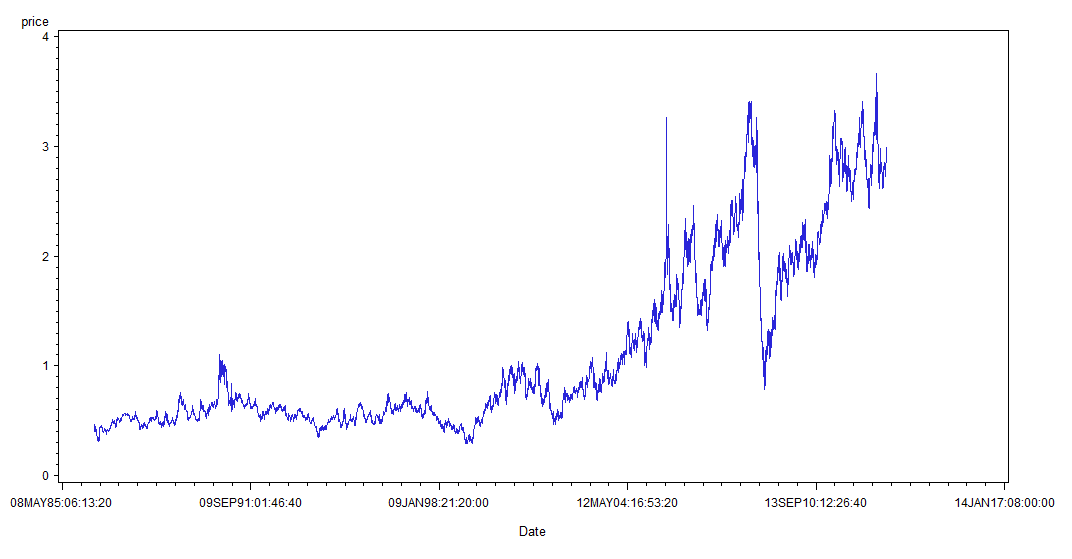
**Problem 1 [16 points] (review week 4 notes)**

You are asked to analyze the daily conventional gasoline spot prices (in dollars per gallon) Gallon). Gasoline price have impact on daily life. High prices results in high transportation and heating costs, and hence, higher prices for food and services. The data file oilprice.csv contains two variables: date and price, and they are obtained from the US Energy Information Administration website at <http://www.eia.gov>.

***a. Create a time plot of gasoline spot prices and discuss the trends over time.***

ANSWER:



Generally, the price is increasing in trend with fluctuation around short period. In year 2004, the price broke the up limit, 1 dollar more or less, for itself. Since then, the price has moved around 1 and 3.5 dollars per gallon with a sharp plummet in year 2008. Now, the price is stable at 3 dollars per gallon (Why the price in Chicago is 4 dollars! I hate the damn tax!! Really!!)

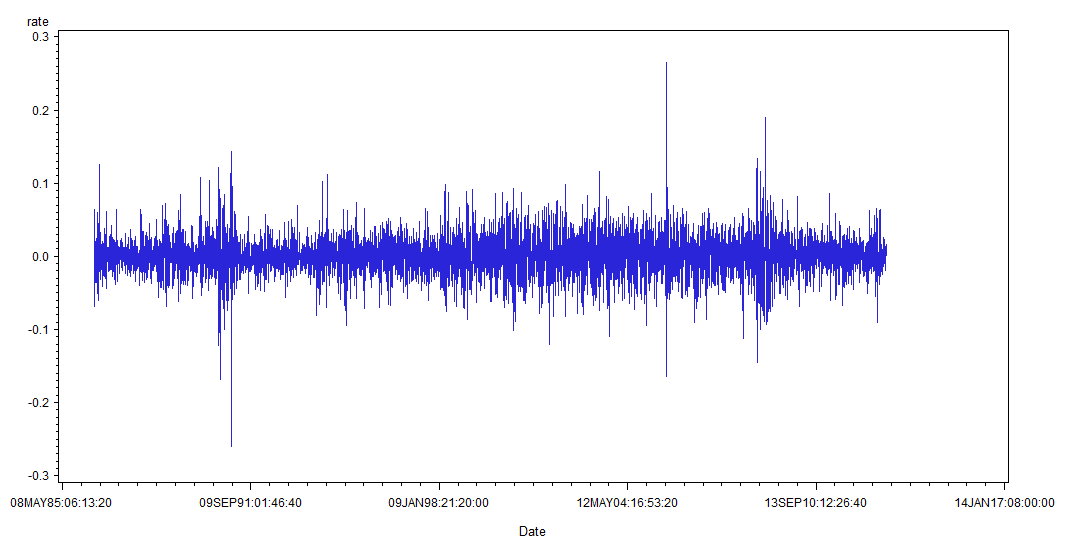
***b. Compute the new time series for price changes\_ rate = (pt-pt-1)/pt-1.***

ANSWER:

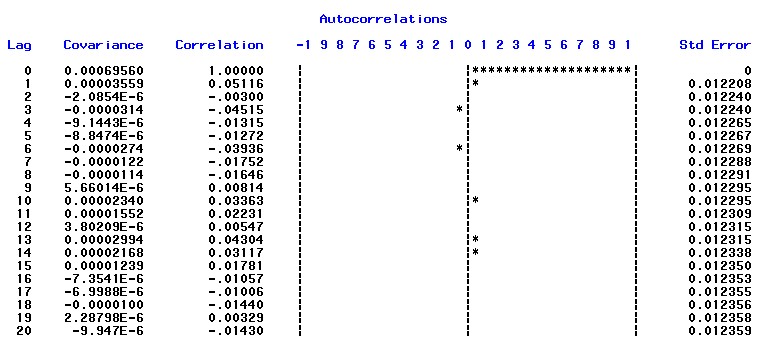
In the code!

***c. Analyze the dynamic behavior of the price changes using the time plot of RATE, and its autocorrelation values (ACFs) (20 lags).***

ANSWER:



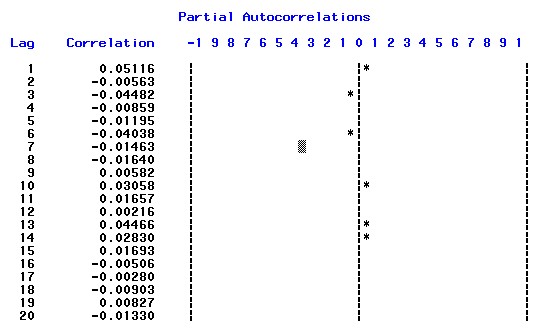
From the time plot of rate, the mean of the rate is around 0.0. The variance fluctuates within 0.1 to -0.1, and there are many outliers outside.

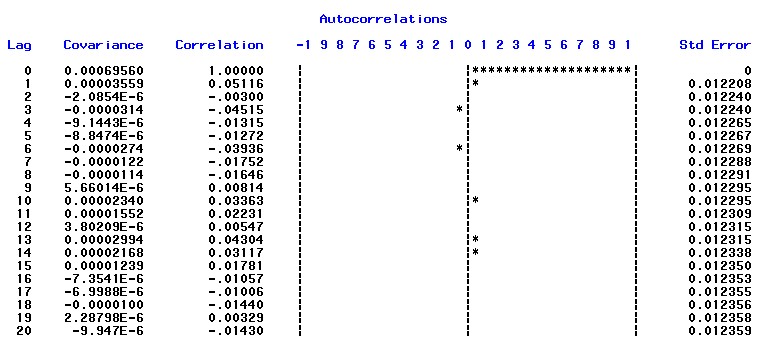


The ACF graph does not show the two standard error lines. However, by checking the number, we could see that lag-1, lag-3, lag-6, lag-10, lag-13 and lag-14 have ACF values that are larger than two standard errors, though not very significantly different than two standard errors.

***d. Analyze the ACF and the PACF functions and determine which model is more appropriate to model this time series. Does the process show AR effects or MA effects?***

ANSWER:





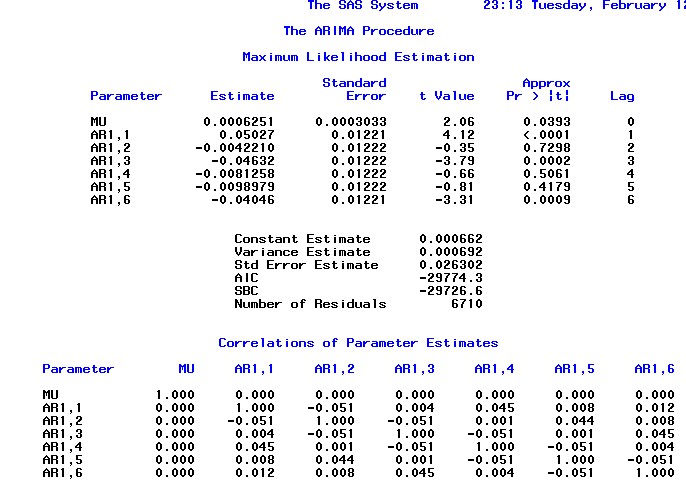
From the ACF and PACF plots, we could see that there are even no two standard error lines on them, I think maybe it is because the value of two standard error is very small.

If I was forced to pick a model, I would choose a AR or MA model with 1, 3, 6, 10, 13, 14 lag (that is also a very awkward model to me) .But I think ACF and

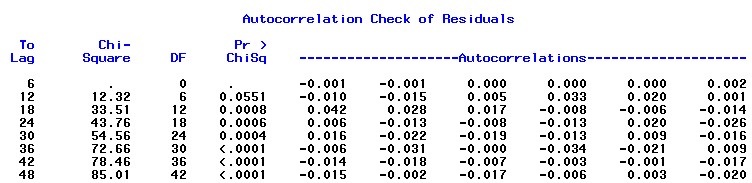
PACF do not tell too much information about which model to choose. I think we will need more steps to see which effects.

***e. Fit an AR(6) model: run model diagnostics to examine the significance of model coefficients and analyze the residuals. If model contains non-significant parameters, remove the parameters and fit the model again. Stop when you reach a satisfactory model.***

ANSWER:



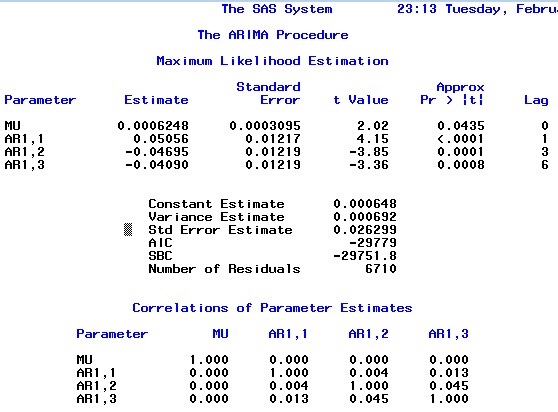
From the result above, we could see that AR(6) model is not very good. Three parameters are not significant.



Residual analysis also shows that some of them are correlated!

Next step for me is to test AR model with 1, 3, 6 lag as parameter since they are significant on the t-test.

Results of AR model with 1, 3, 6 lag is below:



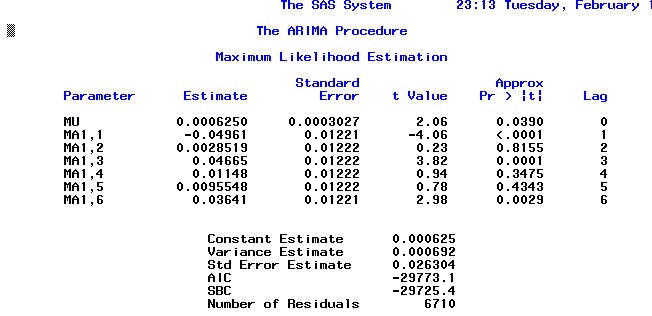
The three parameters are all significant, although the residual analysis of it also looks not better than AR(6)!

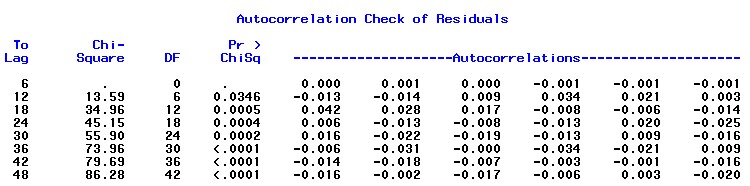
Test for AR model with 1-lag and 3-lag shows a similar result while I reject it because AIC and BIC criteria worsens and residual still does not improve by that!

So my preferred model is Rt = 0.0006+0.051\*Rt-1-0.047\*Rt-3-0.041\*Rt-6.

***f. Fit an MA(6) model run model diagnostics to examine the significance of model coefficients and analyze the residuals. If model contains non-significant parameters, remove the parameters and fit the model again. Stop when you reach a satisfactory model.***

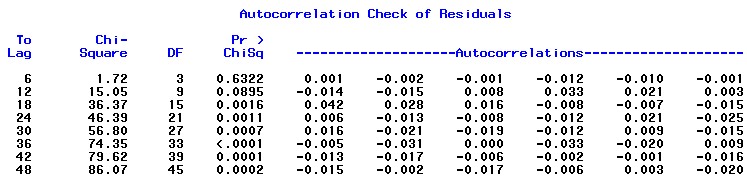
ANSWER:

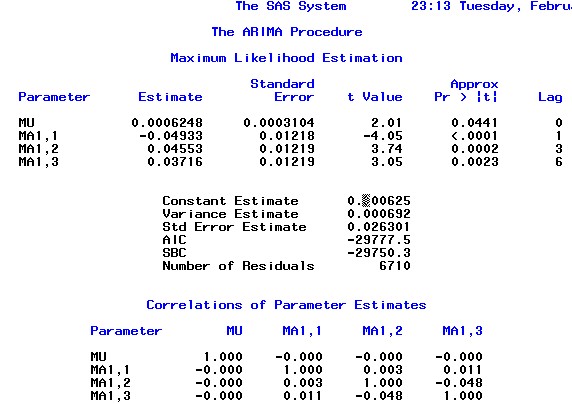




From the result, we could see that MA(6) model also has some insignificant parameters. Also, the residual test shows that there is evidence to reject the hypo that residuals are non-correlated.

Also, from the significance test of parameter, I will test MA model with 1-lag, 3-lag and 6-lag. The result looks better in light of parameter significance.





In MA model`s selection, I prefer Rt  = 0.000625+at+0.05\*at-1-0.046\*at-3-0.037\*at-6.

***h. Identify the best model based on the results of your analysis. Write down the formula for the selected model.***

ANSWER:

I think both AR and MA models are good, I cannot tell which one is better.

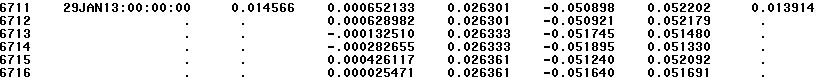
MA model, Rt  = 0.000625+at+0.05\*at-1-0.046\*at-3-0.037\*at-6.

AR model, Rt = 0.0006+0.051\*Rt-1-0.047\*Rt-3-0.041\*Rt-6.

***i. Using the selected final model, compute the 5 step-ahead forecasts with origin in the last observation.***

ANSWER:

C:\Users\felix\Dropbox\courses\csc 425\homework_3\1-7.jpg

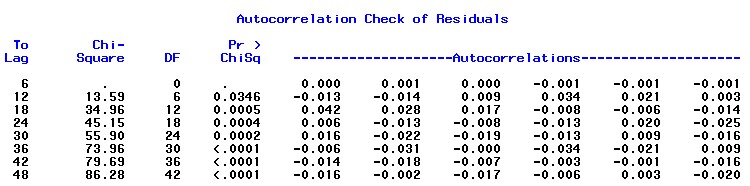
******

Here is my result of computing 5 step-ahead. For some reason I do not know, I cannot plot it in SAS, it always tells me NOTE: 6710 observation(s) outside the axis range for the L95 \* Date request.

***j. Discuss possible weaknesses of your model and analyses.***

ANSWER:

The residuals of my models are not uncorrelated, we could see from below that some of them are correlated told by L-B test of white noise.



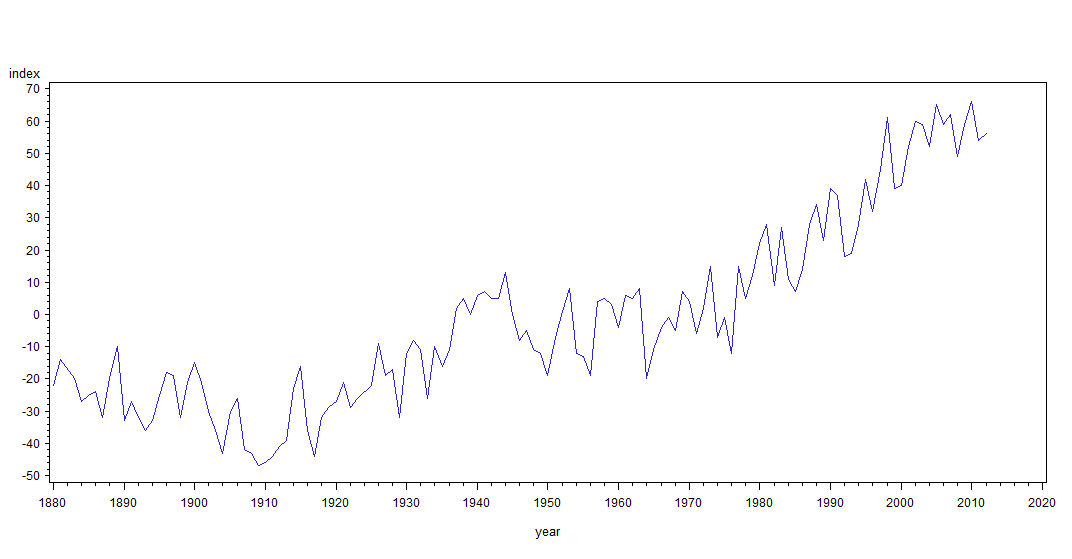
Also, the autocorrelation function shows very small values. I am not very convinced that the series data is not white noise.

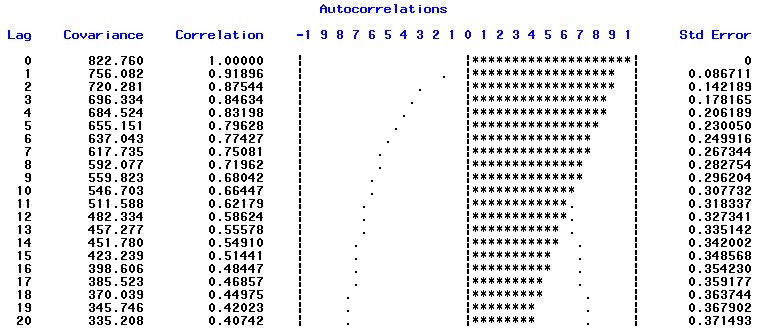
***Problem 2 [12 points] (to be covered in Week 5)***

***Global temperatures are rising. Analyze the time series of Land-Ocean Temperature Index in degrees Celsius published at the NASA-GISS (Goddard Institute for Space Studies). The temperature index measures temperature anomalies that are computed relative to the base period 1951-1980. The reason to work with anomalies, rather than absolute temperatures that absolute temperature varies markedly in short distances, while monthly or annual temperature anomalies are representative of a much larger region (from GISS website).***

***1. Plot the temperature index time series and its ACFs (20 lags). Analyze trends and patterns shown by the data.***

ANSWER:



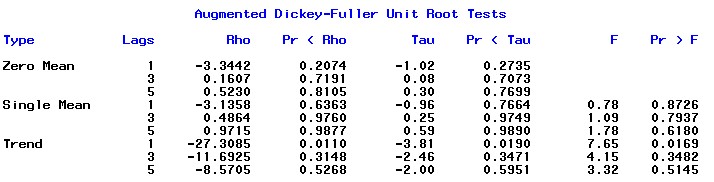


From the result we could see that the temperature in survey is in an uptrend, it is clearly not a weak stationary series. It shows a strong autocorrelation and its memory lasts for a very long period, maybe there is unit root non-stationary.

***2. Analyze if the series is stationary using both the ACF function and the Dickey Fuller test to check if TS is unit-root non-stationary.***

ANSWER:

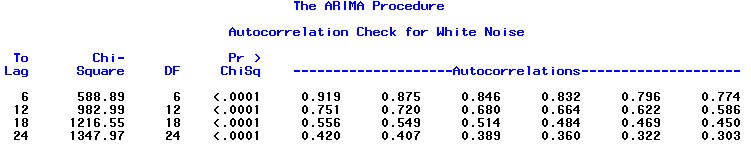
From the ACFs, it looks like unit root non-stationary, since the memory lasts long. However, we have to prove it by D-F test.



In this case, I assume single mean, the result shows we can not reject the null hypo, saying that maybe there is unit root.

***3. Test if TS is white noise using the Ljung - Box test***

ANSWER:

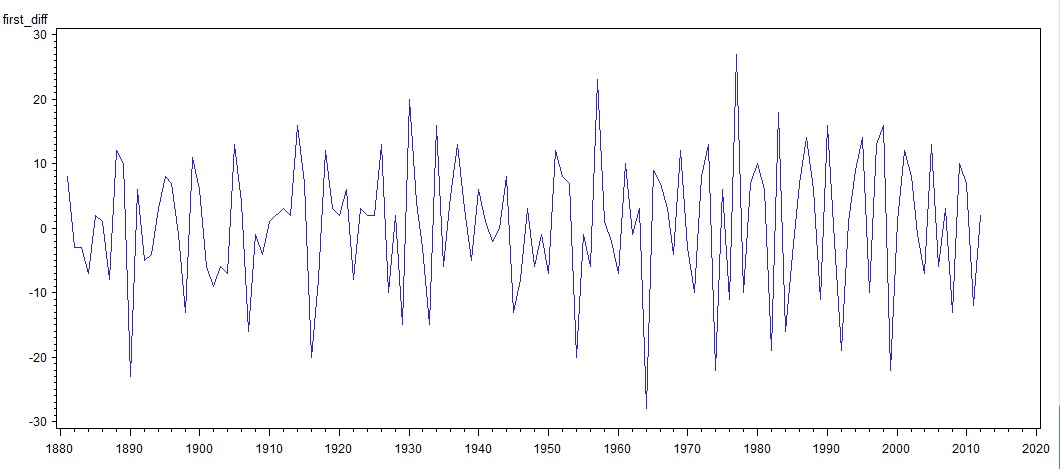


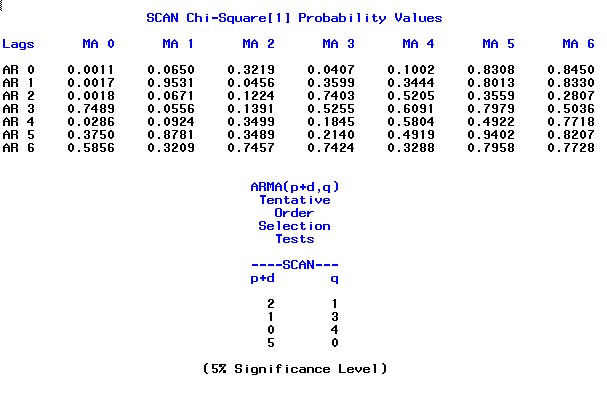
The series is clearly not white noise. The correlations are large though it decays. The null hypo is rejected as well.

***4. Use order selection methods such as EACF or BIC to identify the order of the “best” AR/MA/ARMA model for the temperature index TS. Note that if you determine that the TS is non-stationary, you should use an ARIMA model on the first differenced time series.***

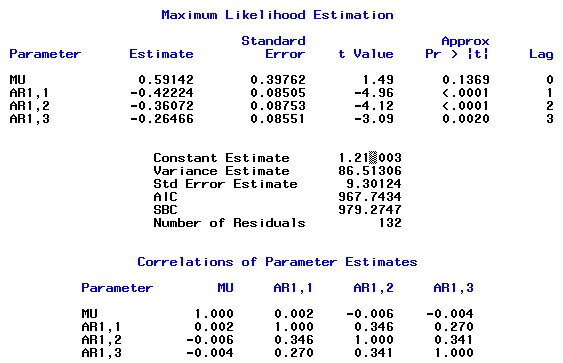
ANSWER:

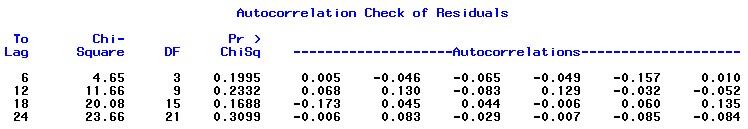
By getting the first difference, I get the time plot, which looks better.





From the result, I have several options to test for the order of p and q. After running these orders of p and q, I prefer p=3, q=0 (gotten by p=4 , q=0 then delete lag-4 coef since it is not significant). since other two models have correlated residuals or insignificant coefficients. Also, this model has the lowest AIC.





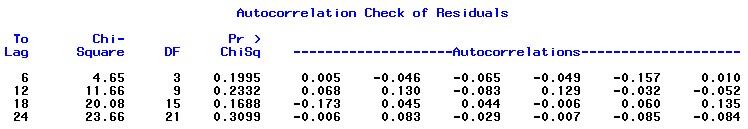
***5. Fit the selected ARIMA model, and write down the result.***

ANSWER:

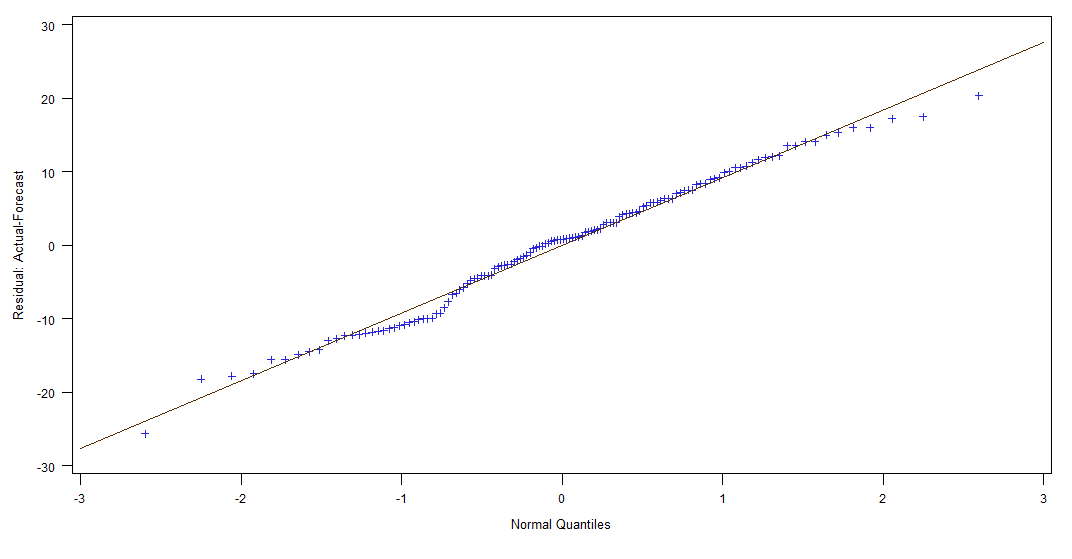
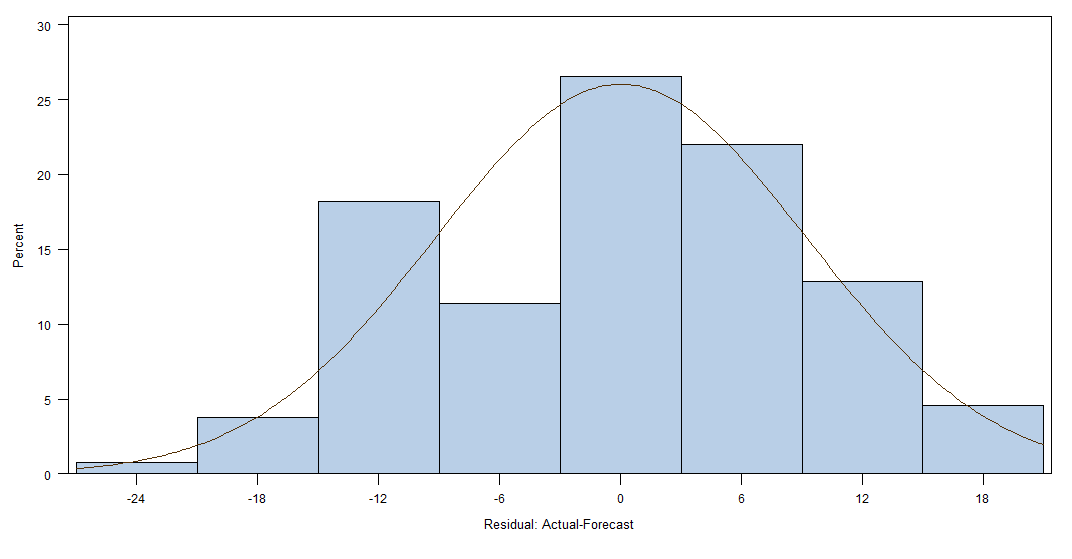
(1-B)Rt = 1.21-0.422\*(1-B)Rt-1-0.361\*(1-B)Rt-2-0.265\*(1-B)\*Rt-3+at

***6. Use residual analysis to check the model.***

ANSWER:



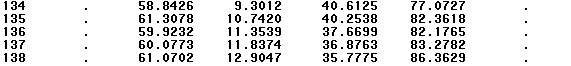
From the result, we could see that we could not reject the null hypo that residuals are uncorrelated, a good case for us. However, the histogram shows that the residuals are not perfectly normally distributed, a small flaw for us.

******

***7. Compute up to 4-step ahead forecasts of the fitted model with origin at the end of the data, i.e. 2012. Write down the forecasts and their standard errors.***

ANSWER:

C:\Users\felix\Dropbox\courses\csc 425\homework_3\2-9-1.jpg



Shown by the forecast column.