Advanced Financial Technology CA2

Jiageng Ding
Faculty of Engineering
University of Bristol
Bristol BS8 1UB, U.K.
lt22041@bristol.ac.uk

I. Data Overview

A. Data Set and Visualization

This experiment uses the close stock price of Microsoft from 31/Dec/2021 to 30/Dec/2022 as the data set.

Figure 1 demonstrates the overall downward trend in Microsoft's closing price over a year, with a drop of about \$100. However, in March, July and November there were brief increases lasting around 30 days respectively, with each increase around \$30.

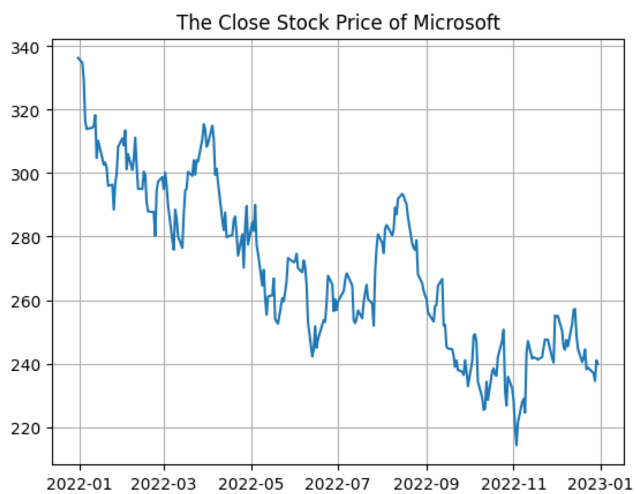

Fig. 1. Close stock price of Microsoft from 31/Dec/2021 to 30/Dec/2022.

B. Stationary Time Series

Stationary process is a stochastic process in which the joint probability distribution does not change with time, and it usually has the following significant characteristics.

- Means do not change over time, i.e., there is no overall upward or downward trend.
- Variances do not change over time, i.e., the magnitude of data change is essentially the uniform.
- The covariance is only related to the time lag and independent of time.
- No seasonal variation.

It can be seen from Figure 1 that this data set is not stationary, as it show an overall decreasing trend and no stable amplitude.

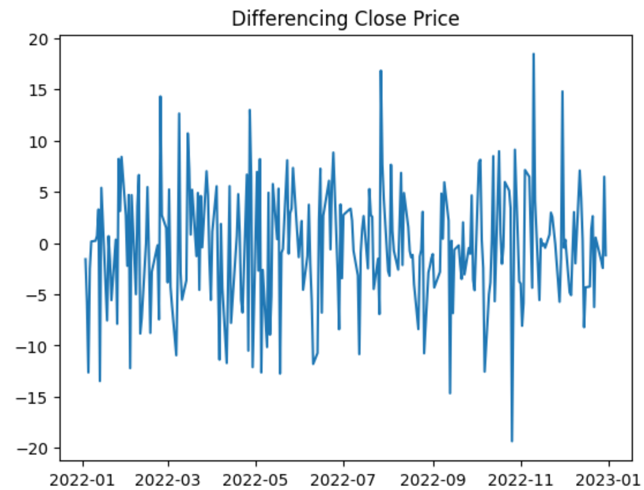

Fig. 2. Differenced close stock price of Microsoft from 31/Dec/2021 to 30/Dec/2022.

After performing one differencing, it can be seen from Figure 2 that the mean and variance are relatively stable. That means after one differencing this dataset could be a stationary process.

II. Experiment Design

Firstly, we need to determine whether the time series is white noise or not. The autocorrelation coefficients of white noise are smaller in different lags, i.e., they are located in the shaded part of the ACF plot. From left plot in figure 4, we can see that more than 5% of the autocorrelation coefficients lie outside the shaded region, so this time series is not white noise.

On the other hand, Figure 3 shows the autocorrelation coefficients of the differenced time series. It can be seen that almost all of the autocorrelation coefficients are within the shaded part. Therefore, autoregressive predictions cannot be applied to this differenced series.

Secondly, the order (p) of the autoregressive model is determined by calculating the partial autocorrelation coefficient. It can be seen from Figure 4 that the value of PACF is larger when $lag = 1$ and the value of PACF is close to 0 when $lag \geq 1$ (located in the shading part). Therefore, we consider that $lag = 1$ is the most favorable for the current model, i.e., the AR(1) model is to be used.

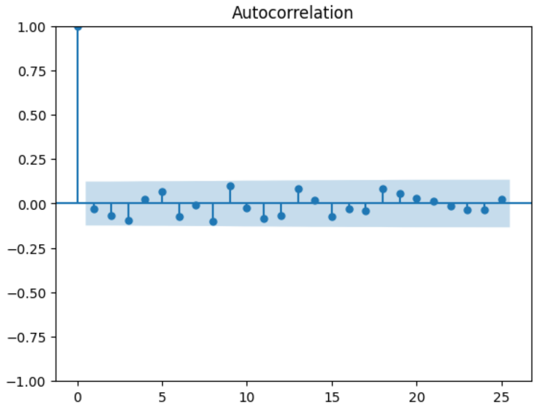

Fig. 3. Autocorrelation for differencing close price.

Finally, we split the date into train set which is used to train AR model and test set for testing prediction. The top 70% of the data from the data set are selected as the train set and the remaining 30% data are used as the test set. Then, linear regression was used on the training set, which resulted in the coefficient ϕ and intercept c in equation 1.

$$Y_t = c + \phi Y_{t-1} + e_t \tag{1}$$

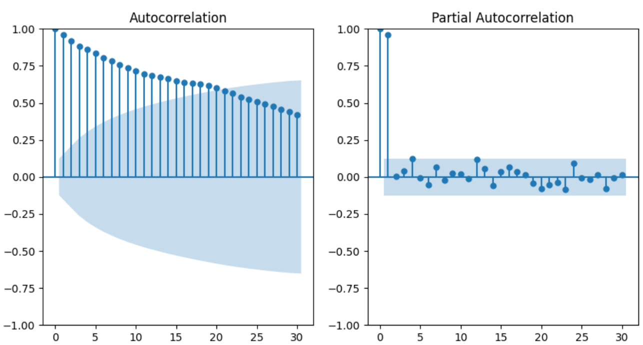

Fig. 4. Autocorrelation and Partial Autocorrelation from lag 0 to lag 30.

Then we can make some prediction with this model. Using the above parameters and equation 1 to predict the test set, the predicted and true values are shown in Figure 5. It can be seen that this model can predict the overall trend well.

However, figure 5 shows that the predicted values at the same time are higher than the true values. This is due to the overall decreasing trend of the data set and due to the fact that the test set is located in the lower interval. In addition, because of the lagging nature of the AR model, the predicted values appear later than the true time for the same values.

III. Conclusion

The AR model is based on the linear model, so it can handle large-scale data quickly and efficiently. In additional, the parameters obtained by the linear model have a strong realistic meaning.

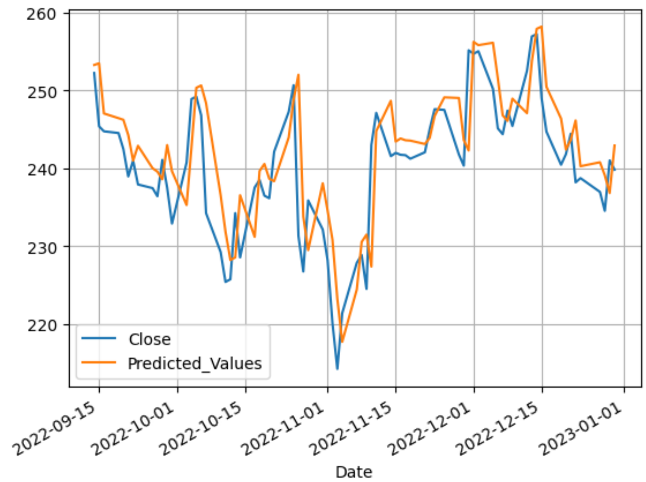

Fig. 5. Real price and predicted price in test set.

On the other hand, since the AR model predicts future data based on historical data, this leads to the fact that it can not handle well time series with upward or downward trends (non-stationary process) like those in this experiment. In addition when the sample size is small, the predicted values may be overfitted. [1]

The above drawbacks can be solved by narrowing the range of values through logarithmic operations and by differencing to be stationary. For random variation problems that cannot be handled by AR models, we can try to use moving average models for training. Similarly, for more complex data set, the nonlinear model can be used for training and predicting. [2]

For seasonal data, we can use the strategy of separating seasonal features from time series. [3] It is also known as the Warping Functional Autoregressive Model (see in [4]) proposed by Chen, Marron, and Zhang in 2019.

References

[1] C. M. Hurvich and C.-L. Tsai, "Regression and time series model selection in small samples," *Biometrika*, vol. 76, no. 2, pp. 297–307, 1989.

[2] J. G. De Gooijer and K. Kumar, "Some recent developments in non-linear time series modelling, testing, and forecasting," *International Journal of Forecasting*, vol. 8, no. 2, pp. 135–156, 1992. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016920709290115P

[3] J. G. De Gooijer and K. Kumar, "Some recent developments in non-linear time series modelling, testing, and forecasting," *International Journal of Forecasting*, vol. 8, no. 2, pp. 135–156, 1992.

[4] Y. Chen, J. Marron, and J. Zhang, "Modeling seasonality and serial dependence of electricity price curves with warping functional autoregressive dynamics," *The Annals of Applied Statistics*, vol. 13, no. 3, pp. 1590–1616, 2019.