3.5.4. **Two simple averaging models for decomposition.**

$\boldsymbol{x}$	$= [x_1, \ldots, x_n]^T$	...	uniform sample vector of the time series $X = \{X_t \vert t \in \mathbb{Z}\}$
$\boldsymbol{e}$	$= [e_1, \ldots, e_n]^T$	...	random errors
t	$\in \{1, 2, \ldots, n\}$		

We are assuming both models:

- additive: $X_t = Tr_t + Sz_t + C_t + E_t$ (AM)

  or

- multiplicative: $X_t = Tr_t \, Sz_t \, C_t \, E_t$ (MM).

Let d be the period length of Sz_t, and $n = md$. We introduce a matrix $m \times d$

$$X = \begin{bmatrix} x_1 & \ldots & x_k & \ldots & x_d \\ x_{d+1} & \ldots & x_{d+k} & \ldots & x_{2d} \\ \vdots & \ldots & \vdots & \ldots & \vdots \end{bmatrix} = [x_{j,k}],$$

where $x_{j,k} = x_{(j-1)d+k}$ (e.g. j =year, k=month).
Let us denote as `vec` the operator rearranging columns of a matrix downwards into one column vector. Then we can write $\boldsymbol{x} = \mathrm{vec}(X^T)$.
MATLAB: $X = \texttt{reshape}(\boldsymbol{x}, d, m)$.'

The j-row $X(j, :)$ stands for samples from j-th cycle of Sz_t including Tr_t, C_t and E_t.
The k-th column $X(:, k)$ gathers samples which are at k-th position within each cycle, again including all components Tr_t, C_t and E_t.

Let $E = [e_{j,k}]$ be the matrix of errors defined alike.

The seasonal component Sz_t is fully determined by its values in one cycle $\boldsymbol{s} := (s_1, \ldots, s_d) := (Sz_1, \ldots, Sz_d)$ with mean $\bar{s} := \frac{s_1 + \cdots + s_d}{d}$ satisfying:

$$\bar{s} = \begin{cases} 0 & \text{v AM} \\ 1 & \text{v MM} \end{cases} \tag{3.5.2}$$

We are going to find estimates $\widehat{s}_k$ fulfilling (3.5.2).

3.5.4.1. **The small trend method** [MATLAB: `szsmtr`]

Assumptions:
$Tr_t + C_t$, or $Tr_t C_t$ should be approximately constant within each cycle of Sz_t:

$$\left.\begin{array}{lllll} \text{AM:} & x_{j,k} - s_k - e_{j,k} & = & Tr_t + C_t & \approx & m_j \\ \text{MM:} & x_{j,k}/(s_k\, e_{j,k}) & = & Tr_t\, C_t & \approx & m_j \end{array}\right\} \text{ in } j\text{-th period}, \; t = (j-1)d + k.$$

Algorithm:

(1) $\widehat{m}_j = \frac{1}{d}\sum_{k=1}^{d} x_{j,k}$ ...estimate of m_j for $j = 1, \dots, m$.
(2) for $k = 1, \dots, d$ we compute:

$$\text{AM:} \quad \widehat{s}_k = \frac{1}{m}\sum_{j=1}^{m} \underbrace{(x_{j,k} - \widehat{m}_j)}_{\approx\, s_k + e_{j,k}}$$

$$\text{MM:} \quad \widehat{s}_k = \frac{1}{m}\sum_{j=1}^{m} \underbrace{x_{j,k}/\widehat{m}_j}_{\approx\, s_k\, e_{j,k}}$$

It is easy to see that these estimates fulfil (3.5.2):

$$\text{AM:} \quad \overline{\widehat{s}} = \frac{1}{d}\sum_{k=1}^{d}\left(\frac{1}{m}\sum_{j=1}^{m}(x_{j,k} - \widehat{m}_j)\right) = \frac{1}{m}\sum_{j=1}^{m}\left(\underbrace{\frac{1}{d}\sum_{k=1}^{d} x_{j,k}}_{\widehat{m}_j} - \underbrace{\frac{1}{d}\sum_{k=1}^{d}\widehat{m}_j}_{\widehat{m}_j}\right) = 0$$

$$\text{MM:} \quad \overline{\widehat{s}} = \frac{1}{d}\sum_{k=1}^{d}\left(\frac{1}{m}\sum_{j=1}^{m}\frac{x_{j,k}}{\widehat{m}_j}\right) = \frac{1}{m}\sum_{j=1}^{m}\underbrace{\frac{1}{\widehat{m}_j}\underbrace{\frac{1}{d}\sum_{k=1}^{d} x_{j,k}}_{\widehat{m}_j}}_{1} = \frac{1}{m}\, m = 1.$$

To be continued with the steps of 3.5.4.3.

3.5.4.2. **Moving average method** [MATLAB: `szma`]

We introduce vector $\boldsymbol{w} = (w_{-q}, \ldots, w_q)^T$ of equal weights for the common moving average filter:

$$\boldsymbol{w} = \frac{1}{d}(\underbrace{1, 1, \ldots, 1}_{(2q+1)\times})^T \quad \text{for odd} \quad d = 2q + 1,$$
$$\boldsymbol{w} = \frac{1}{d}(\underbrace{1/2, 1, \ldots, 1, 1/2}_{(2q+1)\times})^T \quad \text{for even} \quad d = 2q.$$

We start with the moving average operation:

$$\widehat{m}_t = \begin{cases} \sum_{\tau=-q}^{q} w_\tau x_{t+\tau} & \text{for} \quad q+1 \leq t \leq n-q, \\ x_t & \text{for} \quad 1 \leq t \leq q \text{ or } n-q+1 \leq t \leq n. \end{cases}$$

By the construction of weights, exactly the whole seasonal cycle is being averaged at each position of vector $\boldsymbol{w}$. In addition to suppressing the noise component e_t, this averaging eliminates the seasonal component Sz_t as well due to its zero/unit mean in the AM/MM layout.

That is why the the following assumptions are justified:

Assumptions:

$$\left.\begin{array}{lllll} \text{AM:} & x_t - Sz_t - e_t & = & Tr_t + C_t & \approx \widehat{m}_t \\ \text{MM:} & x_t/(Sz_t\, e_t) & = & Tr_t\, C_t & \approx \widehat{m}_t \end{array}\right\} \text{for } q+1 \leq t \leq n-q.$$

Algorithm:

(1) After having the vector of values smoothed by the moving average operation we rearrange it into a matrix of size $m \times d$ with cycles row-by-row, following the analogy with $\boldsymbol{x}$:
$\widehat{\boldsymbol{m}} = (\widehat{m}_1, \ldots, \widehat{m}_n)^T =: \text{vec}(\widehat{M}^T)$, where $\widehat{M} =: [\widehat{m}_{j,k}]$.
Then $X - \widehat{M}$ in AM, or $X./\widehat{M} = [x_{j,k}/\widehat{m}_{j,k}]$ in MM, have the structure shown below, where there are q zeros, or ones at the beginning of the first, and at the end of the last cycle.

For odd d:

$$X - \widehat{M} = \begin{bmatrix} 0 & \dots & 0 & \bullet & \bullet & \dots & \bullet \\ & \vdots & & & & \vdots & \\ \bullet & \dots & \bullet & \bullet & 0 & \dots & 0 \end{bmatrix},$$

or

$$X./\widehat{M} = \begin{bmatrix} 1 & \dots & 1 & \bullet & \bullet & \dots & \bullet \\ & \vdots & & & & \vdots & \\ \bullet & \dots & \bullet & \bullet & 1 & \dots & 1 \end{bmatrix}.$$

For even d:

$$X - \widehat{M} = \begin{bmatrix} 0 & \dots & 0 & \bullet & \dots & \bullet \\ & \vdots & & & \vdots & \\ \bullet & \dots & \bullet & 0 & \dots & 0 \end{bmatrix},$$

or

$$X./\widehat{M} = \begin{bmatrix} 1 & \dots & 1 & \bullet & \dots & \bullet \\ & \vdots & & & \vdots & \\ \bullet & \dots & \bullet & 1 & \dots & 1 \end{bmatrix}.$$

(2) For $k = 1, \dots, d$ is computed:

in AM: $\widetilde{s}_k = \frac{1}{m-\delta_k} \sum_{j=1}^{m} \underbrace{(x_{j,k} - \widehat{m}_{j,k})}_{\approx\ s_k + e_{j,k}},$

in MM: $\widetilde{s}_k = \frac{1}{m-\delta_k} [(\sum_{j=1}^{m} \underbrace{x_{j,k}/\widehat{m}_{j,k}}_{\approx\ s_k\, e_{j,k}}) - \delta_k].$

where $\delta_k = \begin{cases} 0 & \text{for } k = \frac{d+1}{2} \text{ (with odd } d \text{ and } k = q+1) \\ 1 & \text{otherwise} \end{cases}$.

(3) $\overline{\overline{s}} = \frac{1}{d} \sum_{k=1}^{d} \widetilde{s}_k$.

(4) We have to center the estimates to assure the validity of (3.5.2):

$\widehat{s}_k = \widetilde{s}_k - \overline{\overline{s}}$ in AM, or $\widehat{s}_k = \widetilde{s}_k / \overline{\overline{s}}$ in MM for $k = 1, \dots, d$.

Again we continue with the steps of 3.5.4.3.

3.5.4.3. **The separation of Tr_t and Sz_t**

(5) Periodization: $\widehat{Sz}_{k+jd} = \widehat{s}_k$ for $k = 1, \ldots, d$ and $j = 0, 1, \ldots, m-1$.

(6) Separating the seasonal component Sz_t:
$y_t := x_t - \widehat{Sz}_t \approx Tr_t + C_t + e_t$ in AM, or $y_t := x_t / \widehat{Sz}_t \approx Tr_t\, C_t\, e_t$ in MM for $t = 1, \ldots, n$.

(7) In y_t the component $Tr_t + C_t$, or $Tr_t\, C_t$ of y_t (possibly after logarithmic transformation in MM) is estimated using appropriate method (e.g. that of section 3.5.1 or 3.5.2). Just only trend may be estimated putting $q = 0$ in 3.5.1 or using smoothing techniques described in next sections, and then individually separate the cyclic component C_t:

(8) Separating the cyclic component C_t:
$C_t + e_t \approx y_t - \widehat{Tr}_t$ in AM, or $C_t\, e_t \approx y_t / \widehat{Tr}_t$ in MM.

(9) Analyzing periodic components of $C_t + e_t$
or log-transformed $C_t\, e_t$ in MM, using periodogram and periodicity tests of 3.4.3. Now these tests have chance to be sensitive to harmonics of the weak cyclic component, which is no more dominated by the presence the other components. Step (7) may be then repeated to obtain final decomposition of Tr_t and C_t with improved model for C_t.

3.5.4.4. **Eliminating Sz_t by differencing at lag d** (see 3.5.3(1)).

Assumptions:
$x_t = Sz_t + m_t + e_t$, where $m_t = Tr_t + C_t$, and the period length of C_t is significantly longer that that of Sz_t.
Algorithm:

(1) $y_t := x_t - x_{t-d} = \underbrace{Sz_t - Sz_{t-d}}_{\approx 0} + \underbrace{m_t - m_{t-d}}_{\widetilde{m}_t} + \underbrace{e_t - e_{t-d}}_{\widetilde{e}_t},$
$t = d+1, \ldots, n$.

(2) We estimate $\widetilde{m}_t$ as of 3.5.4.3(7)–(9) where, of course, $\widetilde{m}_t = m_t - m_{t-d}$ are to be interpreted like seasonal fluctuations.