Application of Multi-Factor Financial Models in Asset Pricing

Sihan Liang
The Chinese University of Hong Kong, Shenzhen, School of Management and Economics
sihanliang@link.cuhk.edu.cn

Abstract. This paper explores the theoretical underpinnings and practical applications of multi-factor models in stock portfolio management. It outlines the theoretical frameworks of multi-factor asset pricing models such as APT theory, Fama-French three-factor, and five-factor models, analyzing how various risk factors influence asset returns. The study then discusses methods for identifying core risk factors empirically, constructing portfolios, optimizing weight allocations, and evaluating investment performance. Empirical testing using data from the CSI 300 and CSI 800 indices confirms that optimized multi-factor portfolios achieve significant excess returns while managing risk effectively, showcasing the practical utility of this quantitative investment framework in the Chinese A-share market.

Keywords: Multi-Factor Models; Portfolio Management; Asset Pricing; Risk Factors; Empirical Research; Quantitative Investment.

1. Introduction

This paper comprehensively explores multi-factor asset pricing models, which underpin the construction of diversified portfolios aimed at achieving excess returns [1]. It details the theoretical framework and empirical applications of these models in stock portfolio management, reviewing how factors like market, size, value, quality, and investment influence asset returns. The study discusses methods for identifying core risk factors, constructing portfolios, optimizing weight allocations, and evaluating performance through empirical case studies. By bridging theory with practical application, this research advances quantitative investment theories and practices [2].

2. Construction Methods of Multi-Factor Models

2.1 Multi-Factor Model Framework

The framework of multi-factor models typically uses multiple linear regression, treating asset returns as the dependent variable and multiple explanatory variables (i.e., risk factors) as independent variables to construct a regression equation[3]. This framework is widely applied in the analysis of both time-series and cross-sectional data.

\[ R_i = \alpha_i + \beta_{i1}F_1 + \beta_{i2}F_2 + \ldots + \beta_{in}F_n + \epsilon_i \]

where \( R_i \) is the return on asset i. \( \alpha_i \) is the intercept term for asset i, representing the expected excess return of the asset. \( \beta_{ij} \) is the sensitivity or exposure of asset i to factor j. \( F_j \) is the j-th risk factor (e.g., market risk, size factor, value factor). \( \epsilon_i \) is the error term, representing random fluctuations not explained by the model.

2.2 Extraction Methods of Risk Factors

In financial analysis, identifying the risk factors that affect asset returns is a key step in modeling. Common statistical methods for this purpose include Principal Component Analysis (PCA) and Factor Analysis, both of which aim to extract valuable information from a large amount of data[4]. Principal Component Analysis is a statistical technique used to extract key information from
multiple correlated variables by reducing the dimensionality of the data while maximizing the variability in the dataset. The formula for PCA is as follows:

\[ c = \frac{1}{n-1} X^TX \]
\[ c = VLV^T \]
\[ X = \frac{X - \mu}{\sigma} \]

where X is the original data, \( \mu \) and \( \sigma \) are the mean and standard deviation of the data., \( X^T \) is the transpose of \( X \). V is the matrix composed of eigenvectors, and L is the diagonal matrix composed of eigenvalues. Select the eigenvectors corresponding to the largest eigenvalues. These eigenvectors represent the directions of maximum variance in the data and are often considered as new risk factors[5]. Factor Analysis assumes that the variability of observed variables can be explained by several unobserved latent factors and specific factors. To increase the interpretability of the model, orthogonal or oblique rotation can be applied to the factor loadings.

### 2.3 Model Estimation Methods

After constructing the model framework and determining the risk factors, it is necessary to use appropriate methods to estimate the model parameters[6]. The most common method is Ordinary Least Squares (OLS) estimation, which aims to minimize the sum of squared residuals to solve for the model coefficients. Weighted Least Squares (WLS) assigns different weights to heteroscedastic data to obtain more accurate estimates under heteroscedastic conditions. For cases where there is multicollinearity among the independent variables, regularization methods such as Ridge Regression can be used to avoid overfitting the model.

\[ \tilde{\beta} = (X^TX)^{-1}X^TY \]
\[ \tilde{\beta}_{WLS} = (X^TWX)^{-1}X^TWY \]
\[ \tilde{\beta}_{ridge} = (X^TX + \lambda I)^{-1}X^TY \]

where X is the design matrix and Y is the response variable, W is a diagonal weight matrix, typically the reciprocal of the residual variance, \( \lambda \) is the regularization parameter and I is the identity matrix.

### 3. Establishment of Multi-Factor Financial Models

#### 3.1 Data Sources and Processing

In this study, the constituents of the CSI 300 index from January 2019 to April 2023 were selected as the research sample. Data were sourced from the CSMAR and Wind Information databases[7]. The original data set included monthly individual stock returns, market capitalization, book-to-market ratio, investor sentiment indicators, and over 300 other potential influencing variables. After conducting data cleaning tasks such as imputing missing values and handling outliers, the number of valid observations was obtained.

#### 3.2 Factor Extraction and Model Estimation

In complex financial market analysis, the study utilized machine learning algorithms such as Random Forest and LASSO regression to identify the top 30 explanatory variables from a pool of
over 300 original variables [8]. Principal Component Analysis (PCA) was then applied to these 30 variables to extract 5 principal components, which collectively explained 72.6% of the total variance. These components were utilized as independent variables in constructing the Fama-French five-factor model, employing Ordinary Least Squares (OLS) regression to analyze their impact on individual stock returns, as detailed in Figure 1 of the study.

Figure 1: Parameter Estimation Results of the Fama-French Five-Factor Model
The overall R-squared of the model is 65.3%, with an adjusted R-squared of 63.7%, indicating a strong explanatory power. Analysis of the factor loading matrix shows that PC1 mainly corresponds to market risk premium, PC2 corresponds to size effect, and PC3 covers information related to value and quality factors.

3.3 Model Testing and Optimization
A detailed examination and optimization analysis were conducted on the established investment model. Through residual analysis, rolling window backtesting, and other methods, it was found that the model performed robustly overall, with some volatility in extreme market conditions[9]. However, it generally captured the impact of different risk factors on returns well. To further optimize the model, we attempted regularization regressions such as Ridge Regression and ElasticNet. The results showed that ElasticNet slightly outperformed the model overall under certain parameter settings, as shown in Figure 2:

Figure 2: Comparison of Adjusted R-squared for ElasticNet Regression Model
Based on the comparison and optimization of various regression models, this study ultimately selected the ElasticNet regression model with optimal penalty coefficients (alpha=0.3, lambda=0.05) as the analytical tool. The ElasticNet model combines the features of Lasso and Ridge regression, applying both L1 and L2 penalties to the coefficients, aimed at overcoming multicollinearity issues while retaining the variable selection functionality[10]. To validate the effectiveness and improvements of the ElasticNet model, its backtesting results were compared with those of the traditional Ordinary Least Squares (OLS) regression model. The specific comparison results are shown in Table 2:

<table>
<thead>
<tr>
<th>Metric</th>
<th>OLS Model</th>
<th>ElasticNet Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annualized Return</td>
<td>11.20%</td>
<td>11.50%</td>
</tr>
<tr>
<td>Annualized Volatility</td>
<td>17.80%</td>
<td>16.90%</td>
</tr>
<tr>
<td>Annualized Sharpe Ratio</td>
<td>0.63</td>
<td>0.68</td>
</tr>
</tbody>
</table>
It is evident from the table that the ElasticNet model outperforms the OLS model across multiple key performance indicators. Specifically, the ElasticNet model increased the annualized return to 11.50%, compared to the OLS model's 11.20%, demonstrating better return performance. More importantly, the ElasticNet model excelled in reducing portfolio volatility, with the annualized volatility dropping to 16.90%, a significant improvement from the OLS model's 17.80%. The annualized Sharpe ratio improved from 0.63 in the OLS model to 0.68, indicating that the ElasticNet model offers higher efficiency in risk-adjusted returns. It is clear that the ElasticNet model has managed to both reduce volatility and increase return levels, with an improved annualized Sharpe ratio.

4. Empirical Application of Multi-Factor Models in Stock Portfolio Management

4.1 Research Sample and Data

In this study, the constituents of the CSI 800 index from January 2019 to April 2023 were selected as the research sample. The data were sourced from the CSMAR and Wind Information databases. The CSI 800 index reflects the concentration of small and medium-sized companies in the Shanghai and Shenzhen stock markets, with a broad coverage that can better represent the overall characteristics of the Chinese A-share market. The original data set included monthly individual stock returns, market capitalization, book-to-market ratio, turnover rate, and over 200 other potential influencing variables. After data cleaning and processing, the final number of valid observations was obtained as summarized in Table 3:

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Valid Stocks</th>
<th>Number of Valid Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019</td>
<td>723</td>
<td>8,676</td>
</tr>
<tr>
<td>2020</td>
<td>749</td>
<td>8,988</td>
</tr>
<tr>
<td>2021</td>
<td>768</td>
<td>9,216</td>
</tr>
<tr>
<td>2022</td>
<td>782</td>
<td>9,384</td>
</tr>
<tr>
<td>2023</td>
<td>785</td>
<td>3,140</td>
</tr>
<tr>
<td>Total</td>
<td>3,807</td>
<td>39,404</td>
</tr>
</tbody>
</table>

4.2 Identification of Risk Factors

This study utilized the random forest algorithm to identify 30 key explanatory factors from a pool of over 200 variables, crucial for understanding stock price fluctuations. To refine focus on essential risk factors, Principal Component Analysis (PCA) was employed. PCA extracted the most critical information through variable correlations, revealing that the first five principal components captured a substantial portion of market variance.

4.3 Portfolio Construction

This study enhanced the Fama-French five-factor model by incorporating additional risk factors such as growth/value, liquidity, and profitability quality, identified through Principal Component Analysis. This comprehensive approach provides a deeper framework for investment decisions. Using time series regression, the study quantified each stock's exposure to these factors, crucial for constructing portfolios scientifically. Stocks were ranked based on their exposure levels and equally
weighted into five groups, creating portfolios with distinct risk profiles. This method not only emphasizes the influence of various risk factors but also enables performance comparison across different risk configurations, detailed in Table 4.

Table 4: Composition of 5 Investment Portfolios with Stock Codes and Weights

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>600199</td>
<td>0.035</td>
<td>600831</td>
<td>0.038</td>
<td>600309</td>
<td>0.042</td>
<td>600583</td>
<td>0.041</td>
<td>600597</td>
<td>0.044</td>
</tr>
<tr>
<td>600372</td>
<td>0.032</td>
<td>600695</td>
<td>0.035</td>
<td>600688</td>
<td>0.039</td>
<td>600703</td>
<td>0.037</td>
<td>600505</td>
<td>0.04</td>
</tr>
<tr>
<td>600426</td>
<td>0.031</td>
<td>600866</td>
<td>0.034</td>
<td>600816</td>
<td>0.037</td>
<td>600111</td>
<td>0.035</td>
<td>600898</td>
<td>0.038</td>
</tr>
<tr>
<td>600585</td>
<td>0.029</td>
<td>600990</td>
<td>0.032</td>
<td>600036</td>
<td>0.035</td>
<td>600277</td>
<td>0.033</td>
<td>600886</td>
<td>0.036</td>
</tr>
<tr>
<td>600637</td>
<td>0.027</td>
<td>601858</td>
<td>0.03</td>
<td>600638</td>
<td>0.033</td>
<td>600893</td>
<td>0.031</td>
<td>600967</td>
<td>0.034</td>
</tr>
</tbody>
</table>

4.4 Portfolio Optimization

In this study, we used the Most Risk Premium model to optimize the above 5 initial investment portfolios. The goal was to maximize expected excess returns under a predefined risk budget constraint. The optimized weights obtained are shown in Table 5:

5: Weighting of the 5 Optimized Investment Portfolios

<table>
<thead>
<tr>
<th>Portfolio 1</th>
<th>Weights</th>
<th>Portfolio 2</th>
<th>Weights</th>
<th>Portfolio 3</th>
<th>Weights</th>
<th>Portfolio 4</th>
<th>Weights</th>
<th>Portfolio 5</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>600199</td>
<td>0.028</td>
<td>600831</td>
<td>0.045</td>
<td>600309</td>
<td>0.065</td>
<td>600583</td>
<td>0.057</td>
<td>600597</td>
<td>0.062</td>
</tr>
<tr>
<td>600372</td>
<td>0.025</td>
<td>600695</td>
<td>0.039</td>
<td>600688</td>
<td>0.051</td>
<td>600703</td>
<td>0.049</td>
<td>600505</td>
<td>0.055</td>
</tr>
<tr>
<td>600426</td>
<td>0.023</td>
<td>600866</td>
<td>0.037</td>
<td>600816</td>
<td>0.048</td>
<td>600111</td>
<td>0.044</td>
<td>600898</td>
<td>0.05</td>
</tr>
<tr>
<td>600585</td>
<td>0.021</td>
<td>600990</td>
<td>0.035</td>
<td>600036</td>
<td>0.042</td>
<td>600277</td>
<td>0.04</td>
<td>600886</td>
<td>0.046</td>
</tr>
<tr>
<td>600637</td>
<td>0.019</td>
<td>601858</td>
<td>0.032</td>
<td>600638</td>
<td>0.039</td>
<td>600893</td>
<td>0.037</td>
<td>600967</td>
<td>0.043</td>
</tr>
</tbody>
</table>

The optimized final portfolio shows increased exposure to small-cap, high-investment growth stocks compared to the initial portfolio, aiming to capture higher expected returns associated with size and investment factors. Additionally, it enhances exposure to low market beta and high book-to-market ratio factors, targeting excess return premiums from defensive value stocks. Conversely, the portfolio reduces exposure to the profitability quality factor, managing risks linked to potential earnings manipulation.

4.5 Investment Performance Evaluation

The study optimized five initial multi-factor investment portfolios using the Most Risk Premium model to maximize expected excess returns within a specified risk budget. Annualized backtesting showed these portfolios significantly outperformed the CSI 800 benchmark, achieving an annualized return of 18.6% compared to the index's 13.9%. Moreover, the portfolios demonstrated lower annualized volatility at 17.2% versus the index's 19.5%, indicating superior risk management. Their Sharpe ratio of 1.08 exceeded the index's 0.71, highlighting better risk-adjusted returns.
Additionally, the portfolios exhibited stronger risk resistance with a maximum drawdown of -21.3%, significantly lower than the index's -28.7%.

5. Conclusion

This paper explores multi-factor models in stock portfolio management, focusing on theoretical frameworks such as APT, Fama-French three-factor, and five-factor models to analyze how different risk factors impact asset returns. It details methods for identifying core risk factors, constructing portfolios, optimizing weights, and evaluating performance using CSI 300 and CSI 800 indices as case studies. Employing machine learning and regression models, it extracts factors like market, size, value, quality, and investment. Empirical results demonstrate these portfolios achieve significant excess returns over benchmarks while effectively managing risk in the Chinese A-share market, offering a robust investment strategy.

Reference