Introduction

The project will use the stock data of China Construction Bank (0939.HK), which the closing price of the stock.

The dataset has 246 observation in total. It will be using the data of closing price to find out the suitable ARIMA model and to estimate the parameters of the models in this project. Also,

After the calculation and the process of model identification part, an AR(1) model may be suitable for the dataset. The data used will be shown below with the residuals when the AR(1) model are fitted to the data .

Close	AR(1)-Residuals	Close	AR(1)-Residuals	Close	AR(1)-Residuals
6.27	-0.00571	7.07	-0.01041	6.96	-0.04162
6.29	0.017123	6.98	-0.08071	6.91	-0.04239
6.41	0.117427	6.99	0.017919	6.91	0.006855
6.42	0.009252	7.08	0.098071	6.99	0.086855
6.42	-0.0006	7.12	0.04944	6.93	-0.05193
6.57	0.149404	7.12	0.010048	6.96	0.037159
6.57	0.001685	7.08	-0.02995	6.99	0.037615
6.58	0.011685	6.89	-0.18056	6.82	-0.16193
6.49	-0.08816	6.95	0.066551	6.78	-0.03451
6.6	0.110468	7.06	0.117463	6.65	-0.12512
6.57	-0.02786	7	-0.05086	6.44	-0.2071
6.56	-0.00832	6.97	-0.02178	6.47	0.029708
6.63	0.071533	7	0.037767	6.33	-0.13984
6.69	0.062597	7.05	0.058223	6.3	-0.03196
6.65	-0.03649	7.07	0.028983	6.33	0.027579
6.67	0.022901	7.02	-0.04071	6.29	-0.04196
6.7	0.033205	6.95	-0.06147	6.29	-0.00257
6.82	0.123662	6.94	-0.00254	6.31	0.017427
6.88	0.065486	6.81	-0.12269	6.29	-0.02227
6.92	0.046399	6.8	-0.00467	6.23	-0.06257
6.94	0.027007	6.86	0.065182	6.25	0.016515
7.01	0.077311	6.82	-0.03391	6.21	-0.04318
7	-0.00162	6.73	-0.08451	6.19	-0.02379
6.95	-0.04178	6.9	0.174118	6.15	-0.04409
6.91	-0.03254	6.88	-0.0133	6.2	0.045299
6.95	0.046855	6.92	0.046399	6.2	-0.00394
6.94	-0.00254	6.96	0.047007	6.18	-0.02394
6.99	0.057311	7.01	0.057615	6.14	-0.04425
6.98	-0.00193	6.99	-0.01162	6.15	0.005147
6.8	-0.17208	6.92	-0.06193	6.18	0.025299
6.9	0.105182	6.87	-0.04299	6.31	0.125755
6.85	-0.0433	6.93	0.066247	6.39	0.077731

6.89	0.045942	6.91	-0.01284	6.31	-0.08105
6.94	0.056551	7.18	0.276855	6.32	0.007731
6.94	0.007311	7.16	-0.00904	6.33	0.007884
7.14	0.207311	7.06	-0.08934	6.35	0.018036
7.09	-0.03965	7.01	-0.04086	6.42	0.06834
Close	AR(1)-Residuals	Close	AR(1)-Residuals	Close	AR(1)-Residuals
6.54	0.119404	5.73	0.018456	5.99	0.002714
6.6	0.061229	5.64	-0.10109	5.92	-0.07713
6.57	-0.02786	5.64	-0.01246	5.95	0.021801
6.59	0.021685	5.71	0.057544	5.96	0.002257
6.51	-0.07801	5.73	0.008608	6	0.03241
6.57	0.060773	5.77	0.028912	6.13	0.123018
6.68	0.111685	5.73	-0.05048	6.22	0.084994
6.73	0.053357	5.77	0.028912	6.26	0.036363
6.39	-0.33588	5.78	-0.00048	6.28	0.016971
6.31	-0.08105	5.79	-0.00033	6.31	0.027275
6.3	-0.01227	5.75	-0.05018	6.26	-0.05227
6.27	-0.03242	5.78	0.019216	6.27	0.006971
6.19	-0.08288	5.8	0.009673	6.3	0.027123
6.14	-0.05409	5.77	-0.04002	6.29	-0.01242
6.14	-0.00485	5.82	0.039521	6.35	0.057427
6.21	0.065147	5.81	-0.01972	6.29	-0.06166
6.25	0.036211	5.74	-0.07987	6.31	0.017427
6.27	0.016819	5.89	0.139064	6.31	-0.00227
6.25	-0.02288	5.99	0.091345	6.23	-0.08227
6.22	-0.03318	6.03	0.032866	6.31	0.076515
6.19	-0.03364	6.03	-0.00653	6.36	0.047731
6.25	0.055907	6.03	-0.00653	6.48	0.118492
6.16	-0.09318	6.2	0.163474	6.5	0.020316
6.18	0.015451	6.17	-0.03394	6.45	-0.04938
6.17	-0.01425	6.24	0.065603	6.56	0.10986
6.19	0.015603	6.12	-0.12333	6.47	-0.08847
6.2	0.005907	6.02	-0.10516	6.27	-0.19984

6.13	-0.07394	6	-0.02668	6.31	0.037123
6.15	0.014994	5.96	-0.04698	6.21	-0.10227
6.07	-0.0847	5.97	0.00241	6.18	-0.03379
6.04	-0.03592	5.96	-0.01744	6.21	0.025755
5.93	-0.11637	5.94	-0.02759	6.31	0.096211
5.78	-0.15805	5.9	-0.04789	6.42	0.107731
5.71	-0.08033	5.96	0.051497	6.4	-0.0206
5.71	-0.01139	5.92	-0.04759	6.27	-0.1309
5.71	-0.01139	5.98	0.051801	6.33	0.057123
5.7	-0.02139	5.98	-0.00729	6.4	0.068036

Close	AR(1)-Residuals
6.4	-0.0009
6.42	0.0191
6.4	-0.0206
6.23	-0.1709
6.23	-0.00348
6.2	-0.03348
6.17	-0.03394
6.2	0.025603
6.29	0.086059
6.3	0.007427
6.31	0.007579
6.37	0.057731
6.4	0.028644
6.51	0.1091
6.46	-0.04923
6.59	0.130012
6.63	0.041989
6.64	0.012597
6.62	-0.01725
6.68	0.062445
6.63	-0.04664
6.76	0.132597

6.75	-0.00543
6.73	-0.01558

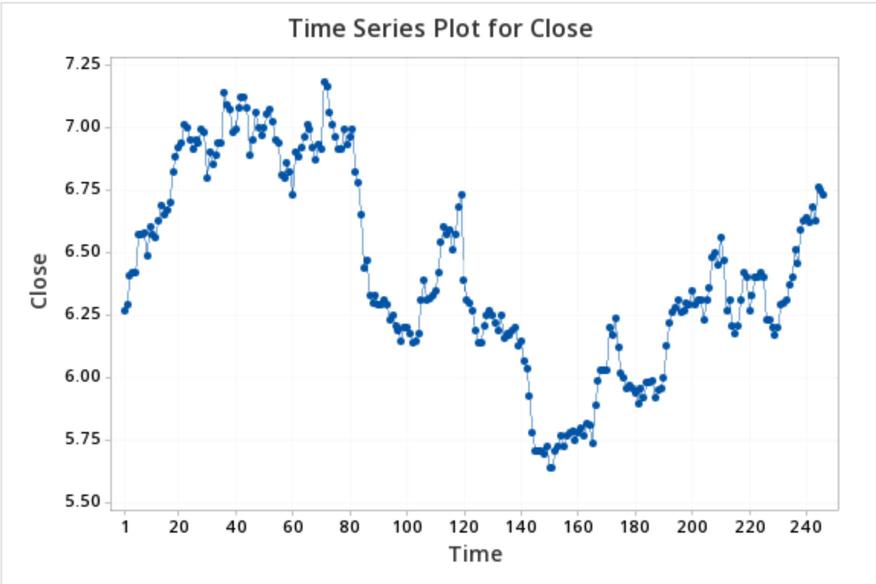
Methodology

I mainly use Minitab to assist with the progress of mini project. I will compare the correlogram and the partial correlogram of the data by Minitab to identify the reasonable ARIMA model and estimate the parameters based on the selected ARIMA model. Besides, I will check the residuals of whether they are consistent with white noise or not.

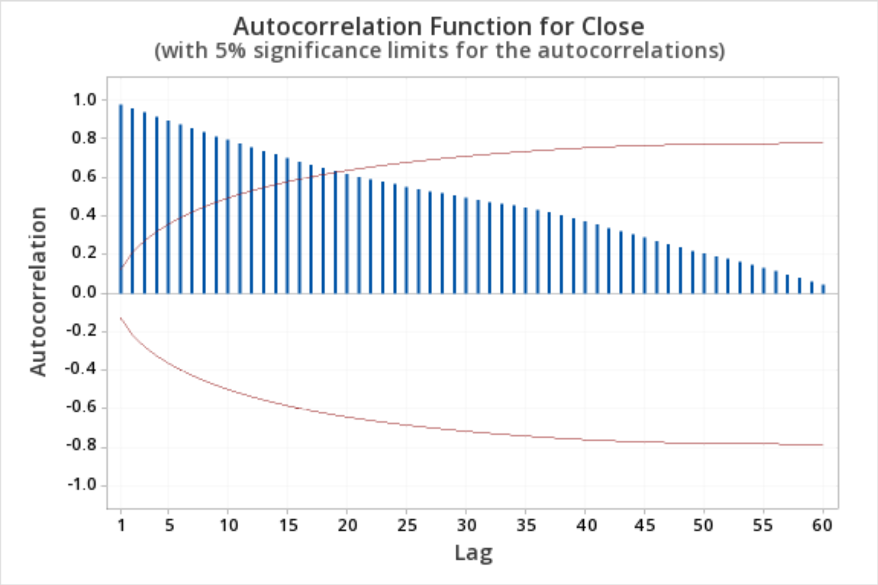
Discussion

<u>Identify the model</u>

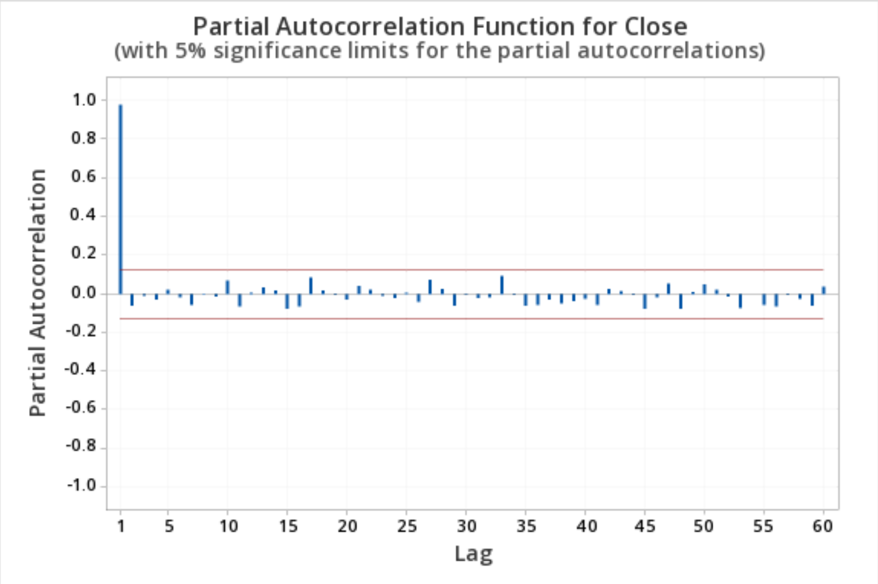
The time series plot shows that there is a non-zero mean in the data that we will discuss later. Therefore, we know there will be a constant term in the model.



Then, plotting the ACF and PACF correlogram can help us to identify the ARIMA model. The figure shows below which has the ACF curve for the closing price. By the convergent, we will use $\pm\frac{2}{\sqrt{n}}$ to identify the conventional error. However, we use the 5% significance limits which is the red lines in the below figure instead of $\pm\frac{2}{\sqrt{n}}$ to be the boundary at this time. We will assume the values are significantly small when the values lie within the red lines. We can see the decaying pattern in the ACF of below correlogram.



In the PACF, we can see an cut off after lag 1. At lag 1, it has large value in magnitude and all the remaining lag will be within the red lines in the below figure. The redlines also recognize that the values are significantly small if the values lie within the boundaries.



By the ACF and the PACF correlogram, AR(1) model may be suitable as the PACF cuts off after lag 1 and ACF dies down exponentially in two correlogram.

<u>Fitted ARIMA model</u>

By using iteration method to fit the ARIMA model, which is AR(1) at this case, we can see various information from the below. We can see that the SSE will be stable after 9 iteration, which means that it does not contain any significant improvement to the SSE after that and the parameters in the iteration 9 can be treated as the parameters with smallest SSE in the model. Therefore, 0.985 and 0.098 will be the estimates of the parameters will be minimizing the SSE.

Estimates at Each Iteration

Iteration	SSE	Parameters	
0	33.5167	0.100	5.876
1	23.5208	0.250	4.896
2	15.3773	0.400	3.917
3	9.0856	0.550	2.938
4	4.6454	0.700	1.958
5	2.0562	0.850	0.979
6	1.3104	0.977	0.153
7	1.3071	0.984	0.105
8	1.3070	0.985	0.099
9	1.3070	0.985	0.098

Relative change in each estimate less than 0.001

The coefficient in the AR(1) model will be 0.9848 and the constant term is 0.09821 in the statistical information of final estimates of parameters below. The mean of x is 6.459 which is the non-zero mean we mentioned before. We can also see the standard error to the coefficients and the t test statistics by the value of the coefficient over the value of standard error. All the p-value are significantly small as the p-value is less than the significance level of 0.05. Therefore, it is significant for the coefficient estimate.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.9848	0.0118	83.71	0.000
Constant	0.09821	0.00468	20.99	0.000
Mean	6.459	0.308		

We can see that the SSE is 1.3070 and the MSE is 0.0053524 when the degree of freedom is 244 as 2 parameters in the model.

Number of observations: 246

Residual Sums of Squares

DF	SS	MS
244	1.30598	0.0053524

Back forecasts excluded

<u>Residuals Analysis</u>

Although the sample size is 246 which is not belongs to small size, using Ljung-Box test instead of Box-Pierce test can give more precise results. Therefore, we will use the Ljung-Box test to conclude the result of the residuals at last.

We will check whether there is any correlation still remaining in the residuals by Ljung-Box test. We will use the ACF values up to lag12 to do the test and we can see that the Chi-square test statisics is 4.19 with the degree of freedom, which is 10 (lags minus the numbers of pararmetors that are estimated) in the below figure.

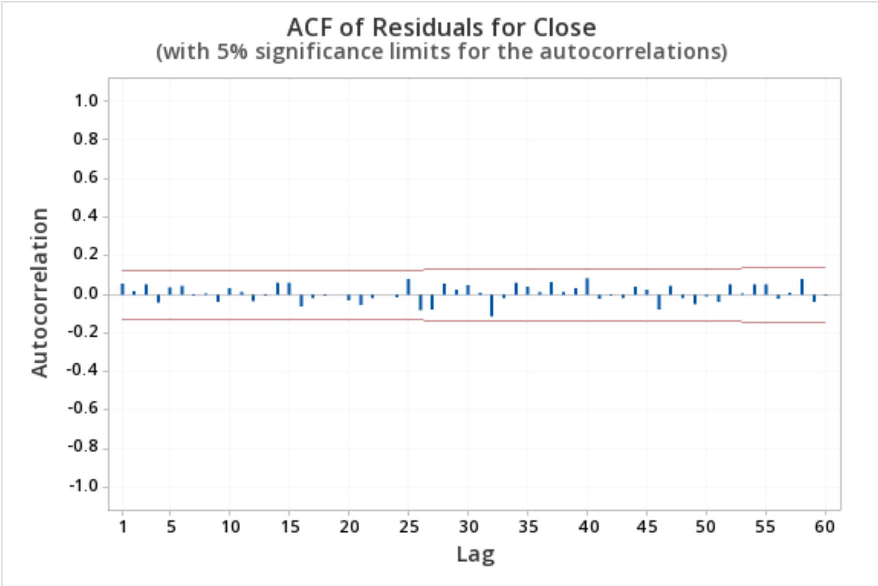
H_0: the residuals are uncorrelated (white noise);
H_1: there are autocorrelations in the residual series.
Q_{LB} (up to twelve terms) $= 4.19$.
As $\chi^2_{10,0.05} = 18.307$, the LB statistic is not significant at 5% level. Since $Q_{LB} = 14.64990319$ is smaller than 18.307, we do not reject null hypothesis at $\alpha = 0.05$ and the residuals are uncorrelated which also means the estimated residual series is consistent with white noise.

The p-value from the distribution up to lag 12 is 0.939 which is greater than 0.05 and none of the correlation for the autocorrelation function of the residuals are significant. It means that residuals are independent.

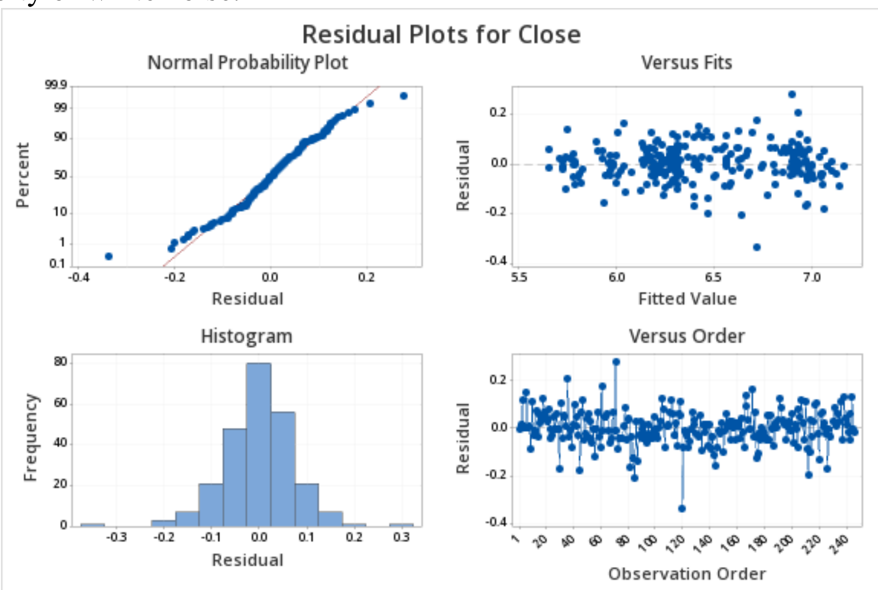
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	4.19	8.33	20.90	28.36
DF	10	22	34	46
P-Value	0.939	0.996	0.962	0.981

With the ACF of the residuals, we can see that all the values lied within the red lines in the correlogram of residuals below and we can treat it as zero.



Finally, we can figure out the supporting evidence whether the residual series is consistent with white noise by using the residual plots for closing price which are shown below. In the plot of Residual Versus Fits, the points are concentrated around zero which shows that it tends to have zero mean. Also, we cannot see any special pattern in this plot too. We can see the frequency of residual in the histogram that it forms a bell-shape and tends to have normal distribution too. All the information fit the property of white noise.



Conclusion

After the model identification process, AR(1) model may be suitable for the data. The model will be like

$$X_t = \mu + \phi_1 X_{t-1} + \ a_t = 0.09821 + 0.9848X_{t-1} + \ a_t.$$

In Ljung-Box test, the test statistic is $Q_{LB} = 14.64990319$ at the significate level which means that the residuals are uncorrelated, and the estimated residual series is consistent with white noise.