***MA (Moving average)***

**Processo MA (1)**

- Considere o processo MA (1)

$$y_t = c + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

em que $\varepsilon_t$ é um ruído branco Gaussiano $N(0, \sigma^2)$ .

- Os seus momentos ***condicionais*** são dados por :

  - $E(y_t \mid \mathbf{Y_{t-1}}) = c + \theta_1 \varepsilon_{t-1}$
  - $Var(y_t \mid \mathbf{Y_{t-1}}) = \sigma^2$
  - $f(y_t \mid \mathbf{Y_{t-1}}) \sim N(c + \theta_1 \varepsilon_{t-1}, \sigma^2)$

- Os momentos ***incondicionais*** podem ser obtidos de forma direta:

$$E((y_t) = ( c + \theta_1 \varepsilon_{t-1} + \varepsilon_t ) = c = \mu$$

$$\begin{aligned} Var(y_t) = E(y_t - \mu)^2 &= E\left[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)\right]^2 \\ &= \sigma^2 (1 + \theta_1^2) \\ &= \gamma(0) \end{aligned}$$

## Invertibilidade de processos MA (1)

- Um modelo de séries temporais é dito **invertível** se é capaz de expressar o distúrbio aleatório corrente como função do valor corrente e de valores passados da série e dos parâmetros do modelo:

$$\varepsilon_t = g(y_t, \theta, Y_{t-1})$$

- Modelos AR são por construção invertíveis. (verifique !)

- Já para os processos MA (1), tem-se:

$$y_t = c + \theta_1 \varepsilon_{t-1} + \varepsilon_t = \mu + (1 + \theta_1 L)\varepsilon_t$$

$$(1 - (-\theta_1)L)\varepsilon_t = y_t - \mu$$

- Dividindo ambos os lados da última equação por $(1 - (-\theta_1)L)$, chega-se a:

$$\varepsilon_t = (1 - (-\theta_1)L)^{-1}(y_t - \mu)$$

- Sob as hipóteses de que $|(-\theta_1)| < 1$ e $t \to \infty$, tem-se que:

$$(1 - (-\theta_1)L)^{-1} = \sum_{j=0}^{\infty} (-\theta_1)^j L^j$$

e a equação anterior pode ser reescrita como:

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta_1)^j \; (y_{t-j} - \mu) \sim AR(\infty)$$

- Note que a condição de invertibilidade de um processo MA (1) equivale à condição de estacionariedade de um processo AR (1). Note também que um processo MA(1) é equivalente a um processo AR ($\infty$).

## Função de Autocorrelação para processos MA (1)

As equações do processo para os tempos **k** e **t-k** são dadas por:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$y_{t-k} = \mu + \theta_1 \varepsilon_{t-k-1} + \varepsilon_{t-k}$$

Avaliando o valor esperado do produto das equações:

$$\begin{aligned}
\gamma(k) &= E\left[(y_t - \mu)(y_{t-k} - \mu)\right] = \\
&= E\left[(\theta_1 \varepsilon_{t-1} + \varepsilon_t)(\theta_1 \varepsilon_{t-k-1} + \varepsilon_{t-k})\right] = \\
&= \theta_1^2 E(\varepsilon_{t-1}\varepsilon_{t-k-1}) + \theta_1 E(\varepsilon_{t-1}\varepsilon_{t-k}) + \theta_1 E(\varepsilon_t \; \varepsilon_{t-k-1}) + E(\varepsilon_t \; \varepsilon_{t-k})
\end{aligned}$$

- Observe que para **k>1**, a expressão encontrada acima é nula, pois $\varepsilon_t$ é um ruído branco que portanto não exibe correlação com seus próprios *lags*. Para k = 1, tem-se:

$$\gamma(1)= \theta_1 \sigma^2$$

- Finalmente, a FAC para processos MA puros é dada por:

$$\rho(k) = \gamma(k) / \gamma(0) = \begin{cases} \dfrac{\theta_1}{(1+\theta_1^2)} & , \quad k = 1 \\ 0 & , \quad k > 1 \end{cases}$$

- Essa é uma **característica geral** de **processos MA puros de ordem p**:

  a autocorrelação para lags **maiores que p** é sempre **nula**, o que permite identificarmos esse tipo de modelo simplesmente através do **correlograma** da série, como veremos adiante.

- Observe também que não há necessidade de imposição de nenhuma restrição sobre os parâmetros para que o modelo seja assintoticamente estacionário. Os modelos MA são ditos **trivialmente estacionários**.

## FACP para processos MA (1)

- Dado que um processo MA (1) pode ser escrito como um modelo autoregressivo de ordem infinita, segue que a função de autocorrelação parcial de processo MA (1) terá **decaimento exponencial** ou de **senóide amortecida**, dependendo da raiz do seu polinômio característico.

## Previsão em processos MA (1)

- A previsão pontual é dada por:

$$y_t = c + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$y_{t+k} = c + \theta_1 \varepsilon_{t+k-1} + \varepsilon_{t+k}, \; k=1,2,...$$

$$\begin{aligned} \hat{y}_{t+k\,|\,t} &= E(y_{t+k} | Y_t) \\ &= \mu + \theta_1 E(\varepsilon_{t+k-1}|Y_t) + E(\varepsilon_{t+k}|Y_t) \\ &= \mu + \theta_1 E(\varepsilon_{t+k-1}|Y_t) \\ &= \begin{cases} \mu + \theta_1 \varepsilon_t & , k = 1 \\ \mu & , k > 2 \end{cases} \end{aligned}$$

- Observe que na prática $\varepsilon_t$ deve ser estimado, o que pode ser feito através da expressão AR para o processo MA.

- De modo semelhante o que vimos para processo auto-regressivos, podemos construir um intervalo de confiança para essa previsão:

$$\mathrm{MSE}(\hat{y}_{t+k|t}) = E\left[(y_{t+k} - \hat{y}_{t+k|t})^2 \mid \mathbf{Y_t}\right]$$
$$= \begin{cases} \sigma^2 & , k = 1 \\ (1+\theta_1^2)\sigma^2 & , k > 1 \end{cases}$$

possibilitando a construção do IC para previsão de k passos:

$$\hat{y}_{t+k|t} \pm z^{\alpha/2}\sqrt{\mathrm{MSE}(\hat{y}_{t+k|t})}$$

**Estimação dos parâmetros**

- A estimação dos parâmetros desconhecidos de um modelo MA(1), $\theta = (\mu, \theta_1, \sigma^2)$, $\mu \in \mathbb{R}$, $\theta_1 \in (-1,1)$, $\sigma^2 \in \mathbb{R}^+$, é efetuada por **máxima verossimilhança**, não podendo ser realizada via mínimos quadrados, como veremos a seguir.

- Para um processo MA (1), assumindo que o termo aleatório segue um processo ruído branco, a densidade condicional de $y_t$ é dada por:

$$p(y_t|\mathbf{Y_{t-1}}) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp[-\frac{1}{2\sigma^2}(y_t - (\mu + \theta_1\varepsilon_{t-1})]^2 \quad t = 1,2,...,T$$

em que

$$\varepsilon_t = y_t - (\mu + \theta_1 \varepsilon_{t-1})$$
$$= (y_t - \mu) - \theta_1(y_{t-1} - \mu) + \theta_1^2(y_{t-2} - \mu) + ... + (-1)^t \theta_1^t \varepsilon_0$$

- Admitindo que o modelo é invertível, $|\theta_1| < 1$ e se $t \to \infty$, esta condição é equivalente a assumir $\varepsilon_o= 0$.

- Assim sendo a log-verossimilhança condicional em $\varepsilon_o= 0$, toma forma:

$$
\begin{aligned}
l(\theta) &= \ln\left(\prod_{t=1}^{T} p(y_t|Y_{t-1},\varepsilon_0 = 0)\right) \\
&= \ln\left(\prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(y_t - (\mu + \theta_1\varepsilon_{t-1})]^2\right) \\
&= \ln\left(\left(\frac{1}{2\pi\sigma^2}\right)^{T/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2(\mu,\theta_1)\right)\right) \\
&= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2(\mu,\theta_1)
\end{aligned}
$$

- A aplicação das condições de 1ª ordem fornece as equações para se obter os estimadores de MV:

$$\partial l / \partial\sigma^2 = 0, \quad \partial l / \partial\theta_1 = 0 \quad e \quad \partial l / \partial\mu = 0$$

(as derivadas devem ser avaliadas nos valores estimados dos outros parâmetros, caso se aplique.)

- Supondo que dispomos dos estimadores de MV para $\mu$ e $\theta_1$, o estimador para $\sigma^2$ é facilmente obtido, possuindo expressão:

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2(\hat{\mu},\hat{\theta}_1)$$

- Entretanto, é fácil checar que para os dois outros parâmetros as equações não apresentarão solução analítica, necessitando assim de algoritmo de otimização numérica para a sua solução.

- Observe que maximizar a função de verossimilhança equivale a minimizar a soma do quadrados dos resíduos.

**Generalização para processos MA(q)**

- Um processo MA(q) é definido por:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim N(0,\sigma^2)$$

$$y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\, \varepsilon_t$$

$$y_t - \mu = \Theta_q(L)\varepsilon_t$$

$$\Theta_q(L) = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)$$

- A distribuição condicional é dada por:

$$f(y_t \mid \mathbf{Y_{t-1}}) \sim N(\mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}, \sigma^2)$$

- A distribuição incondicional é dada por:

$$f(y_t) \sim N(\mu, \sigma^2(1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2))$$

já que:

$$E(y_t) = E(\mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t) = \mu$$

$$Var(y_t) = E(y_t - \mu)^2 = E(\theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t)^2$$

$$= \sigma^2 (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2) = \gamma(0)$$

- A autocovariância será dada por:

$$\gamma_k = E\left[(y_t - \mu)(y_{t-k} - \mu)\right]$$
$$= E\left[\left(\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q}\right)\left(\varepsilon_{t-k} + \theta_1\varepsilon_{t-k-1} + ... + \theta_q\varepsilon_{t-k-q}\right)\right]$$
$$= \begin{cases} \sigma^2\left(\theta_k + \theta_{k+1}\theta_1 + ... + \theta_q\theta_{q-k}\right) & 0 < k \le q \\ 0 & k > q \end{cases}$$

- Portanto, a função de autocorrelação será dada por:

$$\rho_k = \begin{cases} \dfrac{\left(\theta_k + \theta_{k+1}\theta_1 + ... + \theta_q\theta_{q-k}\right)}{\left(1 + \theta_1^2 + ... + \theta_q^2\right)} & 0 < k \le q \\ 0 & k > q \end{cases}$$

- Um pouco de reflexão nos diz que a condição de invertibilidade de um processo MA pode ser investigada através das raízes do polinômio MA:

$$(1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\, \varepsilon_t = y_t - \mu$$

$$\left(1-\lambda_1 L\right)\left(1-\lambda_2 L\right)...\left(1-\lambda_p L\right)\, \varepsilon_t = y_t - \mu$$

em que os $\lambda$'s são as raízes do polinômio de grau q

$$f(z) = \left(z^p + \theta_1 z^{p-1} + \theta_2 z^{p-2} + ... + \theta_q\right)$$

- Portanto, teremos que obter as raízes de um polinômio de grau q para investigar invertibilidade de um processo MA.

- Na prática, a mesma metodologia utilizada para investigar a estacionariedade de um processo AR(p) pode ser utilizada para investigar a invertibilidade de um processo MA(q).

- Define-se a matriz F:

$$\mathbf{F} = \begin{bmatrix} \theta_1 & \theta_2 & \theta_3 & \cdots & \theta_q \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

- Investigar a condição de invertibilidade de um processo MA(q) é equivalente a calcular os autovalores da matriz F, checando a condição:

$$|\lambda_i| < 1, \; i = 1,2, \ldots, q$$

- Quanto à estimação dos parâmetros do modelo, temos que a densidade condicional é dada por:

$$p(y_t|\mathbf{Y_{t-1}}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}\varepsilon_t^2] \quad t = 1,2,\ldots,T$$

$$\varepsilon_t = y_t - (\mu + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}).$$

- Assim sendo a log-verossimilhança condicional em $\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \ldots = \varepsilon_{-q} = 0$, toma forma:

$$l(\theta) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2(\mu, \theta_1, \theta_2, \ldots, \theta_q), \text{ com}$$

$\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \ldots = \varepsilon_{-q+1} = 0.$

- Como já indicado no caso de processos MA (1), métodos numéricos de otimização são utilizados para resolver as condições de primeira ordem e obter as estimativas dos parâmetros do modelo.

***Dualidade AR x MA:***

A dualidade entre os processos AR e MA se reproduz nas "assinaturas" reveladas na FAC e na FACP:

| | **Processo AR(P)** | **Processo MA(q)** |
|---|---|---|
| **FAC** | Decai exponencialmente ou como senóide amortecida | Corte além da ordem q do processo: $\rho_k \neq 0, \quad k = 1, 2, ..., q$ <br> $\rho_k = 0, \quad k = q+1, q+2, ...$ |
| **FACP** | Corte além da ordem p do processo: $\phi_{kk} \neq 0, \quad k = 1, 2, ..., p$ <br> $\phi_{kk} = 0, \quad k = p+1, p+2, ...$ | Decai exponencialmente ou como senóide amortecida |

Identificação dos processos:

- Se a FAC decai exponencialmente, é um indício de que o processo pode ser AR. Nesse caso, a FACP ajuda a determinar a ordem do processo.

- Se a FAC apresenta um corte abrupto depois de poucas defasagens, é um indício de que o processo pode ser MA. Isso se confirma se a FACP decai exponencialmente.

## Significância estatística da FAC e da FACP

- Viu-se que a identificação de um modelo ARMA depende das características apresentadas pelas FAC e FACP.

- Na prática só dispomos de *estimativas amostrais* dessas duas funções, obtidas a partir dos dados da ST sendo investigada.

- Como dispomos de estimativas, que são variáveis aleatórias, devemos efetuar um teste de significância, com a hipótese nula de que os valores populacionais da FAC e FACP são nulos.

- Para grandes amostras, pode-se mostrar que, sob a hipótese nula de um ruído branco:

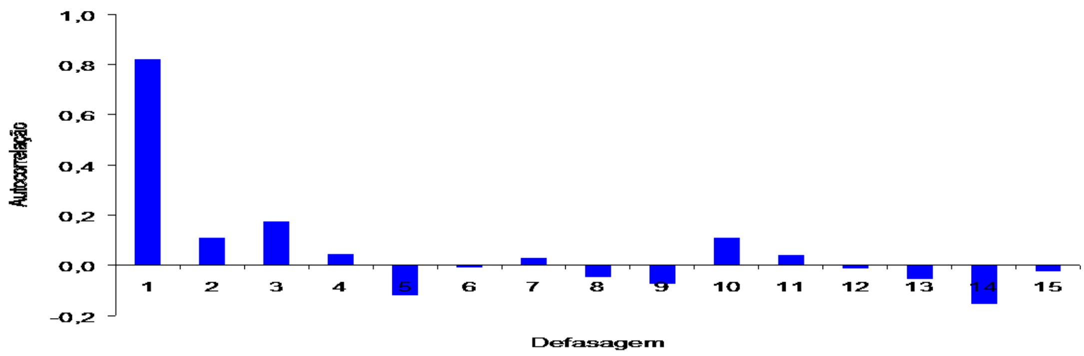
$$\text{FAC: } r_k \sim N(0, 1/T)$$
$$\text{FACP: } \hat{\varphi}_{kk} \sim N(0, 1/T)$$

- Assim, um intervalo de confiança de aproximadamente 95% para uma autocorrelação ou autocorrelação parcial amostral é dado por $\pm 2/\sqrt{T}$. Se $r_k$ ou $\hat{\varphi}_{kk}$ estiver fora do intervalo, é uma indicação de processo AR ou MA presente.

## Exemplo

- Qual a ordem de auto-regressão da série de exportações trimestrais brasileiras de derivados de petróleo (1977:1 a 1999:3)?

- A FAC estimada sugere um processo AR puro, pois ela decai exponencialmente. Por outro lado, sua FACP apresenta (ver figura) um corte em $k$ = 1:




- No entanto, o gráfico mostra autocorrelações parciais diferentes de zero (embora menores) em $k$ > 1. Será mesmo um processo AR (1) ou de ordem superior?

$$N = 91 \quad \Rightarrow \quad 2/\sqrt{91} = 0{,}21$$

- Como a única autocorrelação parcial fora do intervalo ±0,21 é a primeira, podemos concluir ao nível de aproximadamente 5% que o processo é AR (1).

- A idéia é a mesma para identificar um MA($q$), mas o teste se faz sobre as autocorrelações (não parciais).