# HW 9

Brandon Fenton, Allison Theobold and Matthew Pettigrew

Due: December 7, 2016

A record exists of the ice on and ice off dates for the Baikal Lake in Siberia from 1869 to 1996. We will focus on the duration of ice being present on the lake, which is a count of days for each year. There are no missing values in the time series.

Note that these are yearly observations and that the response is number of days of ice on the lake in each year. A researcher might be interested in finding some sort of long-term periodic behavior in these results as well as assessing evidence related to a trend.

1) Make a nice looking time series plot of the time series.
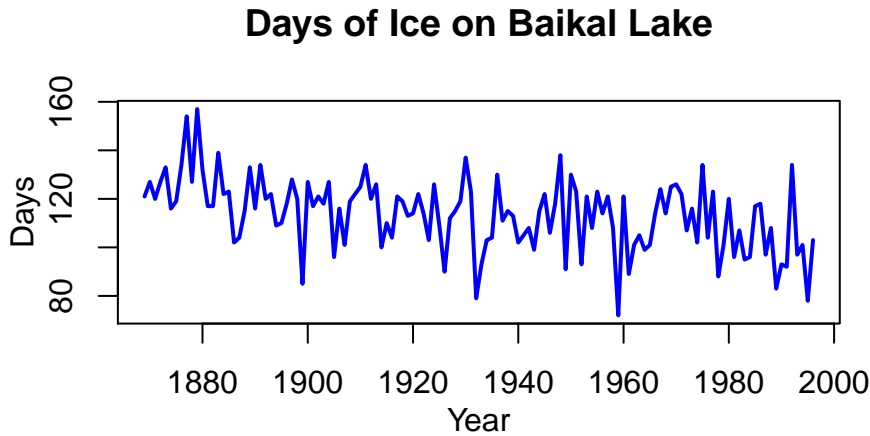
2) Fit an ARMA(17,0) using the arima function and describe the roots discussing

a) stationarity and

b) potential quasi-periodicities.
   To help with (b), you should make a table and sort the roots based on the potential QP.

Table 1: Table continues below

	Roots	Mod(Roots)	Period	Stationarity
6	-1.021+0.202i	1.041	2.133	Stationary
10	-1.021-0.202i	1.041	2.133	Stationary
14	-0.8942-0.5638i	1.057	2.436	Stationary
2	-0.8942+0.5638i	1.057	2.436	Stationary
7	-0.5934-0.9016i	1.079	2.918	Stationary
13	-0.5934+0.9016i	1.079	2.918	Stationary
3	-0.353-1.026i	1.085	3.302	Stationary
9	-0.353+1.026i	1.085	3.302	Stationary
5	0.153+1.095i	1.105	4.389	Stationary

	Roots	Mod(Roots)	Period	Stationarity
15	0.153-1.095i	1.105	4.389	Stationary
1	0.52+0.908i	1.046	5.98	Stationary
11	0.52-0.908i	1.046	5.98	Stationary
4	0.8299-0.7067i	1.09	8.907	Stationary
16	0.8299+0.7067i	1.09	8.907	Stationary
17	1.067-0.36i	1.126	19.31	Stationary
12	1.067+0.36i	1.126	19.31	Stationary
8	1.024-0i	1.024	4.537e+16	Stationary

	Quasi-Period
6	2.13 years
10	2.13 years
14	2.44 years
2	2.44 years
7	2.92 years
13	2.92 years
3	3.3 years
9	3.3 years
5	4.39 years
15	4.39 years
1	5.98 years
11	5.98 years
4	8.91 years
16	8.91 years
17	19.3 years
12	19.3 years
8	45370000000000000 years

The table above summarizes the roots of the AR(17) model, their stationarity, and their quasi-periodicity. There is only one real root, and all of the roots have modulus greater than 1 (therefore stationary). The quasi-periodicity of the roots are also estimated and ordered from least to greatest. The roots with quasi-periods less than about 3 years could be thought to be due to local dependency, and not demonstrating true quasi-periodicity.

3) One method for selecting the order of an ARMA model is to use hypothesis tests on the highest order coefficient and drop down if there is no evidence that the coefficients are different from 0. We can also use confidence intervals to do our test by assessing whether our value of interest under the null is inside or outside of the interval. Even though you are using a 95% confidence interval, you are still doing a 5% significance level test. The confint function can be used to generate confidence intervals for coefficients from many models in R, including arima-fit models. Fit an AR(18) model and use the confidence interval to discuss the need for the 18th lag in the model. Then repeat the confidence interval of interest for the previous AR(17) model and discuss the similar result – what does this suggest about the AR(17) model.

We see in the table below that the 95% confidence interval for ϕ_{18} contains 0, therefore leading us to conclude that and AR(17) model may be sufficient. The second entry in the table is a confidence interval for ϕ_{17}, which does not contain 0. Thus, we conclude that an AR(17) model is a better fit for these data.

```
             2.5 %   97.5 %
--------------- ------- --------
 **Lag 18 CI**  -0.1736  0.1909


 **Lag 17 CI**   0.1124   0.4709
---------------------------------
```

4) Report the estimate and 95% confidence interval for the "intercept" that R reports for the AR(17) model. Interpret the estimate and then calculate the actual estimated intercept in the AR(17) model based on the results. Show your work.

Below is the estimate {R} provides for μ, alone with the 95% confidence interval.

The true intercept can be calclated from the above model summary as follows,

$$\theta_0 = \mu(1 - \phi_1 - \phi_2 - \cdots - \phi_17)$$

$$= 113.8428(1 - \sum_{i=1}^{1} 7\phi_i)$$

$$= 17.93473$$

5) For the AR(17) model (do not re-fit the model), how many of the lower order autoregressive coefficients are not detectably different from 0 in this model using the confidence intervals?

	2.5 %	97.5 %
ar1	-0.01159	0.3183
ar2	0.03827	0.375
ar3	-0.1463	0.1944
ar4	-0.2634	0.0806
ar5	-0.152	0.1954
ar6	0.1089	0.4596
ar7	-0.09462	0.2729
ar8	-0.3411	0.02755
ar9	-0.08955	0.2835
ar10	-0.1155	0.2555
ar11	-0.2432	0.1333
ar12	-0.1382	0.231
ar13	-0.1111	0.2579
ar14	-0.2675	0.0959
ar15	-0.1819	0.1859
ar16	-0.3085	0.05235
ar17	0.1124	0.4709

We see in the table above that for the AR(17) model, all but 3 of the coefficient confidence intervals contain 0.

6) To explore long-term QP, you have to consider somewhat higher order AR() models. But sometimes we can get similar fits to the time series from mixed ARMA models. Report the ACF, PACF, and EACF and discuss models that are suggested by these results.

We will discuss the PACF and EACF in class on Tuesday so bring questions to class on using them.

Series bts
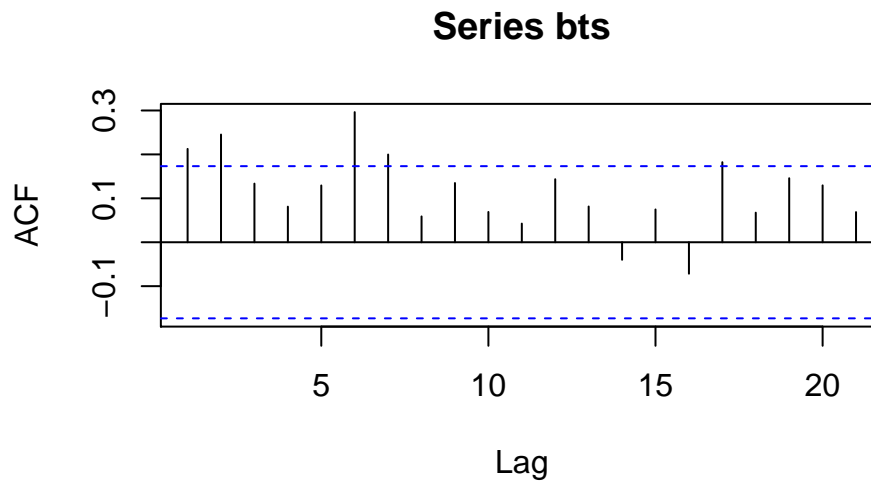
The first two lags have sample acf values which are above the cutoff, but the values for lags greater than two don't really become small enough to warrant the claim that there is dampening or a clear cutoff point after any particular lag number.
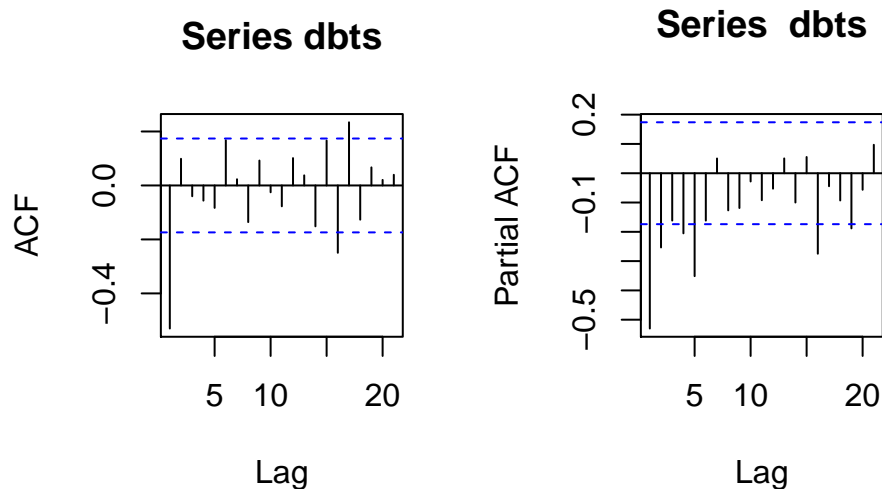
Series bts
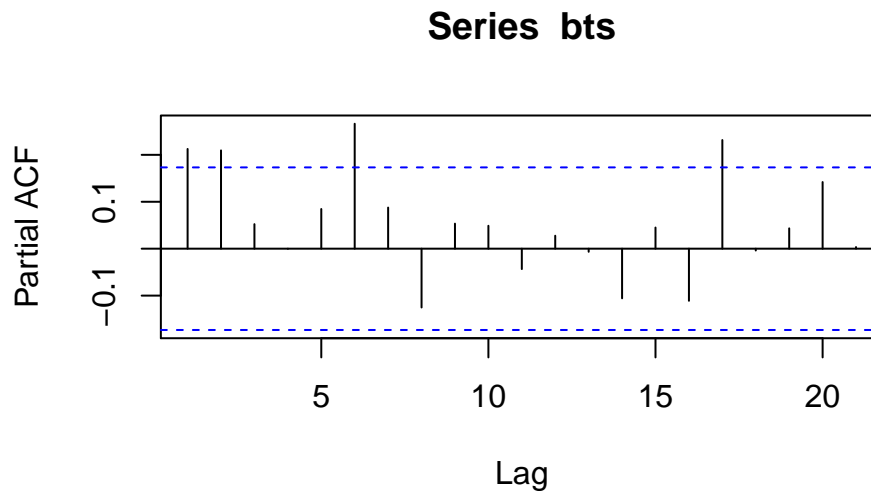

```
eacf(bts)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x o o o o x x o o o  o  o  o  o
## 1 x o o o o o o o o o o  o  o  o  o
## 2 x o o o o o o o o o o  o  o  o  o
## 3 o o o o o o o o o o o  o  o  o  o
## 4 o o o x o o o o o o o  o  o  o  o
## 5 x o o x x o o o o o o  o  o  o  o
## 6 x x x o o o o o o o o  o  o  o  o
```

```
## 7 x o x x o o o o o o o  o   o   o
```

The simplest "clean" model suggested by the eacf output above is an AR(1,1) model, but other models marked with an o in the upper left corner (an MA(2) model, for example) would also be suitable candidates.

7) Consider taking a first difference in the series and repeating the diagnostics from #6. Use the diff function to do the differencing.

Series dbts

Series dbts

```
## AR/MA
##    0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o  o   o   o
## 1 x o o o o o o o o o o  o   o   o
## 2 x o o o o o o o o o o  o   o   o
## 3 x o o x o o o o o o o  o   o   o
## 4 x o x x x o o o o o o  o   o   o
## 5 x x x o o o o o o o o  o   o   o
## 6 x o x x o o o o o o o  o   o   o
## 7 x x x o o o o o o o o  o   o   o
```

The SACF for the first difference has a large spike a lag 1 and then fairly much drops off. There are also other large correlation values at lags 6 and 15 through 17. The SPACF has large spikes at the first two lags and lags 4 and 5. Overall the SPACF exhibits damping behavior. In the EACF table there is a distinct upper left vertex at ARMA(0,1) suggesting an MA(1) model. Based on the three diagnostics an MA(1) seems like the strongest candiate, but larger MA processes could also be considered.

8) This should suggest a particular MA model. Fit that model using arima on the differenced series. In other words, use arima(dbts,order=c(0,0,q)) where you need to determine q. Write out the estimated model based on the model results both in terms of the differenced series and the original time series.

9) Report the confidence interval for the intercept from the model in #8. Discuss what the test using this confidence interval tells you in this model and how it addresses one of the questions of interest to the researcher.

R Code

```
Setup
```

```
Problem 1
```

```
Problem 2
```

```
Problem 3
```

```
Problem 4
```

```
Problem 5
```