

Capítulo 4

Metodología de Box-Jenkins

Econometría II · UFBA

Prof. Pablo Castro

O que veremos

- Revisão de ACF e PACF (preparação)
- Etapa 1: **Identificação**
- Etapa 2: **Estimação** (AIC, BIC, parcimônia)
- Etapa 3: **Diagnóstico** (Ljung-Box, normalidade)

O que é Box-Jenkins?

Um conjunto sistemático de procedimentos para ajustar um modelo **ARIMA** a uma série temporal. Não é uma fórmula – é uma metodologia que exige treino e julgamento.

Preparação

Revisão de ACF e PACF

ACF: Função de Autocorrelação

Definição

A **Função de Autocorrelação** (ACF, *Autocorrelation Function*) mede a correlação entre y_t e a sua defasagem de ordem k , y_{t-k} :

$$\rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)} = \frac{\gamma_k}{\gamma_0}$$

Notação:

- $\gamma_k = \text{Cov}(y_t, y_{t-k})$ – **autocovariância** de ordem k .
- $\gamma_0 = \text{Var}(y_t)$ – variância (é a autocovariância de ordem 0).

A ACF é simplesmente a **correlação** entre observações separadas por k períodos.

Intuição

A ACF é a fotografia da “memória” da série: quanto y_t ainda se parece com y_{t-k} .

Calculando a ACF

Para calcular ρ_k de qualquer processo estacionário, basta seguir 3 passos:

Calculando a ACF

Passo 1. Calcular $\gamma_0 = \text{Var}(y_t)$.

Passo 2. Calcular $\gamma_k = \text{Cov}(y_t, y_{t-k})$ para $k = 1, 2, \dots$

Passo 3. Dividir: $\rho_k = \gamma_k / \gamma_0$.

Ferramentas que vamos usar repetidas vezes:

- Ruído branco ε_t : $\mathbb{E}(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = \sigma^2$, $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ para $t \neq s$.
- Se y_t é estacionário: $\mathbb{E}(y_t) = \mu$ constante e γ_k depende só de k , não de t .
- Quando $\mathbb{E}(y_t) = 0$: $\text{Cov}(y_t, y_{t-k}) = \mathbb{E}(y_t \cdot y_{t-k})$.

Vamos aplicar a 2 processos: AR(1) e MA(1).

Exemplo 1: ACF do AR(1) – apresentação

Processo AR(1)

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1, \quad \varepsilon_t \sim \text{RB}(0, \sigma^2)$$

Premissas que precisamos:

- $\mathbb{E}(y_t) = 0$ (sem constante e com $|\phi| < 1$, a média de longo prazo é zero).
- ε_t é independente de y_{t-1}, y_{t-2}, \dots (causalidade).
- Estacionariedade $\Rightarrow \text{Var}(y_t) = \text{Var}(y_{t-1}) = \gamma_0$ (a mesma variância para todo t).

Roteiro de cálculo:

1. Calcular γ_0 .
2. Calcular γ_1 .
3. Generalizar para γ_k e obter ρ_k .

AR(1): cálculo de $\gamma_0 = \text{Var}(y_t)$

Passo 1. Aplicar variância em $y_t = \phi y_{t-1} + \varepsilon_t$:

$$\text{Var}(y_t) = \text{Var}(\phi y_{t-1} + \varepsilon_t)$$

Passo 2. Como y_{t-1} e ε_t são independentes, $\text{Var}(\text{soma}) = \text{soma das Var}$:

$$\text{Var}(y_t) = \phi^2 \text{Var}(y_{t-1}) + \text{Var}(\varepsilon_t)$$

Passo 3. Usar estacionariedade ($\text{Var}(y_t) = \text{Var}(y_{t-1}) = \gamma_0$) e $\text{Var}(\varepsilon_t) = \sigma^2$:

$$\gamma_0 = \phi^2 \gamma_0 + \sigma^2$$

Passo 4. Isolar γ_0 :

$$\gamma_0 (1 - \phi^2) = \sigma^2 \quad \Longrightarrow \quad \boxed{\gamma_0 = \frac{\sigma^2}{1 - \phi^2}}$$

Observação: aqui já vemos por que precisamos de $|\phi| < 1$ – caso contrário a variância seria infinita ou negativa.

AR(1): cálculo de $\gamma_1 = \text{Cov}(y_t, y_{t-1})$

Passo 1. Como $\mathbb{E}(y_t) = 0$, temos $\gamma_1 = \mathbb{E}(y_t \cdot y_{t-1})$.

Passo 2. Substituir $y_t = \phi y_{t-1} + \varepsilon_t$:

$$\gamma_1 = \mathbb{E}[(\phi y_{t-1} + \varepsilon_t) y_{t-1}]$$

Passo 3. Distribuir dentro da esperança:

$$\gamma_1 = \phi \mathbb{E}(y_{t-1}^2) + \underbrace{\mathbb{E}(\varepsilon_t y_{t-1})}_{=0}$$

(o segundo termo é zero porque ε_t é independente de y_{t-1}).

Passo 4. Reconhecer que $\mathbb{E}(y_{t-1}^2) = \text{Var}(y_{t-1}) = \gamma_0$:

$$\boxed{\gamma_1 = \phi \gamma_0}$$

Ou seja, a autocovariância de ordem 1 é apenas ϕ vezes a variância.

AR(1): generalizando γ_k e obtendo a ACF

Repetindo o mesmo argumento para γ_2 :

$$\gamma_2 = \mathbb{E}(y_t y_{t-2}) = \mathbb{E}[(\phi y_{t-1} + \varepsilon_t) y_{t-2}] = \phi \mathbb{E}(y_{t-1} y_{t-2}) = \phi \gamma_1$$

Logo, $\gamma_2 = \phi \cdot \phi \gamma_0 = \phi^2 \gamma_0$.

Por indução:

$$\gamma_k = \phi^k \gamma_0$$

Dividindo por γ_0 , obtemos a ACF do AR(1):

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k$$

Interpretação

A ACF de um AR(1) **decai geometricamente** à taxa ϕ . Por exemplo, com $\phi = 0,7$: $\rho_1 = 0,7$, $\rho_2 = 0,49$, $\rho_3 \approx 0,34$, $\rho_4 \approx 0,24$, e assim por diante – nunca chega ao zero exato, porém converge.

Exemplo 2: ACF do MA(1) – apresentação

Processo MA(1)

$$y_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim \text{RB}(0, \sigma^2)$$

Observação importante: um MA(q) é **sempre estacionário** – não precisamos de nenhuma restrição sobre θ para garantir estacionariedade.

Premissas que precisamos:

- $\mathbb{E}(y_t) = 0$ (média dos ruídos é zero).
- $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ para $t \neq s$ (ruído branco).

Roteiro: calcular $\gamma_0, \gamma_1, \gamma_2$ e a partir daí a ACF.

MA(1): cálculo de γ_0 e γ_1

Cálculo de $\gamma_0 = \text{Var}(y_t)$:

$$\text{Var}(y_t) = \text{Var}(\varepsilon_t) + \theta^2 \text{Var}(\varepsilon_{t-1}) = \sigma^2 + \theta^2 \sigma^2 = \sigma^2(1 + \theta^2)$$

(o termo cruzado $2\theta \text{Cov}(\varepsilon_t, \varepsilon_{t-1}) = 0$ porque os ruídos são não correlacionados).

Cálculo de $\gamma_1 = \mathbb{E}(y_t y_{t-1})$:

$$\gamma_1 = \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2})]$$

Distribuindo (4 termos):

$$= \underbrace{\mathbb{E}(\varepsilon_t \varepsilon_{t-1})}_{=0} + \theta \underbrace{\mathbb{E}(\varepsilon_t \varepsilon_{t-2})}_{=0} + \theta \underbrace{\mathbb{E}(\varepsilon_{t-1}^2)}_{=\sigma^2} + \theta^2 \underbrace{\mathbb{E}(\varepsilon_{t-1} \varepsilon_{t-2})}_{=0}$$

$$\boxed{\gamma_1 = \theta \sigma^2}$$

MA(1): cálculo de γ_2 e a ACF

Cálculo de $\gamma_2 = \mathbb{E}(y_t y_{t-2})$:

$$\gamma_2 = \mathbb{E}[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-2} + \theta\varepsilon_{t-3})]$$

Todos os 4 termos cruzados envolvem ε s com índices **diferentes**:

$$= \mathbb{E}(\varepsilon_t \varepsilon_{t-2}) + \theta \mathbb{E}(\varepsilon_t \varepsilon_{t-3}) + \theta \mathbb{E}(\varepsilon_{t-1} \varepsilon_{t-2}) + \theta^2 \mathbb{E}(\varepsilon_{t-1} \varepsilon_{t-3}) = 0$$

Mesma lógica para todo $k \geq 2$: só aparecem produtos de ε s com índices diferentes, então $\gamma_k = 0$.

ACF do MA(1):

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\theta \sigma^2}{\sigma^2(1 + \theta^2)} = \frac{\theta}{1 + \theta^2}, \quad \boxed{\rho_k = 0 \text{ para } k \geq 2}$$

Interpretação

A ACF de um MA(1) **corta abruptamente** após o lag 1. Para $\theta = 0,6$: $\rho_1 = 0,6/1,36 \approx 0,44$, e $\rho_k = 0$ para $k \geq 2$. **O q de um MA(q) é o último lag onde a ACF é diferente de zero.**

E o PACF?

ACF: ela não distingue a **correlação direta** da **correlação induzida** pelo “caminho” através dos lags intermediários.

Exemplo: suponha que a inflação do mês t dependa **só** da inflação de $t - 1$ (um AR(1) puro).

Pela ACF que acabamos de calcular, $\rho_2 = \phi^2 \neq 0$. Isso **parece** dizer que y_t depende de y_{t-2} .

Mas essa dependência não é direta! Ela só existe porque:

$$y_t \rightarrow y_{t-1} \rightarrow y_{t-2}$$

ou seja, y_t se relaciona com y_{t-2} **apenas via** y_{t-1} .

A **PACF** responde: *depois de eu controlar pelos lags intermediários, existe alguma correlação direta entre y_t e y_{t-k} ?* Se a resposta for “não”, então o lag k não traz informação nova.

PACF: Função de Autocorrelação Parcial

Definição

A PACF de ordem k , denotada ϕ_{kk} , é a correlação entre y_t e y_{t-k} **depois de remover** o efeito linear de $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$:

$$\phi_{kk} = \text{Corr}(y_t, y_{t-k} \mid y_{t-1}, \dots, y_{t-k+1})$$

Como obter operacionalmente? Para **cada** ordem k , estimamos uma regressão diferente – com k lags – e olhamos o coeficiente do **último lag**.

Notação dos coeficientes: usamos **dois índices**. Em ϕ_{ij} :

- i é a **ordem da regressão** (quantos lags entram no modelo);
- j é a **posição do lag** dentro daquela regressão.

A PACF é sempre ϕ_{kk} : o coeficiente do lag k na regressão que inclui todos os k lags ($i = j = k$).

PACF: a sequência de regressões

Para construir a PACF, fazemos uma regressão diferente para cada ordem k :

Para $k = 1$ (1 regressor):

$$y_t = \phi_{11} y_{t-1} + e_t \quad \Rightarrow \quad \phi_{11} \text{ é a PACF no lag 1.}$$

Para $k = 2$ (2 regressores):

$$y_t = \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + e_t \quad \Rightarrow \quad \phi_{22} \text{ é a PACF no lag 2.}$$

Para $k = 3$ (3 regressores):

$$y_t = \phi_{31} y_{t-1} + \phi_{32} y_{t-2} + \phi_{33} y_{t-3} + e_t \quad \Rightarrow \quad \phi_{33} \text{ é a PACF no lag 3.}$$

Em geral, para k qualquer:

$$y_t = \phi_{k1} y_{t-1} + \phi_{k2} y_{t-2} + \cdots + \phi_{kk} y_{t-k} + e_t$$

A PACF é a **sequência** $\phi_{11}, \phi_{22}, \phi_{33}, \dots$ – sempre o último coeficiente de cada regressão (primeiro índice = segundo índice).

PACF: equações de Yule-Walker

Felizmente, não precisamos rodar uma regressão para cada k . Existe uma fórmula que liga a PACF à ACF – as **equações de Yule-Walker**.

Yule-Walker (forma matricial)

Para cada ordem k , resolvemos o sistema $P_k \phi_k = \rho_k$:

$$\underbrace{\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix}}_{P_k} \underbrace{\begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix}}_{\phi_k} = \underbrace{\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}}_{\rho_k}$$

A PACF de ordem k é o **último elemento** do vetor solução ϕ_k .

Como ler a matriz P_k :

- Diagonal principal só tem 1 ($\rho_0 = 1$).
- Entrada (i, j) é $\rho_{|i-j|}$: depende da *distância* entre os índices. P_k é simétrica.

Exemplo: Yule-Walker para $k = 3$

Vamos visualizar o caso $k = 3$. Para cada linha de P_3 , multiplicamos pelo vetor coluna ϕ_3 elemento a elemento e igualamos à entrada correspondente de ρ_3 .

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \end{pmatrix}$$

Linha 1 (azul): $1 \cdot \phi_{31} + \rho_1 \cdot \phi_{32} + \rho_2 \cdot \phi_{33} = \rho_1$

Linha 2 (vermelho): $\rho_1 \cdot \phi_{31} + 1 \cdot \phi_{32} + \rho_1 \cdot \phi_{33} = \rho_2$

Linha 3 (verde): $\rho_2 \cdot \phi_{31} + \rho_1 \cdot \phi_{32} + 1 \cdot \phi_{33} = \rho_3$

Resultado: um sistema linear de 3 equações em 3 incógnitas $\phi_{31}, \phi_{32}, \phi_{33}$. A PACF de ordem 3 é ϕ_{33} .

PACF do AR(1): cálculo passo a passo ($k = 1$)

Lembre da ACF do AR(1): $\rho_k = \phi^k$, ou seja, $\rho_1 = \phi$, $\rho_2 = \phi^2$, $\rho_3 = \phi^3$, etc.

Yule-Walker para $k = 1$: o sistema é apenas 1×1 :

$$(1) \cdot (\phi_{11}) = (\rho_1) \iff \phi_{11} = \rho_1$$

Substituindo $\rho_1 = \phi$:

$$\boxed{\phi_{11} = \phi}$$

Para $\phi = 0,7$

$\phi_{11} = 0,7$. No lag 1, a PACF coincide com a ACF – já que não há lag intermediário para “descontar”.

PACF do AR(1): cálculo passo a passo ($k = 2$)

Yule-Walker para $k = 2$: sistema 2×2 :

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

Substituindo $\rho_1 = \phi$ e $\rho_2 = \phi^2$:

$$\begin{pmatrix} 1 & \phi \\ \phi & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \phi \\ \phi^2 \end{pmatrix}$$

Passo 1 – Abrir o produto linha a linha:

$$\text{Linha 1: } \phi_{21} + \phi \phi_{22} = \phi \quad \text{Linha 2: } \phi \phi_{21} + \phi_{22} = \phi^2$$

Passo 2 – Isolar ϕ_{21} na Linha 1:

$$\phi_{21} = \phi - \phi \phi_{22}$$

PACF do AR(1): cálculo passo a passo ($k = 2$)

Passo 3 – Substituir na Linha 2:

$$\phi(\phi - \phi\phi_{22}) + \phi_{22} = \phi^2 \iff \phi^2 - \phi^2\phi_{22} + \phi_{22} = \phi^2$$

Passo 4 – Cancelar ϕ^2 dos dois lados e isolar ϕ_{22} :

$$\phi_{22}(1 - \phi^2) = 0 \implies \boxed{\phi_{22} = 0}$$

(como $|\phi| < 1$, $1 - \phi^2 \neq 0$).

Interpretação: a defasagem de ordem 2 **não traz informação nova** sobre y_t , uma vez que já controlamos por y_{t-1} – exatamente o que esperaríamos de um AR(1) puro.

PACF do AR(1): generalizando para $k \geq 3$

Yule-Walker para $k = 3$ (com $\rho_1 = \phi$, $\rho_2 = \phi^2$, $\rho_3 = \phi^3$):

$$\begin{pmatrix} 1 & \phi & \phi^2 \\ \phi & 1 & \phi \\ \phi^2 & \phi & 1 \end{pmatrix} \begin{pmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} \phi \\ \phi^2 \\ \phi^3 \end{pmatrix}$$

Tentativa de solução: $\phi_{31} = \phi$, $\phi_{32} = 0$, $\phi_{33} = 0$.

Verificando cada linha:

- Linha 1: $1 \cdot \phi + \phi \cdot 0 + \phi^2 \cdot 0 = \phi \quad \checkmark$
- Linha 2: $\phi \cdot \phi + 1 \cdot 0 + \phi \cdot 0 = \phi^2 \quad \checkmark$
- Linha 3: $\phi^2 \cdot \phi + \phi \cdot 0 + 1 \cdot 0 = \phi^3 \quad \checkmark$

Por indução, para todo $k \geq 2$:

$$\phi_{kk} = 0 \quad \text{para todo } k \geq 2$$

A PACF do AR(1) **corta após o lag 1.**

PACF do MA(1): cálculo passo a passo ($k = 1$)

Lembre da ACF do MA(1):

$$\rho_1 = \frac{\theta}{1 + \theta^2}, \quad \rho_k = 0 \text{ para } k \geq 2.$$

Yule-Walker para $k = 1$: sistema 1×1 :

$$(1) \cdot (\phi_{11}) = (\rho_1) \iff \boxed{\phi_{11} = \rho_1 = \frac{\theta}{1 + \theta^2}}$$

Exemplo: $\theta = 0,6$:

$$\phi_{11} = \frac{0,6}{1 + 0,36} = \frac{0,6}{1,36} \approx 0,441$$

Diferença em relação ao AR(1): lá tínhamos $\phi_{11} = \phi$ – a PACF no lag 1 era o próprio parâmetro do modelo. Aqui, ϕ_{11} é uma função de θ , mas não bate diretamente.

PACF do MA(1): cálculo passo a passo ($k = 2$)

Yule-Walker para $k = 2$:

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix}$$

Substituindo $\rho_2 = 0$:

$$\begin{pmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{21} \\ \phi_{22} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ 0 \end{pmatrix}$$

Passo 1 – Abrir o produto linha a linha:

$$\text{Linha 1: } \phi_{21} + \rho_1 \phi_{22} = \rho_1 \quad \text{Linha 2: } \rho_1 \phi_{21} + \phi_{22} = 0$$

Passo 2 – Isolar ϕ_{22} na Linha 2:

$$\phi_{22} = -\rho_1 \phi_{21}$$

PACF do MA(1): cálculo passo a passo ($k = 2$)

Passo 3 – Substituir na Linha 1:

$$\phi_{21} + \rho_1 (-\rho_1 \phi_{21}) = \rho_1 \iff \phi_{21}(1 - \rho_1^2) = \rho_1$$
$$\phi_{21} = \frac{\rho_1}{1 - \rho_1^2}$$

Passo 4 – Voltar à Linha 2 e obter ϕ_{22} :

$$\phi_{22} = -\rho_1 \cdot \frac{\rho_1}{1 - \rho_1^2} = \boxed{-\frac{\rho_1^2}{1 - \rho_1^2}}$$

Exemplo $\theta = 0,6$ ($\rho_1 \approx 0,441$):

$$\phi_{22} \approx \frac{-(0,441)^2}{1 - (0,441)^2} = \frac{-0,194}{0,806} \approx -0,241$$

Já é diferente de zero. A PACF do MA(1) **não corta** no lag 2.

PACF do MA(1): cálculo passo a passo ($k = 3$)

Yule-Walker para $k = 3$ (com $\rho_2 = \rho_3 = 0$):

$$\begin{pmatrix} 1 & \rho_1 & 0 \\ \rho_1 & 1 & \rho_1 \\ 0 & \rho_1 & 1 \end{pmatrix} \begin{pmatrix} \phi_{31} \\ \phi_{32} \\ \phi_{33} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ 0 \\ 0 \end{pmatrix}$$

Passo 1 – Abrir o produto linha a linha:

$$\text{Linha 1: } \phi_{31} + \rho_1 \phi_{32} = \rho_1$$

$$\text{Linha 2: } \rho_1 \phi_{31} + \phi_{32} + \rho_1 \phi_{33} = 0$$

$$\text{Linha 3: } \rho_1 \phi_{32} + \phi_{33} = 0$$

Passo 2 – Isolar variáveis nas Linhas 1 e 3:

$$(\text{Linha 1}): \phi_{31} = \rho_1 - \rho_1 \phi_{32} \quad (\text{Linha 3}): \phi_{32} = -\frac{\phi_{33}}{\rho_1}$$

PACF do MA(1): cálculo passo a passo ($k = 3$)

Passo 3 – Combinar Linha 1 com Linha 3:

Substituindo $\phi_{32} = -\phi_{33}/\rho_1$ em ϕ_{31} :

$$\phi_{31} = \rho_1 - \rho_1 \cdot \left(-\frac{\phi_{33}}{\rho_1} \right) = \rho_1 + \phi_{33}$$

Passo 4 – Substituir ϕ_{31} e ϕ_{32} na Linha 2:

$$\rho_1(\rho_1 + \phi_{33}) - \frac{\phi_{33}}{\rho_1} + \rho_1 \phi_{33} = 0$$

$$\rho_1^2 + 2\rho_1 \phi_{33} - \frac{\phi_{33}}{\rho_1} = 0$$

Passo 5 – Multiplicar por ρ_1 e isolar ϕ_{33} :

$$\rho_1^3 + (2\rho_1^2 - 1)\phi_{33} = 0 \implies \boxed{\phi_{33} = \frac{\rho_1^3}{1 - 2\rho_1^2}}$$

Exemplo $\theta = 0,6$ ($\rho_1 \approx 0,441$): $\phi_{33} \approx \frac{(0,441)^3}{1 - 2(0,441)^2} = \frac{0,086}{0,611} \approx 0,141$

Por que a PACF do MA(1) nunca zera

Padrão observado:

$$\phi_{11} \approx 0,441, \quad \phi_{22} \approx -0,241, \quad \phi_{33} \approx 0,141, \quad \phi_{44} \approx -0,083, \quad \dots$$

Os valores decaem em módulo e *alternam de sinal*, mas nunca chegam ao zero exato.

Razão: o MA(1) pode ser reescrito como um AR de ordem *infinita*. Partindo de $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ e iterando $\varepsilon_{t-1} = y_{t-1} - \theta\varepsilon_{t-2}$:

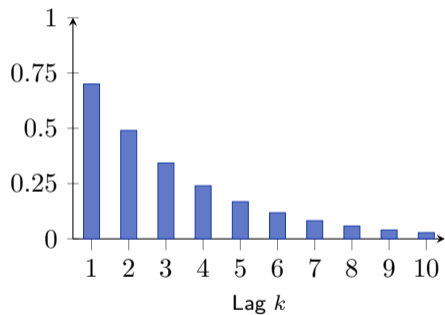
$$y_t = \theta y_{t-1} - \theta^2 y_{t-2} + \theta^3 y_{t-3} - \theta^4 y_{t-4} + \dots + \varepsilon_t$$

Implicação: como há infinitos lags com coeficiente não nulo no modelo verdadeiro, a PACF nunca corta – só decai.

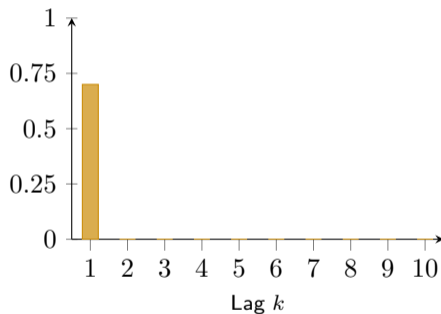
Simetria com o AR(1): no AR(1), a **ACF** é que decai (infinitos lags na representação MA(∞)). Aqui no MA(1), a **PACF** é que decai (infinitos lags na representação AR(∞)).

AR(1) com $\phi = 0,7$: ACF e PACF

ACF: $\rho_k = \phi^k$ (decai)



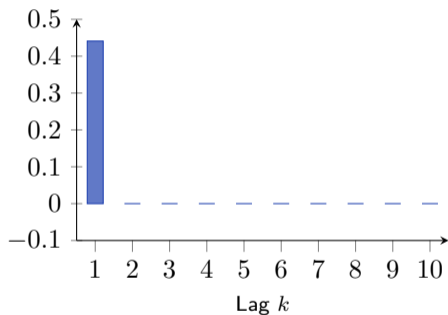
PACF: $\phi_{11} = \phi$, $\phi_{kk} = 0$



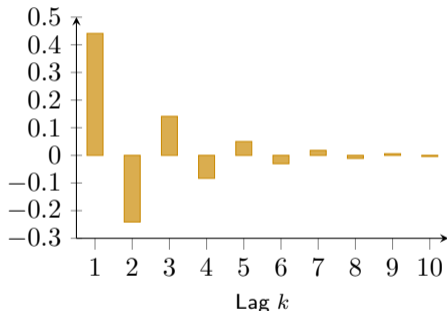
Assinatura de um AR(1): ACF decai gradualmente + PACF corta no lag 1.

MA(1) com $\theta = 0,6$: ACF e PACF

ACF: corta no lag 1



PACF: decai oscilando



Assinatura de um MA(1): ACF corta no lag 1 + PACF decai oscilando.

Como ler ACF e PACF: resumo

Processo	ACF	PACF
$AR(p)$	decai gradualmente	corta após o lag p
$MA(q)$	corta após o lag q	decai gradualmente
$ARMA(p, q)$	decai	decai
Ruído branco	tudo ≈ 0	tudo ≈ 0
Raiz unitária	decai <i> muito </i> devagar	lag 1 próximo de 1

Resumo:

- **AR**: identificado pela PACF (aponta a ordem p).
- **MA**: identificado pela ACF (aponta a ordem q).
- **ARMA**: ambas decaem \rightarrow comparar candidatos por AIC/BIC.

ACF/PACF em dados reais: bandas de confiança

Os cálculos acima dão a ACF e PACF **teóricas**. Em dados, usamos a **estimada**:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Sob $H_0: \rho_k = 0$: $\hat{\rho}_k \approx \mathcal{N}(0, 1/T)$. IC a 95%:

$$\pm \frac{1,96}{\sqrt{T}}$$

São as **linhas tracejadas** no gráfico. Lag *dentro* da banda \Rightarrow não rejeita $\rho_k = 0$.

Cuidados:

- Amostras pequenas ($T < 50$): até processos puros podem ter lags “significativos” por acaso.
- Lag isolado no 12 (mensal) ou 7 (diário): sugere **sazonalidade**, não MA puro.
- ACF que decai *muito* devagar: sugere **raiz unitária**. Diferencie antes.

Próxima aula: aplicar à etapa de **identificação** de Box-Jenkins.

Seção 4.1

A Metodologia de Box-Jenkins

Metodologia de Box-Jenkins: as 3 etapas

Proposta por Box & Jenkins (1976): **procedimento sistemático** para ajustar um modelo ARIMA a uma série temporal.

As 3 etapas

1. **Identificação** – definir a ordem (p, d, q) do modelo.
2. **Estimação** – estimar os coeficientes (ϕ, θ) dos candidatos e escolher o melhor.
3. **Diagnóstico** – verificar se os resíduos se comportam como ruído branco.

Observação: alguns autores acrescentam uma 4ª etapa de **previsão**. Tratamos dela separadamente no Cap. 5.

É um procedimento iterativo: se o diagnóstico falha, voltamos para a identificação e testamos outro modelo.

Obs.: Funções automáticas como `auto.arima` ajudam, mas não substituem julgamento. O procedimento exige treino.

Seção 4.2

Identificação do Modelo

Etapa 1: Identificação

Objetivo: determinar a ordem (p, d, q) do modelo ARIMA.

Procedimento (em 4 passos)

1. **Inspecionar** o gráfico da série: há tendência? sazonalidade?
2. **Testar raiz unitária** com ADF (Cap. 2). Se houver, fica definido o d .
3. **Diferenciar** a série d vezes até torná-la estacionária.
4. **Analisar ACF e PACF** da série estacionária para sugerir p e q .

Resultado: um ou mais modelos *candidatos* – ainda não escolhemos um único.

Não há um único modelo correto. Box-Jenkins é um procedimento de busca: escolhemos o que melhor ajusta, segundo critérios objetivos + julgamento.

Exemplo: identificando o ARIMA do log do PIB

Cenário: ajustar um modelo ARIMA ao log do PIB brasileiro (trimestral).

Passo 1 – Inspeccionar a série:

O log do PIB **crece** ao longo do tempo \Rightarrow presença de tendência.

Passo 2 – Testar raiz unitária (ADF):

ADF com constante e tendência. Resultado:

$$\tau = -1,78 \quad \text{vs.} \quad \tau_{\text{crítico}} = -2,88 \quad (5\%)$$

$|\tau| < |\tau_{\text{crítico}}| \Rightarrow$ **não rejeita** $H_0 \Rightarrow$ há raiz unitária.

Conclusão dos Passos 1-2: $d = 1$ (vamos diferenciar uma vez).

Exemplo: identificando o ARIMA do log do PIB

Passo 3 – Diferenciar:

Calculamos $\Delta y_t = y_t - y_{t-1}$ – a **taxa de crescimento** do PIB. Visualmente, agora a série flutua em torno de um nível constante.

Passo 4 – Analisar ACF e PACF de Δy_t :

- ACF: apenas os primeiros lags significativos.
- PACF: idem.

Candidatos sugeridos (mantendo $d = 1$):

- AR(1) na diferenciada \rightarrow ARIMA(1, 1, 0)
- MA(1) na diferenciada \rightarrow ARIMA(0, 1, 1)
- ARMA(1,1) na diferenciada \rightarrow ARIMA(1, 1, 1)

Próxima etapa: estimar todos e escolher o melhor por AIC/BIC.

Seção 4.3

Estimação do Modelo

Etapa 2: Estimação

Objetivo: estimar os coeficientes (ϕ, θ) de cada candidato e escolher o melhor.

Método de estimação:

- Modelo **só AR**: pode-se estimar por MQO.
- Modelo com componente **MA**: usar **máxima verossimilhança** (ML). Os ε_t não são observáveis, então MQO não se aplica.

Por que estimar vários modelos?

A ACF e PACF dão pistas, mas raramente uma resposta única. Estimamos várias combinações de p e q (até um máximo razoável) e comparamos.

Como comparar?

Usando **critérios de informação**: AIC e BIC.

AIC (*Akaike Information Criterion*)

$$\text{AIC}(p) = \log\left(\frac{\text{SQR}}{T}\right) + \frac{2(1+p)}{T}$$

em que SQR é a Soma dos Quadrados dos Resíduos, T é o tamanho da amostra e p é o número de parâmetros estimados.

Lendo os dois termos:

- 1º termo $\log(\text{SQR}/T)$: mede o **ajuste** aos dados. SQR menor \Rightarrow termo menor.
- 2º termo $2(1+p)/T$: **penaliza** modelos com muitos parâmetros.

Por que penalizar? Mais parâmetros sempre melhoram o ajuste *na amostra*, mas pioram a previsão fora dela (*overfitting*). A penalização força um equilíbrio.

Regra: entre candidatos, escolher o que **minimiza o AIC**.

BIC (*Bayesian Information Criterion*)

$$\text{BIC}(p) = \log\left(\frac{\text{SQR}}{T}\right) + \frac{\log(T)(1+p)}{T}$$

Única diferença para o AIC: a penalização é $\log(T)$ em vez de 2.

Comparação numérica:

- Para $T > 7$: $\log(T) > 2 \Rightarrow$ **BIC penaliza mais.**
- Para $T = 100$: $\log(100) \approx 4,6$ – penalidade $\sim 2,3\times$ a do AIC.
- Resultado: **BIC tende a selecionar modelos menores** que o AIC.

Na prática: reporte os dois e compare. Se discordam, BIC é preferido por sua maior parcimônia. AIC é mais comum em aplicações de previsão.

Parcimônia: o princípio central

Critério da parcimônia

Não inclua parâmetros além do necessário. Box & Jenkins argumentam que um modelo com poucas defasagens muitas vezes explica a série tão bem quanto um modelo de ordem alta.

Por quê?

- Mais parâmetros \Rightarrow melhor ajuste **na amostra** (sempre).
- Mais parâmetros \Rightarrow pior previsão **fora da amostra** (overfitting).
- Mais parâmetros \Rightarrow estimativas com mais variância.

Tradeoff

Modelos com mais parâmetros tendem a apresentar AIC/BIC *menor*, mas não necessariamente são preferíveis. O bom analista equilibra qualidade do ajuste com parcimônia.

Exemplo: PIB brasileiro – comparando candidatos

Voltando ao exemplo do log do PIB. Após diferenciar, testamos todos os $ARIMA(p, 0, q)$ com $p, q \in \{0, 1, 2, 3\}$. AIC obtido:

Modelo	AIC	Modelo	AIC	Modelo	AIC
(0,0,0)	-429,9	(1,0,1)	-437,7	(2,0,2)	-434,4
(0,0,1)	-435,5	(1,0,2)	-436,5	(3,0,0)	-435,7
(0,0,2)	-436,0	(1,0,3)	-434,7	(3,0,1)	-434,7
(0,0,3)	-436,1	(2,0,0)	-437,2	(3,0,2)	-439,1
(1,0,0)	-437,5	(2,0,1)	-436,7	(3,0,3)	-438,2

Lendo a tabela:

- $ARIMA(3, 0, 2)$ minimiza o AIC (-439,1) – mas tem 5 parâmetros.
- $ARIMA(1, 0, 0)$ ($AR(1)$) tem $AIC = -437,5$ – 1 só parâmetro.
- Diferença $\Delta AIC = 1,6$ justifica 4 parâmetros extras? **Não.**

Escolha: $ARIMA(1, 1, 0)$ sobre o log do PIB (lembre que $d = 1$).

Como decidir entre candidatos próximos

Regras práticas:

- Diferença **pequena** no AIC: prefira o mais parcimonioso.
- Diferença **grande** no AIC: prefira o de menor AIC, mesmo com mais parâmetros.
- Não existe regra exata para “pequena” vs. “grande”. Use julgamento.

Exemplo: 1 ponto de AIC com 4 parâmetros extras? Não vale.

Funções automáticas

R: `forecast::auto.arima(y)`

Python: `pmdarima.auto_arima(y)`

Testam várias combinações e retornam o de menor AIC (ou BIC).

Cuidado

Podem errar quando há quebra estrutural, sazonalidade não detectada ou outliers. Sempre **inspecione o resultado**.

Seção 4.4

Diagnóstico

Etapa 3: Diagnóstico

Objetivo: verificar se o modelo escolhido extraiu toda a informação relevante da série.

Como verificar?

Analisando os **resíduos**. Se o modelo é adequado, os resíduos devem se comportar como **ruído branco**: média zero, variância constante e **sem autocorrelação**.

Duas ferramentas:

- **Visual:** ACF dos resíduos – nenhum lag deve sair da banda de confiança.
- **Formal:** teste de **Ljung-Box** – testa H_0 de não autocorrelação.

Se os resíduos ainda têm estrutura, o modelo não capturou tudo. Volte para a identificação e tente outro modelo.

Hipóteses

H_0 : resíduos são i.i.d. (sem autocorrelação)

H_1 : há autocorrelação serial nos resíduos

Estatística do teste (até o lag k):

$$Q(k) = T(T + 2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{T - j}$$

em que $\hat{\rho}_j$ é a autocorrelação estimada dos resíduos no lag j .

Distribuição sob H_0 : $Q(k) \sim \chi_k^2$ aproximadamente.

Regra de decisão:

- $Q > \chi_{1-\alpha, k}^2$ (ou valor- $p < \alpha$): **rejeita** $H_0 \Rightarrow$ há autocorrelação \Rightarrow modelo inadequado.
- Caso contrário: não rejeita \Rightarrow modelo passa no diagnóstico.

Exemplo: diagnóstico do ARIMA(1,1,0) ajustado ao PIB

Aplicando o teste de Ljung-Box aos resíduos com $k = 7$ lags:

Estatística Q	7,66
Graus de liberdade	7
Valor- p	0,36

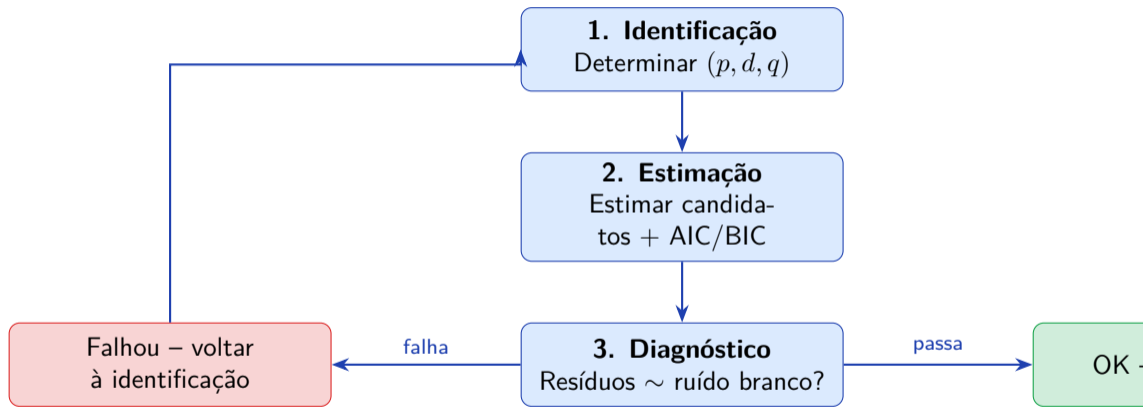
Conclusão: valor- $p = 0,36 > 0,05 \Rightarrow$ **não rejeita** H_0 . Não há evidência de autocorrelação nos resíduos – o modelo passou no diagnóstico.

Outras verificações úteis:

- **Normalidade:** teste de Shapiro-Wilk (importante para IC nas previsões).
- **Inspeção gráfica:** ACF dos resíduos, histograma, gráfico Q-Q.

Modelo final: ARIMA(1, 1, 0) sobre o log do PIB.

Resumo da metodologia



Box-Jenkins é iterativo: se o diagnóstico falha, volte e escolha outro modelo.

Exercícios

Enunciado

A metodologia de Box-Jenkins auxilia a identificar qual processo melhor explica a dinâmica de uma série temporal. Ela é composta pelas etapas:

P – Estimação Q – Diagnóstico R – Previsão S – Identificação

A sequência correta, da primeira para a última, é:

- (a) P, Q, S, R
- (b) P, S, R, Q
- (c) P, R, Q, S
- (d) S, P, Q, R
- (e) S, P, R, Q

Resposta correta: (d) S, P, Q, R

A ordem correta é:

1. **S – Identificação:** determinar (p, d, q) via inspeção, raiz unitária e ACF/PACF.
2. **P – Estimação:** estimar os candidatos e escolher via AIC/BIC.
3. **Q – Diagnóstico:** verificar se os resíduos são ruído branco.
4. **R – Previsão:** (4ª etapa para alguns autores, vista no Cap. 5).

Enunciado

Ajustou-se um AR(2) a uma série de 80 registros mensais:

Parâmetro	Estimativa	Erro-padrão	Estat. t
AR(1)	0,688059	0,09936	6,924
AR(2)	-0,474084	0,10014	-4,733
Constante	39,4475		

Qual o modelo ajustado?

- (a) $Z_t = 39,4475 - 0,474084 Z_{t-2}$
- (b) $Z_t = -0,474084 Z_{t-2}$
- (c) $Z_t = 39,4475 + 0,688059 Z_{t-1}$
- (d) $Z_t = 0,688059 Z_{t-1}$
- (e) $Z_t = 39,4475 + 0,688059 Z_{t-1} - 0,474084 Z_{t-2}$

Resposta correta: (e)

Como é um modelo AR(2), a forma geral é:

$$Z_t = c + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \varepsilon_t$$

Substituindo os coeficientes estimados (com $c = 39,4475$, $\phi_1 = 0,688059$, $\phi_2 = -0,474084$):

$$Z_t = 39,4475 + 0,688059 Z_{t-1} - 0,474084 Z_{t-2}$$

Enunciado

Após remover tendência e sazonalidade, ajustou-se um AR(2) a Y_t :

$$Y_t = 0,4Y_{t-1} - 0,2Y_{t-2} + \varepsilon_t$$

Analise as afirmativas:

- I. As raízes do polinômio característico são complexas, e o modelo **não é invertível**.
- II. O modelo Y_t é **estacionário**.
- III. Suponha $Y_{49} = 1$ e $Y_{50} = 0,75$. A previsão para $t = 51$ é **0,1**.

Item I – F.

“Invertibilidade” é conceito de **MA**, não de AR. Para um AR, falamos de estacionariedade, não de invertibilidade. Item incorreto pelo próprio conceito.

Item II – V.

Passo 1 – Escrever o polinômio característico:

$$1 - 0,4z + 0,2z^2 = 0$$

Passo 2 – Calcular as raízes:

$$z = \frac{0,4 \pm \sqrt{0,16 - 0,8}}{0,4} = \frac{0,4 \pm \sqrt{-0,64}}{0,4} = 1 \pm 2i$$

Passo 3 – Verificar se estão fora do círculo unitário:

$$|z| = \sqrt{1^2 + 2^2} = \sqrt{5} \approx 2,24 > 1 \checkmark$$

⇒ raízes fora do círculo unitário ⇒ **processo estacionário.**

Item III – V.

Passo 1 – Identificar a equação de previsão:

$$\mathbb{E}(Y_{51} | Y_{50}, Y_{49}) = 0,4 Y_{50} - 0,2 Y_{49}$$

(o termo ε_{51} tem esperança zero).

Passo 2 – Substituir os valores:

$$\mathbb{E}(Y_{51}) = 0,4 \cdot 0,75 - 0,2 \cdot 1 = 0,3 - 0,2 = \boxed{0,1}$$

Conclusão: II e III estão corretas.

Enunciado

Considere a tabela de produção anual do bem A:

Ano	Produção	MM3
2016	20	-
2017	22	W
2018	18	X
2019	23	Y
2020	17	Z
2021	20	-

Os valores de W e Y (médias móveis de 3 anos) são, respectivamente:

- (a) 20,0 e 19,3 (b) 20,2 e 20,0 (c) 19,3 e 21,0 (d) 20,0 e 20,2 (e) 18,3 e 21,0

Resposta correta: (a) 20,0 e 19,3

Cálculo de W (média móvel centrada em 2017):

$$W = \frac{20 + 22 + 18}{3} = \frac{60}{3} = 20,0$$

Cálculo de Y (média móvel centrada em 2019):

$$Y = \frac{18 + 23 + 17}{3} = \frac{58}{3} \approx 19,3$$

Lembrete: a média móvel **centrada** usa o ponto t e seus vizinhos $t - 1$ e $t + 1$. Não é usada para previsão – só para suavização (como visto no Cap. 3).

Resumo do Capítulo 4

Conceitos cobertos

- ACF e PACF (revisão e leitura)
- Metodologia de Box-Jenkins em 3 etapas
- **Identificação:** testes de raiz unitária, ACF/PACF
- **Estimação:** AIC, BIC e parcimônia
- **Diagnóstico:** resíduos como ruído branco, Ljung-Box

Para a próxima aula – Cap. 5

- Previsão com modelos ARIMA
- Iteração da equação de previsão
- Previsão dentro vs. fora da amostra
- Medidas de desempenho (RMSE, MAE)

Takeaway: Box-Jenkins é **iterativo**. Identifique, estime, diagnostique. Se falhar, volte. Use AIC/BIC, mas com julgamento.

Dúvidas?

Próxima aula: Previsão (Cap. 5)