Forecasting Time Series Data **STAT-GB 6018**

Forecasting the Stock Price of Microsoft Corporation, Inc.

Future Liang Fall 2023

Contents

1 Introduction

As Jeremy Siegel said, "volatility scares enough people out of the market to generate superior returns for those who stay in." As a result, in this project 1, I will delve into the volatile stock market and try to forecast the stock price of Microsoft Corporation, Inc. (henceforth Microsoft).

The detailed information of this project is summarized in Table 1 below:

Table 1: Project Key Information

Series	Microsoft's Stock Price (Adjusted Close)
Time	October 15, 2022 - October 15, 2023
Frequency	Daily
Observations	250 Trading Days
Method	ARIMA
Model Selection Criteria	Minimization of AIC_c
Forecast Lead	1 Month (30 Trading Days)
Data Source	Yahoo Finance

The project will start by analyzing the potential fitted ARIMA model using the ACF and PACF.

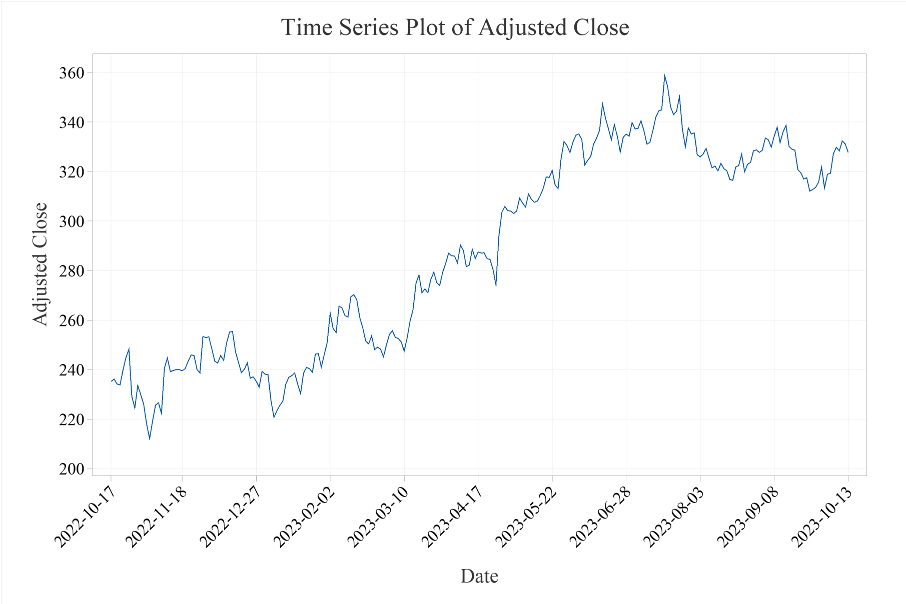
Then, I will do a detailed analysis to find the fitted ARIMA model through the minimization of AIC_c.

Afterwards, I will provide the estimates of the parameters and examine the adequacy of the fitted ARIMA model through the Ljung-Box test as well as the residual plot analysis.

Finally, I will print out the forecasts, forecast intervals, and related graphs with brief comments on feasibility of the forecast.

2 Preliminary Analysis: ACF & PACF

The time series of Adjusted Close of Microsoft is plotted below:

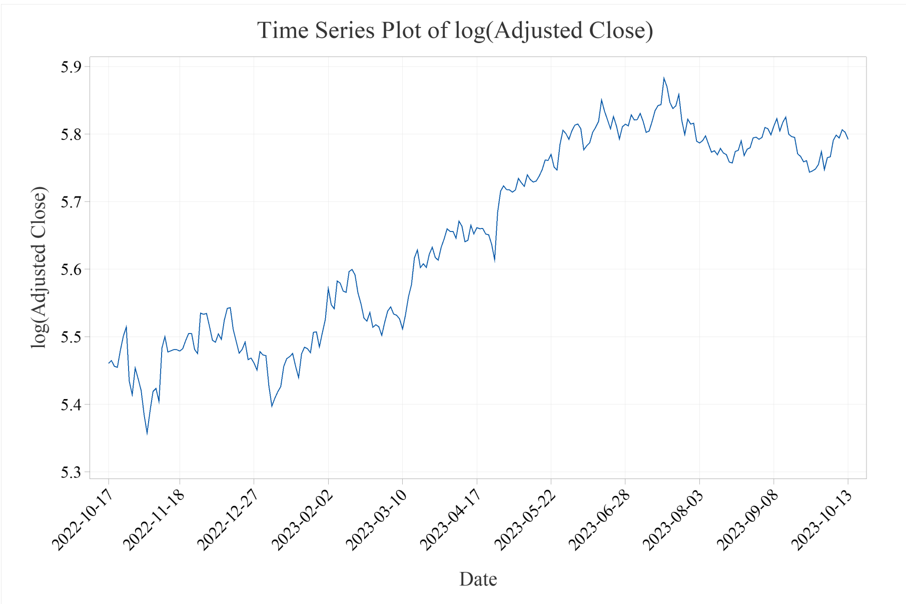


From the graph above, I can see:

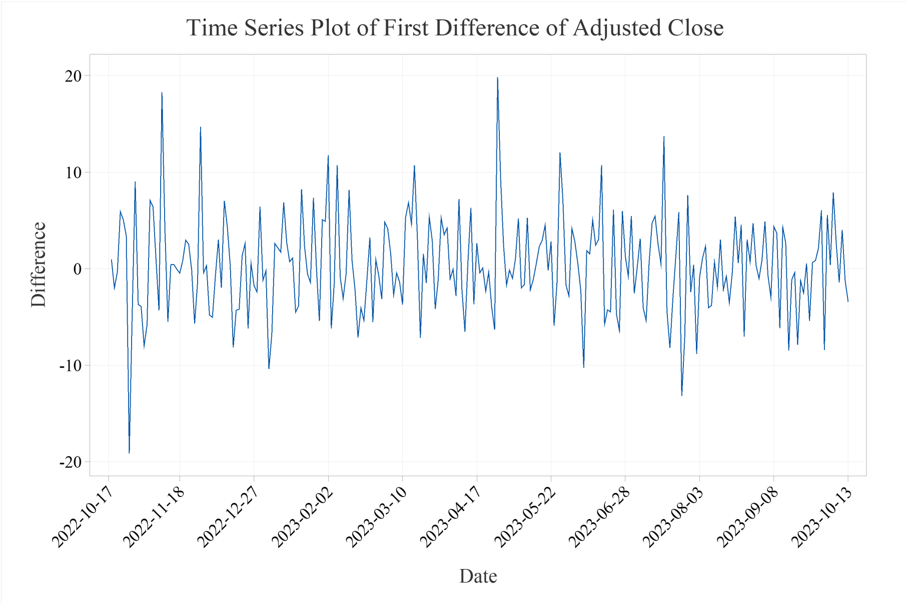
Trend There is a relatively evident upward trend in the series, especially from the beginning until around May 2023, indicating a general increase in stock prices over that time period.

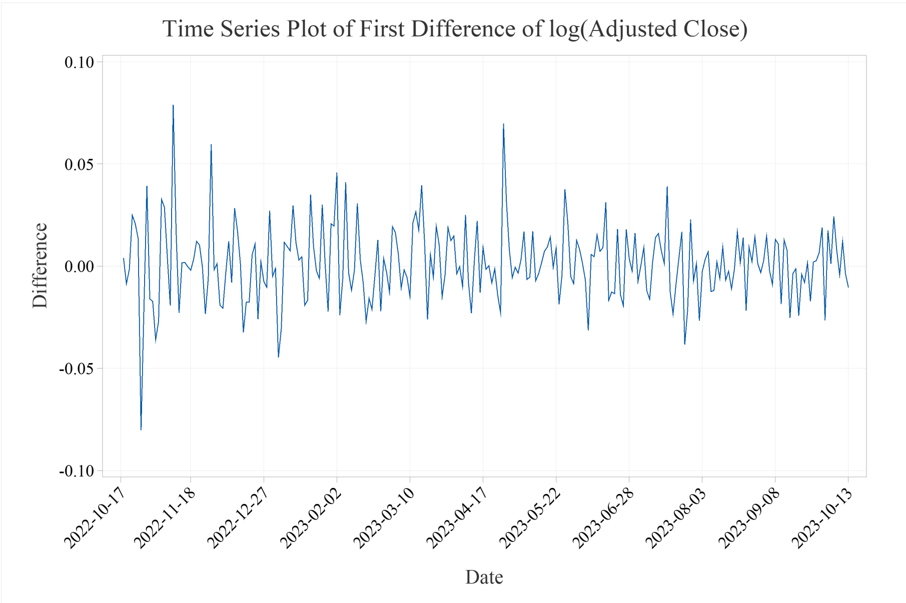
Volatility While the overall trend is upward, there are periods of increased volatility, most notably around the middle of 2023, where I can see sharp peaks and troughs.

The time series of log(Adjusted Close) of Microsoft is plotted below:



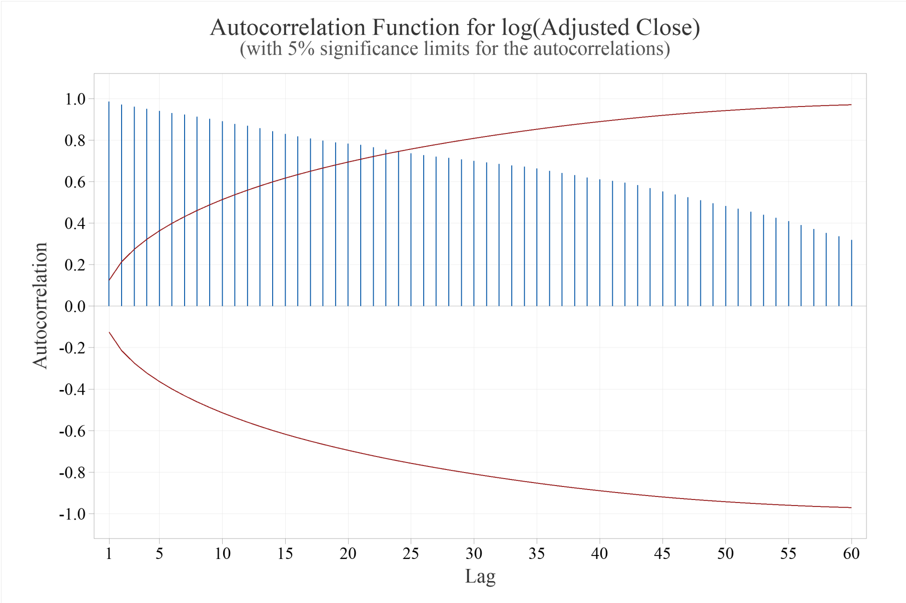
To assess if it is necessary to take log, the difference of Adjusted Close and log(Adjusted Close) of Microsoft are plotted:

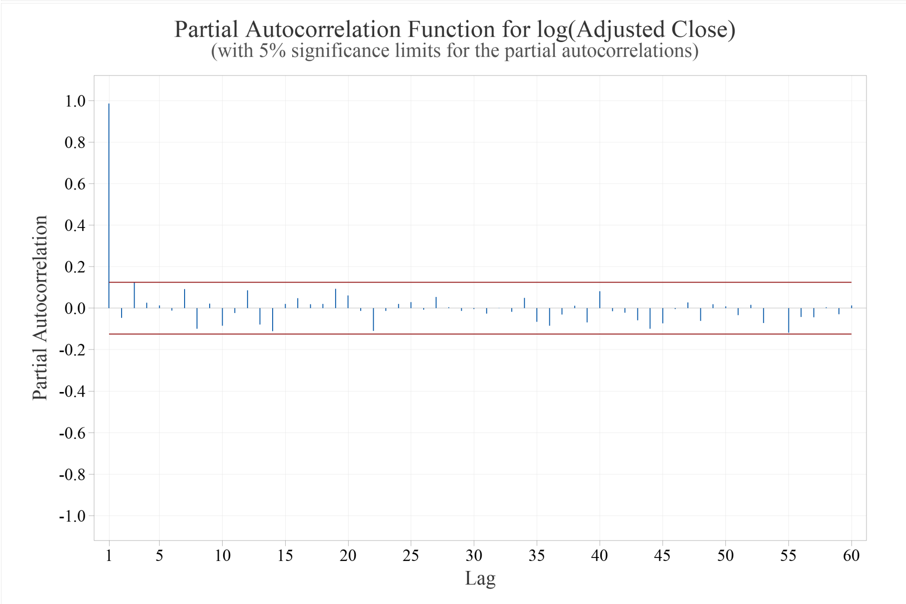




The graphs above show extremely little level-dependent volatility, but I still decide to take log. After all, taking log would not harm.

To fit a potential ARIMA model, I will start by plotting the ACF and PACF.
The ACF and PACF of log(Adjusted Close) of Microsoft is plotted below:

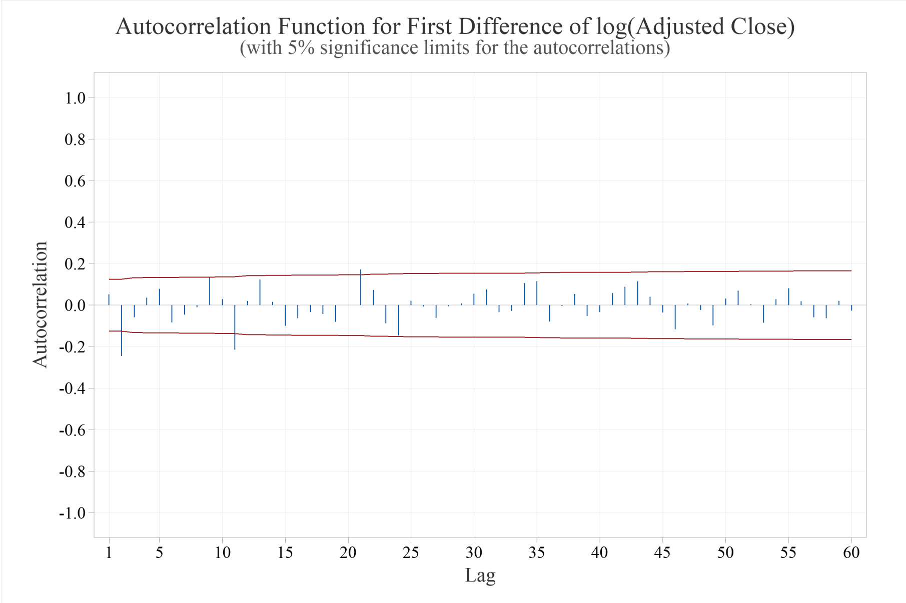


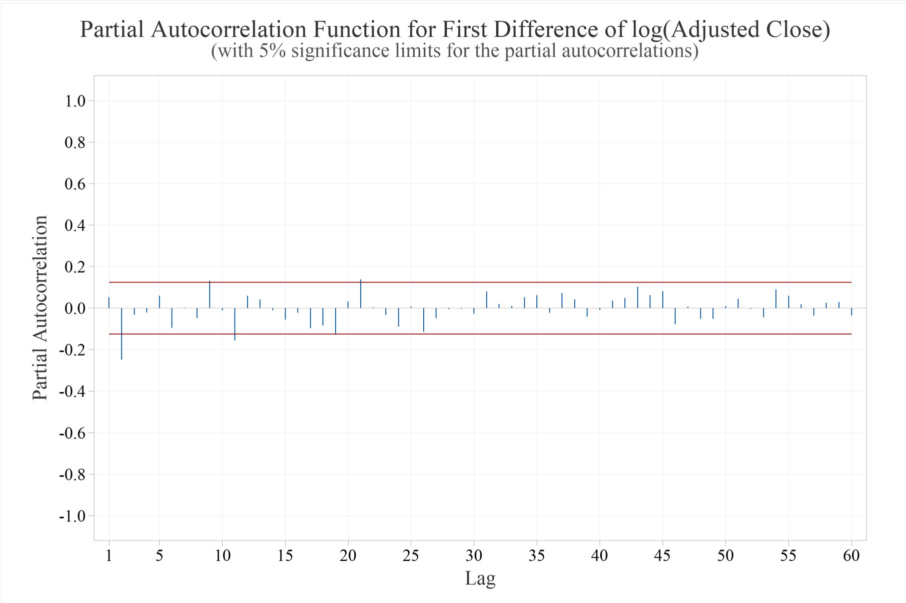


The ACF dies down very slowly, and the PACF has a statistically significant cutoff at lag 1. This indicates that the log(Adjusted Close) is non-stationary, so differencing is needed.

CONTINUE ON THE NEXT PAGE.

The ACF and PACF of the first difference of $\log$(Adjusted Close) of Microsoft is plotted below:





Note that this is the first difference of $\log$(Adjusted Close) of Microsoft, so the d in ARIMA(p,d,q) should be 1.

The ACF suggests the ARIMA(0,1,2) since the ACF has a statistically significant cutoff at lag 2.

The PACF suggests the ARIMA(2,1,0) since the PACF has a statistically significant cutoff at lag 2.

Therefore, based on the ACF and PACF, picking either ARIMA(0,1,2) or ARIMA(2,1,0) might be reasonable, with consideration of parsimony.

3 Technical Analysis: Minimization of AIC_c

The Table 2 below provides the AIC_c of ARIMA(p,d,q) for log(Adjusted Close) of Microsoft **without constant** based on the formula:

$$AIC_c = N\log(\frac{SS}{N}) + 2(p+q+1)\frac{N}{N-p-q-2}$$

Table 2: Model Trial Without Constant

Model	p	d	q	N	SS	AIC_c
1	0	1	0	249	0.0855689	-1983.979428
2	0	1	1	249	0.0850297	-1983.520843
3	0	1	2	249	0.0802580	-1995.852578
4	0	1	3	249	0.0802442	-1993.829244
5	0	1	4	249	0.0802396	-1991.760597
6	0	1	5	249	0.0796153	-1991.605234
7	1	1	0	249	0.0852841	-1982.777175
8	1	1	1	249	0.0825422	-1988.864795
9	1	1	2	249	0.0802440	-1993.830073
10	1	1	3	249	0.0802439	-1991.747318
11	1	1	4	249	0.0799163	-1990.665831
12	1	1	5	249	*	*
13	2	1	0	249	0.0802482	-1995.883015
14	2	1	1	249	0.0802026	-1993.958358
15	2	1	2	249	0.0802010	-1991.880390
16	2	1	3	249	0.0801415	-1989.965014
17	2	1	4	249	0.0784713	-1993.091459
18	2	1	5	249	*	*
19	3	1	0	249	0.0802041	-1993.953971
20	3	1	1	249	0.0802026	-1991.875631
21	3	1	2	249	0.0797492	-1991.186937
22	3	1	3	249	0.0802015	-1987.660982
23	3	1	4	249	0.0789605	-1989.408892
24	3	1	5	249	0.0763741	-1995.548572

Continued on next page

Table 2: Model Trial Without Constant (Continued)

Model	p	d	q	N	SS	AIC_c
25	4	1	0	249	0.0801947	-1991.900018
26	4	1	1	249	0.0801963	-1989.794955
27	4	1	2	249	0.0765714	-1999.194311
28	4	1	3	249	*	*
29	4	1	4	249	0.0759855	-1996.818713
30	4	1	5	249	0.0797507	-1982.605079
31	5	1	0	249	0.0798211	-1990.962408
32	5	1	1	249	0.0791634	-1990.905253
33	5	1	2	249	0.0785163	-1990.813424
34	5	1	3	249	0.0785135	-1988.669298
35	5	1	4	249	0.0758373	-1995.133629
36	5	1	5	249	0.0758253	-1992.983185
					Lowest	-1999.194311

Note: `Minitab` refuses to fit the model with asterisk (*).

Based on the table without constant above, ARIMA(4,1,2) has the lowest AIC_c value, so it is reasonable to choose ARIMA(4,1,2) without constant in this case.

CONTINUE ON THE NEXT PAGE.

The Table 3 below provides the AIC_c of ARIMA(p,d,q) for log(Adjusted Close) of Microsoft **with constant** based on the formula:

$$AIC_c = N\log(\frac{SS}{N}) + 2(p+q+2)\frac{N}{N-p-q-3}$$

Table 3: Model Trial With Constant

Model	p	d	q	N	SS	AIC_c
1	0	1	0	249	0.0851276	-1983.234503
2	0	1	1	249	0.0846695	-1982.528906
3	0	1	2	249	0.0795685	-1995.935080
4	0	1	3	249	0.0795314	-1993.968247
5	0	1	4	249	0.0795123	-1991.927629
6	0	1	5	249	0.0789926	-1991.443062
7	1	1	0	249	0.0848912	-1981.877633
8	1	1	1	249	0.0821738	-1987.912523
9	1	1	2	249	0.0795288	-1993.976300
10	1	1	3	249	0.0795288	-1991.876111
11	1	1	4	249	0.0791710	-1990.881326
12	1	1	5	249	*	*
13	2	1	0	249	0.0796205	-1995.772226
14	2	1	1	249	0.0795331	-1993.962698
15	2	1	2	249	0.0795270	-1991.881654
16	2	1	3	249	0.0794929	-1989.870904
17	2	1	4	249	0.0777534	-1993.244861
18	2	1	5	249	0.0786073	-1988.372141
19	3	1	0	249	0.0795403	-1993.940378
20	3	1	1	249	0.0790694	-1993.318621
21	3	1	2	249	0.0790686	-1991.203465
22	3	1	3	249	0.0789199	-1989.536919
23	3	1	4	249	*	*
24	3	1	5	249	0.0756110	-1995.877649
25	4	1	0	249	0.0795114	-1991.930617
26	4	1	1	249	*	*

Continued on next page

Table 3: Model Trial With Constant (Continued)

Model	p	d	q	N	SS	AIC_c
27	4	1	2	249	0.0764345	-1997.504716
28	4	1	3	249	0.0786974	-1988.086928
29	4	1	4	249	0.0752165	-1997.180339
30	4	1	5	249	0.0749856	-1995.756310
31	5	1	0	249	0.0792165	-1990.738149
32	5	1	1	249	0.0785494	-1990.708577
33	5	1	2	249	0.0779397	-1990.495884
34	5	1	3	249	0.0778858	-1988.496815
35	5	1	4	249	0.0750369	-1995.585862
36	5	1	5	249	*	*
					Lowest	-1997.504716

Note: `Minitab` refuses to fit the model with asterisk (*).

Based on the table with constant above, ARIMA(4,1,2) has the lowest AIC_c value, so it is reasonable to choose ARIMA(4,1,2) with constant in this case.

From Table 2 and Table 3 above, it is reasonable to choose ARIMA(4,1,2) without constant as it has the lowest AIC_c value.
Not surprisingly, `Minitab` also selects ARIMA(4,1,2) without constant as the best fitted model although it uses different metrics to calculate AIC_c.

4 Estimates of Parameters & Examination of Model Adequacy

The estimates of the parameters of ARIMA(4,1,2) without constant are provided by `Minitab` in Table 4 as follow:

Table 4: Parameter Estimates of ARIMA(4,1,2) Without Constant

Type	**Coef**	**SE Coef**	***t*-Value**	***p*-Value**
AR 1	0.1588	0.0702	2.26	0.025
AR 2	-1.1560	0.0643	-17.99	0.000
AR 3	0.0269	0.0665	0.40	0.686
AR 4	-0.1753	0.0645	-2.72	0.007
MA 1	0.1113	0.0291	3.83	0.000
MA 2	-0.96376	0.00585	-164.60	0.000

Assume x_t represents the $\log$(Adjusted Close) of Microsoft.
Set $y_t = x_t - x_{t-1}$.
Then, the complete form of the ARIMA(4,1,2) model without constant can be written as:

$$y_t = 0.1588y_{t-1} - 1.1560y_{t-2} + 0.0269y_{t-3} - 0.1753y_{t-4} + \varepsilon_t - 0.1113\varepsilon_{t-1} + 0.96376\varepsilon_{t-2}$$

The Ljung-Box statistics and related hypothesis test of ARIMA(4,1,2) model without constant are provided by Minitab in Table 5 as follow:

Table 5: Ljung-Box Test Statistics for ARIMA(4,1,2) Without Constant

Lag	12	24	36	48
Chi-Square	9.84	24.56	37.05	57.67
DF	6	18	30	42
***p*-Value**	0.132	0.138	0.176	0.054

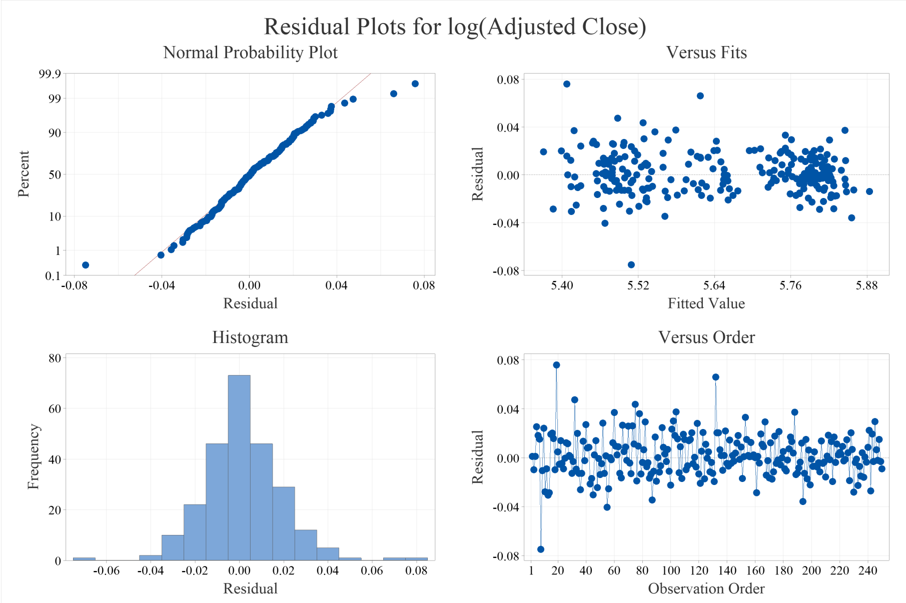
Hypothesis:

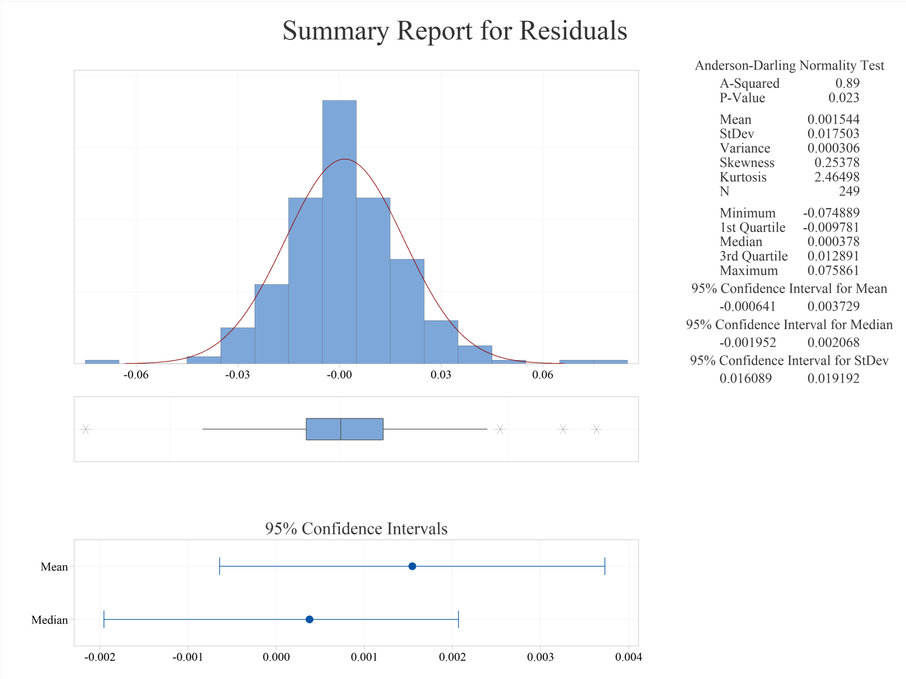
$$\begin{cases} H_0 : \text{The residuals from the model show no autocorrelation, indicating the model is adequate.} \\ H_a : \text{The residuals from the model show autocorrelation, indicating the model is inadequate.} \end{cases}$$

Since the p-values from the Table 5 above are all larger than 0.05, we fail to reject H_0 at 5% significance level.
Therefore, we can conclude that the ARIMA(4,1,2) without constant is an adequate model.

CONTINUE ON THE NEXT PAGE.

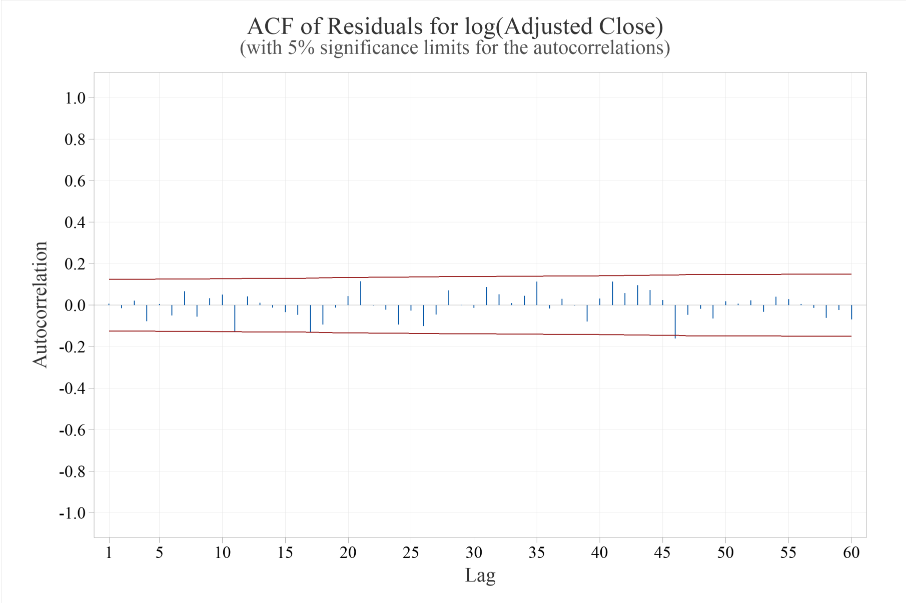
The plots for residuals of log(Adjusted Close) of Microsoft are plotted below:

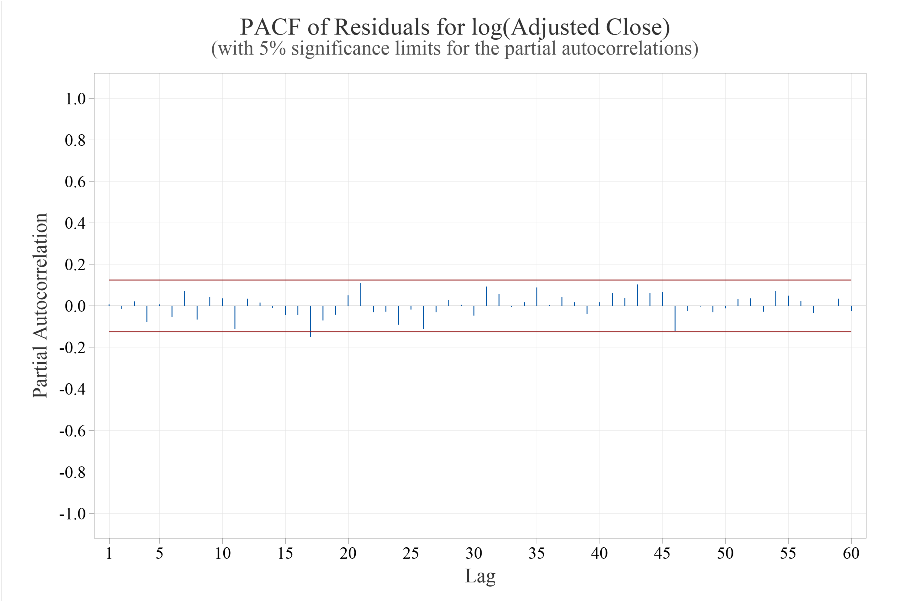




CONTINUE ON THE NEXT PAGE.

The ACF and PACF for residuals of $\log$(Adjusted Close) of Microsoft are plotted below:





The normal probability plot is slightly S-shaped.
The box plot under the histogram shows slight skewness in residual distribution.
The p-value of Anderson-Darling normality test is 0.023, which is lower than 0.05.
As a result, there are statistical evidence that the residuals are not normally distributed, indicating other models like ARIMA/GARCH might be useful in this scenario.

Additionally, the majority of lags remain in the significance bands in the ACF and PACF of the residuals, though there are few breakthroughs, indicating potential autocorrelation in residuals, but not particularly significant.

5 The 1-Month Forecast

The 1-month (30-trading-day) forecasts and 95% forecast intervals for log(Adjusted Close) of Microsoft are provided by `Minitab` in Table 6 as follow:

Table 6: The 1-Month Forecasts & Forecast Intervals for log(Adjusted Close) of Microsoft

Period	Forecast	95% Limits	
		Lower	Upper
251	5.79403	5.75923	5.82883
252	5.79547	5.74507	5.84586
253	5.79395	5.73529	5.85262
254	5.79393	5.72891	5.85894
255	5.79539	5.72416	5.86661
256	5.79536	5.71764	5.87308
257	5.79393	5.71016	5.87770
258	5.79378	5.70502	5.88255
259	5.79515	5.70186	5.88845
260	5.79551	5.69736	5.89366
261	5.79423	5.69117	5.89728
262	5.79367	5.68639	5.90096
263	5.79484	5.68378	5.90589
264	5.79556	5.68052	5.91060
265	5.79454	5.67530	5.91379
266	5.79367	5.67064	5.91670
267	5.79453	5.66818	5.92088
268	5.79552	5.66574	5.92529
269	5.79484	5.66137	5.92832
270	5.79376	5.65681	5.93072
271	5.79425	5.65426	5.93424
272	5.79538	5.65237	5.93840
273	5.79509	5.64876	5.94141
274	5.79394	5.64438	5.94350
275	5.79404	5.64163	5.94644

Continued on next page

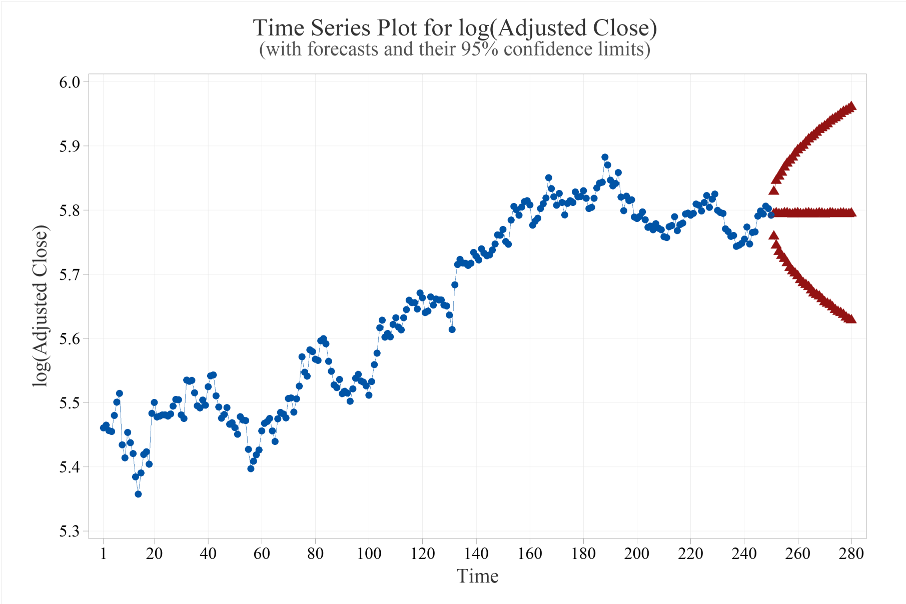
Table 6: The 1-Month Forecasts & Forecast Intervals for log(Adjusted Close) of Microsoft (Continued)

Period	**Forecast**	**95% Limits**	
		Lower	**Upper**
276	5.79518	5.64003	5.95033
277	5.79526	5.63712	5.95340
278	5.79416	5.63299	5.95533
279	5.79390	5.63003	5.95778
280	5.79494	5.62852	5.96135

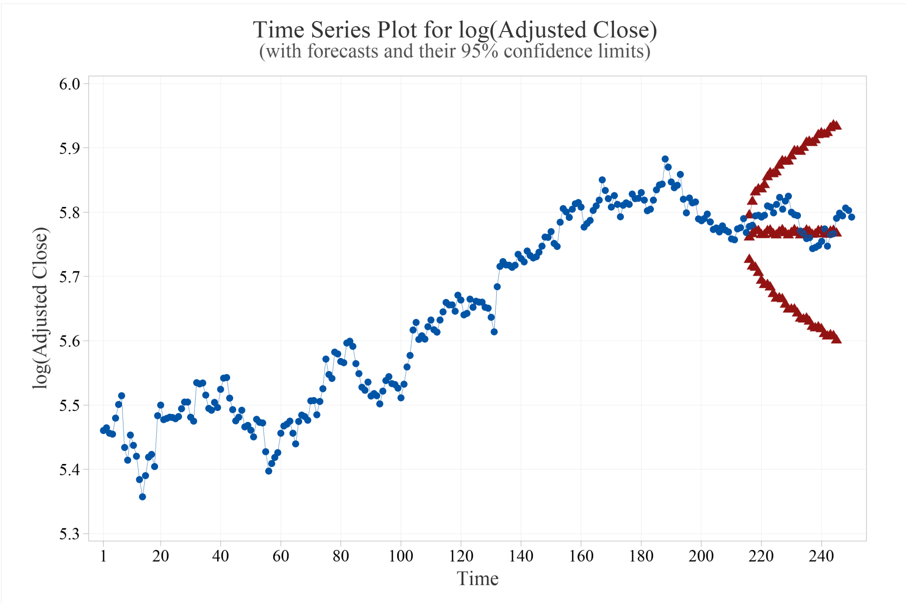
Caveat: For the forecast for Adjusted Close of Microsoft, do the transformation e^{Forecast}.

CONTINUE ON THE NEXT PAGE.

The 1-month (30-trading-day) forecasts and 95% forecast intervals for log(Adjusted Close) of Microsoft are plotted below:



To examine if the forecast interval is too wide, I also plot the backcast from time observation 215 below:



In the backcasting above, the 95% forecast interval captures all the forecasts.
Regarding the width of the forecast interval, I think it is reasonably wide (definitely not excessively wide), indicating some degree of precision of our 1-month (30-trading-day) forecasts for log(Adjusted Closed) of Microsoft.