

# Time Series HW 9

Andrea Mack and Kenny Flagg

December 7, 2016

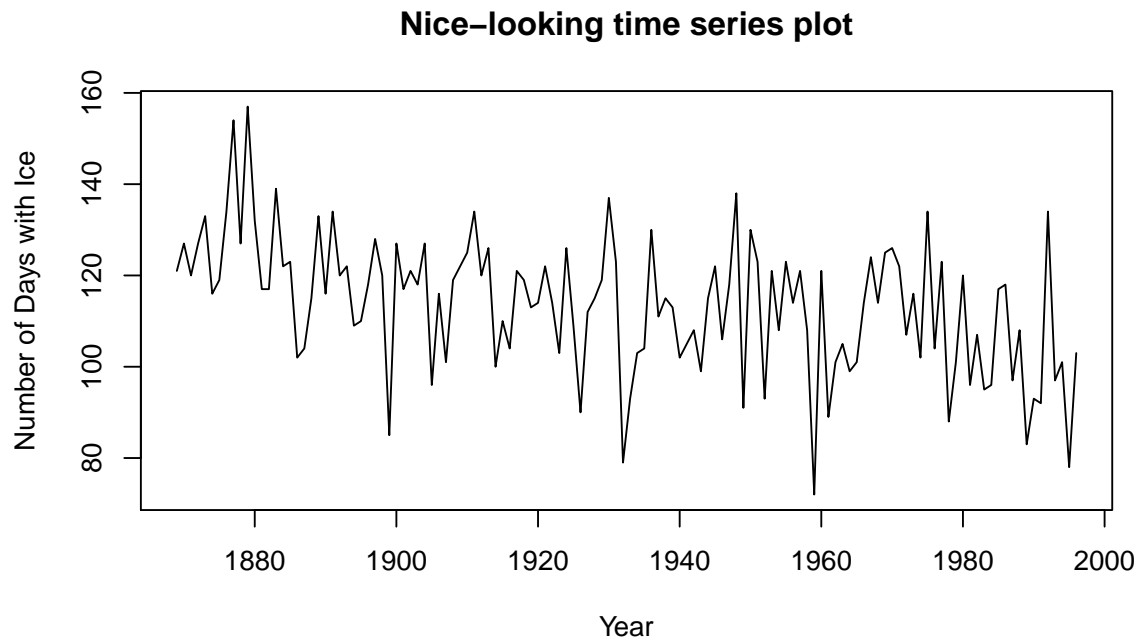
*A record exists of the ice on and ice off dates for the Baikal Lake in Siberia from 1869 to 1996. We will focus on the duration of ice being present on the lake, which is a count of days for each year. There are no missing values in the time series.*

*You can access this data set via:*

```
bts <- read.csv("baikal.csv", header = TRUE)$Days
```

*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*.

	Estimate	SE
ar1	0.15	0.08
ar2	0.21	0.09
ar3	0.02	0.09
ar4	-0.09	0.09
ar5	0.02	0.09
ar6	0.28	0.09
ar7	0.09	0.09
ar8	-0.16	0.09
ar9	0.10	0.10
ar10	0.07	0.09
ar11	-0.05	0.10
ar12	0.05	0.09
ar13	0.07	0.09
ar14	-0.09	0.09
ar15	0.00	0.09
ar16	-0.13	0.09
ar17	0.29	0.09
intercept	113.84	5.54

Root	Modulus	QP
-1.021+0.202i	1.04	2.13
-1.021-0.202i	1.04	2.13
-0.894-0.564i	1.06	2.44
-0.894+0.564i	1.06	2.44
-0.593-0.902i	1.08	2.92
-0.593+0.902i	1.08	2.92
-0.353-1.026i	1.09	3.30
-0.353+1.026i	1.09	3.30
0.153+1.095i	1.11	4.39
0.153-1.095i	1.11	4.39
0.520+0.908i	1.05	5.98
0.520-0.908i	1.05	5.98
0.830-0.707i	1.09	8.91
0.830+0.707i	1.09	8.91
1.067-0.360i	1.13	19.31
1.067+0.360i	1.13	19.31
1.024-0.000i	1.02	

All of the roots have modulus  $> 1$ , implying that the process is stationary. One root is real, and the remaining roots are complex and form 8 conjugate pairs. The process has components with quasi-periods ranging from of 2.13 to 5.98 years that may represent true physical quasi-periodicity. Complex roots resulting in quasi-periods of 8.91 and 19.3 years are also present, but jump quite a bit and may not represent true physical 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?

	Limits AR(18)
2.5 %	-0.17
97.5 %	0.19

Zero is within the 95% confidence interval for the AR(18) term indicating no evidence that it is needed in the model.

	Limits AR(17)
2.5 %	0.11
97.5 %	0.47

The entire 95% confidence interval for the AR(17) term is above zero, indicating a strong evidence that the AR(17) term is positive.

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.

The intercept in the AR(17) model is estimated to be 113.84 with a standard error of 5.54. The associated 95% confidence interval is from 102.9768021 to 124.7087696 when an AR(17) model was fit. That is, we are 95% confident the long run mean when there is no autocorrelation present is between 102.9768021 and 124.7087696 days of ice cover a year.

The average number of days ice was on Baikal Lake per year from 1869 to 1996 was estimated **or population data?** to be 113.85 with an associated 95% confidence interval of 102.9768021 to 124.7087696 days per year. The observed average number of days per year was 113.57, which is very similar to the model estimate.

The model fit was:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \dots + \phi_{17} y_{t-17} + e_t$$

Where  $\phi_0$  is the intercept reported by ARIMA(17, 0, 0). Let  $E[y_t] = \mu$  for all  $t$  and  $t-i$ .

Then,  $\mu = \phi_0 + \phi_1 \mu + \dots + \phi_{17} \mu + 0$

The intercept that represents the long run mean is then  $\mu = \phi_0 \over 1 - \phi_1 - \dots - \phi_{17}$

The actual estimated mean is then:

17.935

*Kenny check:* I believe the difference in interpretation comes in as one is the long run average with autocorrelation present, and the other is not.

The binomial model may be more reasonable for the count data in a finite space.

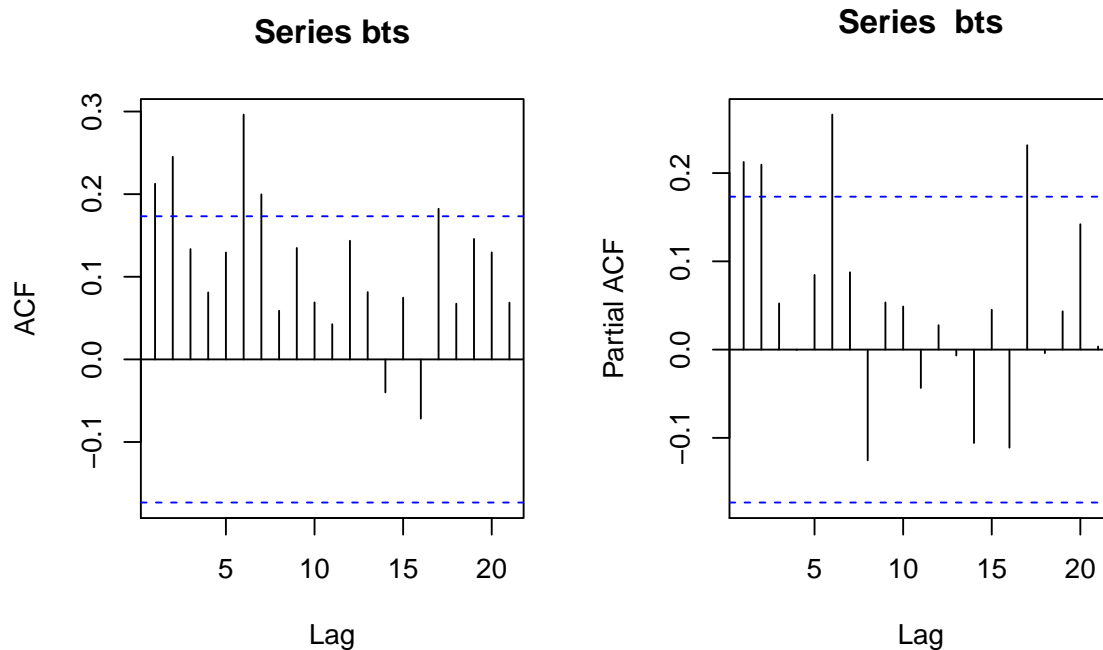
Reference: <http://www.stat.pitt.edu/stoffer/tsa2/Rissues.htm>

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?

	Estimate	Lower 95% Limit	Upper 95% Limit
ar1	0.15	-0.01	0.32
<b>ar2</b>	0.21	0.04	0.38
ar3	0.02	-0.15	0.19
ar4	-0.09	-0.26	0.08
ar5	0.02	-0.15	0.20
<b>ar6</b>	0.28	0.11	0.46
ar7	0.09	-0.09	0.27
ar8	-0.16	-0.34	0.03
ar9	0.10	-0.09	0.28
ar10	0.07	-0.12	0.26
ar11	-0.05	-0.24	0.13
ar12	0.05	-0.14	0.23
ar13	0.07	-0.11	0.26
ar14	-0.09	-0.27	0.10
ar15	0.00	-0.18	0.19
ar16	-0.13	-0.31	0.05
<b>ar17</b>	0.29	0.11	0.47

Of the 16 lower order terms, 14 were not detectably different from 0. Those whose 95% confidence intervals exclude zero are bolded in the table above.

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.

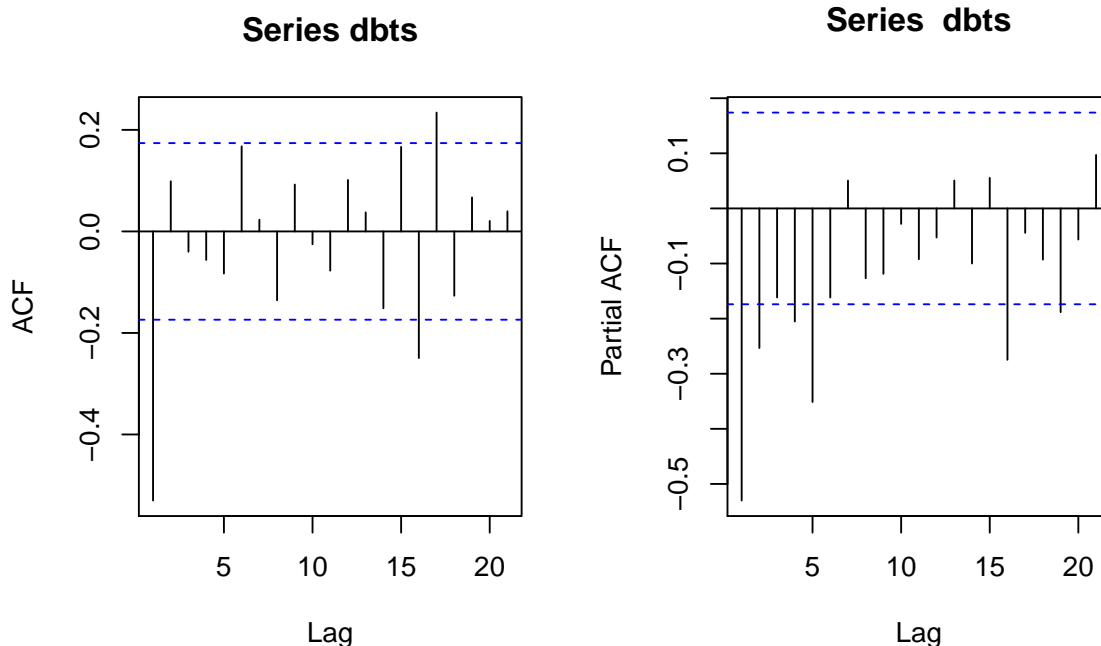


AR/MA														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	o	o	o	x	x	o	o	o	o	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	o	o	o	o	o	o	o	o	o	o	o	o	o	o
4	o	o	o	x	o	o	o	o	o	o	o	o	o	o
5	x	o	o	x	x	o	o	o	o	o	o	o	o	o
6	x	x	x	o	o	o	o	o	o	o	o	o	o	o
7	x	o	x	x	o	o	o	o	o	o	o	o	o	o

The PACF shows significant lags at 1 and 2, so an AR(2) model, while the EACF suggests an AR(3) model. Because the ACF cuts off after lag 2, it suggests an MA(2) model.

Altogether, we would suggest starting with an ARMA(2,0) model.

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



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	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	x	o	o	x	o	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	o
5	x	x	x	o	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	o
7	x	x	x	o	o	o	o	o	o	o	o	o	o	o	o

All three plots indicate an ARMA(0,1) model for the first differences.

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.

Original MA(1) model:  $y_t = 113.57 + 0.151e_{t-1} + e_t$

Differenced MA(1) model:

$$y_{diff_t} = -0.177 + -1e_{diff_{t-1}} + e_{diff_t}$$

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.

The estimated 95% confidence interval for the intercept from the MA(1) model of the data is between  $-0.238$  and  $-0.115$ .