Series 12

1. In this exercise we are going to study state space models and kalman filtering for an MA(1) process which is superposed with some additional noise.

a) First simulate the "obsereved" time series Y_t of length 100, which is the MA(1) process $Z_t = E_t + 0.4E_{t-1}$ superposed with some additional noise V_t (i.e. $Y_t = Z_t + V_t$). Assume that the innovations E_t of the MA(1) process are i.i.d. $\mathcal{N}(0, 0.2)$ distributed and the additional noise V_t i.i.d. $\mathcal{N}(0, 0.3)$.

b) Make a plot of the original MA(1) process Z_t and draw the actually "observed" additional noise process Y_t on top of it.

c) Find a state space formulation of this model, i.e. find matrices/vectors G_t, W_t (resp. w_t^2), F_t, v_t^2 such that

$$X_t = G_t X_{t-1} + W_t \quad \text{with } W_t \sim \mathcal{N}\left(0, w_t^2\right)$$
$$Y_t = F_t X_t + V_t \quad \text{with } V_t \sim \mathcal{N}\left(0, v_t^2\right)$$

Hint: The initial state X_t is

$$X_t = \begin{pmatrix} Z_t \\ E_t \end{pmatrix}$$

d) Estimate the model with the Kalman-Filtering in R.
R-Hint:

```
> ## load the package for kalman filtering
> library(sspir)
> ## state space formulation
> ssf <- SS(y=as.matrix(y),
            Fmat=function(tt,x,phi) {return(matrix(...))},
            Gmat=function(tt,x,phi) {return(matrix(...))},
            Vmat=function(tt,x,phi) {return(matrix(...))},
            Wmat=function(tt,x,phi) {return(matrix(....))},
            m0=matrix(...),
            C0=matrix(...))
> fit <- kfilter(ssf)
> plot(...)
> lines(y, col="red")
> lines(fit$m[,1], col="blue")
```

For the `SS`-Function the follwing matrices from the state space model have to be given:

- `Fmat`: $F_t^\top$ (the transpose of the matrix F_t)
- `Gmat`: the matrix G_t
- `Vmat`: v_t^2 (the covariance matrix of V_t)
- `Wmat`: w_t^2 (the covariance matrix of W_t)
- `m0`: the initial state (i.e. $\mathbf{E}[X_t] = \mathbf{E}\left[\begin{pmatrix} Z_t, & E_t \end{pmatrix}\right]$)
- `C0`: The covariance matrix of the initial state (i.e. $\text{Cov}(X_t)$)

2. In this exercise we will fit a dynamic linear model. We here consider a regression problem where time-varying coefficients may be necessary. The description of the practical situation is as follows: In April 1979 the Albuquerque Police Department began a special enforcement program aimed at countering driving while intoxicated accidents. The program was composed of a squad of police officers, breath alcohol testing (BAT) devices, and a van named batmobile, which housed a BAT device and was used as a mobile station. The data were collected by the Division of Governmental Research of the University of New Mexico under a contract with the National Highway Traffic Safety

Administration of the Department of Transportation to evaluate the batmobile program. (Source: *http://lib.stat.cmu.edu/DASL/Datafiles/batdat.html*)

The data comprise of quarterly figures of traffic accidents and the fuel consumption in the Albuquerque area as a proxy of the driven mileage. The first 29 quarters are the control period, and observations 30 to 52 were recorded during the experimental (batmobile) period. We would naturally assume a time series regression model for the number of accidents:

$$\texttt{ACC}_t = \mu + \alpha \cdot \texttt{FUEL}_t + \beta_{season,t} + E_t$$

The accidents depend on the mileage driven and there is a seasonal effect. In the above model the intercept is assumed to be constant. In the light of the BAT program, we might well replace it by $\mu_t = \mu_{t-1} + \triangle\mu_t$, with $\text{Var}(\triangle\mu_t) = 10$, i.e. some general level of accidents that is time-dependent.

a) Read in the data and plot it.

```
> batmobile <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/batmobile.dat",
                          header=T, quote="\"")
> dat       <- ts(batmobile, start=1, freq=4)
```

b) Before looking at the dynamic linear model with the time-dependent μ_t state the "normal" time series regression model for the number of accidents (without time-varying coefficients).

c) Fit your model in R.
R-Hint: Prepare the data for the regression:

```
> regdat <- cbind(batmobile, season=factor(cycle(dat), lab=c("Q1","Q2","Q3","Q4")))
```

d) Check the residuals of your model. Fit also loess-smoother on the residuals.
R-Help:

```
> times      <- 1:52
> fit.loess <- loess(resid(fit) ~ times, span=0.65, degree=2)
```

e) Now, correct `ACC` for the effect of `FUEL` estimated with your simple linear model and plot the corrected time series (with a loess smoother).
R-Help:

```
> newacc <- ts(batmobile$ACC-coef(...)[2]*batmobile$FUEL, ...)
```

f) Do an STL-analysis of the corrected time series, plot it and look at the acf of the remainder.

g) Perform Exponential Smoothing for the new time series and plot the fitted values.

h) The effect of fuel is removed of the corrected data `newacc`. But still there is a seasonal effect. This we now want to model with a dynamic linear model, where the intercept is time-dependent:

$$\texttt{ACC}_t = \mu_t + \beta_{season,t} + E_t$$

with $\mu_t = \mu_{t-1} + \triangle\mu_t$, with $\text{Var}(\triangle\mu_t) = 10$.
State the state space formulation of this regression model.

i) Fit the state space model (with kalman-filtering) for the (corrected) time series.
R-Hint: Fit a regression model $ACC \sim season$ to get estimates for the unknowm parameters and variances you need in the different matrices of the state space model.

j) Fit a state space model (with kalman-filtering) for the original data.
Hint: Proceed similar as in i).

Exercise hour: Monday, May 19.