Review on Portfolio Optimization with Return Prediction

Introduction

Predicting asset returns and optimizing portfolios are interdependent challenges, where errors in one task can dramatically influence the success of the other. This interdependence is particularly evident in the Markowitz model, where even a 1% error in return prediction can lead to over 50% misallocation in portfolio weights. Such sensitivity often forces practitioners to bypass the Maximum Sharpe Ratio (MSR) portfolio and opt for the General Minimum Variance (GMV) portfolio, which minimizes prediction dependency and focuses on risk reduction. This trade-off highlights a critical insight: any portfolio optimization framework must account for how prediction errors propagate through optimization models. Consequently, the integration of return prediction and portfolio optimization is vital, motivating the need for models that address both tasks simultaneously. Below is a detailed discussion of recent papers that delve into this intersection, with an emphasis on their methodologies, findings, and practical implications.

(Ma et al., 2021)

This paper explores return prediction using machine learning models, including LSTM, DMLP, Random Forest (RF), and Support Vector Regression (SVR), with ARIMA serving as a benchmark. The study uses nine years of daily data (2007–2015) from the China Securities 100 Index and combines these predictions with portfolio optimization methods such as the Mean-Variance (MV) model and the Omega ratio. Random Forest emerges as the best-performing model for prediction, while the Mean-Variance Forecasting (MVF) model—an MV variant incorporating prediction error as introduced in Yu et al. (2020)—produces the best results for portfolio optimization. The Omega ratio, as improved in Kaspos et al. (2014), is also used to address worst-case scenarios under non-Gaussian return assumptions.

A notable aspect of the study is its handling of extreme values. Returns exceeding five times the median are capped to stabilize the dataset by reducing skewness and volatility. This preprocessing step makes returns more Gaussian but could alternatively be integrated within the model itself. Despite these refinements, prediction accuracy remains low, with models achieving less than 50% accuracy in predicting the sign of returns, highlighting the inherent challenges of daily stock market predictions.

The study also incorporates transaction costs in its simulations, which reveal significant variations in portfolio performance due to rebalancing frequency. This practical inclusion underlines the real-world impact of trading costs, particularly for high-frequency strategies. Additionally, while RF outperforms other models, the authors suggest that LSTM may perform better with a richer set of features, leveraging its attention mechanism.

The simulation framework is structured as follows: a rolling 60-day window of historical returns is used to predict next-day returns for all stocks; these predicted returns, along with a rolling 20-day average of prediction errors and a 60-day covariance matrix, are fed into the MVF or Omega ratio models to determine optimal portfolio weights. This workflow integrates prediction and optimization, emphasizing the interdependence of the two tasks.

In summary, RF combined with MVF delivers the best results, while innovations like capped extreme values and transaction cost considerations add practical relevance. However, the study highlights the limitations of current approaches, particularly in feature diversity and prediction accuracy, leaving opportunities for further research.

(Behera et al., 2023)

This paper evaluates stock return prediction using several machine learning models, including Random Forest (RF), Extreme Gradient Boosting (XGBoost), Adaptive Boosting (AdaBoost), Support Vector Machine Regression (SVR), k-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). These predictions are then used for portfolio construction via the Mean-Variance (MV) optimization model. The study focuses on three major stock exchanges: the Bombay Stock Exchange (BSE) in India, the Tokyo Stock Exchange in Japan, and the Shanghai Stock Exchange in China.

The core objective of the study is to identify stocks with high predicted returns to construct an optimized portfolio. Predictions are generated for 50 stocks using all models, and the 30 stocks with the lowest prediction errors are selected for portfolio optimization. However, the paper lacks a simulation of how the portfolios would perform in actual market conditions, missing an opportunity to evaluate real-world applicability.

A limitation of the study is its reliance on past returns as the sole input feature, without discussing feature windows or additional explanatory variables. This narrow input scope likely contributes to underfitting, as indicated by the high Mean Squared Error (MSE) of the models. Similar to other studies, outliers are handled by capping extreme values to reduce volatility, but this approach risks oversimplifying the data and neglecting valuable insights from rare events.

While the methodology is clear, the absence of trading simulations and a more diverse feature set constrains the practical impact of the findings. The study demonstrates the potential of machine learning models in return prediction and portfolio construction but highlights the need for richer feature engineering and real-world performance evaluation.

(Chen et al., 2021)

This paper proposes a hybrid approach for stock price prediction and portfolio optimization using eXtreme Gradient Boosting (XGBoost) enhanced with an Improved Firefly Algorithm (IFA) for prediction and the Mean-Variance (MV) model for portfolio construction. The dataset includes 24 stocks from the Shanghai Stock Exchange, covering the period 2009–2019. The key objective is to leverage machine learning techniques to improve portfolio returns while incorporating predictive accuracy into the optimization process.

For stock prediction, the IFAXGBoost hybrid model optimizes XGBoost hyperparameters using IFA. The algorithm dynamically splits fireflies into elite and normal subgroups and applies chaotic and particle swarm-based strategies to enhance optimization. Input features include log returns over a 15-day rolling window, True Range (TR), Average True Range (ATR), momentum, and Relative Strength Index (RSI). By combining these inputs with the innovative optimization strategy, the model demonstrates improved predictive accuracy, as measured by Mean Squared Error (MSE).

In the portfolio optimization stage, stocks with higher predicted returns are selected, and their weights are determined using the MV model. This step integrates predicted returns with risk metrics to construct an optimized portfolio. The study evaluates the IFAXGBoost + MV framework against traditional methods without stock prediction and alternative prediction models, such as LSTM. While LSTM-based predictions, when paired with MV, perform well, IFAXGBoost delivers the highest Sharpe ratios, even after accounting for transaction costs (0.5%).

The study further highlights the limitations of existing approaches by noting the absence of comparisons with alternative portfolio optimization methods, such as long-term MV holding strategies or other frameworks beyond the MV model. This underscores the need for broader exploration of optimization techniques in future research.

Overall, the paper showcases the efficacy of combining machine learning predictions with robust optimization techniques, with the IFAXGBoost + MV framework outperforming benchmarks like equal-weight (1/n) portfolios and random selection. The inclusion of transaction costs adds practical relevance, though the study's focus on a limited set of optimization methods leaves room for further investigation.

(Ta et al., 2020)

This paper utilizes Long Short-Term Memory (LSTM) networks for return prediction and explores multiple portfolio optimization methods, including Monte Carlo simulation, Mean-Variance optimization, and Equal Weighting (EQ). The dataset consists of ten years of daily prices for 500 large-cap stocks listed on the American stock exchange, with inputs comprising OHLCV data (open, high, low, close, volume).

The study's methodology involves selecting the four best-performing stocks based on the LSTM predictions, measured by their mean and variance. However, this approach raises concerns about diversification, as selecting only a few top-performing stocks may lead to high correlation within the portfolio, contradicting the principle of risk diversification.

A comparison between LSTM and Gated Recurrent Units (GRU) shows that LSTM significantly outperforms GRU in terms of predictive accuracy. Among the portfolio optimization methods, EQ portfolios deliver the highest returns, though this outcome is unsurprising since the approach relies on selecting stocks with the highest predicted returns. This selection process effectively eliminates diversification, favoring concentrated bets on the top stocks.

Despite these limitations, the results demonstrate that the LSTM-based model produces portfolios with returns approximately eight times higher than the S&P 500 over the study period. While this highlights the potential of LSTM for return prediction, the lack of diversification and reliance on high-performing stocks detracts from the robustness of the portfolio optimization framework. Future research could benefit from incorporating broader diversification strategies and a more nuanced evaluation of optimization methods.

(Ma et al., 2020)

This paper examines stock return prediction using Deep Neural Networks (DNNs), Long Short-Term Memory networks (LSTM), and Convolutional Neural Networks (CNN), followed by portfolio optimization where the semi-absolute deviation of prediction errors is used as a risk measure. This innovative risk metric shifts the focus from traditional variance-based measures, emphasizing prediction uncertainty instead. The study focuses on 49 stocks from the Chinese stock market, using data from 2007 to 2015, with OHLCV growth rates as input features.

The performance of these models is evaluated using Equal Weighting (EW) portfolios as benchmarks, with Support Vector Regression (SVR) included for comparison. Among the models, Deep Multilayer Perceptron (DMLP) emerges as the best-performing predictor. Interestingly, the study aligns with findings from other research, where shallow models like XGBoost or DMLP outperform deeper networks such as LSTMs. This suggests that the effectiveness lies in optimization and feature representation rather than leveraging complex memory mechanisms.

One significant contribution of the paper is its focus on prediction error as a risk measure. By maximizing returns and minimizing downside prediction errors, the study presents an alternative approach to traditional Mean-Variance (MV) optimization. However, a limitation is that it only compares its method against EW portfolios rather than traditional MV optimization, which restricts the assessment of its relative efficacy. Additionally, the absence of an efficient frontier in their results indicates the unconventional nature of their risk metric.

Outliers are handled by capping values exceeding five times the median, a strategy aimed at stabilizing the dataset. Despite this, the models exhibit a high Mean Absolute Error (MAE) and less than 50% accuracy in predicting return trends. These outcomes suggest that the shortcomings may stem from shallow feature engineering and limited input diversity.

The paper conducts trading simulations based on desired annual returns, comparing their optimized portfolios against EW benchmarks. Unlike traditional MV frameworks, their method does not show an increase in returns with risk, reflecting the unique dynamics introduced by the semi-absolute deviation risk metric. While the approach highlights the potential of incorporating prediction errors into portfolio optimization, the lack of comparisons with traditional optimization techniques and the limited predictive accuracy underscore areas for further improvement.

(Kelly et al., 2024)

This paper, presented by Bryan Kelly and widely discussed in the field, explores the theoretical impact of increasing model parameterization in machine learning-based return prediction on portfolio performance. While the paper does not rely on experimental or empirical data, it provides a compelling theoretical framework to explain how over-parameterized models can enhance portfolio outcomes, even when the number of training observations is limited and minimal regularization is applied.

The central argument is that machine learning models benefit from incorporating as many plausible predictors as possible and leveraging rich nonlinear relationships rather than relying on simple linear specifications. The paper highlights that these benefits extend even to cases with small datasets, provided the models are accompanied by prudent shrinkage techniques. This approach challenges the conventional caution around over-parameterization, showing that it can lead to better predictions and portfolio construction if properly managed.

Importantly, the authors caution against indiscriminate inclusion of predictors, emphasizing the need for relevance and alignment with the problem domain. By focusing on plausible predictors and leveraging highly parameterized nonlinear models, the theoretical framework demonstrates how machine learning can overcome traditional limitations in financial modeling, such as data scarcity and linear assumptions.

The paper's key contribution lies in its recommendation to embrace model complexity while maintaining rigor in feature selection and regularization. This perspective aligns with broader trends in machine learning, where rich models with carefully curated features outperform simpler models, particularly in contexts with nonlinear relationships like return prediction and portfolio optimization.Our results are not a license to add arbitrary predictors to a model. Instead, we recommend (i) including all plausibly relevant predictors and (ii) using rich nonlinear models rather than simple linear specifications. Doing so confers prediction and portfolio benefits, even when training data are scarce, particu larly when accompanied by prudent shrinkage. Even when the number of raw predictors is small, gains are achieved using those predictors in highly param eterized nonlinear prediction models.

Rooms for improvement:

1. Feature Engineering: Across studies, feature inputs are often limited to historical returns. Expanding feature sets to include technical indicators, macroeconomic variables, and inter-market data could enhance model performance.

2. Incorporating Prediction Errors: Several papers demonstrate the potential of integrating prediction errors into optimization frameworks. This represents a promising shift from traditional Mean-Variance approaches.

3. Multi-Market Indices: Current research is narrowly focused on single-market datasets. Incorporating multi-market indices (e.g., stocks, crypto, commodities) could provide more realistic diversification strategies.

References:

Behera, J., Pasayat, A. K., Behera, H., & Kumar, P. (2023). Prediction based mean-value-at-risk portfolio *optimization using machine learning regression algorithms for multi-national stock markets. Engineering Applications of Artificial Intelligence, 120, 105843.*

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. Applied Soft Computing, 100, 106943.

Kelly, B., Malamud, S., & Zhou, K. (2024). The virtue of complexity in return prediction. The Journal of Finance, 79(1), 459-503.

Ma, Y., Han, R., & Wang, W. (2020). Prediction-based portfolio optimization models using deep neural networks. Ieee Access, 8, 115393-115405.

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications, 165, 113973.

Ta, V.-D., Liu, C.-M., & Tadesse, D. A. (2020). Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. Applied Sciences, 10(2), 437.