**Bull Market or Bear Market: Recession Time Series Prediction for Q1 2024**

John Vincent Deniega, Ravita Kartawinata, Gabi Rivera

Master of Science in Applied Data Science, University of San Diego

***Data Preprocessing***

Data for the stock IDs were pulled from Yahoo Finance Library at 5-year span. The API ticker function retrieves comprehensive daily logs of the stock market movements and metadata from opening prices, high, low, and closing prices as well as volume. No missing data values were observed during initial time series plotting approach on market days. Missing values occurred on weekends and holiday dates. These missing values were then imputed and filled in by propagating the last valid value to the next. Then the time series datasets were converted to a specified frequency in order to be used and fed into time series exploratory tools and models. Data smoothing and differencing were rationalized during exploratory data analysis.

***Exploratory Data Analysis***

Determination of stationarity status was the first step of the platform process. Augmented Dickey-Fuller (ADF) test was employed and based of from the p-values greater than 0.05 significance level results for SPY and AMZN, the time series datasets were both determined not stationary. This is visualized in Figures 1a and 2a from the original time series plot. Figures 1b and 2b shows the next approach which was to subject the time series to (STL) Seasonal-Trend Decomposition using Locally Estimated Scatterplot Smoothing (LOESS) regression. SPY and AMZN’s trend and seasonal components were clearly parsed in both cases. During transformation to remove the trend and seasonal elements for model exploration, first degree differencing of the series was utilized. ADF tests of p-values lower than 0.05 significance level confirms that the time series were converted to stationary datasets. Figures 1c-d and 2c-d presents the autocorrelation (ACF) and partial correlation (PACF) plots of SPY and AMZN. For SPY, ACF suggest lags of [1, 2, 3, 8, 9, 11, 14, 21, 58, 63, 70] can be used to explore AR p parameter value and PACF suggest lags of [1, 2, 3, 8, 9, 11, 14, 21, 58, 63, 70] can be used for AR p parameter value. For AMZN, ACF suggest lags of [1,6,20,30] can be used for MA q parameter value and PACF suggest lags of [1,6,20,30,31] can be used for AR p parameter value.

As shown in Figure 3, additional EDA was performed for AMZN with anomaly detection using STL regression data. The residual was taken and subjected to ±3 standard deviation threshold to reveal anomalous spike and dips from the original stock curve. It is noticeable that two red marks were detected during the 2020 recession when the overall stock market was crashing. Amazon did well during that recession period compared to majority of the market movers. Year 2022 seems to be having increased anomalies up to the most recent stock decline.

**Figure 1**

*SPY Exploratory Data Analysis Plots*

A graph showing a line graph

Description automatically generatedA graph of a graph

Description automatically generated with medium confidence A graph with blue dots

Description automatically generated A graph with blue dots

Description automatically generated

*Note.* (A. Top left) Time Series Plot of SPY 5-year Closing Price. (B. Top right) STL decomposision of SPY time series using LOESS technique. (C. Bottom left) SPY ACF plot. (D. Bottom right) SPY PACF plot.

**Figure 2**

*AMZN Exploratory Data Analysis Plots*

A graph of a stock market

Description automatically generatedA line graph of different types of data

Description automatically generated with medium confidence

A graph with blue dots

Description automatically generated A graph with blue dots and numbers

Description automatically generated

*Note.* (A. Top left) Time Series Plot of AMZN 5-year Closing Price. (B. Top right) AMZN decomposision of AMZN time series using LOESS technique. (C. Bottom left) AMZN ACF plot. (D. Bottom right) AMZN PACF plot.

**Figure 3**

*AMZN Exploratory Data Analysis Plots: Anomaly Detection*

A graph showing a wave of data

Description automatically generated with medium confidenceA graph of a graph showing the growth of the stock market

Description automatically generated

*Note.* (A. Left) Residual from STL regression between ±3 standard deviation threshold. (B. Right) AMZN closing price curve with marked anomalies.