A Study on Backtest Metrics for Financial Analysis

Simge YILDIRIM[*1], Yunus SANTUR[1]

[1]*Software engineering / institute of science, Fırat University, Turkey*

[*]*(yildirimsimge8@gmail.com.com) Email of the corresponding author*

*Abstract –*Backtesting would be an example of a situation that allows the doctor to have an idea of the diagnosis by questioning the patient's history of the disease and considering a few complaints he or she has had. Backward performance metrics play an important role in the evaluation process of a trading strategy. When evaluating a trading strategy, key metrics such as sharp rate, maximum drop, profit factor, number of trades, calm rate, volatility are calculated and analyzed. Data of AAPL, AMZN, SPY, GOOG shares are obtained for this particular period presented. A simple trading strategy was applied on this structure and retrospective performance metrics were measured on it. By comparing the results of these measurements, it is observed how important and effective these measurements are on a trading strategy. It is concluded that the criteria in these observations do not always give the same result and vary from strategy to strategy, from data to data. It was concluded that performance metrics and the number of transactions should be evaluated together when performing backtesting on a trading strategy. It is recommended to consider the results of the presented study in order to make a profitable estimation with less risk.

Keywords – Backtest, Sharpe Ratio, Profit Factor, Time Series, Financial Forecasting

I. INTRODUCTION

Backtesting predicts how successful a financial model or strategy might have been in the past, based on historical data in the financial field. The main purpose of this backtesting is to evaluate the performance of the strategy in different market conditions and to identify possible problems. It is a metric used when backtesting strategy, evaluating its performance, and in financial models. Backtest is frequently preferred by professionals and investors in the process of investing in financial markets. Backtest not only measures past performance, but also predicts future performance. However, since future predictions are made based on past data, the accuracy of future predictions does not give absolute assurance.

When examining the performance of a strategy, it is necessary to pay attention to the profit factor (pf), sharp ratio (saharpe ratio (sr)), maximum reduction (maxDD) criteria.

A. Profit Factor (PF)

The profit factor is defined as gross profit divided by gross loss (including commissions) for the entire trading period. This performance metric correlates profit per unit of risk with values greater than one indicating a profitable system. [1.] PF is the measure calculated by the ratio of total net profit to net loss. In profitable strategies, PF has a value greater than 1. A profit factor of 1 indicates that the system or strategy is at par. A PF less than 1 indicates a loss instead of a profit.

B. Sharpe Ratio (SR)

SR is a measure that measures the return adjusted for the risk situation of the strategy. A high SR

indicates that the strategy generates more returns for some risk situations. If the SR is greater than 1, it is considered good, if it is greater than 2, very good, and if it is greater than 3, it is considered excellent [2]. The SR equation is calculated as in (1).

$$SR = \frac{E(R_t) - R_f}{\sqrt{Var(R_t)}} \qquad (1)$$

"R_f" stands for risk-free investment and "R_t" stands for average return on investment [4].

C. Maximum Drop (MaxDD)

The maximum bearish metric refers to the "worst case scenario" for a trading period. It measures the largest distance or loss from the previous stock peak. [3] MaxDD is the maximum to minimum loss a strategy sees over a given period of time. In other words, it is the difference between the maximum and minimum values of a stock. The larger the MaxDD, the greater the risk in the strategy.

Backtest metrics have an important place in helping investors evaluate the performance of a strategy in different conditions, apply the appropriate strategy and make decisions on this issue.

II. MATERIALS AND METHOD

In this study, a comparison of Backtest results was made. Performance criteria such as Sharpe ratio, maximum decrease, profit factor was determined. It is aimed to make a more realistic comparison between the results by paying attention to the fact that the data are in the same time interval. The same coin size was used for comparison. Since seeing the results of one strategy during the 2020 Covid-19 pandemic and seeing the results of another strategy in a normal period will lead to different results, time intervals with similar market conditions were preferred in order to avoid making wrong decisions. The flow chart of the process sequence applied in the presented study is shown in Figure 1.
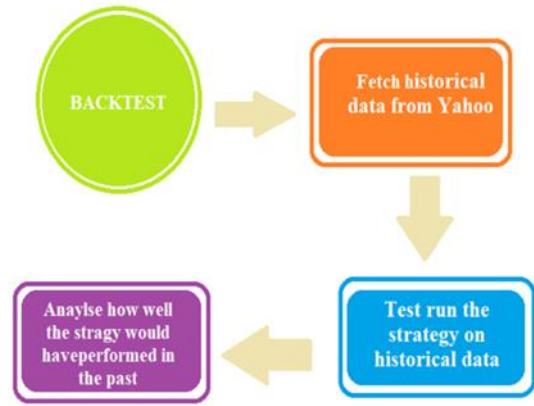

Fig. 1. The flow charts

It is quite important and valuable to use a trading strategy as soon as possible. There are many readily available libraries that provide the opportunity to make these assessments. However, these evaluations of operating functions have been made through the use of these existing libraries through the Python programming language. Even if ready-made libraries do the same work, it is more exciting to be sure of the reliability and evaluation with the function he wrote. First, a strategy is aimed to be able to implement functions and evaluate retrospective performance metrics [7]. More than one strategy has been tested in order to determine whether it is affected by the strategy or not.

Python programming language was preferred in the presented study. Python's libraries such as pandas, numpy, matplotlib, yfinance are loaded and included in the project. The study was carried out in the notebook environment of Google Colab. The code block for the process of obtaining historical data of preferred stocks through Yahoo finance is given in Figure 2.


```python
import pandas as pd
import numpy as np
import yfinance as yf
# hisse senedi verilerini al
tickers = ["AAPL", "AMZN", "GOOG","SPY"]
start_date = "2010-01-01"
end_date = "2023-03-08"
data = yf.download(tickers, start=start_date, end=end_date)['Adj Close']


[*********************100%***********************]  4 of 4 completed
```

Fig. 2. Stock closing prices loaded

This is how he uses the impressions of "AAPL, GOOG, SPY, AMZN" that everyone can access for free. From this part, AAPL, APPLE Inc. is a technology company that designs and develops computer software and online services. GOOG is Google's parent company, Alphabet Inc., which provides online advertising services, search engines, and cloud computing services. SPY stands for the SPDR S&P 500 ETF Trust high-restricted US stock, an exchange-traded fund [8]. AMZN is an e-commerce and cloud computing company that offers a wide variety of products and services to owners and businesses representing Amazon.com, Inc. The closing price data of the aforementioned stocks are shown in the Figure 3[4].
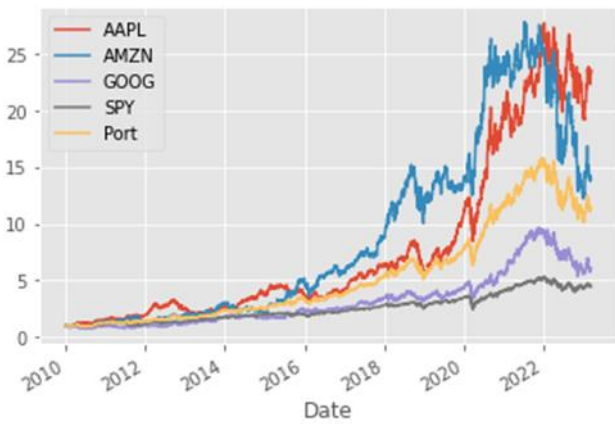

Fig. 3. Stock closing prices

Start date "2010-01-01" and the end date is set as "2023-03-08". The sharpe ratio of the data was calculated using the formula (1).

It has the best Sharp, with AAPL 0.98315 in the first strategy to the execution strategy. In terms of AAPL profit factor, it has the highest value with 1.03606 compared to other papers. With SPY 0.41344 value MaxDD, followed by 0.42452 value, AAPL looks at the smallest MaxDD values. With 76 transactions, AAPL has the highest number of transactions, and SPY has the lowest transaction fee with 67. In the second strategy, the SPY card has the highest hash rate of 0.27535. The profit factor of the SPY package is 1.1909 with outstanding higher value. With AMZN 0.12413 it is the paper with the smallest MaxDD value. Judging by the transaction fee, the highest transaction is SPY with 214, while the lowest is GOOG card with 184. An investor or financial person should not only consider performance metrics when backtesting a paper. With the presented values, it is aimed that process monitoring should be taken into consideration as

well as performance metrics. It is a necessity to experience great risks and losses if the decision is made only by looking at performance metrics. In order to avoid such a situation, when measuring any sentiment on a trading strategy retrospectively, it is necessary to evaluate the overall transaction along with the performance metrics [9].

The strategies and performance metrics performed in this study are comprehensively demonstrated. The table and graph containing the data of Strategy-1 are shown in Figure 4.
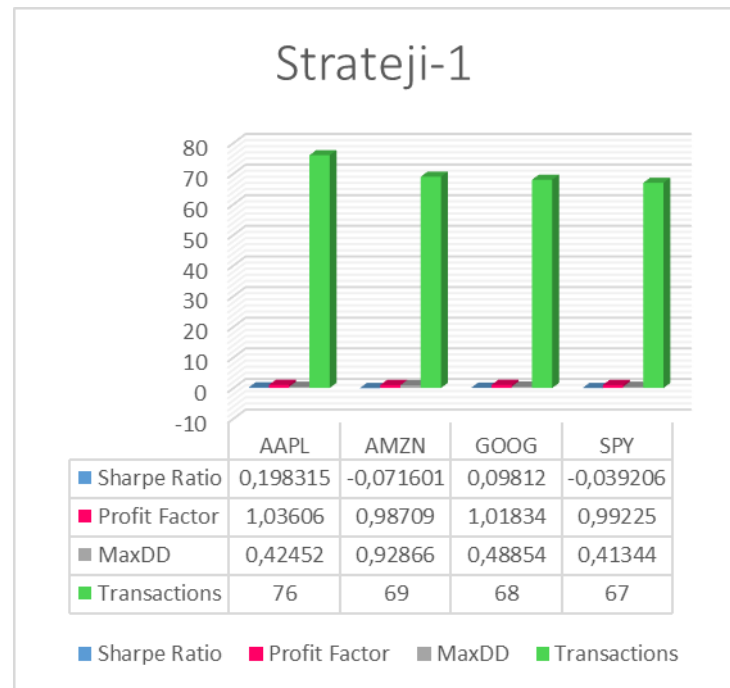

	AAPL	AMZN	GOOG	SPY
Sharpe Ratio	0,198315	-0,071601	0,09812	-0,039206
Profit Factor	1,03606	0,98709	1,01834	0,99225
MaxDD	0,42452	0,92866	0,48854	0,41344
Transactions	76	69	68	67

Fig. 4. Strateji-1

The table and graph containing the data of Strategy-2 are shown in Figure 5.

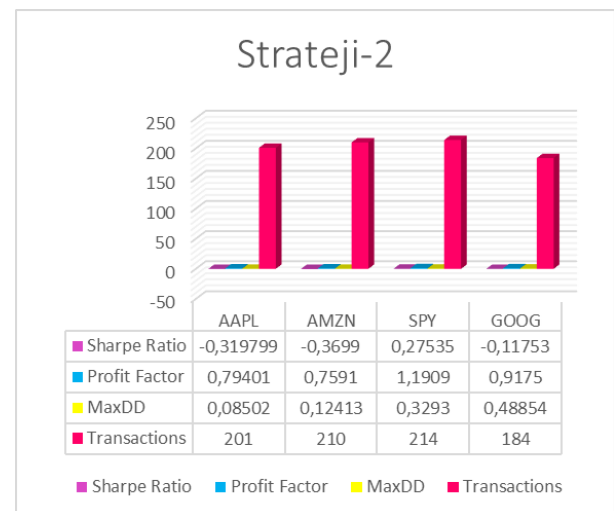|	AAPL	AMZN	SPY	GOOG
Sharpe Ratio	-0,319799	-0,3699	0,27535	-0,11753
Profit Factor	0,79401	0,7591	1,1909	0,9175
MaxDD	0,08502	0,12413	0,3293	0,48854
Transactions	201	210	214	184

Fig. 5. Strateji-2

Profit factor is a preferred performance measurement metric used to reduce the profit or loss of a trading strategy. Profit factor is the value obtained by cross-sectioning total profit to total loss [10]. A calculated value greater than 1 means that the trade chain is a profitable business for the individual. This was calculated as a profit for stocks separately.

If compared in terms of profit factor according to Strategy-1, it is concluded that AAPL will be a more profitable choice than other stocks. If compared to Strategy-2 in terms of profit factor, SPY would be a more accurate choice. The maximum drop usually has a negative range of values. When an evaluation is made in terms of MaxDD, it would be a better choice to base the smallest value [5]. In the presented study, SPY and AAPL in strategies are in a smaller value range in terms of MaxDD compared to other stocks.

It's possible to steal a stock's strategy by looking at Sharpe. The cognate calculations that take these ratios into account show that AAPL sentiment has the highest ratio. It's possible to steal a stock's strategy by looking at Sharpe. These ratios commented that the AAPL sense gave good results within the strategy used. A Sharpe asset of 3 means that his strategy is great, but not a perfect result with this unused strategy [11]. Future chapters aim at the components and outcomes of the strategy to achieve more excellent results.

There are many ready-made libraries for backtesting. However, in this study, necessary calculations were made by creating functions with codes based on formulas without using existing libraries [6]. The purpose of doing it this way is to make the comparison of performance metrics more realistic when backtesting.

III. RESULTS

In the study, SPY with the best 0.27535 sharpness ratio, AAPL with the lowest MaxDD values -0.085022 and AAPL with 1.3606 profit factor were determined. The smallest number of transactions, which also protects transaction groups, belongs to 67 SPY cards. That the results in front of the trading eye will show the expected results. Because if the person has decided only by looking at the performance metrics, it is expected that he will prefer the AAPL case, but it will be a more correct decision to choose the SPY bag when it is natural to

evaluate the number of transactions in order to make a correct guess and take less risk. The retrospective test performance metrics applied for strategy-1 and strategy-2 in the study vary. The obtained materials and the method are explained in detail with graphics and tables. It would be inaccurate to consider a single backtesting performance metric of a trading strategy. Even if all metrics are calculated and analysed and the necessary strategies are implemented, there is always risk. However, there is no risk-free profit [12].

IV. DISCUSSION

When the data obtained from the study are evaluated, it is seen that the results differ according to the strategy and stock. The necessary literature has been made and no similar study has been found. It is thought that better results can be achieved by applying this situation in the future.

V. CONCLUSION

As a result of the study, the importance of performance metrics and trading behaviours when backtesting a trading strategy has been understood. These performance metrics are different among themselves, they show the results obtained. Changes were made according to the strategy used in the operation of the results.

REFERENCES

[1] (2023) TradeStation. [Online]. Available: https://help.tradestation.com/10_00/eng/tradestationhelp /subsystems/spr_topics/report/profit_factor__strategy_p erformance_report_.htm

[2] (2023) İnvestopedia. What is a good Sharp ratio? [Online]. Available: https://www.investopedia.com/ask/answers/010815/wha t-good-sharpe-ratio.asp

[3] (2023) CFA Institute. [Online]. Available: "Sculpting Investment Portfolios: Maximum Drawdown and Optimal Portfolio Strategy, https://blogs.cfainstitute.org/investor/2013/02/12/sculpti ng-investment-portfolios-maximum-drawdown-and-optimal-portfolio-strategy/

[4] Olorunnimbe, K., & Viktor, H. (2023). Deep learning in the stock market—a systematic survey of practice, backtesting, and applications. Artificial Intelligence Review, 56(3), 2057-2109

[5] Feng, S., Wang, N., & Zychowicz, E. J. (2017). Sentiment and the performance of technical indicators. The Journal of Portfolio Management, 43(3), 112-125.

[6] Scognamiglio, S., & Marino, M. (2022). Backtesting stochastic mortality models by prediction interval-based metrics. Quality & Quantity, 1-23.

[7] Wong, L. (2019). Introductory Backtesting Notes for Quantitative Trading Strategies.

[8] Tayalı, S. T. (2020). A novel backtesting methodology for clustering in mean–variance portfolio optimization. Knowledge-Based Systems, 209, 106454.

[9] Pitera, M., & Schmidt, T. (2022). Estimating and backtesting risk under heavy tails. Insurance: Mathematics and Economics, 104, 1-14.

[10] Lindquist, W. B., Rachev, S. T., Hu, Y., & Shirvani, A. (2022). Backtesting. In Advanced REIT Portfolio Optimization: Innovative Tools for Risk Management (pp. 113-129). Cham: Springer International Publishing.

[11] Banulescu-Radu, D., Hurlin, C., Leymarie, J., & Scaillet, O. (2021). Backtesting marginal expected shortfall and related systemic risk measures. Management science, 67(9), 5730-5754.

[12] Christoffersen, P. (2008). Backtesting. Available at SSRN 2044825.