Forecasting Time Series Project 1

Yoon Tae Park(yp2201@nyu.edu)

Source of the data:

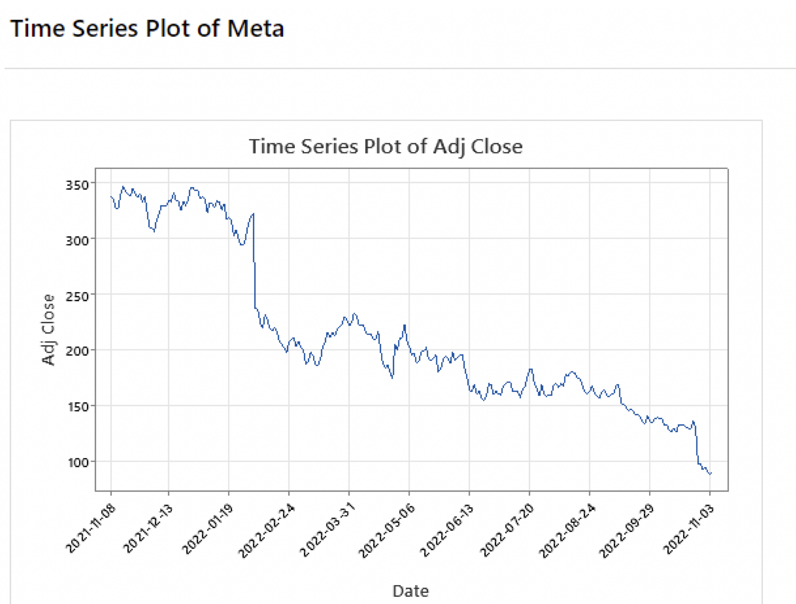
META 1 year (21/11/08~22/11/04)

https://finance.yahoo.com/quote/META/history?period1=1636156800&period2=1667692800&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true

Start by plotting the data

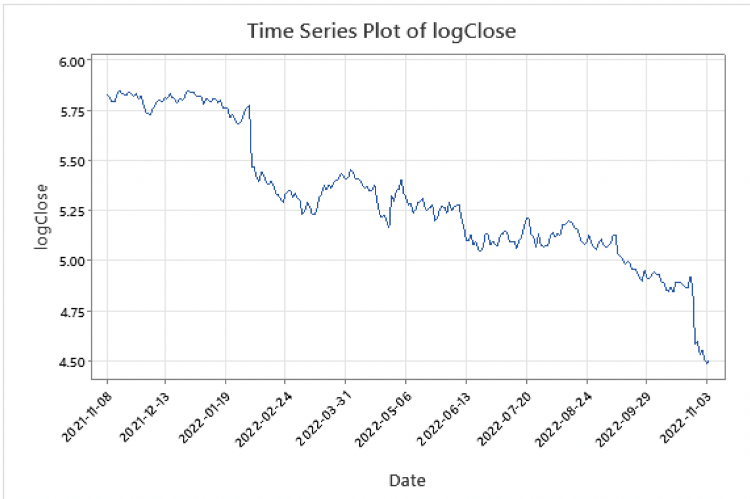
Here, I've plotted a time series plot of 1) adj close price, 2) log adj close price, 3) diff close price, 4) diff log close price.

[Time series of Meta adj close price]



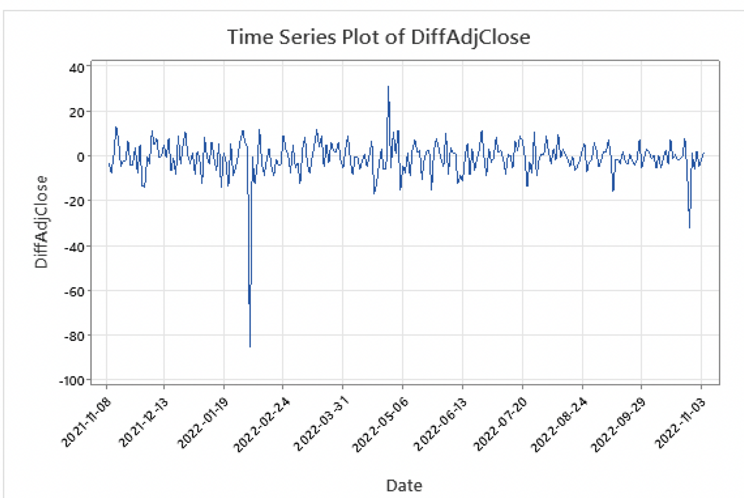
[Time series of Meta log adj close price]

Time Series Plot of logMeta



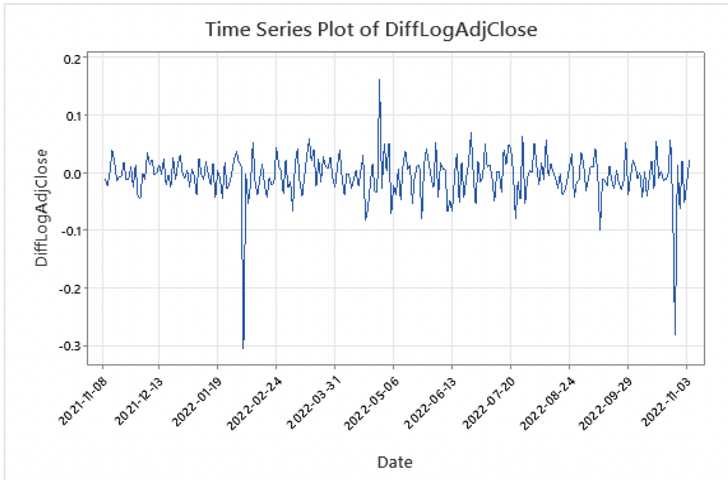
[Time series of Meta difference adj close price]

Time Series Plot of DiffAdjClose



[Time series of Meta difference log adj close price]

Time Series Plot of DiffLogAdjClose



Briefly describe any patterns you see

First, I've checked the Meta Adj close price series and didn't find any strong stationary or mean-reverting situation. Log Meta Adj close price series showed similar patterns, as adj close price series don't show any exponential increase/decrease. Both plots show a constant decrease of the prices, evidenced by the recent economic recession.

Then, I took the 1st difference, and figured out that the difference of both Meta adj close price and log Meta adj close price series are stationary and mean-reverting.

Also, as shown in the plots, there is no clear seasonality.

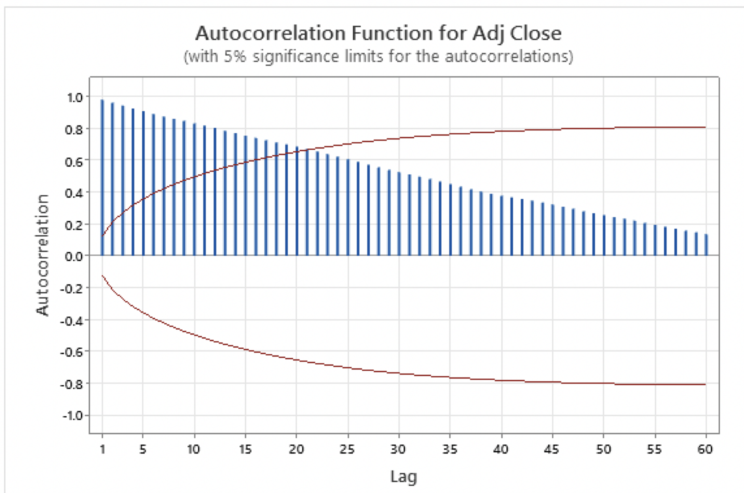
Decide whether you need to take logs.

I didn't take logs, as the stock price itself doesn't show any exponential increase/decrease. Also, as looking at the level of dependency of volatility on both difference of stock price and log stock price, both plots don't show a big difference. Rather, the difference of stock price shows less level of dependency of volatility. Therefore, I didn't take logs.

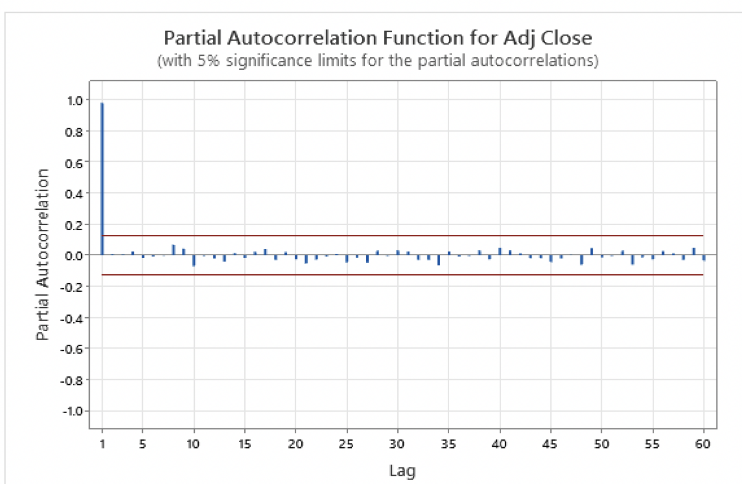
Use the ACF, PACF and AICC to select an ARIMA model, and to decide whether a constant term should be included in the model. Present the parameter estimate printout for the selected model.

I've checked ACF and PACF of Meta adj close price series to see any statistically significant lag point. Here, I've found that there is statistically significant at lag 1 for PACF, and ACF shows lie down, given the difference = 0. (Which makes AR(1), given d=0)

[ACF of Meta stock adj close price, d=0]

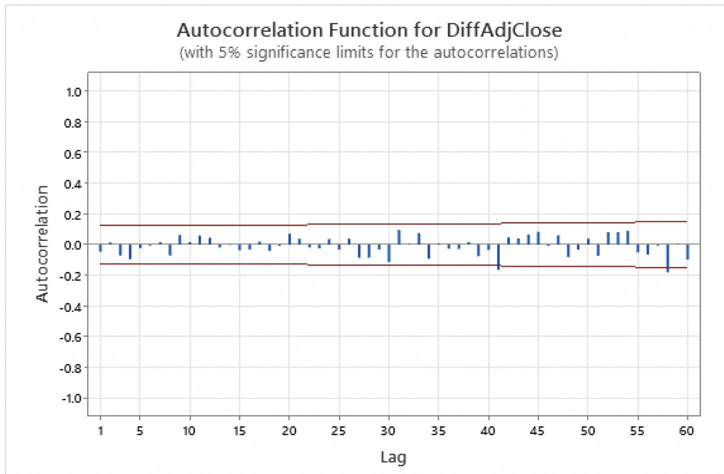


[PACF of Meta stock adj close price, d=0]

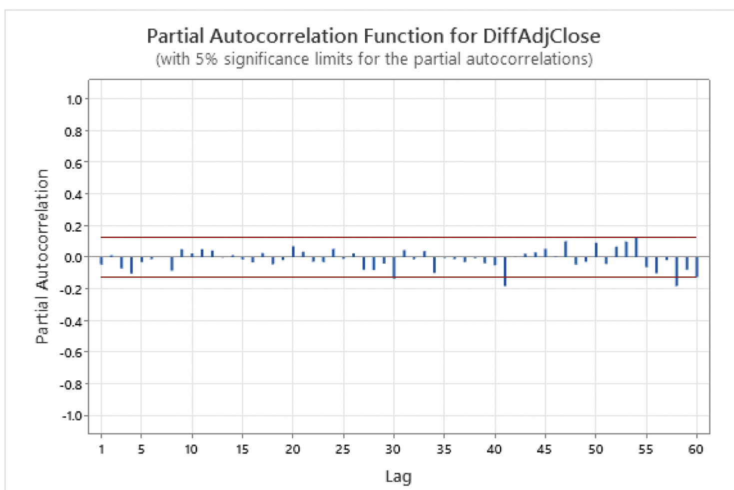


I’ve also checked ACF and PACF for the difference of Meta adj close price. I’ve found that both PACF and ACF don't have any statistically significant lags, meaning that data could be a white noise, and maybe a random walk model can be applied, given the difference = 1.

[ACF of Meta stock adj close price, d=1]



[PACF of Meta stock adj close price, d=1]



However, we can’t conclude any ARIMA model by just combining AR and MA models. Therefore, I’ve created various combinations of ARIMA models and compared AICc. (p = 0, 1, 2/ d = 0, 1 / q = 0, 1, 2). I also compared the cases with a constant and without a constant.

Based on the AICc table, I found that ARIMA (0,1,0) with a constant shows the lowest AICc (1784.99) among other ARIMA models. However, I've selected the next-best non-ARIMA(0,1,0) model as Minitab 21 doesn't really provide sufficient information on random walk models. For example, ARIMA(0,1,0) with constant doesn't have any p-values for model parameters. Therefore I've selected **ARIMA(1,1,0) with constant** as the best model, showing next best AICc score (1786.53)

[AICc, d = 1, With constant]

Model Selection

Model (d = 1)	LogLikelihood	AICc	AIC	BIC
p = 0, q = 0*	-890.472	1784.99	1784.94	1791.99
p = 1, q = 0	-890.214	1786.53	1786.43	1796.99
p = 0, q = 1	-890.219	1786.54	1786.44	1797.00
p = 1, q = 1	-889.479	1787.12	1786.96	1801.04
p = 2, q = 0	-890.203	1788.57	1788.41	1802.49
p = 0, q = 2	-890.218	1788.60	1788.44	1802.52
p = 1, q = 2	-889.475	1789.20	1788.95	1806.56
p = 2, q = 1	-890.161	1790.57	1790.32	1807.93
p = 2, q = 2	-891.820	1793.89	1793.64	1811.25

[AICc, d = 1, Without constant]

Model Selection

Model (d = 1)	LogLikelihood	AICc	AIC	BIC
p = 0, q = 0*	-892.150	1786.32	1786.30	1789.82
p = 1, q = 0	-892.025	1788.10	1788.05	1795.09
p = 0, q = 1	-892.031	1788.11	1788.06	1795.10
p = 1, q = 1	-891.936	1789.97	1789.87	1800.44
p = 2, q = 0	-891.949	1790.00	1789.90	1800.46
p = 0, q = 2	-891.958	1790.01	1789.92	1800.48
p = 1, q = 2	-891.875	1791.91	1791.75	1805.84
p = 2, q = 1	-891.897	1791.96	1791.79	1805.88
p = 2, q = 2	-891.820	1793.89	1793.64	1811.25

For deciding to include a constant term, I've checked the p-value of the model, and found that both p-values are not statistically significant, which is higher than 0.05. While taking p-values as a reference, we should be more focusing on the AICc scores, as we are not actually doing hypothesis testing, and therefore cannot be a determinator on decision. When comparing the AICc scores, ARIMA(1,1,0) with constant shows a better AICc score compared to ARIMA(1,1,0) without constant. Therefore, I've included a constant term.

[Best model (ARIMA(1,1,0) with constant) parameters]

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.0458	0.0634	-0.72	0.472
Constant	-1.037	0.541	-1.92	0.056

[Reference: ARIMA(0,1,0) with constant not showing any p-values]

Final Estimates of Parameters

Random walk model with constant = -0.991320

Write the complete form of the fitted model.

Ans) $x_t = -0.0458x_{t-1} + \varepsilon_t - 1.037$, *where* x_t *= diff(Meta adj close price)*

For the selected model, comment on the Ljung-Box statistics, plot the residuals and the ACF and PACF of the residuals. Comment briefly on any problems revealed by this diagnostic checking.

The model seems to be adequate. We can see that all p-values are greater than 0.05 at all Chi-Square lag points from Ljung-Box statistics. ACF and PACF residual shows no statistically significant lags, showing that the residual is a white noise and therefore providing good evidence that my best model well fits given data points.

On this diagnostic checking, we can only conclude that the model is adequate if p-values at all Chi-Square lag points are greater than 0.05. So, I can say that my model is adequate based on the diagnostic checking. However, I cannot further analyze which model is more adequate or appropriate, given different p-values that are greater than 0.05. For now, we just rely on AICc scores to select the best model, and this diagnostic checking is a sanity check, which is limited to checking the minimum adequacy of the model .

[Ljung-Box statistics]

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

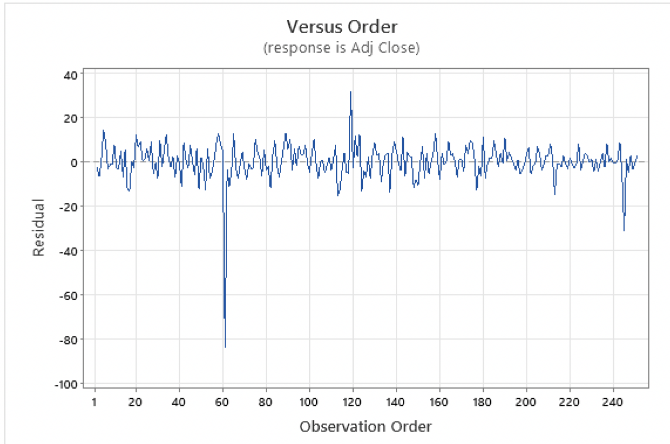
Lag	12	24	36	48
Chi-Square	8.23	11.99	27.42	46.20
DF	10	22	34	46
P-Value	0.606	0.958	0.781	0.464

[Residual sum of squares]

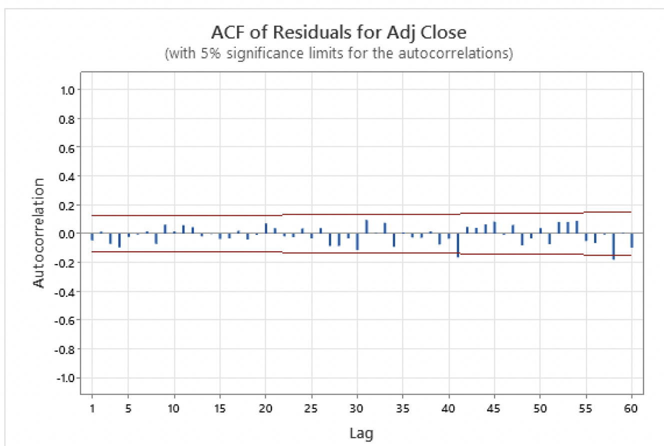
Residual Sums of Squares

DF	SS	MS
248	18128.8	73.1001

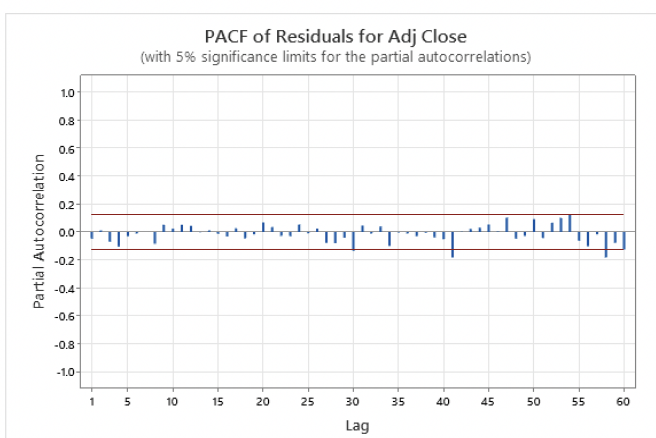
[Time series of Residual plots]



[ACF of the residuals]



[PACF of the residuals]



Plot the data, together with the forecasts (at lead times 1-50 or further, if possible), and the 95% forecast intervals. Comment briefly on whether the forecasts seem reasonable, and on whether the forecast intervals seem excessively wide.

The forecast seems generally reasonable as a model is showing a slightly negative slope from the AR1 parameter and having a negative constant, while having a difference of the data. Therefore, the forecast showing negative downward slope makes sense.

The forecast intervals seem excessively wide as the lower bound is less than zero, which ignores the assumption that the stock price is at least greater than 0. (While negative values make sense if we think of this as a mathematical equation)

This happens as the 95% forecast intervals get wider as the lead time gets increased, and makes a rough prediction for further forecasts. This can be explained as we are only given relatively short lead time compared to the entire time the data is spread.

[Plotting data with forecast at lead time 50 with 95% forecast intervals]