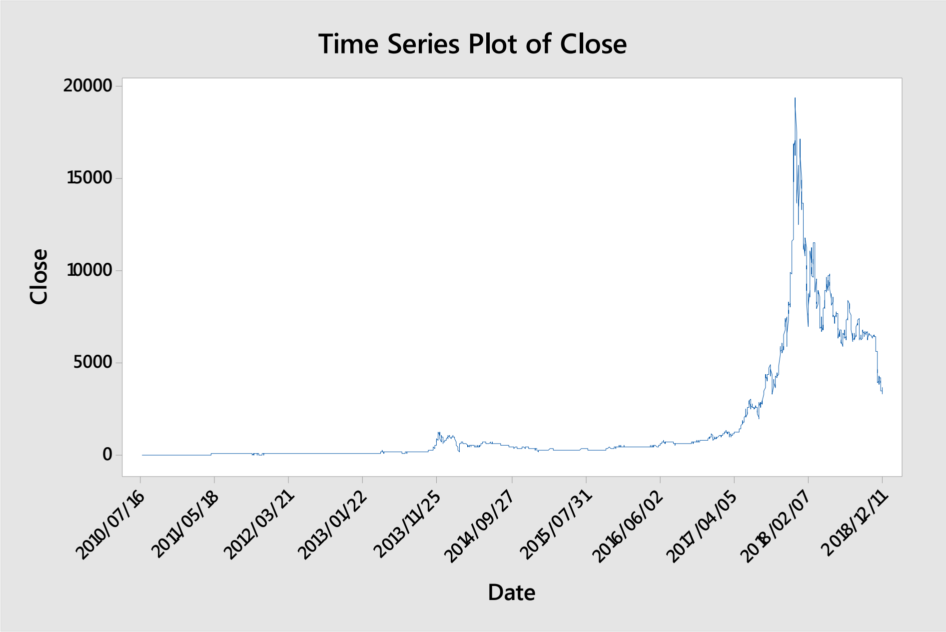
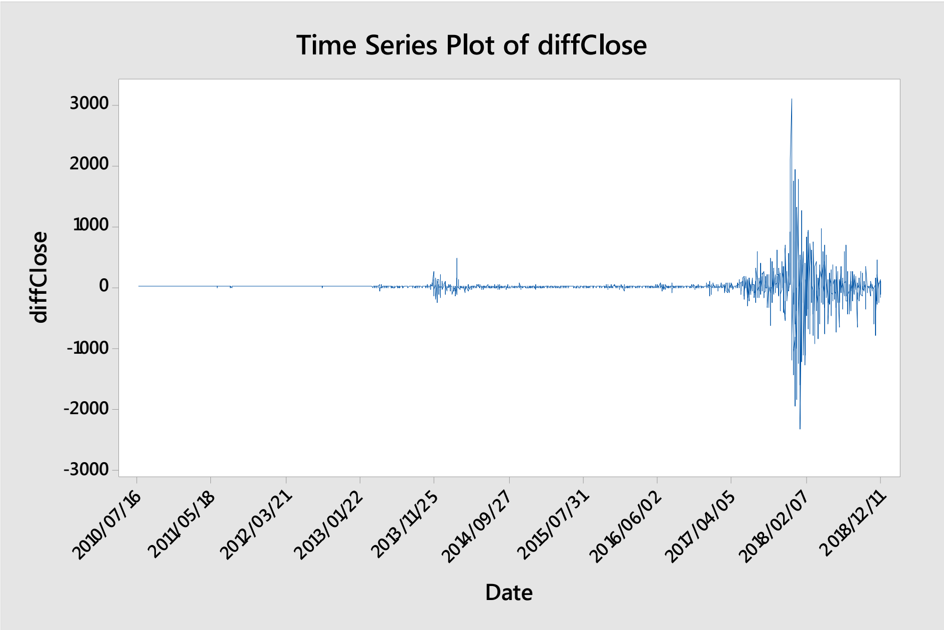
Forecasting Time Series Data Project **Tianze Lin tl2718**

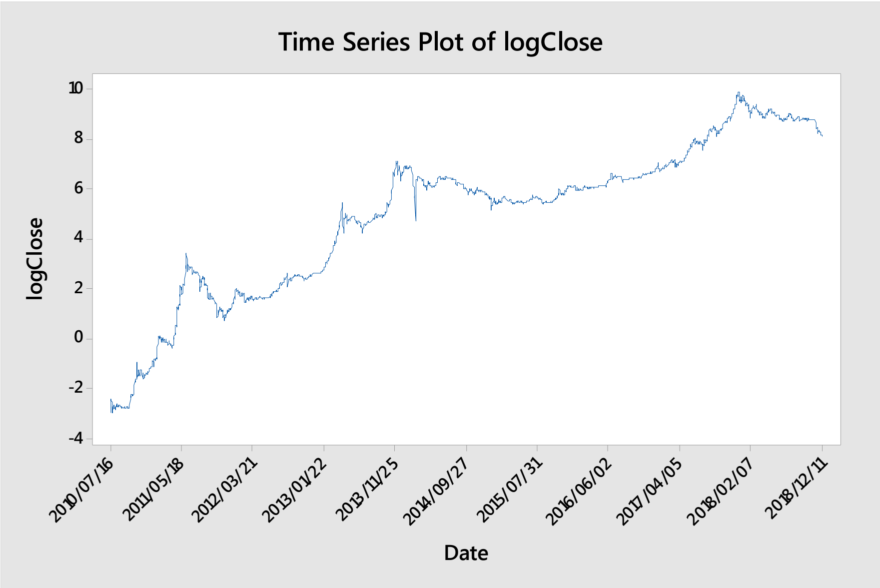
In this project, the value of a bitcoin in USD from July 16th, 2010 to December 13th, 2018 will be analysed. The number of observations in this dataset is 3072. The dataset is available on https://finance.yahoo.com/quote/BTC-USD?p=BTC-USD.

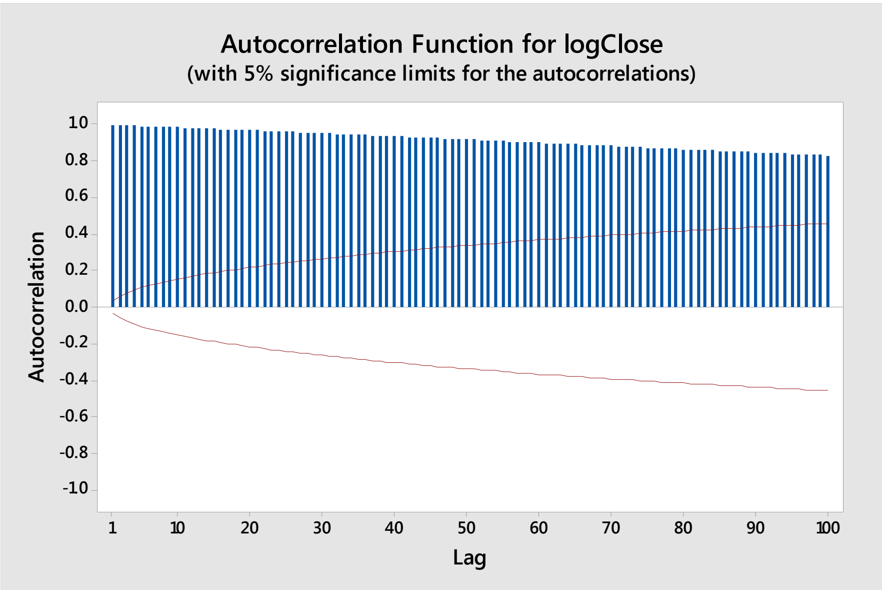
First, plot the closing price and the first difference of closing price.

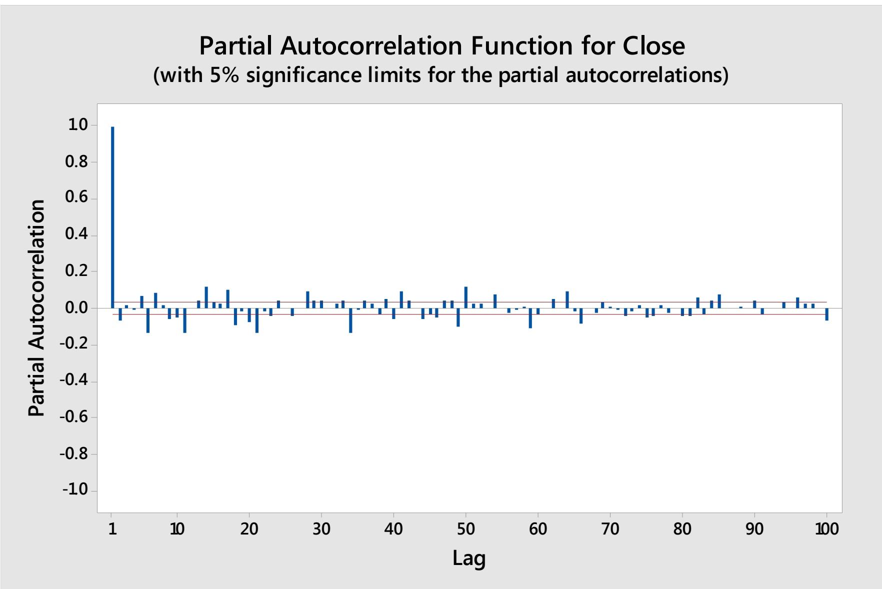




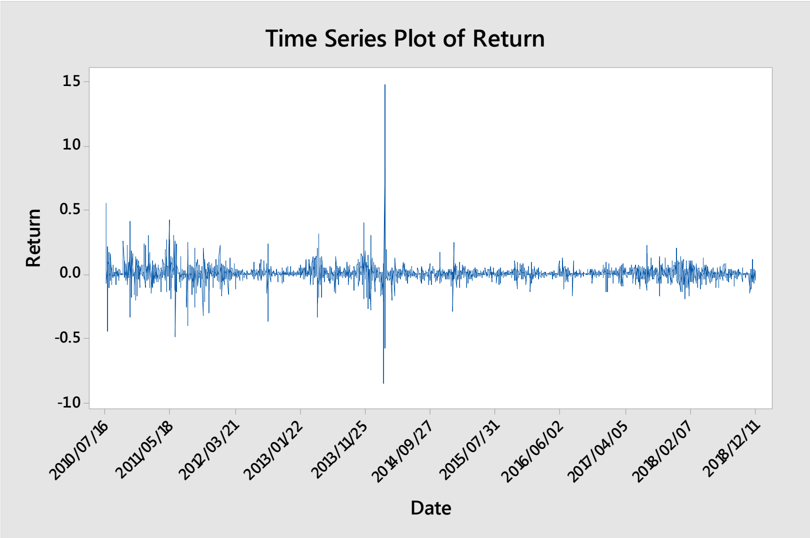
From time series plots above, there is strong evidence that the volatility of sales is dependent on the level of sales. When the closing price rises, the volatility also rises. Furthermore, the change in natural logs of closing prices is approximately equal to the return. Therefore, we should work with the log of closing price.
Next, plot the log of closing price and the ACF and PACF of log of closing price.

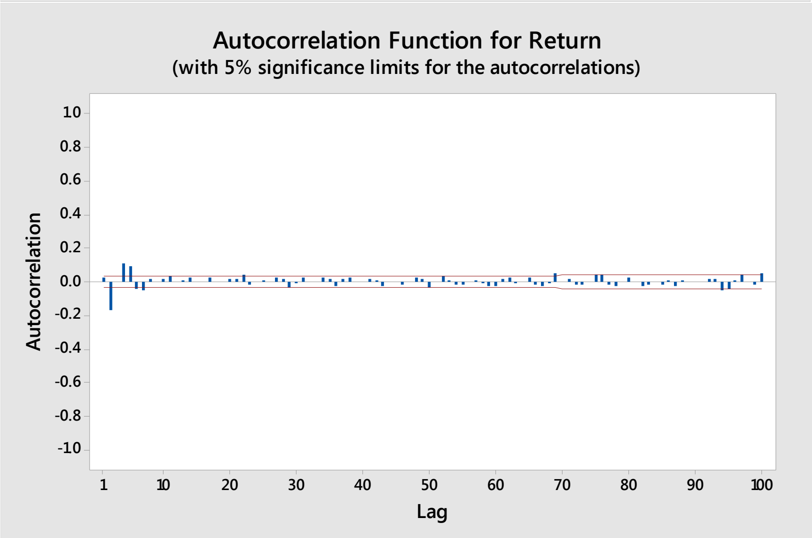

Time Series Plot of logClose
logClose
Date
10
8
6
4
2
0
-2
-4
2010/07/16
2011/05/18
2012/03/21
2013/01/22
2013/11/25
2014/09/27
2015/07/31
2016/06/02
2017/04/05
2018/02/07
2018/12/11

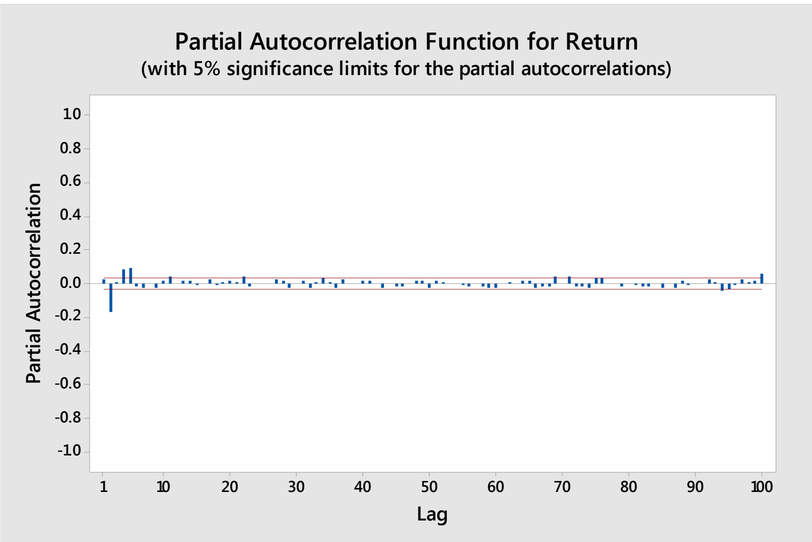

Autocorrelation Function for logClose
(with 5% significance limits for the autocorrelations)
Autocorrelation
Lag
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
10
20
30
40
50
60
70
80
90
100


Partial Autocorrelation Function for Close
(with 5% significance limits for the partial autocorrelations)
Partial Autocorrelation
Lag
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
1
10
20
30
40
50
60
70
80
90
100

ACF decreases very slowly and PACF cuts off beyond lag 1, so we should difference the data. From time series plots, ACF and PACF, there is no strong seasonal pattern, so the seasonal component may not exist.
Next, plot the time series, ACF and PACF of differenced log of closing price (return).







The autocorrelations are small and pattern less, and the time series looks mean-reverting, so higher order of differencing is not needed. ACF and PACF don't provide enough information on selecting p and q of ARIMA model, so AICc should be used to select p and q.

With Consant				Without Consant			
p	q	SS	AICc	p	q	SS	AICc
0	0	14.4226	-16461.5	0	0	14.46276	-16451
0	1	14.4036	-16463.6	0	1	14.4403	-16453.7
0	2	14.033	-16541.6	0	2	14.0828	-16528.7
1	0	14.4104	-16462.1	1	0	14.4482	-16452.1
1	1	14.2598	-16492.4	1	1	14.2959	-16482.6
1	2	14.032	-16539.8	1	2	14.0802	-16527.3
2	0	13.9664	-16556.2	2	0	14.0181	-16542.9
2	1	13.9652	-16554.5	2	1	N/A	N/A
2	2	13.8399	-16580.1	2	2	13.8894	-16567.2

AICc suggests an ARIMA (2,1,2) model with constant. Fit the model in Minitab.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.2120	0.0626	3.39	0.001
AR 2	-0.6228	0.0628	-9.92	0.000
MA 1	0.1990	0.0709	2.81	0.005
MA 2	-0.4622	0.0712	-6.49	0.000
Constant	0.00508	0.00153	3.32	0.001

The p-values of all the coefficients and the constant term are smaller than 0.05, so all the coefficients and constant term are statistically significant.

Let x_t be the return. Then the complete form of our fitted model is

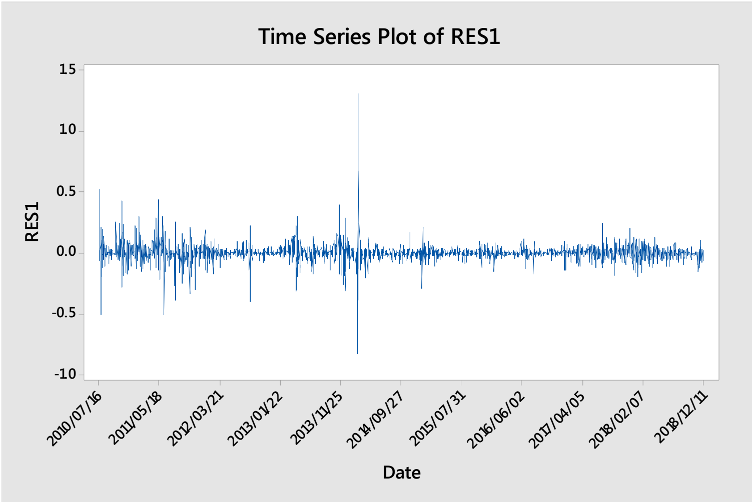
$$x_t = 0.212x_{t-1} - 0.6228x_{t-2} + \varepsilon_t - 0.199\varepsilon_{t-1} + 0.4622\varepsilon_{t-2} + 0.00508$$

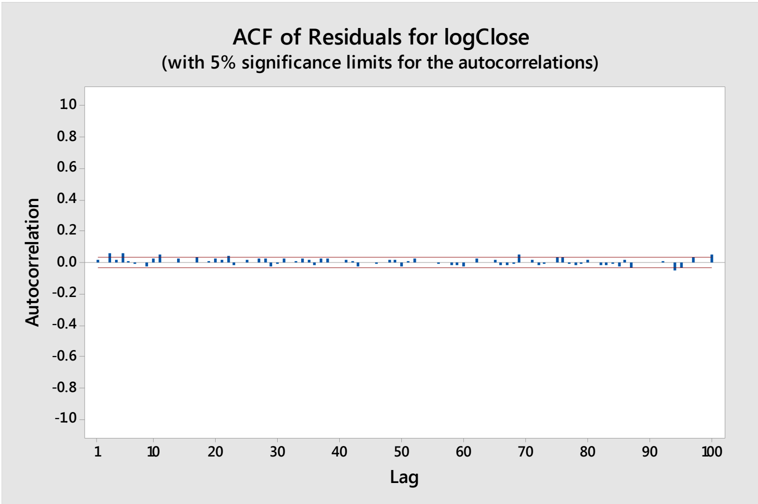
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

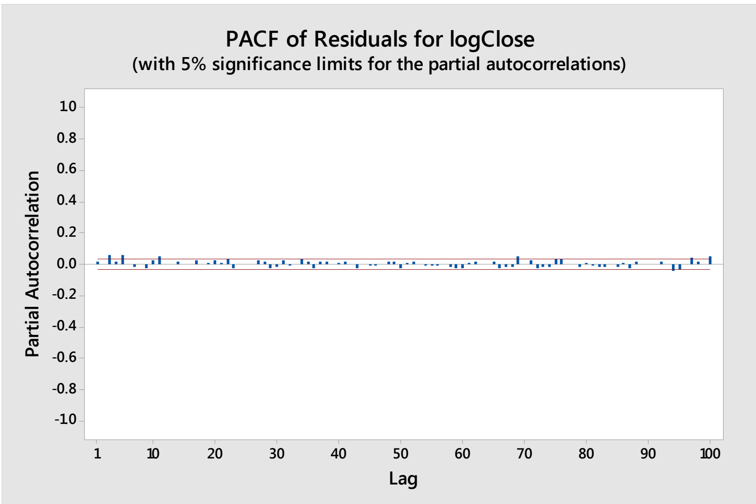
Lag	12	24	36	48
Chi-Square	36.25	50.72	64.82	73.91
DF	7	19	31	43
P-Value	0.000	0.000	0.000	0.002

The Box-Pierce Chi-Square Statistic shows that all the p-values at different lags are smaller than 0.05, so there is strong evidence that our model is inadequate at the significance level of 0.05.

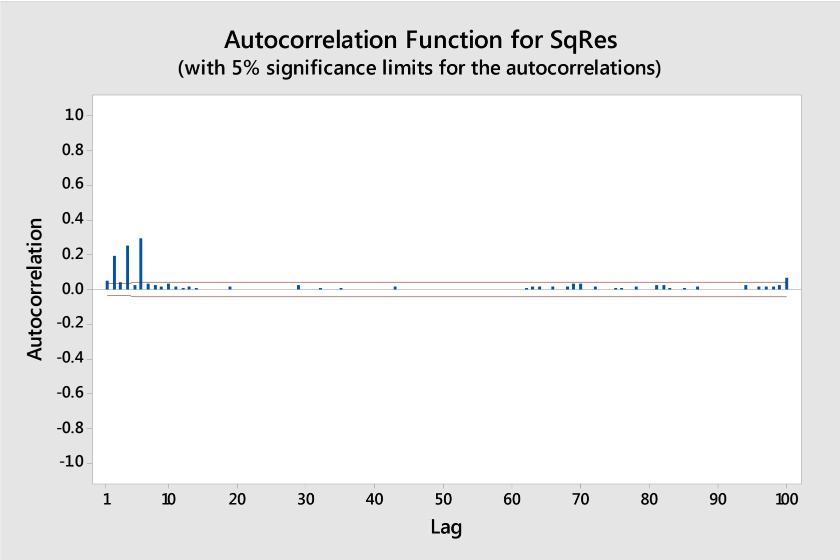
The one-step forecast is 8.10150. The 95% forecast interval is (7.96979, 8.23321). Next, let's plot the residuals from fitted model and the ACF and PACF of residuals.

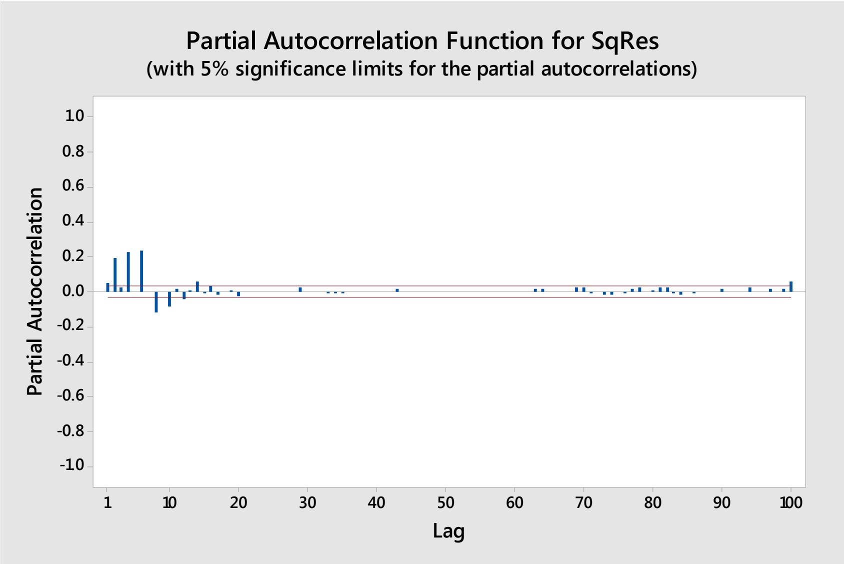






The time series plot of residuals looks stationary, and ACF and PACF are not statistically significant at lag 1 and lag 2, so the residuals might be uncorrelated.
Next, plot ACF and PACF of squared residuals to check if there is conditional heteroskedasticity.





Both ACF and PACF of squared residuals are statistically significant from lag 1 to lag 4. There is strong evidence that conditional heteroskedasticity exists.

Save the residuals. Read the residuals by R and try to fit an ARCH or GARCH model. Compute the AICc for ARCH (0) to ARCH (10) and GARCH (1,1).

```
> y=-0.5*3071*(1+log(2*pi*mean(x^2)))
> -2*y+2*(0+1)*3071/(3071-0-2)
[1] -7873.054

> for(i in 1:10){
+ model <- garch(x,c(0,i),trace=F)
+ print(-2*logLik(model)+2*(i+1)*3071/(3071-i-2))
+ }
'log Lik.' -9171.798 (df=2)
'log Lik.' -9833.732 (df=3)
'log Lik.' -10004.88 (df=4)
'log Lik.' -10141.06 (df=5)
'log Lik.' -10207.76 (df=6)
'log Lik.' -10222.82 (df=7)
'log Lik.' -10243.8 (df=8)
'log Lik.' -10295.32 (df=9)
'log Lik.' -10301.77 (df=10)
'log Lik.' -10325.1 (df=11)
```

```
> model <- garch(x,c(1,1),trace=F)
Warning message:
In sqrt(pred$e) : NaNs produced
> print(-2*logLik(model)+2*(2+1)*3071/(3071-2-2))
'log Lik.' -10357.17 (df=3)
```

AICc selects the GARCH (1,1) model.

```
> logLik(model)
'log Lik.' 5181.589 (df=3)
> summary(model)

Call:
garch(x = x, order = c(1, 1), trace = F)

Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max 
-6.47008 -0.42388 -0.03494  0.39375  7.56096 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0 5.884e-05   3.152e-06    18.67   <2e-16 ***
a1 2.348e-01   6.691e-03    35.09   <2e-16 ***
b1 7.931e-01   3.758e-03   211.04   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the summary of the model, all the parameters are statistically significant,

The complete form of the fitted model is $h_t = 5.844 \times 10^{-5} + 0.2348 \times \varepsilon_{t-1}{}^2 + 0.7931 \times h_{t-1}$.

The unconditional variance of the shocks cannot be computed by formula a0/(1-(a1+b1)) because the sum of a1 and b1 is greater than 1.

The sample variance of residuals is 0.0045.

```
> var(x)
[1] 0.004508095
```

The 95% one step ahead ARIMA-GARCH forecast interval for the log closing price is ($f_{t,1}$+1.96*sqrt(ht+1), $f_{t,1}$-1.96*sqrt(ht+1)).

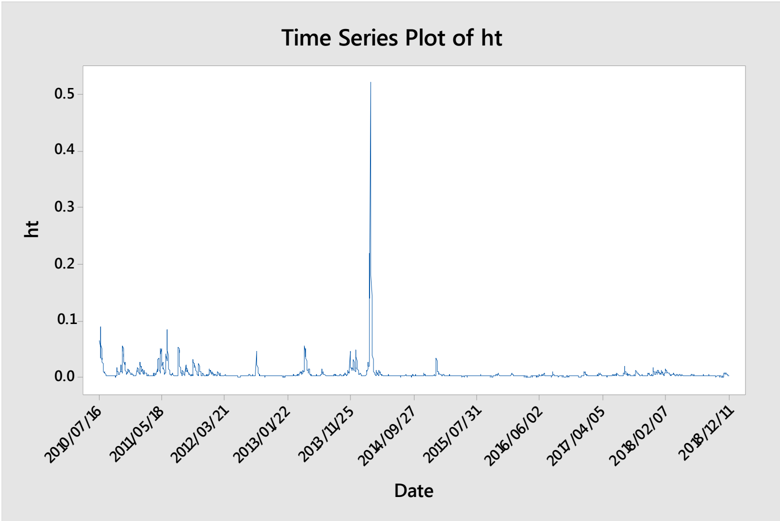
$h_{t+1} = 5.844 \times 10^{-5} + 0.2348 \times \varepsilon_t{}^2 + 0.7931 \times h_t = 0.0025$

The 95% forecast interval is (8.0035, 8.1995), which is narrower than ARIMA-only 95% one-step forecast interval.

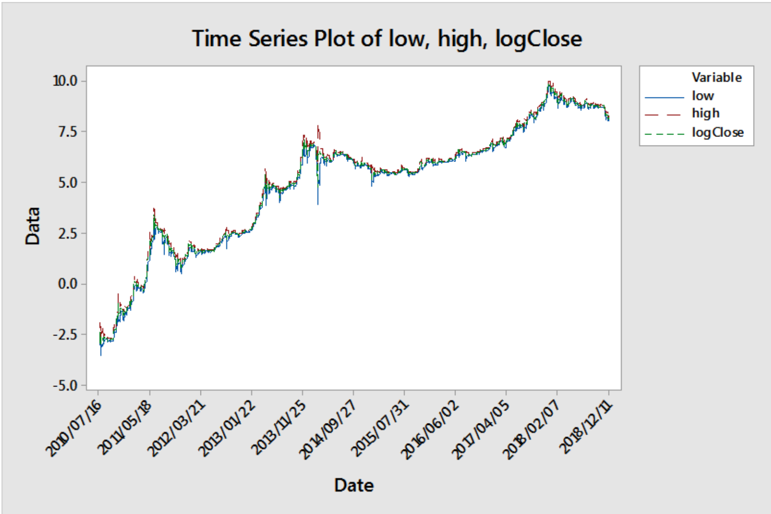
```
> ht = model$fit[,1]^2
> ht1 <- 5.844*10^(-5)+0.2348*x[3071]*x[3071]+0.7931*ht[3071]
> ht1
[1] 0.002501156
> 8.1015+1.96*sqrt(ht1)
[1] 8.199523
> 8.1015-1.96*sqrt(ht1)
[1] 8.003477
```

The 5th percentile of the conditional distribution of the next period's log closing price is $8.1015 - 1.645 \times \sqrt{h_{t+1}} = 8.0192$

```
> 8.1015-1.645*sqrt(ht1)
[1] 8.019231
```

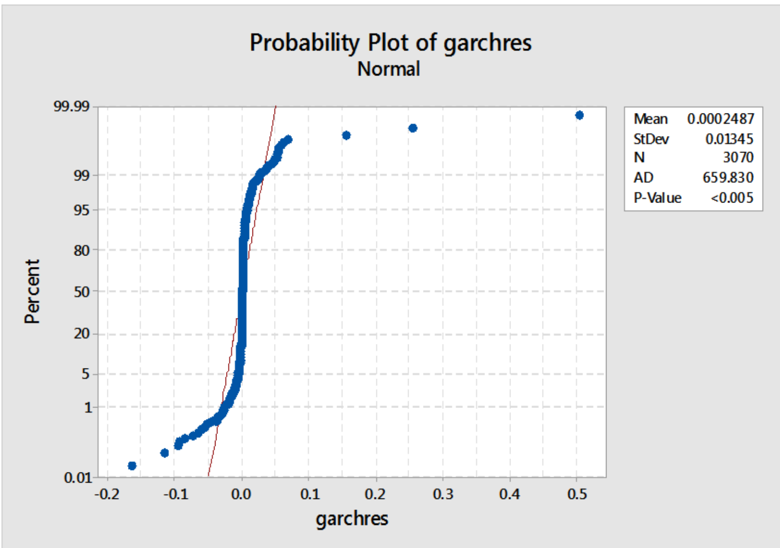


Compare the plot of ht and return, the highly volatile periods of ht agree with the highly volatile periods of log closing price.



From this time series plot, the forecast intervals can predict the log of closing price since the log of closing price doesn't exceed our forecast intervals significantly.

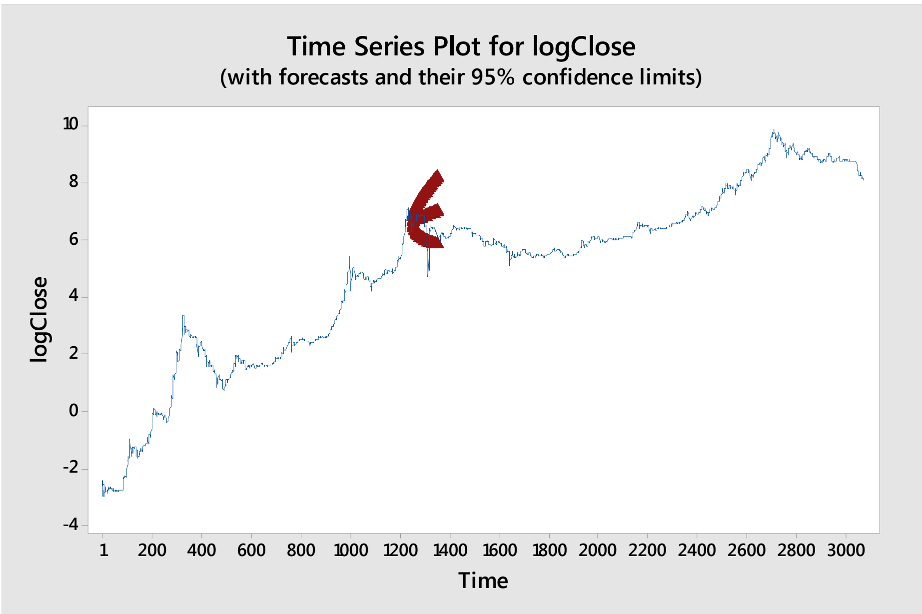
Next, compute the residuals of GARCH model.



Anderson-Darling Normality Test	
A-Squared	130.48
P-Value	<0.005
Mean	-0.4274
StDev	8.6538
Variance	74.8881
Skewness	-0.4290
Kurtosis	12.9996
N	3070
Minimum	-72.0914
1st Quartile	-3.2242
Median	-0.1806
3rd Quartile	2.4676
Maximum	78.6174

From normality plot and normality test, the residuals do not follow a normal distribution, and the high kurtosis adequately describes the long tailedness of the data.

sum (abs(garchres)>1.96) = 0
The intervals fail 0% of the time.



If we construct an ARIMA (2,1,2) model at observation 1250 and obtain 30-step-ahead forecast intervals, we can see that the ARIMA forecast intervals are too narrow. From the time series plot of “logClose, low and high”, the ARIME-GARCH model can adapt to the changing volatility, and ARIMA-GARCH forecast intervals are more useful than ARIMA intervals because residuals haven’t exceeded 1.96 in absolute value so far.

Finally, check the effectiveness of one step ahead forecast interval. The closing price of a bitcoin on December 14th, 2018 is 3,235.48. The log of 3235.48 is 8.0819. This observation falls into both ARIMA forecast interval and the ARIMA-GARCH interval. Both intervals are effective on December 14th, 2018.