**Group Assignment 1**

**Presented by: Qi (Daniel) ZHENG Student ID: G44426724**

**Xinyue ZHAO Student ID: G27620695**

**Sizhe WU Student ID: G26890251**

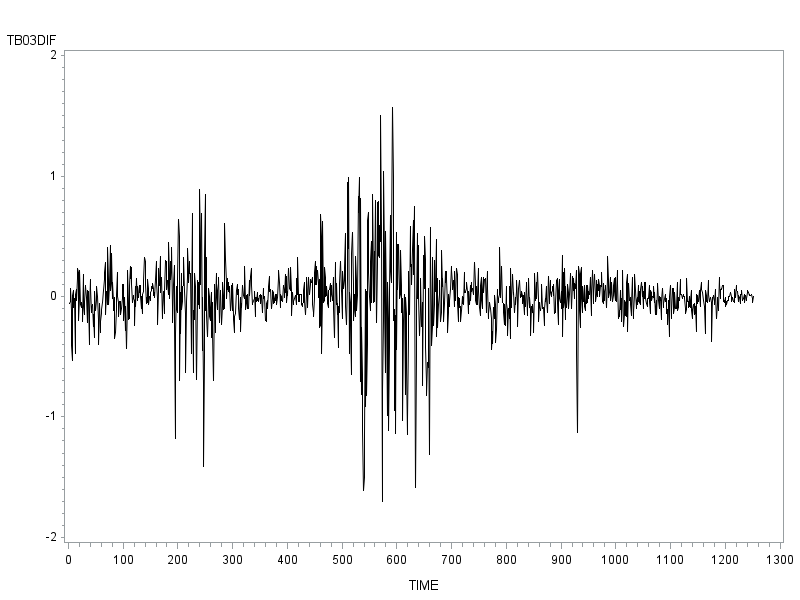
**Ruibo WANG Student ID: G24256043**

**Yingyu LIN Student ID: G48092483**

**Question 1**

**Time series graph for weekly changes:**

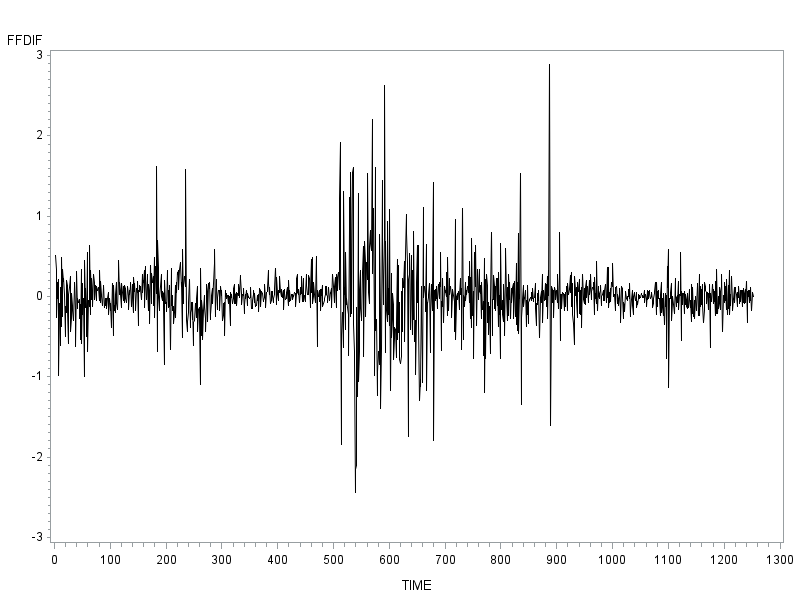
Graph 1: FFDIF VS TIME Graph 2: TB03DIF VS TIME



P1

P2

P3

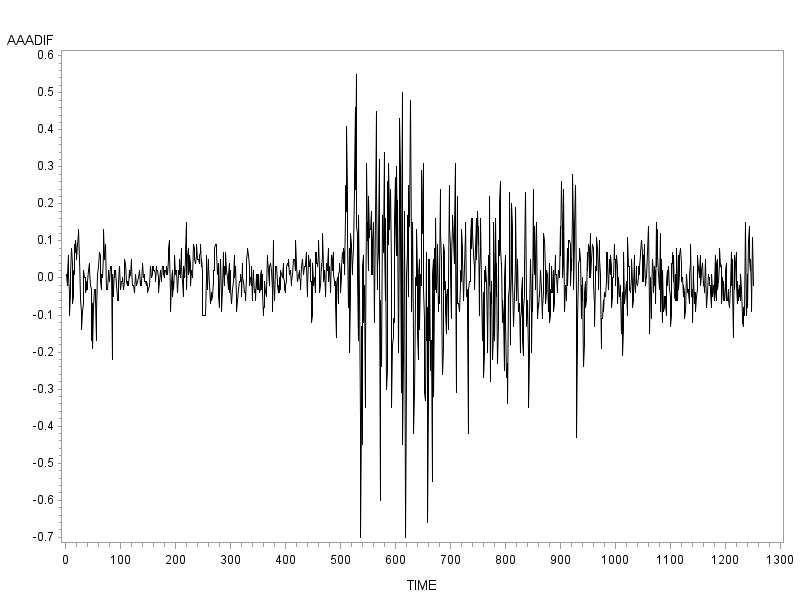


P1

P2

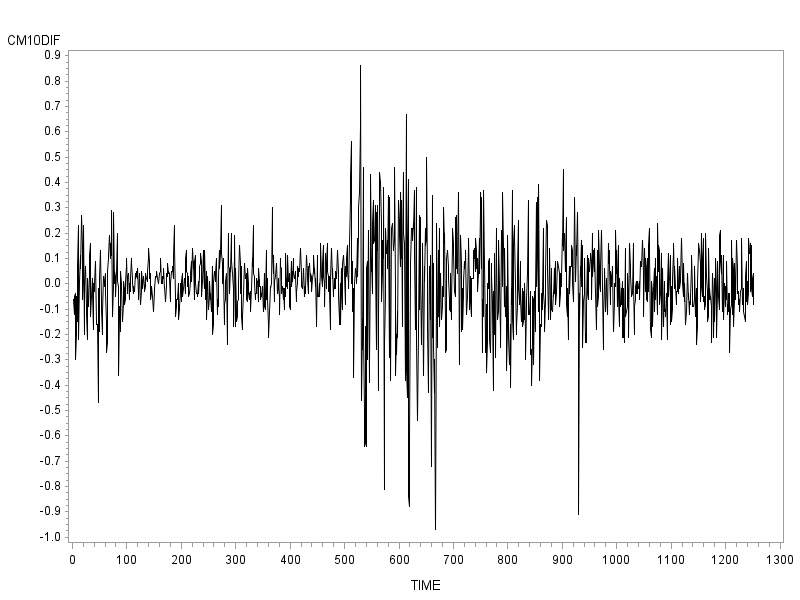
P3

Graph 3: CM10DIF VS TIME Graph 4: AAADIF VS TIME



P2

P3

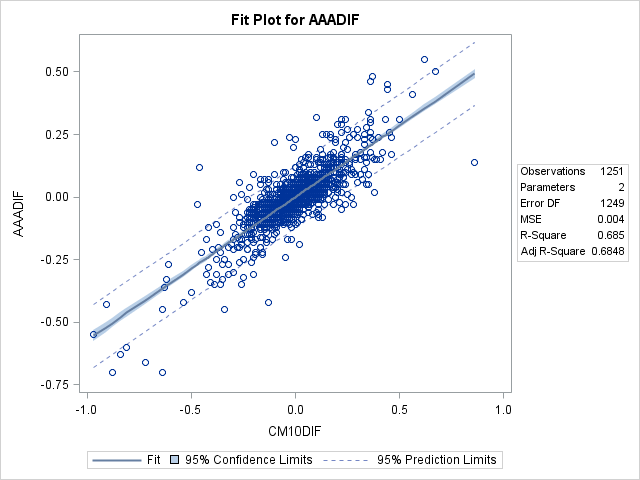


P2

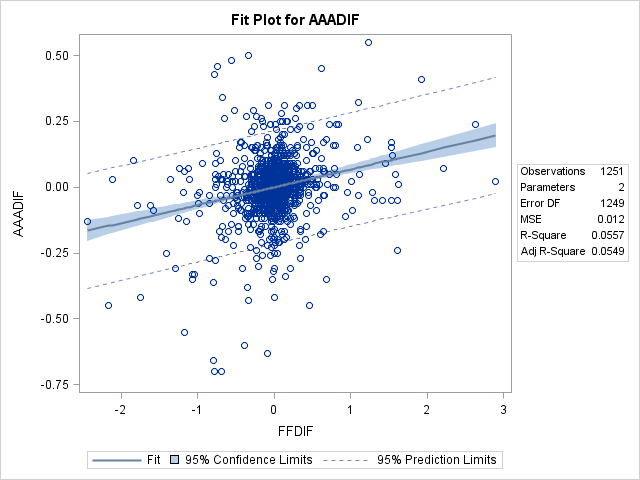
P3

Across the time period, there are three highly volatile regions identifiable in the first-order differential plots, their peaks marked P1, P2, and P3. However, effect of P1 is only present in Graph 1 and Graph 2.

**Scatter plot:**

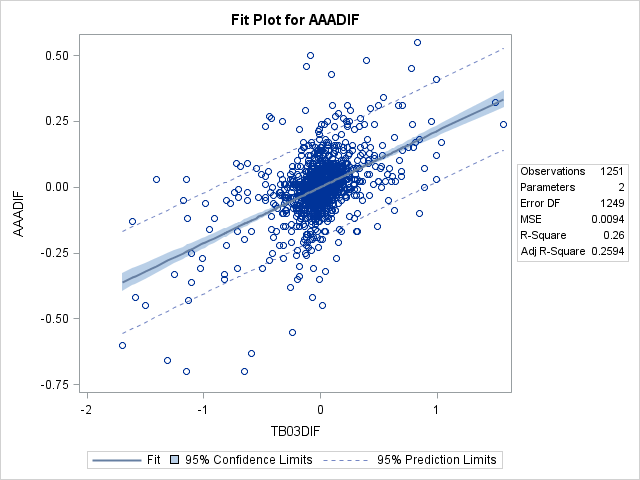
Graph 1

By observing the scatter plot, we can conclude that the graph is positively correlated, that is, the larger values of the horizontal variable CM10DIF (independent variable) are positively associated with larger values of the vertical variable AAADIF (dependent variable), and vice versa. Also, we would like to mention that the sample points are highly concentrated in middle part of the graph. In terms of the outliers, we did observe individual points that fall outside the prediction interval, which could have influence on correlation.

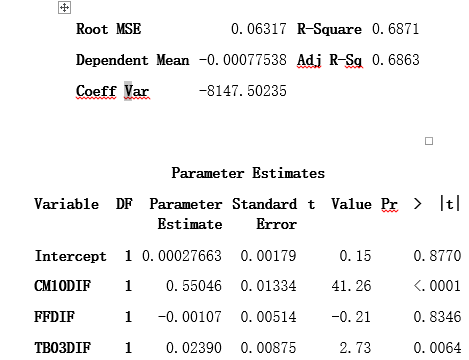
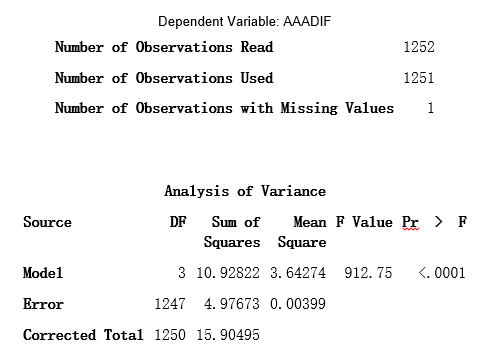
Graph 2

By observing the scatter plot, we can conclude that the data behave in a positively correlated way, which means the larger values of the FFDIF (independent variable) the larger values of the AAADIF (dependent variable), and vice versa. However, the correlation here is not as strong as the correlation between AAADIF and CM10DIF in the first graph.

Graph 3

By observing the scatter plot, we can conclude that the data behave in a positively correlated way, which means the larger values of the TB03DIF (independent variable) the larger values of the AAADIF (dependent variable), and vice versa. However, the correlation here is not as strong as the between AAADIF and CM10DIF in the first graph, but is slightly stronger than the correlation between AAADIF and FFDIF in the second graph.

**Question 2**



**Estimated regression equation:**

Y= 0.00027663 + 0.55046∙CM10DIF -0.00107∙FFDIF+0.02390∙TB03DIF

The parameter estimate of CM10DIF, namely the slope of regression line between AAADIF and CM10DIF, is 0.55046, which means every unit increase of CM10DIF will increase 0.55046 unit of DV (AAADIF), holding FFDIF and TB03DIF constant.

**Question 3**:

Hypothesis testing for **CM10DIF** α = 0.05

H0: β1 = 0

Ha: β1 ≠ 0

Observing from the parameter estimates table, we rejected the null hypothesis (H0) for CM10DIF at 95% confidence level since p (β1 = 0) < α (= 0.05); therefore, there is significant evidence to prove that β1 is different from zero. Hence we concluded that β1 for CM10DIF is a significant coefficient for the regression model.

Hypothesis testing for **FFDIF** α = 0.05

H0: β2 = 0

Ha: β2 ≠ 0

Observing from the parameter estimates table, we retained the null hypothesis (H0) for FFDIF at 95% confidence level since p (β2 = 0) >α (= 0.05); therefore, there is no significant evidence to prove that β1 is different from zero. Hence we concluded that β2 for FFDIF is not a significant coefficient for the regression model, and can thus be replaced by a zero.

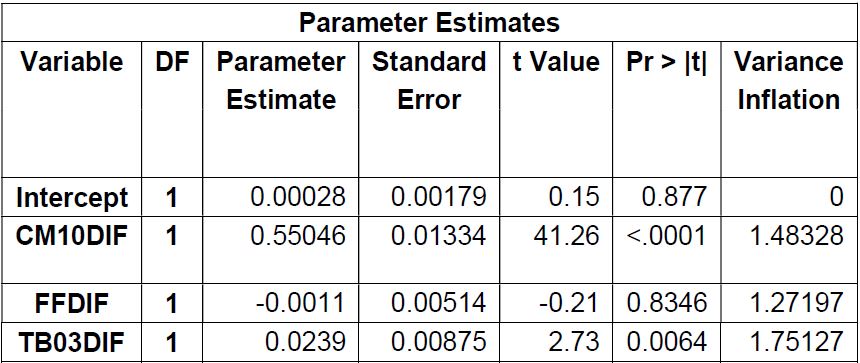
Hypothesis testing for **TB03DIF**  α = 0.05

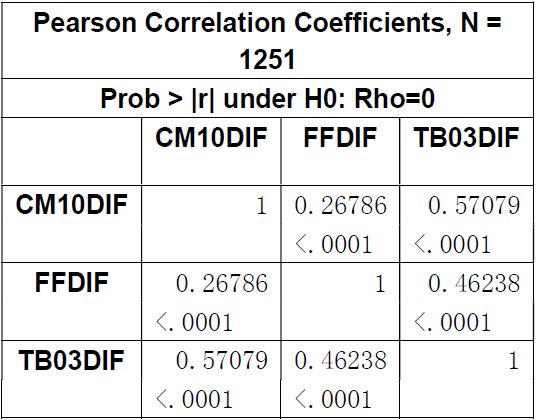
H0: β3 = 0

Ha: β3 ≠ 0

Observing from the parameter estimates table, we rejected the null hypothesis (H0) for TB03DIF at 95% confidence level since p (β3 = 0) <α (= 0.05); therefore, there is significant evidence to prove that β3 is different from zero. Hence we concluded that β3 for TB03DIF is a significant coefficient for the regression model.

**Question 4: Multicollinearity**

According to the variance inflation factors (VIF’s) from the table to the left, we obtained VIF1 = 1.48328, VIF2 = 1.27197, and VIF3 = 1.75127, all of which smaller than 10; therefore, we concluded there is no significant evidence to suggest that multicollinearity exists. In other words, there is no indication of multicollinearity between any combinations of the independent variables.

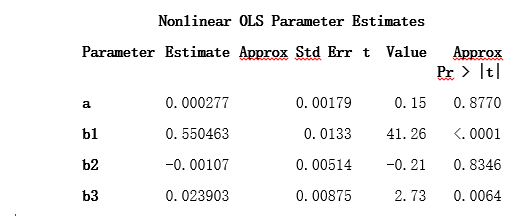


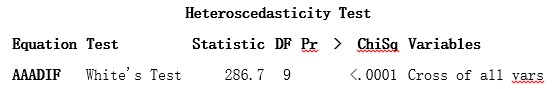
Also this conclusion has been verified by the correlation table provided to the left:

As we can observe, all the correlations between 2 variables in this table are relatively small and trivial, with the largest correlation coefficient 0.57079 (between TB03DIF and CM10DIF), and the smallest 0.26786 (between CM10DIF and FFDIF); therefore, there is no highly correlated variables. Hence, it is less likely to observe an indication of multicollinearity in this data set.

It is also worth mentioning that correlation is not the most reliable measurement of multicollinearity, but the VIF is.

**Question 5: Heteroskedasticity**





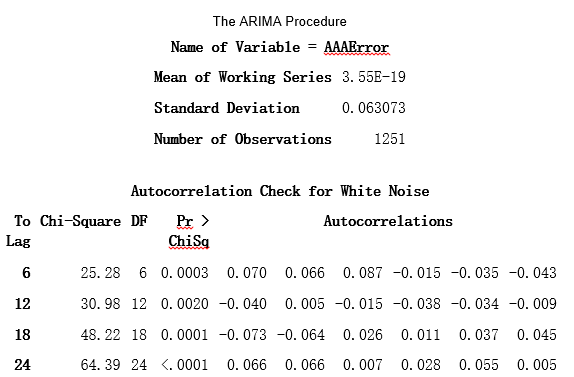
White’s test α = 0.05

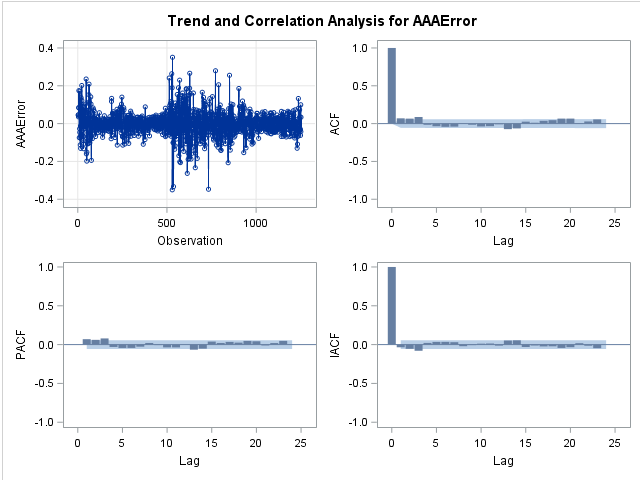
H0: homoskedasticity (constant variance for all residuals)

Ha: heteroskedasticity (non-constant variance for all residuals)

Observing from the heteroscedasticity test table, we rejected the null hypothesis (H0) for White’s at 95% confidence level since statistic nR2 (= 286.7) >> critical value = 16.92, and the p-value (<0.0001) << α (=0.05). Therefore, there is significant evidence to prove heteroskedasticity. Hence the variance of error term does change with the independent variables.

**Question 6:**





H0: ρk = 0 ( k=1, 2, …, 24);  α = 0.05

Ha:  i, s.t. ρi ≠ 0 (i=1, 2, …, 24);

Observing from autocorrelation table, we can reject the null hypothesis for residual at 95% confidence level since the p-value (< 0.0001) <α (= 0.05); therefore, there is significant evidence to prove that at least one ρ is not equal to 0. Hence we concluded that residuals are auto-correlated at some lag (up to 24 lags).