Series 2

1. Reconsider the dataset `hstart` (Series 1: Problem 4) and recall that this dataset contains monthly data on the start of residential construction in the USA within the time frame of January 1966 to January 1974. The data have undergone some transformation unknown to us (perhaps an index over some baseline value has been calculated, or perhaps the data are to be read as $x \cdot 10^?$ construction permits). In our opinion, this makes these data a good didactic choice for illustrating the theory.

(Source: U. S. Bureau of the Census, Construction Reports.)

Importing the data (without `header=T`!) and preparing them:

```
> hstart <- read.table("http://stat.ethz.ch/Teaching/Datasets/WBL /hstart.dat)
> hstart <- ts(hstart[,1], start=1966, frequency=12)
```

a) Parametric Models

Decompose the time series into its components using a parametric model. Plot the time series, including fitted values, and comment on any differences. Choose the order of the polynomial according to how good it fits the real data. Compare the orders 3, 4 and 6.

R hint:

```
> Time <- 1:length(hstart)
> Months <- factor(rep(month.name, length(hstart)/12), levels=month.name)
> H.lm  <- lm(hstart ~ Months + Time + I(Time^2) + I(Time^3) + ...)
> H.fit <- ts(fitted(H.lm), start=1966, freq=12)
> lines(H.fit, lty=3, col=2)                    # H.fit is added to the plot from part a)
```

If you want to compare the different polynomials you can either plot them and compare how good they fit the real data or plot the residuals, `hstart-H.fit` versus Time and check if there is still a structure. Moreover, compare your favorite parametric model with the STL-Decomposition (Series 1: Problem 4.c) using the residuals vs. time plots.

b) The special filter

$$Y_t = \frac{1}{24}\Big(X_{t-6} + 2X_{t-5} + \ldots + 2X_t + \ldots + 2X_{t+5} + X_{t+6}\Big)$$

can be used for computing a trend estimate. Plot this, the STL trend and the data in a single plot. What are the differences between this Fit and the previous ones? What is better, what is worse?

R hint:

```
> plot(hstart,lty=3)
> H.filt <- filter(hstart, c(1,rep(2,11),1)/24 )
> trend <- H.stl$time.series[,2]
> lines(trend, col=3)
> lines(H.filt, lty=2, col=2)
```

Labelling:

```
> legend(1966, 235, legend=c("Time series","Filter","STL"),
+                                                  col=c(1,2,3), lty=c(3,2,1))
```

2. The performance of a machine that analyses the creatine content of human muscular tissue is to be investigated using 157 known samples. These samples are given to the machine one at a time for it to determine their creatin content.

The data are from an investigation into the correct functioning of automated analysis machines. You can find them in the dataset

`http://stat.ethz.ch/Teaching/Datasets/WBL/kreatin.dat` .

In this exercise, we will focus on one of the variables in this dataset, namely `gehalt` (content).

a) Which stochastic model should this series of data follow if the machine is working correctly?

b) Use the time series plot, the autocorrelations and the partial autocorrelations to determine whether or not these data fit the ideal model found in Part a).

R hints:

Converting the data frame (`d.creatine`) to a time series:

```
> t.creatine <- ts(d.creatine[, 2], start = 1, frequency = 1)
```

Plotting ACF and PACF:

```
> acf(t.creatine, plot = TRUE)
> acf(t.creatine, type = "partial", plot = TRUE)
```

3. In this exercise, we consider the `AirPassengers` dataset, a time series indicating the monthly numbers of international airline passengers departing from the USA in the years 1949 to 1960. We use different methods to decompose the time series into trend, seasonal effect and remainder (cf. Series 1, Exercise 4) and compare the remainders by looking at their correlogram.

a) The `AirPassengers` dataset is a sample time series provided by R ; you don't have to read it in. Look at a plot of the time series:

```
> plot(AirPassengers)
```

Why is the correlogram of this time series not meaningful? Explain in a few sentences.

b) Decompose the time series into trend, seasonal component and remainder using the R function `decompose()`; plot the remainder and its correlogram. Interpret the plots in few sentences.

R hints:

```
> airpass.decomp <- decompose(AirPassengers, type = "multiplicative")
> plot(...)
> acf(..., na.action = na.pass, plot = TRUE)
```

See Series 1, Exercise 4.c for hints on extracting the remainder from the object `airpass.decomp`, or use the R -help: `?decompose`. The function uses a filter to estimate the trend; therefore, the first and the last few entries of the decomposition are not defined (value `NA` in R). For this reason we have to use the parameter `na.action = na.pass`, otherwise R complains about missing values.

c) Decompose the log-transformed time series using the R function `stl()`. Estimate the seasonal effect once by averaging (parameter `s.window = "periodic"`) and once by choosing an appropriate smoothing window (parameter `s.window = ...`, where you have to choose an odd integer; cf. Series 1, Exercise 4.c). To determine an appropriate smoothing window, you can look at the `monthplot()` of the seasonal component.

For both estimation approaches (averaging and smoothing window), plot the remainder and its correlogram, and comment on the plots.

R hints:

```
> airpass.stl <- stl(log(AirPassengers), s.window = ...)
> plot(...)
> acf(..., plot = TRUE)
```

d) Explain why you used the parameter `type = "multiplicative"` in Task b), and why you log-transformed the time series before performing an `stl()` decomposition in Task c).

e) Use the differencing approach. Choose $lag = 12$ in order to get rid of the trend and period. Plot the new timeseries and acf. Compare to the previous methods.