Introduction to ARMA processes

Fulvio Corsi

SNS Pisa

Definitions

- Stochastic process (univariate): sequence of random variables $\{Y_t; \quad t \in \mathbb{N}; \text{or} \quad t \in \mathbb{Z}\}$
- Second order process: $\mathbb{E}[Y_t^2] < +\infty \qquad \forall t$
- Mean $\mu_t = \mathbb{E}[Y_t]$
- Variance $\sigma_t^2 = \mathbb{E}(Y_t - \mu_t)^2$
- Autocovariance $\gamma_t(k) \equiv Cov(Y_t, Y_{t-k}) = \mathbb{E}(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k})$ hence $\sigma_t^2 \equiv \gamma_t(0)$
- Autocorrelation

$$\rho_t(k) \equiv Corr(Y_t, Y_{t-k}) = \frac{\gamma_t(k)}{\sqrt{\gamma_t(0)\gamma_{t-k}(0)}}, \qquad -1 \leq \rho_t(k) \leq 1$$

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

Stationarity and Ergodicity

- Strict stationarity:

$$(Y_1, Y_2, ..., Y_n) \stackrel{d}{=} (Y_{1+k}, Y_{2+k}, ..., Y_{n+k}) \qquad \text{for any integer} \qquad n > 1, k$$

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

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

- Interpretation:
  - unconditional mean and variance are constant
  - mean reversion
  - shocks are transient
  - covariance between Y_t and Y_{t-k} tends to 0 as $k \to \infty$

Withe noise

- weak (uncorrelated)
  - $\mathbb{E}(\epsilon_t) = 0 \qquad \forall t$
  - $V(\epsilon_t) = \sigma^2 \qquad \forall t$
  - $Corr(\epsilon_t, \epsilon_s) = 0 \qquad \forall s \neq t$
- strong (independence)
  - $\epsilon_t \sim I.I.D.(0, \sigma^2)$
- Gaussian (weak=strong)
  - $\epsilon_t \sim N.I.D.(0, \sigma^2)$

Lag operator

- the **Lag operator** is defined as:
$$LY_t \equiv Y_{t-1}$$
- is a linear operator:
$$\begin{aligned} L(\beta Y_t) &= \beta \cdot LY_t = \beta Y_{t-1} \\ L(X_t + Y_t) &= LX_t + LY_t = X_{t-1} + Y_{t-1} \end{aligned}$$
- and admits power exponent, for instance:
$$\begin{aligned} L^2 Y_t &= L(LY_t) = LY_{t-1} = Y_{t-2} \\ L^k Y_t &= Y_{t-k} \\ L^{-1} Y_t &= Y_{t+1} \end{aligned}$$
- Some examples:
$$\begin{aligned} \Delta Y_t &= Y_t - Y_{t-1} = Y_t - LY_t = (1-L)Y_t \\ y_t &= (\theta_1 + \theta_2 L)LY_t = (\theta_1 L + \theta_2 L^2)Y_t = \theta_1 Y_{t-1} + \theta_2 Y_{t-2} \end{aligned}$$
- Expression like
$$(\theta_0 + \theta_1 L + \theta_2 L^2 + \dots + \theta_n L^n)$$
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 withe noises ϵ_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 $\mathbb{E}Y_t = 0$

$$\begin{aligned}
\gamma(0) &= \mathbb{E}Y_tY_t = \mathbb{E}(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_t + \theta\epsilon_{t-1}) = \sigma^2(1 + \theta^2); \\
\gamma(1) &= \mathbb{E}Y_tY_{t-1} = \mathbb{E}(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2}) = \sigma^2\theta; \\
\gamma(k) &= \mathbb{E}Y_tY_{t-k} = \mathbb{E}(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-k} + \theta\epsilon_{t-k-1}) = 0 \qquad \forall k > 1
\end{aligned}$$

  and,

$$\begin{aligned}
\rho(1) &= \frac{\gamma(1)}{\gamma(0)} = \frac{\theta}{1 + \theta^2} \\
\rho(k) &= \frac{\gamma(k)}{\gamma(0)} = 0 \qquad \forall k > 1
\end{aligned}$$

- hence, while a withe 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 **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

$$\begin{aligned}
\gamma(0) &= \sigma^2(1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2) \\
\gamma(k) &= \sigma^2 \sum_{j=0}^{q-k} \theta_j\theta_{j+k} \qquad \forall k \leq q \\
&= 0 \qquad \forall k > q
\end{aligned}$$

  and

$$\begin{aligned}
\rho(k) &= \frac{\sum_{j=0}^{q-k} \theta_j\theta_{j+k}}{1 + \sum_{j=1}^{q} \theta_j^2} \qquad \forall k \leq q \\
&= 0 \qquad \forall k > q
\end{aligned}$$

- Hence, an MA(q) is **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 **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(∞) process.

  If an MA process can be written as an AR(∞) 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(∞) by inverting $\theta(L)$.

- Invertibility also has important practical consequence in application. In fact, given that the ϵ_t are **not observable** they have to be reconstructed from the observed Y's through the AR(∞) representation.

Inverting lag polynomials

suppose you want to invert the generic lag polynomial

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$$

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

$$(1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)(\varphi_0 + \varphi_1 L + \varphi_2 L^2 + \ldots) = 1$$

by matching the coefficients of $L^0, L^1, \ldots, L^i$ in both sides we can obtain the φ_i recursively

$$\begin{array}{rcll}
\varphi_0 & = & 1 & \\
\varphi_1 + \varphi_0\theta_1 & = & 0 & \Rightarrow \varphi_1 = -\theta_1 \\
\varphi_2 + \varphi_1\theta_1 + \varphi_0\theta_2 & = & 0 & \Rightarrow \varphi_2 = \theta_1^2 - \theta_2 \\
\ldots & = & 0 & \ldots \\
\varphi_i + \varphi_{i-1}\theta_1 + \ldots + \varphi_0\theta_i & = & 0 & \ldots
\end{array}$$

Auto-Regressive Process (AR)

- A general AR process is defined as
$$\phi(L)Y_t = \epsilon_t$$
It is always invertible but not always stationary.
- Example: AR(1)
$$(1-\phi L)Y_t = \epsilon_t \qquad \text{or} \qquad Y_t = \phi Y_{t-1} + \epsilon_t$$
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 $|\phi| < 1$.

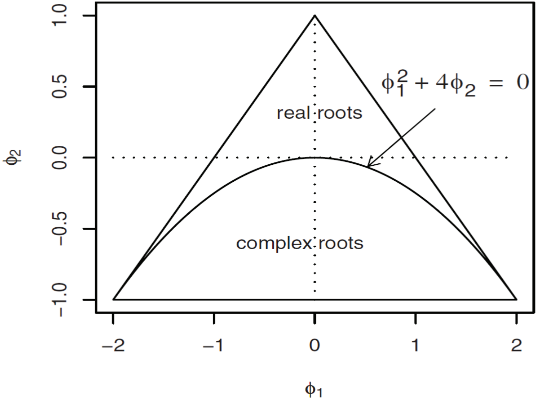
  From this representation we can apply the general formula of MA to compute $\gamma(\cdot)$ and $\rho(\cdot)$. In particular,
$$\rho(k) = \phi^{|k|} \qquad \forall k$$
i.e. monotonic exponential decay for $\phi > 0$ and exponentially damped oscillatory decay for $\phi < 0$.
- In general an AR(p) process
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$
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)

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

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

$$\phi_1 + \phi_2 < 1, \qquad \phi_2 - \phi_1 < 1 \qquad \text{and} \qquad |\phi_2| > 1$$



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

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

$$
X_t = F \quad X_{t-1} + v_t
$$

Hence, the expected value of X_t satisfy,

$$
\mathbb{E}X_t = F\, X_{t-1} \qquad \text{and} \qquad \mathbb{E}X_{t+j} = F^{j+1}\, X_{t-1}
$$

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

$$
\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - \ldots - \phi_{p-1}\lambda - \phi_p \quad = \quad 0.
$$

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

$$
1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_{p-1}z^{p-1} - \phi_p z^p \quad = \quad 0
$$

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

$$
z_1 = \lambda_1^{-1}, \quad z_2 = \lambda_2^{-1}, \quad \ldots \quad , z_p = \lambda_p^{-1}
$$

Partial Autocorrelation Function (PACF)

for an AR(p) process, the k–lag ACF ρ_k can be interpreted as simple regression

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

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

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

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

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

Importantly, AR(p) processes are “p–partially correlated” $\Rightarrow$ identification of AR order

ARMA(p,q)

- 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

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

lie outside the unit circle and, hence, admits the MA(∞) representation:

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

- the process is invertible if all the roots of

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

lie outside the unit circle and, hence, admits the AR(∞) 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.
  $\Rightarrow$ It is then very convenient to use the **"prediction–error"** decomposition of the likelihood:

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

- For example for the AR(1)

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

  the full log-Likelihood can be written as

$$l(\phi) = \underbrace{f_{Y_1}(y_1; \phi)}_{\text{marginal } 1^{st} \text{ obs}} + \underbrace{\sum_{t=2}^{T} f_{Y_t | Y_{t-1}}(y_t | y_{t-1}; \phi)}_{\substack{\text{conditional likelihood} \\ \text{under normality OLS=MLE}}} = f_{Y_1}(y_1; \phi) - \frac{T}{2} \log(2\pi) - \sum_{t=1}^{T} \log \sigma^2 - \frac{1}{2} \sum_{t=2}^{T} \frac{(y_t - \phi y_{t-1})^2}{\sigma^2}$$

  Hence, maximizing the conditional likelihood for ϕ 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 $\Rightarrow$ 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

$$l(\phi) = \underbrace{f_{Y_1}(y_1;\phi)}_{\text{marginal } 1^{st} \text{ obs}} + \underbrace{\sum_{t=2}^{T} f_{Y_t|Y_{t-1}}(y_t|y_{t-1};\phi)}_{\text{conditional likelihood}} = f_{Y_1}(y_1;\phi) - \frac{T}{2}\log(2\pi) - \sum_{t=1}^{T}\log\sigma^2 - \frac{1}{2}\sum_{t=2}^{T}\frac{(y_t - \theta\epsilon_{t-1})^2}{\sigma^2}$$

However, now the ϵ are not observed, I can only observe y. Hence, we have to recover ϵ 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.

Estimation of ARMA models

- For a general 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 $\epsilon_{t-1}, \ldots, \epsilon_{t-q} \Rightarrow \mathbb{E}[\epsilon|X] \neq 0 \Rightarrow$ OLS not consistent.

$\rightarrow$ 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)

$$MSE(\widehat{Y}_t) \equiv \mathbb{E}(Y_t - \widehat{Y}_t)^2$$

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:

$$\begin{aligned}
\mathbb{E}[Y_{t+1} - g(X_t)]^2 &= \mathbb{E}[Y_{t+1} - \mathbb{E}(Y_{t+1}|X_t) + \mathbb{E}(Y_{t+1}|X_t) - g(X_t)]^2 \\
&= \mathbb{E}[Y_{t+1} - \mathbb{E}(Y_{t+1}|X_t)]^2 \\
&+ \mathbb{E}[\mathbb{E}(Y_{t+1}|X_t) - g(X_t)]^2 \\
&+ \underbrace{2\mathbb{E}\{[Y_{t+1} - \mathbb{E}(Y_{t+1}|X_t)][\mathbb{E}(Y_{t+1}|X_t) - g(X_t)]\}}_{=0} \\
&\geq \mathbb{E}[Y_{t+1} - \mathbb{E}(Y_{t+1}|X_t)]^2
\end{aligned}$$

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

Prediction with ARMA models: AR(1) example

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

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

$$\begin{aligned}
\mathbb{E}_t(Y_{t+1}) &= \mathbb{E}_t(\phi Y_t + \epsilon_{t+1}) &&= \phi Y_t \\
\mathbb{E}_t(Y_{t+2}) &= \mathbb{E}_t(\phi^2 Y_t + \phi\epsilon_{t+1} + \epsilon_{t+2}) &&= \phi^2 Y_t \\
\mathbb{E}_t(Y_{t+k}) &= \dots &&= \phi^k Y_t
\end{aligned}$$

with

$$\begin{aligned}
Var_t(Y_{t+1}) &= Var_t(\phi Y_t + \epsilon_{t+1}) &&= \sigma^2 \\
Var_t(Y_{t+2}) &= Var_t(\phi^2 Y_t + \phi\epsilon_{t+1} + \epsilon_{t+2}) &&= (1+\phi^2)\sigma^2 \\
Var_t(Y_{t+k}) &= \dots &&= (1+\phi^2+\phi^4+\dots+\phi^{2(k-1)})\sigma^2
\end{aligned}$$

Notice that

$$\begin{aligned}
\lim_{k\to\infty} \mathbb{E}_t(Y_{t+k}) &= 0 = \mathbb{E}(Y_t) \\
\lim_{k\to\infty} Var_t(Y_{t+k}) &= \sum_{j=0}^{\infty} \phi^{2j}\sigma^2 = \frac{\sigma^2}{1-\phi^2} = Var(Y_t)
\end{aligned}$$

In general

$$Y_{t+k} = \underbrace{\{\text{function of future values}\}}_{\text{determine } Var_t(Y_t t+k)} + \underbrace{\{\text{function of past values}\}}_{\text{determine } \mathbb{E}_t(Y_{t+k})}$$

Prediction with ARMA models

- write the model in its AR(∞) representation:

$$\eta(L)(Y_t - \mu) = \epsilon_t$$

- then the optimal prediction of Y_{t+s} is given by

$$\mathbb{E}[Y_{t+s}|Y_t, Y_{t-1}, \ldots] = \mu + \left[\frac{\eta(L)^{-1}}{L^s}\right]_+ \eta(L)(Y_t - \mu) \quad \text{with} \quad \left[L^k\right]_+ = 0 \quad \text{for} \quad k < 0$$

  which is known as Wiener-Kolmogorov prediction formula.

- In the case of an AR(p) process the prediction formula can also be written as

$$\mathbb{E}[Y_{t+s}|Y_t, Y_{t-1}, \ldots] = \mu + f_{11}^{(s)}(Y_t - \mu) + f_{12}^{(s)}(Y_{t-1} - \mu) + \ldots + f_{1p}^{(s)}(Y_{t-p+1} - \mu)$$

  where $f_{11}^{(j)}$ 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

Wold Theorem: **any** mean zero covariance stationary process can be represented in the form

$$Y_t = \underbrace{\sum_{j=0}^{\infty} \varphi_j \epsilon_{t-j}}_{\text{purely random}} + \underbrace{k_j}_{\text{purely determ}}$$

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
- ϵ_{t-j} and k_j are uncorrelated.

Box-Jenkins Approach

- **check for stationarity**: if not try different transformation (ex differentiation$\rightarrow$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 $= T \log \hat{\sigma}_e^2 + 2\, m$
  - BIC/Schwarz $= T \log \hat{\sigma}_e^2 + m\ \log\ T$

    with σ_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. $(1-L)^2 Y_t$

    $$\Downarrow$$

    then the process Y_t is said to be integrated of order 2 or $I(2)$.
  - $I(0)$ indicate a stationary process.

ARFIMA

- 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

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

  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(∞) 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

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