# RTP Exercise Sheet

## Series 4

### Exercise 4.1

Similar to exercise 3.1, we start with some simulations. Thus, we would like to use this exercise to simulate several time series by means of an ARMA model. Please perform the same steps as in exercise 3.1 for the following models:

a) ARMA(1,2) model with coefficients $\alpha_1 = 0.75$, $\beta_1 = -0.3$ and $\beta_2 = 0.25$.

b) ARMA(2,1) model with coefficients $\alpha_1 = 0.75$, $\alpha_2 = -0.3$ and $\beta_1 = 0.25$.

c) Why is it not possible to simulate the ARIMA(2,1,2) model with coefficients $\alpha_1 = 0.5, \alpha_2 = 0.5$, $\beta_1 = -0.4$ and $\beta_2 = 0.3$ with $d = 1$ with arima.sim?

d) What is the equivalent ARMA model to the ARIMA(2,1,2) model in task c)? Hint: Determine the equivalent ARMA process.

The innovation $E_t$ shall follow a standard normal distribution $\mathcal{N}(0; 1)$ in every model.

### Exercise 4.2

In this exercise, we look at the time series *sunspotarea*, which is available in the package *fpp*. It contains yearly data about the area of sunspots averaged over all days of the year (in units of millionths of a hemisphere). Sunspots are magnetic regions that appear as dark spots on the surface of the sun.

a) Plot the time series. Why does it make sense to log-transform the time series?

b) Choose a suitable AR-model only based on the first 100 observations (1875 - 1974) of the log-transformed series.

   R-Hint: window()

### Exercise 4.3

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. A short description of the data is given in the following table.

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

| | |
|---|---|
| PERIOD | Index of 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 |

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 = TRUE)
d.beluga <- ts(d.beluga)
```

a) Fit the model

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

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(p) model should be assumed for the residuals. Determine the order p of this model, and estimate the parameters $\alpha_1,...,\alpha_p$. R-Hint:

```
r.burg <- ar(...)
```

c) Estimate the regression coefficients and the AR parameters using Generalized Least Squares with Maximum Likelihood estimation. R-Hint:

```
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))
```

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, ...))

d) Optional: Simplify the model if possible.

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

Hint: Cochrane-Orcutt Method.

f) Optional: 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().

## Exercise 4.4

During their yearly spring melt, glaciers deposit layers of sand and mud. These annual sediments, known as varves, can be reconstructed in New England for the whole time between the beginning (about 12'600 years ago) till the end (6'000 years ago) of glacial retreat. From these varves, approximations of paleoclimatic parameters can be computed, such as temperature (a warmer year yields more sediment).

In the dataset *varve.dat*, you will find 350 annual sediment diameters (contained in lines 201 through 550) starting at 11'660 years ago. After loading these data, first construct a time series object from them:

```r
t.url <- "http://stat.ethz.ch/Teaching/Datasets/WBL/varve.dat"
d.varve <- ts(scan(t.url)[201:550])
```

Comment: The procedure scan() is a more general data loading function than read.table(). We use it here to avoid putting the data into a data frame. Do not worry about the exact choice of procedure for reading data here: simply believe us when we say that scan() does what we need, or read the help file.

a) It is advisable to log-transform the time series. Why?

b) Is the log-transformed time series stationary? If not, how can you make this time series stationary?

c) Choose a suitable model that fits the data. Does your model fit? Analyze the residuals and comment on your decision.

   R-Hint:

```r
r.varve <- arima(log(d.varve), order = c(..., ..., ...))
```

d) Write down the model you chose in c) with its estimated coefficients.

*Disclaimer: Parts of the exercises are adopted from 'Applied Time Series Analysis' course at ETHZ by Marcel Dettling.*