HOMEWORK\_TWO by Zhenyang Lu

**Problem 1 [6 pts]**

**Consider the following AR(2) time series process: rt  = 0.01+0.3rt-2+at, where {at} is a Gaussian white noise series with mean zero and constant variance σ2=0.02. (Hint: Read section 2.4.1 and 2.4.4 in your book on properties of AR models before you work on this problem)**

**a) What is the mean of the time series rt?**

ANSWER:

Mean of a time series is the expected value of a time series data. Here, in this case, for an AR(2) model, E(rt) = Φ0/ 1-Φ1-Φ2 = 0.01/1-0-0.3 = 0.014.

**b) Determine if the AR(2) model is stationary. Explain.**

ANSWER:

The characteristic second-order equation of this time series is 1-0.3X2 = 0, two solutions for it are 1.8 and -1.8, whose absolute values are larger than 1, saying that the time series is stationary.

**c) Compute the lag-1 and lag-2 autocorrelations of rt (HINT: check week 3 slides)**

ANSWER:

Lag-1 = cov(rt , rt-1) / var(rt) and Lag-2 = cov(rt , rt-2) / var(rt). However, in this case, we know that ρ1 = Φ1/ 1-Φ2 = 0, ρ2 = Φ1 ρ1 + Φ2 ρ0 =0.3. So, Lag-1 is zero, lag-2 is 0.3.

**d) Assume that r100= - 0.01 and r99 = 0.02. Compute the 1-step and 2-step ahead forecasts of the AR(2) series at the forecast origin t=100.**

ANSWER:

1-step ahead forecast, r101 = 0.01+0.3\*r99 = 0.016

2-step ahead forecast, r102 = 0.01+0.03\*r100=0.0097

**e) [Extracredit: 1 pts] What is the variance of the time series rt?**

ANSWER:

 = 0.02/(1-0.3\*0.3) = 0.022 (according to Wikipedia)

**Problem 2 [12 points]**

**Consider the Unemployment Insurance Weekly Claims (claims) from January 1990 to January 2013 in the file icsa.csv, obtained from the Federal Reserve Bank at St Louis. Build an AR(p) time series model for the series as described below. Review week 3 examples before analyzing this dataset.**

**a) Import the data either in R or SAS (Hint: be careful…the data are comma-delimited. In R you should create a time series object using the zoo() function.**

ANSWER:

See sourcecode.

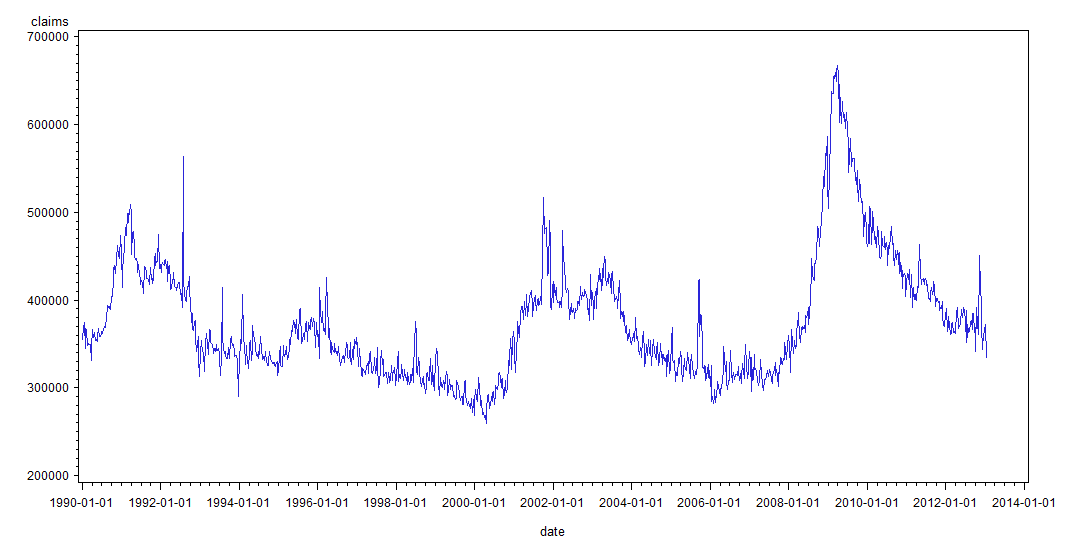
**b) Compute the growth rate of claims (ratechg) variable as the percentage change : (xt-xt-1)/xt-1**

ANSWER:

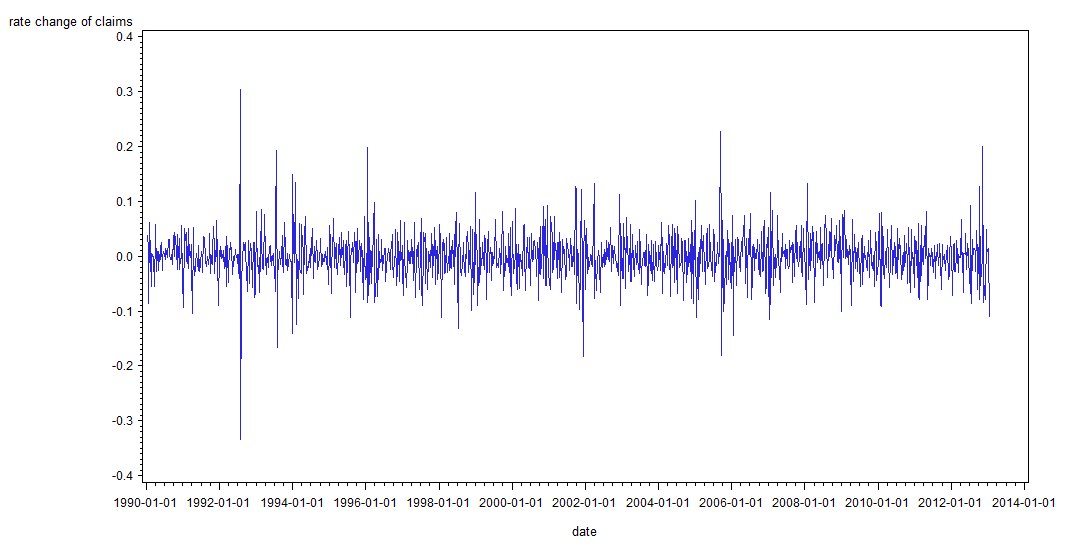
See sourcecode.

**c) Create time plots for claims and for ratechg. Analyze the time trends displayed by the plots?**

ANSWER:



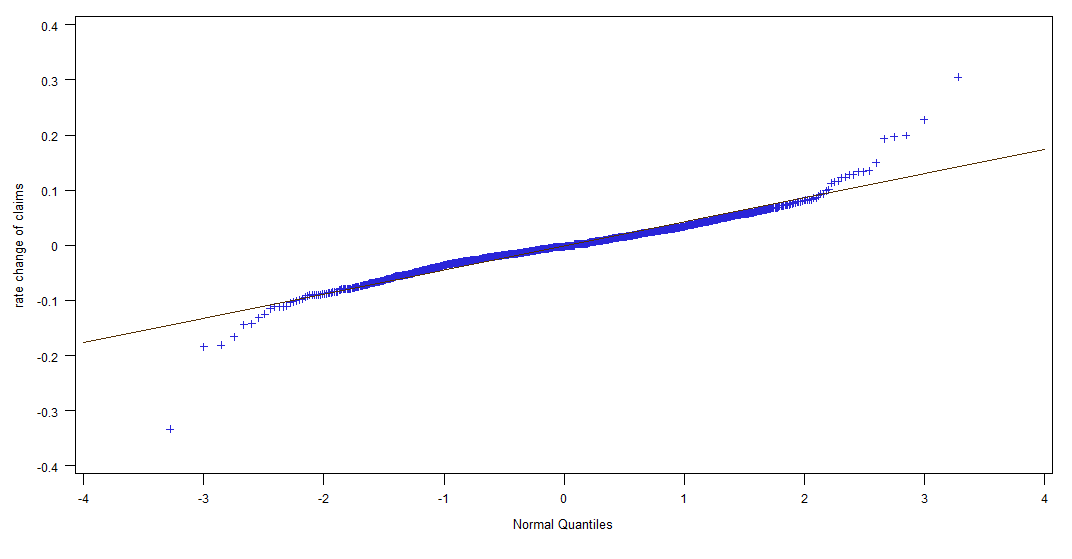
From the graph, we could see that, usually the claims recycle by around 8-10 years, for example, the claims decline from 1991 to 2000 and from 2000 to 2008, the claims go from plummet to peak and then to plummet again. In sum, the claims center around 300000 and 500000, with two short excess beyond 500000.



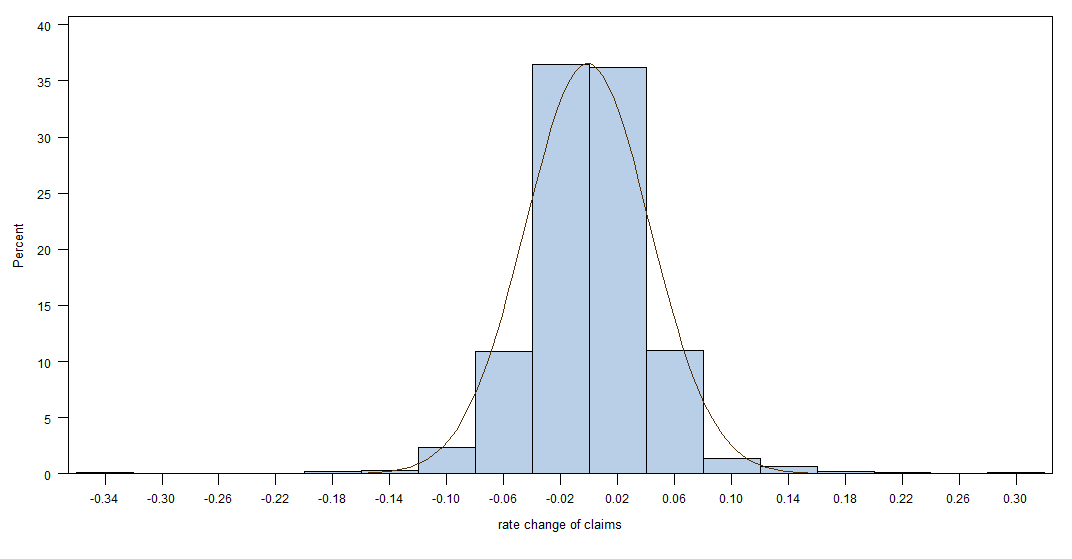
From the time plot of rate change of claims, the mean of it is around 0.0 with variance around 0.1 to 0.2, except for some extreme outliers. If we ignore the outliers, we could say it is weak stationary, showing that the expected rate change and volatility is time invariant.

**d) Analyze the distribution of ratechg. Can you assume that ratechg is normally distributed?**

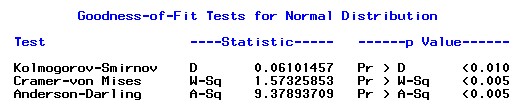
ANSWER:



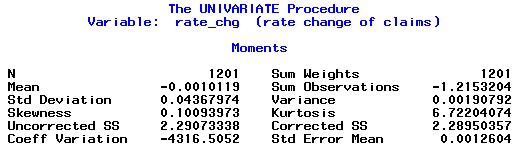
From the QQplot we see that generally the dots lie on the line, except for large tails on both sides. We need to further analyst other data to confirm it is normally distributed.



However, the histogram tells us that possibly it is not normally distributed, since too many observations are centered around the interval between -0.02 and 0.02. Also, there are extreme values on both sides. The K-S test also confirms that it is not normally distributed.

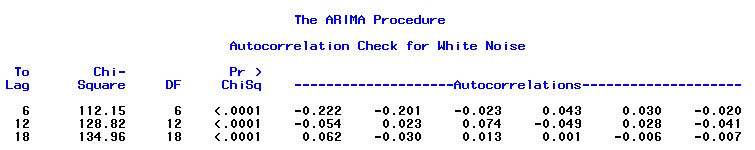


The Kurtosis is too large to be a normal distribution, that explains why there are extreme values on both sides! See below.



**e) Is the time series of rate changes (ratechg)serially correlated? Use the Ljung Box test.**

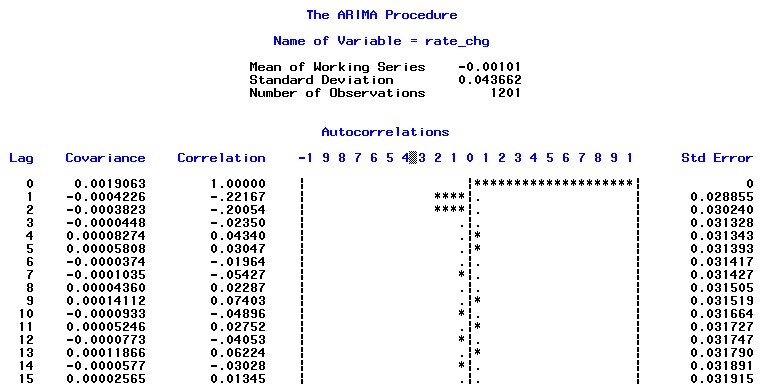
ANSWER:



The L-B test result shows that, null hypo has been rejected since at least three lag periods are significantly different from zero.

**f) Analyze the first 15 lags of ACF for ratechg. Draw conclusions.**

ANSWER:



From the result, we could see that there are evidently serial correlation prove can be identified, autocorrelations for lag 1 and 2 period are significantly different than zero.

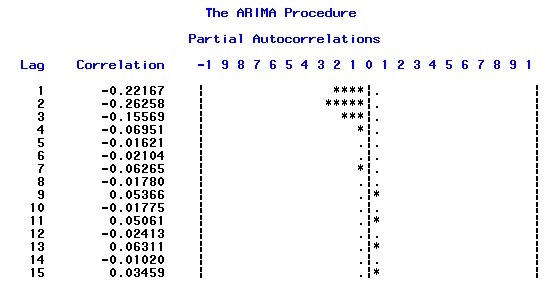
**g) Is there evidence that the time series ratechg is non stationarity?**

ANSWER:

I do not think there is evidence show that it is nonstationary, since as shown by question c), the time plot of rete\_chg is clearly time invariant for its mean and variance. Plus, question f) has shown that the ACF for 15 lags decays very fast, lag 3 has reached zero.

**h) Analyze the first 15 lags of PACF of ratechg. If you want to fit an AR(p) model, what order would you use based on the PACF? Why?**

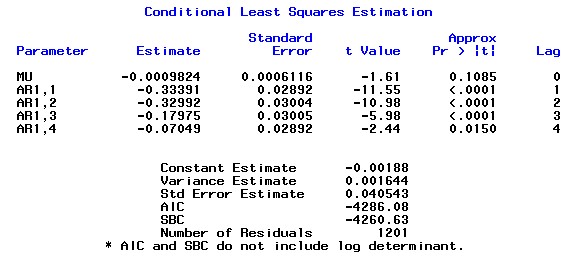
ANSWER:

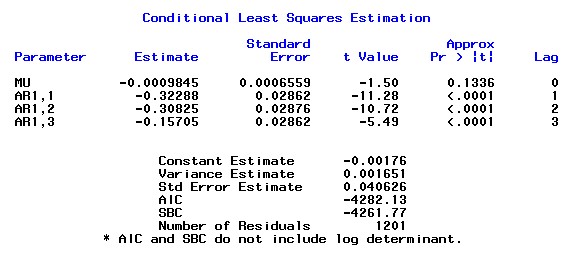


From the result, I would say 3 order or 4 order is best suited to be modeled. Since beyond that, the autocorrelation is not significantly different than zero, which means it is meaningless to add more independent variables.

**i) Fit an AR(p) model for the data.**

ANSWER:



****

Above is the AR(3) model and AR(4) model. Although the AIC and SBC do not change too much on AR(4) and AR(3), for this step, I would suggest AR(3) model, since AR1,4 is not significantly different from zero, shown by its t- value.

Thus, AR(3) is **rt = -0.00176-0.33228rt-1-0.30825rt-2-0.15705rt-3**

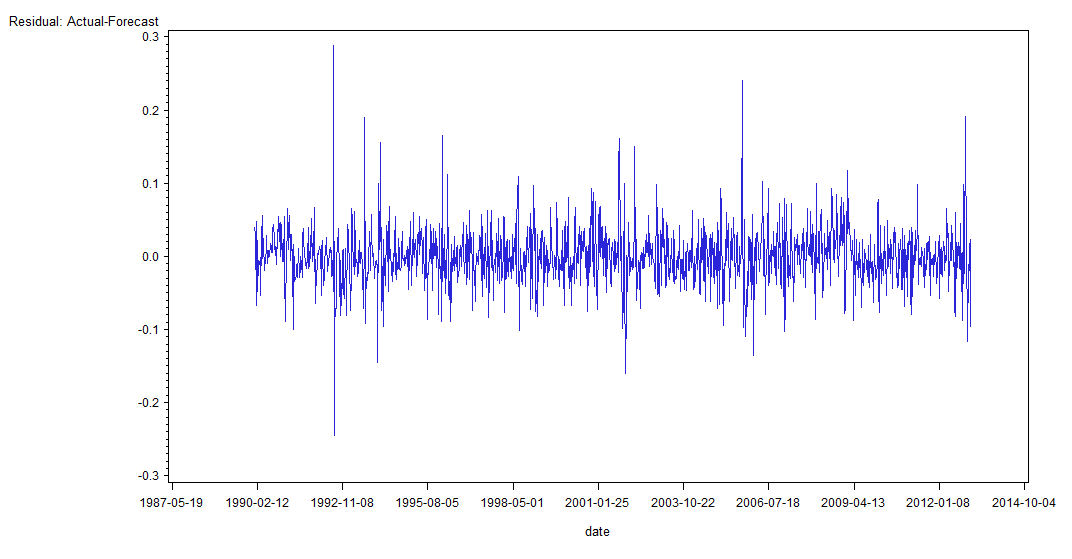
**j) Examine the significance of the model coefficients, and analyze the residuals to select a good model for the data. If the AR(p) model selected above is not appropriate, fit another AR(p) model of different order.**

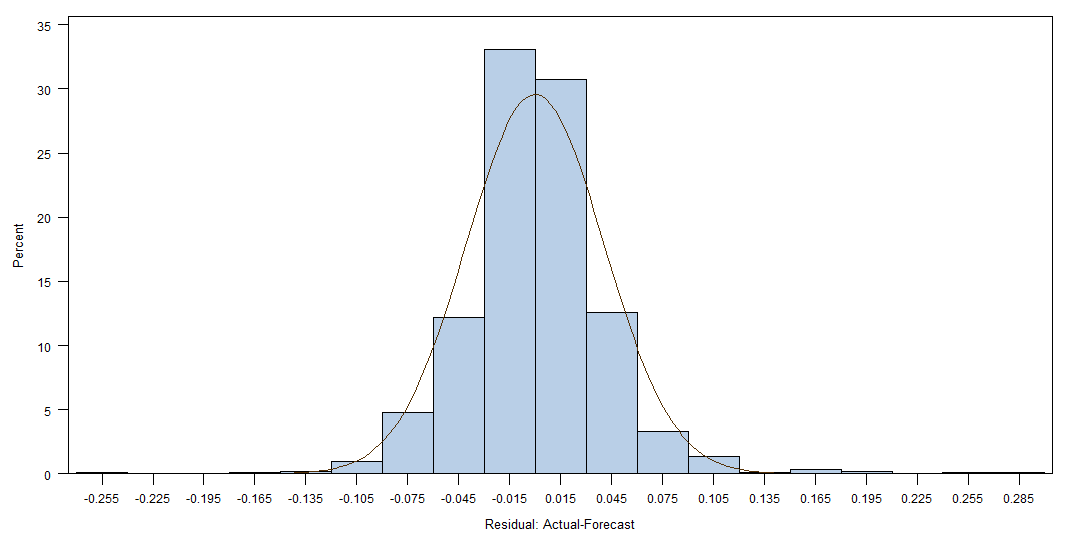
ANSWER:

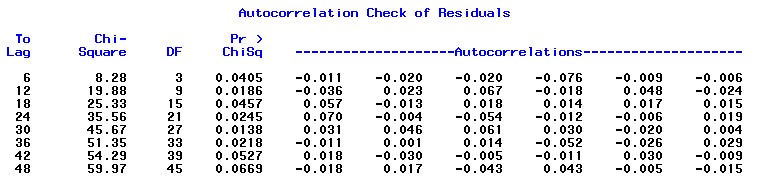
From question i), I have suggested AR(3) model, since AR1,4 is not significantly different from zero, shown by its t- value. For other AR1,I, I = 1,2,3, t-test shows they are all significantly different from zero.

After analyzing residual for AR(4) and AR(2), we could see that residual for those two are identical to AR(3), whose residual is not normally distributed, saying that AR(4) and AR(2) are not superior than AR(3) model, so I will use AR(3) model.

Below I will discuss the residual analysis for AR(3),

****

Time plot of residual centered around 0.0 and variance about 0.1. While the histogram shows it is possibly not a normal one.



This shows that residuals are serially correlated, lags that are significantly different from zero are shown above. So my model is not flawless though it is the best I could do.

**k) Write down the expression of the selected Ar(p) model, and discuss the serial dependence represented by the model.**

ANSWER:

AR(3) is **rt = -0.00176-0.33228rt-1-0.30825rt-2-0.15705rt-3**

The model shows that rt is dependent on the previous three value. They are inversely correlated with rt.