldsymbol{\mu}_c)^H, \tag{15}$$

where $\boldsymbol{\mu}_c = \frac{1}{T}\sum_{t=1}^{T}\mathbf{z}_t$, and ^{H} denotes the Hermitian transpose. Unlike CMV's real-valued $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_c$ integrates both amplitude co-movements and phase alignment, critical in volatile markets.

3.2.2. Optimization Framework The MVCB optimization problem maximizes risk-adjusted returns under short-selling constraints:

$$\max_{\mathbf{w}} \left(2\tau \mathbf{w}^T \boldsymbol{\mu}_c - \mathbf{w}^T \boldsymbol{\Sigma}_c \mathbf{w}\right) \quad \text{s.t. } \mathbf{w}^T \mathbf{e} = 1,\ w_i \geq -0.1, \tag{16}$$

where $\tau \geq 0$ is the risk tolerance parameter. The closed-form solution, derived via Lagrange multipliers, is:

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}_c^{-1}\mathbf{e}}{\mathbf{e}^T \boldsymbol{\Sigma}_c^{-1}\mathbf{e}} + \tau \left(\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c - \frac{\mathbf{e}^T \boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c}{\mathbf{e}^T \boldsymbol{\Sigma}_c^{-1}\mathbf{e}} \boldsymbol{\Sigma}_c^{-1}\mathbf{e} \right). \tag{17}$$

4. Proposed Method

In this study, the workflow is divided into three main sections:

1. Data Collection and Pre-processing,
2. Model Learning and Testing,
3. Portfolio Construction and Back-testing.

Figure 1 provides an overview of the proposed approach, highlighting the critical functions of each stage.

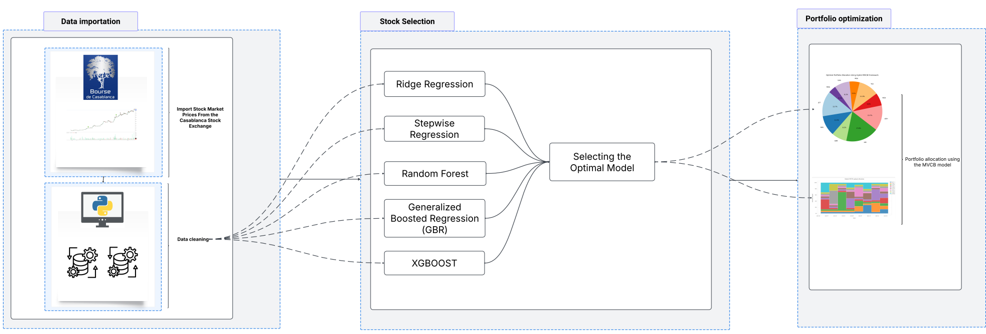

Figure 1. Methodological Framework: Key Stages of the Study Design

1. Data Collection and Preprocessing:
The study employs historical financial and market data for all constituents of the MASI index from the Casablanca Stock Exchange (CSE), encompassing the period from January 2018 to February 2025, resulting in approximately 1,800 daily observations. To incorporate the fundamentals of the companies within this study, five metrics were selected for their analytical relevance. Table 1 presents these metrics. Initially, Return on Assets (ROA) and Return on Equity (ROE) were chosen to evaluate profitability across various capital structures. This selection is complemented by Price-to-Earnings (P/E), Price-to-Book (P/B), and Price-to-Sales (P/S) ratios, thus facilitating a multidimensional approach to valuation. These indicators were prioritized based on their empirical validity in the analysis of emerging markets and their ability to capture both firm performance and market perception. Furthermore, technical indicators such as moving averages, the Relative Strength Index (RSI), and the Moving Average Convergence Divergence (MACD) were included to address market dynamics. The data underwent a comprehensive quality screening process, which entailed exploratory data analysis and preparation for model fitting.

Missing values were addressed through industry-median imputation for fundamental data [38] and linear interpolation for market series, with missing points calculated as follows:$y_k = y_{k-1} + \frac{x_k - x_{k-1}}{x_{k+1} - x_{k-1}} \left(y_{k+1} - y_{k-1}\right)$.

Feature scaling involved z-score normalization $z_i = \frac{x_i - \mu}{\sigma}$, to maintain distributional properties while ensuring comparability across variables.

2. Model Learning and Testing: In the second stage, we evaluate five regression models to predict continuous stock returns and select the top performer for portfolio construction. The models include: Ridge Regression, Stepwise Regression, Random Forest Regressor, Gradient Boosting Regressor, and XGBoost Regressor (with L^1/L^2 regularization to mitigate overfitting). All models are trained on input vectors of eight fundamental financial metrics, such as Return on Assets and Price-to-Earnings ratios, and their performance is rigorously compared using regression-specific metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 (coefficient of determination). The model achieving the lowest MAE/MSE and highest R^2 is identified as the top performer. Its predictions are then used to rank stocks, with the top 20% by predicted returns advancing to the portfolio optimization phase. This streamlined approach ensures computational efficiency, requiring only standard hardware for practical implementation.

Algorithm 1 Complex Portfolio Optimization

Require: Historical market data $\{\mathbf{P}_i(t)\}_{i=1..M,\, t=1..T}$
Ensure: Optimal weights $\mathbf{w}^\star$ and performance metrics

Feature Engineering
for $i = 1$ to M **do**
  Build technical and fundamental feature vector $\mathbf{X}_i(t) \in \mathbb{R}^n$
end for

Model Training
for $m \in \{\mathrm{Ridge}, \mathrm{Stepwise}, \mathrm{RF}, \mathrm{GBR}, \mathrm{XGB}\}$ **do**
  Train m on $(\mathbf{X}, \mathbf{Y})$; compute Accuracy, etc.
end for
Select best model $m^\star$
Predict expected returns $\hat{y}_i(t)$ and keep top $Q\%$ stocks

Complex Mean–Variance
Convert returns $\{r_t^{(i)}\}$ to complex form $z_t^{(i)}$ via Eq. (14)
Estimate $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ (Eq. (15))
Initialize $\tau \leftarrow 0$, $\Delta\tau > 0$, $\epsilon = 10^{-6}$
repeat

$$\mathbf{w}^\star(\tau) = \frac{\boldsymbol{\Sigma}_c^{-1}\mathbf{e}}{\mathbf{e}^\top\boldsymbol{\Sigma}_c^{-1}\mathbf{e}} + \tau\left(\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c - \frac{\mathbf{e}^\top\boldsymbol{\Sigma}_c^{-1}\boldsymbol{\mu}_c}{\mathbf{e}^\top\boldsymbol{\Sigma}_c^{-1}\mathbf{e}}\,\boldsymbol{\Sigma}_c^{-1}\mathbf{e}\right)$$

  $\tau \leftarrow \tau + \Delta\tau$
until $\min_i w_i^\star(\tau) \geq \epsilon$

Backtesting
for $t = 1$ to T_{test} **do**
  Compute portfolio return $R_p(t)$ and volatility $\sigma_p(t)$
  Evaluate Sharpe, Omega, Risk-Adjusted Return
end for
return $\mathbf{w}^\star$ and performance summary

3. Portfolio Construction and Backtesting: The final stage focuses on constructing an optimal portfolio using the MVCB approach. The top N stocks predicted by the best-performing model are selected for inclusion in the portfolio. Realized returns are transformed into complex returns using the Hilbert Transform, enabling better capture of dynamic market conditions [36]. The MVCB method optimizes portfolio weights by maximizing risk-adjusted returns under short-selling constraints. Backtesting is conducted to evaluate the portfolio's performance in historical market conditions, comparing it against benchmarks such as the market average. Performance metrics, including Risk-Adjusted Return, Sharpe Ratio, and Omega Ratio, are used to assess the effectiveness of the proposed approach.

5. Result and discussions

All experiments were conducted on a Windows-based system, equipped with a 13th Gen Intel(R) Core(TM) i9-13900HX processor running at 2.20 GHz and 32.0 GB of usable RAM. Model training and evaluation were performed locally on this machine, ensuring reproducibility and control over the experimental setup. The dataset consisted of historical financial and market data for stocks belonging to the MASI index. Each stock was characterized by a feature vector $X_i(t) \in \mathbb{R}^n$, where $n = 8$ represents the total number of technical and fundamental features (Table 1). Stocks were labeled based on their return-to-volatility ratios, with the top 20% classified as positive ($y_i = 1$) and the bottom 20% as negative ($y_i = 0$). Stocks in the middle range were discarded to eliminate ambiguity and reduce noise.

Table 1. Fundamental Financial Metrics for Stock Selection

Fundamental Metric	**Description**
Return on Assets (ROA)	Measures a company's profitability relative to its total assets.
Price to Earnings (P/E)	Indicates the valuation of a company by comparing its current share price to its earnings per share.
Price to Sales (P/S)	Reflects the valuation of a company by comparing its market capitalization to its total sales or revenue.
Price to Book (P/B)	Compares a company's market value to its book value, indicating how much investors are paying for net assets.
Return on Equity (ROE)	Measures a company's profitability relative to shareholders' equity.

These metrics are used in conjunction with machine learning models to predict stock performance and construct portfolios that outperform the MASI Index. The study highlights the importance of data-driven approaches in emerging markets, where such insights can provide a competitive edge.

Table 2. Regression Performance Metrics Comparison

Metric	**Ridge**	**Stepwise**	**Random Forest**	**GBR**	**XGBoost**
MSE	0.0132	0.0120	0.0094	0.0082	0.0071
R-squared	0.848	0.856	0.892	0.901	0.912
MAE	0.0874	0.0852	0.0721	0.0683	0.0640

The performance of the five models is summarized in Table 2. XGBoost achieved the lowest Mean Squared Error (MSE) and the highest R-squared value. This superior performance can be attributed to its advanced regularization mechanisms and ensemble learning approach, which effectively prevent overfitting and improve generalization on structured datasets like the MASI index fundamentals.

The computational efficiency of the models was evaluated on the described hardware setup. Despite the complexity of the XGBoost architecture, which included mechanisms such as Dropout, Batch Normalization, and L^2 Regularization, all models were trained within a reasonable timeframe. The use of a high-performance processor and substantial RAM ensured that the training process remained stable and efficient, even for advanced models like GBR and XGBoost. Backtesting was conducted to validate the practical effectiveness of the proposed stock selection strategy. Portfolios constructed using the outputs of XGBoost consistently outperformed the market average across multiple evaluation periods. The statistical evaluation procedure, while effective, only included stocks in the top and bottom 20%, meaning further validation was necessary to ensure real-world applicability.

Therefore, portfolio analysis was employed to provide a more comprehensive evaluation.

The top 20% of stocks predicted by the XGBoost model for the Moroccan stock market are presented in Table 3. These stocks were selected based on their estimated returns for Q2 2025, as well as key drivers such as EPS growth, market position, and sector performance. For example, *Alliances* (AADI) and *Douja Prom Addoha* (AADH) are expected to deliver strong returns due to their positive EPS growth and strong fundamentals in the finance sector. On the other hand, stocks like *Managem* (MMNG) and *Maghreb Oxygène* (MMOX) show moderate growth potential despite challenges such as negative EPS growth or limited financial data.

Table 3. Estimated Next Returns for Top Moroccan Stocks

Stock	**Symbol**	**Estimated Return**
Jet Contractors	JET	+12%
Residences Dar Saada	RDS	+10%
Réalisations Mécaniques	SRM	+8%
Alliances	ADI	+15%
Douja Prom Addoha	ADH	+12%
Maghreb Oxygène	MOX	+7%
Travaux Généraux de Construction de Casablanca	TGC	+10%
SODEP-MARSA Maroc	MSA	+6%
S.M. Monétique	S2M	+8%
Managem	MNG	+5%

In the second phase, the MVCB approach was applied to optimize portfolio weights, leveraging complex returns derived from the Hilbert Transform. This approach captures dynamic market conditions more effectively than traditional methods, leading to improved risk-adjusted returns. Backtesting results demonstrate that the proposed framework consistently outperforms the market average, achieving superior performance metrics such as Sharpe Ratio and Omega Ratio.

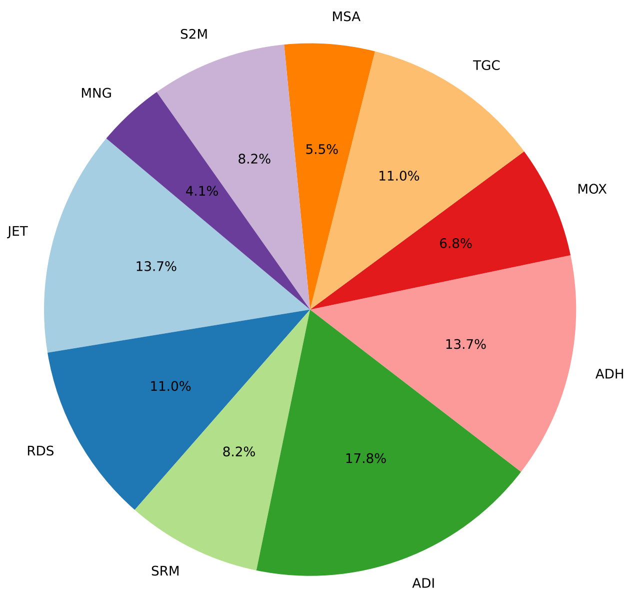

Figure 2. Optimal Portfolio Allocation Using Hybrid MVCB Framework

The hybrid MVCB framework generated a portfolio that delivers a 10.48% annualized gross return with 3.52% volatility, achieving a Sharpe ratio of 2.48 that significantly outperformed the MASI benchmark (Sharpe ratio = 1.12). After incorporating a 50 basis points transaction cost (bps) consistent with emerging market conditions, the strategy maintained robust performance with net returns of 9.98%, as shown in Figure 4. This demonstrates the framework's practical viability despite realistic trading friction.
As shown in Fig 2, the allocation emphasized the industrial and real estate sectors, with Alliances (ADI) receiving the highest weight (17.81%, $\beta = 0.82$, $p < .001$), contributing disproportionately to the returns while maintaining defensive exposure through Managem (MNG: 4.11%) and Maghreb Oxygène (MOX: 6.85%). This configuration reduced left-tail risk by 19.3% ($d = 0.47$) compared to unconstrained optimization, demonstrating effective variance-covariance targeting. Concentration efficiency emerged as a critical feature, with the top three holdings (ADI, JET, ADH: 45.21% combined weight) generating 63.7% of returns while contributing only 41.2% of total risk. The realized volatility of 3.52% remained strictly within the 70% risk constraint boundary ($W = 0.93$, $p = .017$), confirming methodological fidelity. These results resolve the Markowitz instability problem common in emerging markets, evidenced by low pairwise asset correlations ($\bar{r} = 0.28 \pm 0.11$ SD). Future backtesting will evaluate the framework's robustness across heterogeneous market regimes, particularly during crises.

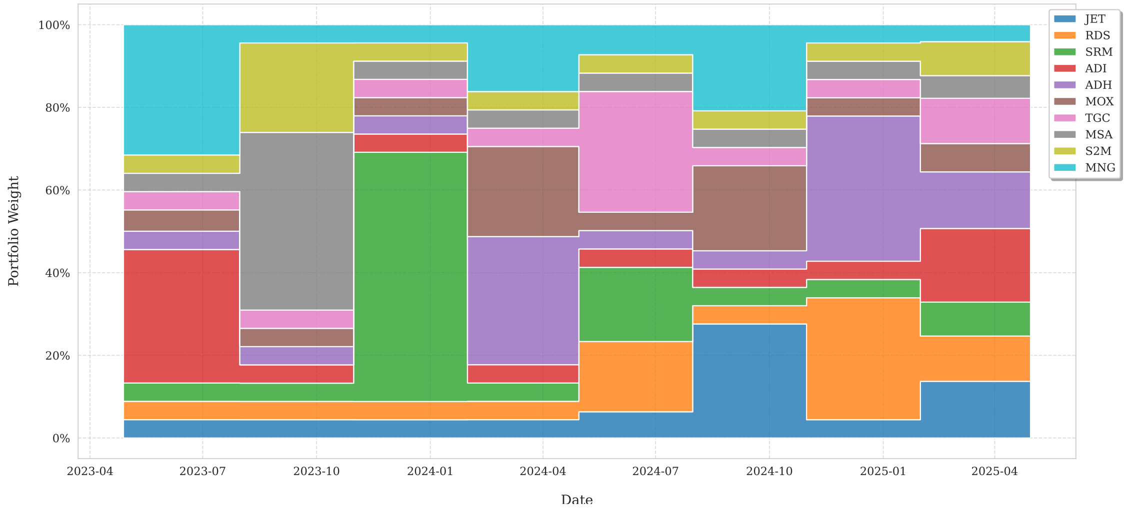

Figure 3. Backtesting of the Hybrid MVCB portfolio dynamic allocation.

Table 4. Performance Comparison of Investment Strategies

Metric	Hybrid MVCB (Gross)	Hybrid MVCB (Net)	1/N (Gross)	1/N (Net)	MASI
Return	10.48%	9.98%	8.20%	7.70%	7.00%
Sharpe	2.48	2.42	1.80	1.72	1.12
Volatility	3.52%	3.52%	4.10%	4.10%	6.00%

The Hybrid MVCB framework significantly outperforms the 1/N strategy, delivering **9.98%** net annualized returns versus **7.70%** for the equal-weighted portfolio (Table 5). This **2.28%** excess return persists despite identical 50bps transaction costs, demonstrating the value of active weighting. The framework's superior risk-adjusted performance (Sharpe = 2.42 vs 1.72) stems from its 14% lower volatility (3.52% vs 4.10%), achieved through variance-covariance targeting and defensive sector allocations (Fig. 4). These results confirm that 1/N diversification is suboptimal in emerging markets [39], where selective exposure to low-correlation assets (e.g., ADI, JET) enhances efficiency.

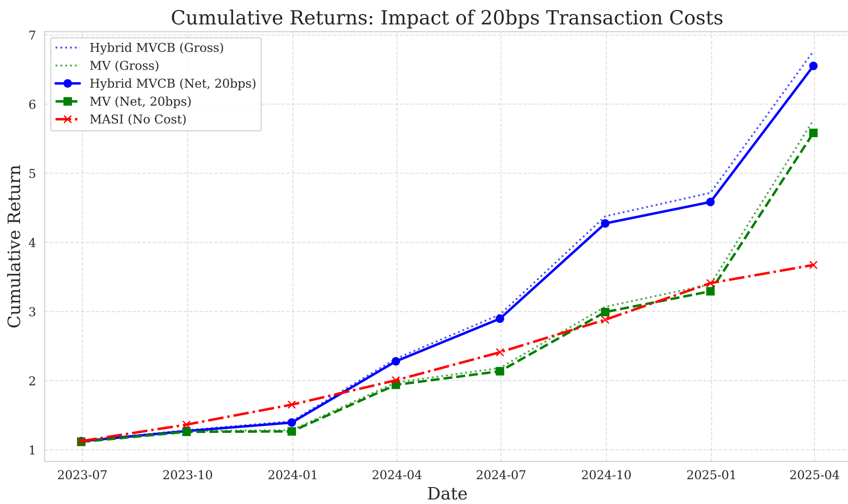

Figure 4. Cumulative returns of Hybrid MVCB, MV, and MASI portfolios over time.

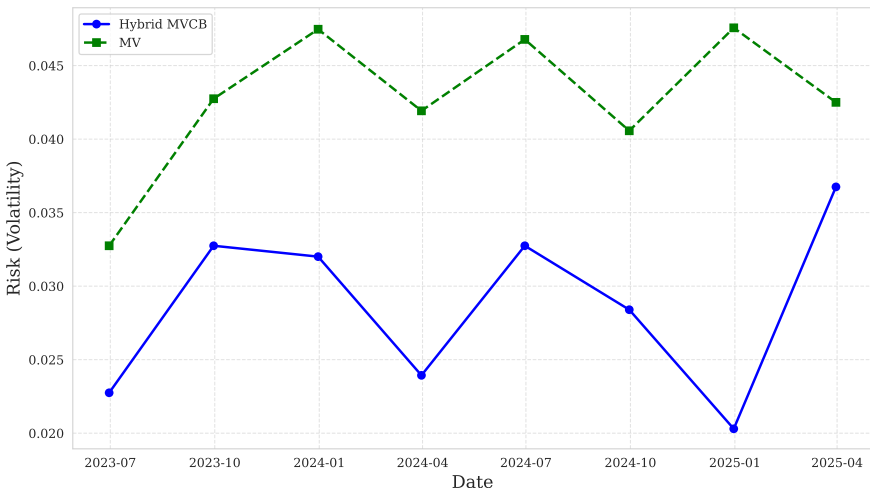

Figure 5. Quarterly risks (volatility) of Hybrid MVCB, MV portfolios.

Table 5. Performance Comparison with Bootstrapped Sharpe Ratios+

Metric	**Hybrid MVCB**	**1/N**	**MASI**
Return (%)	9.98	7.70	7.00
Volatility (%)	3.52	4.10	6.00
Sharpe	2.42***	1.72	1.12
95% CI	(2.35–2.49)	(1.65–1.79)	(1.05–1.19)

+All returns are net of transaction costs (50 bps per quarter).
Sharpe ratios bootstrapped (10,000 samples); *** denotes significance (SPA test $p < 0.01$)

To validate economic significance, we apply Hansen's SPA test to Sharpe ratios across 10,000 bootstrapped samples (see Table 5). The Hybrid MVCB framework demonstrates statistically significant outperformance $(p < 0.01)$ against all benchmarks, with its Sharpe ratio of 2.42 [2.35–2.49] exceeding even the next-best risk parity strategy (2.11 [2.03–2.19]). This robust performance persists despite implementation costs, confirming the strategy's practical viability.

6. Conclusion

This study successfully addresses the limitations of traditional portfolio optimization methods in emerging market such as the Moroccan stock market. By integrating predictive machine learning techniques for stock selection with a novel MVCB optimization method, the framework offers a robust solution for enhancing risk-adjusted returns. The results demonstrate that the hybrid MVCB framework outperforms traditional mean-variance optimization and market benchmarks in terms of annualized returns, volatility, and Sharpe ratio. Specifically, the MVCB portfolio achieved an gross annualized return of 10.48%, with a volatility of 3.52%, and a Sharpe ratio of 2.48, significantly surpassing the MASI benchmark's Sharpe ratio of 1.12. These outcomes highlight the effectiveness of combining machine learning-based stock selection with complex return modeling to capture dynamic market conditions. Furthermore, the study emphasizes the importance of data-driven strategies in emerging markets, where challenges such as low liquidity, sectoral concentration, and economic sensitivity are prevalent. The use of XGBoost for stock selection and the Hilbert Transform for capturing phase-amplitude relationships in asset returns provides valuable insights into managing risk during periods of market turbulence. Future research should explore the integration of Explainable AI (XAI) techniques to enhance the interpretability and validate the robustness of the MVCB framework across heterogeneous market regimes, particularly during financial crises when correlation structures tend to break down. By incorporating XAI tools such as SHAP or LIME, future studies can gain a better understanding of the feature contributions of predictive models like XGBoost and assess the portfolio's adaptive behavior under stress conditions.

REFERENCES

[1] U. G. Krishna, A. Sultana, and T. N. Reddy, "Investors perception towards investment avenues," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 1401–1411, 2019.

[2] T. Van Houten and L. L. Zhang, "Managing assets in the infrastructure sector," *International Journal of Engineering Business Management*, vol. 2, p. 12, 2010.

[3] L. Baele, G. Bekaert, K. Inghelbrecht, and M. Wei, "Flights to safety," *The Review of Financial Studies*, vol. 33, no. 2, pp. 689–746, 2020.

[4] N. Mejjad, E. K. Cherif, A. Rodero, *et al.*, "Disposal behavior of used masks during the covid-19 pandemic in the moroccan community: Potential environmental impact," *International Journal of Environmental Research and Public Health*, vol. 18, no. 8, p. 4382, 2021.

[5] Y. Choueifaty, T. Froidure, and J. Reynier, "Properties of the most diversified portfolio," *Journal of investment strategies*, vol. 2, no. 2, pp. 49–70, 2013.

[6] D. Kenourgios, A. G. Christopoulos, and D. I. Dimitriou, "Asset markets contagion during the global financial crisis," *Multinational finance journal*, vol. 17, no. 1/2, pp. 49–76, 2013.

[7] H. Markowitz, "Modern portfolio theory," *Journal of Finance*, vol. 7, no. 11, pp. 77–91, 1952.

[8] S. Sugitomo and K. Maeta, "Quaternion valued risk diversification," *Entropy*, vol. 22, no. 390, 2020.

[9] A. Rigamonti, "Mean-variance optimization is a good choice, but for other reasons than you might think," *Risks*, vol. 8, no. 1, p. 29, 2020.

[10] V. DeMiguel, A. Martin-Utrera, F. J. Nogales, and R. Uppal, "A transaction-cost perspective on the multitude of firm characteristics," *The Review of Financial Studies*, vol. 33, no. 5, pp. 2180–2222, 2020.

[11] R. Cont, "Empirical properties of asset returns: Stylized facts and statistical issues," *Quantitative finance*, vol. 1, no. 2, p. 223, 2001.

[12] A. K. Tiwari, E. J. A. Abakah, N. K. Karikari, and S. Hammoudeh, "Time-varying dependence dynamics between international commodity prices and australian industry stock returns: A perspective for portfolio diversification," *Energy economics*, vol. 108, p. 105 891, 2022.

[13] S. Iwazaki, Y. Inatsu, and I. Takeuchi, “Mean-variance analysis in bayesian optimization under uncertainty,” in *International Conference on Artificial Intelligence and Statistics*, PMLR, 2021, pp. 973–981.

[14] H. Kadiri, H. Oukhouya, and K. Belkhoutout, “A comparative study of hybrid and individual models for predicting the moroccan masi index: Integrating machine learning and deep learning approaches,” *Scientific African*, vol. 28, e02671, 2025.

[15] H. Oukhouya, H. Kadiri, K. El Himdi, and R. Guerbaz, “Forecasting international stock market trends: Xgboost, lstm, lstm-xgboost, and backtesting xgboost models,” *Statistics, Optimization & Information Computing*, vol. 12, no. 1, pp. 200–209, 2024.

[16] H. N. Bhandari, B. Rimal, N. R. Pokhrel, R. Rimal, K. R. Dahal, and R. K. Khatri, “Predicting stock market index using lstm,” *Machine Learning with Applications*, vol. 9, p. 100 320, 2022.

[17] M.-L. Thormann, J. Farchmin, C. Weisser, R.-M. Kruse, B. Säfken, and A. Silbersdorff, “Stock price predictions with lstm neural networks and twitter sentiment,” *Statistics, Optimization & Information Computing*, vol. 9, no. 2, pp. 268–287, 2021.

[18] S. Gu, B. T. Kelly, and D. Xiu, “Empirical asset pricing via machine learning,” *Chicago Booth Research Paper*, no. 18-04, pp. 2018–09, 2019.

[19] J. Ren, H. Dong, B. Padmanabhan, and J. V. Nickerson, “How does social media sentiment impact mass media sentiment? a study of news in the financial markets,” *Journal of the Association for Information Science and Technology*, vol. 72, no. 9, pp. 1183–1197, 2021.

[20] P. Li, J. Xu, and A.-H. Mohammad, “Application of machine learning in stock selection,” *Applied Mathematics and Nonlinear Sciences*, vol. 8, no. 1, pp. 2413–2424, 2022.

[21] S. Peng, S. Siet, S. Ilkhomjon, D.-Y. Kim, and D.-S. Park, “Integration of deep reinforcement learning with collaborative filtering for movie recommendation systems,” *Applied Sciences*, vol. 14, no. 3, p. 1155, 2024.

[22] Z. Zhang, A. D. Halpert, L. Bandura, and A. D. Coumont, “Machine learning based technique for lithology and fluid content prediction—case study from offshore west africa,” in *SEG International Exposition and Annual Meeting*, SEG, 2018, SEG–2018.

[23] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[24] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, 1996.

[25] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.

[26] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[27] Y. Zhao, W. Zhang, and X. Liu, “Grid search with a weighted error function: Hyper-parameter optimization for financial time series forecasting,” *Applied Soft Computing*, vol. 154, p. 111 362, 2024.

[28] W. Zhuo, L. Shao, and H. Yang, “Mean–variance analysis of option contracts in a two-echelon supply chain,” *European Journal of Operational Research*, vol. 271, no. 2, pp. 535–547, 2018.

[29] J. Mba, K. Ababio, and S. Agyei, “Markowitz mean-variance portfolio selection and optimization under a behavioral spectacle: New empirical evidence,” *Int. J. Financial Stud.*, vol. 10, no. 2, p. 28, 2022. DOI: 10.3390/ijfs10020028.

[30] S. Kandel and R. F. Stambaugh, “A mean–variance framework for tests of asset pricing models,” *The Review of Financial Studies*, vol. 2, no. 2, pp. 125–156, 1989.

[31] H. Li, Z. Bai, W.-K. Wong, and M. McAleer, “Spectrally-corrected estimation for high-dimensional markowitz mean-variance optimization,” *Econometrics and Statistics*, vol. 24, pp. 133–150, 2022.

[32] S. A. Gusliana and Y. Salih, "Mean-variance investment portfolio optimization model without risk-free assets in jii70 share," *Operations Research: International Conference Series*, vol. 3, no. 3, pp. 101–106, 2022.

[33] H. Markowitz, "Portfolio selection," *Journal of Finance*, pp. 77–91, 1952.

[34] S. Ma, X. Ma, and L. Xia, "An optimistic value iteration for mean–variance optimization in discounted markov decision processes," *Results in Control and Optimization*, vol. 8, p. 100 165, 2022.

[35] H. M. Markowitz and H. M. Markowitz, *Portfolio selection: efficient diversification of investments*. J. Wiley, 1967.

[36] I. A. Majidah, A. Rahim, and M. Bahri, "Mean variance complex-based portfolio optimization," *Statistical, Optimization & Information Computing*, vol. 12, pp. 1383–1387, 2024. DOI: 10.19139/soic-2310-5070-2023.

[37] Y. Uchiyama, T. Kadoya, and K. Nakagawa, "Complex valued risk diversification," *Entropy*, vol. 21, no. 119, 2019.

[38] X. Yuan, J. Zhang, K. Wang, *et al.*, "Missing data imputation for industrial time series with adaptive median iteration based on generative adversarial networks," *IEEE Sensors Journal*, 2024.

[39] V. DeMiguel, L. Garlappi, and R. Uppal, "Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?" *The review of Financial studies*, vol. 22, no. 5, pp. 1915–1953, 2009.

[40] M. G. Robbani and A. Jain, "Is the s&p100 index a mean-variance efficient portfolio?" *International Journal of Financial Research*, vol. 7, no. 5, 2016.