Dept of Mathematics and Statistics
King Fahd University of Petroleum & Minerals

STAT460: Time Series

Dr. Mohammad H. Omar

Final Exam Term 162 FORM A
Wednesday May 24 2017
12.30pm-3.00pm

Name____ ID#:__________ Serial #:___

Instructions.

1. Please turn off your cell phones and place them under your chair. Any student caught with mobile phones on during the exam will be considered under the **cheating rules** of the University.
2. If you need to leave the room, please do so quietly so not to disturb others taking the test. No two person can leave the room at the same time. No extra time will be provided for the time missed outside the classroom.
3. Only materials provided by the instructor can be present on the table during the exam.
4. Do not spend too much time on any one question. If a question seems too difficult, leave it and go on.
5. Use the blank portions of each page for your work. Extra blank pages can be provided if necessary. If you use an extra page, indicate clearly what problem you are working on.
6. Only answers supported by work will be considered. Unsupported guesses will not be graded.
7. While every attempt is made to avoid defective questions, sometimes they do occur. In the rare event that you believe a question is defective, the instructor cannot give you any guidance beyond these instructions.
8. Mobile calculators, I-pad, or communicable devices are disallowed. Use regular scientific calculators or financail calculators only. Write important steps to arrive at the solution of the following problems.

The test is 150 minutes, GOOD LUCK, and you may begin now!

Question	Total Marks	Marks Obtained	Comments
1	4+3+3+4=14		
2	4		
3	6		
4	3+3+4=10		
5	5+3=8		
6	5		
7	3+3+3+3+2=14		
8	4+3+5=12		
9	5+5+3=13		
10	3+3+4+4=14		
Total	100		

Extra blank page

1. (4+3+3+4=14 points) The data file **wages** contains monthly values of the average hourly wages (in dollars) for workers in the U.S. apparel and textile products industry for July 1981 through June 1987. Consider the following least squares fit of a **quadratic time trend**.

```
> wages.lm2=lm(wages~time(wages)+I(time(wages)^2)); summary(wages.lm2)
```

```
      Call:
      lm(formula = wages ~ time(wages) + I(time(wages)^2))

Residuals:
      Min        1Q    Median        3Q       Max
-0.148318 -0.041440  0.001563  0.050089  0.139839

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -8.495e+04  1.019e+04  -8.336 4.87e-12 ***
time(wages)       8.534e+01  1.027e+01   8.309 5.44e-12 ***
I(time(wages)^2) -2.143e-02  2.588e-03  -8.282 6.10e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05889 on 69 degrees of freedom
Multiple R-Squared: 0.9864,     Adjusted R-squared: 0.986
F-statistic:  2494 on 2 and 69 DF,  p-value: < 2.2e-16
```

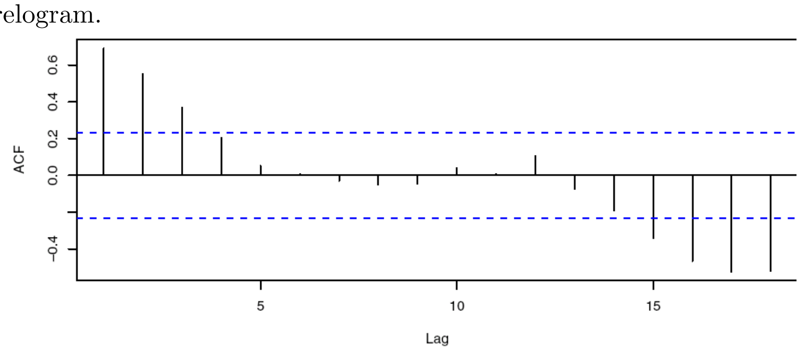
(a) Evaluate the goodness of fit of the **quadratic time trend** model.

(b) The following provides a **runs test** on the standardized residuals. Interpret the results.

```
> runs(rstudent(wages.lm2))
```

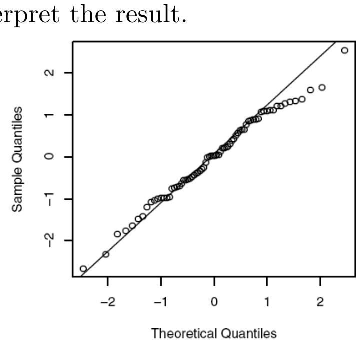
```
$pvalue
[1] 1.56e-07
$observed.runs
[1] 15
$expected.runs
[1] 36.75
$n1
[1] 33
$n2
[1] 39
$k
[1] 0
```

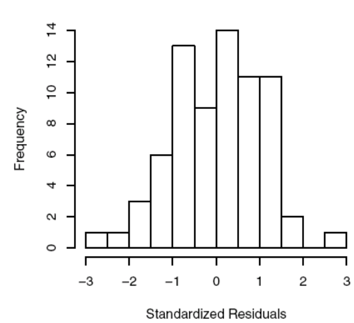
(c) The following are the **sample autocorrelations** for the standardized residuals. Interpret this residual correlogram.




```
> acf(rstudent(wages.lm2))
```

(d) Consider histogram, normal probability plot, and Shapiro test of the standardized residuals below. Interpret the result.






```
> qqnorm(rstudent(wages.lm2)); qqline(rstudent(wages.lm2))
> hist(rstudent(wages.lm2),xlab='Standardized Residuals')
> shapiro.test(rstudent(wages.lm2)
```

Shapiro-Wilk normality test
data: rstudent(wages.lm2)
W = 0.9887, p-value = 0.7693

2. (4 points). Consider an MA(6) model with

$$\theta_1 = 0.5, \theta_2 = -0.25, \theta_3 = 0.125, \theta_4 = -0.0625, \theta_5 = 0.03125, \text{ and } \theta_6 = -0.015625.$$

Find a much simpler model that has nearly the same $\psi-$weights.

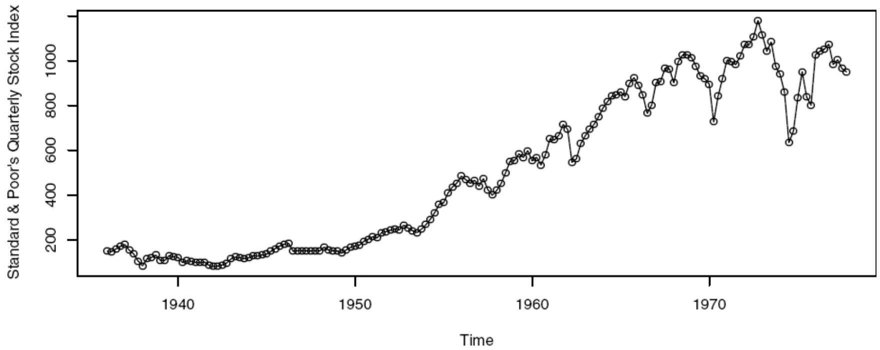
3. (6 points) The sample ACF for a series and its first difference are given in the following table. Here $n = 100$.

lag	1	2	3	4	5	6
ACF for Y_t	0.97	0.97	0.93	0.85	0.80	0.71
ACF for ∇Y_t	-0.42	0.18	-0.02	0.07	-0.10	-0.09

Based on this information alone, which ARIMA model(s) would we consider for the series?

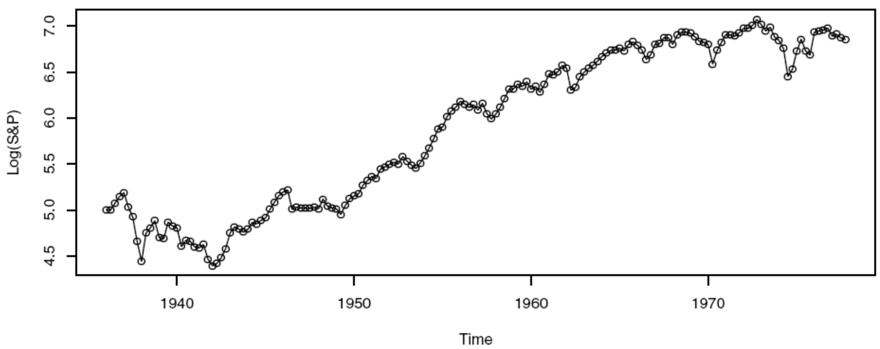
4. (3+3+4=10 points) The data file **SP** contains quarterly Standard & Poor's Composite Index stock price values from the first quarter of 1936 through the fourth quarter of 1977.

(a) The time series plot for these data is displayed below. Interpret any visible patterns.



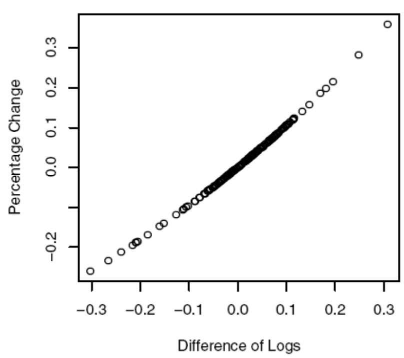

```
> data(SP); plot(SP,type='o',ylab='Standard & Poor\'s Quarterly Stock Index')
```

(b) Now take natural logarithms of the quarterly values and the time series plot of the transformed values.is displayed below. Describe the effect of the logarithms on the behavior of the series.




```
> plot(log(SP),type='o',ylab='Log(S&P)')
```

(c) We now calculate the (fractional) relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them to the differences of (natural) logarithms, $\nabla log(Y_t)$ as given below. How do they compare for smaller values and for larger values?




```
> percentage=na.omit((SP-zlag(SP))/zlag(SP))
> win.graph(width=3,height=3,pointsize=8)
> plot(x=diff(log(SP))[-1],y=percentage[-1], ylab='Percentage Change',
   xlab='Difference of Logs')
> cor(diff(log(SP))[-1],percentage[-1])
```

cor(diff(log(SP))[-1],percentage) =0.996.

5. (5+3=8 points) Consider an MA(1) process for which it is *known* that the process mean is zero. Based on a series of length $n = 3$, we observe $Y_1 = 0$, $Y_2 = -1$, and $Y_3 = \frac{1}{2}$.

(a) Show that the **conditional least-squares** estimate of θ is $\frac{1}{2}$.

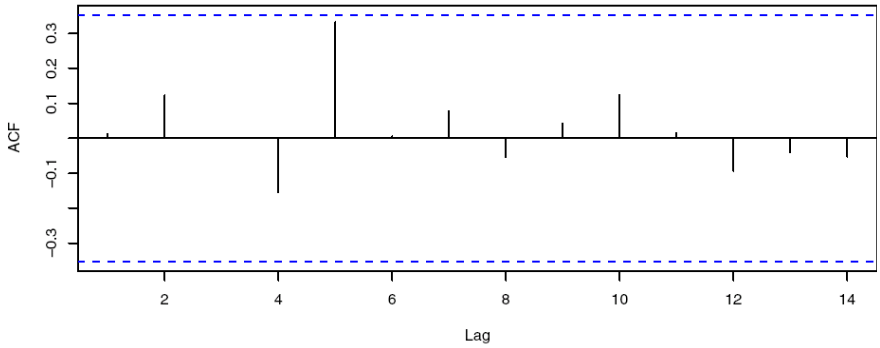
(b) Find an estimate of the **noise variance**. (Hint: Iterative methods are not needed in this simple case.)

.

6. (5 points) For an AR(1) model with $\phi \approx 0.4$ and $n = 100$, the lag 1 sample autocorrelation of the residuals is 0.5. Should we consider this unusual? Why or why not?

7. (3+3+3+3+2=14 points) Fit an AR(3) model by maximum likelihood to the square root of the hare abundance series (filename hare).

(a) The following is the ACF plot of the sample residuals. Comment on the size of the correlations.




```
> data(hare); model=arima(sqrt(hare),order=c(3,0,0))
> win.graph(width=6.5,height=3,pointsize=8); acf(rstandard(model))
```

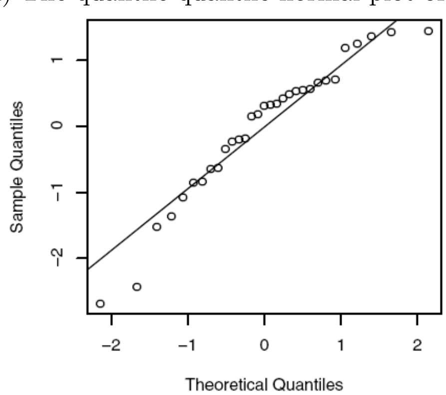
(b) The following provides the Ljung-Box statistic. Does this statistic support the AR(3) specification?

```
> LB.test(model,lag=9)

        Box-Ljung test

        data:  residuals from  model
        X-squared = 6.2475, df = 6, p-value = 0.3960
```

(c) Perform a runs test on the residuals and comment on the results.

```
> runs(rstandard(model))

       $pvalue
       [1] 0.602
       $observed.runs
       [1] 18
       $expected.runs
       [1] 16.09677
       $n1
       [1] 13
       $n2
       [1] 18
       $k
       [1] 0
```

(d) The quantile-quantile normal plot of the residuals is displayed below. Comment on the plot.




```
> win.graph(width=3,height=3,pointsize=8)
> qqnorm(residuals(model)); qqline(residuals(model))
```

(e) The Shapiro-Wilk test is performed on the residuals. What do you conclude.

```
> shapiro.test(residuals(model))
        Shapiro-Wilk normality test

        data:  residuals(model)
        W = 0.9351, p-value = 0.06043
```

8. (4+3+5=12 points) Suppose that annual sales (in millions of dollars) of ACME Corporation follow the AR(2) model $Y_t = 5 + 1.1Y_{t-1} - 0.5Y_{t-2} + e_t$ with $\sigma_e^2 = 2$.

(a) If sales for 2005 to 2007 are as given below, forecast sales for 2008 and 2009.

t	2005	2006	2007
sales in million \$ Y_t	\$9 mil	\$11 mil	\$10 mil

(b) Show that for this model.$\psi_1 = 1.1$.

(c) Calculate 95% prediction limits for your forecast in part (a) for 2008.

9. (5+5+3=13 points) Identify the following as certain multiplicative seasonal ARIMA models:

(a) $Y_t = 0.5Y_{t-1} + Y_{t-4} - 0.5Y_{t-5} + e_t - 0.3e_{t-1}$.

(b) $Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + e_t - 0.5e_{t-1} - 0.5e_{t-12} + 0.25e_{t-13}$

(c) If

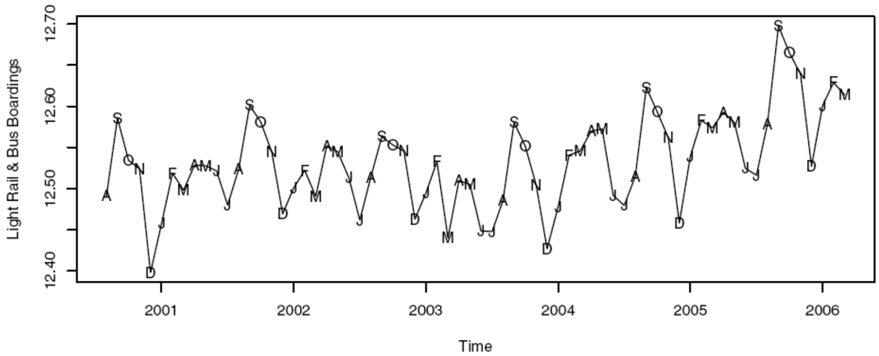
t	last but 4	last but 3	last but 2	last but 1	last
Y_t	91	89	91	92	90

with $e_t = -1$, provide the **1 lead forecast** for the model in part (a).

.

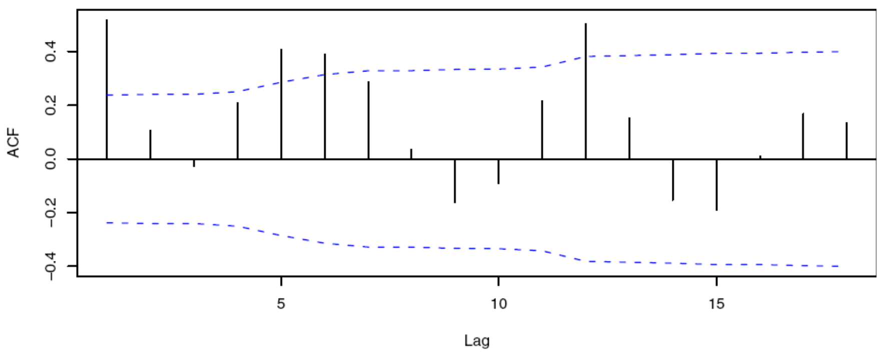
10. (3+3+4+4=14 points) The file named **boardings** contains monthly data on the number of people who boarded transit vehicles (mostly light rail trains and city buses) in the Denver, Colorado, region for August 2000 through December 2005.

(a) The time series plot for these data is produced below with plotting symbols. Does a stationary model seem reasonable?




```
> data(boardings); series=boardings[,1]
> plot(series,type='l',ylab='Light Rail & Bus Boardings')
> points(series,x=time(series),pch=as.vector(season(series)))
```

(b) The sample ACF for this series are calculates and plotted below. At which lags do you have significant autocorrelation?




```
> acf(as.vector(series),ci.type='ma')
```

(c) An $\text{ARMA}(0,3) \times (1,0)_{12}$ model below is fitted to these data. Assess the significance of the estimated coefficients.

```
> model=arima(series,order=c(0,0,3),seasonal=list(order=c(1,0,0),period=12)); model

Call:
arima(x = series, order = c(0, 0, 3), seasonal = list(order = c(1, 0, 0), period
= 12))

Coefficients:
         ma1     ma2     ma3    sar1  intercept
      0.7290  0.6116  0.2950  0.8776    12.5455
s.e.  0.1186  0.1172  0.1118  0.0507     0.0354

sigma^2 estimated as 0.0006542:  log likelihood = 143.54,  aic = -277.09
```

(d) Another model, $\text{ARMA}(0,4) \times (1,0)_{12}$ model is fitted to the data as given below. In comparison to part (d), **interpret** the results.

```
> model2=arima(series,order=c(0,0,4),seasonal=list(order=c(1,0,0),period=12)); model2

Call:
arima(x = series, order = c(0, 0, 4), seasonal = list(order = c(1, 0, 0), period
= 12))

Coefficients:
         ma1     ma2     ma3     ma4    sar1  intercept
      0.7277  0.6686  0.4244  0.1414  0.8918    12.5459
s.e.  0.1212  0.1327  0.1681  0.1228  0.0445     0.0419

sigma^2 estimated as 0.0006279:  log likelihood = 144.22,  aic = -276.45
```

END OF TEST PAPER