Forecast Time Series Data Project1
Alison Yao (yy2564)

Data Description

This project uses the monthly stock price of Netflix from 2002-06-01 to 2022-03-01 (238 observations) for time series forecast. The 2022-03-01 data is the latest entry because I obtained the data in late March from [Yahoo Finance](). By the time this report is graded, there may or may not be data for 2022-04-01, but I am not able to obtain it at the time being.

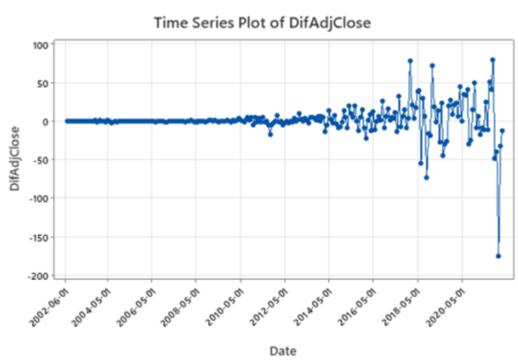
The head of the table is shown below:

NFLX_5y

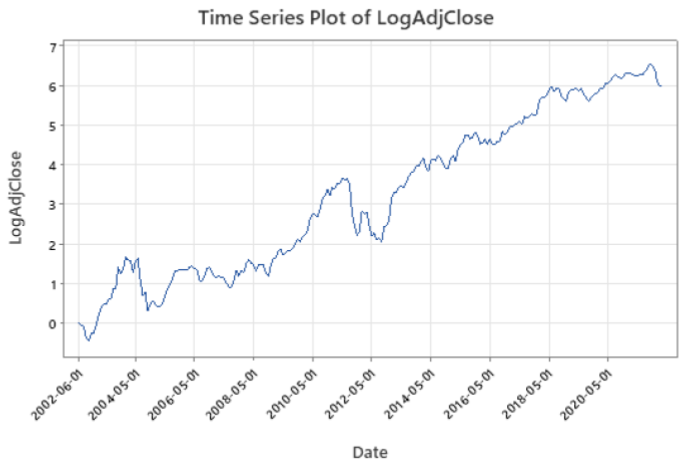
Date	Open	High	Low	Close	Adj Close	Volume
2017-04-01	146.699997	153.520004	138.660004	152.199997	152.199997	149779000
2017-05-01	151.910004	164.750000	151.610001	163.070007	163.070007	116795800
2017-06-01	163.520004	166.869995	147.300003	149.410004	149.410004	135699700
2017-07-01	149.800003	191.500000	144.250000	181.660004	181.660004	185144700
2017-08-01	182.490005	184.619995	164.229996	174.710007	174.710007	136534400

Time Series Plot

First, we plot the time series plot of adjusted closing stock price. The Adj Close stock price grows exponentially over time as a general trend. Surprisingly, while Adj Close did not drop during COVID time (probably because streaming was a favorable entertainment during lockdown), it is declining sharply in recent months, starting from 2021-10-01.



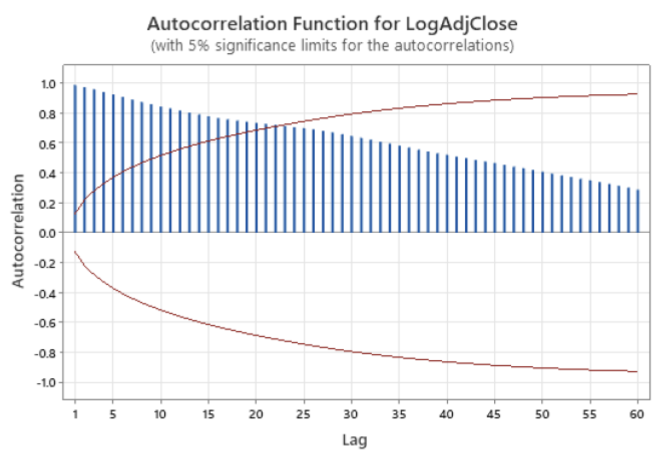
There is strong evidence of level-dependent volatility, as the stock price is relatively stable in early years but much more volatile in recent years, which is especially easy to see if we take the difference of Adj Close price. Therefore, we take the natural log of adjusted closing stock price and plot another time series plot.



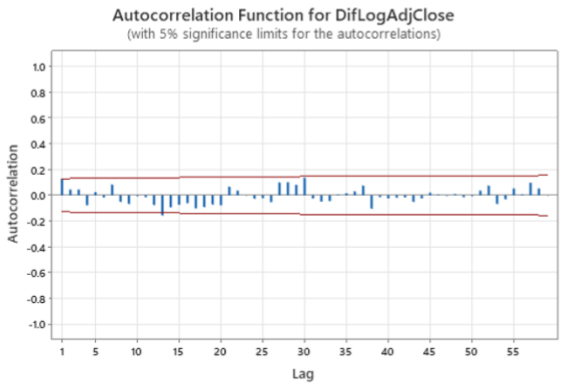
After taking log, we get rid of level-dependent volatility and linearize the exponential trend at the same time.

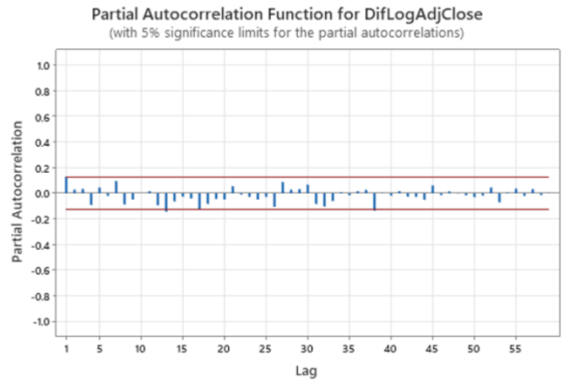
Model Selection

Now, for the first step of choosing an ARIMA model, we need to determine if we should difference the data and how many times should we difference the data if we need to. We can look at the ACF plot of LogAdjClose.

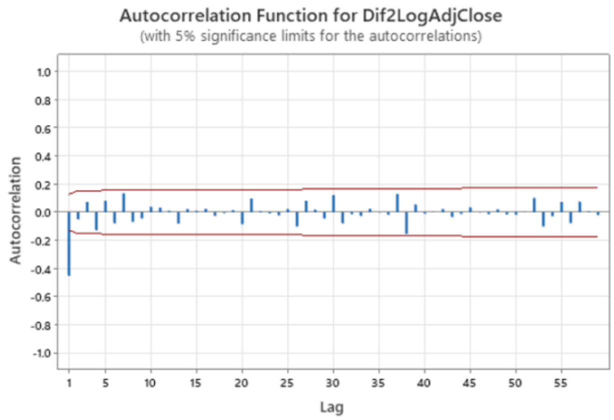


The hanging behavior is strong evidence that we should difference the data. Differencing once gives us:





Differencing twice shows strong evidence of over-differencing as lag-1 is significantly negative and close to -0.5.



Therefore, we conclude that we only need to difference once to make the data stationary, so we choose $d = 1$ in our ARIMA model.

As for the values of p and q, we can check the ACF and PACF of DifLogAdjClose, both of which are statistically significant at lag 1. Therefore, we choose p,q = 0,1 and test the following models:

We know N = n – d = 238 – 1 = 237.

Without Constant					With Constant				
p	d	q	SS	AICc	p	d	q	SS	AICc
0	1	0	6.10529	-865.14335	0	1	0	5.95598	-868.97718
1	1	0	5.96747	-868.52041	1	1	0	5.85750	-870.87692
0	1	1	5.97965	-868.03717	0	1	1	5.86326	-870.64398
1	1	1	5.94664	-867.29740	1	1	1	5.85116	-869.06418

The ARIMA(1,1,0) model with constant yields the smallest AICc -870.87692, so we choose ARIMA(1,1,0) with constant for all following procedures.

Estimation

The Minitab output for ARIMA(1,1,0) with constant is:

ARIMA Model: LogAdjClose

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.1286	0.0647	1.99	0.048
Constant	0.0218	0.0103	2.13	0.035

Differencing: 1 regular difference

Number of observations: Original series 238, after differencing 237

Both the AR1 parameter and the constant are statistically significant, with p-values less than 0.05.

If we denote $\{x_t\}$ as the time series of Adj Close, $\{y_t\}$ as LogAdjClose and $\{z_t\}$ as DifLogAdjClose. LogAdjClose fits the ARIMA(1,1,0) model with a constant. The best estimate of the AR1 coefficient is 0.1286 and that of the constant is 0.0218. Therefore, the fitted model is $z_t = 0.0218 + 0.1286 z_{t-1} + \epsilon_t$ where $z_t = y_t - y_{t-1} = \log x_t - \log x_{t-1}$.

Diagnostic Checking

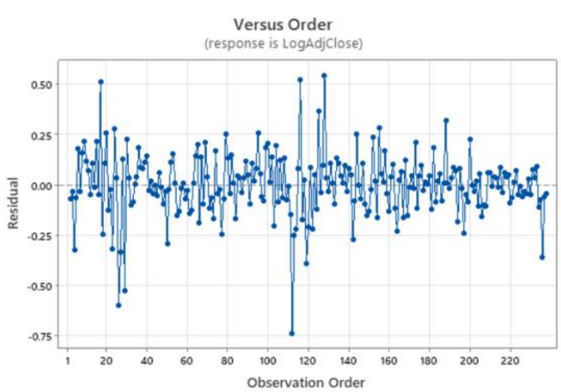
The output of the Ljung-Box test is:

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	7.95	22.51	35.05	42.25
DF	10	22	34	46
P-Value	0.634	0.430	0.418	0.630

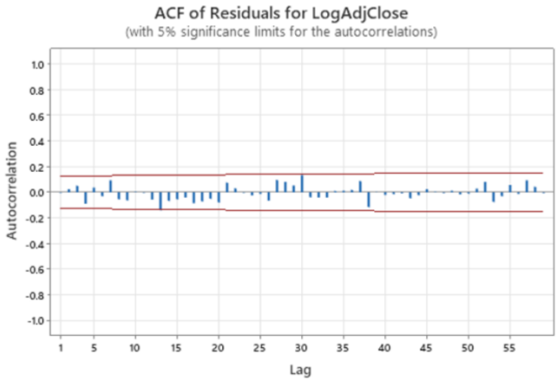
Since all 4 p-values are greater than 0.05, we fail to reject the null hypothesis that the model is inadequate. Therefore, we conclude that the model is adequate.

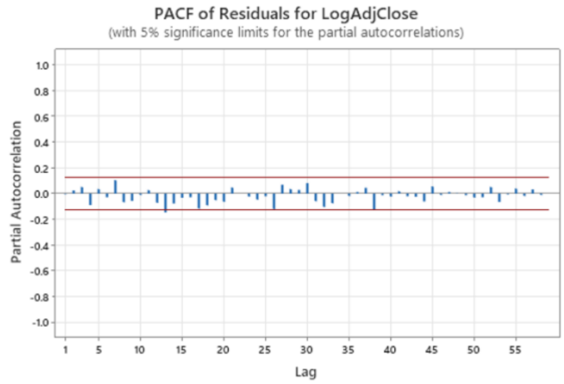
The plot of the residuals is as follows:



The residuals look quite random, with only a few with absolute values over 0.5.

The ACF and PACF of the residuals are as follows:





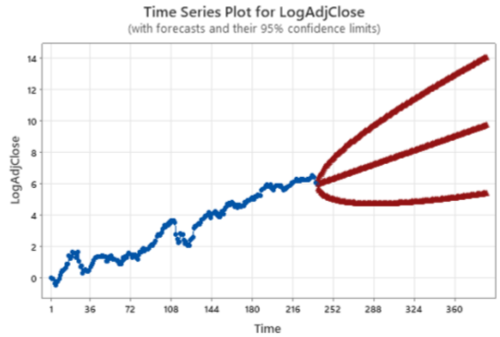
The residuals look uncorrelated. The plots do not indicate much inadequacy in the model.

Forecasting

The forecasts and 95% forecast intervals for 1-150 are as follows:

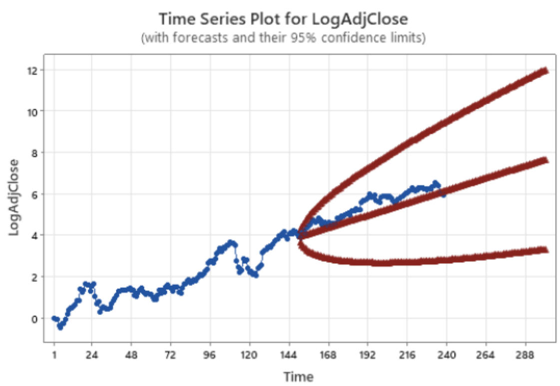
Forecasts from period 238

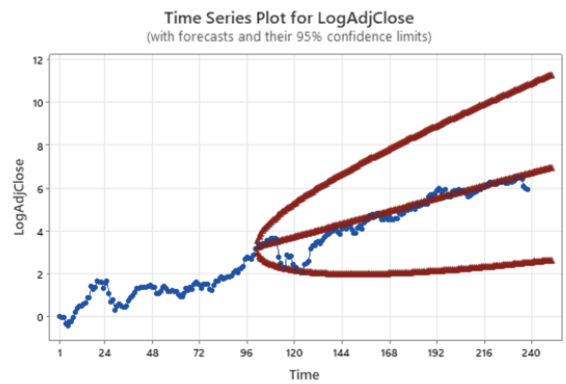
		95% Limits		
Period	Forecast	Lower	Upper	Actual
239	5.96579	5.65629	6.2753	
240	5.98990	5.52320	6.4566	
241	6.01481	5.42878	6.6008	
242	6.03982	5.35460	6.7250	
243	6.06484	5.29303	6.8366	
244	6.08986	5.24025	6.9395	
245	6.11488	5.19401	7.0357	
246	6.13990	5.15291	7.1269	
247	6.16492	5.11596	7.2139	
248	6.18994	5.08248	7.2974	
249	6.21496	5.05194	7.3780	
250	6.23999	5.02393	7.4560	
251	6.26501	4.99814	7.5319	
252	6.29003	4.97431	7.6057	
253	6.31505	4.95223	7.6779	
254	6.34007	4.93173	7.7484	
255	6.36509	4.91265	7.8175	
256	6.39011	4.89487	7.8854	
257	6.41514	4.87829	7.9520	
258	6.44016	4.86280	8.0175	
259	6.46518	4.84832	8.0820	
260	6.49020	4.83479	8.1456	
261	6.51522	4.82213	8.2083	
262	6.54024	4.81030	8.2702	
263	6.56526	4.79923	8.3313	
264	6.59029	4.78889	8.3917	
265	6.61531	4.77923	8.4514	
266	6.64033	4.77021	8.5104	
267	6.66535	4.76180	8.5689	
268	6.69037	4.75397	8.6268	
269	6.71539	4.74668	8.6841	
270	6.74041	4.73992	8.7409	
271	6.76544	4.73365	8.7972	
272	6.79046	4.72786	8.8531	
273	6.81548	4.72253	8.9084	
274	6.84050	4.71762	8.9634	
275	6.86552	4.71313	9.0179	
276	6.89054	4.70905	9.0720	
277	6.91557	4.70534	9.1258	
278	6.94059	4.70200	9.1792	
279	6.96561	4.69902	9.2322	
280	6.99063	4.69638	9.2849	
281	7.01565	4.69407	9.3372	
282	7.04067	4.69208	9.3893	
283	7.06569	4.69039	9.4410	
284	7.09072	4.68901	9.4924	
285	7.11574	4.68791	9.5436	
286	7.14076	4.68708	9.5944	
287	7.16578	4.68653	9.6450	
288	7.19080	4.68624	9.6954	
289	7.21582	4.68620	9.7454	
290	7.24084	4.68640	9.7953	
291	7.26587	4.68685	9.8449	
292	7.29089	4.68753	9.8942	
293	7.31591	4.68843	9.9434	
294	7.34093	4.68955	9.9923	
295	7.36595	4.69089	10.0410	
296	7.39097	4.69243	10.0895	
297	7.41599	4.69418	10.1378	
298	7.44102	4.69612	10.1859	
299	7.46604	4.69826	10.2338	
300	7.49106	4.70058	10.2815	
301	7.51608	4.70309	10.3291	
302	7.54110	4.70578	10.3764	
303	7.56612	4.70864	10.4236	
304	7.59115	4.71167	10.4706	
305	7.61617	4.71486	10.5175	
306	7.64119	4.71822	10.5642	
307	7.66621	4.72174	10.6107	
308	7.69123	4.72542	10.6570	
309	7.71625	4.72925	10.7033	
310	7.74127	4.73323	10.7493	
311	7.76630	4.73735	10.7952	
312	7.79132	4.74162	10.8410	
313	7.81634	4.74602	10.8867	
314	7.84136	4.75057	10.9322	
315	7.86638	4.75525	10.9775	
316	7.89140	4.76006	11.0227	
317	7.91642	4.76500	11.0678	
318	7.94145	4.77007	11.1128	
319	7.96647	4.77526	11.1577	
320	7.99149	4.78058	11.2024	
321	8.01651	4.78601	11.2470	
322	8.04153	4.79157	11.2915	
323	8.06655	4.79724	11.3359	
324	8.09157	4.80302	11.3801	
325	8.11660	4.80892	11.4243	
326	8.14162	4.81492	11.4683	
327	8.16664	4.82104	11.5122	
328	8.19166	4.82726	11.5561	
329	8.21668	4.83358	11.5998	
330	8.24170	4.84001	11.6434	
331	8.26673	4.84654	11.6869	
332	8.29175	4.85316	11.7303	
333	8.31677	4.85989	11.7736	
334	8.34179	4.86671	11.8169	
335	8.36681	4.87363	11.8600	
336	8.39183	4.88064	11.9030	
337	8.41685	4.88774	11.9460	
338	8.44188	4.89493	11.9888	
339	8.46690	4.90221	12.0316	
340	8.49192	4.90958	12.0743	
341	8.51694	4.91704	12.1168	
342	8.54196	4.92458	12.1593	
343	8.56698	4.93220	12.2018	
344	8.59200	4.93991	12.2441	
345	8.61703	4.94770	12.2863	
346	8.64205	4.95557	12.3285	
347	8.66707	4.96352	12.3706	
348	8.69209	4.97155	12.4126	
349	8.71711	4.97966	12.4546	
350	8.74213	4.98784	12.4964	
351	8.76715	4.99610	12.5382	
352	8.79218	5.00443	12.5799	
353	8.81720	5.01283	12.6216	
354	8.84222	5.02131	12.6631	
355	8.86724	5.02986	12.7046	
356	8.89226	5.03847	12.7460	
357	8.91728	5.04716	12.7874	
358	8.94230	5.05592	12.8287	
359	8.96733	5.06474	12.8699	
360	8.99235	5.07363	12.9111	
361	9.01737	5.08259	12.9521	
362	9.04239	5.09161	12.9932	
363	9.06741	5.10070	13.0341	
364	9.09243	5.10985	13.0750	
365	9.11746	5.11907	13.1158	
366	9.14248	5.12834	13.1566	
367	9.16750	5.13768	13.1973	
368	9.19252	5.14708	13.2380	
369	9.21754	5.15654	13.2785	
370	9.24256	5.16606	13.3191	
371	9.26758	5.17563	13.3595	
372	9.29261	5.18527	13.3999	
373	9.31763	5.19496	13.4403	
374	9.34265	5.20471	13.4806	
375	9.36767	5.21451	13.5208	
376	9.39269	5.22437	13.5610	
377	9.41771	5.23429	13.6011	
378	9.44273	5.24426	13.6412	
379	9.46776	5.25428	13.6812	
380	9.49278	5.26436	13.7212	
381	9.51780	5.27449	13.7611	
382	9.54282	5.28467	13.8010	
383	9.56784	5.29491	13.8408	
384	9.59286	5.30519	13.8805	
385	9.61788	5.31552	13.9202	
386	9.64291	5.32591	13.9599	
387	9.66793	5.33634	13.9995	
388	9.69295	5.34683	14.0391	

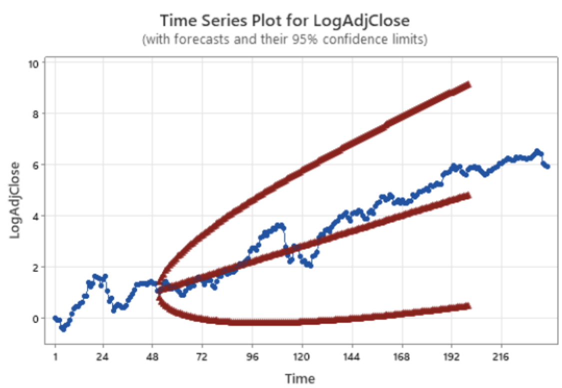


The forecast looks reasonable. We can see that the point estimate shows a linear trend because the model is with a constant and $d = 1$. The forecast interval has a tendency to blow up because LogAdjClose is not stationary.

According to the back-testing at t = 150, 100 and 50, the forecast intervals are wide enough to cover all datapoints, but the forecast intervals seem excessively wide.







<u>Conclusions</u>

Since our best model is an ARIMA(1,1,0) with a constant/drift and the constant term is greater than 0, we conclude that Netflix is a good company to buy and hold. We can do an out-of-sample test for the forecast in the future when we have future data.