MIDTERM SAMPLE PROBLEMS

- **Problem A**

1. Suppose we have the following ADF test result on a time series. What conclusion can we make about that time series and why?(4 pts)

```
adfTest(gdp.return,lags=10)

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 10
  STATISTIC:
    Dickey-Fuller: -0.8671
  P VALUE:
    0.3364
```

2. State whether the following model is causal, and whether it is invertible. Explain why.

$$X_t = Z_t - 1.5Z_{t-1} + 0.5Z_{t-2}$$

3. Suppose **after** conducting one regular differencing and one seasonal differencing on series X_t, we perform seasonal ARMA analysis on the **resulting series**. The following are the outputs of the seasonal ARMA model. Assume the seasonal period is 4.

```
      ar1     ma1      sma1
      0.2569  -0.9302  -0.753
s.e.  0.0787   0.0361   0.058
```

(a) Write down the estimated model for X_t.

(b) Write down the AR and MA polynomials of the model.

4. Obtain all non-zero serial correlations $\rho(h)$ of the model $X_t = (1 - 0.2B)(1 - 0.3B^2)Z_t$ where Z_t is i.i.d. normal with mean zero and variance 3. Show your computation steps.

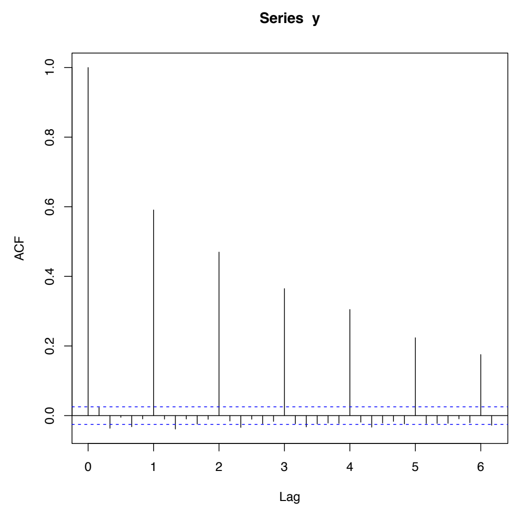
5. Rewrite the following models without B and Δ

$$(1 - \phi_1 B^3)\Delta X_t = (1 + \theta_1 B)(1 + \theta_2 B^2)Z_t$$

6. Rewrite the following models using only B, Δ, X_t, and Z_t.

$$Y_t = X_t - X_{t-3}; \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + Z_t + \theta_1 Z_{t-1}$$

7. Based on the ACF plot, which of the following is true and why?
(A) This is an MA process. (B) This is an AR process. (C) This is a seasonal MA process. (D) This is a seasonal AR process



8. Suppose the monthly simple returns of a portfolio for Oct., Nov., and Dec. of 2011 were 0.04, 0.03, and -0.02 respectively. What is the **quarterly log return** of the portfolio for last quarter of 2011? What is the **quarterly simple return** of the portfolio for last quarter of 2011?

- **Problem B** :

  Consider the GNP data we used in class and the output attached. Answer the following questions:

  1. Judging from the first two Box-Ljung tests, is there any significant serial correlation in the GNP data? Is there any significant serial correlation in the residuals of model one.

  2. Write down the three estimated models.

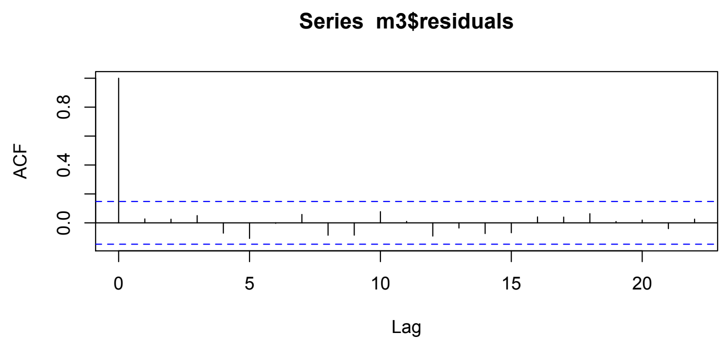
  3. Judging from the estimated coefficients from model two, what action needs to be taken and why?

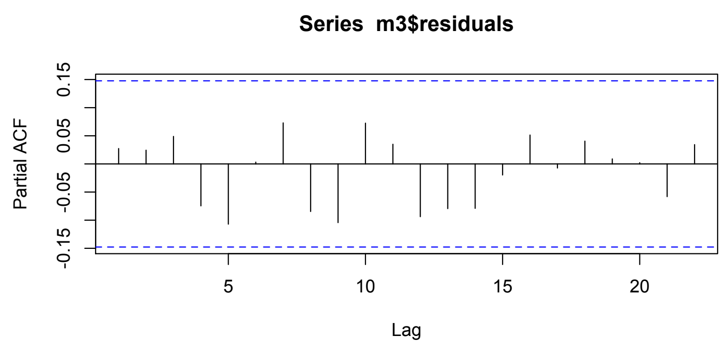
  4. Based on the residual analysis shown in the output, is model three satisfactory? Why and why not?

  5. All things considered, which one of these three models are the best? Explain.

```
> da=read.table("dgnp82.txt")
> x=da[,1]
> Box.test(x,lag=10,type="Ljung")
Box-Ljung test
data:  x
X-squared = 43.2345, df = 10, p-value = 4.515e-06

#####  Model one
> m1=arima(x,order=c(3,0,0))
> m1
Call:
arima(x = x, order = c(3, 0, 0))
Coefficients:
         ar1     ar2      ar3  intercept
      0.3480  0.1793  -0.1423     0.0077
s.e.  0.0745  0.0778   0.0745     0.0012
sigma^2 estimated as 9.427e-05:  log likelihood = 565.84,  aic = -1121.68

> Box.test(m1$residuals,lag=10,type="Ljung")
Box-Ljung test
data:  m1$residuals
X-squared = 7.0169, df = 10, p-value = 0.7239

#####  Model two
> m2=arima(x,order=c(5,0,2))
> m2
Call:
arima(x = x, order = c(5, 0, 2))
Coefficients:
         ar1      ar2     ar3     ar4      ar5      ma1     ma2
      0.6148  -0.7242  0.1184  0.0829  -0.0710  -0.2821  0.8596
s.e.  0.2064   0.2150  0.1087  0.1264   0.0845   0.1886  0.1437
      intercept
         0.0077
s.e.     0.0012
sigma^2 estimated as 9.101e-05:  log likelihood = 568.77,  aic = -1119.55
> Box.test(m2$residuals,lag=12,type="Ljung")
Box-Ljung test
data:  m2$residuals
X-squared = 3.7481, df = 12, p-value = 0.9876

##### Model three
> m3=arima(x,order=c(0,0,2))
> m3
Call:
arima(x = x, order = c(0, 0, 2))
Coefficients:
         ma1     ma2  intercept
      0.3121  0.2714     0.0077
s.e.  0.0736  0.0678     0.0012
sigma^2 estimated as 9.504e-05:  log likelihood = 565.14,  aic = -1122.29
> Box.test(m3$residuals,lag=12,type="Ljung")
Box-Ljung test
data:  m3$residuals
```

```
X-squared = 9.8517, df = 12, p-value = 0.629
> acf(m3$residuals)
> pacf(m3$residuals)
```





- **Problem C** : This problem involves the analysis of daily log return of SP500 from Jan 2006 to Jan 2012.

  1. First we look at the outputs of simple statistics calculated for sp500 daily log returns. Answer the next two questions:

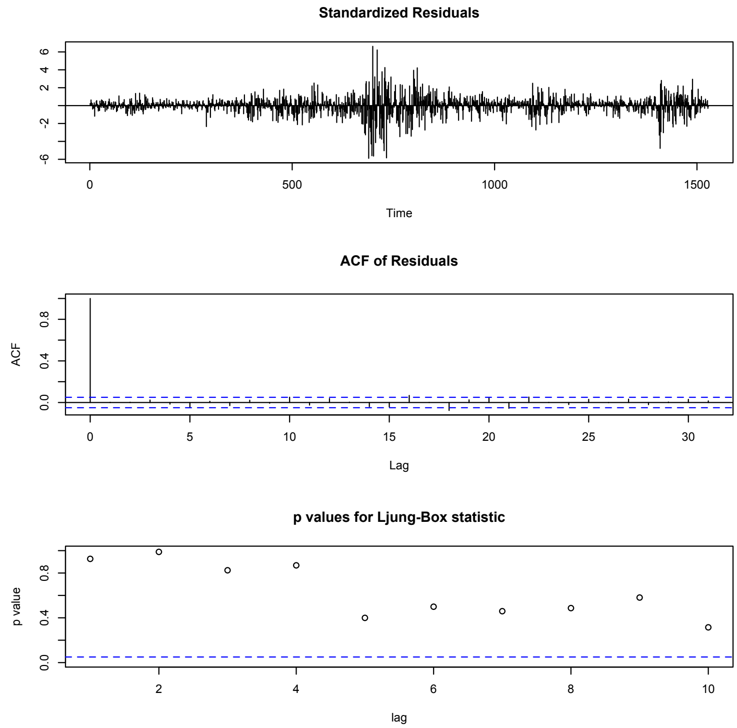
```
> basicStats(sp500.logreturn)
            sp500.logreturn
nobs            1528.000000
NAs                0.000000
Minimum           -0.094695
Maximum            0.109572
1. Quartile       -0.005430
3. Quartile        0.006319
Mean               0.000024
Median             0.000857
Sum                0.036776
SE Mean            0.000396
LCL Mean          -0.000753
UCL Mean           0.000801
Variance           0.000240
Stdev              0.015485
Skewness          -0.276724
Kurtosis           7.927383
```

     (a) Based on the output, is the mean of the log return significantly different from zero at 5% level? Write down the test-statistics and draw conclusion. $Z_{0.025} = 1.96$, $Z_{0.05} = 1.65$?

     (b) Which statistics in the output is usually used for checking if the log return is heavy tailed? Write down the numerical value of that statistics.

  2. Next we fit an AR(2) model to the data. Judging from the tsdiag plot of the model, is the model adequate? Why? Pay special attention to the first plot.

```
> model1=arima(sp500.logreturn ,order=c(2,0,0))
> tsdiag(model1)
```



3. We then perform two Ljung-Box tests on the residual series of model1. Judging from the outputs, answer the following:

```
> Box.test(model1$res, 10, type="Ljung")
Box-Ljung test
data:  model1$res
X-squared = 11.5698, df = 10, p-value = 0.3149
> Box.test((model1$res)^2, 10, type="Ljung")
Box-Ljung test
data:  (model1$res)^2
X-squared = 1376.594, df = 10, p-value < 2.2e-16
```

(a) does there appear to be significant autocorrelation among the residual series. Why?

(b) is the residual series independent. Why?

(c) are there any ARCH-effects in the log returns. Why?

- **Problem D**

  Consider the AR(2) model

$$X_t = 0.02 + 0.7X_{t-1} - 0.1X_{t-2} + Z_t,$$

  where Z_t is normal with mean 0 and variance 2.

  1. What is $E(X_t)$?

  2. What is the AR polynomial for this model? What are the roots of the polynomial?

  3. Suppose that $X_{100} = 1.2$ and $X_{99} = 0.5$. What is the 1-step ahead prediction of X_{101} at forecast origin T=100?

Consider the following model $X_t = 2t + 0.5Z_t + 1.2Z_{t-1}$ where Z_t is normal with mean 0 and variance 2.

4. Write down the form of 2-step ahead forecast when forecast origin is at t. Namely $E(X_{t+2}|\mathcal{F}_t)$.

5. Assume that $X_{100} = 10$, what is the 500 step ahead prediction when the forecasting origin is 100?

6. Write down the 1-step forecasting error $U_t(1) = E(X_{t+1}|\mathcal{F}_t) - X_{t+1}$.

7. Compute the variance of 1-step ahead forecast error at the forecast origin 100.