CA2 – Advanced Financial Technology

Cameron McEwan – cm17521@bristol.ac.uk

1. (5 points) Show a chart to visualize the closed stock price from 04/Jan/2021 to 30/Dec/2021 and write some observations based on this chart.

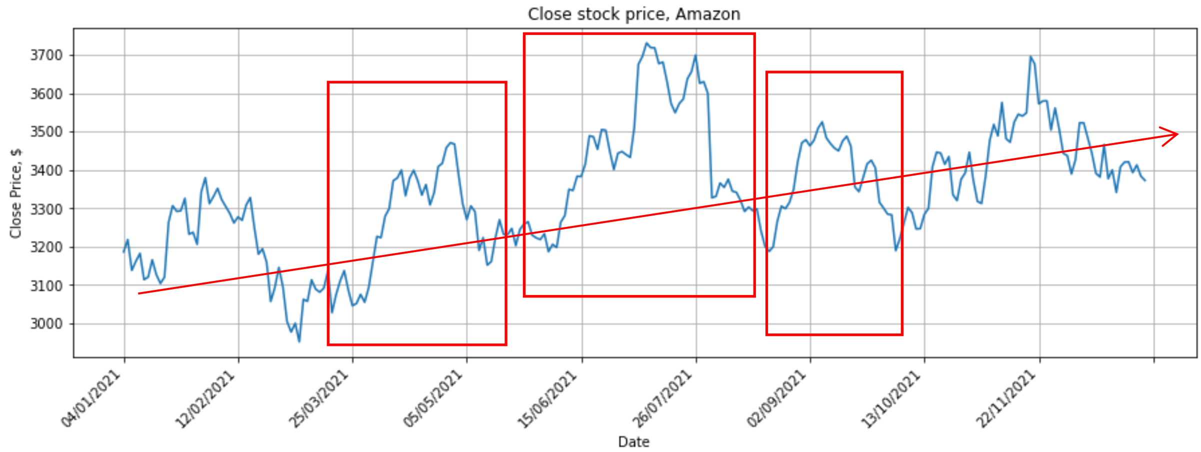


Figure 1: Line graph showing the time series closed stock price of Amazon (AMZN) in the period between 04/01/2021 and 30/12/2021.

By observing Figure 1, you can see that the closed AMZN stock price experiences several trends and patterns. Firstly, the graph experiences peaks and dips in a cyclical pattern. The periods of these cycles appear to be similar but not equal, approximately 2-3 months. Secondly, the chart shows an overall upward trend in the close stock price.

2. (5 points) What are the characteristics of a stationary time series? Is this time series data in AMZN.csv likely to be stationary and why? Check it with the naked eye.

In a stationary time series, the statistical properties (mean, variance, covariance) of the process generating the time series don't change over time. That is, for a stock price, the price at any time is random and can be different from the previous year, but the mean, variance and covariance remain the same for any time period. The price is independent of time in a stationary time series.

The AMZN close price time series is likely to not be stationary because the stock prices often experience various patterns and trends which mean that the price is dependent on time. By observing the graph (as discussed in Q1), the AMZN time series appears to show an overall upward trend, which suggests a non-stationary mean, as well as a cyclical pattern, which suggests a non-stationary variance/covariance.

3. (5 points) Summarise the process of calculating autocorrelation function and partial autocorrelation function with mathematical formulations.

The autocorrelation function (ACF) is the coefficient of correlation between two values in a time series. ACF is the correlation of a signal with a lagged copy of itself as a function of the lag, k. First calculate the sample mean, and then calculate the autocorrelation. For a time series, y_t, the ACF is given by:

$$r_k = \frac{\sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

where r_k is the autocorrelation for lag, k. A lag 1 autocorrelation is the correlation between values that are one time period apart. These correlations comprise of both the direct and indirect correlations. Finally estimate the variance
The partial autocorrelation function (PACF) seeks to remove these indirect correlations for selecting appropriate order of time-series statistical models (e.g., AR). Whilst the ACF finds correlations with lagged results, the PACF finds the correlations of the residuals.

The PACF is estimated via a set of OLS regressions [1]:

$$y_{t,j} = \phi_{j,1} y_{t-1} + \phi_{j,2} y_{t-2} + \ldots + \phi_{j,j} y_{t-j} + \epsilon_t$$

Where $\phi j,j$ is estimated via OLS with the standard $\beta = (X'X)-1X'Y$ coefficients.

So, for example, if you would like the first order PACF:
$y_{t,1} = \phi_{1,1}\, y_{t-1} + \epsilon t$
and the coefficient you want is the $\phi_{1,1}$ given by OLS: $\phi_{1,1} = Cov(y_t - 1, y_t) / Var(y_t)$ (and assuming weak stationarity).
The second order PACF would be the $\phi_{2,2}$ coefficient of:
$y_{t,2} = \phi_{2,1}\, y_t - 1 + \phi_{2,2} y_t - 2 + \epsilon_t$

4. (10 points) Plot the predicted and true stock price with the autoregressive (AR) model on a test set and describe your observations from this plot. Requirements: 1) You need to list key steps and their results (e.g., PACF plot) to predict the stock price. You will lose marks if you only show the final plot about prediction; 2) You need to split this dataset into a train (70%) set and a test (30%) set. Use the train set to train an autoregressive (AR) model and make a prediction on the test set; 3) You need to implement the AR model from scratch rather than using the existing Python library.

Import necessary libraries. Read AMZN.csv file into a dataframe. Using the function from Week 16 lab for the autoregressive model: Set test size to 0.7 and split into train and test

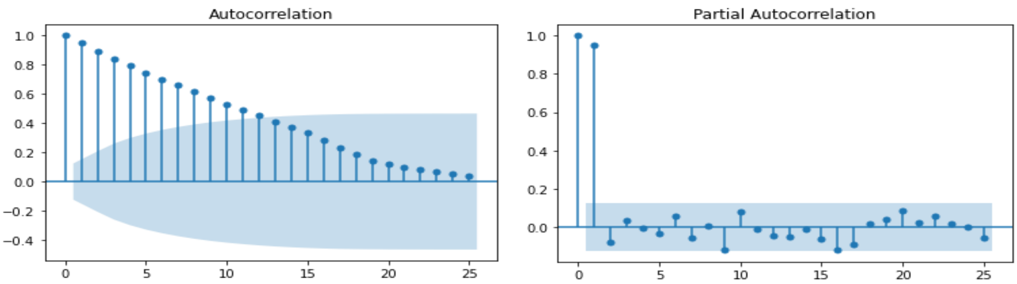
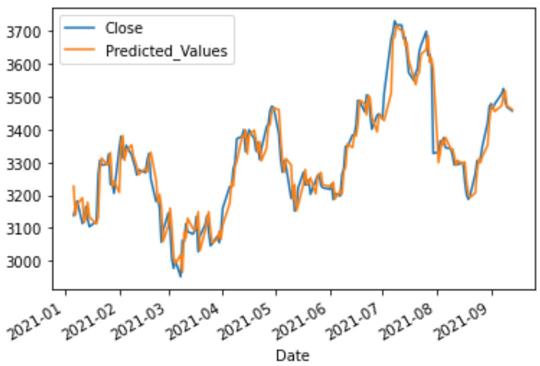


Figure 2: Plot of the autocorrelation function (ACF) (LEFT); and plot of the partial autocorrelation function (PACF) (RIGHT) on the AMZN.csv time series. On the PACF plot, values with a correlation far from zero show significant correlation between residuals for the lag value.

sets. Plot the partial autocorrelation function (PACF) for the time series, to decide which lag value to use in the AR model.

Fit the AR model by calling the function defined in Week 16 lab, which splits the data and returns the value of the RMSE on the train and test set. From observing the PACF plot, lag values of 1, 2, 3, 4 and 10 were all tested for comparison and **lag = 3** gave the lowest **RMSE of 51.6567** (4 d.p.).



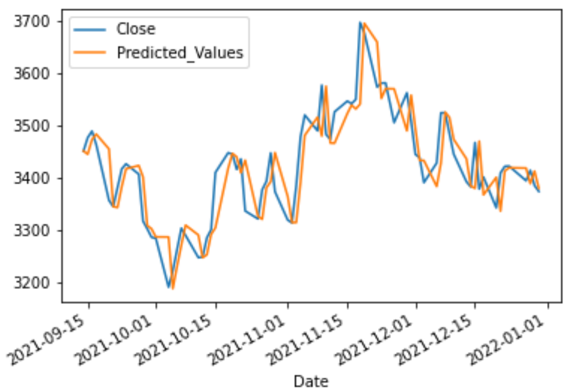


Figure 3: Plot showing the actual Close values between 04/01/21 and 30/12/21 and the predicted values for the train set (LEFT) and the test set (RIGHT).

5. (5 points) What are the strengths and weaknesses of the AR model? Are there some ways to overcome such weaknesses?

Autoregressive models can be used to tell if there is a lack of randomness in the time series. Not much information is required and it can also be used very effectively to forecast any recurring patterns in the data, however it relies on the fact that historical patterns will continue to repeat in the future. Predictions can be made at a range of time periods. Limitations include the fact that the autocorrelation function should not be less than 0.5 else the results are said to be less accurate. When a dataset is significantly affected by social aspects then an AR model is not suitable – a vector AR model would be more suitable because it used to predict multiple time series variables [2].
Autoregressive models cannot model distributions whose next-symbol probability is hard to compute and better alternatives for more complex time series distributions would be energy-based models (which give up efficient sampling) and latent-variable autoregressive models (which give up efficient scoring of a given string) [3].

Bibliography

[1] https://stats.stackexchange.com/questions/493807/how-to-calculate-the-acf-and-pacf-for-time-series

[2] https://thebusinessprofessor.com/en_US/research-analysis-decision-science/autoregression-definition

[3] Lin, C. J. (2020). *Limitations of autoregressive models and their alternatives.*