Exploring and Contrasting Classical and Machine Learning Based Portfolio Optimization Strategies

By: Tejal Borade, Louise Henningsson, Jharold Montoya, Chinmay Thakre and Viet Tran

Boston University
Professor Andrew Lyasoff
MF730 Dynamic Portfolio Theory
Date: December 2023

Contents

1 Introduction

Portfolio optimization is the process of selecting an optimal portfolio according to some objective and has played a pivotal role in finance, specifically with the introduction of modern portfolio theory by Harry Markowitz in 1952. Markowitz's model assumes that an investor's objective is to maximize expected returns given some level of risk - this model leads to the efficient frontier. Each point on this frontier represents a well-diversified portfolio that achieves the optimal balance between desired expected return and allowable risk. Notably, any portfolio yielding a higher expected return must inherently carry excessive risk, introducing a crucial trade-off in portfolio optimization.

However, as financial landscapes evolve, the mathematical challenges of optimizing portfolios become more pronounced, particularly when dealing with non-Gaussian return distributions. Despite the efficiency of the Markowitz model, its limitations in capturing higher moments of return distribution become evident, especially in the presence of volatility.

Against this historical backdrop, this academic paper seeks to explore and contrast classical and machine learning-based portfolio optimization strategies. Delving into the intricacies of both traditional mathematical models and machine learning approaches, we aim to provide insights into the dynamic interplay between financial theory and technological advancements in the pursuit of effective asset portfolio management.

2 Exploratory Data Analysis

For our portfolio optimization analysis, we gathered daily adjusted close price and volume data for various ETFs over a 10-year period starting from 12/09/2013 to 12/09/2023. The data collected made up of two portfolios on which we performed the various optimization techniques. Each portfolio was curated to represent distinct sectors and regional exposures.

The first portfolio, comprising the ETF tickers [EWC, IVV, EWO, EWK, EDEN, EFNL, EWQ, EWG, EIRL, EIS, EWI, EWN, ENOR, EWP, EWD, EWL, EWU, EWA, EWH, EWJ, ENZL, EWS], offers a comprehensive representation of international equity markets. These tickers span a range of global regions, including Canada (EWC), the United States (IVV), Austria (EWO), Belgium (EWK), and many others. By incorporating diverse geographical exposures, this portfolio aims to capture opportunities and manage risks associated with a broad spectrum of international markets.

The second portfolio is constructed with a focus on major sectors within the United States, comprising the ETF tickers [XLY, XLP, XLE, XLF, XLV, XLI, XLB, XLK, XLU]. This portfolio emphasizes sector-specific dynamics within the U.S. equity market. From Consumer Discretionary (XLY) to Utilities (XLU), each ticker represents a distinctive sector, offering insights into the performance and optimization challenges specific to the domestic economic landscape.

The inclusion of various sectors and regions in the global equity portfolio as well as the sectoral concentration in the U.S. equity portfolio enables a nuanced analysis of portfolio optimization techniques in the context of global market dynamics as well as U.S. sectoral trends and market conditions.

3 Optimization Models

In our project, we chose to explore four different optimization strategies to construct a portfolio. The purpose of portfolio optimization is to find the optimal weights to allocate to a diversified set of assets to build a portfolio satisfactory to a set of objectives decided by the investor. The most well-known objective of portfolio optimization is to maximize the expected return of the portfolio given the level of risk an investor is willing to take. We will examine the most widely used optimization technique combined with three different ways of optimizing a portfolio and compare our optimization results to each other and to the results gathered from using machine learning techniques.

3.1 Maximize Portfolio Returns

The first optimization model that was tested was a strategy where we maximized the portfolio returns given a level of risk. The mathematical formula for this optimization problem is the following:

$$R_p = \Sigma_{i=1}^{n} w_i \cdot R_i$$

$$\Sigma_{i=1}^{n} w_i = 1$$

subject to:

$$0 \leq w_i \leq 1 \; for i = 1, 2, ..., n$$

Here, n represents the number of assets in our portfolio and R represents the matrix of returns that we receive from our assets. Furthermore, we implemented the constraint that all asset weights need to be positive, i.e., no short selling is allowed in this optimization problem.

3.2 Minimize Variance

The second optimization model tested was a model to minimize the portfolio variance. This optimization focuses on minimizing the risk of our given portfolio based on a level of returns. The formula implemented to minimize the portfolio variance has the following solution:

$$w = argmin(w^T \Sigma w)$$

$$\Sigma_{i=1}^{n} w_i = 1$$

subject to:

$$0 \leq w_i \leq 1 \ for i = 1, 2, ..., n$$

Where w is the vector of weights and Σ is the covariance matrix. $w^T \Sigma w$ then represents the portfolio variance that we are trying to minimize. In this minimum variance portfolio, we also implemented a constraint of no short selling of our assets.

3.3 Maximize Sharpe Ratio

Next, an optimization model to maximize the Sharpe ratio of the portfolio was used. Maximizing the Sharpe ratio involves both maximizing the expected returns and minimizing the variance, i.e., the risk of the portfolio. Where the Sharpe ratio we want to maximize is given by:

$$Sharpe = (\mu^T w - r_f)/\sqrt{w^T \Sigma w}$$

$$\Sigma_{i=1}^{n} w_i = 1$$

subject to:

$$0 \leq w_i \leq 1 \ for i = 1, 2, ..., n$$

Similarly, $w^T \Sigma w$ represents the variance of the portfolio and $\mu^T w - r_f$ represents the return less the risk-free rate of the portfolio. The portfolio is subject to the constraint of no short selling as well and the weights have to sum to 1.

3.4 Maximize Decorrelation

Lastly, we implemented an optimization model that would maximize the de-correlation between our assets. The fundamental assumptions of this model imply that all assets in the portfolio have similar returns and volatilities but different correlations. This method to maximize the correlation among our assets is a way to further diversify away market risk in which our returns will not be as susceptible to fluctuations in the market. Under specific conditions when the assets in a portfolio have the same volatility, this model will maximize the diversification ratio. Similar to the minimum variance model, the closed-form solution is as follows:

$$w = argmin(w^T Aw)$$

$$\Sigma_{i=1}^{n} w_i = 1$$

subject to:

$$0 \leq w_i \leq 1 \; fori = 1, 2, ..., n$$

Where w is the vector of weights and A is the correlation matrix. The de-correlation maximizing model is closely related to the minimize variance model, however, the minimization of variance is performed on the correlation matrix as opposed to the covariance matrix due to the heterogeneous correlations.

4 Machine Learning Models

4.1 Principal Component Analysis

Principal Component Analysis (PCA) is a mathematical technique used for dimensionality reduction and understanding the underlying structure of data. In the context of portfolio optimization, PCA can help identify principal components that capture the most significant sources of variation in historical returns. Following are the mathematical formulas for PCA and how they relate to portfolio weights calculating the covariance matrix with data on historical returns:

$$\Sigma = \frac{1}{n-1} \times (X - \bar{X})^T \times (X - \bar{X})$$

In this equation, X is a $m \times n$ matrix of the historical returns where m represents the number of assets, n represents the number of time periods, and $\bar{X}$ is the mean vector of the returns. The covariance matrix Σ is then decomposed into its eigenvalues and eigenvectors.

$$\Sigma v = \lambda v$$

$$PC_i = v_i$$

Where v is the eigenvector of Σ and λ represents the corresponding eigenvalues. Further, PC further represents eigenvectors of the covariance matrix and the direction of maximum variance where the i-th PC is the principal component.

After obtaining the principal components through the previous calculations, the following calculation is used to calculate the optimal portfolio weights:

$$optimal\_weights = \frac{PC_{selected}}{\sum_i |PC_{selected}|}$$

Here, $PC_{selected}$ is the selected principal component (in this case, PC_1 specified by pc_id) and $\sum_i |PC_{selected}$ calculates the sum of absolute values of the selected principal component's elements used for normalization. The portfolio weights obtained through this calculation represent a linear combination of assets based on the selected principal component's relationship with the historical returns, instead of explicitly minimizing risk or maximizing returns.

The optimal weights calculation normalizes the weights based on the absolute values of the selected principal component's loadings, effectively allocating portfolio weights according to the direction of the maximum variance captured by our principal component in previous calculations.

4.2 Hierarchical Risk Parity

The Hierarchical Risk Parity (HRP) method, developed by Marcos Lopez de Prado in 2016, is a portfolio optimization technique that uses a hierarchical clustering algorithm to construct a diversified portfolio. This is done by allocating weights to assets based on their correlation or covariance structure. The goal of the HRP model is to minimize the portfolio's overall risk by effectively distributing the risk among the assets. Similarly to the PCA model, the first step of the HRP model is to calculate the covariance matrix:

$$\Sigma = \frac{1}{n-1} \times (X - \bar{X})^T \times (X - \bar{X})$$

In this equation, X is a $m \times n$ matrix of the historical returns where m represents the number of assets, n represents the number of time periods, and $\bar{X}$ is the mean vector of the returns.

The HRP model then uses hierarchical clustering where the assets are initially grouped into clusters based on their covariance matrix and correlation. Thereafter, within each cluster, the covariance matrix of another measure of risk is calculated using the returns of the assets in that cluster. This next matrix characterizes the risk structure among the specific assets in within that cluster.

The ultimate goal of the HRP model is to determine the weights that achieve risk parity. When achieving risk parity, in our case, the assets in our portfolio contribute equally to the total risk within each cluster with different weights. The risk contribution determines the weight of each cluster to the final portfolio.

4.3 Holt-Winters Smoothing

The Holt-Winters smoothing technique in this research paper leverages exponential smoothing methods to dynamically adjust asset weights in response to historical time series data in order to achieve the best possible risk-adjusted return. Smoothing is a statistical technique used to reduce the impact of random variation in a time series data set by creating a smoothed version of a data set by removing fluctuations and noise to highlight underlying trends. Exponential smoothing is another widely used technique, assigning exponentially decreasing importance (i.e., weights) to past observations and giving more importance to recent data. Holt-Winters exponential smoothing extends this method to handle level (l_t), trend (t_t), and seasonality (s_t) in time series data.

The formula for the forecast is:

$$Y_{t+h} = (l_t + h \cdot b_t) \cdot s_{t+h-m}$$

where h is the forecast horizon, and m is the length of the seasonal cycle.

The model is calibrated using historical data to estimate the initial level (l_0), trend (t_0), and seasonality (s_0) parameters. The smoothing equations are as follows:

$$l_t = \alpha \cdot (Y_t - s_{t-m}) + (1 - \alpha) \cdot (l_{t-1} + t_{t-1})$$

$$t_t = \beta \cdot (l_t - l_{t-1}) + (1 - \beta) \cdot t_{t-1}$$

$$s_t = \gamma \cdot (Y_t - l_t) + (1 - \gamma) \cdot s_{t-m}$$

The parameters , , and are smoothing parameters that control the influence of the recent observations on the corresponding components. They can be adjusted to balance the responsiveness of the model to new data and the smoothness of the resulting forecasts.

The optimized portfolio is then constructed by selecting a subset of assets based on the forecasts generated by the Holt-Winters model, emphasizing assets with favorable risk-return profiles. For each asset within the constructed portfolio, the Holt-Winters smoothing method is applied to the time series data, fitting the model to the data and generating forecasts for the defined horizon. The forecast of each asset is calculated by the difference between the final and initial forecast values, providing insight into the expected performance and directional movement of each asset over the horizon.

The weights assigned to each selected asset are determined based on the forecasted trajectory, reflecting the anticipated future performance of the assets. The methodology concludes by returning the optimized weights for the assets in the portfolio.

4.4 Autoencoder

Autoencoder is a type of artificial neural network used for unsupervised learning and dimensionality reduction. The algorithm is designed as follows:

Encoding: The encoder takes the historical returns data $\mathbf{x}$ and compresses it to a lower-dimensional representation $\mathbf{h}$ by mapping the data through one hidden layer. This compressed representation is known as the bottleneck (or latent space) and captures the most important features of the input data.

$$\mathbf{h} = f_{encoder}(\mathbf{x})$$

Here, $f_{encoder}$ represents the mapping function of the encoder and is defined as a dense layer with ReLU activation.

Decoding: The decoder then takes the compressed representation of the latent space $\mathbf{h}$ and reconstructs the original input data $\hat{\mathbf{x}}$, aiming to minimize the difference between the original returns data and the reconstructed returns using the Mean Square Error (MSE) loss function. The communal information (i.e., the regenerated data) is calculated for each asset in the portfolio.

$$\hat{\mathbf{x}} = f_{decoder}(\mathbf{h})$$

Here, $f_{decoder}$ is the mapping function of the decoder and is defined as another dense layer with linear activation.

These mapping functions define how the autoencoder transforms the input data into a compressed representation and reconstructs it back to the original input.

Finally, the optimal weights for the portfolio are determined based on the communal information and differences between the input data and the reconstructed data. The relative differences between the original returns and their reconstructions by the autoencoder contribute to the calculation of weights, aiming for efficient asset allocation in the portfolio.

5 Methodology

The outline for the methodology employed for portfolio optimization focuses on various strategies and agents implemented to see which portfolio gives better performance than an equal-weighted portfolio which was used as a benchmark.

The methodology implemented started with loading the necessary libraries into the code to read the data. The data collected was loaded into the optimization codes directly through Yahoo Finance. In our code, a class is used to pre-process and clean the data combined with converting the daily returns to log returns.

Once the data was clean, all the different optimization and Machine Learning strategies were defined separately. Each method is discussed more in detail above from which the formulas were implemented according to the maximizing returns, minimizing variance, maximizing Sharpe ratio, maximizing de-correlation, principal component analysis, hierarchical risk parity, Holt-Winters Smoothing, and the autoencoder model. From each of these models, we extracted different optimal weights based on the model's assumption.

The 10-year data collected on the selected ETFs and portfolios were then incorporated into the model where the output was the optimally weighted portfolios under the different scenarios. The results of the different models were represented graphically individually against an equal-weighted benchmark of the portfolios collected. Other performance metrics such as cumulative returns, Sharpe ratio, volatility, alpha, and beta were also calculated for each portfolio. From there we were able to evaluate the various models' performances against the benchmark and each other.

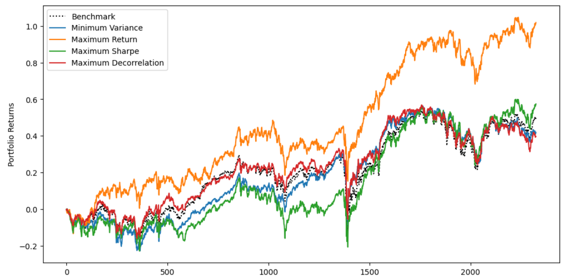
6 Results

In our examination of various portfolio optimization techniques for the Global DM Equity Portfolio, we observe distinct performance metrics across the strategies. Notably, the Maximum Return optimization technique stands out with a mean return of 0.0004 and a Sharpe Ratio of 0.5753, exhibiting its ability to achieve a favorable balance between risk and return. Additionally, the Autoencoder strategy shows strong performance in regards to the Sharpe Ratio at 0.4202, underscoring its effectiveness amongst the machine learning-based approaches to achieve better risk-adjusted returns.

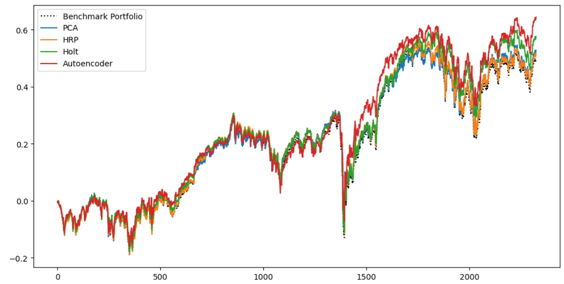
However, it is important to note, that even though our Maximum Return model produced favorable results the weight allocation this model produced allocated the whole weight to one specific ETF. The Maximum Sharpe Ratio and Minimum Variance models produced more diversified portfolios allocating weight to more assets, however, allocating a significant weight to one asset and small weights to the other assets in the portfolio. Comparatively, the Maximum Decorrelation portfolio provided the most diversification among the basic optimization models where the weights of the assets in the portfolio were allocated more evenly, but, even this model excluded more than half of the assets from the portfolio.

It is noteworthy that despite the classical and widely recognized Minimum Variance strategy demonstrating reduced volatility (0.0101), it achieves a slightly lower Sharpe Ratio of 0.2820 compared to the Maximum Return and Autoencoder strategies. This highlights the classic trade-off involved in portfolio optimization, where achieving a maximum return may yield superior risk-adjusted performance.

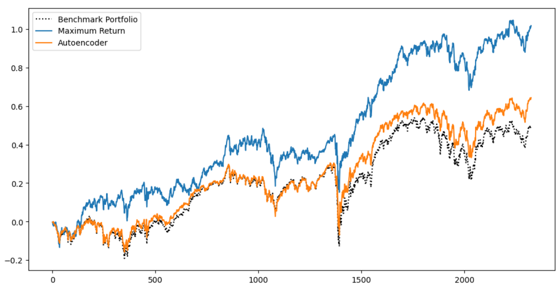
When looking at the results obtained by the Machine Learning models in the Global DM Equity Portfolio, the weight allocation in the portfolio is more diversified with the most equal weights being allocated to the assets in the PCA model. Throughout the Machine Learning models, all of the assets are assigned a weight in the portfolio and therefore included. Notable however, is that even though the Autoencoder model allocates weight to all of the assets in the portfolio, the model allocates most of the weight to one specific ETF with the remaining assets having below 5 percent weight allocation in the portfolio.



(a) Standard Approach



(b) ML Approach



(c) Best from two approaches

Figure 1: Global equity Portfolio

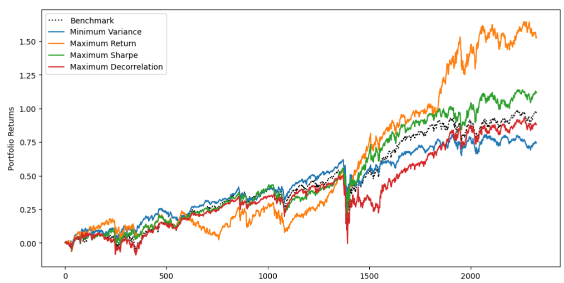
Stats	Benchmark	Minimal Variance	Maximal Returns	Maximal Sharpe	Maximal Decorrelation	PCA	HRP	Holt Smoothing	Autoencoder
Mean	0.0002	0.0002	0.0004	0.0002	0.0002	0.0002	0.0002	0.0002	0.0003
Volatility	0.0112	0.0101	0.0121	0.0116	0.0103	0.0110	0.0114	0.0113	0.0109
Sharpe Ratio	0.3064	0.2820	0.5753	0.3374	0.2644	0.3300	0.3130	0.3526	0.4202
Alpha	0.0000	0.0000	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000	0.0001
Beta	1.0000	0.8351	0.9007	0.9148	0.8767	0.9748	1.0123	0.9947	0.9608

Table 1: Global Equity Portfolio Stats

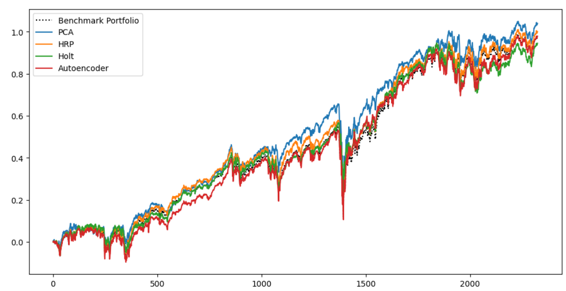
In our analysis of portfolio optimization strategies for the U.S. Sector Equity, the Maximum Sharpe strategy emerged as the standout performer, achieving a Sharpe Ratio of 0.7005. The Autoencoder method also achieved a robust Sharpe Ratio of 0.5451. The classical approach of Minimum Variance demonstrated a reduction in volatility (0.0100) but resulted in a slightly lower Sharpe Ratio of 0.5065 compared to the Maximum Sharpe and Autoencoder strategies. Moreover, the decorrelation-focused strategies (Maximum Decorrelation and PCA), yielded competitive results in regard to the Sharpe Ratio at 0.5015 and 0.6634, respectively.

Similarly to the Global Equity Portfolio, the U.S. Sector Equity Portfolio behaved similarly in our models. The Maximum Return model only allocated equal weight to two of the ETFs in the portfolio while the Minimum Variance and Maximum Sharpe Ratio models allocated weight to more assets in the portfolio, they allocated the majority of the weight to one specific asset in the portfolio. The Maximum Decorrelation model was efficient in allocating the weights of the assets in the portfolio including all but one asset into the portfolio.

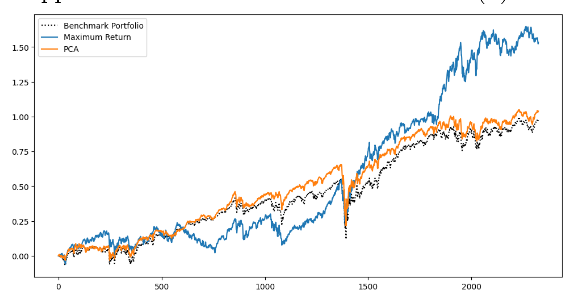
Analyzing the Machine Learning Optimizations, similar to the Global Equity Portfolio, the models were sufficient in their diversification of the weight allocation, including all of the assets in the allocation. In the U.S. Equity Portfolio, the HRP model allocated the majority of the weight to four assets in the portfolio and smaller weights to the remaining assets which was also similar to the Autoencoder model. The PCA model allocated similar weights to all the assets in the portfolio while the Holt-Winters model allocated a large weight to one ETF in the portfolio and smaller weights to the remaining assets.



(a) Standard Approach



(b) ML Approach



(c) Best from two approaches

Figure 2: US equity Portfolio

Stats	Benchmark	Minimal Variance	Maximal Returns	Maximal Sharpe	Maximal Decorrelation	PCA	HRP	Holt Smoothing	Autoencoder
Mean	0.0004	0.0003	0.0007	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004
Volatility	0.0110	0.0100	0.0150	0.0121	0.0120	0.0107	0.0113	0.0112	0.0113
Sharpe Ratio	0.6024	0.5065	0.7005	0.6324	0.5015	0.6634	0.6039	0.5781	0.5451
Alpha	0.0000	0.0000	0.0002	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000
Beta	1.0000	0.8061	0.9813	0.9286	1.0211	0.9363	1.0211	0.9947	1.0056

Table 2: US Sector Equity Portfolio Stats

7 Conclusion

In conclusion, this academic paper explores and contrasts classical and machine learning-based portfolio optimization strategies. We investigate various optimization techniques, ranging from traditional mathematical models to advanced machine learning algorithms, using two distinct portfolios: a Global DM Equity Portfolio and a U.S. Sector Equity Portfolio.

The classical optimization models, including Minimum Variance, Maximum Return, Maximum Sharpe Ratio, and Maximum Decorrelation, demonstrate different characteristics in terms of risk, return, and diversification. Each model has its strengths and weaknesses, with trade-offs between maximizing returns, minimizing risk, and achieving a balanced Sharpe Ratio.

On the other hand, machine learning-based models, such as Principal Component Analysis (PCA), Hierarchical Risk Parity (HRP), Holt-Winters Smoothing, and Autoencoder, bring a new perspective to portfolio optimization. These models leverage data-driven approaches to capture patterns, relationships, and trends in historical returns, offering potential advantages in adapting to changing market conditions.

An important area for improvement in the presented portfolio optimization models is the incorporation of upper-bound constraints on the weights. This adjustment can prevent corner solutions where the optimization assigns extremely high weight to a single asset while allocating negligible weights to others. Including an upper bound constraint on weights is a prudent step to enhance the practicality and feasibility of the portfolio allocation. Such constraints align with real-world investment scenarios where an investor typically avoids concentrating their entire portfolio on a single asset due to diversification and risk management considerations.

Based on these results of both the Global DM Equity Portfolio and the U.S. Equity Portfolio, it is evident that Markowitz's classical approaches, such as the Maximum Return and Maximum Sharpe strategies, continue to play a significant role in portfolio optimization.

Results indicate that different models excel in different aspects. The Maximum Return and Autoencoder models stand out in terms of the Sharpe Ratio in both portfolios, showcasing their potential for achieving superior risk-adjusted returns. The Minimum Variance and Maximum Decorrelation models exhibit effective risk reduction but may sacrifice some returns. The machine learning models, particularly PCA and HRP, demonstrate diversification benefits and provide competitive risk-return profiles.

However, it's crucial to note that the effectiveness of each model depends on the investor's preferences, risk tolerance, and market expectations. Portfolio optimization is inherently a complex task with various factors influencing decision-making. Moreover, the dynamic nature of financial markets requires continuous monitoring and adaptation of portfolio strategies.

In practice, investors should consider a combination of classical and machine learning-based approaches, leveraging the strengths of each to construct robust and adaptive portfolios. Continuous research, back-testing, and refinement are essential for ensuring the reliability and effectiveness of portfolio optimization strategies.

8 References

Alex, Honchar. ”AI for portfolio management: from Markowitz to Reinforcement Learning.” September 28th, 2019. https://medium.com/neurons-lab/ai-for-portfolio-management-from-markowitz-to-reinforcement-learning-cffedcbba566. December 2023.

Conlon, Thomas, John Cotter och Iason Kynigakis. ”Machine Learning and Factor-Based Portfolio Opti mization.” July 8th, 2021. https://arxiv.org/pdf/2107.13866.pdf. December 2023.

Lim, Qing, Qi Cao och Chai Quek. ”Dynamic portfolio rebalancing through reinforcement learning.” 2022. https://link.springer.com/article/10.1007/s00521-021-06853-3. December 2023.

Ranchnog. ”Deep-Portfolio-Management.” September 27th, 2019. GitHub. December 2023.

Reed, Eric. ”A Guide to Portfolio Optimization Strategies.” March 3rd, 2023. https://smartasset.com/investing/guide-portfolio-optimization-strategies. December 2023.

ReSolve Asset Management. ”Portfolio Optimization: Simple Versus Optimal Methods.” https://investresolve.com/portfolio-optimization-simple-optimal-methods/. December 2023.