An Empirical Study of the Markowitz Mean-Variance Model and the CAPM Model in the Stock Market

Rupeng Song[1], Chingfei Luo[2], Meilan Qiu[1*]

[1] School of Mathematics and Statistics, Huizhou University, Guangdong, Huizhou, 516007, China.
[2] School of Statistics, Beijing Normal University, Beijing, 100875, China.

* Corresponding author. Email: 1154512519@qq.com (R.S.); 202211011043@mail.bnu.edu.cn (C.L.); qiumeilan@hzu.edu.cn (M.Q.)
Manuscript submitted October 13, 2024; accepted February 10, 2025; published June 23, 2025.
doi: 10.17706/ijapm.2025.15.1.51-68

Abstract: This paper focuses on the practical application of Markowitz's mean-variance model and Capital Asset Pricing Model (CAPM) in the stock market. Stock data was obtained using the Tushare interface, data preprocessing and visualization were performed by using Python. Multiple stocks from the Shanghai and Shenzhen markets were analyzed using the mean-variance model, and an efficient frontier chart was created. Through empirical research, the optimal investment portfolio with the highest Sharpe ratio and lowest variance was identified, and their expected returns and volatility were compared. Finally, the CAPM model was fitted using the least square methods to evaluate the returns and volatility of individual stocks in relation to the market. Besides, the Fama-French three-factor model, which is a developmental model based on CAPM, was also used for the analysis. This study provides valuable investment strategy references for investors. It addresses the concerns of investors regarding the evaluation of portfolio returns and risks, achieving optimal asset allocation. Furthermore, it deepens the understanding of the significance of Markowitz's mean-variance model and the CAPM model in financial risk management.

Keywords: Markowitz mean-variance model, capital asset pricing model, Fama-French three-factor model, optimal portfolio, least square method, Sharpe ratio

1. Introduction

In recent years, as our people's incomes have been growing, their awareness of and enthusiasm for investment have also been increasing. People are increasingly inclined to invest in stocks and funds, wanting to grow their wealth, but wanting to minimize the risk and maximize the expected return as much as possible. So how do you get higher returns with manageable risk? This requires asset allocation, which is a portfolio issue. One of the four investment principles of investment science is diversification to reduce risk. The systematic study of diversification is known as portfolio theory [1]. This theory was pioneered by American economist Harry M. Markowitz in the early 1950s. The theoretical basis of the model is that investors in the investment market need to carefully consider the expected return and risk level of each security, as well as the degree of correlation between the securities, in order to select the optimal investment portfolio, thereby reducing investment risk, at the same time, it is necessary to take into account one's own risk-tolerance ability and investment preferences, in order to ensure that the investment strategy is in line with the individual's objectives. And Capital Asset Pricing Model (CAPM) [2] is built on the basis of portfolio theory, which simplifies the complex portfolio calculations into simple criteria for the fair assessment of asset prices, so that investors can more easily choose high-quality investment targets. However, the CAPM model faces several limitations, such as the difficulty in determining the value of β, the neglect of factors like market frictions and irrational behavior, and the exclusive focus on systematic risk while ignoring the specific risk of individual stocks. These limitations may make the CAPM model not as effective and accurate in practice as it is in theory. In the early 1990s, Eugene Fama and Ken French introduced the classic three-factor model [3], which added two dimensions-market capitalization and value-as risk factors to the CAPM model. As a result, the three-factor model re-examines the CAPM model and provides explanations for the drivers of long-term stock returns.

Since the introduction of the mean-variance model, there have been countless academic studies centered around this model. For example, Sun *et al*. [4] obtained the optimal portfolio with the largest Sharpe ratio and the optimal portfolio with the smallest variance through empirical research, compared and analyzed their expected return, standard deviation and Sharpe ratio, and gave the effective boundary of the asset portfolio. Zola [5] explored the issue of investors' investment strategies in the investment decision-making process based on the mean-variance model. Song and Chen [6] empirically analyze China's home furnishing industry based on the CAPM model, which is analyzed and verified using the CAPM model, so as to determine whether the CAPM model can reasonably analyze the price of assets. Li [7] used the BJS time series test to empirically investigate the traditional CAPM model and analyze its applicability to CSI 300 index constituents. Luo [8] applied the Fama-French three-factor model in China's A-share market and explained its performance and applicability comprehensively. Drawing on existing research results, this paper establishes a portfolio model by combining the Markowitz mean-variance model with the CAPM model for the data of twenty stocks from 2014 to 2024 on the Shenzhen Stock Exchange and the Shanghai Stock Exchange, and realizes the mining and analysis of the data using Python. The weights, risks and expected returns of each security corresponding to different conditions are finally derived, while the efficient frontier is drawn and investment recommendations are given for each asset in the portfolio. Finally, we will conduct a contrastive analysis of the three models based on the results of our experiments.

2. Preliminary Theoretical of Models

2.1. Markowitz Mean-Variance Modeling

The two core issues that investors value most are risk and expected return, so when assessing the risk and return of an asset investment, a balance needs to be found and an optimal asset allocation strategy adopted to achieve it. In his article, Markowitz discusses how the mean of risky assets can be used to predict expected returns and the variance or standard deviation can be used to assess the level of risk as a way to help investors choose the optimal portfolio of assets and make investment decisions. The main goal of the theory is to help investors make optimal investment decisions in the face of different risks and expected returns, in other words, to maximize investment returns or minimize investment risks.

2.1.1. Mean-variance model

The calculation formula is as follows:

Step one, expected rate of return on the investment portfolio [9]. The expected rate of return on a portfolio represents the likelihood that an investor will earn an average return on the portfolio. The overall expected rate of return on the portfolio depends on the combined effect of the expected rates of return on the various securities. The expression of formula:

$$E(R)_p = \sum_{i=1}^{n} \omega_i R_i \tag{1}$$

where $E(R)_p$ represents the expected return of the portfolio, R_i is the expected return of the i^{th} security, ω_i is the investment weight of the i^{th} security.

Step two, risks to the investment portfolio [10]. The degree of volatility of the various securities in a portfolio and their interrelationships have an impact on the degree of risk diversification in the portfolio. The standard deviation is used to measure the magnitude of the change in returns for each security, while the correlation coefficient is used to measure the correlation between the returns of two securities. The formula for its calculation is

$$\sigma_p^2 = \sum_{i=0}^{n}\sum_{j=0}^{n}\omega_i\omega_j Cov\left(R_i, R_j\right) = \sum_{i=0}^{n}\sum_{j=0}^{n}\omega_i\omega_j\rho_{ij}\sigma_i\sigma_j \tag{2}$$

where σ_p^2 is the variance of the portfolio, σ_i is the standard deviation of security *i*, σ_j is the standard deviation of security *j*, ω_i is the investment weight of security *i*, ω_j is the investment weight of security *j*, $Cov\left(R_i, R_j\right)$ is the covariance between security *i* and *j*, ρ_{ij} is the correlation coefficient between security *i* and security *j*, i.e., $\rho_{ij} = \frac{Cov\left(R_i, R_j\right)}{\sigma_i\sigma_j}$.

Eq. (2) reveals that the diversification effect of a portfolio is mainly determined by a combination of the correlation between two securities, the volatility of different securities, and the investment ratio. Portfolio risk can only be effectively reduced if less correlated securities are selected for the portfolio, and the correlation coefficient has a more significant effect on portfolio risk than the standard deviation.

2.1.2. Efficient boundary model

An investor typically chooses a portfolio mix from viable pools that will deliver the highest possible return with the lowest possible risk. The efficient set is also called the efficient frontier, it is a subset of the feasible set of portfolios that satisfy the following two conditions [11], (i) the portfolio with the lowest risk at a certain rate of return, (ii) the portfolio with the highest rate of return at a certain risk. Two important definitions in the model are the efficient asset portfolio and the minimum variance portfolio. A portfolio is an efficient mean-variance portfolio if it has the highest expected return subject to a specific variance constraint and has the smallest variance subject to a specific expected return requirement, then it is an efficient mean-variance portfolio. The set consisting of all mean-variance efficient asset portfolios is defined as the efficient frontier. A portfolio of assets is a minimum variance portfolio if it minimizes risk at a given level of expected return. The left side of the effective boundary is the outward curved shape formed by the upper edge of the feasible region, and the point of the minimum variance combination lies on the left side of the effective boundary, i.e., on the left boundary in Fig. 1.

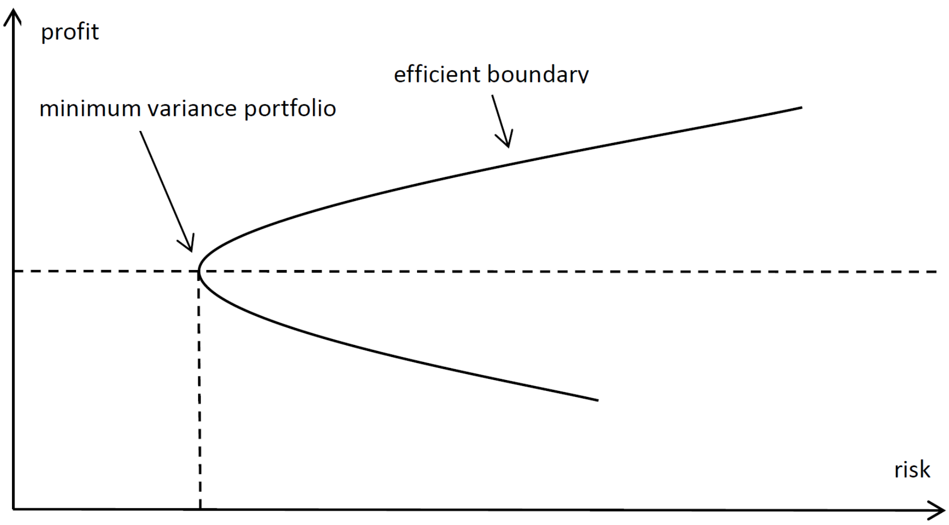

Fig. 1. Effective boundary model.

2.2. Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) was developed based on the mean-variance theory, which is based on Markowitz's assumptions that the stock market is efficient, assets are infinitely allocable, an investor can buy any stock, an investor chooses a portfolio based on the mean and the variance, an investor is risk averse and will never be satisfied, there are risk-free assets, and an investor is free to perform operations such as borrowing and lending [1]. Markowitz's portfolio theory, while having some theoretical value, is difficult to apply in practice because its calculations are too complex. Sharpe innovated on the theory by adding new assumptions, thus making the model more concise. It assumes that capital markets are completely perfect, with no transaction costs, that information is very liquid and easily accessible, that all investors can borrow at the same interest rate and have the same fixed investment horizon and expectations, and that investors have the same understanding of expected returns, standard deviation, and security covariance. Based on the above assumptions, the model can be expressed as

$$E(R_i) = R_f + \beta_{im}\big(E(R_m) - R_f\big),$$

where $E(R_i)$, R_f represents the expected rate and risk-free rate of return on a stock or portfolio respectively, $E(R_m)$ denotes the portfolio of the return on market, and β_{im} is a systematic risk measure for a stock or portfolio.

According to the Capital asset pricing model, it can be concluded that the determination of the expected rate of return requires the following three main factors. First, the risk-free rate of return R_f, it represents the time value of money and is usually based on the bank interest rate of three-month time deposits or the interest rate of one-year treasury bonds. Second, the market risk premium $E(R_m)–R_f$ it reflects the difference between the return on a market portfolio and the risk-free rate, which captures the reward due per unit of risk. Third, the risk factor β, it is the measurement of how a security responsive to the market portfolio fluctuations and is calculated by using the formulation $\beta = \frac{\sigma_i M}{\sigma_M^2}$, where $\sigma_i M$ denotes the covariance of any security S_i for the market portfolio M, σ_M^2 is the market risk.

Of course, the CAPM model has another significant factor to analyze, known as the asset's alpha, denot-ed by α and expressed as $\alpha = E(R_a)–E(R_e)$. It is calculated as the difference between the expected rate of return on actual investments and the expected rate of return in market equilibrium. The reasonableness of asset pricing can be judged according to the changes in the following factors: First, when α is greater than zero, it implies that asset prices are undervalued because the expected rate of return on real investment is higher than the expected rate of return in the market, and a higher rate of return can only be obtained by purchasing the asset at a relatively low price. Second, when α is less than zero, the expected return on real investment is lower than the expected return in the market equilibrium, which implies that the price of the asset is overvalued, and the investor is not able to obtain an adequate return. Third, when α is equal to zero, the expected return on real investment is equal to the expected rate of return in the market equilibrium, which suggests that the price of the commodity is considered to be a reasonable price [11].

In summary, the cornerstone of modern financial market price theory is the Capital Asset Pricing Model (CAPM), which has a wide range of practical applications in the areas of asset valuation, cost of capital budgeting, and resource allocation, despite its many assumptions and conditions that are difficult to fully satisfy. The CAPM model is widely recognized in the field of securities theory, which focuses on analyzing the sensitivity between security returns and changes in market portfolio returns to help investors determine whether the additional returns they receive are matched by the associated risks.

2.3. Fama-French Three-Factor Model

The Capital Asset Pricing Model suggests that the return on a stock is only linearly related to the systematic risk of the stock market as a whole. That is, the expected return of a stock is only related to the systematic risk of the market. However, Banz found that a stock's return is also linked to its market value. In a series of subsequent studies, book-to-market ratio (BE/ME), price-earnings ratio inverse (E/P), and several other metrics have been shown to explain stock price movements, indicating that stock prices are influenced by a range of risk factors [12].

Fama and French demonstrated that a three-factor model [13] can be developed to explain stock returns. The model suggests that the excess return of a portfolio (including individual stocks) can be explained by its exposure to three factors: the market asset portfolio R_m–R_f, the market capitalization factor (SMB), and the book-to-market ratio factor (HML). Although more than three factors have now been developed, these three are the most basic and important, and all other factors are based on it or even influenced by it. This multifactor equilibrium pricing model can be expressed as:

$$E(R_{it}) - R_{ft} = \beta_i[(E(R_{mt}) - R_{ft}] + s_i E(SMB_t) + h_i E(HMI_t),$$

where R_{ft} denotes the risk-free rate of return at time *t*, R_{mt} denotes the market rate of return at time *t* and R_{it} denotes the return on asset *i* at time *t*, respectively. SMB_t is the simulated portfolio return of the market capitalization factor at time t and HMI_t is the simulated portfolio return of the book-to-market ratio factor at time *t*. β, s_i, and h_i represents the coefficients of the three factors, respectively, and the regression model is expressed as follows:

$$R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_i SMB_t + h_i HMI_t + \varepsilon_{it}$$

3. Mean-Variance Modeling for Portfolio Data Analysis

There are five main steps to analyze the portfolio data. First, obtain past trading data of stocks using financial databases. Second, preprocess the data and store it. Next, use data visualization techniques to show the basic direction of the stocks. Then, calculate metrics such as the stock's annualized return, the volatility of its annualized return, the covariance between stocks and the correlation coefficients. Finally, construct 10,000 groups of stochastic portfolios, draw the efficient frontier, and find the return, volatility, and portfolio weight parameters of the optimal portfolio when the Sharpe ratio has maximum and minimum variance.

3.1. Sample Stock Selection and Data Acquisition

3.1.1. Stock selection

The stocks selected in this article include: Shanghai Stock Exchange (SSE) and Shenzhen Stock Exchange (SZSE) stock markets. For the benchmark index, the CSI 300 index was selected to represent the A-share market. On the one hand, it is a more recognized stock market index in the industry, and on the other hand, there are corresponding futures products tracking the CSI 300 index. Based on the results of the social network method analysis [14] and the performance of fund positions, it can be seen that influential fund managers prefer to hold positions in stocks with large outstanding market capitalization and low share price volatility.In addition, large-capitalization stocks have the advantages of good stability, low price volatility, strong financing ability, and suitability for medium- to long-term investment [15]. Therefore, in order to better represent the portfolio, satisfy the public investors as well as to be able to analyze it with more data, this paper selects stocks from China's Shanghai and Shenzhen stock markets based on the size of the market capitalization, and a total of 20 stocks are selected. Due to the need to analyze data over a longer period of time, the six newly listed stocks in the top 20 by market capitalization were excluded, including China Mobile, CATL, CNOOC, China Telecom, PSBC, and Foxconn. Six more stocks are added based on market capitalization, including Wuliangye, BOCOM, Citic Securities, East Money, Zijin Mining, and CIB. Based on data availability, the years 2014 through 2024 were selected as the sample period.

3.1.2. Data acquisition

Business analysis through data collection is an activity that many securities firms engage in, with influential module libraries such as Tushare, Rqdatac, and Windpy. Using Tushare, a free open source financial data interface, we can easily access historical stock trading data. Tushare covers a large amount of financial data, including basic stock lists, listed company information, trading calendars, daily quotes, mid- and high-frequency minute quotes, compounding factors, income statements, balance sheets, cash flow statements and other data, which can satisfy most of the quantitative trader data needs, and it is a suitable tool for all types of investment and financial research. This article primarily analyzes 20 stocks, which include the following: ICBC (601398), Kweichow Moutai (600519), ABC (601288), PetroChina (601857), CCB (601939), BOC (601988), China Life (601628), China Merchants Bank (600036), Ping An (601318), China Shenhua (601088), BYD (002594), Yangtze Power (600900), Sinopec (600028), Midea (000333), Wuliangye (000858), BOCOM (601328), Citic Securities (600030), East Money (300059), Zijin Mining (601899), and CIB (601166). Tushare's third-party financial database needs to be installed and upgraded to the latest before you can access the data. Before getting the data in Python you also have to go to the Tushare website to register as a Tushare community user and go to the user center to view your personal Token credentials. Next the number of the stock you want to find can be searched in its data tool. After the above steps are completed, you can use Python to get the relevant stock data. This paper focuses on obtaining the closing price of each trading day of the twenty stocks listed above from October 9, 2014 to October 9, 2024, a total of 48,216 pieces of trading data.

First we need to import the Tushare database and initialize the pro interface, then define a list to store the desired stock codes. Subsequently, you can use the for loop statement to iterate through each stock and get the ticker symbol, trade date as well as the closing price. Table 1 shows the first sixteen data in the acquisition data from which we can observe the acquisition data.

Table 1. Acquired Raw Stock Data

unnamed:0	ts_code	trade_date	close
0	601398.SH	20241009	6.04
1	601398.SH	20241008	6.12
2	601398.SH	20240930	6.18
3	601398.SH	20240930	6.01
4	601398.SH	20240926	6.20
...	...	...	...

3.2. Preprocessing of Stock Data

Since the data provided by Tushare did not meet the needs of our subsequent analysis, we needed to preprocess the data before proceeding with the analysis [16]. We chose to use the Pandas library to improve the efficiency of data processing, specifically including deleting rows, composite indexes, rows to columns, and other operations on the raw data. After data preprocessing, the data is consolidated into 2,433 data items that contain the date, the stock code, and the closing price for each trading day corresponding to each stock code. The final data is shown in Table 2 below.

Table 2. Processed Data

transaction date	000333.SZ	000858.SZ	002594.SZ	300059.SZ	600028.SH
09/10/2014	20.23	18.62	50.32	15.57	5.31
10/10/2014	20.05	18.35	49.5	15.04	5.27
11/10/2014	19.92	18.02	49.08	15.59	5.19
12/10/2014	19.87	17.94	48.25	15.54	5.2
...	...	...	...	...	...

3.3. Visualization of Stock Data Trends

Using the Matplotlib data visualization library in Python, the changes in stock prices between October 2014 and October 2024 are presented. By first normalizing the stock prices by October 2014, and subsequently applying the plot() function to read and visualize the stock data, we can observe the basic movements of the twenty stocks over a period of less than ten years. The results are shown in Fig. 2 below.

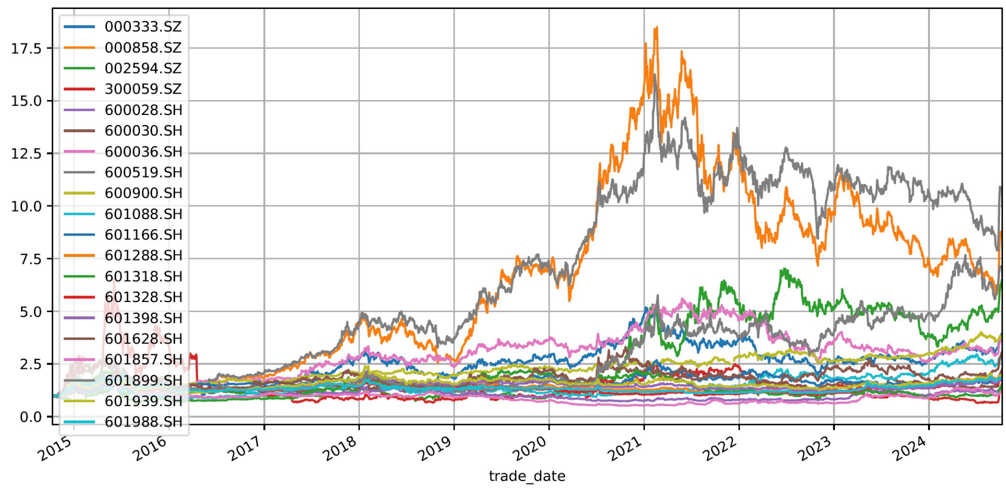

Fig. 2. The fundamental trends of the 20 selected stocks from October 2014 to October 2024.

3.4. Calculate Some Parameters of the Stock

We will use the mean function, the volatility function, the covariance function, and the correlation function to estimate for each stock their annualized average stock price return, annualized stock return volatility, covariance with each other, and correlation coefficients. The results of the annualized average returns on stock prices that we obtained using the mean function can be seen in Table 3.

Table 3. Average Annualized Return on Stock Price

Stock code	Average annualized return on stock price
000333.SZ	1.151363
000858.SZ	2.021022
002594.SZ	1.674662
300059.SZ	−0.201253
600028.SH	0.882634
600030.SH	1.027935
600036.SH	0.914212
600519.SH	2.434457
600900.SH	1.093581
601088.SH	1.597046
601166.SH	0.411518
601288.SH	0.838829
601318.SH	0.921685

601328.SH	0.470896
601398.SH	0.553960
601628.SH	1.483615
601857.SH	0.467429
601899.SH	1.705597
601939.SH	0.660290
601988.SH	0.620487

The results of the annual volatility of stock returns that we obtained using the volatility function can be seen in Table 4.

Table 4. Annual Volatility of Stock Returns

Stock code	Average annualized return on stock price
000333.SZ	1.176512
000858.SZ	1.160584
002594.SZ	1.324901
300059.SZ	1.730840
600028.SH	0.753373
600030.SH	1.069792
600036.SH	0.898313
600519.SH	0.941481
600900.SH	0.606754
601088.SH	1.017578
601166.SH	0.813331
601288.SH	0.592542
601318.SH	0.889959
601328.SH	0.643043
601398.SH	0.598717
601628.SH	1.148327
601857.SH	0.846626
601899.SH	1.208303
601939.SH	0.709728
601988.SH	0.628622

The results of covariance and correlation coefficients for each stock that we obtained using covariance function and correlation function can be referred to as shown in Tables 5 and 6:

Table 5. Covariance Matrix for the 5 of 20 Stocks

Covariance matrix	000333.SZ	000858.SZ	002594.SZ	300059.SZ	600028.SH	...
000333.SZ	1.384179	0.683485	0.446898	0.641710	0.200278	...
000858.SZ	0.683485	1.346955	0.557920	0.744513	0.157421	...
002594.SZ	0.446898	0.557920	1.755362	0.824342	0.094099	...
300059.SZ	0.641710	0.744513	0.824342	2.995807	0.287379	...
600028.SH	0.200278	0.157421	0.094099	0.287379	0.567571	
...	...	...	...	...	...	...

Table 6. Correlation Coefficient Matrix for the 5 of 20 Stocks

Covariance coefficient matrix	000333.SZ	000858.SZ	002594.SZ	300059.SZ	600028.SH	...
000333.SZ	1.000000	0.500560	0.286701	0.315127	0.225957	...
000858.SZ	0.500560	1.000000	0.362838	0.370628	0.180043	...
002594.SZ	0.286701	0.362838	1.000000	0.359474	0.094274	...
300059.SZ	0.315127	0.370628	0.359474	1.000000	0.220388	...
600028.SH	0.225957	0.180043	0.094274	0.220388	1.000000	
...	...	...	...	...	...	...

3.5. Simulation of Large Randomized Portfolios

In order to better understand the returns and volatility of different portfolios, we used a stochastic function to generate 10,000 sets of random weights, each containing sixteen numbers representing the percentage of each stock in the overall investment. We apply the random random function from the Numpy library to generate the stock weight matrix. Next, we need to measure the return and volatility of these 10,000 portfolios. Finally, with the help of the Matplotlib library, we present the dataset containing 10,000 data points as in Fig. 3, so that we can visualize the performance of the portfolio and perform the necessary analysis.

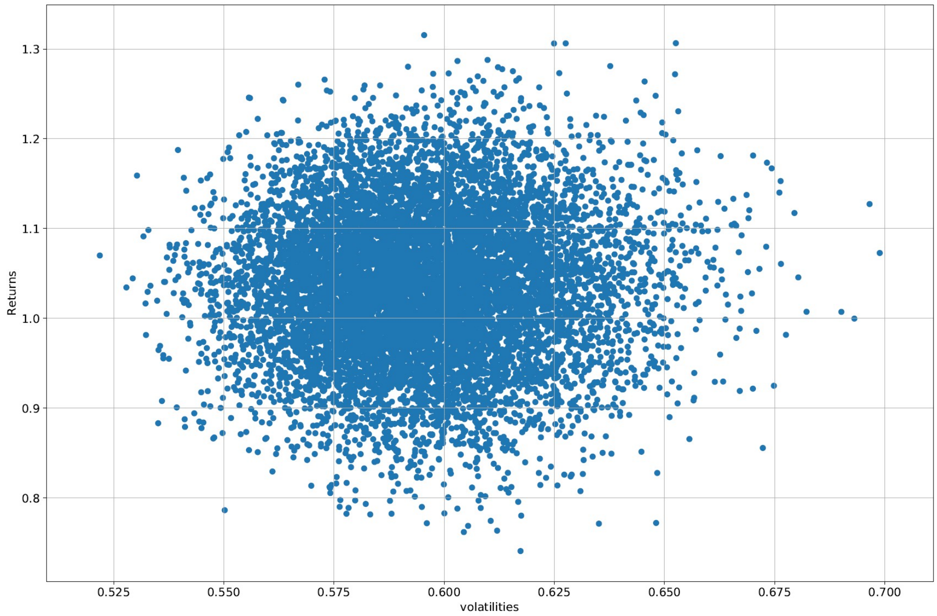

Fig. 3. The relationship between volatility and returns of different Portfolios.

3.6. Find the Optimal Portfolio That Minimizes Variance

Minimizing the risk of a portfolio of assets involves choosing a portfolio with minimum variance, i.e., the minimum variance when assets are combined in different proportions [17]. In building a portfolio of risky assets, portfolios of assets with different levels of risk and return need to be allocated to a range that is shaped like a bullet. Minimizing the variance of an asset portfolio implies that between the asset portfolios corresponding to the two ends of the bullet-shaped range, there exists an optimal asset portfolio with the same level of return but less risk. Of all the minimum variance asset portfolios, the portfolio at the leftmost end of the range has the smallest variance and is therefore called the minimum variance asset portfolio, which consists of the combination of all the end asset portfolios in the set. Seeking a portfolio of assets that minimizes variance is not necessarily the same as seeking a portfolio of assets that minimizes variance, only a portfolio of assets that minimizes variance can truly achieve the goal of minimizing variance. The minimum variance asset portfolio is not necessarily the optimal portfolio because the lowest risk tends to correspond to the lowest expected return at the efficient frontier [18]. Firstly, we introduce the optimize board in the Scipy library, then we define a function for solving optimization and a function for solving minimum variance, and set the constraints, i.e., the sum of the weights is equal to 1, and solve the problem through the minimize function, and the starting parameter list adopts a uniform distribution, i.e., all the weights of the 20 stocks are 0.1, and the final optimal portfolio weight vector with the minimum variance is [0.00000000e+00, 2.91126926e–17, 6.60140062e–02, 0.00000000e+00, 5.39138005e–02, 4.40489565e–19, 3.39165337e–17, 9.55159124e–02, 3.43927427e–01, 5.94045542e–17, 0.00000000e+00, 2.24485622e–01, 3.62789786e–17, 2.49915515e–02, 1.79349088e–01, 8.70413374e–18, 1.51282123e–18, 1.08767195e–02, 0.00000000e+00, 9.25872838e–04], the variance of the portfolio is 0.45479074486140575, and the expected return of the portfolio is 1.0853305515935308.

3.7. Find the Optimal Portfolio When the Sharpe Ratio is Maximized

The Sharpe Ratio [13], also known as the Sharpe Index, is a standardized metric for evaluating a fund's performance, which reflects a risk-adjusted rate of return. The Sharpe Ratio is an indicator used to assess the return performance of an underlying investment, such as a fund, relative to a risk-free rate. The magnitude of the indicator reflects the relationship between a fund's risk-taking ability and the level of return, with a higher value implying a more favorable risk-to-return ratio and vice versa. The primary role of the Sharpe ratio is to measure the relationship between a portfolio's risk-taking ability and excess return performance. It lacks a fixed standard to measure value and can only be shown by comparing it to other products. The expression for the Sharpe ratio is: Sharpe Ratio is equal to $\frac{E(R_p)-R_f}{\sigma_p}$. In this paper, we will look for the portfolios with the largest Sharpe ratios out of the 10,000 random portfolios listed above. First, we define the median interest rate of 2.9% for ten-year Treasury bond rates from 2014 to 2024 as the risk-free rate of return. Additionally, we define the number of assets as twenty. Next, we define a statistics function to record the expected return, standard deviation, and Sharpe ratio of the portfolio [3]. In order to achieve the goal of minimizing the negative value of the Sharpe ratio, we use the optimize module in the Scipy library, and under the constraint that the sum of the weights is 1, the minimize function is used to carry out the calculations, and at the same time, the boundary constraints are defined, where we constrain the range of values of the weights of each asset to be 0 to 0.1, and after the constraints are completed, the solution can be carried out for the planning. The final optimal portfolio weight vector, which maximizes the Sharpe ratio, is [0.06529844, 0.1, 0.1, –0.16009687, 0.1, 0.1, 0.08829217, 0.1, 0.1, 0.1, –0.22924, 0.1, –0.00876801, 0.02829231, 0.1, 0.1, –0.08377804, 0.1, 0.1, 0.1]. The expected return, volatility, and Sharpe ratio of the current portfolio are 1.71927608, 0.5993914 and 2.81998719, respectively. These values are based on the specified weight ranges, which should be allocated equally among the twenty stocks.

4. Equity Investment Data Analysis for Capital Asset Pricing Models

The Capital Asset Pricing Model (CAPM) has important implications for the valuation of assets, the cost of capital budgeting, and the allocation of resources. Through the use of the capital asset pricing model, it is possible to assess whether the value of assets in the securities market is reasonably priced in order to detect price errors in the market. An important application in asset allocation is to select different securities or portfolios and predict the coefficients based on the market trend, which can help us to get higher returns or avoid market risks.

4.1. Examples of Calculating the CAPM Model for 20 Assets

In this section, we will analyze one of the above 20 stocks and compare the returns and volatility of a single stock to the market for determining whether it is worth investing, and all the results are obtained by using Python. We use the CAPM model for stock analysis in the same way as the mean-variance model used above to obtain the data, from the open source financial database Tushare. First, we need to calculate the daily risk-free rate. In this article, we assume that the risk-free rate is 2.9% (as defined in Section 3.7), and since the CAPM model is analyzed using daily returns, the average daily risk-free rate is demanded. The formula for converting the annual interest rate to a daily rate is $R_{\mathrm{fday}} = (1 + R_{\mathrm{fyear}})^{\frac{1}{365}} - 1$, The calculated daily interest rate is 7.94×10^{-5}.

Next, the CAPM model equation $E(R_i) = R_f + \beta_{im}(E(R_m) - R_{f)}$ is utilized to calculate the market risk premium and the stock excess return, respectively. We use the CSI 300 index returns to represent the market index returns, with data collected from 2014 to 2024. The results are shown below in Table 7.

Table 7. Calculated Market Risk Premium

Trade_date	ICBC	KweichowMoutai	ABC	PetroChina	CCB	...
09/10/2014	−0.002879	−0.004579	−0.000079	0.003721	−0.000079	...
10/10/2014	−0.002879	−0.006279	−0.004079	−0.006479	−0.004979	...
13/10/2014	−0.008579	−0.018479	−0.008179	−0.005179	−0.007479	...
14/10/2014	−0.000079	−0.007679	−0.000079	−0.001379	0.004921	...
15/10/2014	−0.000079	0.015521	0.004021	0.001221	−0.000079	
...	...	...	...	...	...	...

Then, the scatter plots of risk premiums are drawn to see the general relationship between the CSI 300 and the market index return. The results of the execution are shown below in Fig. 4.

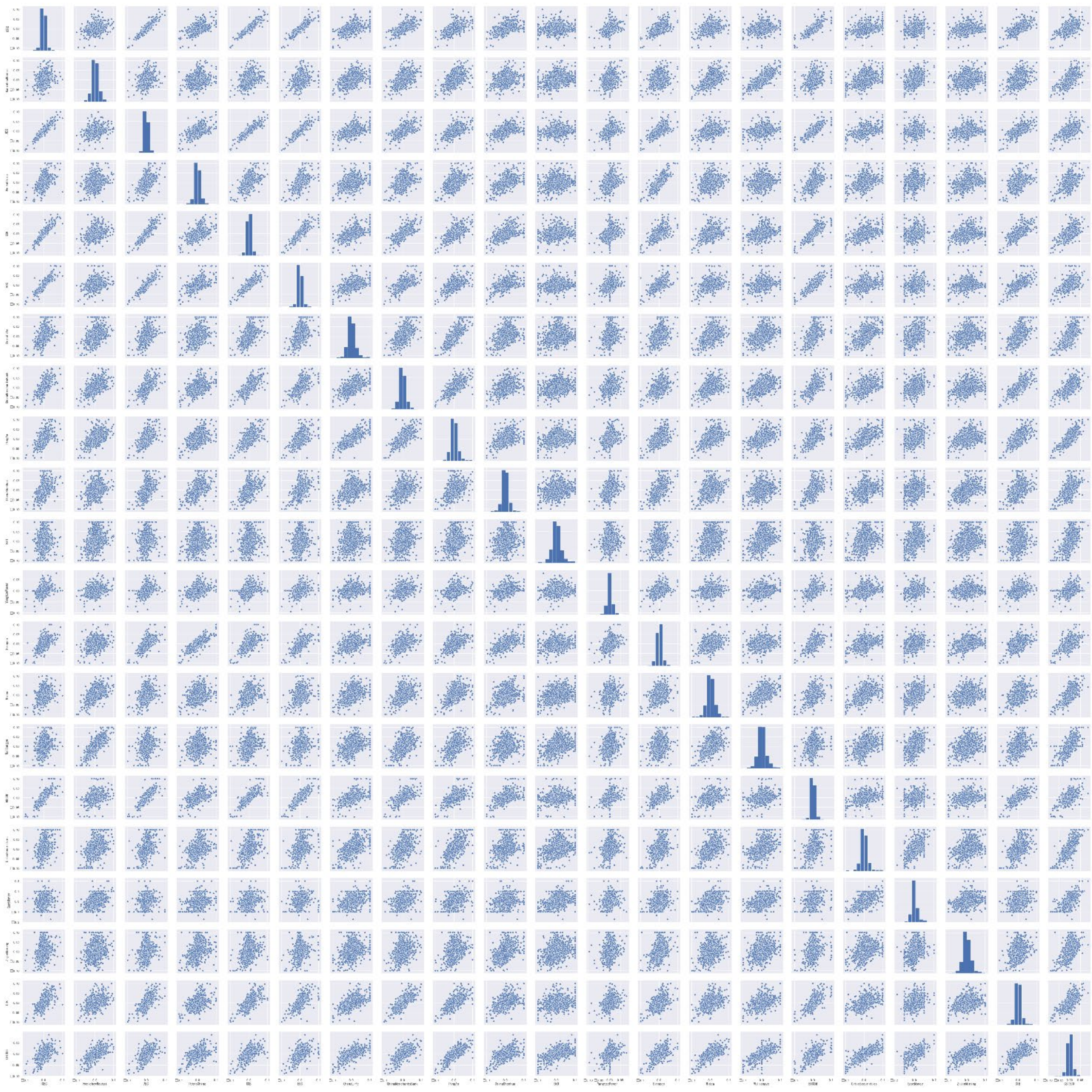

Fig. 4. The relationship between the CSI 300 and the market risk premium for 20 stocks.

Finally, linear regression calculations were performed using the least squares method to fit the CAPM model. The regression results for ICBC are listed in Table 8 below.

Table 8. Regression Results for ICBC (601398)

OLS	coef	std err	t	$P > \lvert t \rvert$	[0.025	0.975]
const	0.0004	0.000	1.608	0.108	–8.31e–05	0.001
CSI300	0.4601	0.017	27.501	0.000	0.427	0.493

According to the regression results for ICBC showed in Table 8, we can get that α = 0.0004, which suggests that in addition to the gains from broader market fluctuations, own-value can generate an additional 0.04 percent, and α is greater than zero indicated that the asset is undervalued. We also know that β = 0.4601, it shows that the volatility of ICBC is more than 0.46 percent compared to the broader market index. We also obtained the simulation results for KweichowMoutai(600519) and listed in Table 9 As is can be

Table 9. Regression results for KweichowMoutai (600519)

OLS	coef	std err	t	$P > \lvert t \rvert$	[0.025	0.975]
const	0.0009	0.000	3.137	0.002	0.000	0.002
CSI300	0.8780	0.021	40.846	0.000	0.836	0.920

Seen from Table 9, we have α = 0.0009, this means that the intercept term is insignificant. And the value of β is 0.8780, it indicates that if the broader market rises by 10%, the KweichowMoutai will expect to rise by 8.78%.

Similarly, repeating the same steps for the rest of 18 stocks, we can obtain the following results showed in the following Table 10,

Table10. Regression Results for the Rest of 18 Stocks

Stock code	α	β
601288.SH	0.0004	0.4757
601857.SH	0.0001	0.6306
601939.SH	0.0004	0.6175
601988.SH	0.0004	0.4877
601628.SH	0.0004	1.1111
600036.SH	0.0006	0.8634
601318.SH	0.0004	1.0198
601088.SH	0.0007	0.7899
002594.SZ	0.0008	1.1215
600900.SH	0.0006	0.3027
600028.SH	0.0003	0.6308
000333.SZ	0.0008	1.0035
000858.SZ	0.0009	1.1387
601328.SH	0.0004	0.5967
600030.SH	0.0004	1.2788
300059.SZ	0.0012	1.5904
601899.SH	0.0009	0.9274
601166.SH	0.0003	0.7928

From the results in Table 10, it can be seen that the values of α are all slightly flat or marginally overvalued. Among them, East Money has the largest α and the largest excess return for the same volatility scenario of the market. Besides, Citic Securities, Wuliangye, BYD and China Life have a large values of β, which indicates a relatively high risk among the twenty stocks. At the same time, we obtained the smaller β

values of Yangtze Power, ABC and BOC, this indicates that the three stocks have a less risky. Therefore, investors can choose to invest in stocks based on their own risk preferences. Investors who pursue high risk and high return can choose East Money, while those who are conservative can focus on Yangtze Power. Furthermore, if investors want to diversify their investments and build investment portfolios, they can allocate their holdings based on this indicator [19].

4.2. Fama-French Three-Factor Model

Based on the theoretical arguments above, we know that the CAPM model has many flaws for empirical analysis. So, instead, we use the optimization model of the CAPM model, the three-factor model, to conduct an empirical analysis of the above 20 stocks. In this paper, we only analyze these three factors at the most basic level. Although the other mined factors appear to be valid factors, many of them are in strong covariance with these three factors. The data we utilized was also procured from Tushare, spanning from October 2014 to October 2024. In this section, we obtained three factor data from the factor model data collected by the Central University of Finance and Economics. These three factors are: market risk premium $(R_M - R_F)$, small market capitalization minus large market capitalization (SMB), and high book-to-market ratio portfolio returns minus low book-to-market ratio portfolio returns (HML). The market risk premium is the return on the market portfolio minus the risk-free rate. The SMB calculates the difference between the returns of a portfolio of stocks with small market capitalization and the returns of growth stocks. The HML calculates the difference between the returns of a portfolio of value stocks and the returns of growth stocks.

Here we use the daily stock closing price data obtained above. Total of 2,432 trading days. Next we calculate the daily returns for each stock and combine the factor data with the daily returns. Then, we can also draw scatter plots and heat maps to compare the correlation between returns and the other three factors. For example, the fitting results of the ICBC(601398) are shown in Figs. 5 and 6.

We can see that the linear relationship between yield and the other three factors may not be very obvious. However, the correlation between yield and the market risk premium factor is somewhat larger. Images are always less intuitive than data. As a result, we are going to use the accurate data to analyze. There are four common investment return metrics used in the following calculation, including total returns, max drawdown, information ratio, and Sharpe ratio which has been described above. Total Returns is the aggregate return over the entire back testing time period of the strategy, and the calculate formulation is $\frac{PV_{end}-PV_{start}}{PV_{start}} \times 100\%$, where PV_{end} denotes the total final stock and cash value of the strategy, PV_{start} denotes the total value of the strategy's starting stock and cash. Max drawdown is the percentage decline from the highest value of the net worth to the greatest loss experienced over the entire time period. It is closely related to losses, meaning a smaller max drawdown is preferable. The calculation is given by Max$(P_x - P_y)/ P_x$, where P_x is the highest value of the net worth and P_y is the value at a subsequent point in time. The information ratio measures the excess return per unit of excess risk: i.e., the higher the ratio, the better the performance. The formulation is expressed as $\frac{R_p-R_m}{\sigma_t}$, where R_p denotes the annualized return of the strategy, R_m refers to the benchmark annualized rate of return and σ_t represents the annualized standard deviation of the difference in daily returns between the strategy and the benchmark. All the results are shown in Table 11:

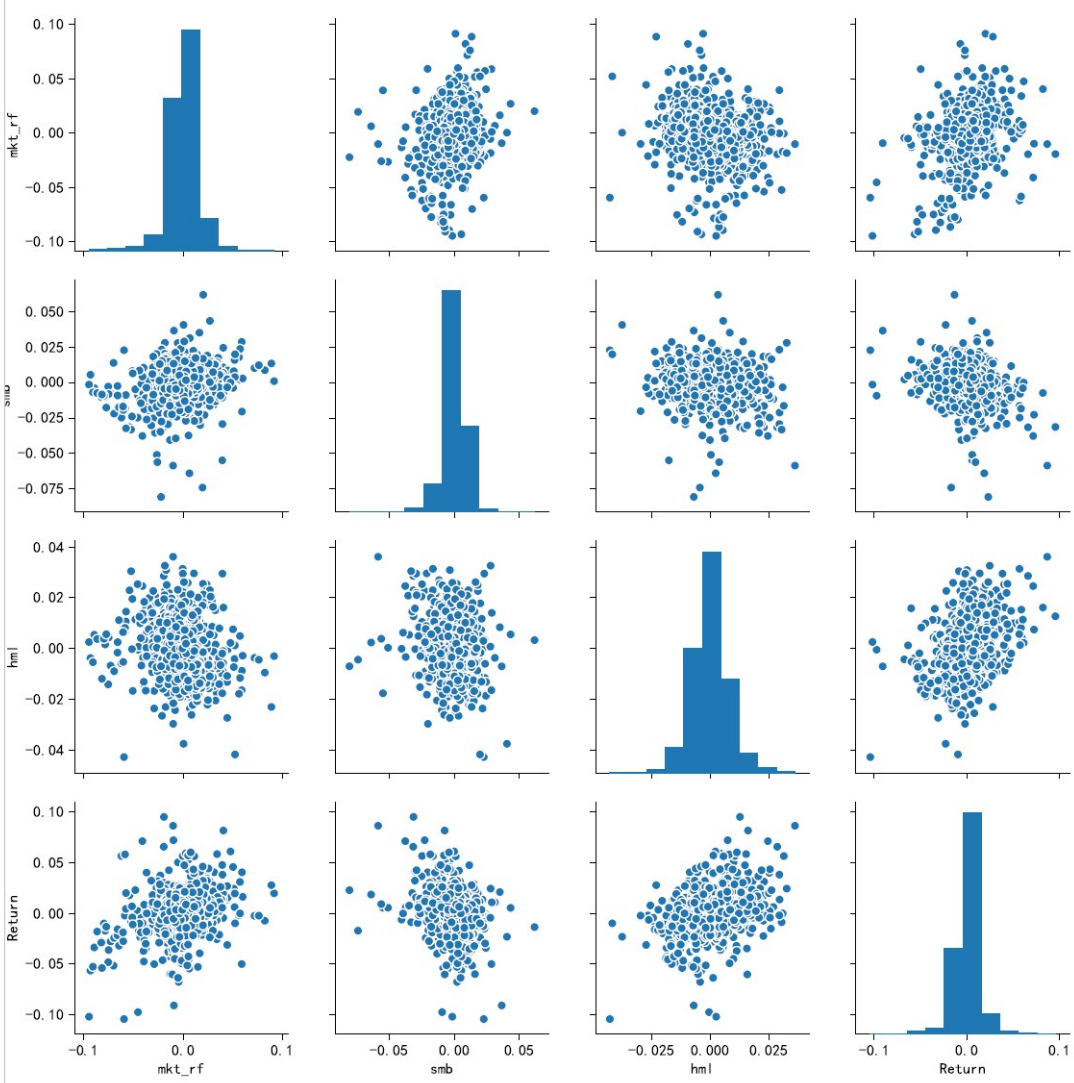

Fig. 5. The scatter plot.

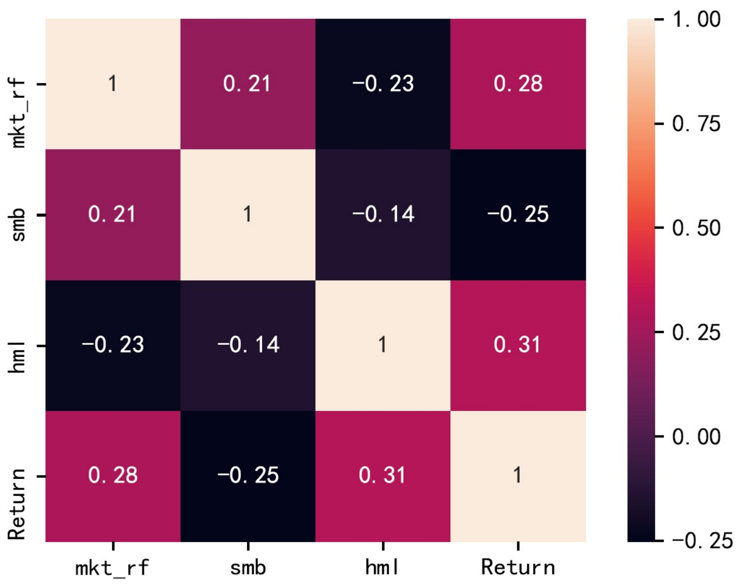

Fig. 6. The heat map.

Table 11. Indicators of the 20 Stocks

Stock code	α	β	SMB	HML	total returns	max draw-down	Sharpe ratio	information ratio
sh.601398	0.000070	0.372938	−0.441232	0.639511	0.715909	0.476129	−2.105626	−2.470176e+06
sh.600519	0.000934	0.681864	−0.883204	−0.439357	8.969064	0.515186	−1.470597	−1.725861e+06
sh.601288	0.000084	0.404903	−0.405572	0.691987	0.875502	0.427368	−2.124978	−2.492904e+06
sh.601857	−0.000303	0.692531	−0.358875	1.097309	0.077806	0.723549	−1.581914	−1.855781e+06
sh.601939	0.000051	0.521717	−0.469032	0.788609	0.916870	0.465851	−1.802729	−2.114973e+06
sh.601988	0.000045	0.423723	−0.377873	0.739064	0.781481	0.467857	−1.984734	−2.328402e+06
sh.601628	0.000163	0.856570	−0.659994	0.564085	1.774522	0.535797	−1.211525	−1.421845e+06
sh.600036	0.000345	0.646894	−0.707125	0.499099	2.541627	0.541538	−1.507065	−1.768450e+06
sh.601318	−0.000161	0.923390	−0.650102	0.562067	0.329356	0.730493	−1.296863	−1.521654e+06
sh.601088	0.000010	0.862009	−0.557856	1.271162	1.611650	0.505896	−1.320356	−1.549423e+06
sz.002594	0.000562	1.038059	−0.304847	−0.587721	5.074921	0.553947	−1.016024	−1.192942e+06
sh.600900	0.000425	0.284882	−0.309150	0.403923	2.646232	0.243028	−2.261837	−2.653602e+06
sh.600028	−0.000232	0.675229	−0.352901	1.037419	0.239171	0.552480	−1.736064	−2.036621e+06
sz.000333	0.000379	0.889650	−0.782799	0.000034	2.707860	0.624872	−1.239947	−1.455281e+06
sz.000858	0.000773	0.941769	−0.931254	−0.476504	6.931257	0.687311	−1.210466	−1.420896e+06
sh.601328	−0.000009	0.511230	−0.426829	0.748018	0.643357	0.544681	−1.888812	−2.215873e+06
sh.600030	−0.000007	1.084718	−0.398528	0.439392	1.266617	0.648334	−1.238803	−1.453750e+06
sz.300059	−0.000125	1.415365	−0.214936	−0.524733	0.599229	0.902180	−0.804270	−9.443943e+05
sh.601899	0.000430	0.965050	−0.359533	0.675726	5.665354	0.622563	−1.143139	−1.341946e+06
sh.601166	0.000019	0.636995	−0.530452	0.661888	0.801549	0.491165	−1.633524	−1.916512e+06

4.3. Summary of Investment Strategies

The traditional Capital Asset Pricing Model (CAPM) facilitates the examination of the quantitative relationship between capital gains and risk, allowing for an assessment of the reasonableness of a stock's price. A high risk coupled with a low price, or vice versa, it suggests that the price may be unreasonable and could be subject to change. Excess returns for a stock can be calculated using the values of alpha (α) and beta (β). Experimental results indicate that, according to the original CAPM model, East Money, Zijin Mining, and Wuliangye exhibit high alpha values, whereas PetroChina, Sinopec, and China Construction Bank (CIB) display low alpha values, other stocks possess similar alpha values.

In contrast, the three-factor model reveals that Kweichow Moutai, Wuliangye, and Zijin Mining also have higher alpha values, while PetroChina and Sinopec maintain lower alpha values. Notably, the beta values derived from both models are quite similar. Although the CAPM model has theoretical limitations due to its less rigorous assumptions, and the Fama-French three-factor model is often regarded as more suitable for empirical applications, the experimental results show that both models provide remarkably similar analyses of these 20 stocks. This minimal difference does not decisively ascertain which model is superior.

However, we can conduct an investment strategy analysis based on the common conclusions drawn from both models. Based on the principles of high returns and low risk, Wuliangye and Zijin Mining are identified as the most favorable stocks for investment, while PetroChina and Sinopec are considered the least attractive options. In summary, the conclusions regarding stock selection are derived from a thorough analysis of the data provided by the models.

Technical analysis aligns with the weak-form efficient market hypothesis, and investors are also influenced by public and insider information based on their access and analytical skills. Furthermore, investment choices vary due to different risk appetites, preferences for certain companies, and irrational factors. Nonetheless, the conclusions drawn from this model offer valuable guidance for investors.

4.4. Comparative Analysis in Practical Applications

From the efficient frontier illustrated in Fig. 3, it is evident that our selected portfolio is optimal. To further validate its effectiveness, we can compare the experimental results with those of a portfolio that reflects the average diversification of these 20 stocks. The expected return, standard deviation, and Sharpe ratio of the equally weighted portfolio are calculated to be 1.03649823, 0.58646498 and 1.71791712, respectively. The portfolio with equal proportional diversification demonstrates a negligible difference in risk compared to the optimal risky portfolio identified using the mean-variance model; however, the return difference is approximately seven percent. Additionally, we can compare these results with the annualized returns of ETF funds in the market, which have also exhibited moderate performance [20]. The fitting effectiveness of the CAPM model and the Fama-French three-factor model [21] can be evaluated using the following indicators:

- **R-squared:** The R-squared is also known as the coefficient of determination, represents the proportion of the variance for a dependent variable that is explained by one or more independent variables in a regression model. It is a key indicator for assessing a model's goodness of fit. R-squared ranges from 0 to 1, the closer the R^2-value is to 1, the better the model fits the data.
- **Adj.R-squared:** The adjusted R-squared is a modified version of R-squared and serves as an im-portant statistical measure for evaluating the goodness of fitting a model in multiple linear regression analysis. Compared to R-squared, the adjusted R-squared more accurately reflects the model's true predictive ability when estimating the variance of errors. Generally speaking, the closer the adjusted R^2-value is to 1, the better the model's interpretation of the data. Consequently, when comparing different models, we tend to prefer the one with a higher adjusted R-squared value.
- **Statistical Significance:** The parameter estimates of the model should be statistically significant. The t-test, sometimes referred to as the F-test, is a statistical method used to ascertain the significance of the parameters within a given dataset. If the *p*-value of a parameter estimate is less than the significance level (e.g., 0.05), it is considered statistically significant, indicating a good model fit [22]. The results of the fitting are presented in Tables 12 and 13 below.

Table 12. TheCAPM model regression indicators

Stock code	R-squared	Adj.R-squared	t	$P > \|t\|$	F-statistic	Prob(F-statistic)
600519.SH	0.402	0.402	3.137	0.002	1668	1.91e–279
600036.SH	0.410	0.410	2.042	0.041	1727	5.71e–287
000333.SZ	0.417	0.417	2.282	0.023	1778	1.85e–293
000858.SZ	0.453	0.453	2.469	0.014	2060	0

Table 13. The Fama-French three-factor model regression indicators

Stock code	R-squared	Adj.R-squared	t	$P > \|t\|$	F-statistic	Prob(F-statistic)
600519.SH	0.424	0.424	3.084	0.002	596	3.07e–290
600036.SH	0.331	0.330	1.083	0.279	400.4	2.81e–211
000333.SZ	0.353	0.352	0.965	0.335	440.7	1.58e–228
000858.SZ	0.470	0.469	2.188	0.029	716.7	0

From the figure, we can conclude that, compared to the Fama-French three-factor model, the R-squared (R^2) value and adjusted R-squared value of the Capital Asset Pricing Model (CAPM) are relatively larger, the p-value of the CAPM is relatively smaller, and the F-statistic of the CAPM is larger. Therefore, we conclude that the CAPM fits better than the Fama-French three-factor model for the 20 stocks analyzed in this paper. Although the Fama-French three-factor model incorporates more factors than the CAPM, the experimental results indicate that the CAPM is more applicable to long-term data analysis of large-capitalization stocks in the Chinese A-share market. In other words, investors should pay more attention to market risk and fundamentals when making long-term value investments, while the Fama-French three-factor model may be more suitable for short-term investments [23].

5. Conclusions

Based on the results presented in the experimental data, it can be concluded that using Markowitz's portfolio theory, it is possible to find ways of investing in portfolios consisting of multiple assets that minimize risk or maximize the Sharpe ratio, while also determining the efficient frontier. These results can help investors better manage risk and earn higher returns. Investors should also consider their own actual situation and risk tolerance as a basis for making rational investment decisions. Fitting the CAPM model using the least squares method, combining it with the Fama-French three-factor model, and comparing it with market indicators can provide valuable references for investors choosing individual securities or a portfolio of securities. The use of Python greatly simplifies the computation of the expected rate of return and the variance of the portfolio in the Markowitz's portfolio theory, and it can also find out the optimal portfolio quickly, which is very important for the application of Markowitz's portfolio theory in China's financial market. This is valuable for applying Markowitz's portfolio theory in the Chinese financial market. At the same time, it can also be used to compare the return and risk factors of a portfolio or a single security with the market, which is also crucial for the development of capital asset pricing models. The combination of traditional financial theories and emerging programming languages not only speeds up the solution of financial problems, but also improves the efficiency of the solution. This undoubtedly brings new opportunities for financial analysts as well as emerging technology talents, who should seize this opportunity to break through their own bottlenecks by utilizing the combined technology and contribute to the development of the financial industry. In addition, for emerging technology talents, they can also practice their own skills and play their own value in the field of financial analysis, which is also a good opportunity for learning and development.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Rupeng Song: Writing—Original Draft, Writing—review & editing, Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Visualization. Chingfei Luo: Writing—review & editing. Meilan Qiu: Supervision, Project administration, Funding acquisition, Writing—review & editing. All authors had approved the final version.

Funding

This work is supported by the Ministry of Education's "Blue Fire Program" (Huizhou) Industry-University-Research Joint Innovation Fund 2018 Annual Project: High-power Intelligent Dimming Control System(CXZJHZ201812), the Guangdong Philosophy and Social Sciences Planning Discipline Co-construction Project (GD23XSH27), Guangdong Provincial Department of Education Key Discipline Research Ability Improvement Project (2022ZDJS055), the Huizhou Philosophy and Social Sciences Project (2022ZX046, 2023ZX056) and Huizhou University Higher Education Teaching Quality and Teaching Reform Project ([2022]163).

References

[1] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, *7(1)*, 77–91. https://doi.org/10.2307/2975974

[2] Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19(3)*, 425–442.

[3] Fama, E. F., & French, R. K. (2000). Forecasting profitability and earnings. *The Journal of Business*, *73(2)*, 161–175. https://doi.org/10.1086/209656
[4] Sun, L. P. (2020). An empirical study of Markowitz's portfolio theory based on Python. *Times Finance, (25)*, 46–50.
[5] Zuo, L. (2021). Study on portfolio strategy based on mean-variance model. *Journal of Financial and Economic Theory*, *(05)*, 9–16.
[6] Song, Y. W., & Chen, L. H. (2022). An empirical study of the capital asset pricing model. *CO-Oerativeconomy and Science*, *(03)*, 71–73.
[7] Li, K. X. (2022). A study on the applicability of the capital asset pricing model in China's stock market. *Economic Outlook The Bohai Sea*, *(04)*, 157–160.
[8] Luo, X. L. (2016). An empirical study of stock returns in China's A-share market-based on the Fama-French three-factor model. *China Price Journal*, *12*, 34–37.
[9] Li, A. (2017). Mathematical planning in unsystematic risk portfolio investment. PhD thesis, North University of China.
[10] Yu, L. X., & Li, Y. (2013). Optimal portfolio analysis based on Markowitz theory. *Finance and Accounting Monthly*, *(22)*, 53–55.
[11] He, X. N. (2020). *Investment Principle and Applications* (4th ed.). China Machine.
[12] Banz, R. W. (1983). The relationship between return and market value of common stocks. *Journal of Financial Economics*, *9*, 3–18.
[13] Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management*, *21(1)*, 49–58.
[14] Lin, S., & Luo, Y. (2022). A study of fund managers' investment behavior from the perspective of social network analysis approach. *Modern Finance and Economics-Journal of Tianjin University of Finance and Economics*, *42(04)*, 79–95.
[15] Shao, X. Y., Han, C. L., & Qiao, Y. B. (2019). Effectiveness analysis of CAPM series models-a comparative analysis based on industry data of Shanghai, Shenzhen and Hong Kong stock markets. *Review of Investment Studies*, *38(10)*, 115–132.
[16] Tsang, T. S., & Chan, L. W. (2021). Data analysis of portfolio returns and volatility based on python technology. *Computer Knowledge and Technology*, *17(32)*, 29–31.
[17] Zhang, Y. M., & Cheng, X. J. (1992). On seeking minimum variance sets. *Journal of Shanghai Maritime University*, *(04)*, 36–40.
[18] Wu, X. Q. (2020). *Securities Investment* (5th ed.). China Renmin University.
[19] Gao, R., & Sha, Y. Z. (2019). Which is the best: A comparison of asset pricing factor models in Chinese mutual fund industry. *Economic Modelling*, *83(C)*, 8–16.
[20] Simaan, Y. (2024). *Markowitz and the CAPM. Annals of Operations Research*. Advance online publication.
[21] Hanauer, M. X., Maarten, J., Laurens, S., *et al.* (2024). Factor models for Chinese A-shares. *International Review of Financial Analysis*, *91*, 1–21.
[22] Feng, W. H., Zhang, X. F., & Chen, Y. S. (2023). Linear regression estimation using intraday high frequency data. *Aims Mathematics, 8(6)*, 13123–13133.
[23] Duan, Y. J., & Li, K. L. (2021). Research on the application of Fama and French three-factor and five-factor models in American industry. *Journal of Physics: Conference Series*, *1865(4)*, 1–5.