The Investment Strategy Optimization based on BL Stock Price Selection based on Arima and Time Series fitting based on Monte Carlo and Optimization Strategy

Li Yuan [1, †], Xuan Zhou [2, †]
Liy127@gmail.edu, Xz321@duke.edu

[1]School of engineering University of Dayton Dayton, Ohio

[2]Fuqua School of Business Duke University Durham, North Carolina

[†]These authors contributed equally.

Abstract— Investment strategy optimization is largely discussed in the financial industry and academia as investors seek to maximize the Return with limited input. This paper discusses how potentially profitable stocks can be selected using statistical models to form a portfolio and then use the predicted prices to find the optimal allocation strategy. We first use the ARIMA Model to predict the stock price on the following 21st day. We further verify the stock selection with Monte Carlo Simulation predicting the price ranges which the stocks will fall into in the next year. Then Black-Litterman Model is used to optimize the asset allocation. We select five stocks with non-negative returns based on the results to form the target portfolio: TSLA, KO, AMD, NKE, and ORCL. According to our simulation, the stocks have an annual profit of 3%, 2%, 7,8%, 3.4%, and 5.1%, respectively. The 95% confidence interval of those five stocks is small which demonstrates that the stock we choose has low risk. For instance, the confidence interval for the TSLA is (-0.05,0.06). Next, we get the weight of each stock to be 27.92%, 0%, 39.61%, 10.10%, and 22.38%, respectively. This asset allocation is the optimal choice. Comparing to the equal weight stock selection, although the equal weight stock selection shows lower variance, our asset allocation has higher profit. In conclusion, we show that the above five stocks will have positive profits in both the short and long run.

Keywords: Investment strategy optimization; financial modeling; ARIMA Model; Monte Carlo Simulation; Black-Litterman Model; stock allocation

1 INTRODUCTION

Asset allocation and Stock Price are predictable among the most important and popular tasks in finance [1]. This problem is hard because there are too many factors influencing the stock price. For instance, the paper by Kurihet al.t al [2] studies that the factor such as gross domestics income, interest rate, and monthly supply [2] influences the stock price and Arzoo, Shamail (2011) gives a case study on Karachi Stock Exchange [3]. There are also microeconomics factors that influence the Stock Price. For example, Pražák.Tomáš. (2020) points out that liquidity rate is a significant factor in Stock Price [4].

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2326922

Starting from 1990, people have been using Stochastic Differential Equation Models [5,6], statistical methods [7,8], and machine learning [9] technics to deal with this issue. Assuming that the log price of the stock is Gaussian Random Walk, people use the geometric Brownian motion model to model the stock price and derive the Black-Scholes Equation [6,7]. Additionally, people could add a jump on the stock price model to make the jump-diffusion process or Levy Process more reasonable [10]. On the other hand, people also use the deep learning approach to predict the stock price [9].

However, in practice, both methods do have some drawbacks. For the Geometric Brownian motion approach, it might not be perfect to assume that the log price is simply a random walk as that the stock price is not a Markov Process For the deep learning approach. However,h the deep learning model might give the researcher reasonable results if building a good architecture and turning the parameter well. It still does not have a relatively perfect theoretical foundation. For instance, the optimal solution may converge to the local-optimal solution instead of the global-optimal solution because finding the optimal global solution is NP-hard [11]. Our paper uses the ARIMA-Model to predict the stock price in the short term and select stocks from the whole pool based on the prediction [9]. This model has a very well-developed theory, and it is a more generalized approach than simply assuming the log price is a random walk. Using the stock's passing data, we could determine the appropriate p,q,d for the ARIMA model.

After that, we will deal with the asset allocation problem. The asset allocation problem is not trivial because one must balance maximizing the profit and minimizing the risk. And the best way of dealing with this problem is to switch the asset allocation problem to find the optimal solution. Noticing that Markov Inequality and Chebyshev [12] inequality demonstrate that variance depicts the risk of the problem and the Expected Return represents the profit, we can use mean-variance optimization to model our asset allocation problem. However, this model is sensitive [13]. Best and Garuer (1991) show that if one perturbs the expected Return a little bit, it will give a quite different result than the unperturbed problem. As a result, we should seek a more robust model. And Black-Litterman Model [14] is the answer for us. Therefore, we will use the Black-Litterman Model to calculate the best asset allocation for selecting stocks. Although the Geometric Brownian Motion is not the best model to model the stock price, we will use Monte Carlo Simulation [15] to simulate the Geometric Brownian Motion, supporting our result for building the Black-Litterman model. Monte Carlo Simulation is one of the best ways to simulate and predict the future price in the long run due to its "randomness" and "deterministic." Randomness is because Monte-Carlo-Simulation gives many results, and we can check the robustness of our results. And the deterministic is by central limit theorem and law of large number; hence the general statistics are controlled by the error and will go to 0 if we simulate more. In the end, we will also use Monte-Carlo-Simulation to simulate the expected Return of our developed portfolio.

Section 2 briefly reviews the definition of the ARIMA model, the Black-Litterman Model, and the Monte-Carlo-Simulation. We also derive the formulas used in the simulation. Section 3 explains the numerical result we get from the model and how we choose the parameter. In section 4, we will give the paper's conclusion and talk about some issues of our method and how we could improve it in the future.

2 METHOD

2.1 Data preparation

The data was derived from Yahoo Finance [16] and Sina Finance [17]. All the data information is from Yahoo Finance except the market capitalization derived from Sina Finance.

We visited the website Yahoo Finance and found the historical price of several companies. Downloading the data, we put them into jupyter notebook and calculated their Return, which is used for further calculation. Since there is trouble to get the market capitalization, we also visited Sina Finance [2] to get the capitalization of the stocks we choose.

2.2 ARIMA-based Trend Prediction

The model we use is ARIMA.

$$(1-\sum_{i=1}^{p}\phi_i L^i)(1-L)^d X_t = (1+\sum_{i=1}^{q}\phi_i L^i)\epsilon_t \tag{1}$$

L^i is the Lag operator and $L^i(X_t) = X_{t-i}$ and ϵ_t are i.i.d random variable.

By programming the code, we drew the graph and calculated the value of the p,d,q based on the result of the graph.

We set 2 years as a period when doing the ARIMA model and draw the time series plot, the autocorrelation plot, and the partial autocorrelation plot to see if the stocks we choose have low value.

After drawing the plot, we get an interval of the stock price on the 21st day. And we also calculate the value of p,d,q based on the result of the graph.

Finally, we chose the five most well-behaved stocks based on their price prediction.

2.3 Price prediction based on Monte Carlo

We use the Monte Carlo method to predict the price of the five stocks we choose to prepare for the View of Black-litterman model.

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{2}$$

S_t is the stock price, μ is the drift term, and it represents the expected Return, σ is the diffusion term, and it represents the risk of the stock. Now, by the Log price definition $V_t = \ln(S_t)$ and the Ito Lemma. We will have the log return to be:

$$dV_t = d(\ln S_t) = (\frac{-1}{2S_t^2}(\sigma S_t)^2 + \frac{1}{S_t}\mu S_t)dt + \frac{1}{S_t}(\sigma)S_t dW_t$$

$$= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t \tag{3}$$

We do the Monte Carlo for 10000 and get a distribution plot of the five stocks. In the plot, we selected the start price and final price using a 95% confidence level. We decided our view and confidence interval with references from the companies' products' market sentiment and conditions

We use the Monte Carlo method again to test the stocks. The purpose of the step is to judge whether the View we made is effective. The judgment is made by comparing it with an equal weight portfolio.

2.4 Asset allocation based on Black-litterman model

We use the prior Return to calculate the posterior expected return and covariance matrix by programming the code.

Posterior Expect Return:

$$E(R) = [(\tau\Sigma)^{-1} + P^T\Omega^{-1}P]^{-1}[(\tau\Sigma)^{-1}\Pi + P^T\Omega^{-1}Q] \tag{4}$$

E(R) is an $K \times 1$ vector of expected returns, where N is the number of assets. Q represents a $K \times 1$ vector of views. P is the $K \times N$ picking matrix which maps views to the universe of assets. Essentially, it tells the model which View corresponds to which asset(s). Ω is the $K \times K$ uncertainty matrix of views. Σ is the $N \times N$ covariance matrix of asset returns. τ denotes a scalar tuning constant.

Posterior Covariance Matrix:

$$\Sigma^{BL} = \Sigma + [(\tau\Sigma)^{-1} + (P^T\Omega^{-1}P)]^{-1} \tag{5}$$

Then, with the view and confidence interval we got from the Monte Carlo process, we calculated the asset allocation.

3 RESULT AND DISCUSSION

3.1 Arima-based Trend Prediction

From the ARIMA analysis, we get the value of p, d, q, and then we predicted the stock price in 21 days. Then, we choose five stocks that probably has positive Return in 21 days. The predicted value is shown in the following plot. (Take TSLA as an example)

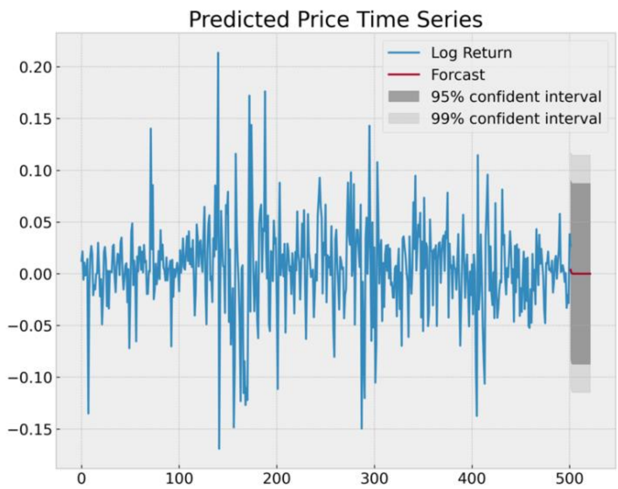

Figure 1. Predicted price time series for TSLA

The blue line visualizes the time series of the price of TSLA. The red line shows the price prediction in the future 21 days. The 95% confident Interval at 21 days is between -0.0873 and 0.0873. The 99% confident Interval at 21 days is between -0.1147 and 0.1147, indicating that TSLA

We use logistics regression to find the best ARIMA(p,d,q) model to fit the log return of the data. To avoid complication, we restrict d to be 0 or 1 and p to be a number between 0 and 5. After finding the best parameter, we use the model to predict the log return (keep in mind that the log return is approximated to equal the true Return when the time step is small because of the Taylor Expansion)

Table 1 Predicted Price Time Series For All 5 Stocks.

No.	Predicted price time series for all five stocks				
	Stock Name	***Return (95% confident interval) %***	***Return (99% confident interval) %***	***p, d, q***	***Return %***
1	TSLA	[-8.73, 8.73]	[-11.47, 11.47]	1, 0, 0	0
2	KO	[-3.25, 3.30]	[-4.28, 4.33]	1, 0, 4	.03
3	AMD	[-5.29, 5.93]	[-7.05, 7.69]	3, 0, 3	.32
4	NKE	[-4.13, 4.33]	[-5.46, 5.66]	3, 0, 3	.10
5	ORCL	[-3.23, 3.23]	[-4.24, 4.24]	0, 0, 1	0

We showed the ranges in both 95% and 99% confidence intervals. And Colum the Return is the result of predicted Return. From d = 0, the plot shows that the time series of the price is rationalized. We judge whether to choose one stock according to whether the Return of each stock is positive or negative. If the Return is negative, the stock will be abandoned. From the table, the returns of the five stocks are not negative. Thus, we choose those stocks.

3.2 Price prediction based on Monte Carlo

After getting these five stocks, we put them into the Monte Carlo process and get an interval of the stock price. (Take 'TSLA' for example)

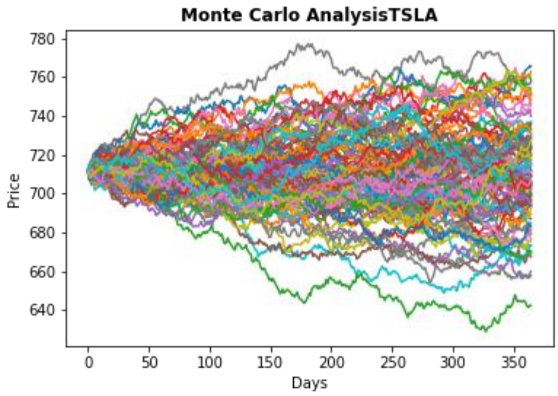

Figure 2. Monte Carlo analysis of TSLA

The figure shows 10000 random routes that the price will be 365 days (derived on Aug 27, 2021). From the density of the distribution, we can tell that the price conforms to the central distribution. It has a centered distribution feature because the random price is formed day by day according to normal distribution.

Since the route in figure 2 is not clear enough to show the result, we transferred the locus plot to the distribution plot.

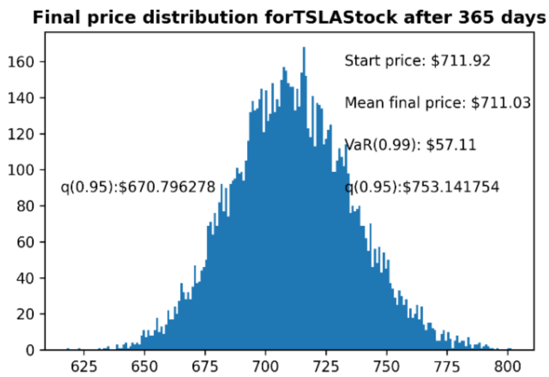

Figure 3. Final price distribution for TSLA stock after 365 days

This plot shows the distribution of each X-axis represents the price, and Y-axis represents the frequency of the price. As mentioned before, this plot shows the feature of normal distribution. It shows that the predicted price will be between $670.8 and $753.1 with a confidence interval of 95%.

Therefore, we can get the Interval of each stock, which the future prices are possibly falling into. The confident Interval is 95%.

Table 2 Result Of Monte Carlo

No.	**Result of Monte Carlo**			
	Stock Name	***Start price***	***End Price(95% confident interval)***	***VaR***
1	TSLA	$711.92	[$670.93, $753.87]	$57.31
2	KO	$55.65	[$54.34, $56.98]	$1.84
3	AMD	$111.40	[$106.03, $120.0]	$8.11
4	NKE	$167.58	[$161.55, $173.35]	$8.43
5	ORCL	$89.35	[$85.17, $93.82]	$5.81

The start price is the price in the date when we use the Monte Carlo model (2021-08-28). VaR is the 'Value at Risk' of each stock, which shows the maximum loss an investment will suffer to a given degree. The result of the End Price is the range of predicted price in 95% confident Interval. Monte Carlo results are in accord with Normal distribution.

3.3 Asset allocation based on Black-litterman model

We choose the price in the Interval by considering the market view of each stock.

Therefore, we get the View and the confident Interval of each stock

- TSLA will increase by 3% after 365 days.
- KO will increase by 2% after 365 days.
- AMD will increase by 7% after 365 days.
- NKE will increase by 3.4% after 365 days.
- ORCL will increase by 5.1% after 365 days.

Table 3 Decision of View and Interval

No.	**Decision of View and Interval**		
	Stock Name	***View***	***Intervals***
1	TSLA	TSLA will increase by 3% after 365 days.	(-0.05, 0.06)
2	KO	KO will increase 2% after 365 days.	(0, 0.02)
3	AMD	AMD will increase 7% after 365 days.	(0, 0.09)
4	NKE	NKE will increase 3.4% after 365 days.	(0, 0.04)
5	ORCL	ORCL will increase 5.1% after 365 days.	(0, 0.06)

The view shows the trend of each stock in 365 days, estimated according to the Monte Carlo results. The Interval shows how much certainty we have on each stock. It is bounds estimated synthetically by specifying one standard deviation confidence interval and the Monte Carlo result. (Data derived on 2021-08-28).

Then input the Views and Intervals and come out the weight of each stock. After calculating the weight, we get the portfolio.

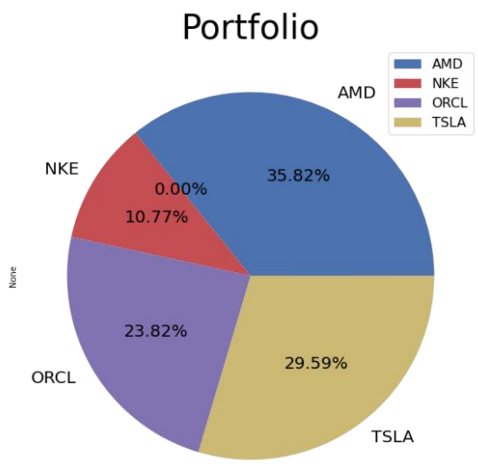

Figure 4. Portfolio of chosen stocks from the B-L model

The proportion of TSLA is 27.92%. The proportion of AMD is 39.61%. The proportion of NKE is 10.10%. The proportion of ORCL is 22.38%. And there is no KO in this portfolio.

With the portfolio, we use Monte Carlo for the second time to see the change of the Price interval and the VaR. We compare our portfolio with an equal weight portfolio.

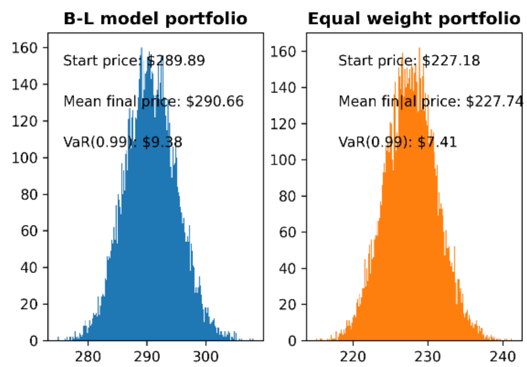

Figure 5. Comparison between the portfolio from the B-L model and the portfolio from the Equal-weight portfolio

The plot above shows the distribution of the Monte Carlo results in the future 365 days. It can be concluded that the mean final price, which is the possible price, has a more conspicuous change from the start price in the B-L model portfolio than in the Equal weight portfolio. It means the B-L model portfolio can probably have a greater return than Equal-weight Portfolio. At the same time, VaR in the B-L model portfolio is higher than the Equal-weight portfolio. That is because the start price of the portfolio is higher than the Equal-weight portfolio.

4 CONCLUSION AND FUTURE WORK

In summary, according to the numerical results, our optimization process provides very reliable and promising returns. Therefore, we believe it is a good approach to select the stocks, build the portfolio and find the best asset allocation. However, due to the limit of time, we do not compare our results to other existing methods, such as the deep learning method. Hopefully, the comparison will be conducted in future research. We may also consider using our method on stock markets in other nations as other countries have different financial systems, impacting the effectiveness of the models we choose. For example, we assume that the stock price is nearly stationary, which means that by taking enough finite difference, we will get a stationary sequence. This might not necessarily be true in other stock markets. When it comes to giving View, the price we get should take more than Monte Carlo into account. Finally, although we combine the short-term model (ARIMA) and the long-term model (Monte Carlo Simulation) to overcome the limitation of using short or long-run predictions, we still have some inevitable errors from these models' limitations.

REFERENCES

[1] Friedman, Benjamin M., et al. "Economic Implications of Extraordinary Movements in Stock Prices." Brookings Papers on Economic Activity, vol. 1989, no. 2, 1989, p. 137. DOI.org (Crossref), https://doi.org/10.2307/2534463.

[2] Arzoo, Shamail. "Political Instability & Its Impact on Karachi Stock Exchange (Impact on Trading Volume and Trading Value, 2011)." SSRN Electronic Journal, 2011. DOI.org (Crossref), https://doi.org/10.2139/ssrn.1801231.

[3] Kurihara, Y. The Relasummarytionship between Exchange Rate and Stock Prices during the Quantitative Easing Policy in Japan. International Journal of Business, 11, 375-386. http://www.craig.csufresno.edu/ijb/Volumes/Volume%2011/V114-3.pdf

[4] Pražák, Tomáš. "The Role of Main Microeconomic Factors on the Stock Prices of Selected Swiss Companies." E-Finanse, vol. 16, no. 1, Mar. 2020, pp. 67–74. DOI.org (Crossref), https://doi.org/10.2478/fiqf-2020-0007.

[5] Yang, Zhijun, and D. Aldous. 2015. Geometric Brownian Motion Model in Financial Market. Berkeley: University of California

[6] Stochastic Calculus for Finance. 2: Continuous-Time Models. Corr. print, Springer, 2008.

[7] D. Banerjee, "Forecasting of Indian Stock Market Using Time-Series ARIMA Model", In: Proceedings of the 2nd International Conference on Business and Information Management, IEEE, 2014, pp.131–135.

[8] Makridakis, Spyros, and Michele Hibon. 1997. Arma models and the box–jenkins methodology. Journal of Forecasting 16: 147–63.

[9] Shen, Jingyi, and M. Omair Shafiq. "Short-Term Stock Market Price Trend Prediction Using a Comprehensive Deep Learning System." Journal of Big Data, vol. 7, no. 1, Dec. 2020, p. 66. DOI.org (Crossref), https://doi.org/10.1186/s40537-020-00333-6

[10] Kou, S. G. "Chapter 2 Jump-Diffusion Models for Asset Pricing in Financial Engineering." Handbooks in Operations Research and Management Science, vol. 15, Elsevier, 2007, pp. 73–116. DOI.org (Crossref), https://doi.org/10.1016/S0927-0507(07)15002-7\

[11] Hochbaum, Dorit S. "Complexity and Algorithms for Nonlinear Optimization Problems." Annals of Operations Research, vol. 153, no. 1, June 2007, pp. 257–96. DOI.org (Crossref), https://doi.org/10.1007/s10479-007-0172-6

[12] Billingsley, Patrick. Probability and Measure. Anniversary ed, Wiley, 2012

[13] Best, Michael J., and Robert R. Grauer. "On the Sensitivity of Mean-Variance-Efficient Portfolios to Changes in Asset Means: Some Analytical and Computational Results." Review of Financial Studies, vol. 4, no. 2, Apr. 1991, pp. 315–42. DOI.org (Crossref), https://doi.org/10.1093/rfs/4.2.315.

[14] Cayirli, Omer. "The Black-Litterman Model: Extensions and Asset Allocation." SSRN Electronic Journal, 2011. DOI.org (Crossref), https://doi.org/10.2139/ssrn.3464770.

[15] Thomopoulos, Nicholas T. Essentials of Monte Carlo Simulation: Statistical Methods for Building Simulation Models. Springer, 2013.

[16] Yahoo finance https://finance.yahoo.com/

[17] Sina finance http://hq.sinajs.cn/list=gb_