How Augmented Dickey Fuller Test, Autocorrelation, and Granger Causality Test Can Help Predict Stock Prices

Literature Review

Jun Lee (jl193)
UIUC jl193@illinois.edu

1 Introduction

In the stock market, predicting stock prices is a common way for investors to earn optimal profits and manage the risks of losses. Investors forecast stock prices in a variety of ways. Investors in the stock market use various methods to predict stock prices and manage their investments. Some investors rely on fundamental analysis, which involves examining a company's financial statements and economic indicators to determine its intrinsic value. Some others put more emphasis on technical analysis, which involves analyzing patterns of historical stock price movement and identifying entry and exit points. In addition, most investors build quantitative models, such as Correlation Forecasting and Statistical Forecasting, to carry out accurate estimates of future price movements based on multiple factors including volatility, interest rates, or even dividend yields. In recent years, two popular models, the Long Short-Term Memory (LSTM) neural network model and Autoregressive Integrated Moving Average (ARIMA) model through ADF Test + Granger Causality Test have been used to predict stock prices, and I will dive deeper into ARIMA model that uses stationary data as an input.

The remainder of this paper is organized as follows: Section 2 provides background information on technical methods (ADF Testing, Granger Causality Testing), Section 3 reviews several sources relevant to my thesis, Section 4 presents the findings of the research and provides future recommendations, and Section 5 finishes with related references.

2 Background

As the target audience for this research paper is the client, who may not possess a strong understanding of the underlying technicalities of statistical models such as the Augmented Dicker Fuller Test (ADF Test) and Granger Causality Test, it is imperative to provide a comprehensive background on these methods. Both methods can be implemented by utilizing a python library called *statsmodels*.

2.1 Augmented Dicker-Fuller Test (ADF Test)

The Augmented Dickey-Fuller test is a unit root test for stationarity of data, which extends the original Dickey-Fuller (DF) test developed by economists Clive W.J. Granger and Robert F. Engie in mid 1980s [4]. The ADF test considers the possibility of autocorrelation in the errors of the model and therefore makes it a more robust test for identifying unit roots [4]. A unit root is a characteristic of a time series that changes randomly over time and therefore cannot be predicted; having unit roots in data can cause unpredictable results in time series. The null hypothesis for ADF test assumes the existence of a unit root in time series data, whereas the alternative hypothesis states that the time series is stationary or trend stationary [2]. The ADF test formula is as follows:

$$ADF(t) = [Y(t) - Y(t-1)] = \alpha + \beta t + \gamma Y(t-1) + \delta * \Delta Y(t-1) + \varepsilon(t)$$

Where: Y(t) being the original time series at time t; α being to the constant in regression; β being the coefficient of time trend variable (t), capturing possible trend in data; γ being coefficient of the lagged value Y(t-1), which represents the degree of autocorrelation in data; δ being coefficient of the differenced time series ΔY(t - 1), capturing degree of mean reversion in data; ε(t) being the residual error of the model.

If the calculated p-value of the hypothesis test is greater than the critical value (coefficients are statistically significant), then the null hypothesis of having unit roots is rejected, indicating that the data is stationary [2]; vice versa.

2.2 Granger Causality Test

The Granger Causality test, developed by American economist Clive W.J. Granger, is closely tied to the concept of cause-and-effect [5]; however, the test differs in that it tests if a variable comes before another in the time series (ex. large amount of turkey sales Granger-cause thanksgiving). Because of the assumption of stationarity in Granger Causality test, conducting ADF test before exploring causality is critical [2]. For a pair of variables X and Y, the Granger Causality test formula is as follows [5]:

$$X(t) = \alpha + \sum(\beta i * X(t-i)) + \sum(\gamma i * Y(t-i)) + \varepsilon 1(t)$$
$$Y(t) = \alpha' + \sum(\delta i * X(t-i)) + \sum(\theta i * Y(t-i)) + \varepsilon 2(t)$$

Where: α and α' being the constants in respective regression equations; βi and γi being coefficients of the lagged values of X & Y up to designated lag order of i, capturing the linear relationship between the two variables' lagged values; $\varepsilon 1(t)$ and $\varepsilon 2(t)$ being the residual error.

The null hypothesis assumes that variable X does not Granger-cause Y [5]. If the coefficients βi is statistically significant in the regression of Y on its lagged values & the lagged values of X, it is implied that X Granger-causes Y [5].

3 Related Work

For the past several decades, a variety of forecasting models were built using mathematical and statistical theories, which includes ARIMA model, to better predict the volatile movements of the stock market and the economy itself. Ariyo et al. [1] used an ARIMA model with applied methods of ADF test and Granger Causality test to predict India's stock market through analysis of causal relationship between Bombay Stock Exchange and macroeconomic indicators. With predefined time-sires variables of 6 variables that have interrelation with American economy and Indian economy. The model is built with significance levels of 1%, 5%, and 10%. Ariyo et al. [1] conducted ADF test before analyzing data to ensure the stationarity of data. Using the data visualization of autocorrelation and lagged periods, F-statistic and probability of Granger Causality were calculated, yielding the result: movement of international market, such as the price change of S&P 500, closely affects Indian domestic stock market (Bombay stock market to be specific). However, the findings are only related to the 6 explored variables, and further research will need to be provided to explore the relationship between indicators and the movements of stock market volatility.

Tae Su Kim et al. [3] designed a causality model to examine the global warming factors in South Korea using self-organizing map and Granger Causality network. The variables used for the model were: CO_2 emission, Population in South Korea, Korean GDP value, and global temperature level. For analysis, Kim checked the stationarity of data (differencing values if not stationary), determined lag periods using partial autocorrelation function, and developed vector autoregression model to conduct Granger Causality test. The result of the test identified relationship between South Korea's GDP growth and CO_2 emission, which has a strong positive linear relationship with global warming index. Tae Su Kim et al. [3] suggested that although CO_2 emission and global temperature are proxies for climate change, unidirectional causality was identified from global temperature and CO_2 emission. However, there are limitations to this study as global warming anomalies have strong relationship with sea surface temperature, and the study only examined the economic conditions of Korea to compare it with global warming measures. Moreover, other chemicals could be accountable for further analysis, including global warming factors such as methane (CH_4), nitrous oxide (NO_2), halocarbons [3]. A detailed study is required to examine the true causal effect.

Mihir Dash [6] examines the relationships between the Indian stock market and the stock markets of Japan, China, Hong Kong, Korea, and other Southeast Asian countries using Augmented Dickey Fuller test and Granger Causality test. The time series data was collected from period of 2000- 2007 prior to the American subprime mortgage crisis in 2008. The findings of the study reveal that the Indian and Japanese stock markets are closely integrated with Southeast Asian markets; however, it was shown that Chinese stock market is relatively isolated from other markets, indicating China's strength in terms of economic stability. Ultimately, the study was successful at contributing to a better understanding of the interconnectedness of such stock markets, but further research is needed to explore the applicability of the Granger causality in financial forecasting.

4 Conclusions & Future Suggestions

Augmented Dickey Fuller test proposed by Mihir Dash [6] and Granger Causality test proposed by Tae Su Kim et al. [3] are the two of the most popular statistical methods to explore the relationship between two time series data and be able to forecast one using another. Augmented Dickey Fuller test proves to be an efficient method to confirm for stationarity of data. Granger Causality test can be usedas a tool for exploring potential causal relationships in time series. Ariyo et al. [1] utilized these methods to build ARIMA model to explore causal relationship between Bombay stock market and international stock markets. Ariyo et al. [1] conducted analysis using BSE sensex, BSE trading volume, Treasury Bill, S&P 500 return, Exchange Rate, and WPI. Mihir Dash [6] demonstrates a similar approach by exploring the relationship between the Indian stock market and Southeast Asian countries. On the other hand, Tae Su Kim et al. [3] applied stationary dataset to analyze Granger causality between global warming index and global warming factors. Tae Su Kim also used self-organizing map to visualize the causality network. I think the research work on this topic can be further expanded. All the methods mentioned have proven to be applicable in theory, but the research sample is limited to the chosen variables only. There are many economic and technical indicators that can be utilized within the analysis to better examine the causal relationships and build statistical models. In addition to ARIMA model, there are methods of modern LSTM models through programming languages and deep learning theories. Further efforts could be made to explore how different economic indicators can have causal relationship with the stock market, or how the accuracies of different statistical models can be compared to maximize the measured validity of stock-market forecasting.

5 Reference

[1] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on

Computer Modelling and Simulation, Cambridge, UK, 2014, pp. 106-112, doi: 10.1109/UKSim.2014.67.

[2] Tripathy, N. (2011). Causal relationship between macro-economic indicators and stock market. ResearchGate. Retrieved April 25, 2023, from https://www.researchgate.net/publication/265414493_Causal_Relationship_between_Macro_Economic_Indicators_and_Stock_Market_in_India.

[3] Thakur Dhakal, Tae-Su Kim, Do-Hun Lee, Gab-Sue Jang. (2023) Examining global warming factors using self-organizing map and Granger causality network: a case from South Korea. Ecological Processes 12:1.

[4] Hamilton, J.D. (1994). Time Series Analysis. Princeton University Press.

[5] Granger, C.W.J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37(3), 424-438.

[6] Mihir Dash, 2015. "A Study of Granger Causality in Asian Stock Markets," Journal of Applied Management and Investments, Department of Business Administration and Corporate Security, International Humanitarian University, vol. 4(3), pages 145-150.