Statistics 153 (Introduction to Time Series) Homework Three

Due on 3 April, 2013

13 March, 2013

1. Consider a dataset of size n generated according to the zero-mean AR(1) model with parameter ϕ. We have seen in class that the Yule-Walker estimate $\hat{\phi}$ of ϕ is approximately normal with mean ϕ and variance $(1-\phi^2)/n$. In this problem, we check this result via simulation. Repeat the following sequence of steps for $n = 20, 60, 100$ and $\phi = 0.2, 0.6, 0.95$ (there are 9 combinations in all)

   (a) Generate 10000 samples each of size n from the AR(1) model with parameter ϕ.

   (b) For each of these 10000 samples, calculate the Yule-Walker estimate of ϕ.

   (c) Report the sample mean and sample variance of these 10000 Yule-Walker estimates. How close are they to ϕ and $(1-\phi^2)/n$ respectively?

   Based on your simulations, comment on the accuracy of this approximation result for the distribution of the Yule-Walker estimates.

2. Consider a dataset of size n generated according to the zero mean MA(2) model with parameter ϕ. We have seen in class that the MLE $(\hat{\theta}_1, \hat{\theta}_2)$ of (θ_1, θ_2) is approximately normal with mean (θ_1, θ_2) and a certain variance-covariance matrix. In this problem, we check this result via simulation. Repeat the following sequence of steps for $n = 20, 60, 100$ and $(\theta_1, \theta_2) = (1.5, 0.75)$ and $(0.95, 0)$ (there are 6 combinations in all)

   (a) Generate 10000 samples each of size n from the MA(2) model with parameters θ_1 and θ_2.

   (b) For each of these 10000 samples, calculate the ML estimates of θ_1 and θ_2.

   (c) You should have 10000 estimates for the pair (θ_1, θ_2). Calculate the mean and variance of these 10000 estimates of θ_1 and θ_2. Also calculate the covariance between the 10000 estimates of θ_1 and θ_2. How close are these values to the values given by the form of the variance-covariance matrix of the ML estimate of (θ_1, θ_2)?

   Based on your simulations, comment on the accuracy of this approximation result for the distribution of the ML estimates for the MA(2) model.

3. A time series data set is plotted in the top panel of Figure 1. Its sample autocorrelation and partial autocorrelation functions are plotted in the middle and bottom panels of Figure 1 respectively.

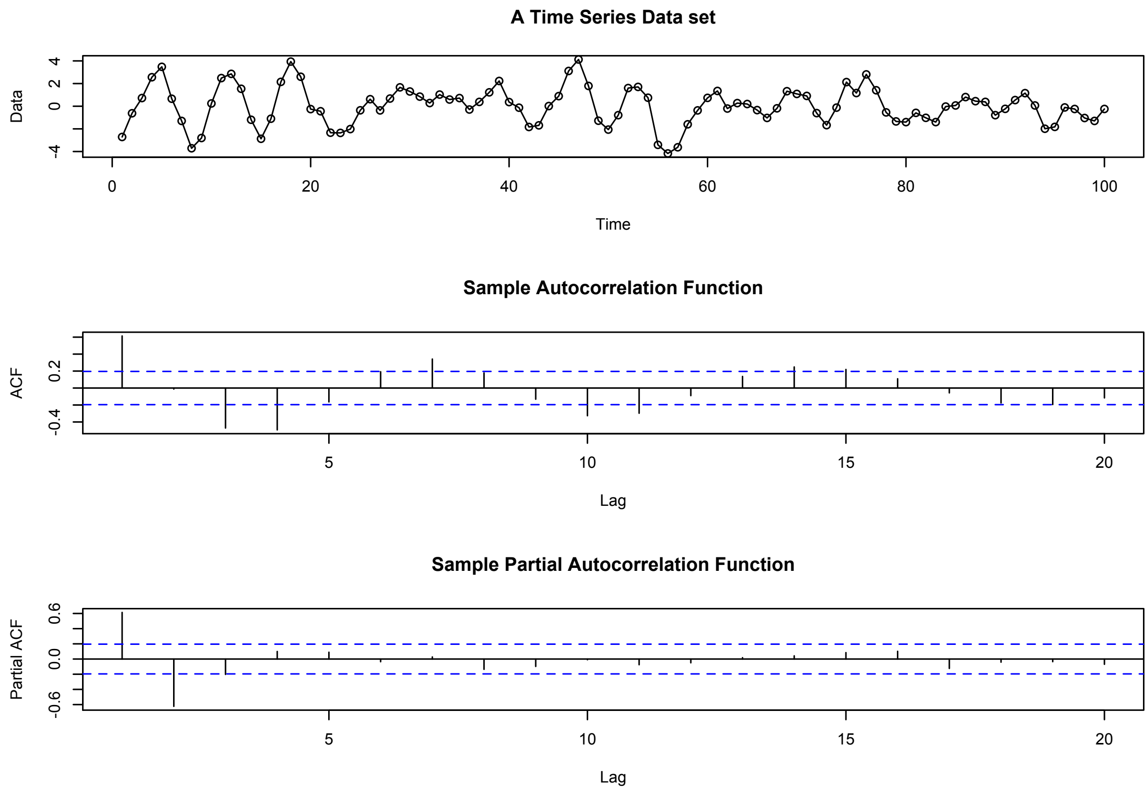

Figure 1: A Time Series Data set along with its sample autocorrelation and partial autocorrelation functions

(a) Is it a good idea to fit the AR(2) model to this dataset? Why or why not?

(b) I fitted the AR(2) model to this dataset using the following function in **R**:

```
ar(dataset, aic = F, order.max = 2)
```

and it gave the following output

```
ar(x = dataset, aic = F, order.max = 2)

Coefficients:
      1         2
   0.9968    XXXXX

Order selected 2  sigma^2 estimated as  1.089
```

The estimate for the second coefficient is missing in the output above. Based on the available information, provide, with reasoning, a good estimate for the second coefficient in the AR(2) model?

4. I generated 1000 observations from the MA(1) model with parameter 0.7 using the following **R** function:

```
dataset = arima.sim(n = 1000, list(ma = 0.7))
```

I then fitted the ARMA(1, 2) model to the data using the function:

```
arima(dataset, order = c(1, 0, 2))
```

which gave the following output:

```
Call:
arima(x = dataset, order = c(1, 0, 2))

Coefficients:
         ar1      ma1      ma2  intercept
      0.7838  -0.0802  -0.5301     0.0430
s.e.  0.2063   0.2107   0.1501     0.0566

sigma^2 estimated as 0.986:  log likelihood = -1412.22,  aic = 2832.45
```

Because the data have been simulated from an MA(1) model with parameter 0.7, I expected the estimated coefficients of `ar1` and `ma2` to be close to zero and the estimated coefficient of `ma1` to be close to 0.7. Explain why this did not happen?

5. A time series data set is plotted in the top panel of Figure 2. Its sample autocorrelation and partial autocorrelation functions are plotted in the middle and bottom panels of Figure 2 respectively.

   For each of the following models given by their **R** functions, provide reasons for using or not using that model for this dataset:

   (a) `arima(dataset, order = c(0, 0, 7))`

   (b) `arima(dataset, order = c(0, 0, 1), seasonal = list(order = c(0, 0, 1), period = 6))`

   (c) `arima(dataset, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1), period = 6))`

6. The following is the `co2` dataset from the package TSA which contains the monthly CO2 level at Alert, Canada from 01/1994 to 12/2004.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1994	363.05	364.18	364.87	364.47	364.32	362.13	356.72	350.88	350.69	356.06	360.09	363.27

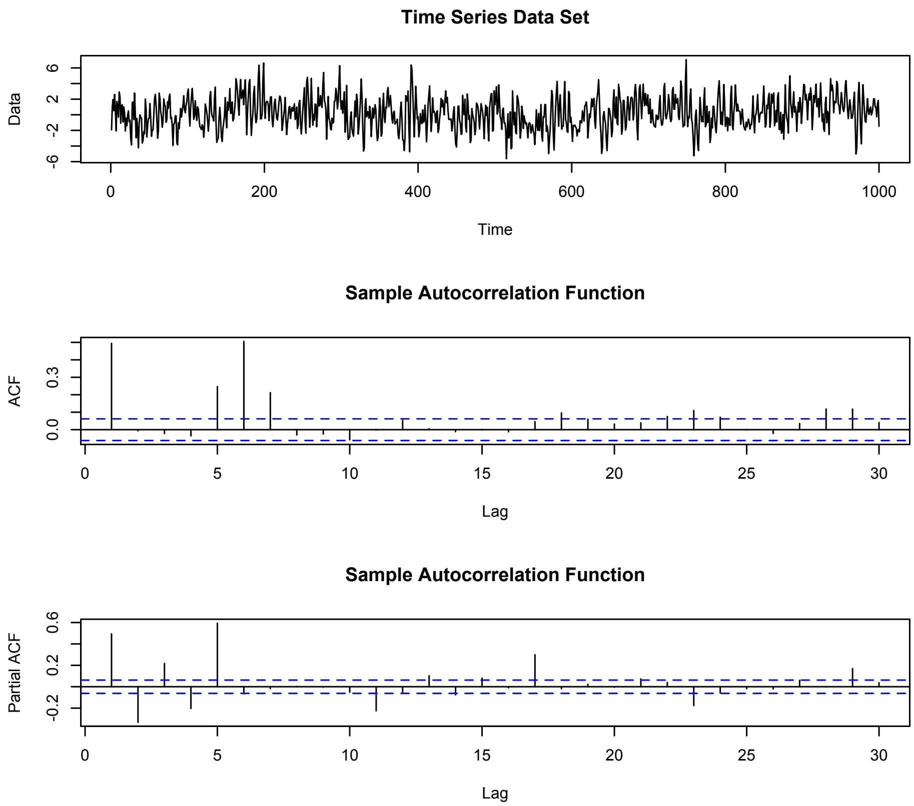

Figure 2: A Time Series Data set along with its sample autocorrelation and partial autocorrelation functions

1995	363.49	364.94	366.72	366.33	365.75	364.32	358.59	352.06	353.45	357.27	362.34	365.65
1996	366.93	366.71	367.63	368.15	369.14	367.33	361.53	356.11	354.51	360.12	363.85	365.52
1997	367.72	369.08	368.17	368.83	369.49	367.57	360.79	355.16	356.01	360.71	364.77	367.81
1998	369.40	370.12	370.88	370.53	371.56	369.28	364.50	357.46	360.54	364.04	368.74	371.58
1999	372.60	373.85	373.75	374.10	374.50	372.04	364.81	359.11	359.65	364.94	369.82	372.62
2000	373.23	375.13	374.83	375.42	376.18	374.01	366.54	360.78	361.77	367.51	370.58	373.37
2001	375.49	375.94	376.42	377.48	377.67	374.78	367.38	361.67	363.39	367.74	373.18	374.41
2002	376.68	377.42	378.27	378.73	379.01	375.95	370.78	364.07	365.36	370.25	374.04	377.99
2003	379.03	379.36	380.90	381.39	382.38	381.02	373.78	367.97	368.55	372.28	377.75	379.99
2004	382.44	382.36	381.58	383.21	383.58	382.59	374.58	368.69	368.55	373.39	378.49	381.62

Consider the following **R** code and output:

```
diff.data = diff(diff(co2, 12))
model = arima(diff.data, order = c(0, 0, 1), seasonal = list(order =
c(0, 0, 1), period = 12))
```

```
predict(model, n.ahead = 1)
$pred
           Jan
2005 -1.168874

$se
           Jan
2005 0.7399812
```

What would be a good prediction for the CO2 level at Alert in January, 2005? Explain.