Asset allocation in representative U.S. stocks

Jinglan Deng[1, †, *], Siyang Lyu[2, †, *]

[1]School of Accountancy Guangdong University of Foreign Studies Guangzhou, China

[2]School of Economics Wuhan University of Technology Wuhan, China

*Corresponding author: 20192005094@gdufs.edu.cn, 298607@whut.edu.cn

†These authors contributed equally.

Keywords: ARMA; Portfolio theory; Monte Carlo simulation

Abstract: This study selected six stocks of companies to help investors diversify their investment. In this paper, we use the ARMA model to predict the returns of the assets. Then modern portfolio theory is adopted to discovery the Maximum Sharpe Ratio Portfolio and the Minimum Volatility Portfolio. The results show that the ARMA model can be surely used to forecast the future return of the asset. Besides, through Monte Carlo simulation, we find that the asset of MSFT accounts for the largest proportion in the two interested portfolios. Finally, based on the asset's weights, we compare the constructed portfolio with the actual market return, and the results show that our portfolios beat the market return and can bring certain financial benefits for investors. To sum up, the results benefit the related investors in financial markets.

1. Introduction

Based on the principle of risk diversification, in securities investment, individuals commonly invest in different securities to construct investment portfolios to avoid risks and obtain more returns. Consequently, studies on portfolio are enduring [1]. Since Markowitz [2] proposed the securities portfolio theory in 1952, most scholars at home and abroad have conducted in-depth research on portfolio management. For example, Sun [3] combined traditional financial theories with emerging programming languages to investigate the calculation of the expected rate of return and portfolio variance in Markowitz's portfolio theory. In addition, Li [4] proposed that the Markowitz theory fully considers the coexistence of risks and returns, etc. In principle, Mean-risk models such as mean-absolute deviation (MAD) model, mean-variance (MV) model and mean CVaR model can be applied to financial instruments such as stocks, bonds, currencies and derivatives as long as the return and risk of assets can be estimated [5-8]. These models are being used in the first step of the asset allocation procedure to determine an optimal proportion of the fund to be allocated to various assets [9-10]. According to the above-mentioned theories, relevant scholars further made some investigations of portfolio based on asset weights. Shahidin [11] forecasted share price by using Geometric Brownian Motion and hence used the Variance-Co-variance to calculate Value at Risk of each stock and stock portfolio. According to the value at risk of stock portfolios of five different industrial products. Their findings indicated that stock portfolios from industrial product trading and service industries are most suitable for risk aversion and risk premium investors. Hoque [12] compared the performance of the Islamic stock portfolio (ISP) and conventional stock portfolio (CSP) of the five industrial sectors and market in Malaysia, the authors found that the risk-sharing ISP is superior to the risk-bearing CSP for better returns at the sector as well as the market level. Khalfaoui [13] used loss functions on the daily Value-at-Risk estimates of a diversified portfolio in three stock indices and proposed a wavelet-based approach which decomposes a given time series on different time horizons. Their findings pointed out that wavelet-based models increase predictive performance of financial forecasting in low scales according to number of violations and failure probabilities for VaR models. In 2012, VaR Model based on J. P. Morgan's RiskMetrics has problem that actual loss exceeded VaR under unstable economic conditions. Therefore, Park [14] proposed a One-factor VaR Model and found that One-factor VaR

Model can solve the problem of the actual loss exceeded VaR. Artini [15] compared the performance of small and medium enterprises (SME) and manufacturing company stock portfolios in the Indonesian, Chinese and Indian capital markets by the Sharpe Index and the significance of differences in average performance in the capital market. Their findings indicated that SME and manufacturing company stock portfolios had relatively better performances in China and India. To sum up, previous literature focused mainly on the performance of the risk of the portfolio. Limited research focused on the return of portfolio. However, as one of the most key characteristics of asset, return should also be focused more. Therefore, this paper explored the return of the portfolio.

By studying the U.S. stock market, the S&P 500 Index and Nasdaq Index are important components of U.S. Capital Markets which reflect the general development of the U.S. economy. Its constituent stocks are sought after by many investors, while other ordinary stocks are neglected. To explore the relationship between stock values and stock types, we chose representative National Association of Securities Dealers Automated Quotations (NASDAQ) component stocks, i.e., Microsoft Corporation (MSFT), NVIDIA Corporation (NVDA), Napco Security Technologies, Inc. (NSSC); and representative S&P Composite 1500 Consumer Discretionary component stocks, i.e., The New York Times Company (NYT), Foot Locker, Inc. (FL). To ensure the assets to be more diversified, Basic Sanitation Company of the State of Sao Paulo-SABESP (SBS) from the New York Stock Exchange (NYSE) Composite component is also selected. And using portfolio theory and Monte Carlo simulation as the theoretical basis, we attempt to build out a value investing strategy in technology and consumer services stock versions to avoid single industry risks. Finally, we got the efficient frontier and found the Maximum Sharpe Ratio Portfolio and the Minimum Volatility Portfolio. The results in this paper can be generalized as follows. First, the ADF test results indicated that the asset historical return series is a steady time series; Second, ARMA model can be used for time series prediction in portfolio; Third, through the Maximum Sharpe Ratio Portfolio and the Minimum Volatility Portfolio, the authors pointed out that MSFT has the largest weight, more attention should be paid to MSFT in the portfolio; Fourth, the constructed portfolio based on the weight we obtained above beat the market. This paper is constructed as follows. Section 2 is the data and methods. Section 3 shows the results and Section 4 concludes the paper.

2. Data and Methods

2.1 Data

We extracted sample data of the U.S. listed companies from yahoo finance (https://finance.yahoo.com/), the sample period is from Jan 1, 2010, to Dec 31, 2020. Additionally, the specific data we select is the daily adjusted closing price, which can give an accurate representation of the firm's equity value beyond the simple market price. To explore the relationship between stock values and stock types, we chose the NASDAQ component stocks, i.e., MSFT, NVDA, NSSC; S&P Composite 1500 Consumer Discretionary component stocks: NYT, FL and NYSE Composite component stocks SBS. And we dropped the null value and performed difference of first-order on the original data to obtain the compounded return of a total of 16608 samples. The descriptive statistics are shown in Table I.

Table 1 Descriptive Statistics of return series

	FL	**MSFT Deviation**	**NSSC**	**NVDA**	**NYT**	**SBS**
Mean	0.0006	0.0008	0.0010	0.0012	0.0005	0.0002
Std	0.0248	0.0160	0.0290	0.0266	0.0228	0.0275
Min	-0.3275	-0.1595	-0.2516	-0.2077	-0.2481	-0.2888
25%	-0.0102	-0.0068	-0.0119	-0.0113	-0.0109	-0.0144
50%	0.0008	0.0007	0.0000	0.0014	0.0006	0.0006
75%	0.0118	0.0087	0.0137	0.0141	0.0118	0.0151
Max	0.2481	0.1329	0.1836	0.2609	0.1197	0.1771

Table I shows that the returns on mean of the six assets are positive, which are 0.0006, 0.0008, 0.0010, 0.0012, 0.0005, 0.0002, respectively. Generally, the return series is stable without abnormal value. NVDA has the largest rally among the six assets, for about 0.2481. And FL has the largest drop among the six assets, for about -0.3275.

2.2 ADF test

Time series need to satisfy the stationary requirement when modeling time series using arma models.And the unit root test is usually used to check the stationary of the time series. Among several methods to certify the existence of unit root, the ADF test, proposed by Said [16] has been widely used because of its advantages. For example, Huo [17] believed that under the circumstance of sufficient sample size, traditional unit root tests such as the ADF test, has good test efficacy. Additionally, WU and DENG [18] found that the traditional unit root test seem to perform better than the quantile unit root test. Thus, in this paper, the ADF test is adopted to test the stationary of the historical stock return series. The specific formulas of ADF test are as follows.

$$\Delta yt = \gamma yt - 1 + \sum_{t=1}^{p} \beta t \Delta yt - 1 + \mu\, t \tag{1}$$

$$\Delta yt = \gamma yt - 1 + \alpha + \sum_{t=1}^{p} \beta t \Delta yt - 1 + \mu t \tag{2}$$

$$\Delta yt = \gamma yt - 1 + \alpha + \delta t + \sum_{t=1}^{p} \beta t \Delta yt - 1 + \mu t \tag{3}$$

Where,α is constant term;δt is time trend term;μt is random disturbance term, we estimated the regression coefficients $\hat{\rho}$ and the standard deviation $\hat{\sigma}$ by using the traditional OLS estimation and constructed the statistic t.The null hypothesis is γ equals to 0.

2.3 ARMA model

ARMA modeling is a kind of quantitative analysis methods to study the laws of development and changes, which reveals the correlation structure and dynamics of series by analyzing their correlations at different moments [19]. The ARMA model takes interference of random fluctuations into account. By using the price historical data to explain stock price time changing pattern, it can make short-term forecasting of future stock prices. Thus, in the financial industry, ARMA model plays a significant role in explaining and forecasting the time series such as stock price [20]. The ARMA model is a combination of autoregressive model (AR) and moving average model (MA).

Among them, the autoregressive model is predicted by a linear combination of past observations and present disturbance values. The mathematical equation of the AR model is,

$$Yt = \varphi 1Yt - 1 + \varphi 2Yt - 2 + \ldots + \varphi pYt - p + et \tag{4}$$

where$p$is the order of the model, $\varphi i(i = 1,2,\ldots,p)$ is the coefficient to be determined, and $et$is the error, and Yt is the smooth time series.

The moving average model is predicted by a linear combination of past and present white noise disturbance values. The mathematical equation of the MA model is,

$$Yt = et - \theta 1et - 1 - \theta 2et - 2 - \ldots - \theta qet - q \tag{5}$$

where$q$is the order of the model,$\theta i(i = 1,2,\ldots,p)$ is the coefficient to be determined, and et is the error, and Yt is the smooth time series.

Therefore, the expression of the autoregressive moving average model is,

$$\begin{gathered} Yt = \varphi 1Yt - 1 + \varphi 2Yt - 2 + \ldots + \varphi pYt - p + \\ et - \theta 1et - 1 - \theta 2et - 2 - \ldots - \theta qet - q \end{gathered} \tag{6}$$

And all the three models are all follow the following default conditions,

$$\begin{gathered} E(e_t) = 0, Var(e_t) = \sigma_e^2, E(e_t e_s) = 0, s \neq t \\ Ex_s e_t = 0, \forall s < t \end{gathered} \tag{7}$$

And the linear function of Y_t with respect to $Y_{t-1}, Y_{t-2}...$ is,

$$Y_t = \sum_{i=0}^{\infty} h_i \left(\sum_{j=0}^{\infty} \pi_j Y_{t-i-j} \right) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} h_i \pi_j Y_{t-i-j} \quad (8)$$

Simplify as,

$$Y_t = \sum_{i=0}^{\infty} C_i Y_{t-1-i} \quad (9)$$

And for $\forall l \geq 1$,the linear function of Y_{t+l} with respect to $Y_{t+l-1}, Y_{t+l-2}...$ is,

$$Y_{t+l} = \sum_{i=0}^{\infty} C_i Y_{t+l-1-i} \quad (10)$$

According to the additivity of linear functions,we can figure out the linear function of Y_{t+l} with respect to $Y_t, Y_{t-1}...$ .And the estimation function is ,

$$\hat{Y}_t(l) = \sum_{i=0}^{\infty} \hat{D}_i Y_{t-i} \quad (11)$$

And the $\hat{Y}_t(l)$ is the predicted value of step l of the time series.

2.4 D-W test

For first-order correlation test, DW test is a classical method to test first-order autocorrelation of series especially for time series, which is proposed by Durbin and Waston [21]. When the sample meets the preconditions, i.e., the sample size is large enough and the random disturbance terms obey normal distribution, the DW test results are excellent to explain autocorrelation [22]. Therefore, to test whether the time series residuals are autocorrelated, DW test is adopted. The process of DW test is shown as follows.

DW statistic,

$$DW = \frac{\sum_{t=2}^{n}(et - et-1)^2}{\sum_{t=2}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 \ -2\sum_{t=2}^{n} etet-1}{\sum_{t=2}^{n} e_t^2} \quad (12)$$

$\sum_{t=2}^{n} e_t^2$ and $\sum_{t=2}^{n} e_{t-1}^2$ can be considered approximately equal when the sample size is large enough, so DW statistic,

$$DW \approx 2\left[1 - \frac{\sum_{t=2}^{n} etet-1}{\sum_{t=2}^{n} e_t^2}\right] \quad (13)$$

Autocorrelation coefficient of error series$\mu 1, \mu 2, ... \mu n$,

$$\varphi = \frac{\sum_{t=2}^{n} \mu t \mu t-1}{\sqrt{\sum_{t=2}^{n} \mu_t^2}\sqrt{\sum_{t=2}^{n} \mu_{t-1}^2}} \quad (14)$$

The estimated autocorrelation coefficient of the error series $\mu 1, \mu 2, ... \mu n$ of the estimate value et,

$$\hat{\varphi} = \frac{\sum_{t=2}^{n} etet-1}{\sqrt{\sum_{t=2}^{n} e_t^2}\sqrt{\sum_{t=2}^{n} e_{t-1}^2}} \quad (15)$$

From the equation (8),

$$\hat{\varphi} \approx \frac{\sum_{t=2}^{n} etet-1}{\sum_{t=2}^{n} e_t^2} \quad (16)$$

So, equation (8) is rewritten,

$$DW \approx 2(1-\hat{\varphi}) \tag{17}$$

Because φ is the autocorrelation coefficient of error μt and $\mu t-1$ and the range of values of φ is -1 to 1, the value range of DW is between 0 and 4. When the DW value is close to 2, φ tends to 0, which means that μt, $\mu t-1$ are not correlated.

2.5 Portfolio Return

Where n is the number of the assets,wi is the weight of the i-th asset in the portfolio,Ri is the return of the i-th asset.

$$Rp = \sum_{i=1}^{n} wiRi \tag{18}$$

And the variance of the portfolio is,

$$D(Rp) = \sum_{i=1}^{n}\sum_{j=1}^{n} wiwjCov(Ri, Rj) \tag{19}$$

Where n is the number of the assets,wi, wj is the weight of the i-th and j-th asset in the portfolio,Ri,Rj is the return of the i-th and j-th asset.

2.6 Ljung-box Test

Ljung-box test is a common method used to test the autocorrelation of time series, which is adapted by Ljung and Box [23] on the basic of Portmanteau statistic. And Ljung-box test is widely used in the application of Econometrics and Time Series [24],

The statistic Q is,

$$Q(m) = T(T+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{T-k} \tag{20}$$

Where T is the sample size,$\hat{\rho}(k)$ is the k -th order autocorrelation coefficient of the sample, m is the number of autocorrelation lags included in the statistic. The $Q(m)$ statistic is asymptotically Chi-Square distributed under the null hypothesis. The null hypothesis is,$\rho 1 = \rho 2 = \ldots = \rho m = 0$ and series are uncorrelated at m-order lags. The p-values above 0.05 indicate the acceptance of the null hypothesis of model accuracy under 95% significant levels.

3. Results

We did the ADF test, ARMA forecasting, DW correlation test and Ljung-box Test. The results are shown in Table II, Table III, Table IV and Table V, respectively.

3.1 ADF test

Table 2 ADF test results

Asset historical return	t-statistic
FL	-19.6095***
MSFT	-19.1705***
NSSC	-42.2815***
NVDA	-16.3316***
NYT	-22.9461***
SBS	-16.0105***

Note: *** represents significance at 1% level.

The above results shows that the significance of the t-statistic is less than 1%, which can indicate that the asset historical return series is a steady time series.

3.2 ARMA forecasting

Table 3 ARMA forecast results

Date	Real return	forecast	lower_ci_95	lower_ci_99	upper_ci_95	upper_ci_99
2020-12-30	-0.0047	-0.0032	-0.0564	-0.0731	0.0500	0.0668
2020-12-31	0.0012	0.0014	-0.0519	-0.0686	0.0546	0.0714
2021-01-04	-0.0562	-0.0045	-0.0578	-0.0745	0.0488	0.0655
2021-01-05	0.0000	-0.0043	-0.0490	-0.0657	0.0576	0.0743
2021-01-06	-0.0137	0.0037	-0.0496	-0.0663	0.0570	0.0738

Note: The ARMA results shown here is for 5 days ahead forecasting of the asset of SBS

Based on the assumption of the ARMA model,we model the ARMA using the stationary time series as above. From Table III, the forecasted return of SBS is close to the real return. And the SBS is the trade code of Basic Sanitation Company of the State of Sao Paulo-SABESP (SBS) from the New York Stock Exchange (NYSE) Composite component. The rise and fall of SBS of the predicted value in the first three days are consistent with the real return. Also, all forecasted values are within the 95% confidence interval.

3.3 D-W correlation test

Table 4 D-W test

Asset model residual	D-W value(d)
FL	2.0128
MSFT	2.0175
NSSC	1.9904
NVDA	2.0142
NYT	1.9931
SBS	1.9960

Table IV shows that all D-W value(d) are close to 2 or equal to 2. Therefore, it indicates that the ARMA model can well forecast financial asset.

3.4 Ljung-box Test

Table 5 Ljung-box Test

Asset model residual	p-value (lag 5)	p-value (lag 10)	p-value (lag 12)
FL	0.9999	0.9991	0.9992
MSFT	0.9985	0.9842	0.9894
NSSC	0.8642	0.9479	0.8709
NVDA	0.4912	0.2307	0.2085
NYT	0.9999	0.9999	0.9999
SBS	0.9999	0.9999	0.9999

Table V shows that almost all p-value from lag order 5 to 12 are close to 1. Only an asset has relatively lower p-value. Thus, this further suggests that the ARMA model can well forecast financial assets. Next, based on the above research methods, we implement the Monte Carlo simulation, and got the efficient frontier and two certain portfolios, i.e., the Minimum Volatility Portfolio and the Maximize Sharpe Ratio Portfolio, in the following Figure 1.

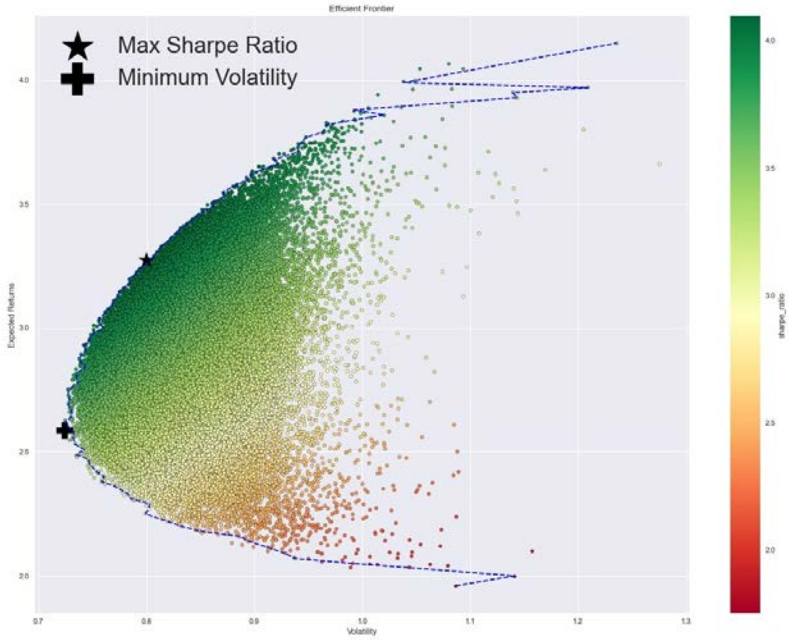

Fig 1 Efficient frontier and the two interested portfolios

The weight for each asset is shown in the following Table VI.

Table 6 Portfolio Optimization

Maximum Sharpe Ratio portfolio						
Asset	***FL***	***MSFT***	***NSSC***	***NVDA***	***NYT***	***SBS***
Weights	13.12%	36.69%	24.08%	22.02%	4.06%	0.02%
Returns	327.59%	***Volatility***	79.95%	***Sharpe_ratio***		409.72%
Minimum Volatility portfolio						
Asset	***FL***	***MSFT***	***NSSC***	***NVDA***	***NYT***	***SBS***
Weights	16.75%	45.75%	11.78%	0.80%	14.55%	10.36%
Returns	258.55%	***Volatility***	72.39%	***Sharpe_ratio***		357.14%

The Sharpe Ratio represents the returns that investors can obtain under a given volatility. The larger the Sharpe ratio, the higher the returns obtained by the given volatility of the portfolio. At the same time, the weight of each asset in the Maximum Sharpe Ratio portfolio can also help investors to choose stocks. Besides, the Minimum Volatility portfolio represents the portfolio composed of multiple assets with the lowest risk to reach the expected return. The above results indicate that the weights of FL, MSFT, NSSC, NVDA, NYT and SBS in the Maximum Sharpe Ratio portfolio are 13.12%, 36.69%, 24.08%, 22.02%, 4.06%, 0.02% respectively. For the Minimum Volatility portfolio, the weights of the six assets are 16.75%, 45.75%, 11.78%, 0.80%, 14.55%, 10.36% separately. It obvious that MSFT has the largest weight for both different portfolios. Therefore, more attention should be paid to MSFT in the portfolio.

3.5 Further discussion

The above portfolio optimization results are obtained from certain six asset. Thus, it is puzzled whether the constructed portfolios beat the market. In this section, we do further explorations to compare the portfolio. Specifically, we collect daily returns for the selected assets and the return of market index for the same time from 2010 to 2020. The results for the comparison are shown as follows in Figure 2. Obviously, the constructed portfolio based on the weight we obtained above beat the market, which means that the investors can surely benefit from our portfolio.

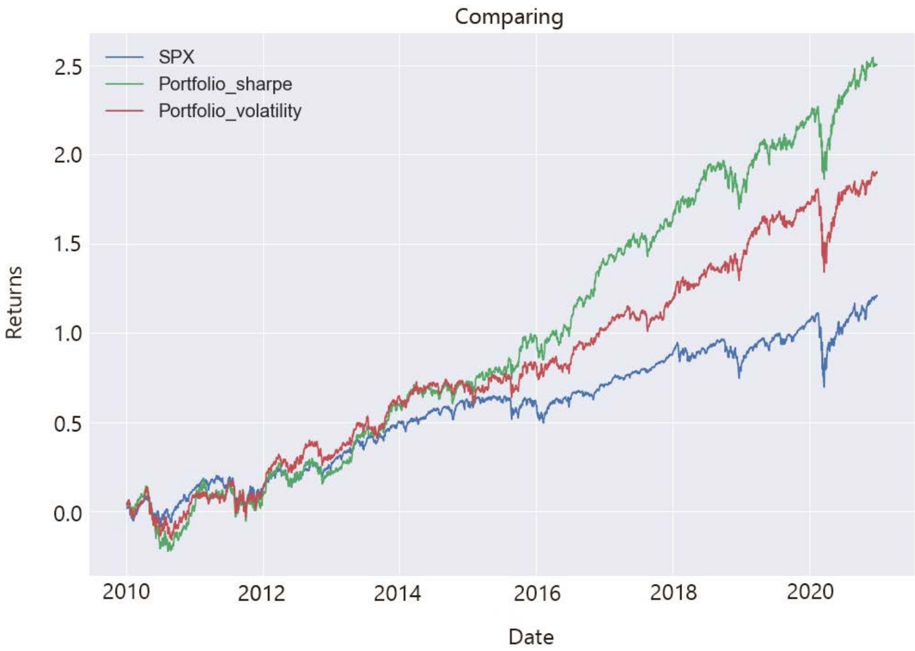

Fig 2 Comparison of the constructed portfolios and the market index return

4. Conclusion

This paper constructs portfolios based on six representative assets from the U.S. stock market. We forecast the asset by ARMA model, and the results show that ARMA can be used more in asset return forecasting. Furthermore, the results from portfolio construction show that MSFT may be an interested company for mean-variance investors who adopt the minimum variance portfolio or the maximum Sharpe Ratio portfolio. Besides, it is identified that our portfolio with certain asset weight can surely beat the market index return and bring economic benefits for the related investors.

However, deficiencies also exist. For example, we only use individual metrics like Sharpe Ratio or volatility to perform the asset optimization, adopting other indicators to make in-depth investigation may be deserved.

References

[1] Ling, J. , Huang, J. Y. , Xie, X. S. , Yang, C. J. , Tan, Y. X. , & Management, S. O. . (2014). A study of portfolio selection for investment in china's stock market based on the arma model and the mean-variance model. Journal of Wuyi University(Natural Science Edition), 26(3), 419-436.

[2] Markowitz, H. M. . (1952). Portfolio selection. The Journal of Finance, 7(1), 77.

[3] Sun, L. B.. (2020). Markowitz's portfolio theory: An empirical study based on Python. Times Finance (25), 46-47+50.

[4] Li, Y. & Yu, L.X.. (2013). Based on analysis of optimal portfolio Markowitz Theory. Finance and Accounting Monthly (22),53-55.

[5] Konno, H. , & Yamazaki, H. . (1991). Mean-absolute deviation portfolio optimization model and its applications to tokyo stock market. Management Science, 37(5), 519-531.

[6] Pardalos, P. M. . (1997). Optimization Techniques for Portfolio Selection. Physica-Verlag HD.

[7] Luenberger, & DavidG. (1973). Introduction to linear and nonlinear programming. Addison-Wesley Pub. Co.

[8] Kato, K. , & Konno, H. . (2007). Studies on a general stock-bond integrated portfolio optimization model. Computational Management ence, 4(1), 41-57.

[9] Arnott, R. D. , Fabozzi, F. J. , & Probus. Asset allocation: a handbook of portfolio policies, strategies and tactics.

[10] Brinson, G. P. , Randolph, H. L. , & Beebower, G. L. . (1995). Determinants of portfolio performance. Financial Analysts Journal, 51(1), 133-138.

[11] Shahidin, A. M. , Othman, S. S. A. & Razali, N. S. M.. (2021). Stock portfolio selection based on investors' risk preference. Journal of Physics: Conference Series, 1988(1).

[12] Hoque, A. , Rakhi, S. , Hassan, K. , & Le, T. . (2021). The performance of stock portfolios: evidence from analysing malaysia case, and implication for open innovation. Social Science Electronic Publishing.

[13] Khalfaoui, R. , & Boutahar, M. . (2012). Portfolio risk evaluation: an approach based on dynamic conditional correlations models and wavelet multiresolution analysis. MPRA Paper.

[14] Park, K. , Ko, K. , & Beak, J. . (2013). An one-factor var model for stock portfolio. Communications of the Korean Statistical Society, 26(3), 471-481.

[15] Artini, L. , & Ni, L. . (2020). Comparison of sme stock portfolio and manufacturing company performance in the indonesian, indian and chinese stock markets. Journal of Economic and Administrative Sciences, ahead-of-print(ahead-of-print).

[16] Said, S. E. , & Dickey, D. A. . (1984). Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika(3), 599-607.

[17] Huo, J. X. . (2011) Random Codfficient Autoregressive Unit Root Tests and Their Application. Tianjin University of Finance and Economics.

[18] Liang, W. U. , & Deng, M. . (2014). The finite sample performance of quantile unit root test and its application. Statistics & Information Forum.

[19] Wang, L. . (2004). Analysis of non-steady time-series forecast for economy based on arma model. Journal of Wuhan University of Technology.

[20] Meucci, A. . (2009). Review of discrete and continuous processes in finance: theory and applications. Social Science Electronic Publishing.

[21] Watson, J. . (1951). Testing for serial correlation in least squares regression. i. Biometrika, 38(1), 159-178..

[22] Liu, M. & Wang, Y.Y.. (2014). Exploring in durbin-watson autocorrelation test. The Journal of Quantitative & Technical Economics.

[23] GRETA, M., LJUNG, GEORGE, E., & P., et al. (1979). The likelihood function of stationary autoregressive-moving average models. Biometrika (https://academic.oup.com/biomet), 66(2), 265-270.

[24] Tsay, R. S. (2009). Analysis of Financial Time Series. China Machine Press (http://www.cmpbook.com).