ASSIGNMENT 1 REPORT

Sam Navdeep K
24b1303

1 Time Series

A time series is a sequence of data points collected successively in accordance with time in equal intervals of time. Examples include: Daily stock price, Monthly sales data, etc.

In time series the sequence of the data points matter and there is a correlation between the data points in general.

1.1 Components of a Time Series

A time series can be decomposed into four main components:

$$Y_t = T_t + S_t + C_t + R_t \tag{1}$$

- **Trend** (T_t): Long-term upward or downward movement
- **Seasonality** (S_t): Repeating patterns over a year in general.
- **Cyclic** (C_t): Long-term fluctuations without fixed frequency
- **Residual** (R_t): Random noise and flucuations in the model which canot be predicted

1.2 Stationarity

A time series is said to be **stationary** if:

- Mean is constant over time
- Variance is constant over time

It becomes crucial to make the model stationery in order for us to use the ARIMA model hence we use the logarithm function to make the variance constant and mean is made constant by using the differencing formula.

1.3 Differencing Formula

$$y'_t = y_t - y_{t-1} \tag{2}$$

y'_t is the differenced value at time t, y_t is the original value at time t, and y_{t-1} is the value at the previous time step. This operation measures the change between consecutive observations. hence we update the value of y_t with y'_t to get constant mean.

2 ARIMA Model

The AutoRegressive Integrated Moving Average (ARIMA) model is a widely used statistical method for time series forecasting.

2.1 Components of the ARIMA Model

The ARIMA model is composed of two fundamental stochastic models: the AutoRegressive (AR) model and the Moving Average (MA) model, combined with differencing to handle non-stationarity.

The AutoRegressive model of order p, denoted as AR(p), is defined as

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \epsilon_t \tag{3}$$

where y_t is the current value of the time series, ϕ_i are the autoregressive coefficients, and ϵ_t is an error term at time t.

The Moving Average model of order q, denoted as MA(q), is given by

$$y_t = c + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \tag{4}$$

where θ_j are the moving average coefficients and ϵ_t represents errors at time t.

3 ACR and PACR plots

The Autocorrelation Function (ACF) measures the correlation between a time series and its lagged values. It helps identify the presence of moving average (MA) components in a time series model.

The Partial Autocorrelation Function (PACF) measures the correlation between y_t and y_{t-k} after removing the linear effects of all intermediate lags 1 to $k-1$. It is primarily used to identify the order of the autoregressive (AR) component.

A sharp cutoff in the ACF indicates an MA(q) process, while a sharp cutoff in the PACF indicates an AR(p) process.

4 Results and plots

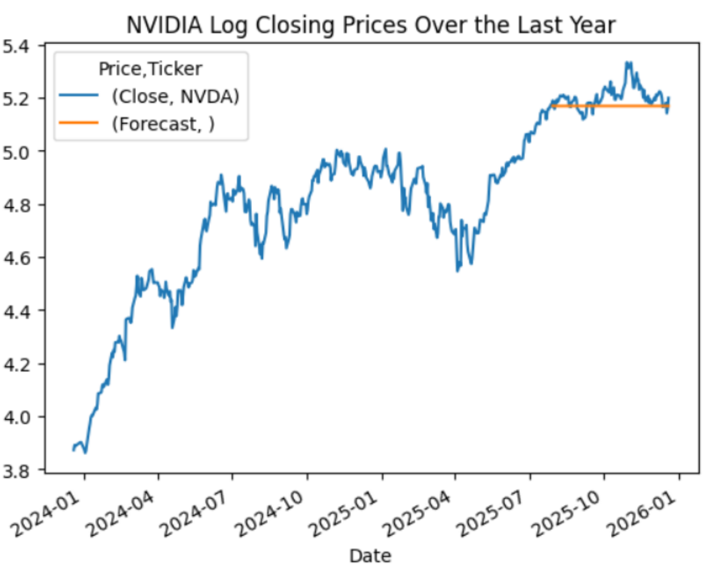

Figure 1: NVIDIA stock price predictions for 4 months

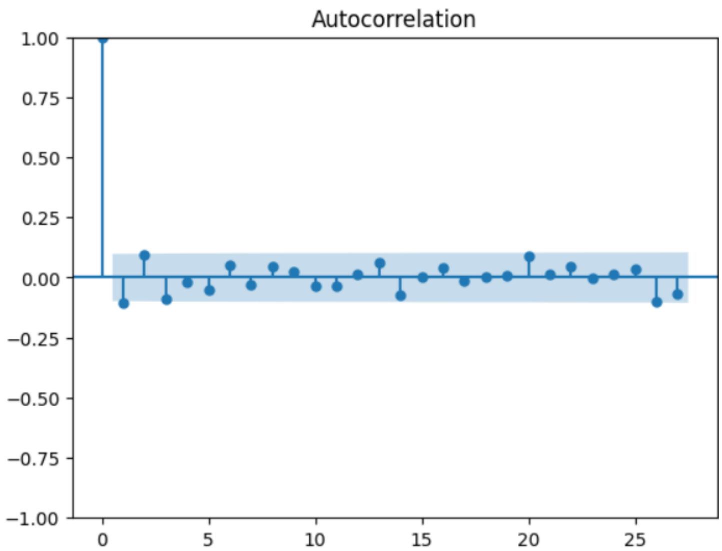

Figure 2: Autocorrelation

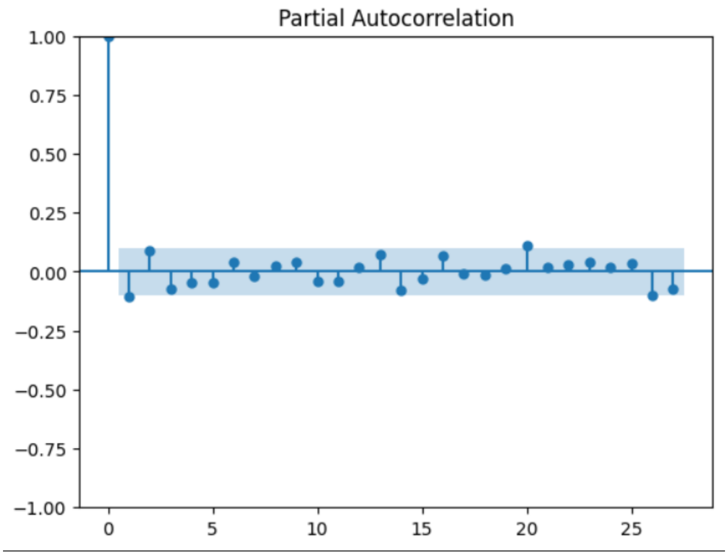

Figure 3: Partial Autocorrelation

```
######### MODEL EVAL ##########
from sklearn.metrics import mean_squared_error, mean_absolute_error, root_mean_squared_error
mse = mean_squared_error(x_test, forecast)
mae = mean_absolute_error(x_test, forecast)
rmse = root_mean_squared_error(x_test, forecast)
print(f'MSE: {mse}, MAE: {mae}, RMSE: {rmse}')
[114]
...  MSE: 0.0030883938151380842, MAE: 0.04230905988000829, RMSE: 0.0555733192740732 74

nrmse = rmse/(x_test.mean())
print(f'NRMSE: {nrmse}')
[115]
...  NRMSE: Ticker
     NVDA    0.010677
     dtype: float64
```

Figure 4: Evaluation metrics

```
Best model:  ARIMA(1,1,1)(0,0,0)[0] intercept
Total fit time: 9.528 seconds
                                SARIMAX Results
```

Figure 5: Best parameters for the model using auto.arima() function

from the plots it is clearly evident that 1,1,1 is the best sutable parameter as the second value is outside the confidence interval for bth ACF and PACF plots.

5 Conclusion

In this work, an ARIMA-based time series forecasting model was developed and evaluated on the test dataset. The performance of the model was assessed using standard regression error metrics, namely Mean Squared Error (MSE), Mean Absolute Error (MAE)and Root Mean Squared Error (RMSE).

The obtained results are as follows: MSE = 0.00309, MAE = 0.04231, RMSE = 0.05557, and NRMSE = 0.01068. The relatively low values of MSE and RMSE indicate that the forecasted values closely follow the actual observations.

.