Enhancing Market Index Performance through Factor-Based Portfolio Construction and Dynamic Risk Assessment

Jingxiang Zou, Zijun Wang, Yanran Xu, Junwen Peng, Pei Zhu

December 2023

Abstract

This project presents a comprehensive exploration of factor-based portfolio construction and dynamic risk assessment in the Chinese equity market. We employ the Fama-French factors model to select stocks and then apply multifaceted methods to estimate covariance matrices, focusing on market-wide risk information. Our methodology includes three parts: stock selection using Fama-French factors, covariance matrix estimation using various advanced techniques, and portfolio optimization through methods like Global Minimum Variance, Risk Parity, and Hierarchical Risk Parity. Backtesting over 10 years reveals that while factor-based approaches yield positive alpha returns, their performance is highly volatile. Optimized portfolios did not consistently outperform a naive 1/N portfolio strategy. The research provides insights into the effectiveness of different investment strategies in the Chinese market and underscores the importance of balancing return generation with risk control. Our findings have implications for portfolio managers and investors seeking enhanced market index performance through quantitative investment strategies.

1 Introduction

As the factors model are widely used in the global equity market, more and more new method have been introduced in order to obtain a higher profit in quantitative investment. Such excess return could be brought by finding more effective factors,and it could also be brought by constructing appropriate portfolio weights. Many researchers have done a lot of researches in this area, including how to estimate accurate and stable covariance matrix by incorporating more useful information which can be applied in computing optimal weights.

Our work consists mainly of 3 parts, the first part is about stock selection based on Fama-French factors model. In this part, we would like to evaluate how such factors can bring the alpha returns beyond the benchmark index and their corresponding risks. In the next part, we constructed complicated approach to obtain covariance matrix by using risk information in the entire market instead of only considering the risk profiles from the assets we have chosen before. This covarince matrix can be decomposed to multifactor risk matrices and idiosyncratic risk matrices. After that, we used optimization methods in order to obtain the optimal allocation weights within a given portfolios. In this part, we attempted several optimization methods such as ordinary minimal variance approach, risk parity method and hierarchy risk parity method. In the final part, we did backtests for the performance of our given selected portfolio based on different optimization method. We also computed and compared corresponding indicators to validate our portfolio from different aspects.

Eventually, we found the following empirical results in Chinese stock market during the past 10 years. The first one is the Fama-French factors can generate positive alpha returns in Chinese market but so volatile. During a certain period, it performs much worse than the benchmark index. It appeared even the index was rising while our portfolio is falling. The second one is that the optimization approach may not able to bring profits within the portfolio. Such empirical results have also been found by some scholars previously.

In this paper, we introduced the principals of the models and optimization methods that have used, described how we conducted our empirical experiments and what data we have used. Finally we also showed our empirical results, made comments on what we have found and how should we do in real investment practice.

2 Multi-Factor Stock Selection

In the realm of portfolio construction, the multi-factor stock selection process emerges as a cornerstone of our investment strategy, aiming to distill a myriad of stocks into a potent blend of top performers. In this section, we outline the multi-factor stock selection process, which involves the calculation of Fama-French factors and daily alpha values to identify the top-performing stocks. Our methodology includes data collection from the MYQUANT platform and stock selection indicator estimation.

2.1 Data Collection and Processing

We initiated the process by collecting historical stock market data for a predefined stock pool of 999 stocks (from SZSE.000001 to SZSE.000999). The data was sourced from the MYQUANT platform and included essential attributes such as closing prices (TCLOSE), total market capitalization (TOTMKTCAP), and price-to-earnings ratio (PETTM).

To ensure consistency and usability in our dataset, we implemented a series of data processing steps. Firstly, we standardized the date columns by converting them to the 'date' format, ensuring uniformity in our time-series data. Secondly, we streamlined the dataset by eliminating unnecessary columns, allowing us to concentrate on the most pertinent information for our analysis. Furthermore, we organized the data chronologically based on the 'date' column, ensuring that our historical records are in the correct temporal order. Lastly, we computed both daily returns and excess returns for each stock, essential metrics for our subsequent analysis. These data processing procedures laid a solid foundation for our multi-factor stock selection methodology.

2.2 Calculation of Fama-French Factors

Our multi-factor stock selection method integrates the Shenzhen Component Index (SZSE.399106) as the market benchmark. We collect historical data of the index, focusing on the daily 'close' price to calculate market returns. These returns are then adjusted by the annualized risk-free rate to derive the excess market return, providing a more accurate representation of market-specific performance.

Alongside this, we need to calculate another two factors: the Small Minus Big (SMB) and High Minus Low (HML) factors. To accomplish this, we segment our comprehensive dataset based on market capitalization (median market cap) and price-to-earnings ratio (PETTM), classifying stocks into six categories: Small Low, Small Medium, Small High, Big Low, Big Medium, and Big High.[3] This classification enables us to capture the nuances of size and value factors across the stock universe.

We then calculate the SMB and HML values for each day in our dataset. SMB is determined by comparing the average returns of small-cap stocks against those of large-cap stocks, while HML is derived from the differential in returns between high and low price-to-earnings stocks. These calculations are conducted across all stock categories for each day, ensuring a dynamic and responsive approach to factor analysis. Finally, we integrate these calculated SMB and HML values into each stock's dataset, augmenting our library with these crucial factor indicators.

2.3 Estimation of Alpha and Stock Select

The estimation of daily alpha values and stock selection is the next pivotal phase in our strategy. We start by preparing the data for regression analysis, where we consider the excess market return, SMB, and HML as independent variables and the excess return of each stock as the dependent variable. Adding a constant for the intercept, we undertake a regression analysis for each day in our dataset.

$$\text{Excess Return} = \alpha + \beta_{\text{Market}} \times \text{Excess Market Return} + \beta_{\text{SMB}} \times \text{SMB} + \beta_{\text{HML}} \times \text{HML} + \epsilon$$

Our approach is iterative and adaptive. For each day, we calculate the alpha value, reflecting the stock's performance over the market return, adjusted for the Fama-French factors. This calculation is nuanced, taking into account the possibility of regression failures or early iterations where the data might be sparse. In such cases, we use the mean of the excess return as a fallback, ensuring robustness in our analysis.

This alpha estimation process allows us to discern the stocks that consistently outperform market expectations when adjusted for size and value factors. Using this insight, we then embark on the stock selection process, focusing on 40 stocks with the lowest alpha values each month. This selection process aims to compile a portfolio of stocks that are not just strong performers but also exhibit stability and resilience when assessed against broader market indicators.

3 Multi-Factor Model Risk Prediction

Investment is a double-edged sword, where investors pursue returns while also bearing risks. A good multi-factor model framework typically comprises three modules: return model, risk model, and performance attribution. This report focuses on the second major function of multi-factor models risk prediction.

When estimating a multi-factor structured risk matrix, adjustments are required for the estimation of the factor covariance matrix and idiosyncratic risk matrix to ensure consistency between in-sample and out-of-sample estimations, thereby enhancing the accuracy of the estimated results. The adjustments include the following:

1. Estimation of the factor covariance matrix: Newey-West autocorrelation adjustment, eigenvalue adjustment.
2. Estimation of the idiosyncratic risk matrix: Newey-West autocorrelation adjustment, structured model adjustment, Bayesian shrinkage adjustment.

3.1 Multi-factor risk forecasting

Investment is a double-edged sword; investors are both seekers of returns and bearers of risks. Compared to visible returns, the invisible risks are often easily overlooked by investors. However, in reality, effective risk control can significantly amplify investors' efforts. Naturally, estimating the future risks of an investment portfolio becomes the central focus of this report.

Markowitz introduced the use of return variance in 1952 to measure the risk of individual assets, ushering in a new era of quantitatively measuring asset risk. However, practical applications revealed that the optimal investment portfolios derived from asset return covariance matrices often perform suboptimally out of sample. Shepard (2009) [7]pointed out that under assumptions of normality and stationarity, due to estimation errors, the risk of optimal investment portfolios obtained using sample covariance matrices is usually underestimated. The relationship between model estimates and true risk satisfies:

$$\sigma_{\text{true}} = \frac{\sigma_{\text{est}}}{1 - \left(\frac{N}{T}\right)}$$

In this context, σ_{true} represents the actual volatility of the optimal investment portfolio, σ_{est} stands for the portfolio risk estimated by the risk model, N denotes the number of assets, and T is the number of observations in the sample. For instance, when using 100 trading days of data to estimate the covariance matrix of returns for 50 stocks, the estimated risk of the optimal investment portfolio is only half of the actual risk. Additionally, due to the non-stationarity of return sequences, excessively long time intervals are not selected for estimating the sample covariance matrix. However, in a market with numerous trading stocks where the number of stocks greatly exceeds the sample's time length, the sample covariance matrix becomes non-invertible, leading to significant estimation errors.

Estimating stock return covariance using a multifactor model requires estimating the covariance matrix between common factors and the covariance matrix of stock-specific risks, significantly reducing the estimation burden. For example, assuming there are 2000 stocks in the market, directly computing their covariance matrix would entail over 200 million computations, while using a multifactor model for estimation would significantly reduce the number of computations and noticeably enhance accuracy.

A good multifactor model framework typically includes the following three modules: 1) Return model: Identifying style factors closely related to stock returns and characterizing the direction and magnitude of each factor's impact on stock returns. 2) Risk model: Introducing structured estimation methods for the covariance matrix of stock returns, reducing the number of estimated parameters while enhancing the robustness and credibility of estimates, aiming to predict future risk levels of investment portfolios. 3) Performance attribution: Combining the return model and risk model, enables the analysis of a portfolio's performance and risk, aiding investors in understanding the sources of returns and the portfolio's exposure to risk.

We particularly focus on the second major function of the multi-factor model forecasting. Leveraging the structured estimation of the covariance matrix of stock returns through the multi-factor model and applying it to forecast the future risk of any given investment portfolio yields credible results. Furthermore, employing various methods to construct the portfolio leads to a significant reduction in risk compared to the benchmark portfolio, thereby significantly increasing the portfolio's Sharpe ratio compared to the benchmark.

3.2 Methods for estimating multifactor risk matrices

This section will introduce the estimation methods for the multifactor structured risk matrix. The multifactor model posits that asset returns consist of returns driven by common factors and asset-specific returns, with individual asset-specific returns being uncorrelated. Consequently, during risk estimation, it becomes necessary to separately estimate the covariance matrix of style factors and the variance matrix of specific risks. To ensure consistency between in-sample and out-of-sample estimations and enhance the accuracy of the estimates, we will employ Newey-West autocorrelation adjustment, eigenvalue adjustment, and volatility bias adjustment for estimating the style factor matrix. Additionally, for estimating stock-specific risks, we will utilize Newey-West autocorrelation adjustment, structured model adjustment, and Bayesian shrinkage adjustment. These adjustment methods and their effects will be systematically explained in the following sections.

3.2.1 Estimation of the covariance matrix of style factors

3.2.1.1 Newey-West Autocorrelation Adjustment The traditional approach directly utilizes the covariance matrix of stock returns to measure the relationships among stocks, treating all data equally important. However, in reality, the market undergoes numerous changes daily, where recent data has a greater impact on the current state. Therefore, we employ the method of exponentially weighted moving average (EWMA) to calculate the daily covariance matrix F^{Raw}, assigning higher weights to data closer to the current date:

$$F_{kl}^{\text{Raw}} = \text{cov}(f_k, f_l)_t = \sum_{s=t-h}^{t} \lambda^{t-s}(f_{k,s} - \bar{f}_k)(f_{l,s} - \bar{f}_l) / \sum_{s=t-h}^{t} \lambda^{t-s}$$

Here, $f_{k,s}$ represents the return of factor k at time s, $\bar{f}_k$ denotes the exponentially weighted average return of factor k over the sample period, h signifies the length of the sample period, and the half-life parameter τ assigns a weight of $1/2$ to the data from $t - \tau$ days to the current day. λ is calculated as $0.5^{1/\tau}$. In practical computations, we take $h = 252$ and $\tau = 90$.

Since we need to forecast the risk for the upcoming month and the correlation matrix of factors is derived from the daily returns of these factors, it's essential to consider the impact of serial correlation among factor returns. We can perform a Newey-West adjustment on F^{Raw} to compute the adjusted matrix F^{NW}. Specifically,

$$F^{\text{NW}} = 21 \cdot \left[F^{\text{RAW}} + \sum_{\Delta=1}^{D} \left(1 - \frac{\Delta}{1+D}\right) (C_{+\Delta}^{(d)} + C_{-\Delta}^{(d)}) \right]$$

Where D represents the lagged time duration, the computation method for $C_{+\Delta}^{(d)}$ and $C_{-\Delta}^{(d)}$ is as follows. The superscript d denotes that the indicator is derived from daily data.

$$C_{kl,+\Delta} = \text{cov}(f_{k,t-\Delta}, f_{l,t}) = \sum_{s=t-h+\Delta}^{t} (f_{k,s-\Delta} - \bar{f}_k)(f_{l,s} - \bar{f}_l) / \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}$$

$$C_{kl,-\Delta} = \text{cov}(f_{k,t}, f_{l,t-\Delta}) = \sum_{s=t-h+\Delta}^{t} (f_{k,s} - \bar{f}_k)(f_{l,s-\Delta} - \bar{f}_l) / \sum_{s=t-h+\Delta}^{t} \lambda^{t-s}$$

In practical computations, we set the lagged time duration for the serial correlation of variance and covariance to D=2. Other parameters are configured with h=252, and a half-life of 90.

3.2.1.2 The eigenvalue adjustment As described earlier, reveals that estimating directly from the covariance matrix would significantly underestimate the risk of the optimal investment portfolio. To address this, Menchero [6]proposed an eigenvalue adjustment method to correct the covariance matrix. The underestimation of risk in the optimal investment portfolio is closely associated with the concept of eigenvalues in the covariance matrix: mathematically, eigenvalues are derived from the eigenvectors of the factor covariance matrix, while economically, they represent uncorrelated investment portfolios.

$$D_0 = U_0^T F^{NW} U_0$$

Where U_0 is an $N \times N$ orthogonal matrix, the kth column of U_0 represents the kth eigenvector of F^{NW}. The N elements in this vector denote the weights of N specific assets in a particular investment portfolio, referred to as the kth eigenportfolio of F^{NW}. As U_0 is an orthogonal matrix, the eigenportfolios are uncorrelated with each other, and the sum of squares of asset weights in each eigenportfolio is equal to 1. D_0 is a diagonal matrix where the kth element on its diagonal represents the variance of the kth eigenportfolio, and its square root represents the risk of the kth eigenportfolio.

Although the true factor covariance matrix is not known, during simulations, F^{NW} can be considered as the "true" covariance matrix, U_0 as the "true" eigenportfolio weights, and D_0 as the "true" eigenfactor variance matrix. In the mth simulation iteration, follow these steps:

(1) Firstly, generate an $N \times T$ simulated matrix b_m representing the factor returns. Each element in the kth row follows a normal distribution with mean 0 and variance $D_0(k)$, the kth diagonal element of D_0. This sets the variance of elements in the kth row as the "true" variance of the kth eigenfactor.

(2) Compute a simulated factor returns matrix of size N×T using the following formula:

$$r_m = U_0 b_m$$

(3) Compute the covariance matrix of simulated factors.

$$F_m^{MC} = cov(r_m, r_m)$$

We can prove that F_m^{MC} is an unbiased estimate of the 'true' covariance matrix F^{NW}.

(4) Perform an eigenvalue decomposition on the simulated covariance matrix:

$$D_m = U_m^T F_m^{MC} U_m$$

Combine the resulting simulated eigenvectors with the 'true' covariance matrix F^{NW} to obtain the true covariance matrix of the simulated eigenvectors:

$$\tilde{D}_m = U_m^T F^{NW} U_m$$

Note: U_m represents simulated eigenvectors and F^{NW} is the 'true' asset return covariance matrix. Hence, $D_{\tilde{m}}$ is not a diagonal matrix. However, the diagonal elements of $D_{\tilde{m}}$ can be considered as the true variances of the kth simulated eigenvector.

After completing these four steps, one simulation is concluded. We perform a total of M simulations and define the simulated risk deviation for the kth eigenvector as:

$$\lambda(k) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\frac{\tilde{D}_m(k)}{D_m(k)}}$$

In the simulation process, we assume that asset returns follow the assumptions of normality and stationarity. However, real financial data often exhibits characteristics of 'fat tails' and 'leptokurtosis.' Therefore, before correcting the covariance matrix, it is necessary to adjust the simulated risk bias appropriately:

$$\gamma_k = a[\lambda(k) - 1] + 1$$

where a is an adjustment coefficient, typically a value slightly greater than 1, and is set to 1.2 in practical calculations. Subsequently, based on the empirical risk bias $\gamma(k)$, the eigenfactor variances are 'de-biased,' leading to the 'de-biased' covariance matrix:

$$\tilde{D_0} = \gamma^2 D_0$$

where γ^2 is a diagonal matrix with the kth diagonal element being $\gamma^2(k)$. Finally, the 'de-biased' factor covariance matrix F^{Eigen} can be obtained through an orthogonal rotation.

$$F^{Eigen} = U_0 \tilde{D_0} {U_0}^T$$

3.2.2 Estimation of the covariance matrix of idiosyncratic risk

In our model, we assume that the idiosyncratic risk of each stock is independent of the style risk factors, and they are also mutually independent. Under this assumption, the idiosyncratic risk covariance matrix $\tilde{\Delta}$ is orthogonal.

3.2.2.1 Newey-West Autocorrelation Adjustment First, we employ the same Newey-West method as in the estimation of the covariance matrix of style factors. We employ the method of exponentially weighted moving average (EWMA) to calculate the daily covariance matrix $\tilde{F}^{\mathrm{Raw}}$, assigning higher weights to data closer to the current date:

$$\tilde{F}_{kl}^{\mathrm{Raw}} = \mathrm{cov}(\tilde{f}_k, \tilde{f}_l)_t = \sum_{s=t-h}^{t} \lambda^{t-s}(\tilde{f}_{k,s} - \bar{\tilde{f}}_k)(\tilde{f}_{l,s} - \bar{\tilde{f}}_l) / \sum_{s=t-h}^{t} \lambda^{t-s}$$

Here, $\tilde{f}_{k,s}$ represents the specific return of stock k at time s, $\bar{\tilde{f}}_k$ denotes the exponentially weighted average return of stock k over the sample period, h signifies the length of the sample period, and the half-life parameter τ assigns a weight of 1/2 to the data from $t - \tau$ days to the current day. λ is calculated as $0.5^{1/\tau}$. In practical computations, we take $h = 252$ and $\tau = 90$.

Since we need to forecast the risk for the upcoming month and the correlation matrix of stocks is derived from the daily returns of these stocks, it's essential to consider the impact of serial correlation among factor returns. We can perform a Newey-West adjustment on $\tilde{F}^{\text{Raw}}$ to compute the adjusted matrix $\tilde{F}^{\text{NW}}$. Specifically,

$$\tilde{F}^{\text{NW}} = 21 \cdot \left[\tilde{F}^{\text{RAW}} + \sum_{\tilde{\Delta}=1}^{D} \left(1 - \frac{\tilde{\Delta}}{1+D}\right) (C_{+\tilde{\Delta}}^{(d)} + C_{-\tilde{\Delta}}^{(d)}) \right]$$

Where D represents the lagged time duration, the computation method for $C_{+\tilde{\Delta}}^{(d)}$ and $C_{-\tilde{\Delta}}^{(d)}$ is as follows. The superscript d denotes that the indicator is derived from daily data.

$$C_{kl,+\tilde{\Delta}} = \text{cov}(\tilde{f}_{k,t-\tilde{\Delta}}, \tilde{f}_{l,t}) = \sum_{s=t-h+\tilde{\Delta}}^{t} (\tilde{f}_{k,s-\tilde{\Delta}} - \bar{\tilde{f}}_k)(\tilde{f}_{l,s} - \bar{\tilde{f}}_l) / \sum_{s=t-h+\tilde{\Delta}}^{t} \lambda^{t-s}$$

$$C_{kl,-\tilde{\Delta}} = \text{cov}(\tilde{f}_{k,t}, \tilde{f}_{l,t-\tilde{\Delta}}) = \sum_{s=t-h+\tilde{\Delta}}^{t} (\tilde{f}_{k,s} - \bar{\tilde{f}}_k)(\tilde{f}_{l,s-\tilde{\Delta}} - \bar{\tilde{f}}_l) / \sum_{s=t-h+\tilde{\Delta}}^{t} \lambda^{t-s}$$

In practical computations, we set the lagged time duration for the serial correlation of variance and covariance to D=5. Other parameters are configured with h=252, and a half-life of 90, which are the same with in the estimation of the covariance matrix of style factors.

3.2.2.2 Structural model adjustment In implementation, there are possible missing data and exotic values of individual stocks' idiosyncratic returns. These data problems can be caused by newly-listed stocks, long-term suspended stocks, and Disclosure of major company events. We employ structural model[4] to handle these issues based on the idea that stocks with the same idiosyncratic risk characteristics may have the same idiosyncratic volatility.

First, we calculate the robust standard deviation$\tilde{\sigma_u}$ as

$$\tilde{\sigma_u} = \frac{1}{1.35} * (Q_3 - Q_1)$$

where Q_1 and Q_3 are the 1/4 and 3/4 quantile of the idiosyncratic return for the stock. In this quantile calculation, we choose the sample time period h=252.

Then we calculate the fat-tail level of the idiosyncratic return Z_u.

$$Z_u = |\frac{\sigma_{u,eq} - \tilde{\sigma_u}}{\tilde{\sigma_u}}|$$

where $\sigma_{u,eq}$ is the sample standard deviation of the idiosyncratic return. Large Z_u implies the existence of exotic values in this idiosyncratic return series.

Finally we introduce the coordination parameter γ, and calculate the adjusted idiosyncratic risk $\hat{\sigma}_u$ as

$$\gamma = \left[min(1, max(0, \frac{h-60}{120})) \right] * [min(1, max(0, exp(1 - Z_u)))]$$

$$\hat{\sigma}_u = \gamma * \sigma_u^{NW} + (1-\gamma) * \sigma_u^{STR}$$

where σ_u^{NW} is the idiosyncratic risk after Newey-West adjustment, γ lies between 0 and 1, and σ_u^{STR} is the structural risk of the stock, which is calculated through regression using all stocks with $\gamma = 1$

$$ln(\sigma_u^{TS}) = \sum_k X_{nk} * b_k + \epsilon_n$$

where X_{nk} is the factor exposure matrix of the stock. After getting the regression coefficients b_k, we calculate the structural risk of all stocks, including those with $\gamma < 1$ as

$$\sigma_n^{STR} = E_0 * exp(\sum_k X_{nk} * b_k)$$

where E_0 is a constant slightly larger than 1. It is used to eliminate the bias caused by the exponential of the residuals. Here we choose $E_0 = 1.05$.

3.2.2.3 Bayesian shrinkage adjustment Based on the reference, the above two estimation methods tend to underestimate the future risk of low-in-sample-volatility stocks while overestimating the future risk of high-in-sample-volatility stocks. To handle this bias, we introduce Bayesian Shrinkage to shrink the idiosyncratic risk of individual stocks towards the weighted average idiosyncratic risk of the group divided by market capital.

The idiosyncratic risk of the stock is calculated by

$$\sigma_n^{SH} = v_n \bar{\sigma}(s_n) + (1 - v_n)\hat{\sigma}_n$$

$$\bar{s_n} = \sum_{n \subset S_n} w_n \hat{\sigma}_n$$

where $\bar{\sigma}(s_n)$ is the Bayesian Prior risk matrix, as the shrinkage target. It represents the weighted average risk of the market capital group where the stock lies. $\hat{\sigma}_n$ is the idiosyncratic risk of the stock after the structural model adjustment. V_n is the shrinkage intensity calculated by

$$v_n = \frac{q|\hat{\sigma}_n - \bar{\sigma}(S_n)|}{\Delta_n(S_n) + q|\hat{\sigma}_n - \bar{\sigma}(S_n)|}$$

where q is the shrinkage parameter set as 1, $\Delta_\sigma(S_n)$ represents the standard variation of the idiosyncratic risk of the market capital group where the stock lies.

$$\Delta_\sigma(S_n) = \sqrt{\frac{1}{N(S_n)} \sum_{n \subset S_n} (\hat{\sigma}_n - \bar{\sigma}(S_n))^2}$$

When the predicted risk of the stock diverges from the average risk of the market capital group, the shrinkage intensity v_n will be larger, and the shrunk risk σ_n^{SH} will give larger weight to the Bayesian prior risk $\bar{\sigma}(s_n)$.

4 Optimization Methods

Based on the preceding analytical efforts, it is reasonable to anticipate that we have arrived at a robust estimation of the covariance matrix for future asset returns. This estimation is integral, as it serves as a foundational input for the subsequent phase of our study, where we allocate values to each asset in the portfolio. To this end, we plan to employ and evaluate a range of optimization methods. This experimental approach will allow us to assess the effectiveness of different strategies in optimizing portfolio allocations, thereby contributing to a more nuanced understanding of asset allocation dynamics along different optimization schemes.

4.1 Global Minimum Variance

Our primary focus is on the widely recognized and classic Markowitz Portfolio Theory, specifically the Global Minimum Variance (GMV) approach. The essence of a GMV portfolio is its composition of assets that collectively exhibit the lowest possible volatility. This characteristic stems from the individual assets' minimal sensitivity to risk, ensuring that their prices are less prone to significant fluctuations. Consequently, the aggregated risk level of the GMV portfolio is lower than the individual risk levels of the constituent stocks. A minimum variance portfolio contains risk by minimizing the variance of portfolio returns:

$$\min_{w} W^T C W$$

with an inherent constraint:

$$\sum w_i = 1, w_i \geq 0$$

We take an interest in GMV not only because of its popularity but also because of its outstanding performance that has been proven in the equity market, where stocks that are less volatile than their counterparts have historically produced comparable or better returns.

4.2 Risk Parity

We then think of Risk Parity, which asserts that the 'risk contribution' of each asset to the portfolio risk is gauged by its partial derivative to the value-based asset weight, which is $\frac{\partial \sigma_p}{\partial w_i}$. [5] This representation is adjusted by multiplying the corresponding weight to become $w_i \frac{\partial \sigma_p}{\partial w_i}$ so that the value represents a portion of σ_p, because:

$$\begin{aligned} \sum w_i \frac{\partial \sigma_p}{\partial w_i} &= \sum w_i \frac{(CW)_i}{\sqrt{W^T CW}} \\ &= \sigma_p \end{aligned}$$

To equate the 'risk contribution' of the assets is to minimize their distinctions. In practice, we seek to minimize their second central moment:

$$\min_{w} \sum (w_i \frac{(CW)_i}{\sqrt{W^T CW}} - \frac{\sigma_p}{N})^2$$

with an inherent constraint:

$$\sum w_i = 1, w_i \geq 0$$

The risk parity approach asserts that when asset allocations are adjusted to the same risk level, the risk parity portfolio can achieve a higher Sharpe ratio and can be more resistant to market downturns than the traditional portfolio. It is believed that Risk parity is vulnerable to significant shifts in correlation regimes. In our project, we shall see if it holds for the Chinese market as well.

4.3 Hierarchy Risk Parity

Hierarchical Risk Parity is a novel portfolio optimization method developed by Marcos Lopez de Prado in 2016[1]. The advantages of this are that it does not require the inversion of the covariance matrix as with traditional mean-variance optimization, and seems to produce diverse portfolios that perform well out of sample.

4.4 The Naive $1\ /\ N$ Method

To provide a contrast, we do no optimization at all. This is represented by assigning equal weights to each of the assets in the portfolio at all times. (DeMiguel etc., 2007) [2] evaluate the out-of-sample performance of the sample-based mean-variance model, and its extensions designed to reduce estimation error, relative to the naive 1/N portfolio. Of the 14 models they evaluate across seven empirical datasets, none is consistently better than the 1/N rule in terms of Sharpe ratio, certainty-equivalent return, and turnover. We would want to see if this assertion also holds for our portfolio.

5 The backtesting procedure and results

5.1 Backtest

The backtesting process encountered significant challenges due to data constraints. The Chinese stock market is known for its limited transparency and restricted access to historical data, which impeded our ability to fully access comprehensive historical price information. This limitation posed a substantial hurdle in calculating the return covariance matrix for selected stocks, thereby impacting the feasibility of subsequent optimization and backtesting procedures.

To mitigate this issue, we sourced the broadest possible dataset on the covariance matrix from the Shenzhen Stock Exchange. The initial phase of our study involved selecting a set of 20 stocks predicted to yield optimal performance. However, due to incomplete data for some of these stocks, we were compelled to introduce a measure of redundancy. Specifically, we randomly chose 15 out of the 20 selected stocks, ensuring that each had a complete dataset, including a fully populated series of asset returns and a comprehensive covariance matrix without missing values.

There were instances where the data limitations were so acute that even this redundancy strategy was insufficient. In such cases, where fewer than 15 stocks had complete data series, we excluded that specific time point from our analysis. Consequently, we did not adjust the asset allocation for that period and instead retained the previous allocation for the following period. Originally, the plan was to adjust the allocations monthly over a ten-year span, encompassing 122 time points. However, due to these data constraints, we had to exclude 15 time points, ultimately conducting adjustments at 107 time points.

5.2 Results

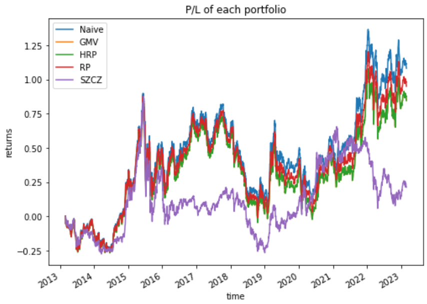
	NAI	GMV	HRP	RP	Benchmark Index
Ann Return	10.81%	9.71%	9.71%	10.21%	5.27%
Ann Vol	25.23%	25.80%	25.80%	25.52%	25.43%
Sharpe	0.4285	0.3765	0.3765	0.4000	0.2071
Max Drawback	45.6%	48.6%	48.6%	47.3%	60.8%
Excess Ann Return	5.6%	4.3%	4.3%	4.9%	NA
Excess Ann Vol	16.6%	16.9%	16.9%	16.8%	NA
Excess Sharp	0.156	0.077	0.077	0.113	NA
Excess Max Drawback	49.7%	52.2%	52.2%	52.1%	NA

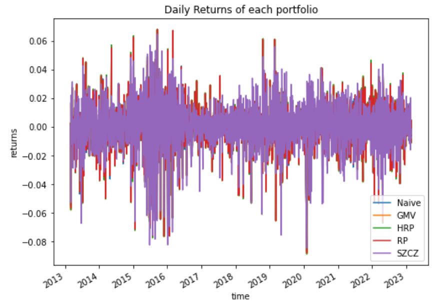
Upon examining the annualized returns of the portfolio, it appears that our management strategies have been notably effective, yielding an annual return exceeding 9 percent across the assets. Notably, the naive 1/N portfolio demonstrated an exceptional performance, achieving a 10.81 percent annual return.

However, when evaluating these results in the context of return volatility and Sharpe Ratio, the perceived success is moderated. While the portfolio Sharpe Ratios substantially surpassed the market index, none exceeded the value of 1. Additionally, all portfolio return volatilities were higher compared to the SZCZ index. This observation suggests that the impressive absolute return performance can be partly attributed to an increase in volatility.

To provide a more comprehensive assessment, we also calculated various excess indicators to further validate our portfolio's performance. These calculations revealed that despite achieving positive excess returns over the 10-year period, there remains a notable risk in terms of relative profit and loss due to significant excess volatility and considerable maximum drawdown. These findings highlight the necessity for more focused attention on the stability of specific factors and the methodologies employed in covariance matrix computation.

It is noteworthy that the superior performance of the Naive 1/N portfolio, in comparison to other optimization methods, corroborates the findings of DeMiguel et al. (2007). Their research posited that most optimization schemes do not consistently surpass the 1/N rule when evaluated against metrics such as the Sharpe ratio, certainty-equivalent return, and portfolio turnover. Our results, indicating the Naive 1/N portfolio's higher returns, lend empirical support to this hypothesis, suggesting that the simplicity of equal weighting often matches or exceeds the complexity of more sophisticated asset allocation strategies in terms of overall performance.





A critical insight from our analysis is that the outperformance of the Naive 1/N method over other methods is predominantly attributable to effective stock selection rather than the optimization process itself. This is evident when analyzing the contribution of each component to the portfolio's excess return. For instance, when comparing the Naive 1/N portfolio to the Global Minimum Variance (GMV) portfolio, it can be deduced that only about 19.85 percent (calculated as $\frac{10.81-9.71}{10.81-5.27}$) of the Naive 1/N portfolio's return can be ascribed to the optimization strategy. In contrast, a substantial majority, approximately 80.15 percent (computed as $\frac{9.71-5.27}{10.81-5.27}$) is attributable to the efficacy of the factor-based stock selection process. This breakdown underscores the significant impact of stock selection on portfolio performance, highlighting its predominance over the allocation method used.

References

[1] Marcos Lopez De Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69, 2016.

[2] Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22(5):1915–1953, 12 2007.

[3] Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

[4] Richard C Grinold and Ronald N Kahn. Active portfolio management. 2000.

[5] Chris Kelliher. *Quantitative Finance with Python: A Practical Guide to Investment Management, Trading, and Financial Engineering*. CRC Press, 2022.

[6] Jose Menchero, Jun Wang, and Doyle Jay Orr. Improving risk forecasts for optimized portfolios. *Financial Analysts Journal*, 68(3):40–50, 2012.

[7] Peter G Shepard. Second order risk. *arXiv preprint arXiv:0908.2455*, 2009.