Solution to Series 2

1. a)
```
> hstart <-read.table("http://stat.ethz.ch/Teaching/Datasets/WBL/hstart.dat")
> hstart <- ts(hstart[,1], start=1966, frequency=12)
> plot(hstart)
```
The time series under investigation is non-stationary (see the plot for Part c). Its non-stationary properties consist of a trend and deterministic seasonal fluctuations (by month).
We can decompose this time series into its trend, its seasonal fluctuations (month effects) and the remainder:
$$X_t = m_t + \alpha_{i\langle t\rangle} + E_t$$

If the trend can be described by a polynomial of degree k, we then have
$$m_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \ldots + \beta_k \cdot t^k$$
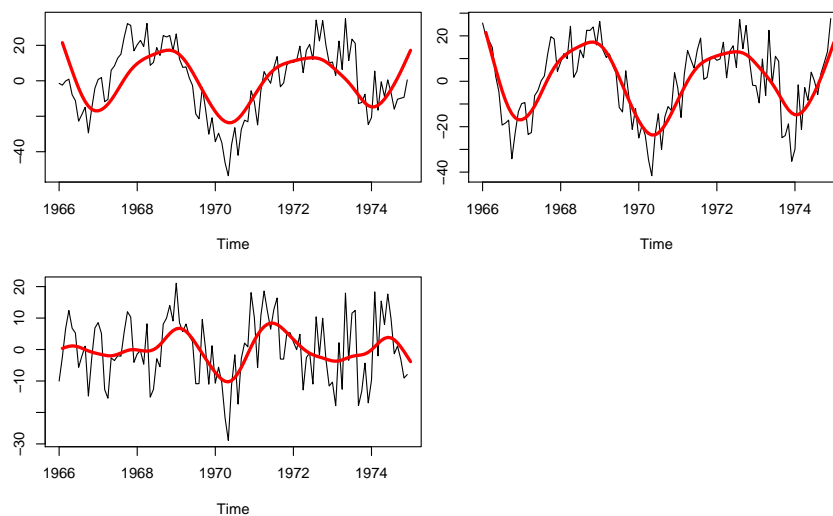
b) Parametric model:
$$X_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \ldots + \beta_6 \cdot t^6 + \alpha_{i\langle t\rangle} + E_t,$$

$$\text{where } t = 1, \ldots, 108 = 12 \cdot 9, \qquad i\langle t\rangle \in \{1, \ldots, 12\}.$$

We model the trend using a polynomial of order 6 and the monthly effects using a factor (see the plot below). The local maximum of the parametric model in the years 1970-1 is not fit well by the parametric model. In some other cases this lack of fit at maxima and minima also appears. However, the lower-order polynomials are even worse at fitting the trend.
```
> Time <- 1:length(hstart)
> Time2 <- Time/12+1966
> Months <- factor(rep(month.name, length(hstart)/12), levels=month.name)
> H.lm6  <- lm(hstart ~ Months + Time + I(Time^2) + I(Time^3)+
   I(Time^4) + I(Time^5) + I(Time^6))
> H.lm3  <- lm(hstart ~ Months + Time + I(Time^2) + I(Time^3))
> H.lm4  <- lm(hstart ~ Months + Time + I(Time^2) + I(Time^3)+I(Time^4))
> H.fit6 <- ts(fitted(H.lm6), start=1966, freq=12)
> H.fit3 <- ts(fitted(H.lm3), start=1966, freq=12)
> H.fit4 <- ts(fitted(H.lm4), start=1966, freq=12)
```


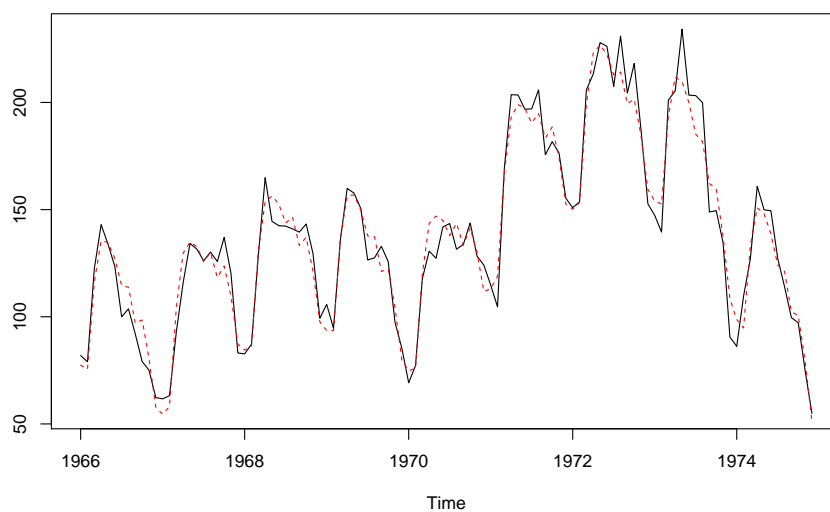**c)** The non-parametric model STL fits the data better than the parametric one does.

```
> H.stl <- stl(hstart, s.window="periodic")
> remainder <- H.stl$time.series[,3]
> plot(hstart)
> H.np <- hstart - remainder
> lines(H.np, col=2, lty=2)
```
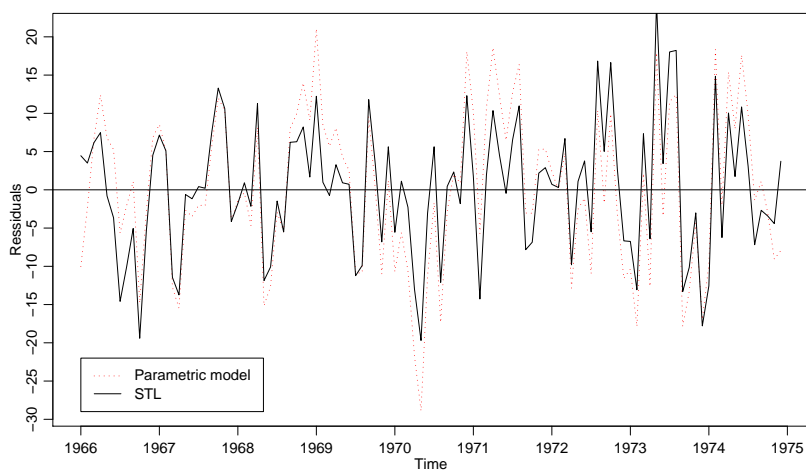

In the residual time series plot shown below, the discrepancies between the two models are quite visible. Over certain intervals of time, the residuals of the parametric model differ from zero in a systematic way.


R commands for creating the residual plots:

```
> plot(ts(resid(H.lm6), start=1966, freq=12), lty=3, col=2, ylab="Residuals")
> lines(remainder)
> abline(h=0)
> legend(1966, -22, legend=c("Parametric model","STL"), col=c(2,1), lty=c(3,1))
```

d) In the following plot, we have the original data, additive seasonal effects, the trend curve and the residuals (remainder) of the STL decomposition, drawn in order from top to bottom.
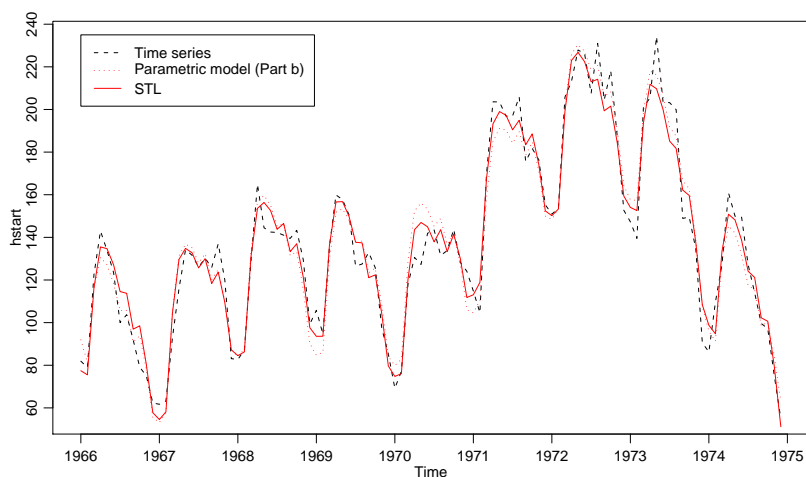
Furthermore, in the "monthplot" of `seasonal.var` (next figure, right side) we see how the monthly effects of May to August increase with the years, while the effects of January, September and December decrease.

So it seems reasonable to take seasonal effects that vary instead of fixed ones. When we remember the data we are dealing with, I think it is plausible that seasonal effects of starts of residential construction should vary over time.
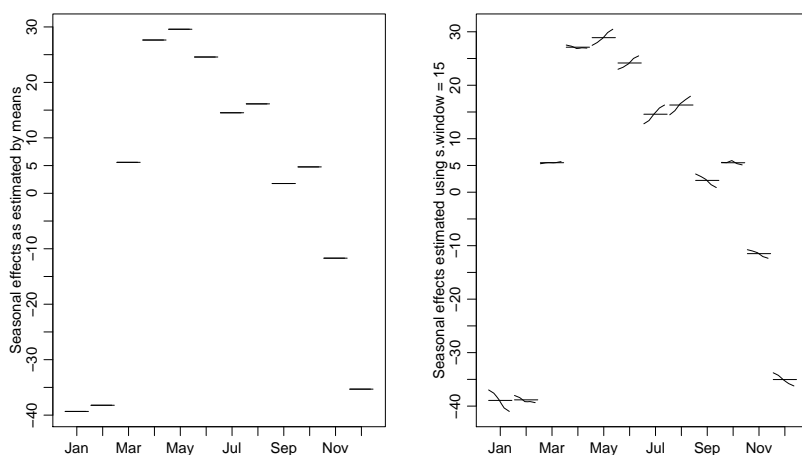

e) The trend line from the special filter is somewhat less smooth than the trend line stemming from the STL decomposition. However, the smoothness of the STL trend line can be steered by the smoothness parameter `t.window` of the procedure `stl`, e.g. setting this parameter to 25 (cf. R hint for Part c).
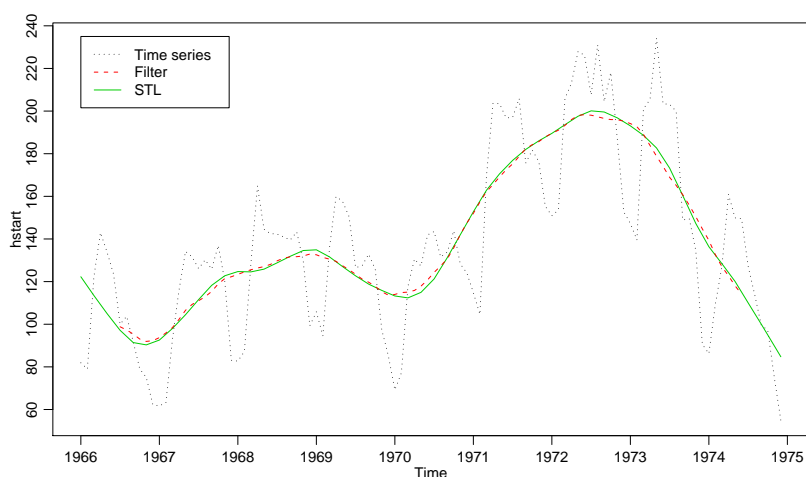
f) First we remove the seasonal fluctuations by computing $Y_t = X_t - X_{t-12}$. In the next plot, we then see that a trend nonetheless remains.

Taking first-order differences of Y_t, i.e.

$$Z_t = Y_t - Y_{t-1} = (X_t - X_{t-12}) - (X_{t-1} - X_{t-13}) \,,$$

finally gives us a stationary series.
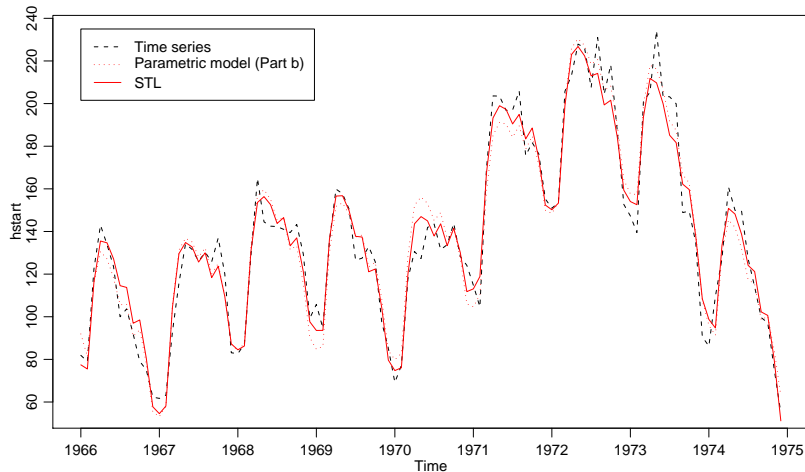
2. **a)** In the plot below (Part c), the seasonally adjusted time series is shown by a dotted line.

b) See the R hints on the problem sheet.

c) At any given time, both filters use an arithmetic mean of the same observations (± 4 time steps). Their difference lies in the weights assigned to these observations: filter 3 uses smaller weights for more distant observations – unlike filter 4, which weights them all equally. Thus filter 3 is better able to track local minima and maxima.
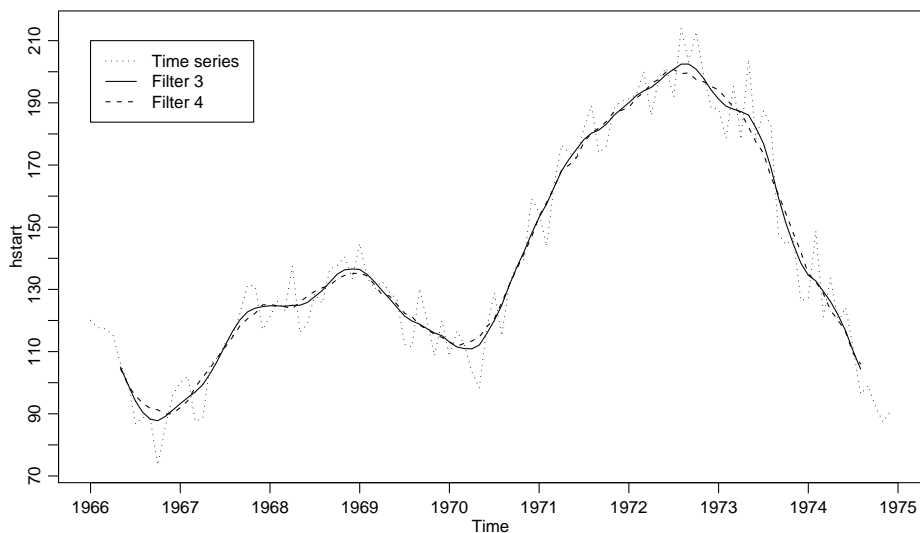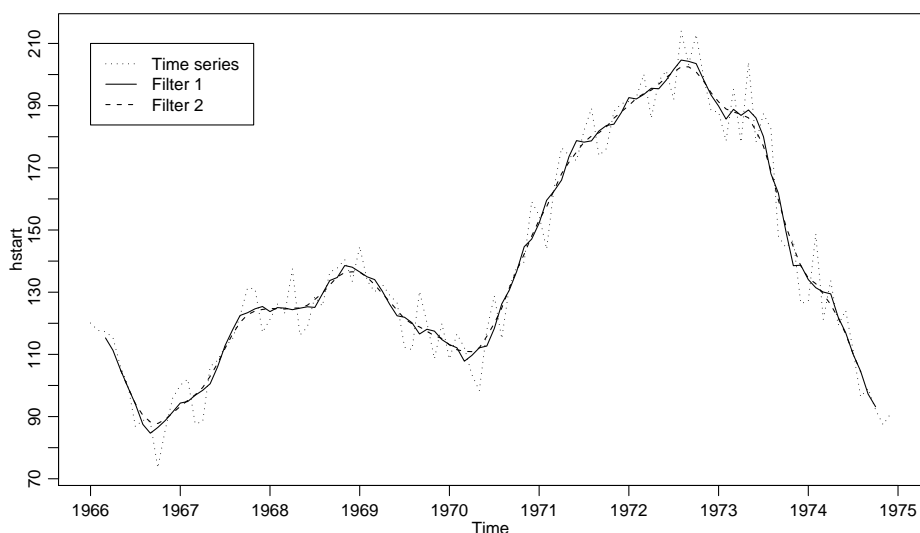
**d)** Filter 2 and 3 are in fact identical. This is even more evident in the following plot that shows only them: here the solid and dashed lines are the same.
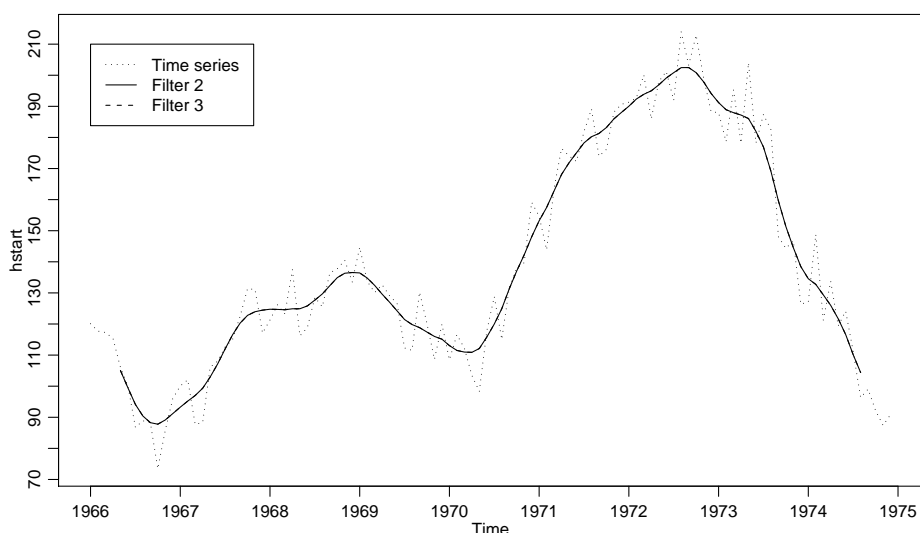
A filter with a wide window and weights decreasing away from the centre (such as filter 3) can thus be written as a multiple application of a filter with constant weights (such as filter 2). This is a commonly used method of avoiding the use of weights, instead only requiring a choice of window sizes for the filters used. You will encounter this technique later when smoothing the periodogram (spectral analysis).

3. The computations of moments and sketches of proofs given here are mainly aimed at those students with a particular mathematical interest. Subsequent theory will partially revisit this topic. Often the stationarity (or lack thereof) of a time series is evident from the standard time series plot. The following plot helps us guess the correct answers for the series we have been given.
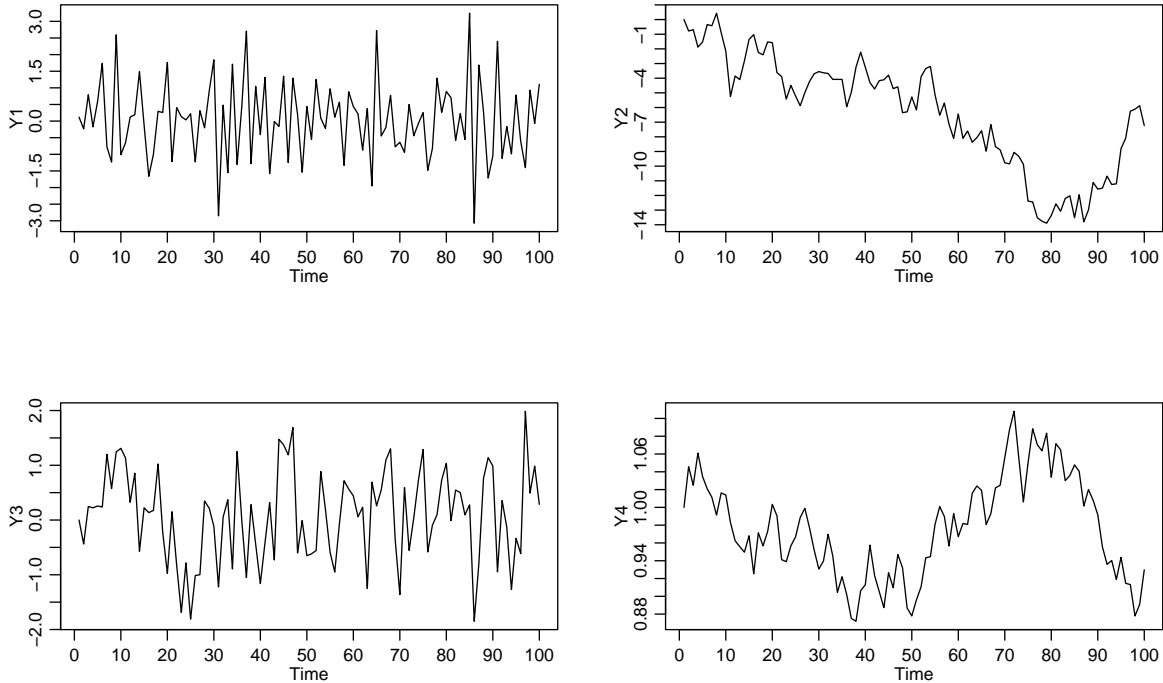
a) $Y_t = E_t - 0.5 \cdot E_{t-1}$ is a stationary time series (cf. the plot shown); to be precise, an MA(1) process (Time Series II). We have

$$E[Y_t] \quad = \quad E[E_t] - 0.5 \cdot E[E_{t-1}] = 0 \ (= \text{ const.})$$

$$\text{Cov}(Y_t, Y_{t+h}) \quad = \quad \begin{cases} -0.5 \cdot \text{Var}(E_{t-1}) = -0.5 & h = -1 \\ \text{Var}(E_t) + 0.5^2 \cdot \text{Var}(E_{t-1}) = 1.25 & h = 0 \\ -0.5 \cdot \text{Var}(E_t) = -0.5 & h = 1 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Thus $\text{Cov}(Y_t, Y_{t+h})$ is independent of t. For those who are interested, here is a quick sketch of the proof of Equation (1):

$$\begin{aligned} \text{Cov}(Y_t, \ Y_{t-1}) \quad &= \quad \text{Cov}(E_t - 0.5 \cdot E_{t-1}, \ E_{t-1} - 0.5 \cdot E_{t-2}) \\ &= \quad \text{Cov}(E_t, \ E_{t-1}) - 0.5 \cdot \text{Cov}(E_t, \ E_{t-2}) \\ &\quad -0.5 \cdot \text{Cov}(E_{t-1}, \ E_{t-1}) + 0.5^2 \cdot \text{Cov}(E_{t-1}, \ E_{t-2}) \\ &= \quad 0 - 0 - 0.5 \cdot \text{Var}(E_{t-1}) + 0 = -0.5 \end{aligned}$$

etc.

b) $Y_t = Y_{t-1} + E_t$ is **not** stationary (cf. the above plot). We have

$$\begin{aligned} E[Y_t] \quad &= \quad E[Y_{t-1}] + E[E_t] = E[Y_{t-1}] = 0 \quad \text{for } Y_0 = 0 \\ \text{Var}(Y_t) \quad &= \quad \text{Var}(Y_{t-1}) + \text{Var}(E_t) = \text{Var}(Y_{t-1}) + 1 \\ &\implies \quad \text{Var}(Y_t) \text{ is not constant.} \end{aligned}$$
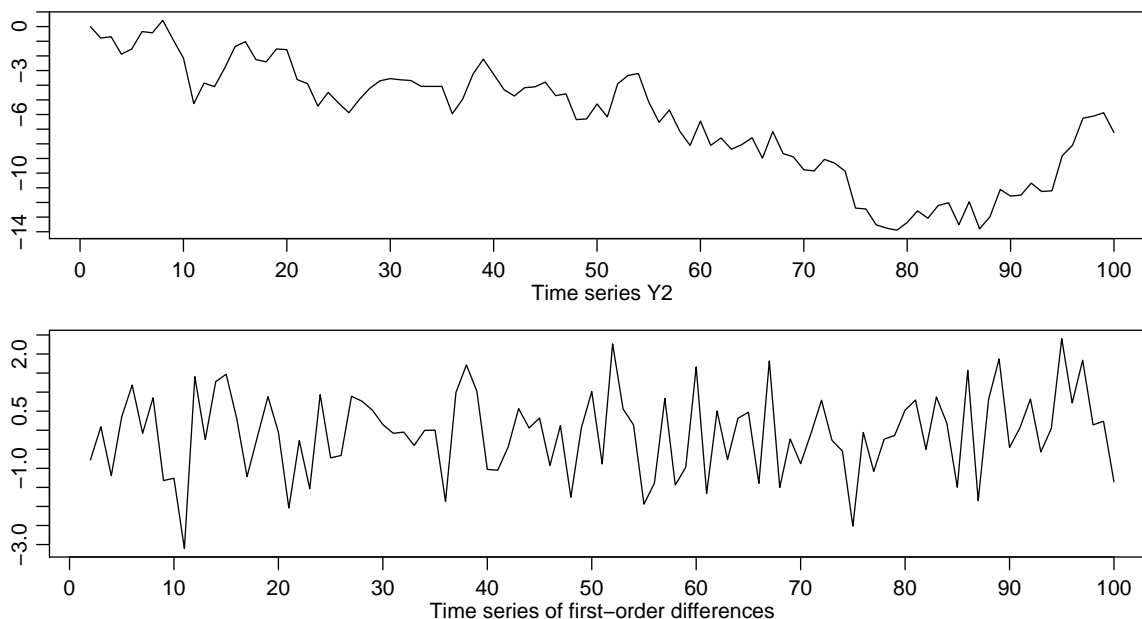
(The variance decreases by 1 at every time step)

As a consequence, $\text{Cov}(Y_t, Y_{t+h})$ is not constant in t (Take e.g. $h = 0$).

Note:

- Y_{t-1} and E_t are uncorrelated random variables, since

$$\text{Cov}(Y_{t-1}, E_t) = \ldots = \text{Cov}(Y_0, E_t) = 0$$

- The time series $X_t := Y_t - Y_{t-1}$ of first-order differences is stationary. We obtain $X_t = E_t$.

Time series Y2

Time series of first–order differences

Commands for drawing this plot:

```
> zeitsim <- ts(zeitsim)
> X2 <- diff(zeitsim[,2], lag=1)
> par(mfrow=c(2,1))
> plot(zeitsim[,2],ylab=" ",xlab="Time series Y2")
> plot(X2,ylab=" ",xlab="Time series of first-order differences")
```

c) $Y_t = \frac{1}{2} \cdot Y_{t-1} + E_t$ is stationary (an AR(1) process – cf. Chapter 5.1). We have

$$E[Y_t] = \frac{1}{2} \cdot E[Y_{t-1}] = 0 \quad \text{for } Y_0 = 0$$

$$\text{Var}(Y_t) = (\frac{1}{2})^2 \cdot \text{Var}(Y_{t-1}) + 1 = 1 + \frac{1}{4} + \frac{1}{16} + \ldots = \sum_{i=0}^{\infty} (\frac{1}{4})^i$$

$$= \frac{1}{1 - \frac{1}{4}} = \frac{4}{3} = const.$$

(Sum of a geometric series)

With some computational effort, we can show that $\text{Cov}(Y_t, Y_{t+h}) = \ldots = 0.5^{|h|} \cdot \text{Var}(Y_t)$. Thus $\text{Cov}(Y_t, Y_{t+h}) = 0.5^{|h|} \cdot \frac{4}{3}$, and $\text{Cov}(Y_t, Y_{t+h})$ is independent of t.

d) $Y_t = Y_{t-1} \cdot E_t$ is **not** stationary, as $\text{Var}(Y_t)$ is not constant. It is not easy to show this mathematically, but it is evident from the time series plot.

For this process, new values are about the same as the old ones $\pm 5\%$.

Note:

- $Y_t^* = Y_t/Y_{t-1}$ is stationary, and $Y_t^* = E_t$.
- $Y_t^{**} = \log(Y_t) - \log(Y_{t-1}) = \log(Y_t^*)$ is also stationary.
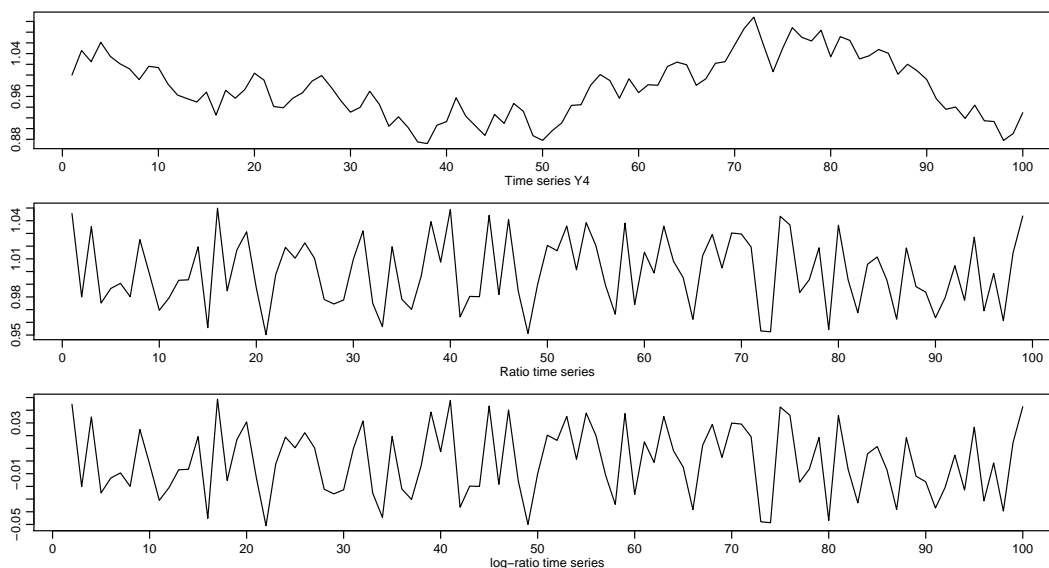
Note that the plots of Y_t^* and Y_t^{**} differ only in terms of scale.

R commands for the plots above:

```
> par(mfrow=c(3,1))
> plot(zeitsim[,4],ylab=" ",xlab="Time series Y4")
> Y4 <- zeitsim[,4]
> ts(X4 <- Y4[-1] / Y4[-100])
> plot(X4,ylab=" ",xlab="Ratio time series")
> X <- log(zeitsim[,4])
> X5 <- diff(X, lag=1)
> plot(X5, ylab=" ", xlab="log-ratio time series")
```

Appendix for R enthusiasts: We can create, and form a matrix out of, the 4 time series in the following way:

```
> e1 <- rnorm(100)
> y1 <- filter(e1, c(1, -.5), side = 1)
> y1[1] <- e1[1]

> e2 <- c(0,rnorm(99))                                    # Y0 must be 0
> y2 <- cumsum(e2)

> e3 <- rnorm(99)
> y3 <- c(0, filter(e3, 0.5, "recursive"))

> e4 <- c(1,runif(n=99, min=0.95, max=1.05))
> y4 <- cumprod(e4)

> zeitsim <- cbind(ts(y1), ts(y2), ts(y3), ts(y4))
```

(Of course there are also other methods, some of them probably more efficient than this one.)