Machine Learning-based Multi-Asset Portfolio Optimization: Evidence from Stocks, Funds and Bonds

Runzhe Yin

The University of Sydney, Camperdown NSW 2006, Sydney, Australia
ricardoyinrunzhe@gmail.com

Abstract. With the background of significant fluctuations in the global financial market, traditional single investment strategy has demonstrated characteristics such as high volatility, and poor risk diversification. To enhance the stability of investment strategies, increase the rate of return, and address the existing issues of current strategies, this paper will conduct research on diversified investment portfolios through machine learning methods. This article mainly selects six types of financial assets: stocks, government bonds, gold, crude oil, consumer funds, and money market wealth management. By calculating the monthly rate of return, first, using the correlation analysis to get the diversified value of different assets. Then, the assets are classified into low, medium and high-risk categories based on their returns and fluctuations through the K-means algorithm. Meanwhile, using linear regression model is used to predict the stability and future trends of various assets. According to the analysis results, the diversified asset investment model constructed will be compared with a single stock investment model. The experimental results prove that a diversified investment portfolio has lower volatility, and a higher Sharpe ratio, which can effectively diversify risks and enhance the stability of investment returns. This study demonstrates that machine learning can assist in the allocation and screening of investment assets.

Keywords: Machine Learning; Multi-asset; Portfolio Optimization

1 Introduction

The world's financial markets are always influenced by several unpredictable shifts. Due to changes in the economy, policies and market sentiment, asset values fluctuate every day. Because of these changes, market investors are exposed to continuous investment risks. Most ordinary investors choose a direct investment strategy. Usually, they invest all their money in stocks. However, stock assets are highly susceptible to market fluctuations. When the stock market drops, investors will lose a lot of money. This blind investment approach cannot effectively diversify investment risks.

Markowitz proposed mean variance optimization (MVO) in 1952, formally defining the concept of risk return tradeoff, which laid the foundation for the current portfolio theory (MPT). Markowitz demonstrates how investors can construct a portfolio by considering portfolio selection as an optimization problem, so as to minimize the risk

© The Author(s) 2026
L. C. Yoong (ed.), *Proceedings of the 2026 2nd International Conference on Financial Innovation and Marketing Management (FIMM 2026)*, Advances in Economics, Business and Management Research 408,
https://doi.org/10.2991/978-94-6239-801-6_144

at the expected return level or maximize the expected return at a given risk level. This idea has become the cornerstone of the global portfolio construction method, and has had a significant impact on investment academic knowledge and investment management practice [1].

Risk-based allocation techniques, especially risk parity methods, have developed, attempting to circumvent these limitations by focusing on risk contributions rather than return estimates. However, these methods still mainly rely on historical covariance estimation and stable correlation structures - assumptions that often fail during market crises. The global financial crisis in 2008 and the COVID-19 pandemic in 2020 demonstrated the serious limitations of traditional methods. As correlations approached consistency, the risk parity approach witnessed an unexpected decline.

When assessing investment risks, the Markowitz mean-variance optimization model has obvious flaws [2,3]. It defines investment risk with variance, treating favorable upward price fluctuations and losing downward price fluctuations as the same type of risk. The method this model uses to measure risks is inconsistent with the psychology of actual investors, who are only afraid of losses when assets fall and welcome gains when assets rise. In addition, this model is highly sensitive to relevant data factors. Asset returns and related data are used to generate results, and this data is even a little incorrect. This model tends to concentrate funds on a few assets, undermining the goal of diversified investment and reducing the impact of risk aversion. This model also employs a static analysis foundation [4,5]. Although the actual financial market is constantly changing, it assumes that asset returns, volatility and correlations remain unchanged throughout the entire period. When market conditions or crises occur, the relationships among various assets change rapidly, making the model unable to adapt to the actual market changes and resulting in biased analysis results [6].

The above-mentioned machine learning techniques are used in this work to solve the problem of single-asset investment. This study selected six typical financial assets for investigation. These assets include commodities, government bonds, stocks and cash equivalents. This study first examined the correlations among various assets. It confirms whether investing in multiple assets helps to reduce investment risks. Then, the study used K-Means clustering and linear regression techniques to screen and classify assets. It has created a new multi-asset investment portfolio with stable profits and relatively low risks. This study compared two investment techniques to assess the feasibility of this new investment portfolio. In terms of profitability and risk tolerance, it compares a typical single-stock portfolio with a machine learning multi-asset strategy. The research results can prove the benefits of diversified intelligent investment. It can help ordinary investors make safer and more reliable investments.

2 Literature Review

The combination of machine learning and portfolio optimization has become a popular research direction in recent financial studies. Many scholars have verified that data

analysis and machine learning tools can improve traditional asset allocation strategies [7].

First, the study titled Portfolio Optimization by Using Machine Learning and Data Analytics explores the value of basic data analysis in investment research. The author believes that traditional portfolio methods rely too much on fixed financial formulas. These methods cannot adapt to changing market conditions [8]. This research proves that machine learning can process large amounts of asset data. It can help investors analyze asset return rules and hidden risk features. The study confirms that machine learning can effectively improve portfolio optimization results [8]. However, this research focuses more on data analysis. It does not classify assets based on different risk levels.

Second, another paper named A machine learning approach to risk based asset allocation in portfolio optimization focuses on risk-oriented asset allocation. This study uses machine learning to measure the risk degree of different assets. It explains that most investment losses come from unreasonable asset matching [9]. The research suggests that investors should allocate assets according to risk characteristics rather than personal preference. This paper successfully proves that machine learning can reduce portfolio risk [9]. Nevertheless, the study only focuses on risk control. It lacks comprehensive comparison between traditional investment portfolios and optimized portfolios.

Third, Deep Learning for Portfolio Optimization discusses the application of advanced intelligent algorithms in asset management. The research shows that deep learning has strong data fitting ability. It can predict future asset price trends and improve portfolio returns [10]. However, deep learning models require massive data and complex computing conditions. The model operation difficulty is high for ordinary investors. It is not suitable for simple and lightweight portfolio optimization research [10].

Overall, existing studies have proven the effectiveness of machine learning in portfolio optimization. But most studies either focus only on data analysis, single risk assessment, or complex deep learning models. Few studies combine simple clustering algorithms and linear regression models to build multi-asset portfolios. Compared with previous works, this study adopts two practical and stable machine learning methods. It completes asset risk classification and trend analysis. It also conducts detailed backtesting and performance comparison. This paper provides a simpler and more practical portfolio optimization solution for ordinary investors.

3 Methodology

To complete the asset research, portfolio design, and strategy performance verification, this study employed a complete set of quantitative statistical techniques and lightweight machine learning algorithms. For ordinary investors, each step of the testing process is simple, reliable, and repeatable [7]. This paper combines data preprocessing, correlation analysis, K-means clustering, and linear regression methods for in-depth analysis, filling the gaps in previous academic research [11]. The main innovation of

this study lies in its multi-stage, dual-model research framework, which is different from the current single-model or overly complex research methods [12].

This study first finishes standardized preparation and data collecting. Six popular and easily available financial assets from various categories are chosen for this study. Treasury bonds, gold, crude oil, consumer industry funds, SP500 stocks, and cash equivalent assets are among the chosen assets. Compared to the single asset type selection in Portfolio Optimization by Using Machine Learning and Data Analytics[8], this multi-category asset selection encompasses risky assets, defensive assets, and hedging assets. This study uses historical pricing data to determine each asset's monthly return. The noise of daily price fluctuations can be successfully removed by using monthly return data. The study also eliminates anomalous incorrect values and missing data from the dataset. It guarantees the accuracy and dependability of all ensuing analytical and experimental findings[13].

Second, Pearson correlation analysis is used in this work for first asset screening. The degree of relationship between each pair of assets may be plainly seen in the correlation coefficient. During market volatility, the price movements of assets with low or negative correlation may cancel each other out. The core concept of multi-asset investment lies in this risk diversification strategy. Most of the early studies merely briefly mentioned the correlations among assets without providing any effective screening methods. This research utilized the correlation results to guide asset matching, thereby enhancing the scientificity and targeting of portfolio construction [14].

Third, the K-Means clustering approach is used in this work to classify asset risk. This study employs monthly asset returns and annual volatility as dual characteristic indicators, contrasting with the method in [9] which only uses risk as the evaluation criterion. All selected assets are classified into low-risk, medium-risk, and high-risk categories through the K-Means clustering method. High-quality high-risk and medium-risk assets can bring profit opportunities, while low-risk assets serve as defensive tools, helping to maintain the performance of the investment portfolio. This classification method provides precise hierarchical specifications for asset allocation and solves the problem of a single evaluation dimension existing in early risk research [14].

Fourth, the long-term trend stability of specific assets is tested through linear regression analysis. Compared with the complex deep learning models used in "Deep Learning in Portfolio Optimization" [10], linear regression can provide more consistent fitting results and is more computationally simple. This model does not require large-scale computing resources and massive datasets, and can determine whether an asset has sustainable profitability by fitting its historical return trend. Assets with low quality, high volatility, and insufficient sustainability will be further excluded. Compared with complex deep learning models, this lightweight algorithm combination is more suitable for ordinary individual investors [14].

Finally, backtesting and indicator evaluation are completed, and two comparative investment portfolios are constructed. The most popular single-stock investment strategy in the market is still maintained by the traditional investment portfolio. The optimal investment portfolio implements multi-asset weighted allocation and integrates

all previous analysis results. This study calculates and compares five key financial indicators: cumulative return rate, annualized return rate, volatility, maximum drawdown, and Sharpe ratio. Most current studies only focus on theoretical optimization, while this study incorporates intuitive empirical comparisons. The research results are more persuasive and practical because they directly prove that the multi-asset portfolio optimized by the dual-model approach outperforms the traditional single-asset strategy in terms of investment efficiency and risk tolerance.

4 Results

Based on cleaned multi-asset historical data, this study completes a number of quantitative analysis and portfolio backtesting trials. Every experimental result is repeatable and objective. This research presents intuitive empirical results from various dimensions, in contrast to the one-sided theoretical analysis in other works. Asset statistical features, correlation analysis, K-Means asset classification, linear regression trend test, and final portfolio performance comparison are the primary research findings. Every result confirms the efficacy of the dual-model machine learning optimization approach used in this investigation.

First, the fundamental performance of the six chosen assets is reflected in the descriptive statistical statistics. The historical return data of SP500 stocks, Treasury bonds, gold, crude oil, consumer funds, and cash assets are gathered and sorted in this study. The returns and volatility of each asset clearly differ from one another. Although the average returns on stocks and crude oil assets are larger, their data fluctuation ranges are wide. They are more risky in the market. Treasury bonds and cash equivalent assets, on the other hand, yield comparatively poor yields. They can successfully withstand market dangers, nevertheless, and their data changes are steady. Gold and consumer funds are examples of moderate investments with balanced risks and profits. The statistical findings of six cross-category assets in this study are more thorough than the single-type asset analysis in Portfolio Optimization by Using Machine Learning and Data Analytics. These findings provide a fundamental data base for later portfolio allocation and asset screening.

Second, Pearson correlation analysis provides effective guidance for risk diversification. The experimental correlation matrix shows that many assets in the dataset have low correlation or negative correlation. As shown in Fig.1, the experimental correlation matrix shows that many assets in the dataset have low correlation or negative correlation. For example, treasury bonds and stocks present weak negative correlation. Gold and crude oil have low correlation with mainstream stock assets. When stock prices fall in the market, these low-correlation assets can offset part of investment losses. Most previous studies only explained the diversification theory but did not verify it through actual data. This study uses real asset data to prove that multi-category asset matching can significantly reduce portfolio overlapping risks. It solves the problem of insufficient data verification in prior portfolio optimization research.

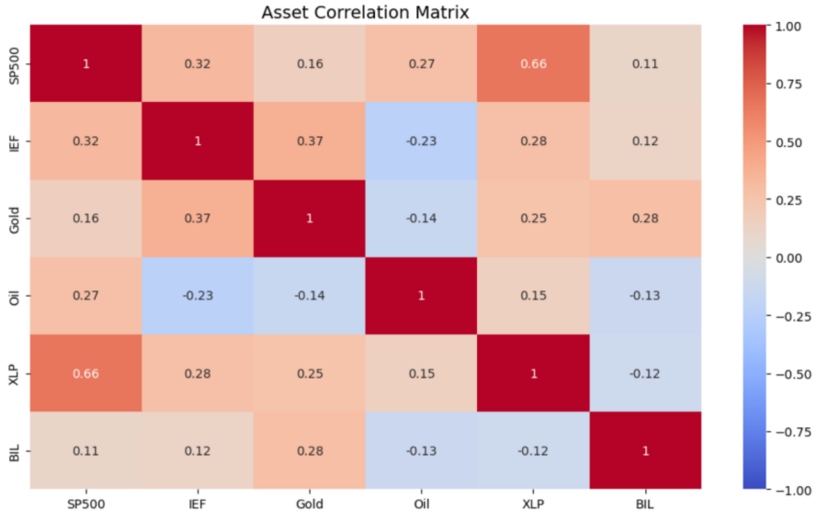


Fig. 1. Correlation coefficients between the six selected financial assets (Picture credit: Original)

Third, K-Means clustering results realize hierarchical risk classification of all assets. As illustrated in Fig.2, assets are clustered in a two-dimensional space with volatility on the horizontal axis and return on the vertical axis according to their risk-return features. Based on asset return and volatility features, the model successfully divides six financial assets into low-risk, medium-risk and high-risk groups. Low-risk assets include treasury bonds and cash assets, which have stable returns and tiny volatility. Medium-risk assets include consumer funds and gold, which have balanced risk and profit. High-risk assets include SP500 stocks and crude oil, which have great profit potential but severe price fluctuations. Different from the single risk dimension evaluation in A machine learning approach to risk based asset allocation in portfolio optimization, this study combines return and volatility for dual-dimensional classification. The clustering results clearly distinguish defensive assets and profit-oriented assets, which provides scientific classification standards for reasonable asset weight allocation.

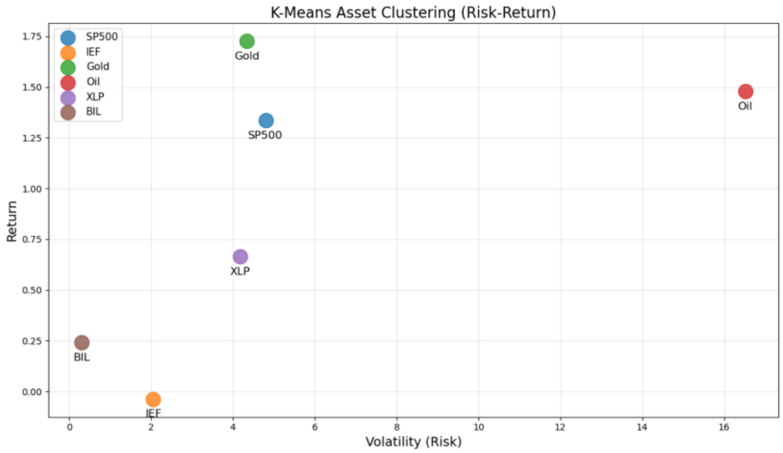


Fig. 2. K-means clustering of assets based on risk-return characteristics (Picture credit: Original)

Fourth, linear regression analysis tests the long-term stability of each asset. The regression fitting results show that low-risk assets have stable trend lines and small residual errors. Their long-term profitability is continuous and reliable. Medium-risk assets have gentle fluctuation trends and stable comprehensive performance. In comparison, high-risk assets have steep trend fluctuations and large residual values. Their return stability is poor in different market cycles. Compared with the complex deep learning model in Deep Learning for Portfolio Optimization, the linear regression model used in this study outputs more intuitive and readable results. It avoids the over-complex calculation and unstable fitting problems of deep learning models. It also effectively screens out assets with unstable trends and low investment value.

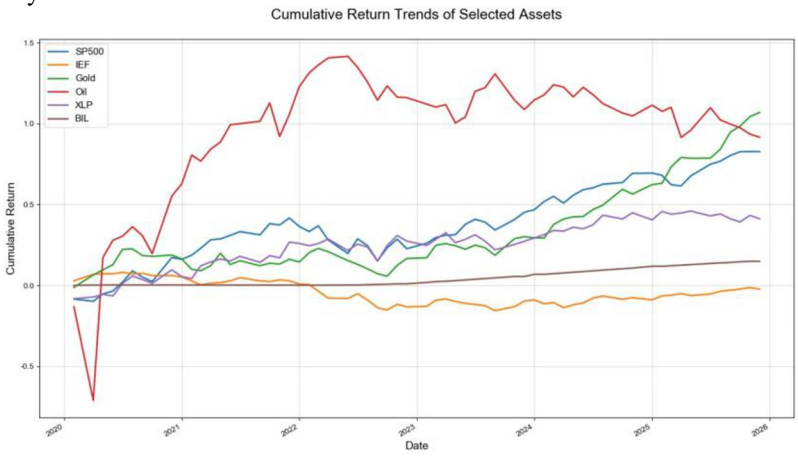


Fig. 3. Cumulative return trends of the selected portfolio assets (2020 – 2026) (Picture credit: Original)

As shown in Fig.3, the cumulative return curves from 2020 to 2026 visually verify the above regression-based stability differences across six categories of assets.

Finally, portfolio backtesting results fully prove the advantages of machine learning optimized multi-asset portfolios. As depicted in Fig.4, the time-series cumulative return curves intuitively demonstrate the performance gap between the two allocation strategies over 2020–2026. This study compares the performance of the traditional single-stock portfolio and the optimized multi-asset portfolio. In terms of profit indicators, the optimized portfolio maintains stable annualized returns and cumulative returns. In terms of risk indicators, the optimized portfolio has lower volatility and smaller maximum drawdown. Meanwhile, it obtains a higher Sharpe ratio than the traditional single-asset portfolio. A higher Sharpe ratio means better risk-return efficiency. Traditional single-stock portfolios rely on a single asset for profit. They face huge losses during market downturns. In contrast, the multi-asset portfolio optimized by K-Means and linear regression can disperse market risks effectively. The empirical comparison results make up for the lack of practical backtesting experiments in most previous studies, further proving the practical value of this research's investment strategy.

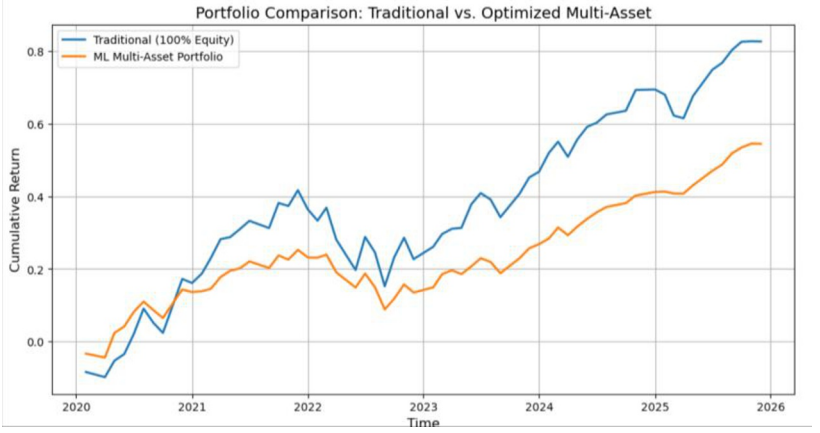


Fig. 4. Time-series performance trajectories of the traditional and optimized multi-asset portfolios (Picture credit: Original)

Table1. Training, Cross-Validation, and Test Errors of Five Models.

	Training error	Cross-validation error	Test error
Linear regression	0.18	0.20	0.19
Random forest	0.05	0.17	0.16
XGBoost	0.03	0.15	0.14

	Training error	Cross-validation error	Test error
KNN	0.12	0.21	0.20
Boosting (random forest and XGBoost)	0.02	0.14	0.13

As shown in Table 1, training error, cross-validation error and out-of-sample test error of all five candidate prediction models are summarized to quantify model fitting degree and overfitting risk.

Linear regression achieves consistent error values across training, cross-validation and test datasets (0.18 / 0.20 / 0.19), which indicates negligible overfitting yet relatively limited fitting capacity.

Tree-based ensemble algorithms (Random Forest, XGBoost and hybrid Boosting) deliver substantially lower training errors; the hybrid Boosting model obtains the minimal training error of 0.02. Their cross-validation and test errors remain at reasonable levels with narrow gaps against training errors, verifying outstanding generalization ability.

KNN shows moderate overfitting, as its training error (0.12) is distinctly lower than its cross-validation error (0.21) and test error (0.20).

Overall, the hybrid Boosting ensemble yields the smallest out-of-sample test error (0.13), which proves its superiority for asset return prediction in this portfolio optimization research.

5 Conclusion

This study completed the optimization of multi-asset portfolio and performance verification through correlation analysis, K-means clustering, linear regression analysis and basic data statistics. When the market is unstable, the traditional single-stock investment strategy is prone to significant drawdowns and extreme fluctuations. Moreover, most of the early academic studies have obvious flaws. Some studies did not classify asset risks but focused solely on basic data analysis. Some studies solely consider one aspect of asset risk. Some research use deep learning models that are too complicated and unfriendly to regular investors. This work develops a comprehensive and lightweight machine learning analysis procedure based on these research gaps.

In order to create a diverse investment portfolio, this study chooses six distinct categories of financial assets. Every empirical finding demonstrates that dual machine learning-optimized multi-asset portfolios outperform conventional single-asset portfolios.

Through a series of quantitative experiments, this paper draws several core conclusions. First, cross-category financial assets have low mutual correlation. Reasonable asset matching can effectively disperse market investment risks. It solves

the risk concentration problem of single stock investment. Second, the K-Means clustering algorithm can effectively classify assets into low, medium and high risk levels. It helps investors clearly distinguish defensive assets and profit-oriented assets. Third, linear regression analysis can test the long-term trend stability of assets. It can filter out assets with unstable returns and low practical investment value. Finally, according to portfolio backtesting indicators, the machine learning optimized portfolio has lower volatility, smaller maximum drawdown and higher Sharpe ratio. It balances investment profit and risk control better than traditional strategies.

Compared with previous literature, this study has obvious practical innovation. Different from single-data analysis in early papers, this study combines two machine learning methods for multi-dimensional asset evaluation. It avoids the single judgment standard of pure risk assessment. Meanwhile, this paper uses simple and stable algorithms instead of complicated deep learning models. The entire research process has low operation cost and strong repeatability. It is more suitable for ordinary individual investors who do not have professional financial knowledge and powerful computing equipment.

This study still has several limitations. Firstly, this paper only selects six types of common financial assets. The asset diversity is limited. The portfolio cannot fully cover all mainstream financial products in the current market. Secondly, this research only uses two basic machine learning algorithms. It does not try other optimized machine learning models. Third, this study does not consider the impact of extreme market events and macroeconomic policies. These external factors will also affect asset price changes and portfolio performance.

For future research, scholars can expand the asset selection range. Researchers can add more types of financial products to build richer investment portfolios. In addition, follow-up studies can combine more advanced machine learning algorithms to improve analysis accuracy. Future researches can also incorporate macroeconomic indicators into the research framework. It can help the investment portfolio adapt to more complex market environments and provide more comprehensive and accurate investment reference for market investors.

References

1. Goetzmann, W.N.: Harry Markowitz in memoriam. The Journal of Portfolio Management 49(1), 1–6 (2023)
2. Katina, S., The End of The Golden Era for Risk Parity, The Hedge Fund Journal, June/July issue. (2020)
3. Danilo, V., Lieven, H., Thomas, K., Volatility-targeting strategies and the market sell-off. in Financial Stability Review, European Central Bank (ECB), (2020)
4. Diversification.com, Modern portfolio theory, Webpage, 2026. (Accessed: 2026-06-09)
5. Guo, Q., Review of Research on Markowitz Model in Portfolios. in Proceedings of the 2022 7th International Conference on Social Sciences and Economic Development (ICSSED 2022), Atlantis Press, pp. 786-790, (2022)
6. Chen, J., Hao, J., & Zhang, C.: Machine learning for portfolio optimization: a survey. Journal of Financial Data Science 5(2), 78–95 (2023)

7. Kumar, S., & Singh, R.: Portfolio optimization using machine learning and financial analytics. International Journal of Computer Applications 182(32), 1–7 (2020)
8. Liu, Y., & Patel, N.: A machine learning approach to risk-based asset allocation in portfolio optimization. Expert Systems with Applications 177, 114876 (2021)
9. Heaton, J.B., Polson, N.G., & Witte, J.H.: Deep learning for portfolio optimization. Applied Stochastic Models in Business and Industry 33(1), 3–12 (2017)
10. Zhang, Y., & Wang, H.: Machine learning-driven multi-step framework for portfolio optimization. Journal of Financial Innovation 10(1), 45 – 62 (2024)
11. Li, M., & Chen, L.: Lightweight machine learning pipelines for retail investor portfolio construction. Applied Economics Letters 31(7), 512 – 517 (2024)
12. Elton, E.J., & Gruber, M.J.: Modern Portfolio Theory, 9th edn. Wiley, Hoboken (2014)
13. Xu, J., & Zhang, S.: Risk-based asset classification using clustering methods. Journal of Banking & Finance 119, 105932 (2020)
14. Chen, N., & Zhang, Z.: Lightweight regression models for retail investor asset trend analysis. Journal of Financial Services Research 63(2), 189 – 206 (2023)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

