Application of Statistics to Portfolio Optimization

Siming Yi*
Department of Science, Nottingham University, Nottingham, United Kingdom
*Corresponding author: ssysy10@nottingham.ac.uk

Abstract. In the 21st century, with the advent of big data, data will emerge as the most powerful tool, enabling investment thoughts or ideas to be transformed into mathematical models. Backed by comprehensive data on real situations for analysis, employing large data sets will become a reasonable and effective approach for investment analysis, in particular, portfolio investment. The aim of this study is to gain a relatively more comprehensive understanding of the application of statistics to investment portfolios. It focuses on the principle of the factor model. The advantages and disadvantages of different factors in different situations are compared. In the experimental part, the same index, such as adjusted R2, is applied to the same data in different models for comparison. The experimental findings demonstrate that the three-factor model can more accurately fit the data and is more useful than the CAPM model. The three-factor model is inferior to the five-factor model, which is better suited to some specific interactions.

Keywords: portfolio optimization; factor model; statistics.

1. Introduction

In today's world, statistics are employed more and more frequently. Statistics can also play a significant role in the securities investment market. Statistics is a subject of general methodology and a tool to understand problems through quantitative analysis [1]. It can carry out structural analysis in the securities investment market. At the same time, the statistical method can also make value predictions to make theoretical pricing of securities issuance and listing prices. The price trend of securities and futures can be analyzed to measure risks. It can be seen that statistics have penetrated many subdivisions of securities investment and have become an indispensable part of it. In particular, statistics has made a significant contribution to portfolio optimization, a topic that has a constant concern of financial engineering [2].

In the complexity and uncertainty of today's financial markets, investors are facing enormous challenges, and how to effectively manage and optimize the investment portfolio has become an important issue. However, statistics provide scientific methods to quantify and interpret the risks and returns of financial markets, providing powerful decision support for investors.

This paper will explore the application of statistics in portfolio optimization. First, this paper introduces commonly used statistical methods, such as mean-variance theory, that can help investors evaluate and manage the risk of their portfolios to optimize asset allocation and risk management strategies. Following that, sophisticated statistical tools are introduced in this research. Those methods can help investors analyze and optimize portfolios more comprehensively and improve the accuracy and reliability of investment decisions. In addition, several cases will be shown to understand these models better.

2. Theoretical Basis

2.1. Mean-Variance Portfolio Optimization (MVPO) Model

2.1.1 Single-objective MVPO model

The mean-variance approach developed by Markowitz is a mathematical framework for portfolio optimization. It aids in the selection of the most efficient portfolio by evaluating different combinations of risky assets based on their expected returns and associated variances. By considering these factors, the mean-variance approach helps investors identify portfolios that offer a favorable trade-off between expected returns and risks [3].

Given a portfolio of n securities with yields of $r_1, r_2.., r_n$. An important problem facing investors is how to assign an appropriate weight to each security w_i, where $i = 1,2,\ldots,n$. The expected value vector R_i reflects the expected return of various securities. Variance σ_i^2 reflects the risk of the first security, and the covariance σ_{ij} reflects the correlation coefficient between the i security and the j security. The portfolio optimization approach can be mathematically modeled as follows:

$$\min \sigma^2 = \sum_{ij=1}^{n} w_i w_j \sigma_{ij} = \sum_{ij=1}^{n} \sigma_i^2 x_i^2 \sum_{i,j=1,i\neq j}^{n} \sigma_{ij} w_i w_j, \tag{1}$$

$$\sum_{i=1}^{n} w_i R_i = r, \tag{2}$$

$$\sum_{i=1}^{n} w_i = 1, \tag{3}$$

here $0 \leq w_i \leq 1, i = 1,2,\ldots,n$. And the supplied Eq. 1's goal is to reduce the amount of risk in the portfolio. Eq. 2 is used to determine the portfolio's projected return rate, and Eq. 3 makes sure that the entire budget is invested wisely. The single mean-variance (MV) model can be reformulated to prioritize maximizing returns while considering a predefined level of risk. This can be achieved by solving a model that aims to minimize risk while targeting a specific level of return or maximizing the return while maintaining a predetermined level of risk. By formulating the MV model in this manner, a portfolio can be obtained that aligns with the investor's desired trade-off between returns and risk.

Some researchers have transitioned from a single-objective approach to a multi-objective model to identify the efficient portfolio from numerous asset combinations within the solution space. By considering all objectives simultaneously, the multi-objective model allows for a comprehensive analysis of various factors when constructing an optimal portfolio. This multi-objective model enables researchers to make more informed decisions in portfolio management.

2.1.2 Multi-objective MVPO model

According to some researchers, MVPO mathematical model can be expressed in the following manner [4]:

$$\begin{cases} \min \boldsymbol{f}(\boldsymbol{x}) = \left(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}) \ldots, f_p(\boldsymbol{x}) \right) \\ \text{subject to: } \boldsymbol{e}(\boldsymbol{x}) = \left(e_1(\boldsymbol{x}), e_2(\boldsymbol{x}) \ldots, e_m(\boldsymbol{x}) \right) \leq \boldsymbol{0} \, . \\ \text{where: } \boldsymbol{x} = \left(x_1, x_2, \ldots, x_n \right) \in \boldsymbol{X} \end{cases} \tag{4}$$

In this model, the set of feasible solutions is determined by the constraints $e(x) \leq 0$. The vector x represents the decision variables or parameters, where $x = (x_1, \ldots, x_n)$, and X denotes the decision space. The value of p varies depending on whether it is a single-objective model (p = 1) or a multi-objective model (p > 1).

2.2. Factor Model (FM)

A FM is used to explain and forecast asset returns. The model makes the assumption that asset returns can be understood as a linear synthesis of a number of fundamental variables. These fundamental factors can be macroeconomic variables, industry indicators, or other security characteristics, and their impact on asset returns is statistically significant. Based on this assumption, factor models allow investors to identify the contribution of specific factors to asset returns and perform optimal portfolio management based on factor sensitivity.

2.2.1 Capital Asset Pricing Model (CAPM)

After the MV model was proposed, a number of economists simplified the original model from an empirical perspective. Sharp proposed the CAPM. The theory of CAPM is based on the situation that investors all use Markowitz's theory for portfolio investment, and the market equilibrium is formed

which is the price completely reflects the value. Currently, a linear model can be used to represent the relationship between projected asset return and risk [5]. The general consensus is that there is a correlation between an asset's projected rate of return and its risk, as evaluated by β. This model has now become the cornerstone of modern finance theory, providing a solid theoretical basis for understanding and analyzing investment decisions.
The details are shown in the following:

$$E(r_p) = r_f + \beta_p[E(r_m) - r_f]. \tag{5}$$

Where $E(r_p)$ is the expected rate of return of the portfolio, r_f represents the risk-free rate. And r_m is the market return.

The link between asset (excess) returns and market (excess) returns is examined by the CAPM using time-series regression. This involves selecting a specific time window, for example, the previous 163 trading days of the US stock market, and conducting a regression analysis to calculate the beta coefficient. Beta measures the sensitivity of an asset or portfolio to market movements within the chosen time frame. The mathematical analytical solution for beta is as follows:

$$\beta_i = \frac{Cov_{i.M}}{\sigma_M^2} \tag{6}$$

Where $Cov_{i.M}$ denotes the covariance between the return of asset i and the market return. σ_M^2 is the market return's variance.

Beta (β) represents the measure of the systematic risk faced by a portfolio or individual stock. It represents the shift in the portfolio's or stock's excess return rate relative to the market as a whole. In other words, beta measures how sensitive the returns on an asset are to changes in the larger market. It can be seen that it actually measures the degree of correlation between assets and markets relative to market volatility, which is also known as factor loading. The excess return portion of the market came to be known as the market factor, laying the theoretical foundation for the burgeoning school of factors that followed.

2.2.2 Fama-French three-factor (FFTF) model

Since the 1970s, academics have gradually found that a certain style of a class of stocks has a higher probability of beating the market. Basu discovered the earnings-to-market capitalization (EP) effect, Banz discovered the small market capitalization (MC) effect, and book-to-market (BM) and debt-to-market (DM) effects were also discovered. The size component (small-minus-big factor, SMB), as well as the value factor (high-minus-low factor, HML), were the explanations for the anomalies based on the ICAPM model that Fama and French came up with using the multi-factor theory of APT [6].

For this reason, the FFTF model has been developed to provide an explanation for the average return on assets. This particular model takes into account three key factors: MC, SMB, and HML, which represent the excess return of a portfolio comprising stocks with a high book-to-market ratio. By incorporating these factors, the Fama-French model seeks to better understand and analyze the sources of asset returns, offering a more comprehensive framework for assessing investment performance.

The correlation between average returns and asset MC is captured by market factors. The average return of smaller businesses is typically higher than that of larger ones.

SMB stands for the earnings spread between portfolios of small- and large-cap companies. It assesses the market performance and relative returns of small and large businesses. The SME index often shows a positive value when small-cap stocks outperform large-cap equities and a negative value when the converse is true. When comparing the returns of a portfolio of stocks with a high book-to-market (B/M) ratio with a portfolio of stocks with a low book-to-market ratio, HML is used to measure the difference. The HML factor aims to distinguish between growth stocks with low asset/market ratios and value firms with high asset/market ratios in terms of performance and relative returns. In general, a positive HML shows that value equities do better than growth companies, whereas a negative HML shows the opposite.

With those three factors, Fama and French rasied the equation which stands for the FFTF model.

$$R_{it} - R_{ft} = \alpha_i + \beta_i * \left(R_{mt} - R_{ft}\right) + \beta_{is} * SMB_t + \beta_{ih} * HML_t + \varepsilon_t . \quad (7)$$

2.2.3 Fama-French five-factor (FFFF) model

In-depth research on the three-factor model anomaly was undertaken by Fama and French, who discovered a high correlation between stock return, stock profitability, and investment growth rate. The book-to-market ratio [7] can be used to illustrate why this relationship is not MC's. The dividend discount model then accounts for profitability and the pace of investment growth. Additional elements like the profitability factor (robust minus weak, RMW), and the investment growth rate factor (conservative minus aggressive, CMA) are introduced in order to increase the FFTF model's explanatory power.

RMW, which is used to represent changes in stock returns due to profitability, is a collection of earnings differentials between a portfolio consisting of highly lucrative and poorly profitable stocks. Profitability is calculated by dividing operating income minus expenses for fiscal year t-1 by the book value at the end of fiscal year $t-1$. Its name is "operating profit."

The CMA is the variation between a portfolio's total asset growth rate between one with a lower growth rate and one with a greater growth rate. It demonstrates how stock returns have changed over time in relation to how quickly total assets have grown across various portfolios. Total new assets in fiscal year $t-1$ divided by total assets at the end of fiscal year $t-2$ is the formula for calculating investments.

$$\begin{gathered} R_{it} - R_{ft} = \alpha_i + \beta_{im} * \left(R_{mt} - R_{ft}\right) + \beta_{is} * SMB_t + \beta_{ih} * HML_t + \beta_{ir} * \\ RMW_t + \beta_{ic} * CMA_t + \varepsilon_{it} \end{gathered} . \quad (8)$$

3. The Contrastive Analysis

In this section, the comparative analysis of different factor models is applied to explore the characteristics of the different models. Next, several cases will be analyzed to identify those factor models.

3.1. Model Comparative Analysis

3.1.1 CAPM and FFTF model

In this case, the aim is to investigate the interpretability of the CAPM model and the three-factor model in developing countries [9]. Most of the studies on these two models are based on developed countries. In order to verify the completeness of these two theories, the case is chosen to explore a portfolio of securities traded in a developing country which is India.

It has selected the data of the Indian index SP BSE 500 from 2003 to 2019 (16 years). For the construction of those factors used in models, this case categorizes the stocks in the index BSE by SIZE and VALUE. For size (MC), it is divided into S (small) and B (big), which represent the bottom 50 percent and the bottom 50 percent respectively. For value, which is based on the book equity to market equity ratio, it is divided into L, M, and H, which represent the bottom 30 percent, the top 3 percent, and the rest middle portion respectively. These two indicators are arranged and combined to form six sets of portfolios. Then, these variables are calculated using six portfolios. Specifically, the market factor, which represents the risk premium, is obtained by comparing the monthly returns of the market against the risk-free rate of return. To account for company size, the SMB factor is utilized as a proxy. For each month throughout the sample period, this entails calculating the difference between the equally-weighted average returns of three portfolios made up of small companies and large stocks. It should be noted that the SMB factor specifically excludes the effects of book-to-market equity (BE/ME). Comparing the excess returns of two portfolios made up of high book value to market capital (BE/ME) stocks against the proportional average returns of firms with a low book-to-market equity ratio on a monthly basis yields the value proxy known as HML (high minus low), or HML. This component of the model aims to capture the value effect while being independent of the influence of company size.

Next, this experiment uses different rubrics to comparatively analyze the CAPM and the three-factor model such as f-statistics, standard deviation, etc. Of particular interest are the values of adjusted R2. The result lists the corresponding adjusted R2 values for different combinations of factors.

Initially, the researchers estimated the coefficients of the CAPM by considering only the market factor as an independent variable. The results demonstrated that the market factor played a predominant role in influencing stock returns. Moving on, the results also presents the outcomes of estimating the Fama-French three-factor model and its variants. By augmenting the market factor with additional factors such as the "value" factor or the "size" factor, the adjusted R2 values improved, indicating an enhanced capability of the model to elucidate asset returns. Notably, the researchers observed that the inclusion of the "size" factor resulted in higher adjusted R2 values for small stocks, implying a more pronounced size effect in this subsample. Conversely, for larger stocks, the introduction of the "value" factor tended to outweigh the impact of the size effect, leading to higher adjusted R2 values.

Comparing the CAPM results with the variation of the Fama-French model incorporating both the SMB (size) and HML (value) factors, it became apparent that the CAPM exhibited a higher adjusted R2 value. This suggested that the market factor played a primary role in explaining excess returns. However, when all three factors were integrated, the adjusted R2 values surpassed those of the CAPM. Furthermore, all three factors were statistically significant, except for the SMB factor in the B/M and B/H6 cases.

The Fama-French three-factor model surpassed the CAPM in terms of comprehensiveness and applicability for the environment of the Indian stock market, according to the findings.

3.1.2 FFTF and FFFF

The purpose of the experiment is to find out which of the FFTF and FFFF models better reflected the stock market before and during COVID-19. For this purpose, 30 industries of the US stock market were selected for portfolio analysis. In this case, OSL(Ordinary Least Squares） estimation was used to analyze each factor in the two models one by one[10]. The differences in the expression of each factor before and after the epidemic are analyzed comparatively.

The study findings revealed that the FFTF model's efficiency increased during the epidemic, although not as significantly as the five-factor model. In the model, the significance levels of the SMB and HML variables increased by 3.7 percent and 14.29 percent, respectively. These results suggest that the FFTF model is highly effective in explaining expected excess market returns.

On the other hand, in the five-factor model, the significance levels of the HML and CMA variables increased by 64.29 percent and 100 percent, respectively, indicating a significant improvement in efficiency during the pandemic. Therefore, it can be concluded that while the five-factor model may not be a suitable tool when the market becomes complex, it becomes an excellent tool when the market becomes simplified. Interestingly, although the HML variable is present in both the three-factor and five-factor models, its significance diminishes in the five-factor model, suggesting it's redundant in that context. This finding aligns with Fama and French's study, which demonstrated that removing HML did not enhance the explanatory power of average returns in U.S. data from 1963 to 2013 [8].

In the FFFF regression, the intercepts of the other FFTF were above zero and exceeded three standard deviations, while the HML regression was below one standard deviation. This implies that HML's larger average returns are already captured by its correlations with the other four factors, particularly the profitability and investment factors.

3.2. Discussion

The FFTF model is superior to the CAPM for industrialized nations like the U.S. stock market. It takes into account the market, size, and value variables that have an impact on cross-sectional dispersion in stock returns. However, its applicability may vary across different markets. For instance, in some specific situations in the Indian market, the value factor impacts equity returns, but the size factor does not. This shows that only certain market circumstances are appropriate for using the FFTF approach.

The FFFF model has been proposed as an alternative to the FFTF model. Studies indicate that the FFFF model offers a marginal improvement and can be considered the "least-bad" model. The FFFF model can more effectively address the issue of price anomalies in Australia, and it can accurately depict the relationship between Australia and the United States in light of their proximity in terms of geography and political economy [11].

The COVID-19 pandemic's effect on the FFFF model's applicability in the American market has been seen to negatively impair its goodness of fit, it is important to highlight. The R-squared value, a measure of explanatory power, experienced a significant drop during the initial three months of the pandemic. Further research is needed to understand the extent of this change in the R-squared value as the outbreak continues.

Valiet asserts that the combination of the new elements makes it more difficult to accurately describe a cross-section of stock returns, acknowledging that the three-factor model differs significantly from the five-factor model [12].

4. Conclusion

Portfolio optimization has consistently remained a subject of significant interest within academic circles. Exploring the most suitable model for the portfolio has garnered considerable attention from researchers. As research in this field has advanced, the associated theoretical models have undergone substantial transformations. Based on the mean variance principle, researchers have produced many models for portfolio optimization, especially, the factor model. This paper mainly introduces the principle of one to five-factor models and relevant comparative analysis cases. By considering the impact of multiple factors on portfolio returns, factor models are able to provide a more comprehensive and systematic approach to explaining and predicting asset performance. Investors can better understand the sources of portfolio risk and make decisions on risk management and asset allocation. Although factor models provide important insights and methods in portfolio applications, there are several challenges and limitations in research. For example, selecting appropriate factors and determining their weights is a complex task that requires careful consideration of the rationality of factor selection and the reliability of data. In addition, the factor model is based on historical data and statistical assumptions, and future changes in the market environment may affect the performance of the model. For example, the situation of developed countries mentioned in this paper may not be applicable to the difference of stock market before and after major events in India, a developing country. Therefore, it is necessary to be careful when using factor model for prediction.

In summary, factor models are a valuable tool in portfolio applications to help investors understand and manage the risk and return characteristics of portfolios. However, more research is required to enhance the factor model's accuracy and adaptability in order to better serve the needs of investors and the current environment.

References

[1] Metaxiotis K, Liagkouras K. Multiobjective evolutionary algorithms for portfolio management: A comprehensive literature review. Expert systems with applications, 2012, 39(14): 11685-11698.

[2] Ni H, Wang Y. Stock index tracking by Pareto efficient genetic algorithm. Applied Soft Computing, 2013, 13(12): 4519-4535.

[3] Rubinstein M. Markowitz's" portfolio selection": A fifty-year retrospective. The Journal of finance, 2002, 57(3): 1041-1045.

[4] Kalayci C B, Ertenlice O, Akbay M A. A comprehensive review of deterministic models and applications for mean-variance portfolio optimization. Expert Systems with Applications, 2019, 125: 345-368.

[5] Rossi M. The capital asset pricing model: a critical literature review. Global Business and Economics Review, 2016, 18(5): 604-617.

[6] Fama E F, French K R. Common risk factors in the returns on stocks and bonds. Journal of financial economics, 1993, 33(1): 3-56.

[7] Fama E F, French K R. A five-factor asset pricing model. Journal of financial economics, 2015, 116(1): 1-22.

[8] Fama E F, French K R. International tests of a five-factor asset pricing model. Journal of financial Economics, 2017, 123(3): 441-463.

[9] Sehrawat N, Kumar A, Nigam N K, et al. Test of capital market integration using Fama-French three-factor model: Empirical evidence from India. Investment Management & Financial Innovations, 2020, 17(2): 113.

[10] Li K, Duan Y. Research on the application of Fama and French three-factor and five-factor models in American industry. Journal of Physics: Conference Series. IOP Publishing, 2021, 1865(4): 042105.

[11] Chiah M, Chai D, Zhong A, et al. A Better Model? An empirical investigation of the Fama–French five-factor model in Australia. International Review of Finance, 2016, 16(4): 595-638.

[12] Connor G, Sehgal S. Tests of the Fama and French model in India. 2001.