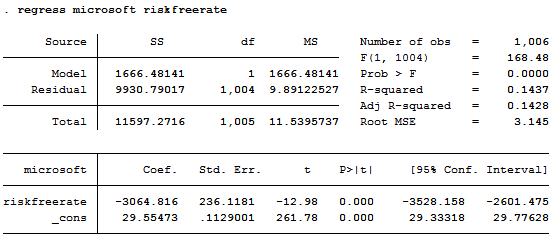
MN-M585: Applied Data Analytics

Student Name

Student ID

**Section 1**

**Question 1**



From the above table, the CAPM model is:

* Y= 29.55 -3064.82R where 29.55 (Constant) is the Risk free rate and -3064.82 is beta value.
* Aggressiveness and defensiveness of a stock depends on the beta value. Here, securities that have beta >1 are very risky and thus referred to aggressive stocks while those with beta less than 1 are less risky and thus called defensive stocks.
* In this case, since beta= -3064.82, then the Microsoft stocks are defensive.

**Question 2**

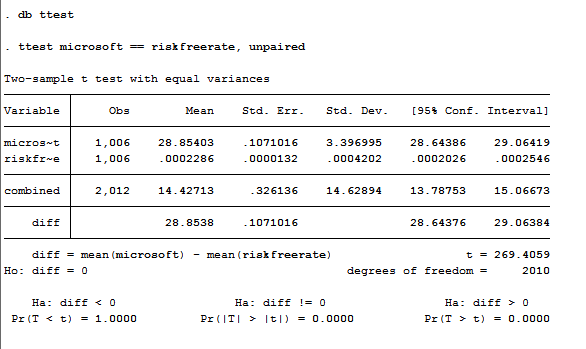
The R squared value also coefficient of determination is 14.37% which is approaching 0. This is a clear indication of the weak variability between the dependent variable- Microsoft stock prices and the independent variable- Microsoft risk free rate and this implies that the regression model’s coefficients do not explain all of the variability around the mean making this model a poor model.

**Question 3**

Ho: Microsoft stock has same volatility as the market portfolio

H1: Microsoft stock does not have the same volatility as the market portfolio

In this case, I will use a two sample t-test assuming unequal variances (two sided test) and below is the table results.



The p value of interest is one in the middle because it indicates a two sided test. The p value is 0 which is below the alpha value 0.05 and thus we reject the null hypothesis and conclude that Microsoft stock does not have the same volatility as the market portfolio.

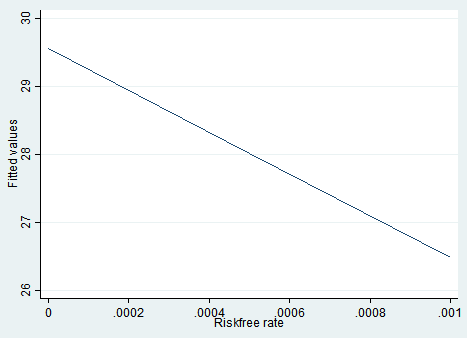
**Question 4**

Some of the assumptions of CLRM - Classical Linear Regression Models include:

* The regression model is linear in parameters, and not linear in variables.
* The explanatory variable is considered non-stochastic
* E ()= 0, Expected value (mean)= 0, Expected mean of the residuals is zero.
* Variance () or E () = s^2, Constant variance, No heteroscedasticity in the residuals or errors
* E (t≠t-1, No Autocorrelation; Errors / Residuals are not correlated with their previous values.
* E ( 0, The error is uncorrelated with the explanatory variable

**Testing for Linearity**

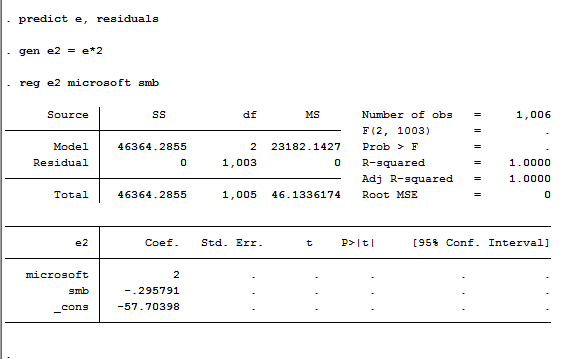
The following is the scatter plot and a line plot of the data.



The line plot shows that the Microsoft prices are positively related to the market prices or volatilities as indicated by the positive line. No violation of this property.

**Test for Heteroscedasticity**

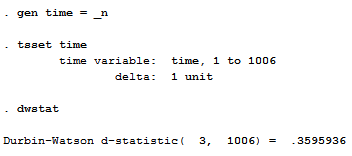
Used to test and determine if the residuals’ variances are constant or not.



There is no value for the Prob > F and thus we can conclude that there is no heteroscedasticity. Thus there is no violation of this property.

**Test for Autocorrelation**

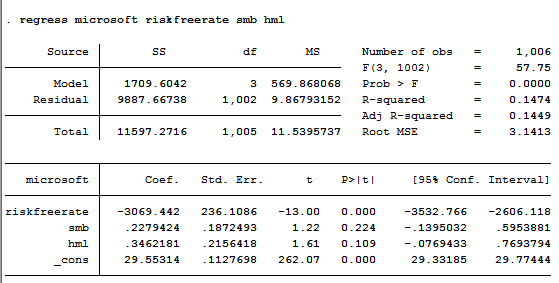
I will use a Durbin Watson statistic. The statistics ranges from 0 to 4. 0<2 value indicates a positive autocorrelation. >2 to 4 indicates negative autocorrelation. 2 indicates no autocorrelation. Below is the stata table result for the same.



From the table, DW d-statistic is 0.36 indicating a positive autocorrelation among the residuals. Therefore, the model violates this assumption.

**Question 5**

* The Fama-French 3-factor model is a CAPM which expands on capital asset pricing model by adding size risk and value risk factors to the market risk factor. The three factors used are SMB- Small Minus Big, HML-High Minus Low and the portfolio’s return less risk free rate return.
* Therefore, the two factors that I will add to my CAPM model will be HML and SMB and the following is the STATA output for the same.



The new CAPM model will be:

Y= 29.55 – 3069.442R + 0.23SMB+ 0.34HML

From the model, the addition of SMB and HML do not affect the risk free rate of the model but it does slightly affect the beta value by reducing it compared to the original model.

**Question 6**

I will select the Fama-French model, Y= 29.55 – 3069.442R + 0.23SMB+ 0.34HML. This is because first, the inclusion of the two extra variables does not affect the risk free rate. HML is used to indicate the value premium or it shows the spread in returns between value stocks and growth stocks. A positive HML value of + 0.34 indicates that Microsoft have high value stocks which shows a good performance and shows the excess returns of Microsoft due to its high book-to-market equity value. The positive SMB value shows that Microsoft stocks are weighted towards small-cap stocks.

Therefore, these variables help in determining the size of the Microsoft stocks’ risk as well as the value of the risk.

**Question 7**

For the goodness of fit, let use the value of R squared and the F statistic. For the new model, R ^2 is 14.74. This shows a weak variability between the independent and dependent variables indicating that the coefficients of the model do not explain all of the variability around the mean making this model a poor model. The F statistic is 57.75 which is greater than 1 and this shows that the model generally is a good fit.

Comparing this model to the CAPM one, R squared value is the same and thus no improvement here. The F value was 168.48 which is greater than the new second model. Thus the CAPM model is more a good fit compared to the Fama-French model by considering the F statistic value.

**Section 2**

**Question 1**

Time series data is a historical data in a given time period. This data is helpful in forecasting future values and mostly used in financial analysis. In this section, I chose Alibaba historical stock prices data from Oct 06, 2014 to Dec 11, 2019. This historical data can be used in determining a suitable model that can be employed and used in forecasting future Alibaba stock prices. Some of the well-known forecasting models are ARMA, ARIMA, ARIMA-GARCH.

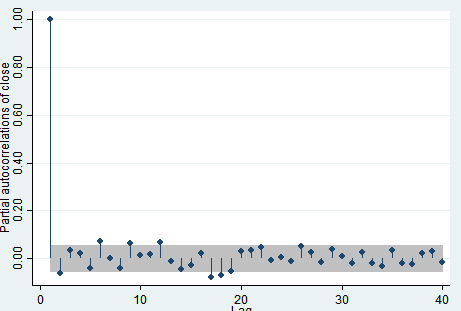
**Question 2**

One cannot fit ARMA model to the daily close prices series. First, the series data have to be converted into time series version before any STATA analysis the series by running the code:

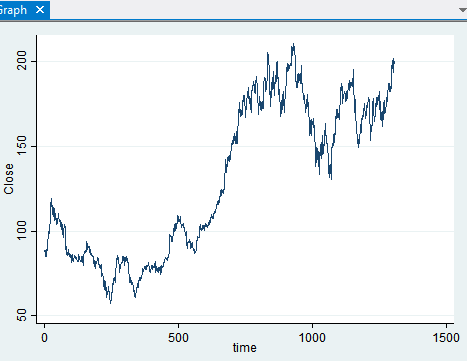
gen time = \_n

tsset time

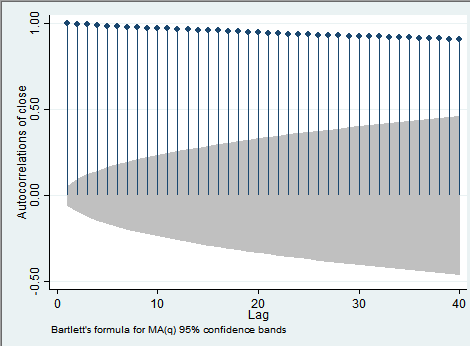
In determining the ARIMA (p, d, q) model, the coefficients p, d, q have to be determined first. To determine the value of p, I plotted Partial Correlogram (pac) graph and determined the first lag or lags outside the confidence bound as shown below. As my order, I chose the first lag hence p=1.



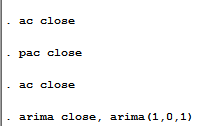
To determine d, I identified the number of times I have to differentiate in order to have a stationary data. In this case, I did not have to differentiate thus d=0 and below is the graph.

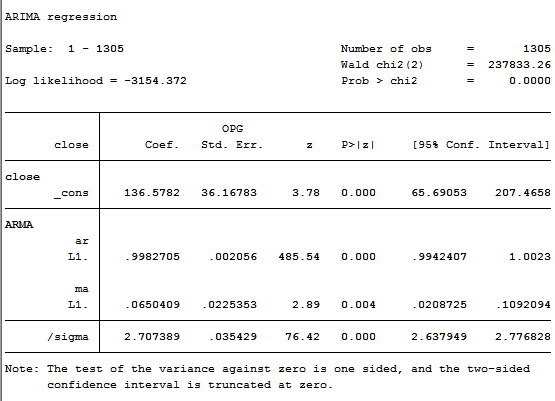


To determine q, I used the Correlogram plot and determined the lags outside the confidence bounds. There were many lags outside the bound and I started with the first one and thus q= 1.



Therefore, the ARIMA/ARMA model in this case is ARIMA (1,0,1).

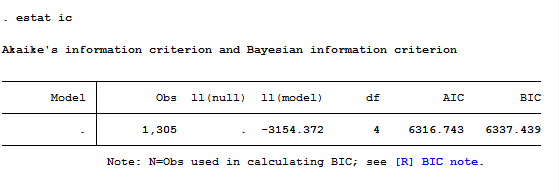




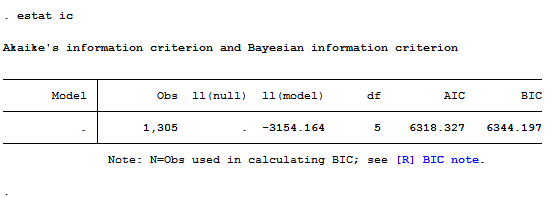
**Question 3**

The following is the AIC and BIC for the various suggested ARIMA models.

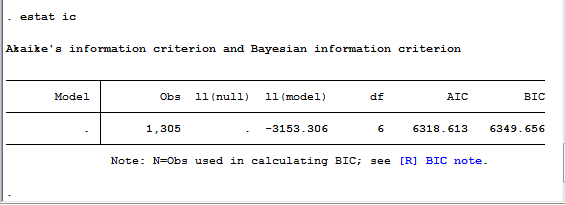
The following is for ARIMA (1,0,1)



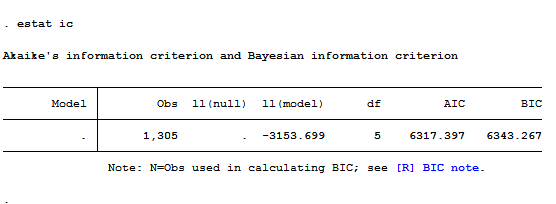
The following is for ARIMA (2,0,0).



The following is for ARIMA (3,0,0).



The following is for ARIMA (3,0,1).

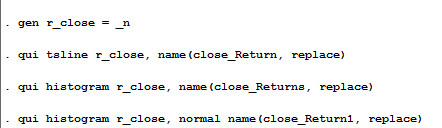


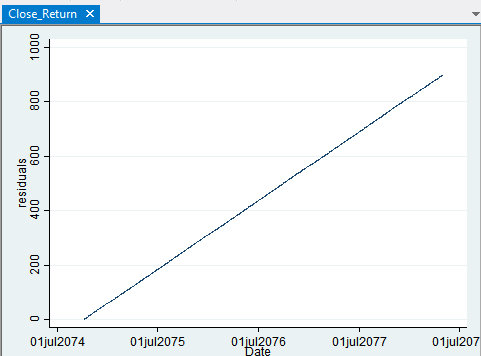
When interpreting the AIC and BIC, either a lower AIC and BIC value indicates a good fit. In most cases, AIC value is used. In this case, ARIMA (1,0,1) has the least values of AIC and BIC and thus ARIMA (1,0,1) is our best model.

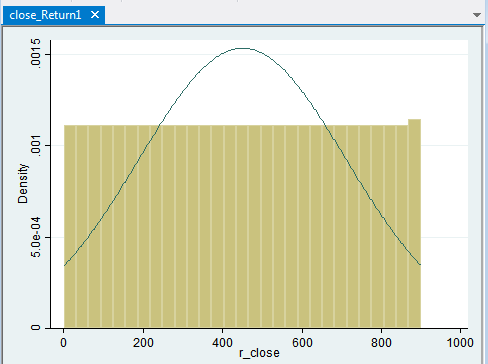
**Question 4**

This will involve first testing for the presence of heteroscedasticity in the residuals as well as the test if the residuals follow a normal distribution. I will also run a test to check if there is presence of ARCH effects or not.

First, let run a test for checking the presence of heteroscedasticity and if the residuals are normally distributed and below is the table results with the codes.

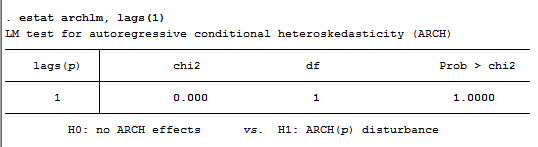






The first graph shows that there is no heteroscedasticity due to the linear line graph, if there was heteroscedasticity, then the graph would resemble that of stationary time plot. The second graph also indicates that the residuals are not normally distributed because as one can see, the histogram bars are level.

Let confirm further by running the presence of ACRCH effects test. Below is the STATA code and table results.



P value is 1 and thus we fail to reject our null hypothesis and conclude that there are no ARCH effects.

Therefore, this data is not fit for GARCH model.

**Question 5**

Because the data does not contain ARCH effects, then it is not fit for GARCH models but rather, ARIMA model will do the best.

**Question 6**

Since we cannot apply GARCH Models test to the data, then forecasts and volatility tests using either of the GARCH models will be difficult.

