Analysis of Annual Return Rate Data of Enterprises Based on R Language

Qi Lin *
International School, Beijing University of Posts and Telecommunications, Beijing 100876, China
*Corresponding author e-mail: linwel@bupt.edu.cn

Abstract. As a simple and practical statistical tool, R language has become more and more widely used in the financial industry. This paper collects the monthly prices of 10 asset companies and uses R language as a tool for data analysis. The 10 asset companies are analyzed for the annual return rate and other indicators. This paper also draws on the MVP and t-copula investment strategies of securities investment to simulate the investment of these 10 asset companies.

1. Introduction
With the increasing popularity of internationalization, the economic interaction between the state and the country has become more and more popular, especially in the lives of ordinary people. Take Sino-US economic and trade as an example, companies such as NIKE, APPLE, AMAZONE have been closely related to the lives of ordinary Chinese people [1]. The specific embodiment is not only that ordinary people can experience and purchase the products of these merchants. In the stock market and financial bond investment, ordinary Chinese people can also invest in these companies. Then how to use the existing enterprise data to evaluate and analyze the investment status of the enterprise is still a more specialized problem. The R language provides an effective help as a convenient programming language.

At present, using R language as a professional and convenient data analysis tool, it has become one of the most commonly used ways to carry out and visualize data in the economic and financial industry. As the most direct embodiment of the economic and financial industry of the country's people's livelihood, the use of R language to analyze the bank's turnover and other data can have very good practical significance [2]. This paper proposed the use of R language to simulate and visualize the audit data of a bank. It mainly utilizes the R language to interact with the Hadoop platform and utilizes the MapReduce model for distributed analysis to meet the requirements of big data analysis requirements of orders of magnitude up to TB and even PB. At the same time, the paper also emphasizes the powerful features of the R language in the data analysis, and uses the common bar chart, line chart, scatter chart, pie chart, etc. to visually display the results of data analysis.

There are good interfaces between R language and other programming languages and databases [3]. For example, you can connect to Oracle, SQL and other databases through ROracle, RSQLServer, etc.; call Python, Java, C through r Python, rJava, Rcpp, etc. Languages such as C++ can also share data with statistics and data analysis software such as Minitab, S, SAS, SPSS, Stata, Systat, Weka, etc. using foreign and other packages [4].
Combining the previously read literature and my undergraduate professional e-commerce, I decided to rate the monthly prices of 10 large and medium-sized companies in five US industries and calculate their monthly returns. At the same time, combined with some financial statistical models in the economic field, such as Minimum Variance Portfolio (MVP), t-copular and other strategies into the investment evaluation of simulation. At the same time, this paper will also use the excellent visual characteristics of R language, using Q-Q diagrams to visualize the analysis results.

2. Data Prepressing

The data for this experiment was from Google Finance. The experiment used Google Finance to investigate the monthly prices of ten large and medium-sized companies in the United States: Abercrombie & Fitch, Amazon, Apple, AT&T, Bank of America, Hormel Foods, JPMorgan Chase, Nike, Tyson Foods, Verizon. I collected the monthly prices of the ten companies from January 2010 to December 2017 (the latest monthly prices were not available from Google Finance) and calculated the monthly prices of these companies from here. At the same time, I also obtained price information for the monthly risk-free rate (T-bill) from January 2010 to December 2017. In the process of selecting data, the ten companies or assets I selected were from five industries, and each industry had two assets. The corresponding situation is as follows:

The 10 assets and the industries they belong to are:
1. Technology: Amazon, Apple
2. Finance: Bank of America, JPMorgan Chase
3. Food: Hormel Foods, Tyson Foods
4. Communication: Verizon, AT&T
5. Clothing and Retail: Nike, Abercrombie & Fitch

After getting the initial data, I wrote the data into the .xlsx and .csv files. The reason is that there are packages in the R language that focus on the processing of files in these two formats. In order to facilitate the implementation in the code, I have written these 10 companies into the file names ANF.csv, AMZN.csv, AAPL.csv, T.csv, BAC.csv, HRL.csv, JPM.csv, NKE.csv, TSN.csv and VZ.csv files. Then I summarized the monthly reports of the 10 companies and assets into a file named data.xlsx, and I added a risk-free rate for the corresponding time in the data.xlsx file. Descriptive data for each asset and enterprise is shown in Table 1.

Table 1 The Enterprises asset descriptive data

| Company              | Mean  | Std. Dev. | Skewness | Kurtosis | Beta  |
|----------------------|-------|-----------|----------|----------|-------|
| Abercrombie & Fitch  | 0.00299 | 0.13633   | 0.3133   | 1.28234  | 1.47259 |
| Amazon               | 0.02677 | 0.07916   | 0.34822  | 0.08772  | 1.22217 |
| Apple                | 0.02174 | 0.07011   | -0.06159 | -0.28187 | 1.07861 |
| AT&T                 | 0.00548 | 0.04408   | -0.34562 | 0.27015  | 0.4015  |
| Bank America         | 0.01162 | 0.09652   | 0.0693   | 0.45918  | 1.5957  |
| Hormel Foods         | 0.01537 | 0.05222   | 0.08834  | 0.62871  | 0.54406 |
| JPMorgan Chase       | 0.01322 | 0.07075   | -0.52216 | 0.98175  | 1.46119 |
| Nike                 | 0.01599 | 0.0593    | -0.37612 | 0.89279  | 0.77353 |
| Tyson Foods          | 0.02148 | 0.07384   | -0.12776 | 1.07155  | 0.625   |
| Verizon              | 0.008   | 0.0479    | 0.20821  | -0.55847 | 0.45526 |
| T-bill               | 0.00017 | 0.00025   | 2.13898  | 3.61347  | *      |

Table 1 provides the first look at our data. It provides descriptive statistics for each asset and T bill. It turns out that Amazon’s average return is at most 0.02677, while Abercrombie & Fitch has an average return of at least 0.00299. In addition, the average returns of companies in the same industry, such as the financial industry (Bank of America and JPMorgan Chase) and the technology industry (Amazon and Apple), are generally close to each other.
The skewness coefficient of all assets is close to zero, indicating that 10 assets are usually symmetrical. The beta coefficient indicates whether an asset is larger or smaller than market fluctuations. According to the survey results, AT&T, Hormel Foods, Nike, Tyson Foods and Verizon all have beta values less than 1, indicating that they are less volatile than the market, so they are non-aggressive assets. On the other hand, the remaining five assets are radical assets. Interestingly, companies from the food and telecommunications industries are non-aggressive, while assets from the banking industry are aggressive. This is reasonable because we expect the banking industry to react more to sudden changes in the market.

3. Method

3.1. MVP algorithm

MVP is the abbreviation for the minimum variance portfolio [5]. The MVP investment algorithm is suitable for risk-averse investors, that is, the risk factor is the first consideration, and the impact of risk on investment is avoided as much as possible. In this kind of investment algorithm, the return of the securities is measured by the income \( E \), and the investment risk is expressed by the variance of the income. The total return of the portfolio is expressed as a weighted average of the expected returns of each asset. The risk of the combined asset is expressed as the variance or standard deviation of the return, then the MVP can be expressed as:

\[
\min \delta^2(r_p) = \sum \sum w_i w_j \text{cov}(r_i, r_j) \tag{1}
\]

\[
E(r_p) = \sum w_i r_i \tag{2}
\]

And:
- \( r_p \) ---- Combined Income
- \( r_i, r_j \) ---- The proceeds of the i-th and j-th assets;
- \( w_i, w_j \) ---- The weight of asset i and asset j in the combination
- \( \delta^2(r_p) \) ---- The variance of the combined income is the overall risk of the combination
- \( \text{cov}(r, r_j) \) ---- Variance between two assets

There is a package for MVP calculations in the R language, which requires only a simple call. The key is to use the R language to retrieve the data of each asset from the data.xlsx file. The algorithm calculates the global MVP, the short MVP and the non-short MVP respectively.

3.2. Copula algorithm

The Copula function describes the correlation between variables [6]. It is actually a function that connects the joint distribution functions with their respective edge distribution functions. Therefore, some people call it a connection function. There are two reasons for using Copula. The first is that Copula is a method for studying the measure of dependence. The second is Copula, which is the starting point for constructing a two-dimensional distribution family, which can be used for multivariate model distribution and stochastic simulation. The Copula function, as a tool for the interdependence mechanism between variables, contains almost all the dependent information of random variables. In the case that it is impossible to determine whether the traditional linear correlation coefficient can correctly measure the correlation between variables, the Copula function is for the variable. Analysis of correlations is useful.

In portfolio investments, the Copula function can be used to assess the investment risk of each asset. The investment risk can be divided into two parts. The first part is the credit risk of each asset itself, and the other part is the risk caused by the related structure between the assets. This requires a function to solve the relationship between a single edge distribution and how far away, Copula can solve such problems.

Similar to MVP, the Copula function is mainly used to evaluate risk in this experiment. Its main algorithm structure has been encapsulated in the R language, mainly for the call of data.
4. Results

Figure 1 shows the monthly price (blue line) and return (red line, percentage) for each asset. In general, for most assets, their monthly prices have increased since 2010. However, we have seen Abercrombie & Fitch’s price drop sharply around 2012. This makes sense, because in 2012, their CEO Mike Jeffries commented that they would exclude those who did not fit the brand, which turned out to be a terrible marketing mistake. In addition, the monthly return on most assets fluctuates between 10% and 20%.

Figure 2 shows the equity curve for each asset (brown line), and the equity curve for the S&P 500 index is covered at the top. The equity curve shows that the growth of each asset from 2010 to 2017 is $1. Overall, for most assets, the stock curve is the same as the pattern we see in the monthly price curve. However, due to Abercrombie & Fitch’s scandal, I tend to see the A&F curve drop. The stock curve of the S&P 500 Index has continued to grow over the past eight years. We note that the stock curves of Apple, Amazon, Hollande and Nike are both higher than the S&P 500’s stock curve, indicating its outstanding performance in the market. On the other hand, Bank of America and JPMorgan Chase have performed poorly compared to companies in other industries.
Figure 3-4 shows the returned box plot and histogram. With the exception of Apple and Verizon, almost all assets have outliers, indicating that their performance is relatively stable. The average returns of the 10 assets are close to each other. In addition, the histogram shows that the distribution of most assets is symmetrical and bell-shaped, so assets may be normally distributed. This can be demonstrated by looking at Figure 5.

Figure 5 shows the normal Q-Q plot for 10 assets. We can see that for each asset, except for a few outliers, most of the points are on the hypothesis line.
In addition, this article installs different assignments for each asset, and the final results are as follows. Alternating distributions are normal distribution, skewed normal distribution, generalized error distribution, skewed generalized error distribution, t-distribution and skewed t-distribution. We choose the model according to the AIC rule. From Table 2, we note that almost all assets have a potential normal distribution, with the exception of Hormel Foods, which has a generalized error distribution. However, since the generalized error distribution family includes all normal distributions, we believe that the results are consistent with our previous findings, so we can say that the normal hypothesis is useful for our further research. We also run Ljung-Box test smoothness and the results prove that the return is generally static.

| Company           | Distribution      | Parameters (mean, sd) |
|-------------------|-------------------|-----------------------|
| Abercrombie & Fitch | Normal            | (0.003, 0.136)        |
| Amazon            | Normal            | (0.027, 0.078)        |
| Apple             | Normal            | (0.022, 0.069)        |
| AT&T              | Normal            | (0.005, 0.044)        |
| Bank America      | Normal            | (0.012, 0.096)        |
| Hormel Foods      | Generalized Error | (0.016, 0.052, nu=1.22) |
| JPMorgan Chase    | Normal            | (0.013, 0.07)         |
| Nike              | Normal            | (0.015, 0.059)        |
| Tyson Foods       | Normal            | (0.023, 0.075)        |
| Verizon           | Normal            | (0.008, 0.048)        |

Figure 6 depicts a paired scatter plot between 10 assets, where the correlation coefficients are also listed. Based on the results, we found a strong correlation between the returns of Verizon and AT&T. This makes sense because they come from the same industry, but beyond that, we don't see a very clear relationship between other assets.
5. Conclusion

Combined with the above data analysis of this experiment, two investment strategies for MVP and Copular can be derived.

5.1. Minimum Variance Portfolio (MVP)

According to Tables 3 and 4, the average annual return of MVP (no shorts) is 0.1536, which is lower than the average annual return of Amazon, Apple, Hormel, JPMorgan Chase, Nike and Tyson. The annual standard deviation of MVP (no shorts) is 0.1063, which is less than the annual standard deviation of all 10 assets. ANF weighs 0, and the top three weights of this portfolio come from AT&T, Hormel Foods and Verizon.

MVP (shorts)

According to Tables 3 and 4, the average annual return for MVP (short) is 0.1644, which is lower than the average annual return of Amazon, Apple, Hormel, Nike and Tyson. The annual standard deviation of MVP (no shorts) is 0.1043, which is less than the annual standard deviation of all 10 assets. The weight of ANF is negative, and the top three weights in this portfolio come from AT&T, Hormel Foods and Apple.

MVP (no shorts) vs MVP (shorts)

The average annual return of MVP (short) is greater than the average return of MVP (no short), and the standard deviation of MVP (short) is less than the standard deviation of MVP (no short). In MVP (short), Abercrombie & Fitch Co (ANF) has a negative weight, but MVP has a weight of 0 (no short). Except for the weight of the ANF, the weights of the remaining nine assets are not much different between MVP (short) and MVP (no short).

Table 3. Comparison of shorts and no shorts.

|                  | MVP(no shorts) | MVP (shorts) |
|------------------|----------------|--------------|
| Monthly Mean Return | 0.0128         | 0.0137       |
| Monthly SD       | 0.0307         | 0.0301       |
| Annual Return    | 0.1536         | 0.1644       |
| Annual SD        | 0.1063         | 0.1043       |
Table 4. The comparison of Annual Mean Return and Annual SD among ten assets.

|          | ANF | AMZN | AAPL | T | BAC | HRL | JPM | NKE | TSN | VZ |
|----------|-----|------|------|---|-----|-----|-----|-----|-----|----|
| Annual Mean Return | 0.0359 | 0.3212 | 0.2608 | 0.0657 | 0.1395 | 0.1845 | 0.1586 | 0.1906 | 0.2578 | 0.0960 |
| Annual SD | 0.4723 | 0.2742 | 0.2429 | 0.1527 | 0.3344 | 0.1809 | 0.2451 | 0.2054 | 0.2558 | 0.1659 |

The effective boundaries (shown in Figure 7) show a set of optimal portfolios that give the highest return on risk or the lowest risk for a given return.

5.2. **Copulas**

This paper attempts to fit various copulas to the data. The results are shown in Table 5 below.

Table 5. The Copula's fitting data

| Company | T   | Normal | Frank | Clayton | Gumbel | Joe  |
|---------|-----|--------|-------|---------|--------|------|
| Loglik  | 52.01 | 49.04  | 44.32 | 40.87   | 41.69  | 31.96 |
| P       | 2    | 1      | 1     | 1       | 1      | 1    |
| AIC     | -100.11 | -96.07 | -86.64 | -79.74  | -81.39 | -61.91 |

Therefore, t-copula gives the smallest AIC, so this paper chooses t-copula to fit the data. The first two of this article are examples. In Fig. 8, the red line indicates the empirical distribution, and the black line indicates the t-copula contour map. T-copula fits the data well and this paper examines the relationship between these assets and t-copula.
Based on the above results, for most assets, the monthly price has increased since 2010, and the average monthly return rate is about 1.5%, most of which is normally allocated. AT&T, Hormel Foods, Nike, Tyson Foods and Verizon's Betas are less than 1, indicating that they are non-aggressive assets and the remaining assets are positive assets.

If the ultimate goal is to minimize risk, MVP (no shorts) produces the best results with an annual risk of 0.1043 and an annual return of 0.1644. AT&T, Hormel food and apples account for a large proportion of weight. If the ultimate goal is to get the maximum return for each risk, then the tangent portfolio (short) is the best, resulting in a 2.2% return. In addition, if I get a $100,000 investment, MVP can effectively reduce VaR and ES, and a tangible portfolio with no shorts will also reduce VaR. Abercrombie & Fitch Co. has the largest VaR and expected deficit. I use t-copula to fit the data because it provides the lowest AIC.

Reference
[1] Zafar, A. (2007). The growing relationship between China and Sub-Saharan Africa: Macroeconomic, trade, investment, and aid links. The World Bank Research Observer, 22(1), 103-130.
[2] Wu Xinxin. Research on Financial Big Data Auditing Application Based on R Language [D]. Nanjing Audit University, 2017.
[3] Pebesma, E., Nüst, D., & Bivand, R. (2012). The R software environment in reproducible geoscientific research. Eos, Transactions American Geophysical Union, 93(16), 163-163.
[4] Ferraro, M. B., & Giordani, P. (2015). A toolbox for fuzzy clustering using the R programming language. Fuzzy Sets and Systems, 279, 1-16.
[5] Scherer, B. (2011). A note on the returns from minimum variance investing. Journal of Empirical Finance, 18(4), 652-660.
[6] Li, D. X. (2000). On default correlation: A copula function approach. The Journal of Fixed Income, 9(4), 43-54.