Jessy Huang

Time Series Data

Professor Hurvich

5/2/22

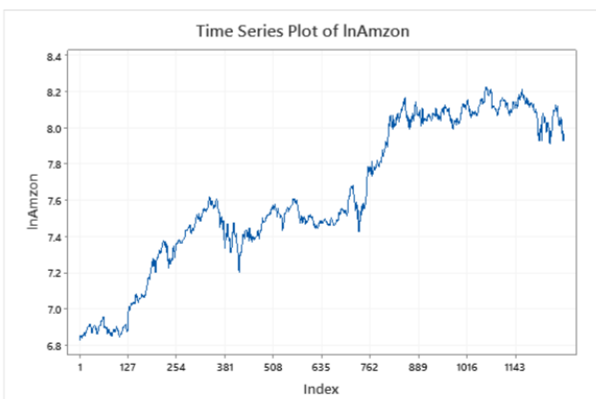
ARIMA-ARCH Model on Time Series Forecasting

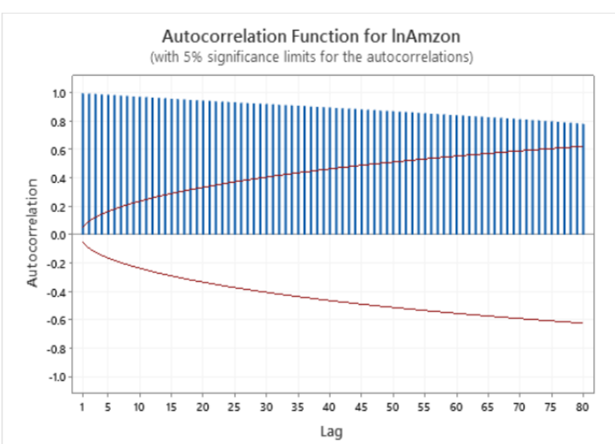
For this project, I chose the data for the historical daily close prices of Amazon (NasdaqGS:AMZN). I obtained the data from Capital IQ and the dates are from April 28, 2017 to April 29, 2022, total n=1270. For the purpose of this project, I removed the last data point and will be using it again to check the performance of the forecast intervals later. I will use minitab for ARIMA and R for ARCH models for this project.

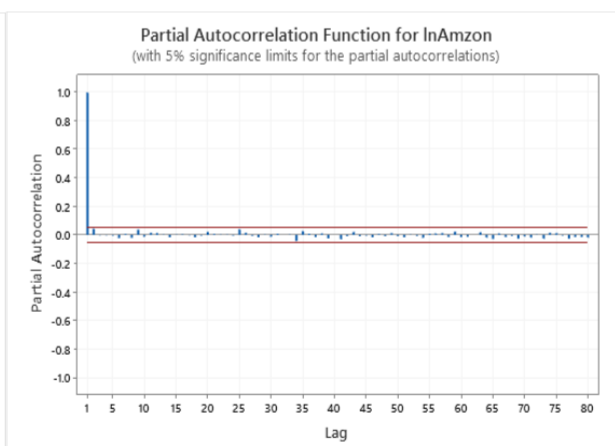
Data URL: https://www.capitaliq.com/CIQDotNet/Charting4/ModernBuilder.aspx?CompanyId=18749&fromC3=11)

Part 1: Time Series Plot of Amazon

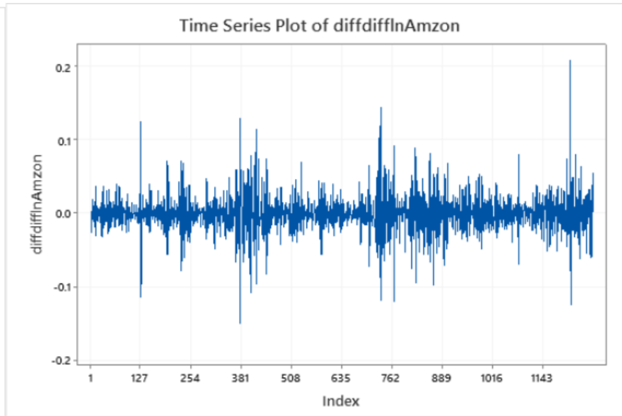
To start, I plotted a time series of the share prices of Amazon. The time series plot has some sharp up and downs throughout the time range. To remove the level dependent volatility, I'll use natural log of the pricing data for the rest of my analysis.

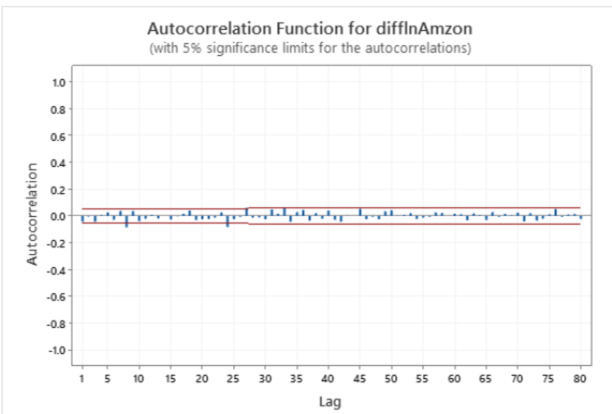


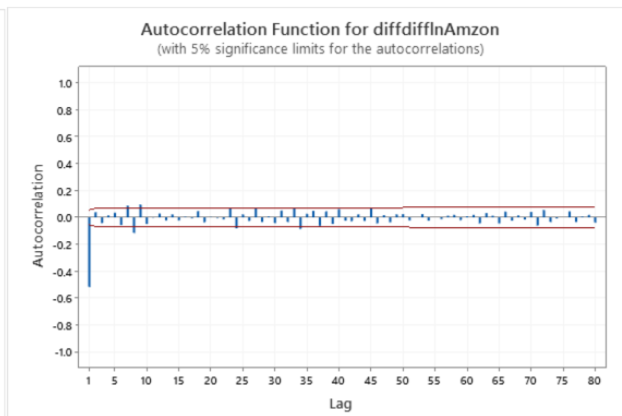


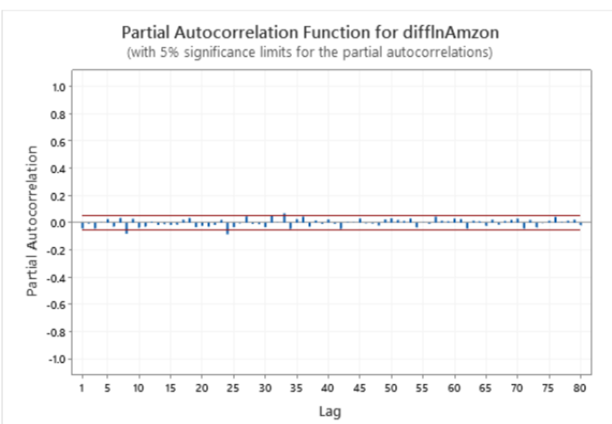


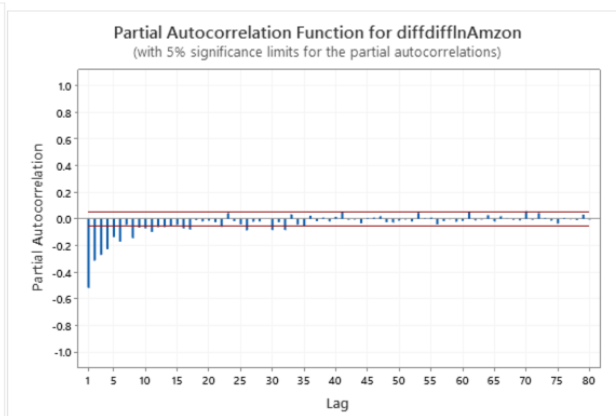
From the time series plot of the natural log of Amazon, we can see that the log-price is not stationary because the series is not mean-reverting. The ACF of the log-price stays around 1 and decreases gradually, and the PACF is almost 1 at lag-1 and cuts off after lag-1. This suggests that the series should be differenced to see if the difference of log-price is stationary.











To determine if I should use the first difference of log-price, I further computed the second difference of log-price as well as its respective ACF and PACF as shown above. We can see that taking the second differentiation seems to be over differencing, as the ACF and PACF for the second differentiation show a strong pattern of negative autocorrelation in some lags. Therefore, we will use d=1 for the ARIMA model.

After the first difference, the series is stationary. The ACF and PACF of the first differences of log-price are insignificant at all lags except lag 8 and lag 24 which are slightly significant.

Considering the parsimony rule in model identification, “The total number of parameters in the model should be as small as possible”, I’ll only consider potential ARIMA models with d=1 and p and q at a maximum of 2. Therefore, this suggests that an ARIMA(2,1,2) model is ideal for the natural log price. However, we will test several options, and decide through calculating the lowest AICc value.

Part 2: Determining the ARIMA Model

To determine if ARIMA(2, 1, 2) is the suitable model for log-price, I calculated all of the AICc values for (p, 1, q) models with and without a constant, with p and q both ranging from 0 to 2. Among the candidates, ARIMA(2,1,2) with constant has the lowest AICc (-9957.22), therefore ARIMA (2,1,2) with constant is chosen.

AICc formula:

$$AIC_C = N \log\left(\frac{SS}{N}\right) + 2(p+q+1)\frac{N}{N-p-q-2} \quad \text{if no constant term in model ,}$$

$$AIC_C = N \log\left(\frac{SS}{N}\right) + 2(p+q+2)\frac{N}{N-p-q-3} \quad \text{if constant term is included .}$$

ARIMA(0,1,0) SS without constant is manually calculated by:

SS $= \sum^{1269}{}_{=1}$ $^2 = 0.494243$

ARIMA(0,1,0) SS with constant is manually calculated by:

SS $= \sum^{1269}{}_{=1}(- ^{-})^2 = 0.493218$

$=$ $-$ $_{-1}$ and is the natural log price of Amazon stock.

ARIMA+C	n	N	p	d	q	SS	AICc
(0,1,0)	1269	1268	0	1	0	0.493218	-9952.33
(0,1,1)	1269	1268	0	1	1	0.492261	-9952.78
(0,1,2)	1269	1268	0	1	2	0.492237	-9950.83
(1,1,0)	1269	1268	1	1	0	0.492273	-9952.75
(1,1,1)	1269	1268	1	1	1	0.491615	-9952.43
(1,1,2)	1269	1268	1	1	2	0.50318	-9920.93
(2,1,0)	1269	1268	2	1	0	0.49226	-9950.77
(2,1,1)	1269	1268	2	1	1	0.492241	-9948.8
(2,1,2)	1269	1268	2	1	2	0.488207	-9957.22

ARIMA+wiout C n		N	p	d	q	SS	AICc
(0,1,0)	1269	1268	0	1	0	0.494243	-9951.7
(0,1,1)	1269	1268	0	1	1	0.493375	-9951.92
(0,1,2)	1269	1268	0	1	2	0.493365	-9949.94
(1,1,0)	1269	1268	1	1	0	0.493382	-9951.91
(1,1,1)	1269	1268	1	1	1	0.49325	-9950.23
(1,1,2)	1269	1268	1	1	2	0.49295	-9948.99
(2,1,0)	1269	1268	2	1	0	0.493377	-9949.91
(2,1,1)	1269	1268	2	1	1	0.493338	-9948
(2,1,2)	1269	1268	2	1	2	0.492791	-9947.39

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-1.7884	0.0121	-147.48	0.000
AR 2	-0.9020	0.0141	-64.10	0.000
MA 1	-1.75522	0.00497	-353.28	0.000
MA 2	-0.8615	0.0128	-67.30	0.000
Constant	0.00331	0.00192	1.72	0.085

Differencing: 1 regular difference

Number of observations: Original series 1269, after differencing 1268

The results show that all coefficients are statistically significant since the t-ratio p-value is less than 0.05, except the constant coefficient whose t-ratio p-value is greater than 0.05, which means that the constant coefficient is statistically insignificant. However, t-ratios are valid if and only if the fitted model is correct. We will never know if the fitted model is correct, so the t-value is unreliable. We can continue with our analysis and interpret the ARIMA(2,1,2) with constant model as:

stands for the first difference of the natural log of Amazon.

Residual Sums of Squares

DF	SS	MS
1263	0.488207	0.0003865

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

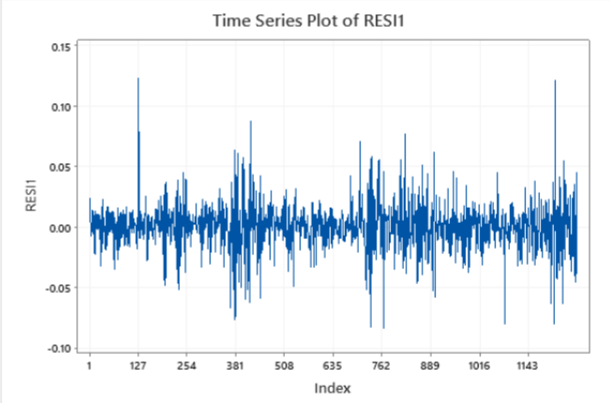
Lag	12	24	36	48
Chi-Square	8.20	23.86	46.06	55.96
DF	7	19	31	43
P-Value	0.315	0.202	0.040	0.089

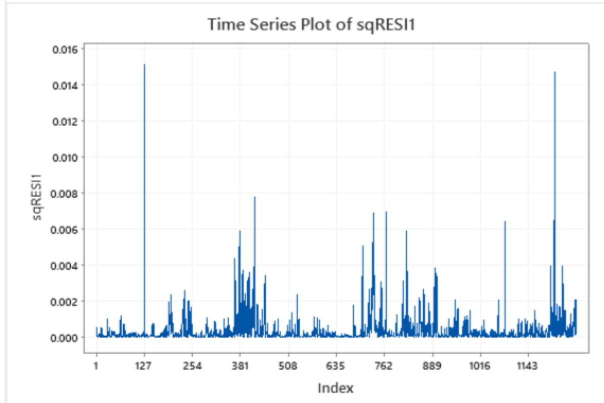
The Ljung-Box results indicate that the model is adequate, we do not need to add more parameters, as every Ljung-Box chi-square values' corresponding p-value is above 0.05 except lag 36 whose p-value is slightly less than 0.05. Overall this is a desirable result.

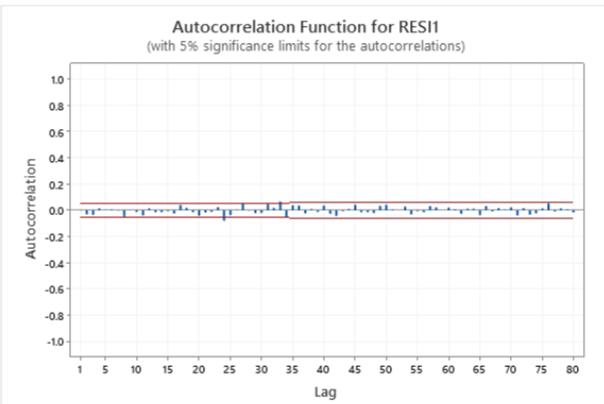
Forecasts from period 1269

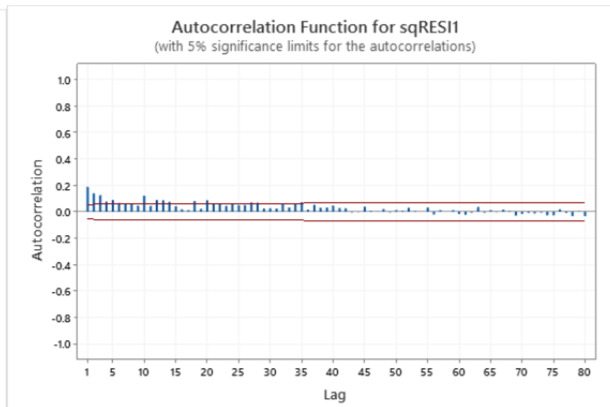
		95% Limits		
Period	Forecast	Lower	Upper	Actual
1270	7.96918	7.93063	8.00772	

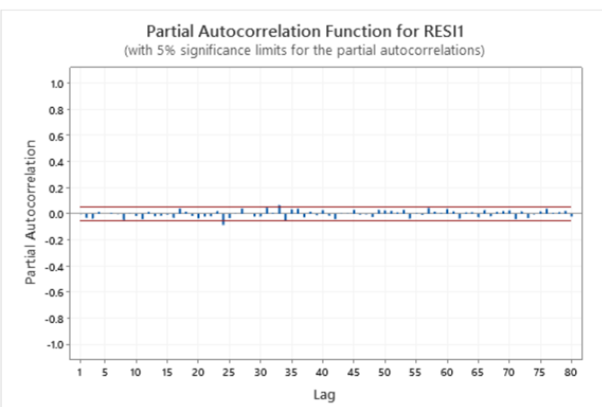
Part 3: Checking for Conditional Heteroscedasticity

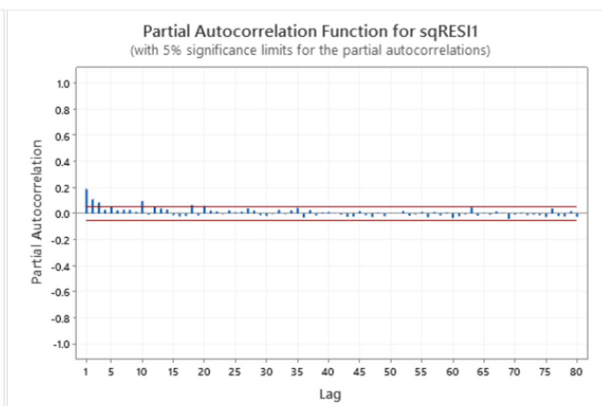












The ACF and PACF of the ARIMA(2,1,2) with constant model residuals are not significant at any lags, this is understood since the residuals is the non-gaussian white noise. The ACF of the squared residuals are significant in many lags at the beginning, then dies down, and PACF of the squared residuals cuts off at lag 3, and is slightly significant at lag 10, which might be the sign of a potential q=4 or q= 5, ARCH model. It indicates that residuals are approximately uncorrelated, but are not independent. The volatility appears like dynamic volatility clustering, appears conditional variance depending on the past and present data, which indicates the evidence of autoregression conditional heteroscedasticity, suggesting a potential ARCH(4) or ARCH(5) model.

Part 4: Finding the Appropriate ARCH(q) or GARCH (1,1) Model

Using R on the residuals from minitab ARIMA(2,1,2) with constant, I calculated AICc of ARCH with q from 0 to 10 and GARCH(1,1). I choose GARCH(1,1) since it has the lowest AICc value (-6574.424).

Loglik for q=0 is manually calculated by the following formula:

```
loglik0 <- -0.5 * N * (1 + log(2 * pi * mean(x^2)))
```

AICc is calculated by the following formula:

```
k <- q + 1
aicc <- -2 * loglik + 2 * k * N / (N - k - 1)
```

ARCH(q)		
q	logLik	AICc
0	3185.429	-6368.855
1	3230.907	-6457.804
2	3255.971	-6505.923
3	3272.508	-6536.984
4	3282.464	-6554.881
5	3285.23	-6558.393
6	3282.525	-6550.96
7	3283.938	-6551.761
8	3277.531	-6536.918
9	3270.706	-6521.237
10	3265.571	-6508.931
garch(1,1)		
	3290.221	-6574.424

GARCH(1,1) model R command and the summary and log likelihood:

```
 >model <- garch(x, c(1,1), trace=FALSE) # ARCH(5)
>logLik(model)
loglik: 3290.221
>summary(model)
Call:
garch(x = x, order = c(1, 1), trace = FALSE)

Model:
GARCH(1,1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.31637 -0.54628 0.01867 0.54583 9.04118

Coefficient(s):
```

```
   Estimate Std. Error t value Pr(>|t|)
a0 3.483e-05   5.192e-06    6.707 1.99e-11   ***
a1 1.606e-01   2.363e-02    6.800 1.05e-11   ***
b1 7.539e-01   3.180e-02    23.706 < 2e-16   ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Diagnostic Tests:
	Jarque Bera Test
data: Residuals
X-squared = 2432.9, df = 2, p-value < 2.2e-16

	Box-Ljung test
data: Squared.Residuals
X-squared = 0.15846, df = 1, p-value = 0.6906
```

Above is the GARCH(1,1) summary from R. In term of t-stats p-value, there is mistake on term “Pr(>|t|)” in R model summary. This should be right tail t-test, therefore, as long as 0.5* Pr(>|t|) is less than 0.05, we can conclude the coefficient is significant. At our case, each coefficient (a0, a1, b1) P-value is <0.05, therefore all coefficient are statistically significant.

GARCH(1,1) model can be defined as:

=3.483* ⁻ + 1.606* ⁻ *ε ₋ + 0.7539* ₋

h is the conditional variance of the ARIMA(2,1,2) with constant model residuals.

The unconditional(marginal) variance of the ARIMA model residuals:

Var (ε)=ɕ/(1-(a1+b1))= **3.483*** ⁻ **/(1-1.606*** ⁻ – . **)=4.0737*** ₋

We can see that the unconditional(marginal) variance of the shocks in the model doesn’t depend on t, it is finite constant. In other word, unconditionally, {ε } process is stationary since a1+b1 = 0.9145 is less than 1, implying the finite variance of the shock.

Part 5: 95% One Step Forecast Interval

Now I am using the following R command to get ht, the conditional variance of the ARIMA model residuals.

```
>ht=model$fit[,1]^2
```

ht(t=1269)= **6.6704*** ₋ ARIMA residual ε (t=1269)=**0.045827**

, **= 3.483* ⁻ + 1.606* ⁻ *0.045827^2+ 0.7539*6.7604* ⁻ =8.7498975* ⁻**

From part 2 minitab ARIMA model , **=7.96918 and** **95% forecast interval:(7.93063,8.00772)**

Hence 95% one step ahead forecast interval for the ARIMA-GARCH model are:

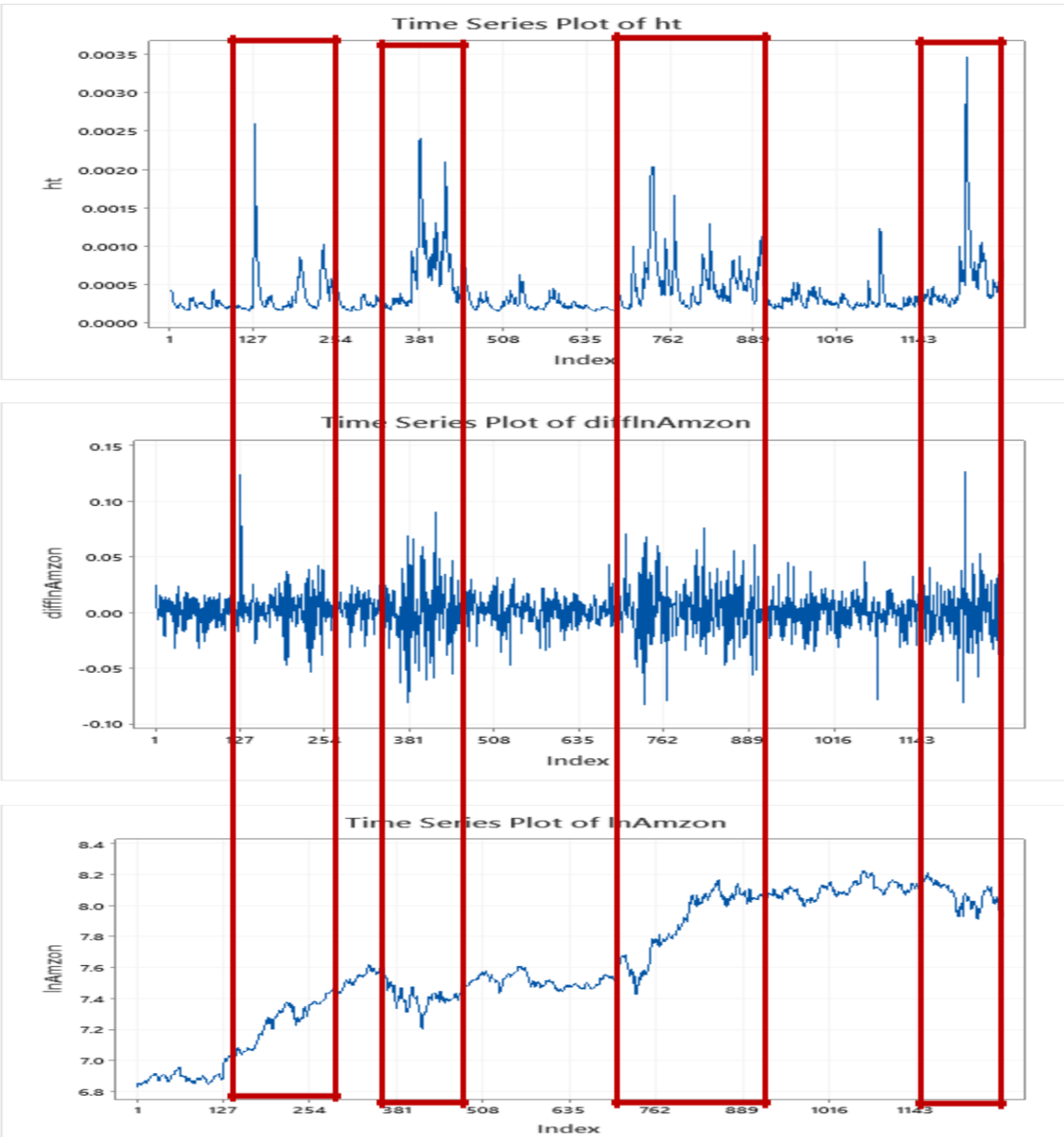
$, \pm . \sqrt{\quad} , = . \quad \pm . \sqrt{8.7498975} * 10^{-4} = (. \quad , . \quad)$

You can see the ARIMA-GARCH one step ahead 95% forecast interval is wider than ARIMA. This shows the ability of ARCH model to better adapt to changing volatility conditions. In other words, the combined model better reflects the recent changes (such as volatility) in the log-prices of Amazon by using the residuals and conditional variances.

This suggests that the ARIMA-GARCH model is a slightly more precise prediction.

Part 6: Conditional Variance Plot

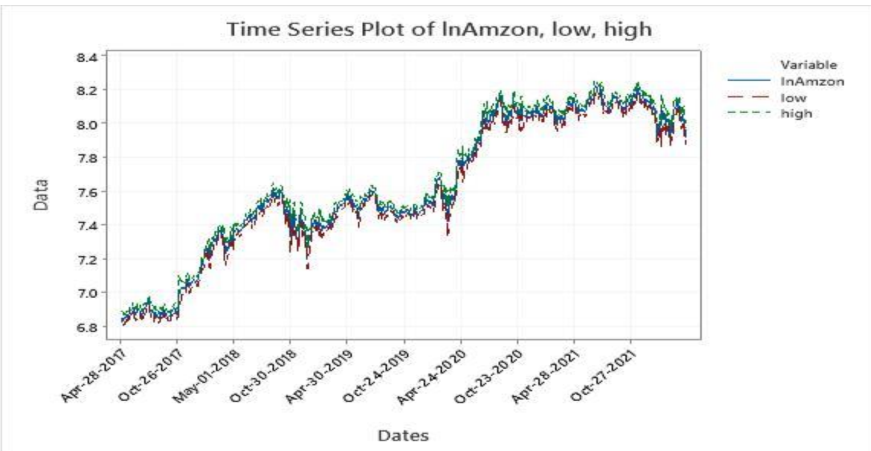
I plotted the fitted conditional variance ht and the differenced logarithmic series versus order in Minitab:



There are several period volatility bursts in ht, such as at t = 130~133, t=381~387, t = 735~737, and t= 1215~1216, where we can see that the periods of high volatility clusters in ht correspond to the observed volatility in the logarithmic series.

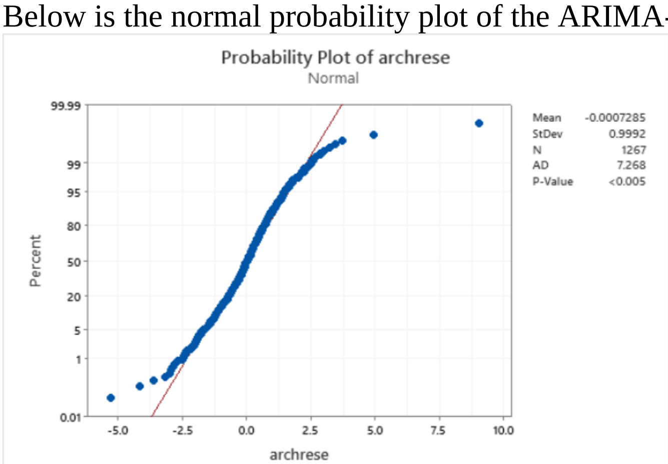
Part 7: Time Series Plot with Forecast Intervals

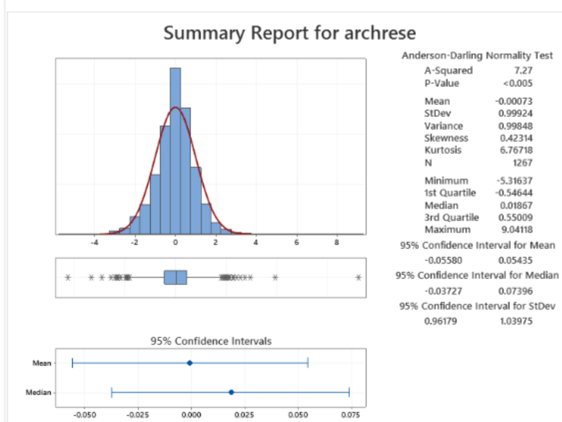
From the fitted data, it appears that the forecast intervals track the data well, but in an actual forecasting context, model might not be entirely realistic, might not fit so well since the ARIMA-ARCH parameters are estimated from the entire data set, not just the observations up to the time at which the forecast is to be constructed.



Part 8: Test: Is ARIMA (2, 1 , 2) with constant – GARCH (1, 1) Model Adequate?

Below is the normal probability plot of the ARIMA-ARCH model's residuals:





The low p-value (<0.005) indicates that the ARIMA-GARCH residuals do not follow a normal distribution. The S-shaped probability plot suggests a fat-tailed distribution instead. The Kurtosis is 6.76718 which is greater than 3. Therefore, the model appears to not have been able to adequately describe the leptokurtosis in the data.

Part 9: Failure of Forecast Intervals and Checking Forecast Intervals

Calculating via Minitab, the 95% prediction interval constructed from the ARIMA-GARCH model failed to cover the daily exchange rate 67 times. With 1269-2=1267 total number of predictions, the rate of failure was 67/1267=5.288%.

1270th (April 29, 2022) Amazon stock price check:

The ARIMA 95% one step ahead forecast interval of 1270th is **(7.93063, 8.00772),** the ARIMA-GARCH model is **(7.9112027, 8.027157),** the actual value of the 1270th observation (**April 29th, 2022**) is **ln(2485.63) = 7.81828**. This point falls outside both the ARIMA and ARIMA-GARCH 95% prediction interval. Both ARIMA and ARIMA-GARCH 95% forecast interval seems to be too narrow, but ARIMA-GARCH 95% prediction interval lower boundary is 7.9112, which is slightly closer to the actual value 7.81828 than ARIMA model’s lowest boundary 7.93063, therefore ARIMA-GARCH seems slightly more accurate than ARIMA.

Based on the news, on April 29, 2022, Amazon stock suffered its biggest one-day drop since 2006, falling 14% after report of quarterly loss. This explains why the 1270th real data does not fall into ARIMA-GARCH 95% 1270th forecast interval. As we know, ARIMA-GARCH is conditional volatility model, in other word, it forecasts the future based on the past and present conditions. On April 29, 2022, Amazon announced its first quarterly loss, which is an unexpected event that cannot be adequately expressed or forecasted using past data. This event introduced new condition which is not included in the model, Hence, the portion introduced by the new condition cannot be forecasted by the model.