Statistics 510 Name: ______________________
Practice Final
Fall 2009

Note that this is not an exhaustive coverage of possible topics on the actual final. It should be used for practice.

---

1. Suppose that you have a time series, $x_1, ..., x_{200}$ that was generated by simulation using the following R command:
`x=arima.sim(list(order = c(0,0,1), ma = 0.5), n = 200).`

   This time series was then fit with two different models using R. Why are the fits both so good? The log likelihoods are almost the same, and the AIC for both models are very close. Moreover, the diagnostic plots in figures 1 and 2 are not only both good but almost identical to one another. Why is this the case?

```
 > sarima(x,0,0,1)
$fit
Call:
arima(x = data, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
    xreg = xmean, include.mean = F)
Coefficients:
         ma1    xmean
      0.4972  -0.1918
s.e.  0.0633   0.0983
sigma^2 estimated as 0.865:  log likelihood = -269.43,  aic = 544.87
$AIC
[1] 0.8750359

> sarima(x,0,1,2)
$fit
Call:
arima(x = data, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
    xreg = constant, include.mean = F)
Coefficients:
          ma1      ma2  constant
      -0.4804  -0.4665   -0.0003
s.e.   0.0663   0.0669    0.0041
sigma^2 estimated as 0.8752:  log likelihood = -270.18,  aic = 548.37
$AIC
[1] 0.896708
```

2. The following questions are all based on data which consist of the number of passenger miles flown domestically in the UK recorded monthly from July 1962 to May 1972 .

   (a) There seems to be an overarching trend in the data. Moving average smoothing should clarify this trend. In figure 3, there is the original data plus the data after using a moving average smoother with window sizes of five, ten, and twelve. Which do you think most clearly shows the overall trend? Why is there such a big difference between windows of size ten and twelve? What effect does moving average smoothing have on the variance of a time series after smoothing?

(b) The original data does not appear stationary. Look at the plots in figure 4. Included in this figure are the plots of the original data and of the differenced data, seasonally differenced data, and the data that has been both differenced and seasonally differenced. Which of these appear stationary? Discuss the plots.

(c) The ACF plots of all of the differenced data can be found in figure 5. Do these plots seem reasonable based on the plots in figure 4? Comment on the ACF of the doubly differenced data and also the PACF of this doubly differenced data in figure 6. What models might be reasonable?

(d) Below you will find the R output for three possible models along with the diagnostics for these fits (in figures 7, 8, and 9). Which model fits best? Are there any problems with thee diagnostics?

```
 > sarima(air,1,1,1,1,1,1,12)
$fit
Call:
arima(x = data, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S))
Coefficients:
         ar1      ma1    sar1     sma1
      0.4389  -0.8273  0.3554  -0.8972
s.e.  0.1488   0.0872  0.1843   0.2843

sigma^2 estimated as 20.45:  log likelihood = -316.08,  aic = 642.15
$AIC
[1] 4.08533

> sarima(air,1,1,1,0,1,1,12)
$fit
Call:
arima(x = data, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S))
Coefficients:
         ar1      ma1    sma1
      0.5014  -0.8533  -0.522
s.e.  0.1314   0.0736   0.146
sigma^2 estimated as 22.43:  log likelihood = -317.5,  aic = 642.99
$AIC
[1] 4.161014
```

```
> sarima(air,1,1,0,0,1,1,12)
$fit
Call:
arima(x = data, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S))
Coefficients:
          ar1     sma1
      -0.2461  -0.6213
s.e.   0.0987   0.1503
sigma^2 estimated as 23.67:  log likelihood = -321.07,  aic = 648.13
$AIC
[1] 4.197762
```

(e) Write out the model (after differencing) that you think is the best.

(f) In figure 10, there are four versions of the periodogram. There is the raw periodogram, followed by spans of 3, 5, and 9. Which do you think is the best representation of the spectral density? How are the orginal data, the periodogram, the spectral density, and the autocovariance function related?

(g) In figure 11, there are four versions of the periodogram all with spans equal to five. But they have tapering of 0, 0.1, 0.25 and 0.5. Comment on the plots. Why do you think there is such a dramatic difference between the plots? What does tapering do?

Figure 1:

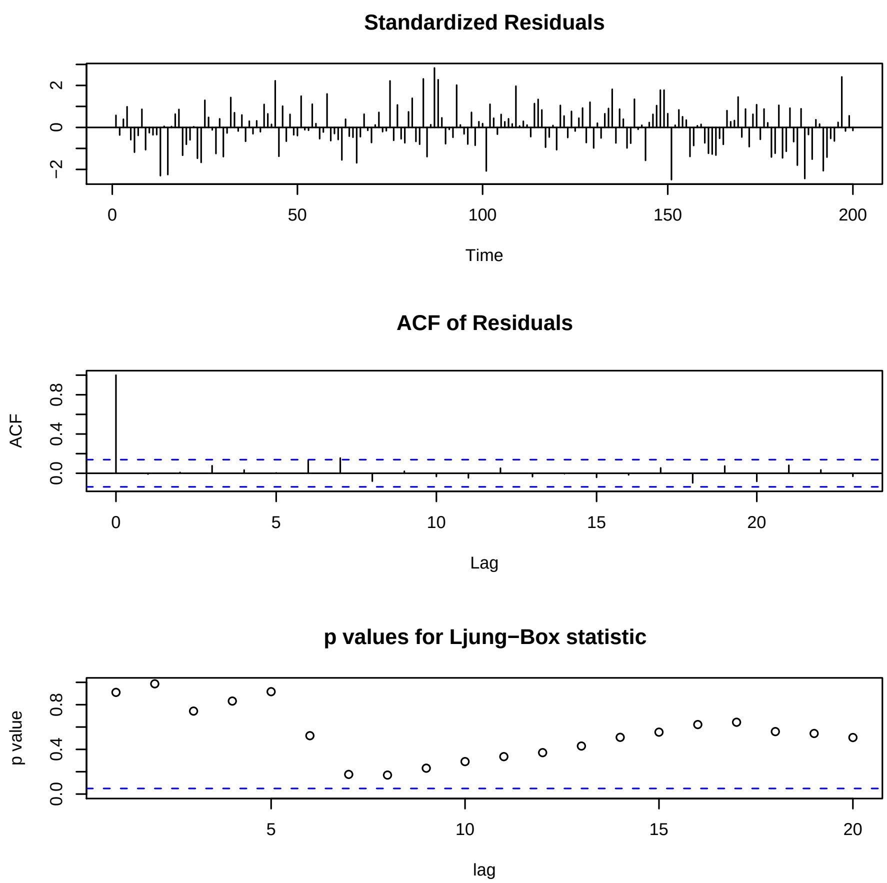

Standardized Residuals
Time
ACF of Residuals
ACF
Lag
p values for Ljung−Box statistic
p value
lag

Figure 2:

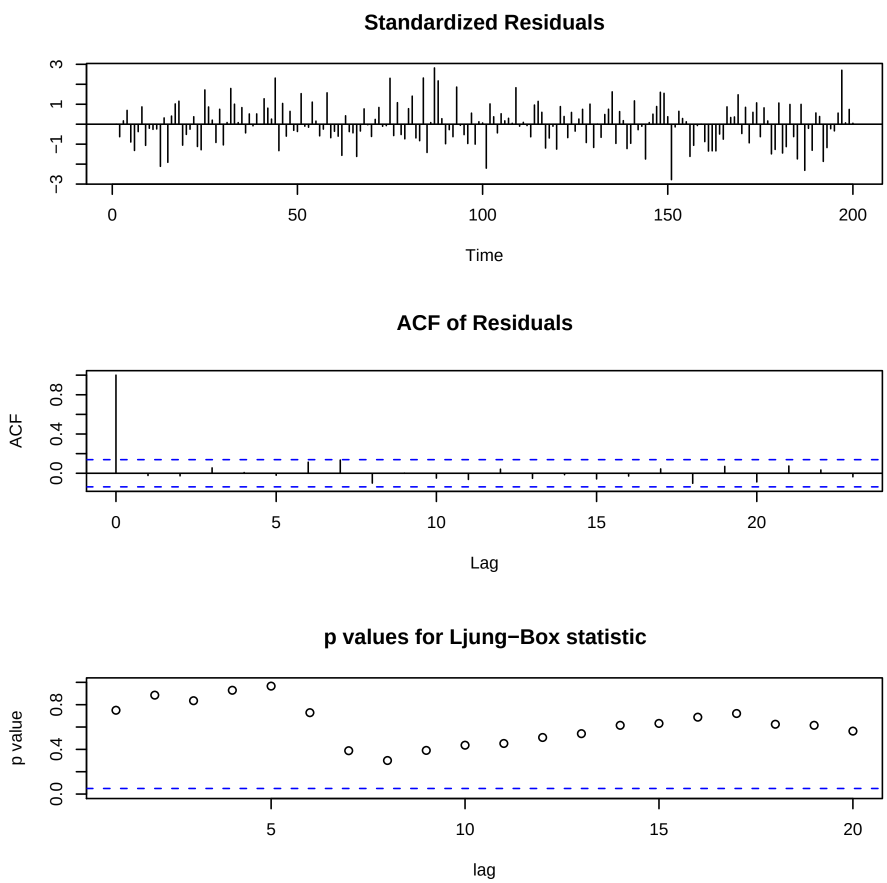

Standardized Residuals
3
1
−1
−3
0
50
100
150
200
Time
ACF of Residuals
ACF
0.0
0.4
0.8
0
5
10
15
20
Lag
p values for Ljung−Box statistic
p value
0.0
0.4
0.8
5
10
15
20
lag

Figure 3:

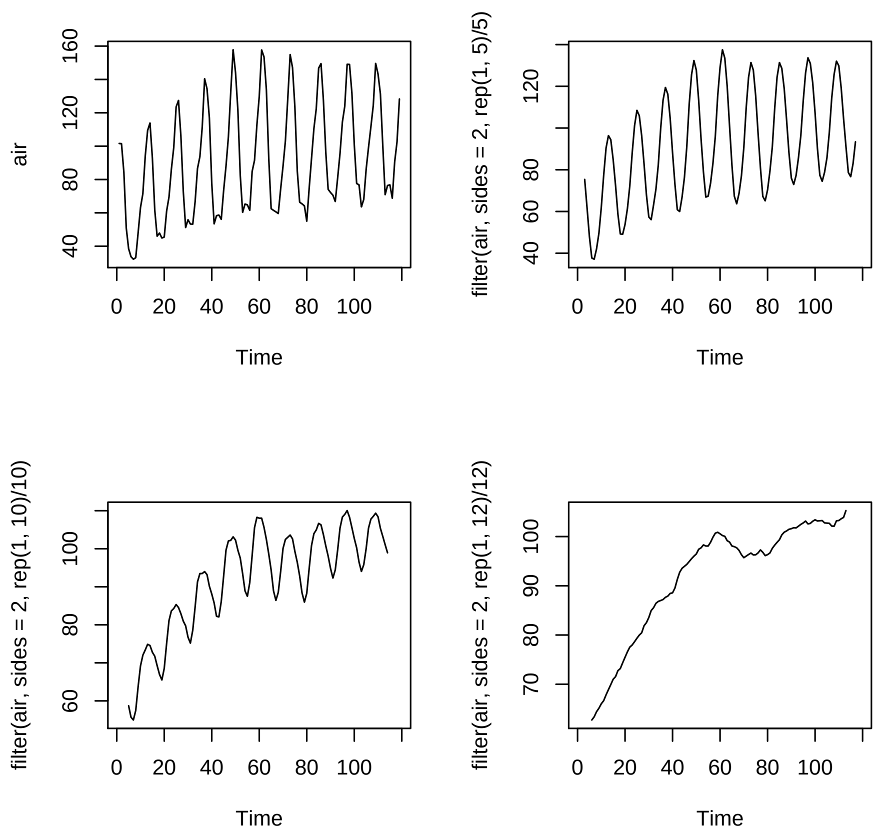

air
filter(air, sides = 2, rep(1, 5)/5)
filter(air, sides = 2, rep(1, 10)/10)
filter(air, sides = 2, rep(1, 12)/12)
Time
Time
Time
Time

Figure 4:

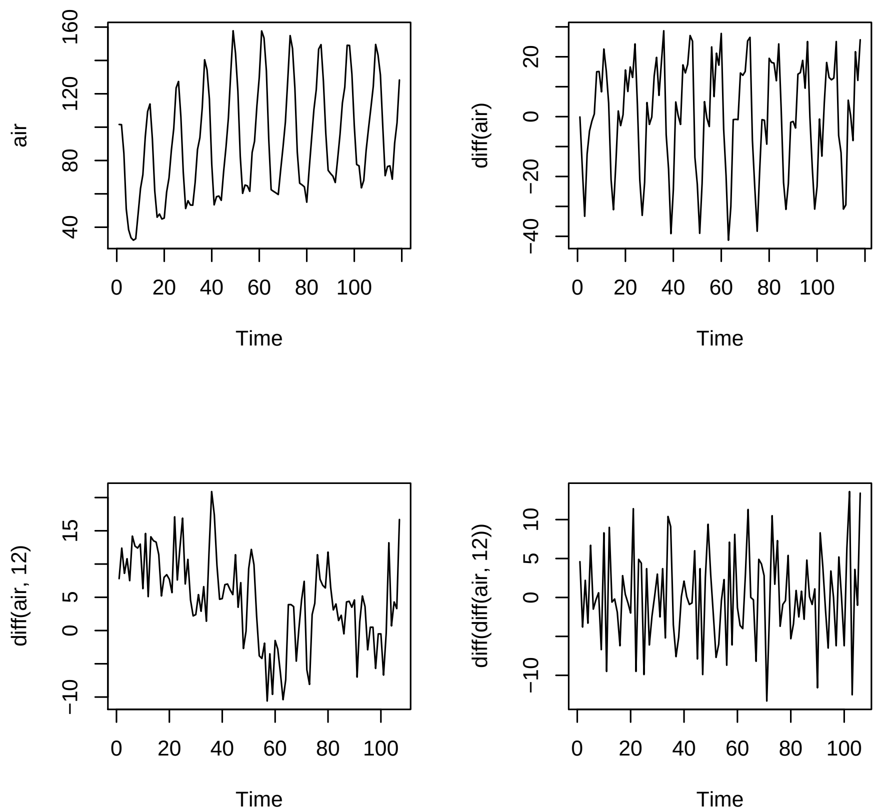

air
160
120
80
40
0 20 40 60 80 100
Time
diff(air)
20
0
−20
−40
0 20 40 60 80 100
Time
diff(air, 12)
15
5
0
−10
0 20 40 60 80 100
Time
diff(diff(air, 12))
10
5
0
−10
0 20 40 60 80 100
Time

Figure 5:

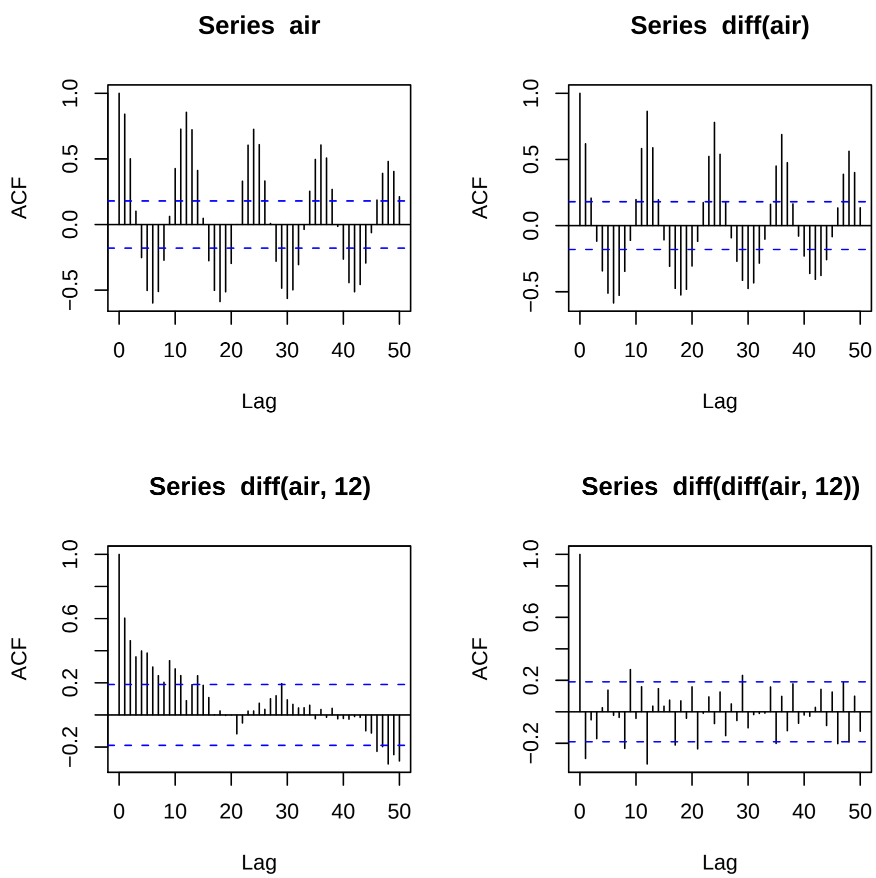

Series air
ACF
Lag
Series diff(air)
ACF
Lag
Series diff(air, 12)
ACF
Lag
Series diff(diff(air, 12))
ACF
Lag

Figure 6:

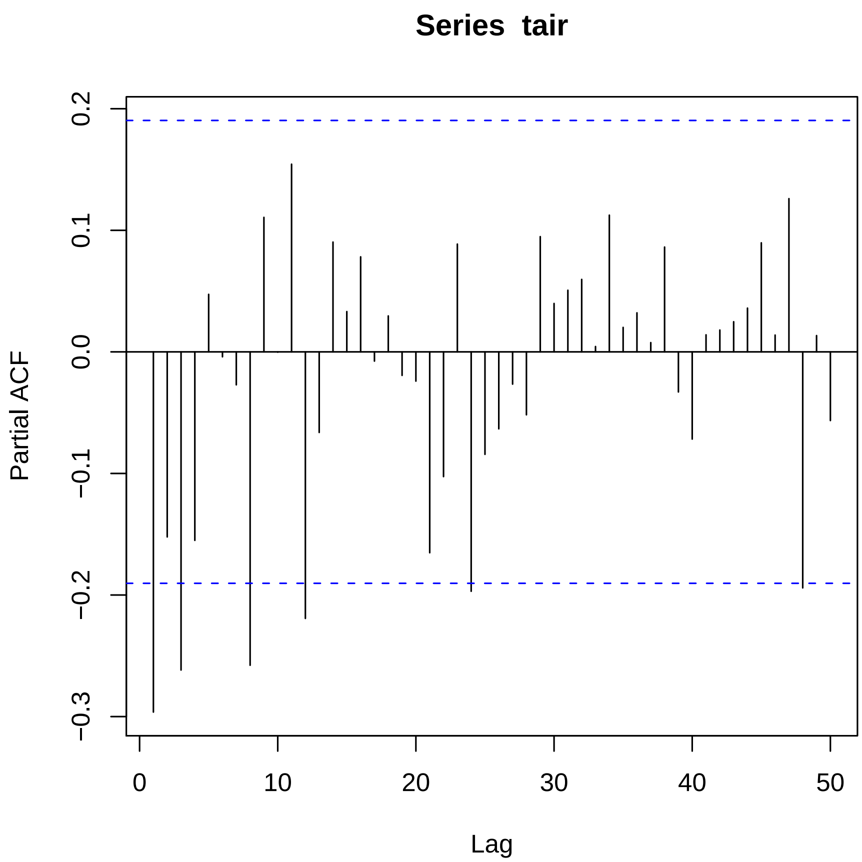


Figure 7:

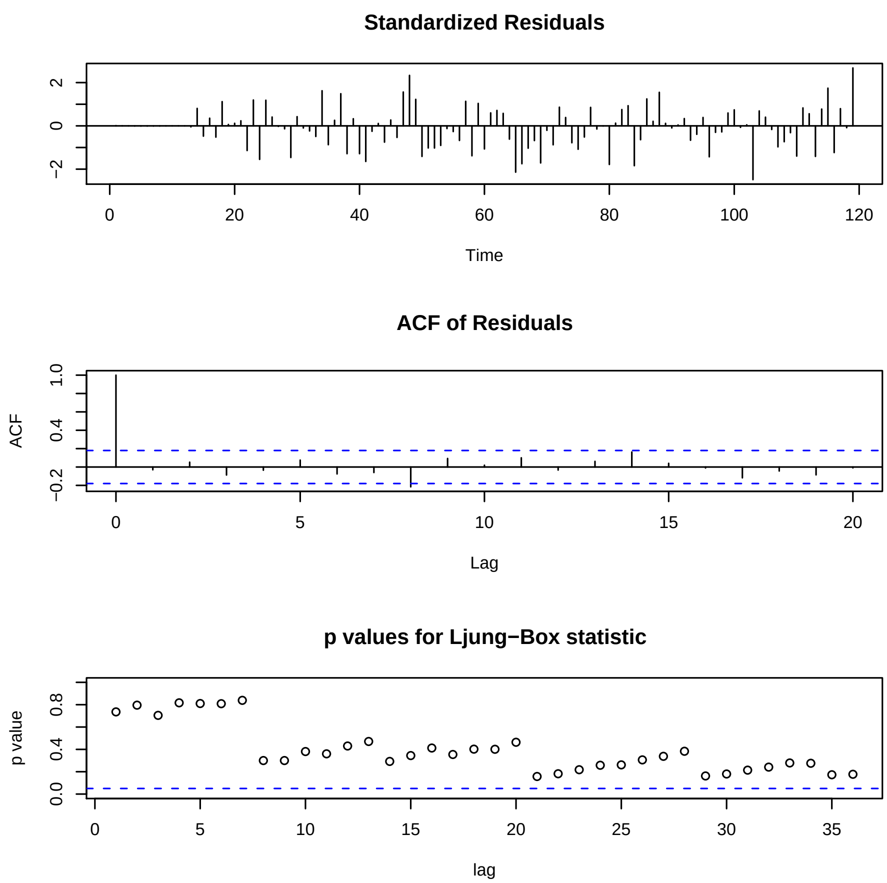

Standardized Residuals
2
0
−2
0
20
40
60
80
100
120
Time
ACF of Residuals
ACF
1.0
0.4
−0.2
0
5
10
15
20
Lag
p values for Ljung−Box statistic
p value
0.8
0.4
0.0
0
5
10
15
20
25
30
35
lag

Figure 8:

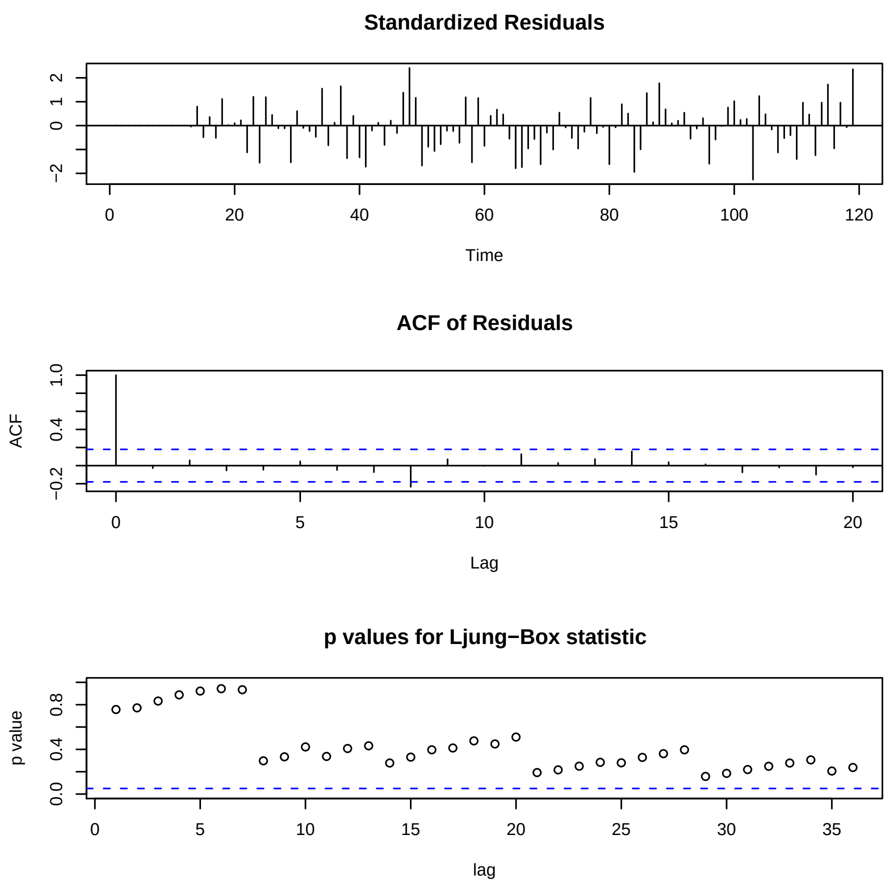

Standardized Residuals
2
1
0
−2
0
20
40
60
80
100
120
Time
ACF of Residuals
ACF
1.0
0.4
−0.2
0
5
10
15
20
Lag
p values for Ljung−Box statistic
p value
0.8
0.4
0.0
0
5
10
15
20
25
30
35
lag

Figure 9:

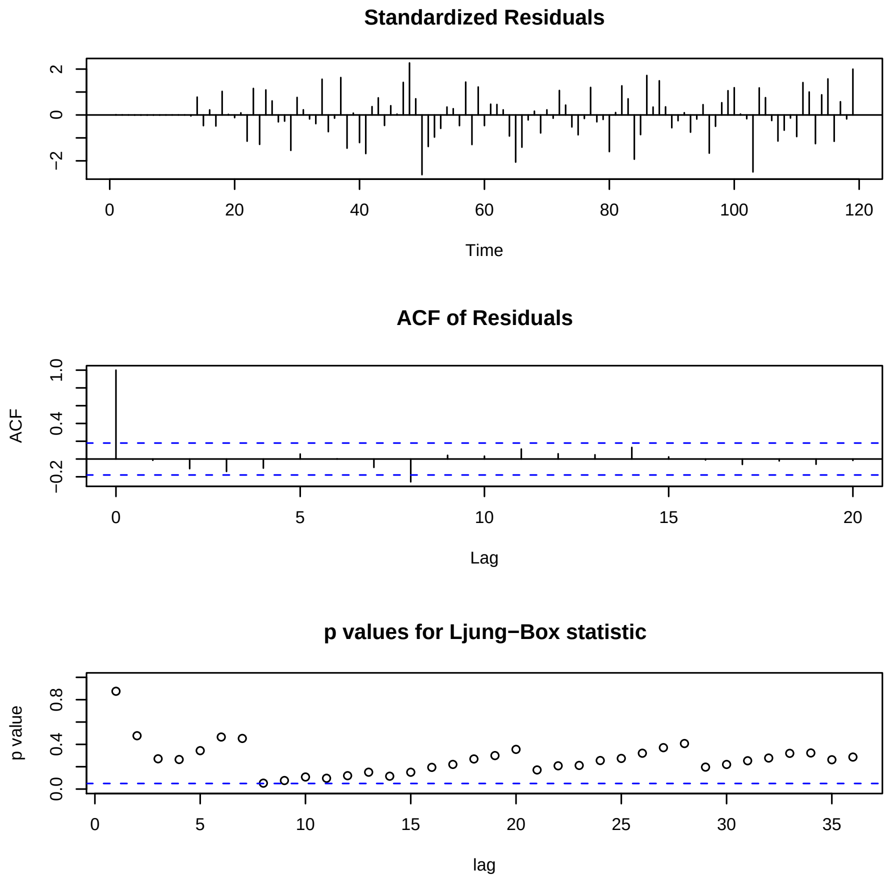

Standardized Residuals
2
0
−2
0
20
40
60
80
100
120
Time
ACF of Residuals
ACF
1.0
0.4
−0.2
0
5
10
15
20
Lag
p values for Ljung−Box statistic
p value
0.8
0.4
0.0
0
5
10
15
20
25
30
35
lag

Figure 10:

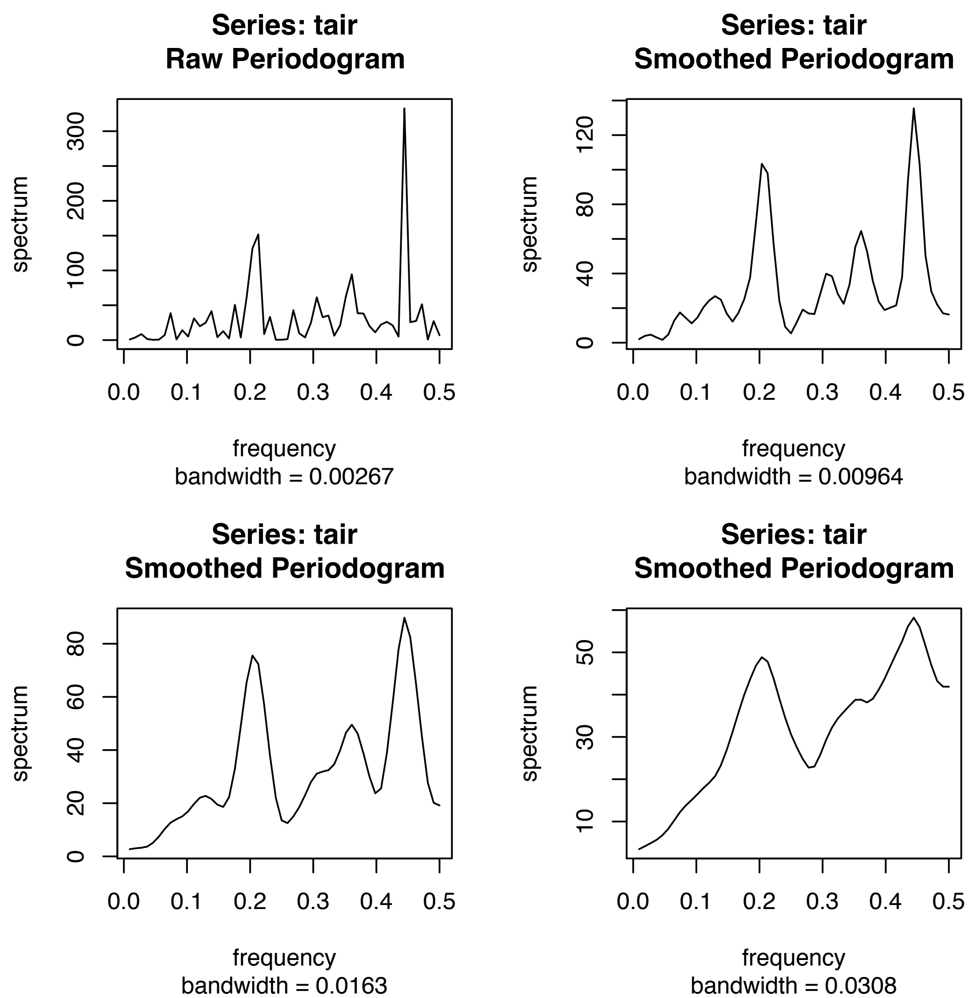
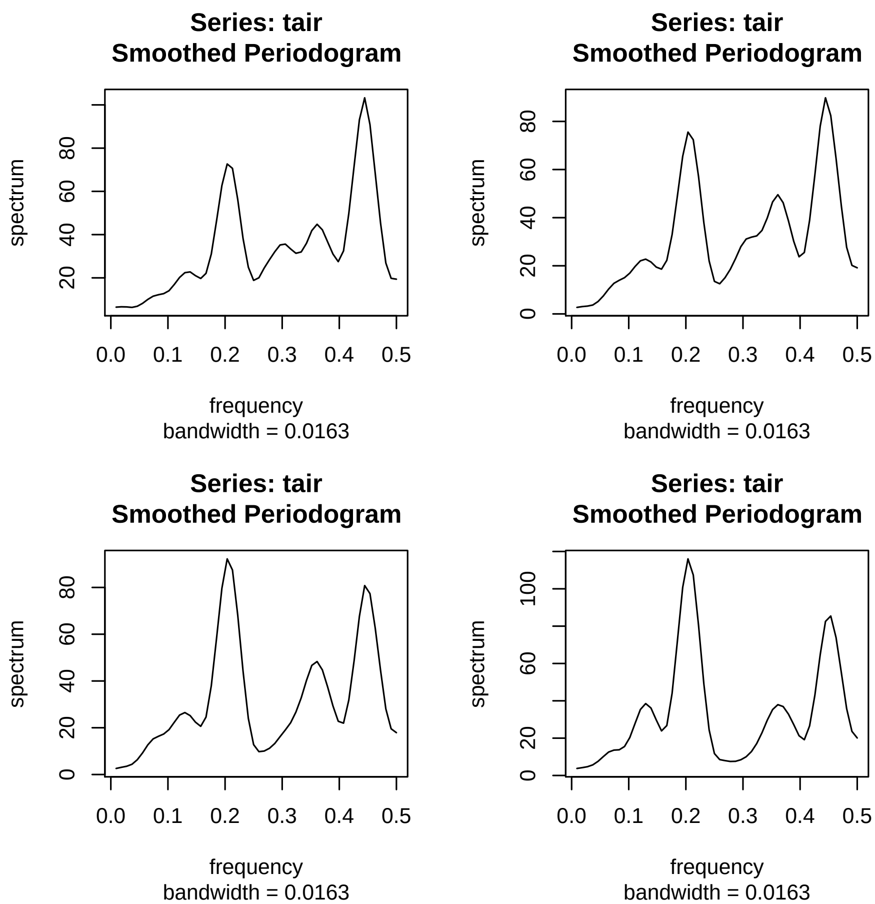

Series: tair
Raw Periodogram
spectrum
0
100
200
300
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.00267
Series: tair
Smoothed Periodogram
spectrum
0
40
80
120
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.00964
Series: tair
Smoothed Periodogram
spectrum
0
20
40
60
80
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0163
Series: tair
Smoothed Periodogram
spectrum
10
30
50
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0308

Figure 11:


Series: tair
Smoothed Periodogram
spectrum
20
40
60
80
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0163
Series: tair
Smoothed Periodogram
spectrum
0
20
40
60
80
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0163
Series: tair
Smoothed Periodogram
spectrum
0
20
40
60
80
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0163
Series: tair
Smoothed Periodogram
spectrum
0
20
60
100
0.0
0.1
0.2
0.3
0.4
0.5
frequency
bandwidth = 0.0163