Stat 5170: Assignment 4

due Feb 17, at start of class

1 R Tutorial

In this tutorial, we will learn how to use several R commands to smooth a time series and remove trends.

1. (a) Start by creating a new folder "assignment4" and download the file "gas.dat". This data file contains monthly heating oil prices from July 1973 to December 1987. Load the data into R using the command

```
gas<-scan("gas.dat")
```

Plot the data. Comment on this time series.

(b) We can take a centered moving average of the data by using the `filter()` command in R. Take the moving average over thirteen observations using the command

```
magas<-filter(gas,sides=2,rep(1,13)/13)
```

We can overlay this moving average over the plot of the data. Type

```
plot(gas)
lines(magas, col="red")
```

You can add an optional argument to change the color of the overlay as a visual aid.

(c) We can also perform a one-sided moving average of the data in R. For example,

```
m2<-filter(gas,sides=1,rep(1,12)/12)
lines(m2, lty=2, col="blue")
legend("topleft", c("centered","1-sided"), lty=c(1,2), col=c("red","blue"))
```

The `legend()` function allows us to overlay a legend.

(d) Plot the residuals (`gas-magas`). Do these residuals appear stationary?

(e) From now on we will be working with the logarithm of the data. Use `lgas<-log(gas)` to do this. Let us now recalculate the moving average using the following commands. We also plot the residuals, and the ACF of the residuals.

```
malgas<-filter(lgas,sides=2,rep(1,13)/13)
plot(lgas)
lines(malgas)
resmalgas<-lgas-malgas
resmalgas<-resmalgas[!is.na(resmalgas)]
plot.ts(resmalgas)
acf(resmalgas)
```

Notice the fourth command, the filter command will register a "NA" when the window will not fit inside the data. Using the fifth command above deletes the "NA" entries. Do the residuals appear more reasonable?

(f) Another method for smoothing the data is kernel smoothing. Use the following commands to perform kernel smoothing and to analyze the residuals.

```
ksmlgas<-ksmooth(1:length(lgas),lgas,"normal",bandwidth=12)$y
par(mfrow=c(3,1))
plot(lgas)
lines(ksmlgas, col="red")
resksmlgas<-ksmlgas-lgas
plot.ts(resksmlgas)
acf(resksmlgas)
```

Adjust the bandwidth in the first command above to see how the curve changes. How does the ACF of the residuals change when using a bandwidth of 3 versus a bandwidth of 12? How does the fitted line change? Hint: to be able to compare, you may want to set `par(mfrow=c(2,3))` and display the plot of the smoothed data, residuals, and the ACF of the residuals for both bandwidths.

2 Assignment

1. (No R required). Suppose that $x_t = w_t + kw_{t-1} + kw_{t-2} + kw_{t-3} + \cdots + kw_0$, for $t > 0$, k constant, and w_i iid $\mathrm{N}(0, \sigma_w^2)$.

   (a) Derive the mean and autocovariance function for $\{x_t\}$. Is $\{x_t\}$ stationary?

   (b) Derive the mean and autocovariance function for $\{\nabla x_t\}$. Is $\{\nabla x_t\}$ stationary?

2. (No R required). Show that ∇^k eliminates a kth order polynomial. **Hint**: try mathematical induction.

3. Load the monthly temperature data for England from 1723 to 1970 (in Celsius) using the following command

```
engtemp<-scan("tpmon.dat",skip=1)
```

   (a) Plot the data and create an ACF. Comment on seasonality. Try doing this only on the first 300 observations. The command `engtemp2<-engtemp[1:300]` will give you the first 300 observations of `engtemp`.

(b) Fit the following model using `lm()`:

$$x_t = \beta_1 + \beta_2 cos\left(2\pi\frac{1}{12}t\right) + \beta_3 sin\left(2\pi\frac{1}{12}t\right) + w_t$$

(Hint: You will need to "create" the variables for the covariates. It may be useful to know that there are R functions `sin()` and `cos()`. For the value of t, use a vector that represents the index for time, e.g.: `(1:length(engtemp)`. So, if I am creating a variable for $cos(\pi t)$, I will type `cos(pi*(1:length(engtemp)))`.) Be sure to use all observations for model fitting. What are the OLS estimates for $\beta_1, \beta_2, \beta_3$?

(c) Plot the residuals of the above fit, and produce the ACF plot. Are the residuals stationary? Are they dependent?

(d) Based on the residuals, comment on whether we have successfully detrended this time series.

4. Use the smoothing techniques introduced in class and above to comment on the trend in the global temperature data. The data can be found in `globtemp.dat`). Try at least three different degrees of smoothing (e.g.: under-smooth, just right, and over-smooth). Based on your plots, what are the effects of under-smoothing and over-smoothing?