Week 3: AR models: ACF, Yule-Walker Equations, PACF. ARMA (p, q) models. Differencing. Non-Stationary models: ARIMA. Transformations.

4.3 AR(p) models: $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t,\ \{Z_t\} \sim WN(0, \sigma_Z^2)$.

Using shift operator $B : BX_t = X_{t-1}$, AR(p) model can be written as

$X_t = \phi_1 BX_t + \phi_2 B^2 X_t + \ldots + \phi_p B^p X_t + Z_t$ or $(1 - \phi_1 B - \ldots - \phi_p B^p)X_t = Z_t$.

Let $\phi(z) \overset{def}{=} 1 - \phi_1 z - \ldots - \phi_p z^p$. Then, AR(p) model is written as $\phi(B)X_t = Z_t$.

AR(p) is stationarity when $\phi(z) \equiv (1 - \phi_1 z - \ldots - \phi_p z^p) \neq 0$ **for all** $|z| \leq 1$**.**

AR(p) is always invertible.

ACF: We show that ACF satisfies the **Yule-Walker equations**:

(4.3.1) $\rho_X(k) - \phi_1 \rho_X(k-1) - \phi_2 \rho_X(k-2) - \ldots - \phi_p \rho_X(k-p) = 0$.

which has a solution of the form

(4.3.2) $\rho_X(k) = A_1 \alpha_1^k + \ldots + A_p \alpha_p^k$, where $\alpha_1, \ldots, \alpha_k$ are the roots of $x^p - \phi_1 x^{p-1} - \ldots - \phi_p = 0$ with $|\alpha_i| < 1$.

Derivation of Yule-Walker equations:

$$\gamma_X(k) = EX_t X_{t-k} = E\{(\phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + Z_t)X_{t-k}\}$$
$$= \phi_1 E(X_{t-1}X_{t-k}) + \phi_2 E(X_{t-2}X_{t-k}) + \ldots + \phi_p E(X_{t-p}X_{t-k}) + E(Z_t X_{t-k}).$$

Note:
- $E(X_{t-1}X_{t-k}) = \gamma_X(k-1);\ \ldots, E(X_{t-p}X_{t-k}) = \gamma_X(k-p)$;
- $E(Z_t X_{t-k}) = 0$ if $k \geq 1$ and σ_Z^2 if $k = 0$. (*from 4.1 of Lecture 4, AR(1) cov calculation).*

Thus,

$$\gamma_X(k) = \begin{cases} \phi_1 \gamma_X(k-1) + \ldots + \phi_p \gamma_X(k-p), & k \geq 1 \\ \phi_1 \gamma_X(1) + \ldots + \phi_p \gamma_X(p) + \sigma_Z^2, & k = 0\ (*) \end{cases}$$

For $k \geq 1$, divide $\gamma_X(k)$ by $\gamma_X(0)$ to get the equation for $\rho_X(k) = \gamma(k)/\gamma(0)$:
(Recall: $(\rho_X(0) = 1, \rho_X(-k) = \rho_X(k)\ )$

- $\rho_X(k) = \phi_1 \rho_X(k-1) + \phi_2 \rho_X(k-2) + \phi_3 \rho_X(k-3) + \ldots + \phi_p \rho_X(k-p)$, that is,

For $k = 1:\quad \rho_X(1) = \phi_1 + \phi_2 \rho_X(1) + \phi_3 \rho_X(2) + \ldots + \phi_p \rho_X(p-1)$;

For $k = 2:\quad \rho_X(2) = \phi_1 \rho_X(1) + \phi_2 + \phi_3 \rho_X(1) + \ldots + \phi_p \rho_X(p-2)$;

$$\ldots$$

For $k = p,\quad \rho_X(p) = \phi_1 \rho_X(p-1) + \phi_2 \rho_X(p-2) + \phi_3 \rho_X(p-3) + \ldots + \phi_p$.

For $k = 0$, substitute $\gamma_X(0) = \sigma_X^2, \gamma_X(k) = \rho_X(k)\gamma_X(0) \equiv \rho_X(k)\sigma_X^2$, into $(*)$ to get:

$$\sigma_X^2 = \phi_1 \rho_X(1)\sigma_X^2 + \ldots + \phi_p \rho_X(p)\sigma_X^2 + \sigma_Z^2.$$

Therefore,

$$\sigma_X^2 = \frac{\sigma_Z^2}{1 - \phi_1 \rho_X(1) - \ldots - \phi_p \rho_X(p)}$$

Note: depending on coefficients, the ACF may have exponential decay, oscillations, sinusoidal behavior, combination of these.

Example 4.3.3 AR(2): For $p = 2$, the system of Yule-Walker equations is:

$\rho_X(1) = \phi_1 + \phi_2 \rho_X(1)$;
$\rho_X(2) = \phi_1 \rho_X(1) + \phi_2$.

From equation 1, find $\rho_X(1) = \frac{\phi_1}{1-\phi_2}$,

From equation 2, find $\rho_X(2) = \frac{\phi_1^2}{1-\phi_2} + \phi_2 = \frac{\phi_1^2 + \phi_2(1-\phi_2)}{1-\phi_2}$.

....

5. Mixed Autoregressive-Moving Average Models ARMA(p, q)

5.1 ARMA(1, 1) (based on §2.3 of [BD]):

$$X_t - \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1}, \ \ |\phi_1| < 1, \ \ |\theta_1| < 1.$$

- **AR(1) and MA(1) are special cases of ARMA(1,1).**
- **When $|\theta_1|$ is close to zero, ARMA(1,1) behaves similar to the behavior of AR(1).**
- **When $|\phi_1|$ is close to zero, it behaves similar to MA(1).**

ARMA(1,1) is invertible if $|\theta_1| < 1$.

Derivation: as in MA(1) model, write

$(1 - \phi_1 B)X_t = (1 + \theta_1 B)Z_t;\ Z_t = (1 + \theta_1 B)^{-1}(1 - \phi_1 B)X_t$.

The last operator represents a convergent series for $|\theta_1| < 1$. $\square$.

For $|\phi_1| < 1$ AR(1,1) is stationary, causal, and can be represented as MA(∞):
$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, **with** $\psi_0 = 1$, $\psi_j = \phi_1^j + \theta_1 \phi_1^{j-1} = \phi_1^{j-1}(\phi_1 + \theta_1)$.

Derivation: From $(1 - \phi_1 B)X_t = (1 + \theta_1 B)Z_t$ and $|\phi_1| < 1$ follows:

$$\begin{aligned} X_t &= (1 - \phi_1 B)^{-1}(1 + \theta_1 B)Z_t = (1 + \phi_1 B + \phi_1^2 B^2 + \ldots)(1 + \theta_1 B)Z_t \\ &= (1 + (\phi_1 + \theta_1)B + (\phi_1^2 + \theta_1 \phi_1)B^2 + \ldots + (\phi_1^j + \theta_1 \phi_1^{j-1})B^j + \ldots)Z_t \\ &= Z_t + (\phi_1 + \theta_1)Z_{t-1} + (\phi_1^2 + \theta_1 \phi_1)Z_{t-2} + \ldots + (\phi_1^j + \theta_1 \phi_1^{j-1})Z_{t-j} + \ldots. \end{aligned}$$

ACF of ARMA(1,1): $\rho_X(k) = \frac{(\phi_1+\theta_1)(1+\phi_1\theta_1)}{(1+2\phi_1\theta_1+\theta_1^2)} \phi_1^{k-1}, k \geq 1$.

Behavior of ACF of ARMA(1,1): a spike at lag 1 (MA(1) part) and then exponential decay as in AR(1).

Derivation: Use MA(∞) representation to write:

$\gamma_X(k) = E[X_t X_{t+k}] = E\Big[\Big(\sum_{j=0}^{\infty}\psi_j Z_{t-j}\Big)\Big(\sum_{m=0}^{\infty}\psi_m Z_{t+k-m}\Big)\Big] = \sigma_Z^2 \sum_{j=0}^{\infty}\psi_j\psi_{j+k}$

b/c $EZ_{t-j}Z_{t+k-m} = \sigma_Z^2$ iff $t-j = t+k-m$ or $m = k+j$. Substitute ψ_j's to get the acf (or check Example 3.2.1 on p. 78 of [BD].)

. . . .

5.2 ARMA(p, q)(§3.1 of [BD]): With $\boldsymbol{\{Z_t\} \sim WN(0, \sigma_Z^2)}$

$$\boldsymbol{X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},} .$$

ARMA(p,q) is stationary if $\phi(z) \equiv 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0$ for all $|z| \leq 1$.

ARMA(p,q) is invertible if $\theta(z) \equiv 1 + \theta_1 z + \ldots + \theta_q z^q \neq 0$ for all $|z| \leq 1$.

Notion of Parsimony (economy of coefficients):

All stationary and invertible ARMA models of finite order can be represented as MA(∞) and as AR(∞). The presentation with finite number of coefficients is more economic and is called parsimonious representation.}

. . . .

6. PACF-Partial Autocorrelation Function ([BD], §3.2.3)

Problem: The property used to identify MA(q) process is that $\rho_X(k) = 0, k > q$. For AR(p) process, the form of the ACF is similar (exponential decay or oscillations) for different p. Thus, we have a problem of how to identify p. We thus define another function – PACF.

Notation: $\alpha(0) = 1, \alpha(n) = \phi_{nn}$ = n-th partial autocorrelation.

Result: For AR(p) process $\alpha(n) = \phi_{nn} = 0$, for all $n > p$.

Thus, PACF plays a major role in identification of the AR model and its order.

Definition of the PACF and derivation of the Result.

Assume that we have a data sample and we try to match it to AR(k) model, where k is unknown, so it can be $k = 1$ or $k = 2$, or $k = 3$, ..., $k = n$ for some n. This means that we write n hypothetical AR models:

$X_t = \phi_{11} X_{t-1} + Z_t$

$X_t = \phi_{21} X_{t-1} + \phi_{22} X_{t-2} + Z_t$

$\ldots$

$X_t = \phi_{n1} X_{t-1} + \phi_{n2} X_{t-2} + \ldots + \phi_{nn} X_{t-n} + Z_t$

$\ldots$

and try to choose one model which fits our sample the best.

Recall that for each model we can write the Yule-Walker equations (4.3.1):

$\rho_X(h) = \phi_{n1}\rho_X(h-1) + \phi_{n2}\rho_X(h-2) + \ldots + \phi_{nn}\rho_X(h-n)$, $h = 1, 2, \ldots, n$
or

$\phi_{n1}\rho_X(h-1) + \phi_{n2}\rho_X(h-2) + \ldots + \phi_{nn}\rho_X(h-n) = \rho_X(h)$, $h = 1, 2, \ldots, n$,

so that for AR(1) model $\phi_{12} = \phi_{13} = \ldots = 0$; for AR(2) model $\phi_{23} = \phi_{24} = \ldots = 0$;... for AR($n$) model $\phi_{n,n+1} = \ldots = 0$, i.e. each equation ends at $\phi_{11}, \phi_{22}, \ldots, \phi_{nn}$.
We denote $\alpha_n = \phi_{nn}$ and call it PACF at lag n.

To find PACF:

- If $n = 1$, the model is $X_t = \phi_{11}X_{t-1} + Z_t$ and the Y-W equations are (recall that $\rho_X(0) = 1$):

$\alpha(1) = \phi_{11} = \rho_X(1)$.

- If $n = 2$, the model is $X_t = \phi_{21}X_{t-1} + \phi_{22}X_{t-2} + Z_t$ and the Y-W equations are:
$\phi_{21}\rho_X(h-1) + \phi_{22}\rho_X(h-2) = \rho_X(h)$ that is a system of equations

$h = 1$: $\phi_{21} + \phi_{22}\rho_X(1) = \rho_X(1)$;

$h = 2$: $\phi_{21}\rho_X(1) + \phi_{22} = \rho_X(2)$. Solution: $\alpha(2) = \phi_{22} = \dfrac{\begin{vmatrix} 1 & \rho_X(1) \\ \rho_X(1) & \rho_X(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho_X(1) \\ \rho_X(1) & 1 \end{vmatrix}} = \frac{\rho_X(2) - (\rho_X(1))^2}{1 - (\rho_X(1))^2}$

- If $n = 3$ we will have three Y-W equations to find $\alpha(3) = \phi_{33}$, etc.

- For general n we have a system of n equations:

$h = 1 : \phi_{n1} + \phi_{n2}\rho_X(1) + \ldots + \phi_{nn}\rho_X(n-1) = \rho_X(1)$
$h = 2 : \phi_{n1}\rho_X(1) + \phi_{n2} + \ldots + \phi_{nn}\rho_X(n-2) = \rho_X(2)$
$\ldots$
$h = n : \phi_{n1}\rho_X(n-1) + \phi_{n2}\rho_X(n-2) + \ldots + \phi_{nn} = \rho_X(n)$.

In matrix form: $R_n \underline{\phi}_n = \underline{\rho}_n$ where

$$\underline{\rho}_n = \begin{pmatrix} \rho_X(1) \\ \rho_X(2) \\ \cdot \\ \cdot \\ \rho_X(n) \end{pmatrix}, \quad \underline{\phi}_n = \begin{pmatrix} \phi_{n1} \\ \phi_{n2} \\ \cdot \\ \cdot \\ \phi_{nn} \end{pmatrix}, \quad R_n = \begin{pmatrix} 1 & \rho_X(1) & \ldots & \rho_X(n-1) \\ \rho_X(1) & 1 & \ldots & \rho_X(n-2) \\ \ldots & \ldots & \ldots & \ldots \\ \rho_X(n-1) & \rho_X(n-2) & \ldots & 1 \end{pmatrix}.$$

For each n we solve the system to recover $\alpha(n) := \phi_{nn}$, the nth PACF, $n = 1, 2, \ldots$.

If $\alpha(n) \equiv \phi_{nn} = 0$ for all $n \geq p+1$ but $\alpha(p) \neq 0$, then the order of the AR is p.

For MA(q) process PACF is dominated by combination of damped exponentials and/or damped sine waves; see MA(1) example below. We notice a duality between AR and MA processes.

6.1 PACF for MA(1):

$\boldsymbol{\alpha(1) = \phi_{11} = \rho_X(1) = \frac{\theta_1}{1+\theta_1^2}}$,
$\boldsymbol{\alpha(2) = \phi_{22} = \frac{\rho_X(2) - (\rho_X(1))^2}{1 - (\rho_X(1))^2} \equiv \frac{-(\rho_X(1))^2}{1 - (\rho_X(1))^2} \equiv \frac{-\theta_1^2}{1+\theta_1^2+\theta^4}}$, $\ldots$ $\boldsymbol{\alpha(k) = \phi_{kk} = \frac{(-1)^{k+1}\theta_1^k}{(1+\theta_1^2+\ldots+\theta_1^{2k})}, k > 0}$.

....

7. Non-Stationary TS. Differencing. (Based on [BD], SS1.5.1.1, 1.5.2 on pp. 25-29)

7.1 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 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.

Definition. Denote by $\nabla^d X_t$ the dth difference of X_t for all t, and define:

$\nabla X_t \equiv \nabla^1 X_t := X_t - X_{t-1}$. We call $\nabla X_t = X_t - X_{t-1}$ "the **differenced** process".
Note: $\nabla X_t = (1-B)X_t$.

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

Note: $\nabla^2 X_t$ can also be derived as follows:

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

Example 7.1.1: $X_t = bt + S_t$, where S_t is a stationary process. (**Linear trend**).

Then for $W_t = \nabla X_t = X_t - X_{t-1}$ 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 7.1.2: $X_t = bt^2 + S_t$, where S_t is a stationary process. (**Quadratic trend**).
Take

$$\begin{aligned} W_t = \nabla^2 X_t &= X_t - 2X_{t-1} + X_{t-2} \\ &= b[t^2 - 2(t-1)^2 + (t-2)^2] + [S_t - 2S_{t-1} + S_{t-2}] \\ &= 2b + (S_t - 2S_{t-1} + S_{t-2}) = \text{stationary time series.} \end{aligned}$$

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 7.1.3. The Random Walk Model. (see 2.1.5 of LNs; ([BD], pp. 7, 14)

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

$X_t = X_{t-1} + Z_t$, $\{Z_t\} \sim WN(0, \sigma_Z^2)$ or

$X_1 = Z_1$, $X_2 = Z_1 + Z_2, \ldots, X_t = Z_1 + \ldots + Z_t = \sum_{j=0}^{t-1} Z_{t-j}$, $t = 1, 2, \ldots$.

Note: $EX_t = 0$, $\gamma_X(t,s) = EX_sX_t = \sigma_Z^2 \min(s,t)$ (calculated in 2.1.5 of Week 1), that is, the Random Walk X_t is non-stationary.

The differenced process $W_t = X_t - X_{t-1} \equiv Z_t$ is stationary with ACF $\rho_W(k) = 0, k > 0$.

. . . .

7.2 ARIMA(p, d, q):

Definition. A non-stationary times series $\{X_t\}$ follows **ARIMA(p,d,q) model** if $W_t = \nabla^d X_t = (1-B)^d X_t$, produced by differencing $\{X_t\}$ d times, is a stationary ARMA(p,q).

Model for ARIMA(p,d,q) process:

$\rightarrow$ Let $W_t = \nabla^d X_t$ be a stationary ARMA(p,q) model: $\phi(B)W_t = \theta(B)Z_t$, or
$W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} - \ldots - \phi_p W_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}$, $\phi(z) \neq 0$, for all $|z| \leq 1$.

$\rightarrow$ then the model for X is $\phi(B)(1-B)^d X_t = \theta(B)Z_t$ or $\phi^*(B)X_t = \theta(B)Z_t$ with $\phi^*(z) = \phi(z)(1-z)^d$.

. . . .

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

7.3.1 Definition. The Classical Decomposition Model assumes that the data is a realization of the process of the type

$$(*) \qquad X_t = m_t + s_t + S_t,$$

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

. . . .

7.3.2. Conversion of seasonal data to a stationary process via differencing at lag s:

Apply **lag s difference: $W_t := \nabla_s X_t := X_t - X_{t-s}$** to model $X_t = m_t + s_t + S_t$ (∗).

Then $W_t = (m_t - m_{t-s})$(trend) + $(S_t - S_{t-s})$(stationary).

Lag s difference $\nabla_s X_t = (1 - B^s)X_t$ removes periodicity (seasonality) with period s. Usually, one needs to additionally difference to remove the trend.

. . . .

Short Summary:

(i) The Classical Decomposition Model: It assumes that the data is a realization of the process of the type

$$(*) \quad \boldsymbol{X_t = m_t + s_t + S_t},$$

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

(ii) Differencing at lag s and an operator ∇ (nabla):
lag s difference : $W_t := \nabla_s X_t := X_t - X_{t-s} = (1 - B^s)X_t$.

lag 1 difference: $W_t := \nabla X_t \equiv \nabla^1 X_t := X_t - X_{t-1} = (1 - B)X_t$; W_t is the differenced process.

dth difference at lag 1: $W_t := \nabla^d X_t = (1 - B)^d X_t$ the dth difference of X_t at lag 1.

We showed (Examples 7.1.1, 7.1.2 and 7.2.2) that

• If the trend m_t is polynomial of order d, then $W_t = \nabla^d X_t$ is stationary (differencing d times eliminates polynomial trend of order d).

• If the data is seasonal with period s, then difference at lag s, $W_t = \nabla_s X_t = (1 - B^s)X_t$, is no longer periodic (seasonal) (differencing at lag s eliminates seasonality with period s.)

• We say that a non-stationary times series $\{X_t\}$ follows **ARIMA(p,d,q) model** if $W_t = \nabla^d X_t$, produced by differencing $\{X_t\}$ d times, is a stationary ARMA(p,q).

(iii) Model for ARIMA(p,d,q) process:

$$\boldsymbol{\phi(B)(1 - B)^d X_t = \theta(B) Z_t \text{ or } \phi^*(B) X_t = \theta(B) Z_t \text{ with } \phi^*(z) = \phi(z)(1 - z)^d.}$$

Then, $W_t = \nabla^d X_t \equiv (1 - B)^d X_t$ follows stationary ARMA(p,q) model:

$$\boldsymbol{\phi(B) W_t = \theta(B) Z_t, \text{ with } \phi(z) \neq 0,\ \theta(z) \neq 0 \text{ for all } |z| \leq 1.}$$

Note: X is nonstationary and the polynomial $\phi^*(z)$ has unit root $z^* = 1$ of order d.

....

7.4 Discussion of ARIMA, effects of differencing, unit roots.

(7.4.1) ARIMA (p,d,q) models are useful for representing data with a trend.

(7.4.2) Typical feature: slowly decaying ACF.

(7.4.3) Unit root in AR part:

For ARIMA process $\{X_t\}$: $\phi^*(B)X_t = \theta(B)Z_t$ with $\phi^*(z) = (1 - z)\phi(z)$, **differencing at lag 1 eliminates the unit root in AR part producing stationary ARMA**:
$W_t = X_t - X_{t-1}$ follows the model: $\phi(B)W_t = \theta(B)Z_t$, $\phi(z) \neq 0$ for $|z| \leq 1$.

(7.4.4) Unit root in MA part: if X_t is described by a model $\phi(B)X_t = \theta^*(B)Z_t$ with $\theta^*(z) = (1 - z)\theta(z)$, then the series was overdifferenced:

let Y_t be an invertible ARMA process such that $\phi(B)Y_t = \theta(B)Z_t$. Let $X_t = \nabla Y_t = Y_t - Y_{t-1}$. Then $\phi(B)X_t = \phi(B)(1 - B)Y_t = (1 - B)\theta(B)Z_t$, i.e.,
differencing invertible ARMA produces a unit root in MA part of the model.

Conclude: if estimates of ϕ and θ have roots close to unit roots, check your differencing.

(7.4.5) Effect of overdifferencing on the variance and ACF:

Consider an MA(1) process: $X_t = (1 + \theta B)Z_t, |\theta_1| < 1$, with

ACF $\rho_X(1) = \frac{\theta}{1+\theta^2} \neq 0$, $\rho_X(k) = 0$, $k \geq 2$ and variance $\gamma_X(0) = (1 + \theta^2)\sigma_Z^2$.

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

$\rightarrow W_t = (1 - B)(1 + \theta B)Z_t \equiv (1 + (\theta - 1)B - \theta B^2)Z_t$.

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

Also, the variance of the differenced process W_t is $\gamma_W(0) = (1 + (1 - \theta)^2 + \theta^2)\sigma_Z^2$. Thus,

$\rightarrow \gamma_W(0) - \gamma_X(0) = (1 - \theta)^2\sigma_Z^2 > 0$, 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.

Conclude: if additional differencing increases variance, it is unnecessary.

. . . .

8. Log Transformations of Data. Box-Cox Transformations. ([BD], §6.2 (a))

Plotting the data may suggest that it is sensible to transform them. Main reason for making transformations:

(i) **To stabilize variance or/and seasonal effect**: when there is a trend and the variance or/and the size of the seasonal effect seems to increase with the mean. Consider logarithmic transformation.

(ii) **To make data normally distributed.** Classical time series model building and forecasting are carried out on the assumption that the data is normally distributed. If it is not the case, one may try transforming data.

The Box-Cox transformation: (U_t here is the original data)

$$f_\lambda(U_t) = \begin{cases} \ln U_t, & \text{if } U_t > 0,\ \lambda = 0; \\ \lambda^{-1}(U_t^\lambda - 1), & \text{if } U_t \geq 0,\ \lambda \neq 0 \end{cases}$$

Most often: $\lambda = 0$, $\lambda = 0.5$ which correspond to $Y_t = \ln U_t$ and $Y_t = \sqrt{U_t}$.

In general, for a family of transformations f_λ we try to find a value λ_0 of λ, such that $Y_t(\lambda_0) = f_{\lambda_0}(V_t)$ is a Gaussian process which can be modeled by ARIMA process. There are maximum likelihood techniques to find the value of λ_0.

Problem with this approach: usually we get Y_t Gaussian for each t, but NOT a Gaussian process. This is why non-Gaussian and non-linear models are important.

NOTES: (a) λ can be negative or positive. R routine boxcox {MASS} plots in (-2,2) interval.
(b) f_λ is defined to be continuous at $\lambda = 0$ when $f_0 = \ln U_t$. To see that, as $\lambda \rightarrow 0$, write:

$$\frac{x^\lambda - 1}{\lambda} = \frac{e^{\lambda \log x} - 1}{\lambda} \approx \frac{(1 + \lambda \log x + (\lambda \log x)^2/2 + \ldots) - 1}{\lambda} \rightarrow \log x.$$