**A Proposal to Time Series Analysis for Forecasting S&P 500 (Standard & Poor’s 500) Index with Macro-Economic Indicators**

Junseong Lee

February 2023

# Introduction

Time series analysis and statistical hypothesis analysis gives investors a way to predict the prices of stocks or derivatives (options) using another time-series data, ultimately providing decision-helping factors for firms to make investment decisions. The time series analysis includes methods like Augmented Dickey Fuller Test, and statistical hypothesis test includes methods like Serial Correlation (also known as Autocorrelation) and Granger Causality Test. All of these tests are critical to forecast the stock data for following reasons:

• The ADF testing computes unit root test for stationarity of data, ensuring that the data is in time-series

• Autocorrelation examines the similarity between observations of a variable as a function of time lag between each observation

• Granger Causality becomes an investigative method to find causality between two time-series variables

The time series analysis and statistical hypothesis analysis are crucial to find leading indicators that ultimately give clues to the future moves of S&P 500 Index or the general US stock market. This strategy is derived from historical data and uses accurate calculations of data to find correlations between S&P 500 Index (ticker: SPX) and a macroeconomic indicator. While technical analysis finds patterns within the candlestick charts by using technical indicators—such as Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Vortex Indicator, Bollinger Bands, and Ichimoku Cloud Strategy—the findings of technical analysis are completely perspective-based (depends on how the person views) and cannot predict the future price moves accurately. Technical analysis assumes that the patterns happened before in history will continue to hold true going forward in the future. Consequently, the goal of time series analysis and statistical hypothesis test is to make predictions of stock prices through computational reasonings. These statistical methods use mathematical formulas to evaluate the nature of provided stock data, and therefore, conducting time-series analysis and statistical hypothesis testing requires high understanding of statistics as well as relative programming experience to deploy such methods.

# Statement of Problem

# The problem in financial forecasting methods is the difficulty of accurately predicting stock prices due to the unpredictable and volatile nature of the stock market. Traditional methods such as technical analysis rely on patterns in historical price data, which may not accurately reflect future market trends. For example, majority of technical indicators, such as Moving Average Convergence Divergence and many others, throw buy and sell signals; however, sometimes the signals can be inaccurate or misleading. Time series analysis, on the other hand, uses statistical methods to analyze historical data and identify correlations with other relevant economic indicators to make predictions about future stock prices. The proposed use of time series analysis and Granger Causality Test in this proposal aims to provide a more accurate and reliable approach to forecasting S&P 500 Index movements, helping investors make more informed investment decisions.

# Objectives

In this paper, I will achieve the following two goals:

1. Explain the basic structure of two time-series and causality tests: ADF Test and Granger Causality Test
2. Use ADF Test and Granger Causality test on real-time historical SPX prices and a set of Macro Economic indicator values to evaluate how different indicators are related to the S&P price move.

This paper attempts to find a data series or macro-economic indicators which could be predictive of asset price movements and therefore actionable by investors. The steps will include:

• Assess potential relationships between S&P Price and Macro-economic indicators.

• Build and fit a model to the data of S&P500 and economic indicators.

• Interpretate the model and find a possible association between them.

With the development of algorithmic trading with artificial intelligence, countless quantitative and finance researchers have come up with different machine learning algorithms and forecast models, such as linear regression, Discounted Cash Flow Model, Dividend Discount Model, and others, to predict future stock prices. Having said that, according to the efficient market hypothesis in finance theory, the stock market is highly reactive to the public information and user sentiment, meaning that historical prices and volume data might not be very useful for forecasting stock prices, and while fundamental analysis (financial statement analysis) can provide some insights to a company or economy’s future prospects, the time-series analysis and statistical hypothesis test are essential to draw accurate conclusions.

# Plan of Action

This section exhibits my plan for obtaining the listed objectives discussed in the section above. My first goal of the project is to explain basic structure of the two methods: ADF Testing and Granger Causality Testing.

Beginning with ADF Test, the null hypothesis for this test is that there is a unit root, and the alternate hypothesis differs slightly according to which equation you’re using. The basic alternate is that the time series is stationary (or trend-stationary). Also, ADF Test and Autocorrelation can be used together.

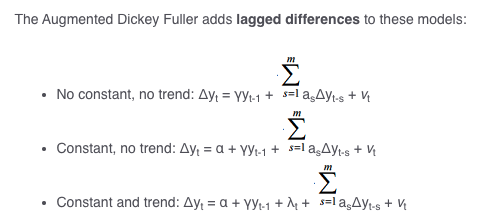


Figure 1. ADF Test Mathematical Formulas

Now, the theoretical concept of Granger Causality is closely tied to the concept of cause-and-effect, as everyone knows; however, they are not the same. For example, in cause-and-effect, variable X is casual to variable Y if X is the cause of Y or vice versa. On the other hand, Granger Causality tests if a variable comes before another in the time series; for example, large sales of turkey Granger-cause Thanksgiving.

Essentially, Granger causality test is an investigative method to find causality between two variables that are in time series. Because the variables must be in time series, conducting ADF Test before doing assessing Granger Causality Test is essential. The method is a probabilistic account of causality; it uses numeric data sets to find patterns of potential correlation. The visualization of Granger Causality is shown below:

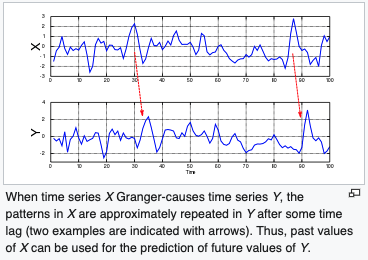


Figure 2. Granger Causality Visualization

Granger Causality Formula is as follows:

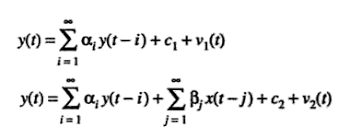
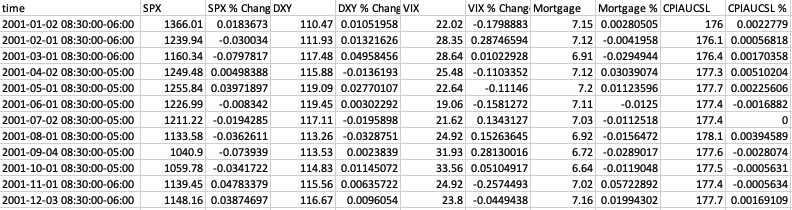


Figure 3. Granger Causality Formula

If the Granger causality test shows that changes in one stock price have a significant effect on another stock price, then this suggests that there is a causal relationship between the two stocks. This information can be used to make predictions about the future performance of one stock based on the past performance of another stock.

The second goal of my research is to apply these two methods on real time data to see how the results perform in being able to predict SPX using a macroeconomic indicator. To achieve this goal, the historical data of SPX I use includes monthly closing price from 2001/1/2 to 2022/11/01, and I combined the SPX data with 15+ macroeconomic indicator data that I extracted from Tradingview.com. (Figure 4) These indicators include: VIX, Mortgage rate, Consumer Price Index, PPI, PMI, 10 yr Treasury Yield, Unemployment rate, Interest Rate, OECD Composite Leading Indicator, Industry Production, Retail Sales, Consumer Confidence Index, S&P Quarterly Sales. I will utilize Python to complete this experiment, including its related libraries, such as NumPy, Matplotlib, Pandas, Statsmodel, Sklearn, and Datetime.

Figure 4: Data (partially shown)

**Management Plan**

I am a senior at the University of Illinois at Urbana-Champaign majoring in Finance + Data Science with minor in Computer Science. In the two years as a college student at University of Illinois, I had the opportunity to participate in many meaningful organizations and as well as projects, including FACES Consulting, Alpha Investment Management Partners, Quant, Data Science and Artificial Intelligence Society, and Association for Quantitative Trading Education. Working in an investment management and data science organizations has provided me with invaluable real-world experience and honed my ability to work effectively by utilizing statistical and financial knowledge. Furthermore, my self-study of the book "Charting and Technical Analysis" by Fred McAllen demonstrates my genuine interest and commitment to learning fundamentals of trading and having a deeper understanding of the financial market. Consequently, I am confident that my strong technical and analytical skills, combined with my positive attitude and strong work ethic, make me a strong candidate to complete this project.

# Conclusion

The proposed research will achieve the following goals: (1) Explain the basic structure of two time-series and causality tests: ADF Test and Granger Causality Test; (2) Use ADF Test and Granger Causality test on real-time historical SPX prices and a set of Macro Economic indicator values to evaluate how different indicators are related to the S&P price move. The research will be presented with a final summary of analysis in SPX prediction using a chosen indicator on April 11, 2023.

Reference

Direct, Science. “Granger Causality Test.” Granger Causality Test - an overview | ScienceDirect Topics. Accessed February 21, 2023. <https://www.sciencedirect.com/topics/social-sciences/granger-causality-test#:~:text=Granger%20causality%20approach.-,The%20Granger%20causality%20test%20is%20a%20statistical%20hypothesis%20test%20for,be%20rejected%20at%20that%20level>.

Thangavelu, Poonkulali. “Top 10 U.S. Economic Indicators.” Investopedia. Investopedia, April 14, 2022. <https://www.investopedia.com/articles/personal-finance/020215/top-ten-us-economic-indicators.asp>.

Fuller, W. A. “ADF -- Augmented Dickey Fuller Test.” Statistics How To, September 17, 2020. <https://www.statisticshowto.com/adf-augmented-dickey-fuller-test/>.

Peer Review Report

In the middle of writing my client proposal. Initially, I began with the topic of proposing a data science career development program to UIUC department of data science (including statistics team, and CS team), but I felt that I wanted to go deeper into what I did for my technical description: time-series analysis. My new proposal idea was to conduct time-series analysis and statistical hypothesis test to find a leading indicator for S&P 500 price moves and propose it to a financial firm for better investment decisions using the conducted test results. I have told all of my peer reviewers about this issue, and they were aware of my plan of switching topic before they evaluated my client proposal.

From the peer review, I mainly noticed that my management plan was vacant because I wasn’t sure how to draft the section. However, I received feedback from Gefei about management plan; he described to me that management plan had to provide needed qualifications (almost like prerequisite for taking a certain class) to have the intellectual capacity to complete this client proposal project. From his words, I have planned to generate my management plan, which was to describe my background and relevant experiences needed to complete the statistical analysis I mentioned in the proposal. I have added that section in order to verify my knowledge and previous experience. Also, besides Gefei Zhao, I received peer review feedback from two other students in my class, and they mentioned that the main point was clear to understand, which was: to explain what time series analysis is and ultimately identify a pattern or an indicator that could be used to create a better understanding of the future patters of stock movements. They also mentioned that the plan of action was extremely superb. While the plan of action section had a lot of technical terms, they told me that I did a great job of explaining certain technical terms that people without experience in financial field or statistical field would not know.

Ultimately, I believe that I had a strong start, with my draft being close to being done. I explained very clearly what the technical terms are, and what kind of statistical, mathematical method I will be using to analyze the data. I was able to benefit from peer review because I gained understanding of my audience from different majors (econ, astrophysics, and others) and realized that my statement of problem and the introduction gave them solid ideas of what I’m trying to do. I am glad that my peers were able to tell me directly what parts I need to work on. Overall, the peer review gave me an opportunity to assess my client proposal from different perspectives.