5. **Non-Stationary TS. Differencing.** (Based on [BD], pp. 28–35)

5.1 **Elimination of trend.** The first step in the analysis of any TS is to plot the data. If the TS behave as if they have no fixed mean level, then the data is non-stationary. Many times, however, although there is no fixed mean, the parts of the series display a certain kind of homogeneity, in sense that the local behavior (on short intervals) for the series is similar, apart from a difference in level and “trend”.

**Example:** Population of US.

This kind of non-stationarity is called **homogeneous**. By suitable **differencing** of such a process we are able to come to a stationary process. This technique is due to Box and Jenkins.

(i) **Def.** Denote: by $\nabla^d X_t$ the $d$th difference of $X_t$ for all $t$, and define: $\nabla X_t \equiv \nabla^1 X_t := X_t - X_{t-1}$, 

$\nabla^2 X_t = \nabla(\nabla X_t) = \nabla(X_t - X_{t-1}) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}$,

Note: $\nabla^2 X_t = (1 - B)^2 X_t$

$\nabla^d X_t = (1 - B)^d X_t = C^0_d X_t - C^1_d X_{t-1} + C^2_d X_{t-2} - \ldots - (-1)^d C^d_d X_{t-d}$.

We call $\nabla X_t = X_t - X_{t-1}$ “the differenced process”.

**Example (i):** $X_t = bt + S_t$, where $S_t$ is a stationary process. (Linear trend).
Then for $W_t = \nabla^2 X_t = X_t - X_{t-1}$ we have:

$$W_t = bt + S_t - b(t - 1) - S_{t-1} = b + (S_t - S_{t-1})$$

Thus, $W_t$ is stationary.

In some cases it is necessary to difference more than one time in order to achieve stationarity.

**Example (ii):** $X_t = bt^2 + S_t$, where $S_t$ is a stationary process. (Quadratic trend).

Then take

$$W_t = \nabla^2 X_t = X_t - 2X_{t-1} + X_{t-2}$$

$$= b[t^2 - 2(t - 1)^2 + (t - 2)^2] + S_t - 2S_{t-1} + S_{t-2}$$

$$= 2b + (S_t - S_{t-1} - S_{t-2})$$

1
= stationary TS.

**Conclude:** If the trend is polynomial of order \( k \), then \( W_t = \nabla^k X_t \) is stationary (differencing \( k \) times eliminates polynomial trend of order \( k \)).

**Example (iii). The Random Walk Model.** ([BD], pp. 9, 17; also example 3.1 (vi) of Lecture 3)

This model is often used to represent non-stationary data (income, price changes, stock prices):

\[
X_t = X_{t-1} + Z_t
\]

or

\[
X_t = \sum_{j=0}^{t-1} Z_{t-j}, \ t = 1, 2, \ldots
\]

Note: \( EX_t = 0, EX_sX_t = \sigma_Z^2 \min(s, t) \) (calculated in 3.1 (vi) of Lecture 3) i.e. the TS is non-stationary. Note also that the differenced process \( W_t = X_t - X_{t-1} \equiv Z_t \) is stationary and that the ACF for \( W \) is \( \rho_k = 0, k > 0 \).

5.2 **Elimination of trend and seasonality via differencing. The Classical Decomposition model.**

**Examples of seasonal data:** Sunspots, Accidental Deaths, etc.

If the data is a realization of the process of the type

\[
(*) X_t = m_t + s_t + S_t,
\]

\( m_t \) = trend component, polynomial of order \( k \)
\( s_t \) = seasonal component, i.e. periodic with period \( d \): \( s_{t+d} = s_t, \sum_{j=1}^{d} s_j = 0 \).
\( S_t \) = stationary process

[The Classical Decomposition Model]

Convert to a stationary process via differencing:

take *lag d difference* : \( W_t := \nabla_d X_t = X_t - X_{t-d} \):

then \( W_t = (m_t - m_{t-d})(\text{trend}) + (S_t - S_{t-d})(\text{stationary}) \). Usually, one needs to difference more to remove the trend.

Note: \( \nabla_d X_t = (1 - B^d)X_t \) removes periodicity (seasonality) with period \( d \).

**(d) ARIMA(p, d, q)**: this is nonstationary model such that after \( d \) difference operations we obtain ARMA(p,q). Recall:

\[
\nabla X_t = X_t - X_{t-1} = (1 - B)X_t
\]

\[
\nabla^2 X_t = \nabla(X_t - X_{t-1}) = X_t - 2X_{t-1} + X_{t-2} = (1 - B)^2 X_t
\]
etc. The model then can be written as:

\[ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} - \ldots - \phi_p W_{t-p} = Z_t - \theta_1 Z_{t-1} - \ldots - \theta_q Z_{t-q} \]

for \( W_t = \nabla^d X_t \). Another ways to write it:

\[ \phi(B)(1 - B)^d X_t = \theta(B) Z_t \]

or

\[ \phi^*(B) X_t = \theta(B) Z_t \text{ with } \phi^*(z) = \phi(z)(1 - z)^d \]

We request: \( \phi(z) \neq 0, \theta(z) \neq 0 \) for all \( |z| \leq 1 \). Note that \( \phi^*(z) \) has unit root of order \( d \).

- Useful for representing data with trend.
- Typical feature: slowly decaying ACF.
- Unit root in AR part: if \( X_t \) comes from the process \( \phi^*(B) X_t = \theta(B) Z_t \) with \( \phi^*(z) = (1 - z)\phi(z) \), then difference to eliminate the unit root: for \( W_t = X_t - X_{t-1} \) we have the model: \( \phi(B) W_t = \theta(B) Z_t \).
- Unit root in MA part: if \( X_t \) comes from the process \( \phi(B) X_t = \theta^*(B) Z_t \) with \( \theta^*(z) = (1 - z)\theta(z) \), then the series was overdifferenced: let \( Y_t \) be an invertible ARMA process: \( \phi(B) Y_t = \theta(B) Z_t \). Let \( X_t = Y_t - Y_{t-1} \). Then \( \phi(B) X_t = \phi(B)(1 - B) Y_t = (1 - B)\theta(B) Z_t \).

Thus, if estimated \( \phi \) and \( \theta \) have roots close to unit roots, check your differencing.

- **Effect of overdifferencing on the variance and ACF**

* Suppose we consider the MA(1) process: \( X_t = (1 + \theta B) Z_t \). Its ACF is given by:

\[ \rho_X(1) = \frac{\theta}{1 + \theta^2}, \quad \rho_X(k) = 0, \quad k \geq 2 \]

The first difference of this process \( W_t = X_t - X_{t-1} \) is the MA(2) process:

\[ W_t = (1 - B)(1 + \theta B) Z_t \equiv (1 - (1 - \theta) B - \theta B^2) Z_t \]

Its ACF \( \rho_W(2) \neq 0 \), i.e. ACF structure is more complicated.

Also, the variance of original process \( \gamma_X(0) = (1 + \theta^2) \sigma_Z^2 \), and the variance of the differenced process \( W_t \) is \( \gamma_W(0) = 2(1 - \theta + \theta^2) \sigma_Z^2 \). Thus, \( \gamma_W(0) - \gamma_X(0) = (1 - \theta)^2 \sigma_Z^2 > 0 \), i.e. the variance of the overdifferenced MA process is larger than that of the original process. Similar calculation can be made for AR(1) process.