Processos ARMA (p,q)

- Um processo ARMA (p,q) conjuga num mesmo modelo a parte AR e a parte MA, sendo dado por:

$$\Phi_p(L)\, y_t = c + \Theta_q(L)\varepsilon_t$$

onde

$$\mu = c/(1 - \varphi_1 - \varphi_2 - ... - \varphi_p)$$

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

$$\Phi_p(L) = 1 - \varphi_1 L - \varphi_2 L^2 - ... - \varphi_p L^p$$

- Os processos AR e MA, visto nas seções anteriores, são casos particulares do processo geral misto descrito acima: AR(p) ~ ARMA(p,0) e MA(q) ~ ARMA(0,q).

Estacionariedade

- A estacionariedade de um processo ARMA (p,q) será determinada pela parte AR do processo, através da análise das raízes do polinômio AR:

$$\Phi_p(L) = 1 - \varphi_1 L - \varphi_2 L^2 - ... - \varphi_p L^p$$

- Todas devem estar fora do círculo unitário. De forma equivalente, as raízes de

$$f(z) = z^p - \varphi_1 z^{p-1} - \varphi_2 z^{p-2} - ... - \varphi_p$$

devem estar todas dentro do círculo unitário.

Invertibilidade

- Como era de se esperar a invertibilidade do processo ARMA será determinada pelo polinômio MA.

- Portanto ela deve ser checada verificando se as raízes do polinômio MA:

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

estão fora do círculo unitário ou de forma equivalente que as raízes de

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

estejam todas dentro do círculo unitário.

Processo ARMA (1,1)

- Considere o modelo ARMA (1,1):

$$y_t = c + \varphi_1 \, y_{t-1} + \theta_1 \, \varepsilon_{t-1} + \varepsilon_t$$

- No modo **condicional**:

  - $E(y_t \mid \mathbf{Y_{t-1}}) = c + \varphi_1 \, y_{t-1} + \theta_1 \, \varepsilon_{t-1}$
  - $Var(y_t \mid \mathbf{Y_{t-1}}) = \sigma^2$
  - $f(y_t \mid \mathbf{Y_{t-1}}) \sim N(c + \varphi_1 \, y_{t-1} + \theta_1 \, \varepsilon_{t-1} \, , \sigma^2)$

- Para obtermos os momentos incondicionais do processo deve-se inicialmente "resolver" a equação de diferenças finitas implicada por um modelo ARMA (1,1):

$$y_t = c + \varphi_1 y_{t-1} + \theta_1 \, \varepsilon_{t-1} + \varepsilon_t$$

$$(1 - \varphi_1 L) \, y_t = c + \theta_1 \, \varepsilon_{t-1} + \varepsilon_t$$

- A solução desta equação é dada pela soma da solução homogênea (y_t^h) mais a solução particular (y_t^p):

$$y_t = y_t^h + y_t^p$$

- A solução da equação homogênea y_t^h é a solução da equação $(1 - \varphi_1 \, L) \, y_t = 0$, que será dada por:

$$y_t^h = A\varphi_1^t$$

- A solução particular y_t^p é obtida resolvendo-se a equação para y_t em termos do polinômio em L:

$$y_t^p = c + \varphi_1 y_{t-1}^p + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$(1 - \varphi_1 L)\, y_t^p = c + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$y_t^p = (1 - \varphi_1 L)^{-1} c + (1 - \varphi_1 L)^{-1}(\theta_1 \varepsilon_{t-1} + \varepsilon_t)$$

- Portanto a solução geral será dada por:

$$y_t = A\varphi_1^t + (1 - \varphi_1 L)^{-1} c + (1 - \varphi_1 L)^{-1} (\theta_1 \varepsilon_{t-1} + \varepsilon_t)$$

- A partir desta expressão podemos obter a média e a variância incondicional:

$$y_t = A\varphi_1^t + (1 - \varphi_1 L)^{-1} c + (1 - \varphi_1 L)^{-1} (\theta_1 E(\varepsilon_{t-1}) + E(\varepsilon_t))$$
$$= A\varphi_1^t + (1 - \varphi_1 L)^{-1} c \Rightarrow \text{ não é estacionário !}$$

- Usando que:

  - $|\varphi_1| < 1$
  - $\frac{1}{1-\varphi_1 L} = \sum_{j=0}^{\infty} (\varphi_1 L)^j, se\ |\varphi_1| < 1$
  - o tempo **t** está suficientemente afastado da origem.

tem-se que:

$$E(y_t) = (1 - \varphi_1 L)^{-1} c = (1 - \varphi_1)^{-1} c = \mu$$

$$\begin{aligned}
Var(y_t) &= E\left[\ y_t - \ \mu \quad ^2\right] \\
&= E\left(\theta_1 \sum_{j=0}^{\infty} \varphi_1^j\ \varepsilon_{t-j-1} + \sum_{j=0}^{\infty} \varphi_1^j\ \varepsilon_{t-j}\right)^2 \\
&= E\left(\sum_{j=0}^{\infty} \psi_j\ \varepsilon_{t-j}\right)^2
\end{aligned}$$

em que:

$$\psi_j = \begin{cases} 1, \ j = 0 \\ (\varphi_1 + \theta_1)\varphi_1^{j-1}, \ j \geq 1 \end{cases}$$

Assim:

$$\begin{aligned}
E\left(\sum_{j=0}^{\infty} \psi_j\ \varepsilon_{t-j}\right)^2 &= E\left(\sum_{j=0}^{\infty} \psi_j^2\ \varepsilon_{t-j}^2 + 2\sum_{i<j}^{\infty} \psi_j \psi_i\ \varepsilon_{t-j}\, \varepsilon_{t-i}\right) \\
&= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 \\
&= \sigma^2 \left[1 + (\varphi_1 + \theta_1)^2 (1 + \varphi_1^2 + \varphi_1^4 + \ldots)\right]
\end{aligned}$$

- Calculando a soma da progressão geométrica, chega-se finalmente à expressão para variância incondicional:

$$\mathrm{Var}(y_t) = \sigma^2 \frac{1+2\varphi_1\theta_1+\theta_1^2}{(1-\varphi_1^2)} = \gamma(0)$$

Função de Autocorrelação

- Trabalhando a equação do modelo:

$$y_t = c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$
$$y_t = (1 - \varphi_1)\mu + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$
$$(y_t - \mu) = \varphi_1 (y_{t-1} - \mu) + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

- Multiplicando os dois lados por $(y_{t-k} - \mu)$:

$$\begin{aligned}
\gamma(k) &= E\,(y_t - \mu)(y_{t-k} - \mu) \\
&= E\Big[(\varphi_1(y_{t-1} - \mu) + \theta_1 \varepsilon_{t-1} + \varepsilon_t)(y_{t-k} - \mu)\Big], \ k = 1,2,... \\
&= \varphi_1 E\,(y_{t-1} - \mu)(y_{t-k} - \mu) + \theta_1 E\,\varepsilon_{t-1}(y_{t-k} - \mu) + E\,\varepsilon_t(y_{t-k} - \mu) \\
&= \varphi_1 \gamma(k-1) + \theta_1 E\,\varepsilon_{t-1}(y_{t-k} - \mu) + E\,\varepsilon_t(y_{t-k} - \mu)
\end{aligned}$$

Utilizando que:

$$E(y_t\varepsilon_s) = \begin{cases} = 0, \text{ se } s > t, \text{pois um choque futuro não pode afetar } y_t, \\ \text{pois, pela representação de Wold, } y_t \text{ é função do valor} \\ \text{presente } \varepsilon_t \text{ e de valores passados de } \varepsilon_t : \varepsilon_{t-1}, \varepsilon_{t-2}, ... \\ \\ \neq 0, \text{ se } s \leq t \end{cases}$$

Quando a expressão deste valor esperado cruzado for diferente de zero ela será obtida utilizando-se a equação do modelo y_t.

- Para k = 1:

$$
\begin{aligned}
\gamma(1) &= \varphi_1 \gamma(0) + \theta_1 \mathrm{E}\ \varepsilon_{t-1}(y_{t-k} - \mu) + \mathrm{E}\ \varepsilon_t(y_{t-k} - \mu) \\
&= \varphi_1 \gamma(0) + \theta_1 \mathrm{E}\left[\varepsilon_{t-1}^2\right] \\
&= \sigma^2 \frac{(1+\varphi_1\theta_1)(\varphi_1+\theta_1)}{(1-\varphi_1^2)}
\end{aligned}
$$

- Para k > 1:

$$\gamma(k) = \varphi_1^{k-1}\ \gamma(k-1)$$

- Finalmente, a FAC de um processo ARMA (1,1) será dada por:

$$\rho(k) = \gamma(k)\ /\ \gamma(0) = \begin{cases} \dfrac{(1+\varphi_1\theta_1)(\varphi_1+\theta_1)}{(1+2\varphi_1\theta_1+\theta_1^2)}, & k=1 \\ \varphi_1^{k-1}\ \rho(1) & ,k>1 \end{cases}$$

- Observe que a partir do *lag* q=1 (da porção MA do processo), a forma da FAC seguirá uma exponencial ou senóide amortecida, que é a forma da FAC para processos AR.

Função de autocorrelação parcial

- Uma vez que o processo ARMA(1,1) é dado por

$$y_t = c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Usando que: $\mu = c/(1-\varphi_1)$, segue que:

$$(y_t - \mu) - \varphi_1 (y_t - \mu) = (1-(-\theta)L)\varepsilon_t$$

$$\varepsilon_t = (1-(-\theta)L)^{-1}[(y_t - \mu) - \varphi_1 (y_{t-1} - \mu)] = \sum_{j=0}^{\infty} (-\theta)^j L^j [(y_t - \mu) - \varphi_1 (y_{t-1} - \mu)]$$

$$\varepsilon_t = (y_t - \mu) - \varphi_1 (y_{t-1} - \mu) + \sum_{j=1}^{\infty} (-\theta)^j L^j [(y_t - \mu) - \varphi_1 (y_{t-1} - \mu)]$$, donde segue que:

$$y_t - \mu = \varphi_1 (y_{t-1} - \mu) + \sum_{j=1}^{\infty} (-1)^{j+1} \theta^j [(y_{t-j} - \mu) - \varphi_1 (y_{t-j-1} - \mu)] + \varepsilon_t$$

- Ou seja trata-se de um processo AR(infinito), onde o coeficiente que multiplica y_{t-1} é φ_1.

- Assim a FACP em k=1 é dada por $\varphi_{11} = \varphi_1$, e além destes termos ela decairá como exponencial/senóide amortecida, pois se comportará como um processo AR(infinito).

Previsão k passos a frente em modelos ARMA(1,1)

- Usando a expressão básica do modelo:

$$y_t = c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

$$y_{t+k} = c + \varphi_1 y_{t+k-1} + \theta_1 \varepsilon_{t+k-1} + \varepsilon_{t+k}$$

$$\hat{y}_{t+k\,|\,t} = E(y_{t+k} |\, \mathbf{Y_t})$$

$$= c + \varphi_1 E(y_{t+k-1} |\, \mathbf{Y_t}) + \theta_1 E(\varepsilon_{t+k-1} |\, \mathbf{Y_t}) + E(\varepsilon_{t+k} |\, \mathbf{Y_t})$$

$$k = 1,\ \hat{y}_{t+1|\,t} = c + \varphi_1 E(y_t |\, \mathbf{Y_t}) + \theta_1 E(\varepsilon_t |\, \mathbf{Y_t}) + E(\varepsilon_{t+1} |\, \mathbf{Y_t})$$

$$= c + \varphi_1 y_t + \theta_1 \varepsilon_t$$

$$k = 2,\ \hat{y}_{t+2\,|\,t} = c + \varphi_1 E(y_{t+1} |\, \mathbf{Y_t}) + \theta_1 E(\varepsilon_{t+1} |\, \mathbf{Y_t}) + E(\varepsilon_{t+2} |\, \mathbf{Y_t})$$

$$= c + \varphi_1 \hat{y}_{t+1|\,t}$$

$$k = 3,\ \hat{y}_{t+3\,|\,t} = c + \varphi_1 E(y_{t+2} |\, \mathbf{Y_t}) + \theta_1 E(\varepsilon_{t+2} |\, \mathbf{Y_t}) + E(\varepsilon_{t+1} |\, \mathbf{Y_t})$$

$$= c + \varphi_1 \hat{y}_{t+2|\,t}$$

$$= c + \varphi_1 c + \varphi_1^2 \hat{y}_{t+1|\,t}$$

- De forma geral:

$$\hat{y}_{t+k\,|\,t} = c\,(1 + \varphi_1 + \ldots + \varphi_1^{k-2}) + \varphi_1^{k-1}\, \hat{y}_{t+1|\,t},\ k = 2,3,\ldots$$

- De forma semelhante aos casos de processo AR eMA puros, podemos construir intervalos de confiança para as previsões pontuais:

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

em que

$$\mathrm{MSE}(\hat{y}_{t+k|t}) = E[(y_{t+k} - \hat{y}_{t+k|t})^2 \mid \mathbf{Y_t})$$

Estimação

- A estimação dos parâmetros desconhecidos de um modelo ARMA(1,1), $\theta = (c, \varphi_1, \theta_1, \sigma^2)$, é efetuada por **máxima verossimilhança**, não podendo ser realizada via mínimos quadrados devido a presença dos valores latentes dos choques aleatórios associados a parte MA(1).

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

$$p(y_t|\mathbf{Y_{t-1}}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(y_t - (c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1})]^2 \text{ ,t =1,2,...,T}$$

em que

$$\varepsilon_t = y_t - (c + \varphi_1 y_{t-1} + \theta_1 \varepsilon_{t-1})$$

- Admitindo que o modelo é invertível, $|\theta_1| < 1$. 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:

$$
\begin{aligned}
l(\theta) &= \ln\left(\prod_{t=2}^{T} p(y_t|Y_{t-1},\varepsilon_0 = 0)\right)\\
&= \ln\left(\prod_{t=2}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{1}{2\sigma^2}(y_t - (c + \varphi_1 y_{t-1} + \theta_1\varepsilon_{t-1})]^2\right)\\
&= \ln\left(\left(\frac{1}{2\pi\sigma^2}\right)^{(T-1)/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t^2(c,\theta_1,\varphi_1)\right)\right)\\
&= -\frac{(T-1)}{2}\ln(2\pi) - \frac{(T-1)}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=2}^{T}\varepsilon_t^2(c,\theta_1,\varphi_1)
\end{aligned}
$$

- Novamente, aplicam-se as condições de 1ª ordem, que fornecem as equações a partir das quais são obtidos os estimadores de MV:

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

- Supondo que dispomos dos estimadores de MV para μ, φ_1 e θ_1, o estimador para σ^2 é facilmente obtido, possuindo expressão:

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

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

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

Identificação

- O processo de identificação consiste na determinação das ordens p e q de um processo ARMA (p,q).

$$y_t = c + \varphi_1\ y_{t-1} + \varphi_2\ y_{t-2} + \ldots + \varphi_p\ y_{t-p} + $$
$$+ \theta_1\ \varepsilon_{t-1} + \theta_1\ \varepsilon_{t-2} + \ldots + \theta_1\ \varepsilon_{t-p} + \varepsilon_t$$

- Como foi visto na seção sobre a **dualidade AR x MA**, para processos puros AR ou MA, estas ordens podem ser inferidas através da inspeção simultânea da FAC e FACP, uma vez que seus comportamentos são bem determinados para esses tipos de processos.

- Entretanto, para processos mistos, a forma da FAC e FACP não nos permite identificar as ordens p e q do processo. Como exemplo, considere um processo ARMA (1,1). A FAC se comporta como a FAC de um processo AR e a FACP como a de um processo MA. Isto será também verdade para outros processos ARMA mistos.

- Mais especificamente, para um processo ARMA(p,q) pode-se mostrar os seguintes tipos de padrões:

  FAC: até o lag k=1,2,....q se comportará como um processo MA, e depois disso, para k= q+1, q+2, ... terá comportamento de um processo AR, decaindo exponencialmente ou como senóide amortecido. Ou seja, visualmente é como se fosse a FAC de um processo AR.

- a **FAC** de um processo ARMA (p,q) começa a decair a partir do **lag q**.

FACP: um processo ARMA(p,q) estacionário e inversível sempre poderá ser escrito como um processo AR(infinito), onde os primeiros p termos defasados são multiplicados pelos coeficientes $\varphi_1, \varphi_2, \cdots, \varphi_p$. Generalizando o resultado do ARMA(1,1), segue que os primeiros p termos da FACP serão aqueles de um processo AR(p), isto é, $\varphi_{kk}=\varphi_k$, k=1,2,...,p, e além destes começará a decair como exponencial ou senóide amortecida. Ou seja, visualmente é como se fosse a FACP de um processo MA.

- a **FACP** de um processo ARMA (p,q) começa a decair a partir do **lag p**.

- É então possível concluir que para um processo misto ARMA(p,q) FAC e FACP decaem exponencialmente e/ou como senóide amortecida, qualquer que sejam p e q.

- Como conseqüência da não unicidade da forma da FAC/FACP para processos mistos, uma ST estacionária real que produz FAC/FACP fora do “padrão” de processos puros, será considerada como sido gerado por um processo misto. Entretanto, não se pode determinar de forma imediata que processo misto em particular a gerou.

- A identificação de processos mistos (ou puros) utilizando ST reais é também auxiliada através de "<u>critérios de informação</u>" (CI).

- Dado um conjunto de modelos possíveis p/ uma ST, os CI selecionam, entre estes modelos, aquele que minimiza, simultaneamente, a <u>discrepância</u> (entre os modelos e os dados) e a <u>complexidade</u>.

- A minimização da discrepância é uma dimensão desejável de um critério de seleção de modelos.

- Funções típicas para medir a discrepância são:

$$D(\hat{\varepsilon}_t) = \begin{cases} -\ S_p(\hat{\theta}) = \sum_{t=1}^{T} |\hat{\varepsilon}_t|^p, \ p=1,2,3,.. \\ -\ \log L(\hat{\theta}) \end{cases}$$

sendo:

- $\hat{\varepsilon}_t$ o resíduo, ou inovação do modelo;

- $L(\hat{\theta})$ = é a verossimilhança avaliada no máximo.

- A lógica da derivação[1] dos CI's fazem com que todos tenham a mesma função de aderência, a qual é dada por $-2 \log L(\hat{\theta})$.

- A complexidade mínima é justificada pelo **Princípio da Parcimônia**: sendo iguais outras características desejáveis, modelos mais simples, isto é, com menor número de parâmetros, devem ser preferidos.

---

[1] Uma derivação formal destes critérios pode ser obtida utilizando a "distância" de Kullback-Leibler.

- A "penalização" da complexidade é expressa pela função de penalização f(k,T), a qual envolve **k**, o número de parâmetros utilizados em cada um dos modelos.

- Formalmente, os CI são expressos pela seguinte relação geral:

$$CI(k) = \text{discrepância} + \text{complexidade}$$
$$= -2 \log L(\hat{\theta}) / T^* + f(k, T^*) / T^*$$

em que:

- k é o número de parâmetros independentes estimados. Para modelos ARMA com média não nula, $k = p + q + 1$;

- T^* = número de obs. efetivamente utilizadas na estimação do modelo;

- $L(\hat{\theta})$ é a verossimilhança avaliada no máximo.

- Os critérios mais conhecidos e suas funções de penalização são dados abaixo:

  - Crit. de Akaike (AIC) : $f(k, T) = 2k / T^*$
  - Crit. de Schwarz (BIC): $f(k, T) = k \ln(T^*) / T^*$

- Na prática: dada uma ST estimamos vários modelos ARMA (p,q), calculamos o AIC/BIC para cada um deles e então escolhemos como o modelo "mais apropriado" para a ST aquele que minimizar um dos critérios.

- Assintoticamente ($T \rightarrow \infty$), AIC tende a sobre-identificar os modelos, enquanto que BIC tende a identificar corretamente o modelo (consistente).

- Considerando a normalidade dos erros do processo ARMA, i.e., que $\varepsilon_t \sim N(0,\sigma^2)$, então segue que:

$$l(\theta) = \log L(\theta) = -\frac{(T-p)}{2}\log(2\pi) - \frac{(T-p)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=p+1}^{T} \varepsilon_t^2(c, \theta\text{'s}, \varphi\text{'s}).$$

Como $\hat{\sigma}^2 = \frac{1}{(T-p)}\sum_{t=p+1}^{T} \hat{\varepsilon}_t^2(\hat{c}, \hat{\theta}\text{'s}, \hat{\varphi}\text{'s})$, subsituindo na expressão acima, temos:

$$-2 \log L(\hat{\theta}) = (T-p) \log\left(\sum_{t=p+1}^{T} \hat{\varepsilon}_t^2(\hat{c}, \hat{\theta}\text{'s}, \hat{\varphi}\text{'s}) \right)$$

De forma que os CI's ficam com expressão dada por:

$$\begin{aligned} CI &= -2 \log L(\hat{\theta}) + \begin{cases} 2k/T^* & \text{(AIC)} \\ k \log T^*/T^* & \text{(BIC)} \end{cases} \\ &= \log\left(\sum_{t=p+1}^{T} \hat{\varepsilon}_t^2(\hat{c}, \hat{\theta}\text{'s}, \hat{\varphi}\text{'s}) \right) + \begin{cases} 2k/T^* & \text{(AIC)} \\ k \log T^*/T^* & \text{(BIC)} \end{cases} \end{aligned}$$

Processos ARIMA(p,d,q)

- Se a série a ser modelada for não-estacionária então devemos faze-la estacionária efetuando um número adequado de diferenciações:

$$z_t = \Delta^d y_t = (1 - L)^d y_t, \text{ ou } y_t \sim I(d) \quad d = 1,2,3,...$$

onde **d** é a <u>ordem de integração</u> da série, isto é, número de diferenciações necessárias para tornar a série estacionária.

- Assim sendo o processo ARMA(p,q) será aplicado à série z_t, o qual irá definir um processo ARIMA(p,d,q) p/ a série y_t.

$$\Phi_p(L)\, z_t = c + \Theta_q(L)\varepsilon_t$$

$$\Phi_p(L)\Delta^d y_t = c + \Theta_q(L)\varepsilon_t$$

$$\varphi_{p+d}(L) y_t = c + \Theta_q(L)\varepsilon_t$$

em que:

- c = é o drift do modelo
- $\Theta_q(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$
- $\Phi_p(L) = 1 - \varphi_1 L - \varphi_2 L^2 - ... - \varphi_p L^p$
- $\varphi_{p+d}(L) = \Phi_p(L)\Delta^d = (1 - L)^d\ \Phi_p(L)$

$$= 1 - \varphi_1 L - \varphi_2 L^2 - ... - \varphi_{p+d} L^{p+d}$$

Exemplos de processos ARIMA(p,d,q):

- ARIMA(0,1,1): $\Delta y_t = (1 + \theta_1 L)\varepsilon_t$
- ARIMA(0,2,2): $\Delta^2 y_t = (1 + \theta_1 L + \theta_2 L^2)\varepsilon_t$
- ARIMA(1,1,1): $(1- \varphi_1 L)\Delta y_t = (1 + \theta_1 L)\varepsilon_t$

- Pode-se mostrar que y_t segue um processo ARIMA(p,d,q) com drift se e somente se y_t segue um processo de tendência polinomial de grau d, superposto a um erro não-estacionário ARIMA(p,d,q).

- Formalmente o resultado anterior pode ser colocado sob a forma:

$$\Phi_p(L)\,\Delta^d y_t = c + \Theta_q(L)\varepsilon_t,\ c = \Phi_p(1)\,\beta_d\ d!$$

$$\Updownarrow$$

$$y_t = \sum_{j=0}^{d} \beta_j t^j + a_t,\quad a_t = \frac{\Theta_q(L)\varepsilon_t,}{\Phi_p(L)\,\Delta^d}$$

- Este comportamento pode ser observado considerando um caso particular.

- Considere que y_t é um processo obtido pela superposição de uma tendência linear determinística a um processo ARIMA(1,1,1). Vamos provar que este processo implica num ARIMA(1,1,1) com drift. A prova completa exige a prova da volta dessa afirmativa, que é deixada para o aluno.

- Temos como hipótese:

$$y_t = (\beta_0 + \beta_1 t) + a_t$$

em que:

$$(1 - \varphi_1 L)\Delta a_t = (1 + \theta_1 L)\varepsilon_t$$

$$a_t = \frac{(1 + \theta_1 L)\varepsilon_t}{(1 - \varphi_1 L)\Delta}$$

- Trabalhando a expressão:

$$y_t = (\beta_0 + \beta_1 t) + \frac{(1 + \theta_1 L)\varepsilon_t}{(1 - \varphi_1 L)\Delta}$$

$$\Delta y_t = \Delta(\beta_0 + \beta_1 t) + \frac{(1 + \theta_1 L)\varepsilon_t}{(1 - \varphi_1 L)}$$

$$\Delta y_t = \beta_1 + \frac{(1 + \theta_1 L)\varepsilon_t}{(1 - \varphi_1 L)}$$

$$(1 - \varphi_1 L)\Delta y_t = (1 - \varphi_1 L)\beta_1 + \frac{(1 - \varphi_1 L)(1 + \theta_1 L)\varepsilon_t}{(1 - \varphi_1 L)}$$

$$(1 - \varphi_1 L)\Delta y_t = (1 - \varphi_1 L)\beta_1 + (1 + \theta_1 L)\varepsilon_t$$

$$(1 - \varphi_1 L)\Delta y_t = c + (1 + \theta_1 L)\varepsilon_t$$

Processo ARIMA(0,1,1)

- Introduzimos um método de previsão para ST desenvolvido na década de 50, conhecido como EWMA (*exponentially weighted moving average*), ou métodos das médias móveis.

- Diferente de um modelo, um método é uma expressão matemática útil que especifica como as previsões são formadas. O EWMA é dado por:

$$\hat{y}_{t+1|t} = \lambda \sum_{j=0}^{\infty} (1-\lambda)^j \, y_{t-j} \qquad ,0 < \lambda < 1 \quad \text{(i)}$$

$$= \lambda \, y_t + \lambda \sum_{j=1}^{\infty} (1-\lambda)^j \, y_{t-j}$$

$$\hat{y}_{t+1|t} = \lambda \, y_t + (1-\lambda)\,\hat{y}_{t\,|t-1} \qquad \text{(ii)}$$

Observe que:

- na formação da previsão quanto mais afastada da origem t for uma observação, menor o seu peso na contribuição da previsão.

- os pesos são normalizados, pois $\sum_{j=0}^{\infty}(1-\lambda)^j \lambda = 1$.

- o fator de decaimento λ controla a variação da previsão: qto + or λ, maior o peso das observações mais recentes e assim sendo a previsão responde mais rapidamente às flutuações da série. Isto é adequado para séries instáveis, com muita variação. De forma análoga, valores pequenos de λ são indicados para séries que apresentam maior estabilidade.

- Na prática, a implementação do EWMA necessita de:

  - valor inicial para a previsão, que pode ser a média aritmética da série;

  - valor para λ, o qual pode ser estimado minimizando um dos seguintes critérios estatísticos: Soma dos Quadrados dos Resíduos (SQR) ou Soma dos Erros Absolutos (SEA):

$$\text{SQR}(\lambda) = \sum_{t=1}^{T}(y_{t+1} - \hat{y}_{t+1|t}(\lambda))^2$$

$$\text{SEA}(\lambda) = \sum_{t=1}^{T}| y_{t+1} - \hat{y}_{t+1|t}(\lambda)|$$

- Através de método numérico apropriado, escolhe-se o λ que minimiza um dos critérios anteriores.

- Pode-se estabelecer uma relação entre o EWMA e a função de previsão de um ARIMA(0,1,1) sem o *drift* c. Trabalhando a expressão desse processo, tem-se:

$$\Delta y_t = (1 + \theta_1 L)\varepsilon_t$$

$$y_t = y_{t-1} + \theta_1\varepsilon_{t-1} + \varepsilon_t \qquad \text{(i)}$$

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

$$\hat{y}_{t+1|t} = E(y_{t+1}|\mathbf{Y_t}) = E(y_t + \theta_1\varepsilon_t + \varepsilon_{t+1}|\mathbf{Y_t})$$

$$\hat{y}_{t+1|t} = y_t + \theta_1\varepsilon_t \qquad \text{(ii)}$$

$$\hat{y}_{t|t-1} = y_{t-1} + \theta_1\varepsilon_{t-1} \therefore y_{t-1} = \hat{y}_{t|t-1} - \theta_1\varepsilon_{t-1} \qquad \text{(iii)}$$

- Substituindo a equação (iii) em (i):

$$y_t = \hat{y}_{t|t-1} - \theta_1\varepsilon_{t-1} + \theta_1\varepsilon_{t-1} + \varepsilon_t = \hat{y}_{t|t-1} + \varepsilon_t$$

$$y_t = \hat{y}_{t|t-1} + \varepsilon_t \qquad \text{(iv)}$$

- De (ii), tem-se que:

$$y_t = \hat{y}_{t+1|t} - \theta_1\varepsilon_t$$

- Igualando essa última expressão a (iv):

$$\hat{y}_{t+1|t} = \hat{y}_{t|t-1} + (1 + \theta_1)\varepsilon_t$$

- Supondo que $\theta_1 \in (-1,0)$, então $\lambda = (1 + \theta_1) \in (0,1)$. Isolando ε_t em (iv), finalmente chegamos a:

$$\hat{y}_{t+1|t} = (1-\lambda)\,\hat{y}_{t|t-1} + \lambda\, y_t$$