Introduction to ARMA processes

Fulvio Corsi

SNS Pisa
Definitions

- Stochastic process (univariate): sequence of random variables \( \{Y_t; \ t \in \mathbb{N}; \text{or} \ t \in \mathbb{Z}\} \)

- Second order process: \( \mathbb{E}[Y_t^2] < +\infty \quad \forall t \)

- Mean \( \mu_t = \mathbb{E}[Y_t] \)

- Variance \( \sigma_t^2 = \mathbb{E}(Y_t - \mu_t)^2 \)

- Autocovariance \( \gamma_t(k) \equiv \text{Cov}(Y_t, Y_{t-k}) = \mathbb{E}(Y_t - \mu_t)(Y_t - k - \mu_t - k) \) hence \( \sigma_t^2 \equiv \gamma_t(0) \)

- Autocorrelation
  \[
  \rho_t(k) \equiv \text{Corr}(Y_t, Y_{t-k}) = \frac{\gamma_t(k)}{\sqrt{\gamma_t(0)\gamma_{t-k}(0)}}, \quad -1 \leq \rho_t(k) \leq 1
  \]

- Partial Autocorrelation \( a_t(k) \equiv \text{Corr}(Y_t, Y_{t-k}|Y_{t-1}, \ldots, Y_{t-k+1}) \)
Stationarity and Ergodicity

- **Strict stationarity:**
  \[(Y_1, Y_2, \ldots, Y_n) \overset{d}{=} (Y_{1+k}, Y_{2+k}, \ldots, Y_{n+k})\]
  for any integer \(n > 1, k\)

- **Weak/second-order/covariance stationarity:**
  \begin{align*}
  &\mathbb{E}[Y_t] = \mu \\
  &\mathbb{E}[(Y_t - \mu)^2] = \sigma^2 < +\infty \text{ (i.e. constant and independent of } t) \\
  &\mathbb{E}[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma(|k|) \text{ (i.e. independent of } t \text{ for each } k) \\
  \Rightarrow &\quad \rho(k) = \frac{\gamma(k)}{\gamma(0)}.
  \end{align*}

- **Ergodicity**
  \begin{align*}
  &\text{Ergodic in mean: } \bar{y} \equiv \frac{1}{T} \sum_{t=1}^{T} Y_t \xrightarrow{p} \mathbb{E}[Y_t] \\
  &\text{Ergodic in second moments: } \frac{1}{T} \sum_{t=1}^{T} (Y_t - \mu)(Y_{t-k} - \mu) \xrightarrow{p} \gamma(k)
  \end{align*}

- **Interpretation:**
  \begin{itemize}
  \item unconditional mean and variance are constant
  \item mean reversion
  \item shocks are transient
  \item covariance between \(Y_t\) and \(Y_{t-k}\) tends to 0 as \(k \to \infty\)
With the noise

- weak (uncorrelated)
  - $\mathbb{E}(\epsilon_t) = 0 \quad \forall t$
  - $V(\epsilon_t) = \sigma^2 \quad \forall t$
  - $\text{Corr}(\epsilon_t, \epsilon_s) = 0 \quad \forall s \neq t$

- strong (independence)
  - $\epsilon_t \sim I.I.D.(0, \sigma^2)$

- Gaussian (weak=strong)
  - $\epsilon_t \sim N.I.D.(0, \sigma^2)$
the Lag operator is defined as:

\[ LY_t \equiv Y_{t-1} \]

is a linear operator:

\[
L(\beta Y_t) = \beta \cdot LY_t = \beta Y_{t-1} \\
L(X_t + Y_t) = LX_t + LY_t = X_{t-1} + Y_{t-1}
\]

and admits power exponent, for instance:

\[
L^2 Y_t = L(LY_t) = LY_{t-1} = Y_{t-2} \\
L^k Y_t = Y_{t-k} \\
L^{-1} Y_t = Y_{t+1}
\]

Some examples:

\[
\Delta Y_t = Y_t - Y_{t-1} = Y_t - LY_t = (1 - L)Y_t \\
y_t = (\theta_1 + \theta_2 L)LY_t = (\theta_1 L + \theta_2 L^2)Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2}
\]

Expression like

\[
(\theta_0 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_n L^n)
\]

with possibly \( n = \infty \), are called lag polynomial and are indicated as \( \theta(L) \)
Moving Average (MA) process

The simplest way to construct a stationary process is to use a lag polynomial \( \theta(L) \) with \( \sum_{j=0}^{\infty} \theta_j^2 < \infty \) to construct a sort of “weighted moving average” of white noises \( \epsilon_t \), i.e.

- **MA(q)**
  \[
  Y_t = \theta(L)\epsilon_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}
  \]

- **Example, MA(1)**
  \[
  Y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t
  \]
  being \( \mathbb{E}Y_t = 0 \)
  \[
  \begin{align*}
  \gamma(0) & = \mathbb{E}Y_tY_t = \mathbb{E}(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_t + \theta\epsilon_{t-1}) = \sigma^2(1 + \theta^2); \\
  \gamma(1) & = \mathbb{E}Y_tY_{t-1} = \mathbb{E}(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2}) = \sigma^2\theta; \\
  \gamma(k) & = \mathbb{E}Y_tY_{t-k} = \mathbb{E}(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-k} + \theta\epsilon_{t-k-1}) = 0 \quad \forall k > 1
  \end{align*}
  \]
  and,
  \[
  \begin{align*}
  \rho(1) & = \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1 + \theta^2} \\
  \rho(k) & = \frac{\gamma(k)}{\gamma(0)} = 0 \quad \forall k > 1
  \end{align*}
  \]
  hence, while a white noise is “0-correlated”, MA(1) is **1-correlated** (i.e. it has only the first correlation \( \rho(1) \) different from zero)
Properties MA(q)

In general for a \textbf{MA}(q) process

\[
Y_t = \theta(L)\epsilon_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \ldots + \theta_q\epsilon_{t-q}
\]

we have

\[
\gamma(0) = \sigma^2 \left( 1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2 \right)
\]

\[
\gamma(k) = \sigma^2 \sum_{j=0}^{q-k} \theta_j\theta_{j+k} \quad \forall k \leq q
\]

\[
\gamma(k) = 0 \quad \forall k > q
\]

and

\[
\rho(k) = \frac{\sum_{j=0}^{q-k} \theta_j\theta_{j+k}}{1 + \sum_{j=1}^{q} \theta_j^2} \quad \forall k \leq q
\]

\[
\rho(k) = 0 \quad \forall k > q
\]

Hence, an MA(q) is \textbf{q-correlated} and it can also be shown that any stationary q-correlated process can be represented as an MA(q).

But, given a q-correlated process, is the MA(q) process unique? In general no, indeed it can be shown that for a q-correlated process there are \(2^q\) possible MA(q) with same autocovariance structure. However, there is only one MA(q) which is \textbf{invertible}. 
Invertibility conditions for MA

- first consider the MA(1) case:
  \[ Y_t = (1 + \theta L)\epsilon_t \]
  given the result
  \[ (1 + \theta L)^{-1} = (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \theta^4 L^4 + \ldots) = \sum_{i=0}^{\infty} (-\theta L)^i \]
  inverting the \( \theta(L) \) lag polynomial, we can write
  \[ (1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \theta^4 L^4 + \ldots)Y_t = \epsilon_t \]
  which can be considered an AR(\( \infty \)) process.

  If an MA process can be written as an AR(\( \infty \)) of this type, such MA representation is said to be **invertible**. For MA(1) process the invertibility condition is given by \( |\theta| < 1 \).

- For a general MA(q) process
  \[ Y_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\epsilon_t \]
  the invertibility conditions are that the roots of the lag polynomial
  \[ 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q = 0 \]
  lie outside the unit circle. Then the MA(q) can be written as an AR(\( \infty \)) by inverting \( \theta(L) \).

- Invertibility also has important practical consequence in application. In fact, given that the \( \epsilon_t \) are **not observable** they have to be reconstructed from the observed \( Y \)'s through the AR(\( \infty \)) representation.
suppose you want to invert the generic lag polynomial

\[
\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q
\]

i.e. finding the series \( \theta(L)^{-1} = \varphi_0 + \varphi_1 L + \varphi_2 L^2 + ... \) such that

\[
(1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)(\varphi_0 + \varphi_1 L + \varphi_2 L^2 + ...) = 1
\]

by matching the coefficients of \( L^0, L^1, ..., L^i \) in both sides we can obtain the \( \varphi_i \) recursively

\[
\begin{align*}
\varphi_0 &= 1 \\
\varphi_1 + \varphi_0 \theta_1 &= 0 \quad \Rightarrow \varphi_1 = -\theta_1 \\
\varphi_2 + \varphi_1 \theta_1 + \varphi_0 \theta_2 &= 0 \quad \Rightarrow \varphi_2 = \theta_1^2 - \theta_2 \\
&\vdots \quad \Rightarrow \varphi_i = \theta_1^{i-1} - \theta_i \\
\varphi_i + \varphi_{i-1} \theta_1 + ... + \varphi_0 \theta_i &= 0 \quad \Rightarrow \varphi_i = \theta_1^{i-1} - \theta_i
\end{align*}
\]
A general AR process is defined as

$$\phi(L)Y_t = \epsilon_t$$

It is always invertible but not always stationary.

Example: AR(1)

$$(1 - \phi L)Y_t = \epsilon_t \quad \text{or} \quad Y_t = \phi Y_{t-1} + \epsilon_t$$

by inverting the lag polynomial $(1 - \phi L)$ the AR(1) can be written as

$$Y_t = (1 - \phi L)^{-1} \epsilon_t = \sum_{i=0}^{\infty} (\phi L)^i \epsilon_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} = MA(\infty)$$

hence the stationarity condition is that $|\phi| < 1$.

From this representation we can apply the general formula of MA to compute $\gamma(\cdot)$ and $\rho(\cdot)$. In particular,

$$\rho(k) = |\phi|^k \quad \forall k$$

i.e. monotonic exponential decay for $\phi > 0$ and exponentially damped oscillatory decay for $\phi < 0$.

In general an AR($p$) process

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

is stationarity if all the roots of the characteristic equation of the lag polynomial

$$1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$

are outside the unit circle.
Example: AR(2)

\[(1 - \phi_1 L - \phi_2 L^2) Y_t = \epsilon_t \quad \text{or} \quad Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t\]

the roots of the quadratic characteristic equation exceed 1 in absolute value if three conditions are satisfied:

\[\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1 \quad \text{and} \quad |\phi_2| > 1\]
State Space Representation of AR(p)

to gain more intuition on the AR stationarity conditions write an AR(p) in its state space form

\[
\begin{bmatrix}
Y_t \\
Y_{t-1} \\
\vdots \\
Y_{t-p+1}
\end{bmatrix} =
\begin{bmatrix}
\phi_1 & \phi_2 & \phi_3 & \cdots & \phi_{p-1} & \phi_p \\
1 & 0 & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 1 & 0 \\
0 & 0 & 0 & \cdots & 1 & 0 \\
\end{bmatrix}
\begin{bmatrix}
Y_{t-1} \\
Y_{t-2} \\
\vdots \\
Y_{t-p}
\end{bmatrix} +
\begin{bmatrix}
\epsilon_t \\
0 \\
\vdots \\
0
\end{bmatrix}
\]

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

Hence, the expected value of \(X_t\) satisfy,

\[\mathbb{E}X_t = F X_{t-1}\]

and \[\mathbb{E}X_{t+j} = F^{j+1} X_{t-1}\]

is a linear map in \(\mathbb{R}^p\) whose dynamic properties are given by the eigenvalues of the matrix \(F\).

The eigenvalues of \(F\) are given by solving the characteristic equation

\[\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \cdots - \phi_{p-1} \lambda - \phi_p = 0.\]

Comparing this with the characteristic equation of the lag polynomial \(\phi(L)\)

\[1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_{p-1} z^{p-1} - \phi_p z^p = 0\]

we can see that the roots of the 2 equations are such that

\[z_1 = \lambda_1^{-1}, \quad z_2 = \lambda_2^{-1}, \quad \ldots, \quad z_p = \lambda_p^{-1}\]
for an AR(p) process, the k–lag ACF $\rho_k$ can be interpreted as simple regression

$$Y_t = \rho_k Y_{t-k} + \text{error},$$

while the k–lag PACF $a_t(k) \equiv \text{Corr}(Y_t, Y_{t-k}|Y_{t-1}, \ldots, Y_{t-k+1})$ can be seen as a multiple regression

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots + a_k Y_{t-k} + \text{error}$$

it can be computed by solving the Yule-Walker system (obtained by multiplying both sides of an AR(p) model by $Y_t, Y_{t-1}, \ldots$, taking expectations, and inverting).

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \ldots & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(k) \end{bmatrix}$$

Importantly, AR(p) processes are “p–partially correlated” ⇒ identification of AR order
An ARMA(p,q) process is defined as

$$\phi(L)Y_t = \theta(L)\epsilon_t$$

where $\phi(L)$ and $\theta(L)$ are $p^{th}$ and $q^{th}$ lag polynomials.

The process is stationary if all the roots of

$$\phi(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_{p-1} z^{p-1} - \phi_p z^p = 0$$

lie outside the unit circle and, hence, admits the MA($\infty$) representation:

$$Y_t = \phi(L)^{-1} \theta(L)\epsilon_t$$

The process is invertible if all the roots of

$$\theta(z) \equiv 1 + \theta_1 z + \theta_2 z^2 + \ldots + \theta_q z^q = 0$$

lie outside the unit circle and, hence, admits the AR($\infty$) representation:

$$\epsilon_t = \theta(L)^{-1} \phi(L)Y_t$$
Estimation of AR models

- In time series the data are usually not i.i.d.
  ⇒ It is then very convenient to use the “prediction–error” decomposition of the likelihood:

\[ L(y_T, y_{T-1}, \ldots, y_1; \theta) = f(y_T|\Omega_{T-1}; \theta) f(y_{T-1}|\Omega_{T-2}; \theta) \ldots f(y_1|\Omega_0; \theta) \]

- For example for the AR(1)

\[ y_t = \phi_1 y_{t-1} + \epsilon_t \]

the full log-Likelihood can be written as

\[ l(\phi) = f_{Y_1}(y_1; \phi) + \sum_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \phi) = f_{Y_1}(y_1; \phi) - \frac{T}{2} \log(2\pi) - \sum_{t=1}^{T} \log \sigma^2 - \frac{1}{2} \sum_{t=2}^{T} \frac{(y_t - \phi y_{t-1})^2}{\sigma^2} \]

Hence, maximizing the conditional likelihood for \( \phi \) is equivalent to minimize

\[ \sum_{t=2}^{T} (y_t - \phi y_{t-1})^2 \]

which is the OLS criteria.

- In general for AR(p) process OLS are consistent and, under gaussianity, asymptotically equivalent to MLE ⇒ asymptotically efficient
Estimation of MA models

For example, for the MA(1)

\[ y_t = \theta \epsilon_{t-1} + \epsilon_t \]

the full log-Likelihood can be written as

\[ l(\phi) = \underbrace{f_Y(y_1; \phi)}_{\text{marginal 1st obs}} + \sum_{t=2}^{T} \underbrace{f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \phi)}_{\text{conditional likelihood}} = f_Y(y_1; \phi) - \frac{T}{2} \log(2\pi) - \sum_{t=1}^{T} \log \sigma^2 - \frac{1}{2} \sum_{t=2}^{T} \frac{(y_t - \theta \epsilon_{t-1})^2}{\sigma^2} \]

However, now the \( \epsilon \) are not observed, I can only observe \( y \). Hence, we have to recover \( \epsilon \) from \( y \) by

\[ \epsilon_t = y_t - \theta \epsilon_{t-1} = (-\theta)^t \epsilon_0 + \sum_{i=1}^{t} (-\theta)^i y_{t-i} \]

as long as the MA is invertible.

So now the minimization of RSS is highly non-linear in \( \theta \) \( \Rightarrow \) MLE or Non-linear Least Square.
For a general \textbf{ARMA}(p, q)

\[ Y_t = \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \]

\( Y_{t-1} \) is correlated with \( \epsilon_{t-1}, \ldots, \epsilon_{t-q} \) \( \Rightarrow \mathbb{E}[\epsilon|X] \neq 0 \Rightarrow \text{OLS not consistent.} \)

\( \rightarrow \text{MLE with numerical optimization procedures.} \)
Optimal Prediction

if the Loss Function of a prediction is a quadratic function of the prediction error i.e. the Mean Square Error (MSE)

\[ \text{MSE}(\hat{Y}_t) \equiv \mathbb{E}(Y_t - \hat{Y}_t)^2 \]

then the optimal prediction of \( Y \) in terms of past values \( X \) is given by the conditional expectation \( \mathbb{E}(Y_{t+1} | X_t) \).

Proof:

\[
\mathbb{E}[Y_{t+1} - g(X_t)]^2 = \mathbb{E}[Y_{t+1} - \mathbb{E}(Y_{t+1} | X_t) + \mathbb{E}(Y_{t+1} | X_t) - g(X_t)]^2 \\
= \mathbb{E}[Y_{t+1} - \mathbb{E}(Y_{t+1} | X_t)]^2 \\
+ \mathbb{E}[(\mathbb{E}(Y_{t+1} | X_t) - g(X_t)]^2 \\
+ 2\mathbb{E}\{|Y_{t+1} - \mathbb{E}(Y_{t+1} | X_t)|[\mathbb{E}(Y_{t+1} | X_t) - g(X_t)]\} \\
= 0 \\
\geq \mathbb{E}[Y_{t+1} - \mathbb{E}(Y_{t+1} | X_t)]^2
\]

If the process \( Y \) is linear or normally distributed the linear projection \( \hat{Y}_t \equiv P(Y_{t+1} | X_t) = \alpha'X \) is the optimal prediction (i.e. the one minimizing the MSE) and \( \alpha' = [\mathbb{E}(X_tX_t')]^{-1}\mathbb{E}(X_t'Y_{t+1}) \approx OLS \)
Prediction with ARMA models: AR(1) example

with \( \mathbb{E}_t(Y_{t+1}) \equiv \mathbb{E}(Y_{t+1}|Y_t, Y_{t-1}, \ldots, \epsilon_t, \epsilon_{t-1}, \ldots) \) and \( \text{Var}_t(Y_{t+1}) \equiv \text{Var}(Y_{t+1}|Y_t, Y_{t-1}, \ldots, \epsilon_t, \epsilon_{t-1}, \ldots) \)

For the AR(1): \( Y_t = \phi Y_{t-1} + \epsilon_t \) we have

\[
\begin{align*}
\mathbb{E}_t(Y_{t+1}) &= \mathbb{E}_t(\phi Y_t + \epsilon_{t+1}) = \phi Y_t \\
\mathbb{E}_t(Y_{t+2}) &= \mathbb{E}_t(\phi^2 Y_t + \phi \epsilon_{t+1} + \epsilon_{t+2}) = \phi^2 Y_t \\
\mathbb{E}_t(Y_{t+k}) &= \ldots = \phi^k Y_t
\end{align*}
\]

with

\[
\begin{align*}
\text{Var}_t(Y_{t+1}) &= \text{Var}_t(\phi Y_t + \epsilon_{t+1}) = \sigma^2 \\
\text{Var}_t(Y_{t+2}) &= \text{Var}_t(\phi^2 Y_t + \phi \epsilon_{t+1} + \epsilon_{t+2}) = (1 + \phi^2)\sigma^2 \\
\text{Var}_t(Y_{t+k}) &= \ldots = (1 + \phi^2 + \phi^4 + \ldots + \phi^{2(k-1)})\sigma^2
\end{align*}
\]

Notice that

\[
\lim_{k \to \infty} \mathbb{E}_t(Y_{t+k}) = 0 = \mathbb{E}(Y_t)
\]

\[
\lim_{k \to \infty} \text{Var}_t(Y_{t+k}) = \sum_{j=0}^{\infty} \phi^{2j} \sigma^2 = \frac{\sigma^2}{1 - \phi^2} = \text{Var}(Y_t)
\]

In general

\[
Y_{t+k} = \underbrace{\{\text{function of future values}\}}_{\text{determine } \text{Var}_t(Y_{t+k})} + \underbrace{\{\text{function of past values}\}}_{\text{determine } \mathbb{E}_t(Y_{t+k})}
\]
write the model in its AR(\(\infty\)) representation:
\[
\eta(L)(Y_t - \mu) = \epsilon_t
\]

then the optimal prediction of \(Y_{t+s}\) is given by
\[
E[Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + \left[ \frac{\eta(L)^{-1}}{L^s} \right] + \eta(L)(Y_t - \mu) \quad \text{with} \quad \left[ L^k \right]_+ = 0 \quad \text{for} \quad k < 0
\]
which is known as Wiener-Kolmogorov prediction formula.

In the case of an AR(p) process the prediction formula can also be written as
\[
E[Y_{t+s}|Y_t, Y_{t-1}, ...] = \mu + f_{11}^{(s)}(Y_t - \mu) + f_{12}^{(s)}(Y_{t-1} - \mu) + ... + f_{1p}^{(s)}(Y_{t-p+1} - \mu)
\]
where \(f_{ij}^{(j)}\) is the element (1, 1) of the matrix \(F^j\) in the state space representation of AR(p).

The easiest way to compute prediction from AR(p) model is, however, through recursive methods.
Wold Theorem: any mean zero covariance stationary process can be represented in the form

\[ Y_t = \sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j} + k_j \]

where

- \( \sum_{j=0}^{\infty} \varphi_j^2 < \infty \) and \( \varphi_0 = 1 \)
- \( \epsilon_t = Y_t - P(Y_t | Y_{t-1}, Y_{t-2}, \ldots) \) are the linear prediction errors
- \( \{ \varphi_j \} \) and \( \{ \epsilon_t \} \) are unique
- \( k_j \) is linearly deterministic
- \( \epsilon_{t-j} \) and \( k_j \) are uncorrelated.
**Box-Jenkins Approach**

- **check for stationarity:** if not try different transformation (ex differentiation → ARIMA models)

- **Identification:**
  - check the autocorrelation (ACF) function: a q-correlated process is an MA(q) model
  - check the partial autocorrelation (PACF) function: a p-partially-correlated process is an AR(p) model

- **Validation:** check the appropriateness of the model by some measure of fit.
  - AIC/Akaike $= T \log \hat{\sigma}_e^2 + 2m$
  - BIC/Schwarz $= T \log \hat{\sigma}_e^2 + m \log T$
    with $\sigma_e^2$ estimation error variance, $m = p + q + 1$ \( n^\circ \) of parameters, and $T n^\circ$ of obs
  - Diagnostic checking of the residuals.
**ARIMA**

- Integrated ARMA model:
  
  - ARIMA(p,1,q) denote a nonstationary process $Y_t$ for which the first difference $Y_t - Y_{t-1} = (1 - L)Y_t$ is a stationary ARMA(p,q) process.
    
    $\Downarrow$
    
    $Y_t$ is said to be integrated of order 1 or $I(1)$.

  - If 2 differentiations of $Y_t$ are necessary to get a stationary process i.e. $(1 - L)^2Y_t$
    
    $\Downarrow$
    
    then the process $Y_t$ is said to be integrated of order 2 or $I(2)$.

  - $I(0)$ indicate a stationary process.
The $k$-difference operator $(1 - L)^n$ with integer $n$ can be generalized to a fractional difference operator $(1 - L)^d$ with $0 < d < 1$ defined by the binomial expansion

$$(1 - L)^d = 1 - dL + d(d - 1)L^2/2! - d(d - 1)(d - 2)L^3/3! + ...$$

obtaining a fractionally integrated process of order $d$ i.e. $I(d)$.

If $d < 0.5$ the process is cov stationary and admits an AR($\infty$) representation.

The usefulness of a fractional filter $(1 - L)^d$ is that it produces hyperbolic decaying autocorrelations i.e. the so called long memory. In fact, for ARFIMA(p,d,q) processes

$$
\phi(L)(1 - L)^d Y_t = \theta(L) \epsilon_t
$$

the autocorrelation functions is proportional to

$$
\rho(k) \approx ck^{2d-1}
$$