Series 6

1. In this exercise we shall apply a Kalman filter to smooth a given time series. From the lecture we know that we can write

$$Y_t = Z_t + E_t$$

with a smooth Z_t by taking Z_t to be a special AR(2) process.

a) Load and plot the time series y:

```
> load(url("ftp://stat.ethz.ch/Teaching/Datasets/WBL/ts-serie9.Rdata"))
```

The time series is now available as the R object y.

b) In the R package sspir, there are functions for defining state space models and the Kalman filter. Specify the model as follows:

```
> library(sspir)
> ss.model <- SS(y = y,
+                x = NA,
+                Fmat = function(tt,x,phi) {
+                        return(matrix(c(1,0))) },
+                Gmat = function(tt,x,phi) {
+                        return(matrix(c(2,1,-1,0), ncol=2)) },
+                Vmat = function(tt,x,phi) {
+                        return(matrix(phi[1])) },
+                Wmat = function(tt,x,phi) {
+                        return(matrix(c(phi[2],0,0,0), ncol=2))},
+                m0 = (matrix(c(0,0),nrow=1)),
+                C0 = (diag(c(10,10))),
+                phi = c(1,0.01)
+                )
```

Use the output of `print(ssm.ar2gl)` to familiarize yourself with the notation used in this package. At the top of the output, you will find the state space model. The authors of the package denote the state vector by θ.

What is the meaning of the values of the unknown parameter vector phi?

Hint: The notation used differs from that utilized in the lecture notes; it also varies between different books and software packages.

c) Use the ML estimates $\phi = (0.9, 0.0001)$ (cf. part f)) to fit the Kalman filter, and smooth the output using the `smoother()` function. Plot the output of the Kalman filter and the smoother.

```
> phi(ss.model) <- c(0.9,0.0001)

> model.filter <- kfilter(ss.model)
> model.smooth <- smoother(model.filter)
> plot(model.filter$y)
> lines(model.filter$m[, 1], lty = 1, col = 2)
> lines(model.smooth$m[, 1], lty = 1, col = 3, lwd = 2)
```

d) Now examine the influence of the error variance in the AR(2) process Z_t (2nd coefficient in phi) by first reducing it to zero and then raising it to 0.06. Plot the filter and smoothing results each time. To do so, adapt the sample code from part c).

e) Plug in the ML estimates once more and vary the initial values m0 and C0. Be unrealistic and plug in eg. large and small values for m0 and C0.

```
> phi(ss.model) <- c(0.9, 0.0001)
> m0(ss.model) <- matrix(..., ncol = 2)
> C0(ss.model) <- diag(...)
```

f) Estimate both parameters using maximum likelihood (ML). To this end, write your own function and note that variances cannot be negative.

Think about what happens in the following R code. If any of the functions is unfamiliar to you, refer to its help file.

```
> f.optim <- function(log.phi)
+ {
+    phi(ss.model) <- exp(log.phi)
+    model.filter <- kfilter(ss.model)
+    return(c(-model.filter$likelihood))
+ }
> log.phi.min <- nlm(f.optim, log(c(1, 0.00001)))
```

2. In this exercise, we continue the investigation of the connection between advertising expenditure and sales started in Series 5, Exercise 1. We again examine the dataset `advert.dat` which contains the annual advertising expenditure (`ADVERT`) and the sales revenue (`SALES`) for a particular brand of vegetable stock (Lydia Pinkham's Vegetable Compound, 1907-1960).

Load and log-transform the data as follows:

```
> d.advert <- read.table("ftp://stat.ethz.ch/Teaching/Datasets/WBL/advert.dat",
+    header = TRUE)
> ts.advert <- ts(log(d.advert[, "ADVERT"]), start = 1907)
> ts.sales <- ts(log(d.advert[, "SALES"]), start = 1907)
```

In the following problems, we will refer to `ts.advert` and `ts.sales` as $X_{1,t}$ and $X_{2,t}$, respectively.

a) First look at the time series plots. Are the series stationary?
R hint: you can either use the generic plot function (first line), or a specific time series plot function (second line):

```
> plot(ts.union(ts.sales, ts.advert), plot.type = "single")
> ts.plot(ts.sales, ts.advert)
```

(Unless the time series have been joined using `ts.union()`, the command `plot(d.sales, d.advert)` will plot both of them against each other. It is easy to see that they are correlated.)

b) Compute the first-order differences $Y_{1,t} := X_{1,t} - X_{1,t-1}$ (`ts.adv.d1`) and $Y_{2,t} := X_{2,t} - X_{2,t-1}$ (`ts.sal.d1`). Are these processes stationary?

c) Regard the transfer function model

$$Y_{2,t} = \sum_{j=0}^{\infty} \nu_j Y_{1,t-j} + E_t \,.$$

Describe this model in words. From a business point of view, what is an illustrative interpretation of this model?

d) Estimate the coefficients ν_j. Proceed as described below; see also the lecture slides. Make an interpretation of the model you estimate. What do you notice about this model?
Instructions:

- First identify the row $Y_{1,t}$. Which model should you fit? Estimate its coefficients using the R function `arima()`.

- Compute D_t and Z_t for the transformed series.
  R hints: Suppose your R object with the ARMA model you used to fit $Y_{1,t}$ is called `r.fit.adv`. D_t then simply contains the residuals:

  ```
  > ts.D <- resid(r.fit.adv)
  ```

  To calculate Z_t, you can extract the AR (or possibly ARMA) coefficients of the fitted model and then use the R function `filter()`. For example, if `alpha` is a vector of AR coefficients, use

```
> ts.Z <- filter(ts.sal.d1, c(1, -alpha), sides = 1)
```
The argument `sides = 1` causes `filter()` to only take the past into account; cf. the help page.

Since you should only consider time points present in both time series, use

```
> ts.trans <- ts.intersect(ts.Z, ts.D)
```

- Draw the cross-correlogram for D_t and Z_t. Give an interpretation of them.
  R hint:

  ```
  > acf(ts.trans, na.action = na.pass)
  ```

- Estimate the coefficients ν_j using the forumla given in the lecture.
  R hints:
  The variance of the errors D_t can be found in `r.fit.adv$sigma2`. The command

  ```
  > acf(..., type = "covariance", na.action = na.pass)$acf[, 1, 2]
  ```

  also returns the cross-covariances between Y_t and Z_t.

Preliminary discussion: Monday, May 09.

Deadline: Monday, May 16.