Series 4

1. Since simulations can be of use in model validation, we would like to use this exercise to simulate several time series by means of an ARMA model.

a) AR(2) model with coefficients $\alpha_1 = 0.9$ and $\alpha_2 = -0.5$.

b) MA(3) model with coefficients $\beta_1 = 0.8$, $\beta_2 = -0.5$ and $\beta_3 = -0.4$.

c) ARMA(1,2) model with coefficients $\alpha_1 = -0.75$, $\beta_1 = -1$ and $\beta_2 = 0.25$.

The error E_t shall follow a standard normal distribution $N(0, 1)$ in every model. For each of the three models, do the following:

(I) First think how the autocorrelations should behave based on the theory.

(II) Use the procedure `ARMAacf()` to compute the *theoretical* autocorrelations and plot them.

R Hints:

```
## Theoretical autocorrelations
> plot(0:30, ARMAacf(ar=c(0.9,-0.5), lag.max=30), type="h",
       ylab="ACF")
## Theoretical partial autocorrelations
> plot(1:30, ARMAacf(ar=c(0.9,-0.5), lag.max=30, pacf=T), type="h",
       ylab="PACF")

> plot(0:30, ARMAacf(ma=c(..., ..., ...), ...), ...)
> ...
```

(III) Now simulate all three models a) - c). Take several different lengths for the time series: $n =$ 200, 500 and 1000. Repeat these simulations several times to develop some intuition on what is “chance” and what is “structure”.

R Hints:
You can use the procedure `arima.sim()` to simulate the time series. The length of the simulated series you can choose by setting the argument `n`, and the model by setting the parameter `model` (to a list!).

```
> r.sim1 <- arima.sim(n=..., model=list(ar=c(0.9,-0.5)))
```

(IV) Take a look at the simulated time series, the acf and the pacf plots.

2. In this exercise, the goal is to choose a suitable model for given data. The time series is available at `http://stat.ethz.ch/Teaching/Datasets/WBL/mcARMA.dat` and can be read in with

```
> dat <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/mcARMA.dat",
                    header = TRUE)
> ts <- ts(dat$x)
```

a) Plot the time series. Does it look stationary?

b) By looking at the ACF and PACF, try fitting different plausible models for this time series using the function `arima()`. You can choose among AR(p), MA(q) and ARMA(p, q) and decide on your favourite by looking at the Residuals, the AIC and the significance of the estimated coefficients (which can be found by constructing confidence intervals based on the `arima()`-output).
R-Hint:
The package `forecast` provides the funtion `tsdisplay()`, which (using the argument `points = FALSE`) provides a nice display of the time series plot, the ACF and the PACF of a time series.

3. There is a study on the development of beluga whales that focusses on the nursing behaviour of mother and calf. During a total of 160 time periods (each lasting 6 hours) subsequent to birth, the following variables were observed for "Hudson", a beluga calf. Zoologists use this data to ascertain the health of this young whale.

`PERIOD`	Index of the time period
`BOUTS`	Square root of the number of nursing bouts
`LOCKONS`	Square root of the number of lock-ons ("docking attempts")
`DAYNIGHT`	Day (= 1, 8am – 8pm) or night (= 0, 8pm – 8am) indicator
`NURSING`	Square root of the number of seconds spent successfully nursing during the period.

A nursing bout is defined as a successful nursing episode where milk was obtained. We would like to model the nursing time by means of the other variables. Count variables have already undergone a square root transformation to stabilize their variance ("first-aid-transformation"). You will find the data in the file `beluga.dat`.

Load the data in the usual way and create a time series matrix:

```
> d.beluga <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/beluga.dat",
header=T)
> d.beluga <- ts(d.beluga)
```

a) Fit the model

$$\texttt{NURSING} = \beta_0 + \beta_1\,\texttt{PERIOD} + \beta_2\,\texttt{BOUTS} + \beta_3\,\texttt{LOCKONS} + \beta_4\,\texttt{DAYNIGHT}$$

using ordinary linear regression. Check the independence of the residuals. What conclusions can zoologists draw from this analysis?

b) Due to the correlations involved, an AR model should be assumed for the residuals. Determine the order p of this model, and estimate the parameters $\alpha_1, \ldots, \alpha_p$.

c) What transformation should you apply to obtain a linear model with independent errors? State it as a formula.

d) (*) How would you perform this transformation (or these transformations) in R? Use the transformed time series to carry out another regression, and look at the correlation structure for the errors!

R-hint: `lag()` Note: depending on your implementation, you might have to rename your variables to something meaningful to fit the transformed dataset.

e) Estimate the regression coefficients and the AR parameters using Maximum Likelihood.

R-hints:

```
> library(nlme)                              # Load the package containing the procedure gls()
> r.bel.gls <- gls(NURSING ~ BOUTS + LOCKONS + DAYNIGHT + PERIOD,
+ data = d.beluga, correlation = corARMA(form= ~ PERIOD,
+ p = r.burg$order, q = 0, fixed = FALSE), method = "ML")
> summary(r.bel.gls)
> d.resid <- ts(resid(r.bel.gls))
> plot(d.resid)
> acf(d.resid)
> pacf(d.resid)
```

To ensure convergence of the algorithm, known estimates of the AR parameters can be passed to `corARMA()` as starting values using the optional argument `values`. In this particular case, this does not change the outcome.(`correlation = corARMA(..., value = r.burg$ar, ...)`)

f) Simplify the model where possible (by leaving out variables from the linear regression).

Preliminary discussion: Monday, April 03.