

Chapter 2

Trend and Seasonal Components

If the plot of a TS reveals an increase of the seasonal and noise fluctuations with the level of the process then some transformation may be necessary before doing any further analysis of the TS. For example, the fluctuations in the UK unemployment data and the sales of an industrial heater data are considerably reduced (evened) by *log* transformation; compare Figures 1.2 and 2.1 and Figures 1.7 and 2.2.

Our aim is to estimate and eliminate the deterministic components $m(t)$ and $s(t)$ in such a way that the random noise Y_t is a stationary process, in the sense that its fluctuations are stable and it has no trend. A formal definition of stationarity will be given in Chapter 4. Such a process can be modelled and analyzed using the well developed theory of stationary TS. All these components are parts of the original TS and their overall analysis leads to the analysis of X_t .

2.1 Elimination of Trend in the Absence of Seasonality

When there is no seasonality, the model (1.1) simplifies to

$$X_t = m_t + Y_t, \quad t = 0, 1, \dots, n, \quad (2.1)$$

where $E(X_t) = m_t$, that is, we assume that $E(Y_t) = 0$.

2.1.1 Least Square Estimation of Trend

We assume that $m(t) = m(t, \boldsymbol{\beta})$ is a function of time t , such as a polynomial of degree k , depending on some unknown parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$. We

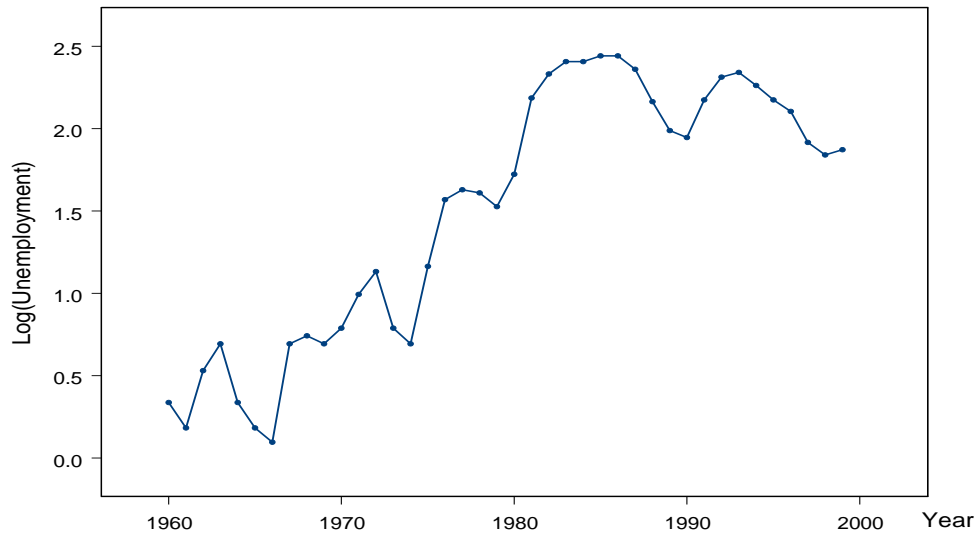


Figure 2.1: Natural logarithm of the UK unemployment data, compare Fig. 1.2

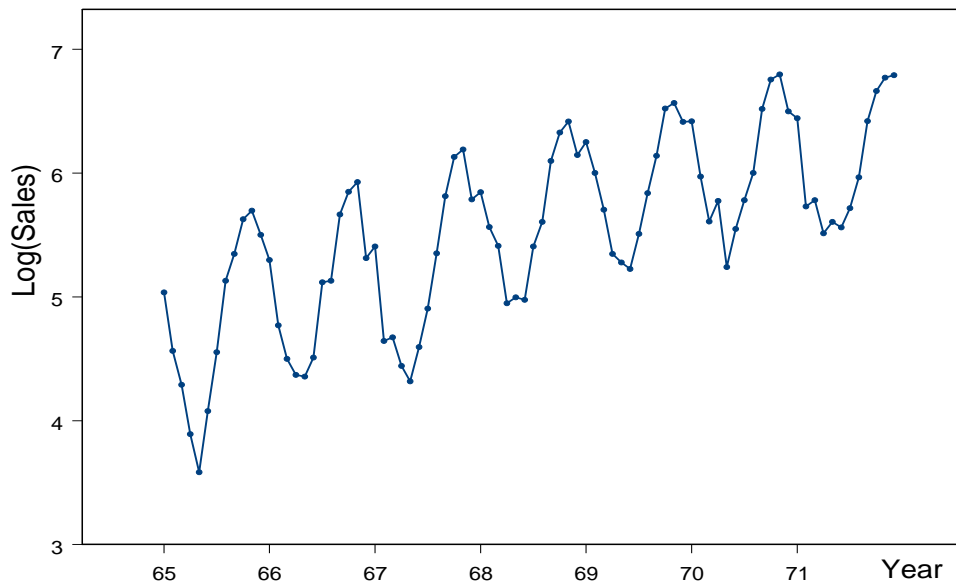


Figure 2.2: Natural logarithm of the data: Sales of an industrial heater.

estimate the parameters by fitting the function to the data x_t , that is by minimizing the sum of squares of differences $x_t - m(t, \boldsymbol{\beta})$. To find a general form of the estimator we work with the rvs X_t rather than their realizations x_t . We find the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ by minimizing the function:

$$S(\boldsymbol{\beta}) = \sum_{t=1}^n (X_t - m(t, \boldsymbol{\beta}))^2. \quad (2.2)$$

For example, consider a linear trend $m(t, \beta_0, \beta_1) = \beta_0 + \beta_1 t$. Then

$$S(\boldsymbol{\beta}) = \sum_{t=1}^n (X_t - \beta_0 - \beta_1 t)^2.$$

To minimize $S(\boldsymbol{\beta})$ with respect to the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ we need to calculate the derivatives of $S(\boldsymbol{\beta})$ with respect to β_i , $i = 0, 1$, compare them to zero and solve the resulting equations, i.e.,

$$\begin{cases} \frac{dS(\boldsymbol{\beta})}{d\beta_0} = -2 \sum_{t=1}^n (X_t - \beta_0 - \beta_1 t) = 0 \\ \frac{dS(\boldsymbol{\beta})}{d\beta_1} = -2 \sum_{t=1}^n (X_t - \beta_0 - \beta_1 t)t = 0. \end{cases}$$

This can be written as

$$\begin{cases} \sum_{t=1}^n X_t - n\beta_0 - \beta_1 \sum_{t=1}^n t = 0 \\ \sum_{t=1}^n X_t t - \beta_0 \sum_{t=1}^n t - \beta_1 \sum_{t=1}^n t^2 = 0. \end{cases} \quad (2.3)$$

Solving the first equation for β_0 we obtain

$$\beta_0 = \frac{1}{n} \sum_{t=1}^n X_t - \beta_1 \frac{1}{n} \sum_{t=1}^n t = \bar{X}_t - \beta_1 \bar{t}.$$

Hence, equations (2.3) can be written as

$$\begin{cases} \beta_0 = \bar{X}_t - \beta_1 \bar{t} \\ n\bar{X}_t \bar{t} - \beta_0 n\bar{t} - \beta_1 n\bar{t}^2 = 0. \end{cases}$$

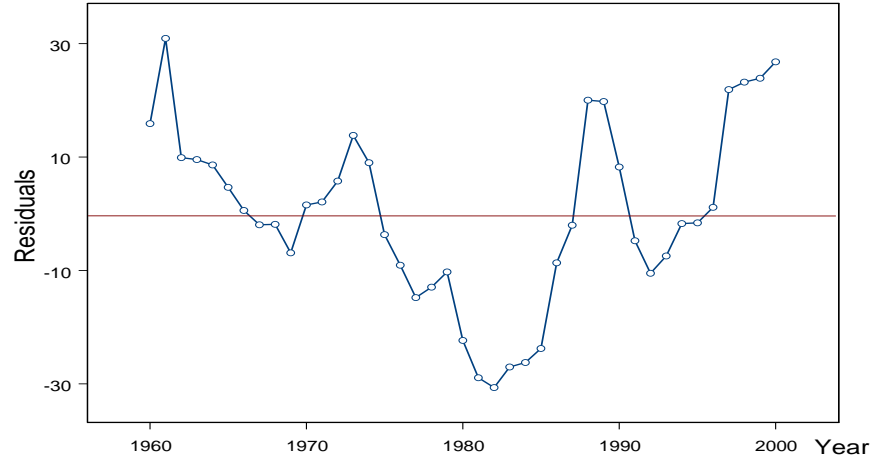


Figure 2.3: Residuals of a linear fit to the UK consumption data; compare Fig. 1.8

Then substituting β_0 to the second equation we can calculate the formula for β_1 and obtain the estimators of both parameters

$$\begin{cases} \hat{\beta}_0 = \bar{X}_t - \hat{\beta}_1 \bar{t} \\ \hat{\beta}_1 = \frac{\overline{X_t t} - \bar{X}_t \bar{t}}{t^2 - (\bar{t})^2}. \end{cases}$$

We need to distinguish an estimator as a function of rvs from an estimate (i.e., a value of the function for a given set of data). This should be obvious from the context.

Fitting a linear trend to the UK consumption data (see Figure 1.8) we obtain the following estimates

$$\hat{\beta}_0 = -12349.7, \quad \hat{\beta}_1 = 6.3754.$$

Hence the estimate of the trend is

$$\widehat{m(t)} = -12349.7 + 6.3754t.$$

The random noise is then estimated as

$$\hat{y}_t = x_t - \hat{m}_t = x_t - (-12349.7 + 6.3754t).$$

The plot (see Figure 2.3) of the residuals does not show any clear trend. However it indicates correlations of the random noise. In this course we will be modelling

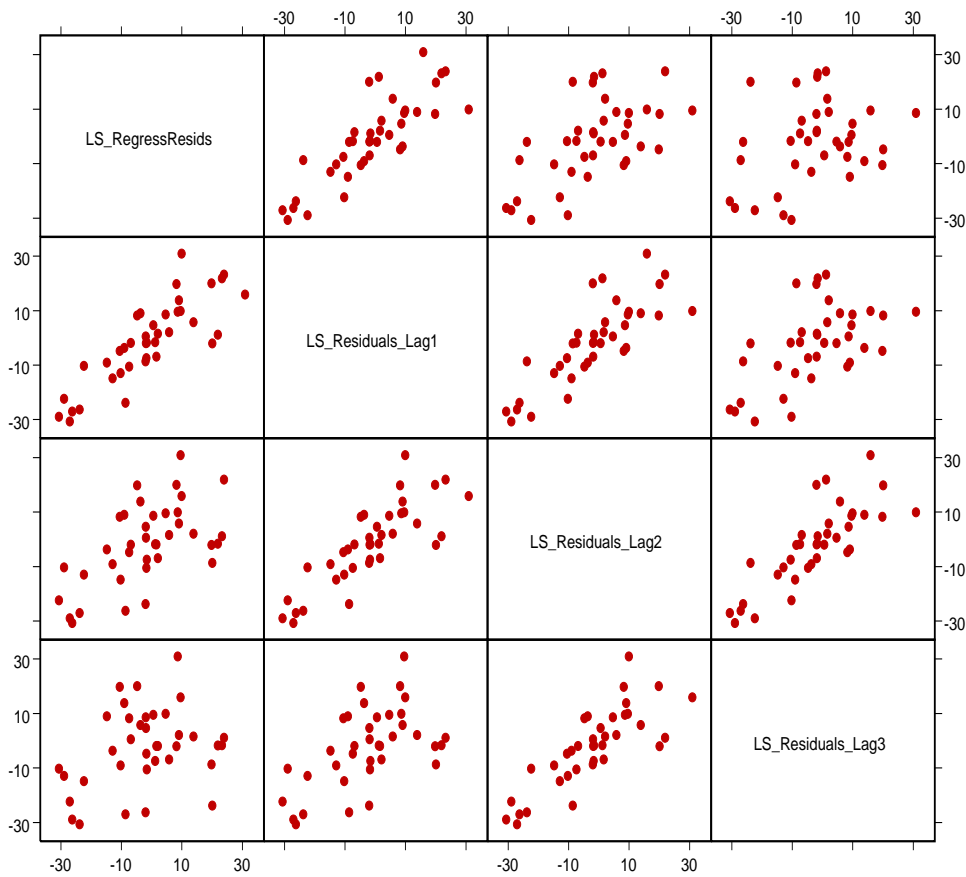


Figure 2.4: Correlation matrix for the residuals of the Least Squares Regression fit; UK consumption data.

this kind of random variables to be able to predict future values of the TS, for example next year values of consumption in the UK.

The matrix plot shows that indeed the residuals are dependent, at least up to two neighboring values.

2.1.2 Smoothing by a Moving Average

Let q be a positive integer. We call

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t+j} \quad (2.4)$$

a **symmetric two-sided moving average** of the process $\{X_t\}$.

Simple example

Take the following realization of $\{X_t\}$:

$$\{x_t\} = \{5 \ 3 \ 2 \ 4 \ 5 \ 6 \ 7 \ 8 \ 3 \ 9\}$$

For $q = 1$ we have the following realizations of W_t :

$$\begin{aligned} w_1 &= ? \\ w_2 &= \frac{1}{3}(5 + 3 + 2) = \frac{10}{3} \\ w_3 &= \frac{1}{3}(3 + 2 + 4) = \frac{9}{3} \\ w_4 &= \frac{1}{3}(2 + 4 + 5) = \frac{11}{3} \\ &\vdots \\ w_9 &= \frac{1}{3}(8 + 3 + 9) = \frac{20}{3} \\ w_{10} &= ? \end{aligned}$$

Equation (2.4) defines W_t for $q+1 \leq t \leq n-q$.

Problem: How to calculate values of W_t for $t \leq q$ and for $t > n-q$?

One possibility is to define $X_t = X_1$ for $t < 1$ and $X_t = X_n$ for $t > n$. Then in the simple example we would get $x_0 = x_1 = 5$ and $x_{11} = x_{10} = 9$ and

$$\begin{aligned} w_1 &= \frac{1}{3}(5 + 5 + 3) = \frac{13}{3} \\ w_{10} &= \frac{1}{3}(3 + 9 + 9) = \frac{21}{3}. \end{aligned}$$

What can we achieve calculating moving average? If $X_t = m_t + Y_t$, then

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^q m_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^q Y_{t+j}.$$

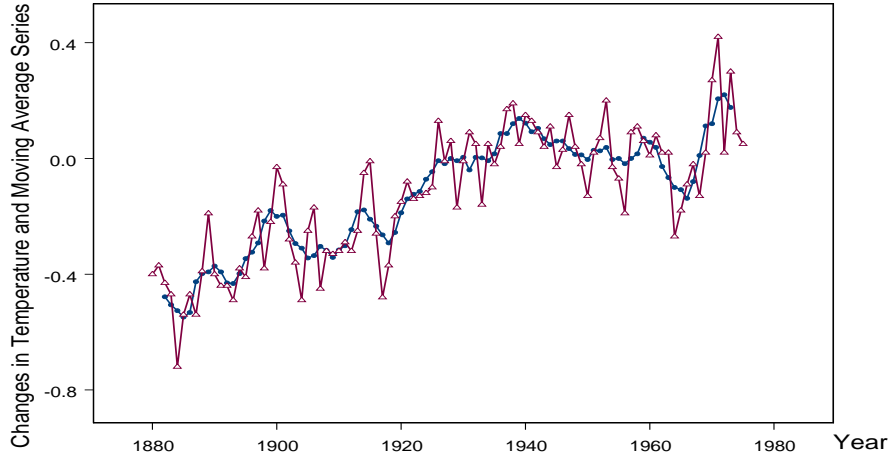


Figure 2.5: Surface air "temperature change" for the globe. The TS data for years 1880-1985 [Degrees Celsius] and a symmetric five elements moving average.

If we can assume that the sum of noise values is close to zero, then W_t evaluates the trend m_t . It is a good evaluation of m_t if, over $[t - q, t + q]$, the trend is approximately linear. It is used as an estimator of trend and we write

$$\hat{m}_t = W_t = \frac{1}{2q+1} \sum_{j=-q}^q X_{t+j} \text{ for } q+1 \leq t \leq n-q.$$

Then, the estimator of the noise is

$$\hat{Y}_t = X_t - W_t.$$

The moving average values w_t for $q = 2$ for the surface air "temperature change" for the globe data are shown in Figure 2.5 and the residuals $y_t = x_t - w_t$ in Figure 2.6. There is no apparent trend in the residuals.

The process $\{W_t\}$ is a linear function (**linear filter**) of random variables $\{X_t\}$. In general, it may be written as

$$W_t = \sum_{j=-\infty}^{\infty} a_j X_{t+j}, \quad (2.5)$$

where

$$a_j = \begin{cases} \frac{1}{2q+1} & \text{for } -q \leq j \leq q, \\ 0 & \text{for } |j| > q, \end{cases}$$

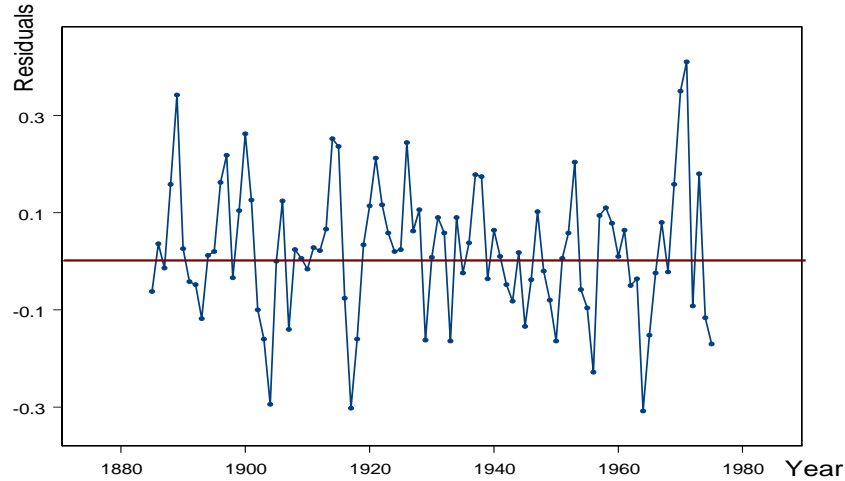


Figure 2.6: Residuals after removing trend calculated as a symmetric five elements moving average. Surface air "temperature change" for the globe data, 1880-1985.

if W_t is a symmetric two-sided ma.

Graphically, it could be presented as in Figure 2.7. It smoothes the data by averaging, that is by removing rapid fluctuations (bursts). It attenuates the noise.

The set of coefficients $\{a_j\}_{j=-q,\dots,q}$ is also called a linear filter.

If

$$m_t = \sum_{j=-q}^q a_j X_{t+j}$$

then we say that the filter $\{a_j\}$ **passes through without distortion**.

The weights $\{a_j\}$ neither have to be all equal nor symmetric. By applying appropriate weights it is possible to include a wide choice of trend functions. Suppose that the trend follows a polynomial of order three and we want to find a seven points filter $\{a_j\}_{j=-3,\dots,3}$, that is a moving average

$$W_t = \sum_{j=-3}^3 a_j X_{t+j}.$$



Figure 2.7: Graphical representation of a linear filter

If the trend is

$$m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3,$$

then a filter which passes through without distortion is such that

$$m_t = W_t \text{ and so } \sum_{j=-3}^3 a_j X_{t+j} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3.$$

We are interested in the middle point, i.e., for $j = 0$. It is convenient to take values of t such that the middle one is zero. For a seven point average they will be

$$t = -3, -2, -1, 0, 1, 2, 3.$$

Then the trend at the middle point, and so the value of the moving average, is

$$m(t = 0) = \beta_0.$$

To find β_0 we may apply the Least Squares method for the seven points. Differentiating the function

$$S(\boldsymbol{\beta}) = \sum_{t=-3}^3 (X_t - (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3))^2$$

with respect to vector $\boldsymbol{\beta}$ gives the following four equations

$$\left\{ \begin{array}{l} \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_0} = -2 \sum_{t=-3}^3 (X_t - (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3)) = 0 \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_1} = -2 \sum_{t=-3}^3 (X_t - (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3)) t = 0 \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_2} = -2 \sum_{t=-3}^3 (X_t - (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3)) t^2 = 0 \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_3} = -2 \sum_{t=-3}^3 (X_t - (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3)) t^3 = 0. \end{array} \right.$$

After simple manipulation we obtain

$$\left\{ \begin{array}{l} \sum_{t=-3}^3 X_t = \sum_{t=-3}^3 (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) \\ \sum_{t=-3}^3 X_t t = \sum_{t=-3}^3 (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) t \\ \sum_{t=-3}^3 X_t t^2 = \sum_{t=-3}^3 (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) t^2 \\ \sum_{t=-3}^3 X_t t^3 = \sum_{t=-3}^3 (\beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3) t^3. \end{array} \right.$$

However, odd powers of t sum to zero, that is

$$\sum_{t=-3}^3 t = \sum_{t=-3}^3 t^3 = \sum_{t=-3}^3 t^5 = 0,$$

what simplifies the equations to

$$\left\{ \begin{array}{l} \sum_{t=-3}^3 X_t = 7\beta_0 + 28\beta_2 \\ \sum_{t=-3}^3 X_t t = 28\beta_1 + 196\beta_3 \\ \sum_{t=-3}^3 X_t t^2 = 28\beta_0 + 196\beta_2 \\ \sum_{t=-3}^3 X_t t^3 = 196\beta_1 + 1588\beta_3 \end{array} \right.$$

of which only equations 1 and 3 involve β_0 , so we can ignore the other two equations. Solving equations 1 and 3 for β_0 then gives

$$\beta_0 = \frac{1}{21}(-2X_{-3} + 3X_{-2} + 6X_{-1} + 7X_0 + 6X_1 + 3X_2 - 2X_3).$$

The trend value for any point is then the weighted average of the seven points of which that point is in the centre,

$$m_t = W_t = \frac{1}{21}(-2X_{t-3} + 3X_{t-2} + 6X_{t-1} + 7X_t + 6X_{t+1} + 3X_{t+2} - 2X_{t+3}).$$

The weights are

$$\left(-\frac{2}{21}, \frac{3}{21}, \frac{6}{21}, \frac{7}{21}, \frac{7}{21}, \frac{3}{21}, -\frac{2}{21}\right) = \frac{1}{21}(-2, 3, 6, 7, 6, 3, -2),$$

which due to the symmetry can be written in a short way, such as

$$\frac{1}{21}(-2, 3, 6, \mathbf{7}),$$

the bold typeface indicating the middle weight.

In general, if we want to find a linear filter $\{a_j\}$ appropriate for fitting a polynomial trend of order k , we have to minimize

$$S(\beta) = \sum_{t=-q}^q (X_t - (\beta_0 + \beta_1 t + \dots + \beta_k t^k))^2$$

and find the solution for β_0 in terms of X_t , $t = -q, \dots, q$.

The following filters are often applied:

- $\{a_j\}$ being successive terms in the expansion $(\frac{1}{2} + \frac{1}{2})^{2q}$. For $q = 1$ we get

$$\{a_j\}_{j=-1,0,1} = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right).$$

For large q the weights approximate to a normal curve.

- So called Spencer's 15-point moving average (used for smoothing mortality statistics to obtain life tables). Here $q = 7$ and the weights are

$$\{a_j\}_{j=-7,\dots,7} = \frac{1}{320}(-3, -6, -5, 3, 21, 46, 67, \mathbf{74}).$$

- The Henderson's moving average, which aims to follow a cubic polynomial trend. For example, for $q = 4$ the Henderson's m.a. is

$$\{a_j\}_{j=-4,\dots,4} = (-0.041, -0.010, -0.119, 0.267, \mathbf{0.330}).$$

Having estimated the trend we can calculate the residuals, as in the previous case, i.e.,

$$\hat{y}_t = x_t - w_t = x_t - \sum_{j=-q}^q a_j x_{t+j} = \sum_{j=-q}^q b_j x_{t+j},$$

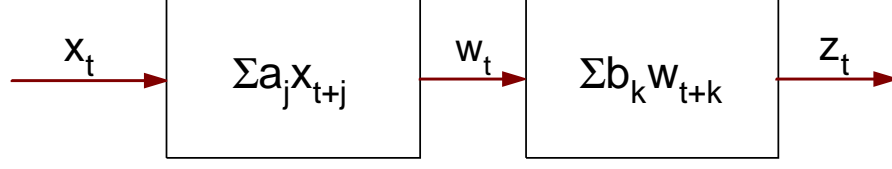


Figure 2.8: Graphical representation of a double linear filter

where

$$b_j = \begin{cases} -a_j, & \text{for } j = -q, \dots, q, j \neq 0 \\ 1 - a_0, & \text{for } j = 0 \end{cases}$$

and if $\sum a_j = 1$ then $\sum b_j = 0$.

Sometimes it may be necessary to apply filters more than once to obtain a stationary process. Two filters might be represented as in Figure 2.8.

The resulting time series $\{Z_t\}$ can be written as

$$Z_t = \sum_{k=-s}^s b_k W_{t+k} = \sum_{k=-s}^s b_k \sum_{j=-q}^q a_j X_{t+k+j} = \sum_{i=-(s+q)}^{s+q} c_i X_{t+i}.$$

For example, for $s = q = 1$ we obtain

$$\begin{aligned} Z_t &= \sum_{k=-1}^1 b_k \sum_{j=-1}^1 a_j X_{t+k+j} = \\ &= b_{-1}a_{-1}X_{t-2} + (b_{-1}a_0 + b_0a_{-1})X_{t-1} + (b_{-1}a_1 + b_0a_0 + b_1a_{-1})X_t + \\ &+ (b_0a_1 + b_1a_0)X_{t+1} + b_1a_1X_{t+2} = \sum_{i=-2}^2 c_i X_{t+i}. \end{aligned}$$

The weights $\{c_i\}$ can be obtained by so called **convolution** of the weights $\{a_j\}$ and $\{b_k\}$, the operation denoted by \star ,

$$\{c_i\} = \{a_j\} \star \{b_k\},$$

where c_i is a sum of all products $a_j b_k$ for which $j + k = i$.

Using this operation in the above example gives

$$\begin{aligned} \{c_i\} &= (a_{-1}, a_0, a_1) \star (b_{-1}, b_0, b_1) \\ &= (a_{-1}b_{-1}, a_0b_{-1} + a_{-1}b_0, a_{-1}b_1 + a_0b_0 + a_1b_{-1}, a_0b_1 + a_1b_0, a_1b_1) \\ &= (c_{-2}, c_{-1}, c_0, c_1, c_2). \end{aligned}$$

If one of the moving averages is not symmetric, we calculate the convoluted weights in the same way, see the example below.

$$\begin{aligned} &(a_{-1}, a_0, a_1) \star (b_0, b_1) = \\ &(a_{-1}b_0, a_0b_0 + a_{-1}b_1, a_0b_1 + a_1b_0, a_1b_1) = \\ &(c_{-1}, c_0, c_1, c_2). \end{aligned}$$

With $\{a_j\} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and $\{b_k\} = (\frac{1}{2}, \frac{1}{2})$ we get

$$\{c_i\} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \star \left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}\right).$$

and the new series Z_t is the following combination of X_t

$$Z_t = \frac{1}{6}X_{t-1} + \frac{1}{3}X_t + \frac{1}{3}X_{t+1} + \frac{1}{6}X_{t+2}.$$

2.1.3 Differencing

Differencing is a trend removing operation by using special kind of a linear filter with weights $(-1, 1)$. This procedure is often repeated until a stationary series is obtained.

We denote the **first (lag 1) difference operator** by ∇ , that is

$$\nabla X_t = X_t - X_{t-1}.$$

Here another operator comes into play, it is called the **backward shift operator**, usually denoted by B , such that

$$BX_t = X_{t-1}.$$

Hence,

$$\nabla X_t = X_t - BX_t = (1 - B)X_t.$$

Note that

$$X_{t-2} = BX_{t-1} = B(BX_t) = B^2X_t.$$

In general

$$B^j X_t = X_{t-j}.$$

Similarly,

$$\begin{aligned} \nabla^2 X_t &= \nabla(\nabla X_t) \\ &= (1 - B)(1 - B)X_t \\ &= (1 - 2B + B^2)X_t \\ &= X_t - 2X_{t-1} + X_{t-2}. \end{aligned}$$

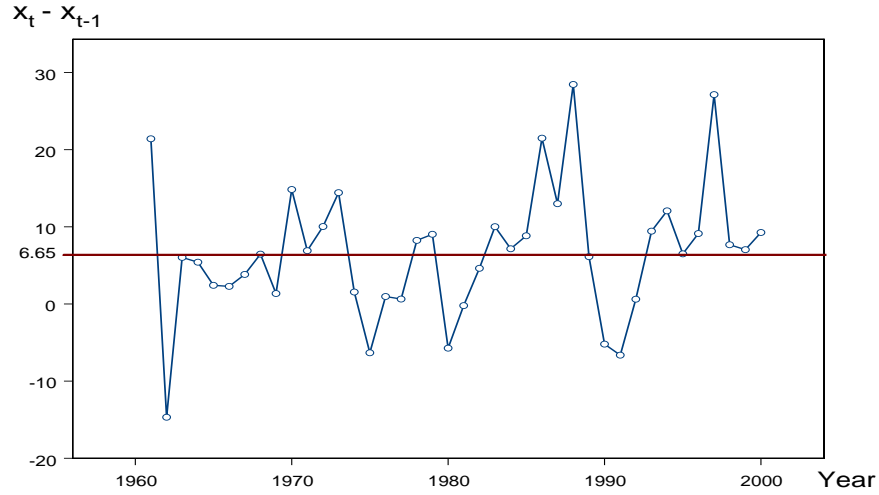


Figure 2.9: The differenced data: UK Consumption, compare Figures 1.1 and 1.8

By the way, this is what we would get by convolution of two linear filters with weights $(-1, 1)$

$$\begin{aligned} (a_{-1}, a_0) \star (b_{-1}, b_0) &= (-1, 1) \star (-1, 1) = \\ &((-1) \cdot (-1), (-1) \cdot 1 + 1 \cdot (-1), 1 \cdot 1) = \\ &(1, -2, 1) = (c_{-2}, c_{-1}, c_0). \end{aligned}$$

In general, we may write

$$\nabla^j X_t = \nabla(\nabla^{j-1} X_t), \quad \text{for } j \geq 1, \quad \nabla^0 X_t = X_t.$$

Assume that a TS model is

$$X_t = m_t + Y_t,$$

where the trend m_t is a polynomial of degree k . Then for $k = 1$ we have

$$\begin{aligned} \nabla X_t &= X_t - X_{t-1} \\ &= m_t + Y_t - (m_{t-1} + Y_{t-1}) \\ &= \beta_0 + \beta_1 t - [\beta_0 + \beta_1(t-1)] + \nabla Y_t \\ &= \beta_1 + \nabla Y_t. \end{aligned}$$

Similarly, for $k = 2$ we obtain

$$\begin{aligned}
 \nabla^2 X_t &= X_t - 2X_{t-1} + X_{t-2} \\
 &= m_t + Y_t - 2(m_{t-1} + Y_{t-1}) + m_{t-2} + Y_{t-2} \\
 &= \beta_0 + \beta_1 t + \beta_2 t^2 - 2[\beta_0 + \beta_1(t-1) + \beta_2(t-1)^2] \\
 &\quad + \beta_0 + \beta_1(t-2) + \beta_2(t-2)^2 + \nabla^2 Y_t \\
 &= 2\beta_2 + \nabla^2 Y_t.
 \end{aligned}$$

For a polynomial trend of degree k we have

$$\nabla^k X_t = k! \beta_k + \nabla^k Y_t.$$

It means that if the noise fluctuates about zero, then k -th differencing of a TS with a polynomial trend of degree k should give a stationary process with mean about $k! \beta_k$.

Example UK consumption data.

The data show a linear trend. The least squares method gives the following estimates

$$\hat{m}(t) = -12349.7 + 6.3754t$$

and we can see from the graph of the differenced data (Figure 2.9) that the mean is close to β_1 and that there is no trend.