

Lista 2

Econometria II – UFBA | Prof. Pablo Castro
Capítulos 4, 5 e 6: Box-Jenkins, Previsão, ARDL e Cointegração

Notação padrão: $\{\varepsilon_t\}$ (ou $\{u_t\}$) é ruído branco com $\mathbb{E}[\varepsilon_t] = 0$ e $\text{Var}(\varepsilon_t) = \sigma^2$; operador defasagem L : $Ly_t = y_{t-1}$.

Questão 1. Cálculo da ACF e PACF – AR(2) e MA(2).

Nas notas do capítulo 4, vimos o cálculo da ACF e PACF do AR(1) e MA(1). Nesse exercício vamos fazer para os casos de ordem 2. Considere os dois processos estacionários abaixo, com $\varepsilon_t \sim \text{RB}(0, \sigma^2)$:

$$\text{AR}(2): y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \text{MA}(2): y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}.$$

Ambos os processos tem $E[y_t] = 0$. Relembre: para calcular a ACF, basta calcular $\gamma_0 = \text{Var}(y_t)$, depois $\gamma_k = \mathbb{E}[y_t y_{t-k}]$ (usando $\mathbb{E}[y_t] = 0$), e dividir $\rho_k = \gamma_k / \gamma_0$. Para calcular a PACF de ordem k , resolva o sistema de Yule-Walker de ordem k .

(a) ACF do AR(2).

- (i) **Variância.** Aplique $\text{Var}(\cdot)$ nos dois lados de $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, usando a independência entre ε_t e os lags de y e a estacionariedade ($\text{Var}(y_t) = \text{Var}(y_{t-j}) = \gamma_0$), para obter a fórmula de γ_0 .
- (ii) **Autocovariâncias.** Calcule $\gamma_k = \mathbb{E}[y_t y_{t-k}]$ substituindo $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$ e usando que $\mathbb{E}[\varepsilon_t y_{t-k}] = 0$ para $k \geq 1$. Mostre que, para $k \geq 1$:

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}.$$

Depois divida por γ_0 para obter a recursão da ACF.

- (iii) **Condições iniciais.** Use $k = 1$ e a simetria $\rho_{-1} = \rho_1$ para mostrar que

$$\rho_1 = \frac{\phi_1}{1 - \phi_2},$$

e então use $k = 2$ para obter $\rho_2 = \phi_2 + \phi_1 \rho_1$.

- (iv) **Valores numéricos.** Para $\phi_1 = 1,0$ e $\phi_2 = -0,5$, calcule ρ_1 , ρ_2 , ρ_3 e ρ_4 usando a recursão.

(b) PACF do AR(2). Para cada ordem k , a PACF ϕ_{kk} é o **último coeficiente** da solução do sistema de Yule-Walker $P_k \phi_k = \rho_k$.

- (i) **Lag 1.** Obtenha a expressão do sistema 1×1
- (ii) **Lag 2.** Monte e resolva o sistema 2×2 :

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

Mostre que $\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ e calcule numericamente.

- (iii) **Lag 3.** Monte o sistema 3×3 com os valores de ρ_1 , ρ_2 , ρ_3 do item (a) e resolva para ϕ_{33} . O que você obtém? Interprete.

(c) ACF do MA(2).

- (i) Calcule $\gamma_0 = \text{Var}(y_t)$ usando a independência dos ε . (Dica: expanda $\text{Var}(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2})$ e use $\text{Cov}(\varepsilon_t, \varepsilon_s) = 0$ para $t \neq s$.)
- (ii) Calcule $\gamma_1 = \mathbb{E}[y_t y_{t-1}]$ e $\gamma_2 = \mathbb{E}[y_t y_{t-2}]$ expandindo o produto $(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}) \times (\varepsilon_{t-k} + \dots)$ e identificando quais termos têm o mesmo índice de ε (esses sobrevivem; os outros são zero).
- (iii) Mostre que $\gamma_k = 0$ para $k \geq 3$.
- (iv) Compute ρ_1 e ρ_2 numericamente para $\theta_1 = 0,7$ e $\theta_2 = -0,4$.

(d) PACF do MA(2). Use os valores de ρ_1 e ρ_2 do item (c) e $\rho_k = 0$ para $k \geq 3$. Calcule ϕ_{11} , ϕ_{22} e ϕ_{33} montando e resolvendo os sistemas de Yule-Walker de ordens 1, 2 e 3.

(e) Tabela comparativa. Complete a tabela com os resultados e explique em uma frase o padrão de cada coluna.

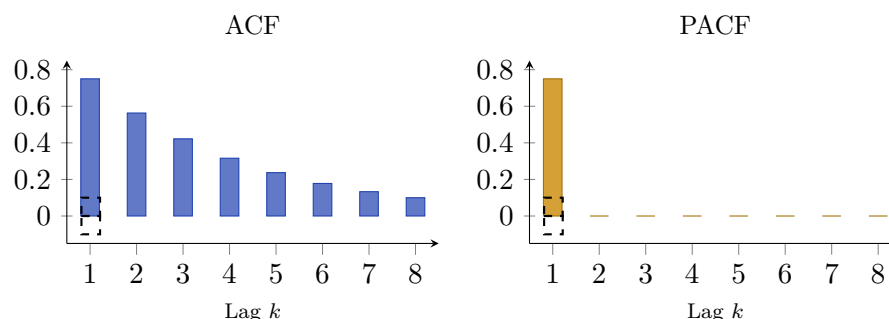
	AR(2) ($\phi_1=1,0$, $\phi_2=-0,5$)		MA(2) ($\theta_1=0,7$, $\theta_2=-0,4$)	
Lag k	ρ_k	ϕ_{kk}	ρ_k	ϕ_{kk}
1				
2				
3				
4				

Questão 2. ACF e PACF – identificação gráfica.

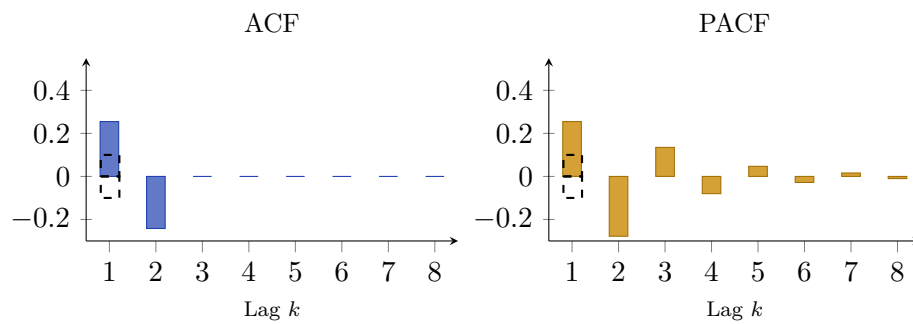
Cada par de gráficos abaixo mostra a ACF (barras azuis) e a PACF (barras douradas) de um processo estacionário gerado por simulação. As linhas tracejadas representam bandas de confiança a 95% ($\pm 1,96/\sqrt{T}$).

Para cada caso, identifique o tipo de processo mais provável ($\text{AR}(p)$, $\text{MA}(q)$, $\text{ARMA}(p, q)$ ou ruído branco) e **justifique** sua escolha com base no comportamento da ACF e da PACF.

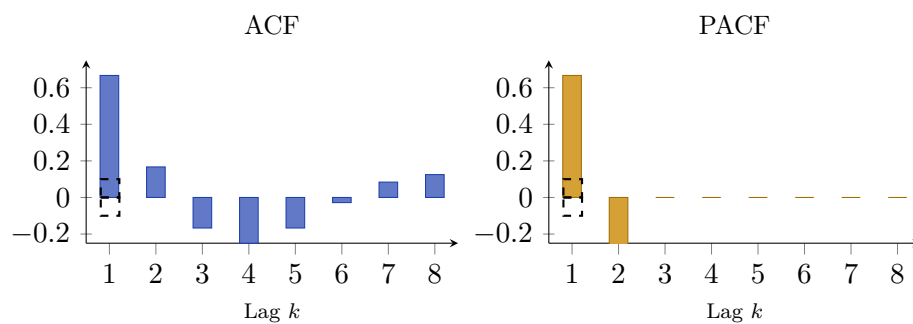
Caso (a):



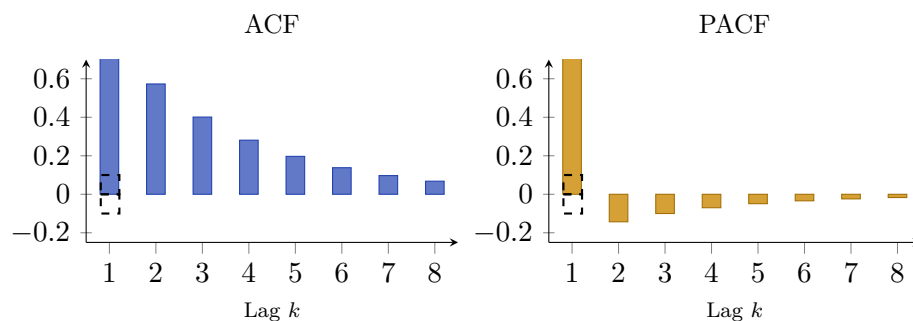
Caso (b):



Caso (c):



Caso (d):



Questão 3. Metodologia de Box-Jenkins.

- (a) Descreva as **três etapas** da metodologia de Box-Jenkins. Para cada etapa, explique o objetivo e as ferramentas utilizadas.
- (b) Um pesquisador ajusta uma série e obtém os seguintes valores de AIC para modelos $\text{ARIMA}(p, 1, q)$ com $p, q \in \{0, 1, 2\}$:

Modelo	AIC	Modelo	AIC	Modelo	AIC
ARIMA(0,1,0)	-310,2	ARIMA(1,1,1)	-318,7	ARIMA(2,1,1)	-319,2
ARIMA(0,1,1)	-316,4	ARIMA(1,1,2)	-317,9	ARIMA(2,1,2)	-318,0
ARIMA(1,1,0)	-317,5	ARIMA(2,1,0)	-316,8	ARIMA(0,1,2)	-315,1

Qual modelo você selecionaria? Justifique com base no princípio da **parcimônia** e no AIC.

- (c) Explique a diferença entre AIC e BIC. Para a mesma amostra de tamanho $T = 200$, qual deles tende a selecionar um modelo com **menos** parâmetros? Calcule a penalidade por parâmetro adicional em cada critério para $T = 200$.

Questão 4. Diagnóstico – Ljung-Box.

Após ajustar um ARIMA(1, 1, 0) a uma série com $T = 120$ observações, os resíduos apresentam as autocorrelações estimadas:

$$\hat{\rho}_1 = 0,08, \quad \hat{\rho}_2 = -0,06, \quad \hat{\rho}_3 = 0,11, \quad \hat{\rho}_4 = 0,03, \quad \hat{\rho}_5 = -0,09.$$

- (a) Escreva a estatística de Ljung-Box $Q(5)$ e calcule seu valor numérico.
- (b) Sob H_0 de ausência de autocorrelação, $Q(5) \sim \chi^2_5$ (aproximadamente). O valor crítico da χ^2 a 5% com 5 graus de liberdade é 11,07. O modelo passa no diagnóstico?
- (c) O que você faria se o modelo **não passasse** no diagnóstico? Descreva o procedimento.

Questão 5. Previsão com AR(1) – recursão e convergência.

Seja o processo AR(1) estimado:

$$y_t = 3 + 0,7 y_{t-1} + u_t, \quad u_t \sim \text{RB}(0, \sigma^2).$$

A última observação disponível é $y_T = 20$.

- (a) Calcule $\hat{y}_{T+1}$, $\hat{y}_{T+2}$ e $\hat{y}_{T+3}$ usando a recursão $\hat{y}_{T+h} = 3 + 0,7 \hat{y}_{T+h-1}$.
- (b) Qual é a média de longo prazo $\bar{y}$ do processo? Verifique que $\hat{y}_{T+h} \rightarrow \bar{y}$ quando $h \rightarrow \infty$.
- (c) Mostre algebricamente que a fórmula geral é

$$\hat{y}_{T+h} = \frac{3}{1 - 0,7} (1 - 0,7^h) + 0,7^h y_T.$$

Use essa fórmula para calcular $\hat{y}_{T+5}$ e $\hat{y}_{T+10}$; compare com $\bar{y}$.

- (d) Calcule a **variância do erro de previsão** $\text{Var}(e_{T+h})$ para $h = 1, 2, 3$. Mostre que ela cresce com h e, para $h \rightarrow \infty$, converge para $\text{Var}(y_t) = \sigma^2/(1 - 0,49)$.

Questão 6. Previsão recursiva do ARMA(1,1).

Considere o modelo ARMA(1,1) com constante:

$$y_t = \mu + \phi y_{t-1} + u_t + \theta u_{t-1}, \quad |\phi| < 1.$$

Regra de previsão ótima (esperança condicional):

- Valores **passados** y_{T-j} ($j \geq 0$): substituir pelo valor observado.
- Erros **passados** u_{T-j} ($j \geq 0$): substituir pelo resíduo estimado $\hat{u}_{T-j}$.
- Erros **futuros** u_{T+j} ($j \geq 1$): substituir por zero ($\mathbb{E}[u_{T+j}|\Omega_T] = 0$).
- Valores **futuros** y_{T+j} ($j \geq 1$): substituir pela previsão $\hat{y}_{T+j}$.

- (a) Mostre que $\hat{y}_{T+1} = \mu + \phi y_T + \theta \hat{u}_T$ e que $\hat{y}_{T+h} = \mu + \phi \hat{y}_{T+h-1}$ para $h \geq 2$. Explique **em uma frase** por que o componente MA desaparece a partir de $h = 2$.
- (b) Usando a recursão do item (a), expresse **explicitamente** $\hat{y}_{T+2}$ e $\hat{y}_{T+3}$ em função apenas de y_T , $\hat{u}_T$, μ e ϕ .

- (c) Generalize: mostre que para qualquer $h \geq 1$,

$$\hat{y}_{T+h} = \frac{\mu(1 - \phi^h)}{1 - \phi} + \phi^h y_T + \phi^{h-1} \theta \hat{u}_T.$$

Dica: use indução ou itere diretamente a recursão $\hat{y}_{T+h} = \mu + \phi \hat{y}_{T+h-1}$ partindo de $\hat{y}_{T+1}$.

- (d) O que acontece com $\hat{y}_{T+h}$ quando $h \rightarrow \infty$? Compare com o AR(1) puro ($\theta = 0$).

Questão 7. Erro de previsão do ARMA(1,1).

Continuando com o modelo ARMA(1,1): $y_t = \mu + \phi y_{t-1} + u_t + \theta u_{t-1}$.

Defina o erro de previsão $e_{T+h} = y_{T+h} - \hat{y}_{T+h}$.

- (a) Mostre que $e_{T+1} = u_{T+1}$. *Dica:* escreva y_{T+1} e subtraia $\hat{y}_{T+1}$, usando que $\hat{u}_T \approx u_T$.
- (b) Mostre que $e_{T+2} = (\phi + \theta) u_{T+1} + u_{T+2}$. *Dica:* use $y_{T+2} = \mu + \phi y_{T+1} + u_{T+2} + \theta u_{T+1}$ e subtraia $\hat{y}_{T+2} = \mu + \phi \hat{y}_{T+1}$.
- (c) Mostre que $e_{T+3} = \phi(\phi + \theta) u_{T+1} + (\phi + \theta) u_{T+2} + u_{T+3}$.
- (d) A partir do padrão dos itens (a)–(c), deduza a expressão geral para e_{T+h} para $h \geq 2$.
- (e) Calcule $\mathbb{E}[e_{T+h}]$ e $\text{Var}(e_{T+h})$ (para $h \geq 2$). Mostre que a previsão é **não viesada** e que $\text{Var}(e_{T+h})$ é crescente em h .
- (f) **Comparação.** Escreva $\text{Var}(e_{T+h})$ para o AR(1) puro ($\theta = 0$) e compare com o ARMA(1,1). Para $\phi > 0$ e $\theta > 0$, qual dos dois tem *maior* incerteza de previsão para $h \geq 2$? Interprete.

Questão 8. Medidas de desempenho – MAE, RMSE, MAPE.

Dois modelos foram ajustados a uma série mensal do IPCA. As previsões e os valores realizados estão na tabela abaixo:

h	Modelo A (ARIMA(1,1,0))			Modelo B (SARIMA)		
	y_{T+h}	$\hat{y}_{T+h}^A$	e_h^A	y_{T+h}	$\hat{y}_{T+h}^B$	e_h^B
1	0,62	0,50		0,62	0,58	
2	0,44	0,50		0,44	0,47	
3	0,71	0,50		0,71	0,65	
4	0,38	0,50		0,38	0,41	
5	0,55	0,50		0,55	0,52	

- (a) Complete a coluna de erros e calcule o MAE, o RMSE e o MAPE para cada modelo.
- (b) Qual modelo você preferiria? Justifique usando mais de uma medida.

Questão 9. ARDL: efeitos de curto e longo prazo.

Estimou-se a seguinte equação para o consumo das famílias brasileiras (C_t , em bilhões de reais) em função da renda disponível (Y_t , em bilhões de reais), dados trimestrais, $T = 80$:

$$\hat{C}_t = 8,5 + 0,40 C_{t-1} + 0,30 Y_t + 0,20 Y_{t-1}.$$

- (a) Identifique este modelo como ARDL(p, m) e indique as ordens p e m .
- (b) Qual é o **efeito de curto prazo** ($\partial C_t / \partial Y_t$)? Interprete economicamente.
- (c) Calcule o **efeito de longo prazo** (ELP) de uma variação permanente de 1 real na renda sobre o consumo.

- (d) Interprete a diferença entre o ECP e o ELP. Qual é o **multiplicador de longo prazo**?

Questão 10. Cointegração e Engle-Granger.

- (a) (*Conceito*) Defina cointegração. Qual é a diferença entre duas séries $I(1)$ **cointegradas** e duas séries $I(1)$ **não cointegradas**? Dê um exemplo econômico de cada caso.
- (b) (*Engle-Granger*) Descreva o procedimento de Engle-Granger em 3 passos para testar cointegração entre y_t e x_t , ambas $I(1)$. Por que os valores críticos do ADF aplicado aos resíduos são **mais negativos** que os do ADF padrão?
- (c) (*V/F com justificativa*) Classifique cada afirmação como Verdadeira (V) ou Falsa (F):
- (i) Se $y_t \sim I(1)$ e $x_t \sim I(1)$ e são cointegradas, os resíduos da regressão $y_t = \mu + \eta x_t + u_t$ são $I(0)$.
 - (ii) Toda regressão envolvendo duas séries $I(1)$ é espúria.
 - (iii) Se $y_t \sim I(1)$ e $x_t \sim I(2)$, pode haver cointegração entre elas.
 - (iv) A regressão de Δy_t em Δx_t é sempre válida, mesmo se y_t e x_t são cointegradas – apenas perdemos informação de longo prazo.
 - (v) Para testar cointegração entre **três** variáveis, o procedimento de Engle-Granger com os mesmos valores críticos de duas variáveis é suficiente.
- (d) (*Numérico*) Uma pesquisadora obteve os seguintes resultados ADF. A última coluna representa o valor crítico relevante para o teste de hipótese:

Série	Estatística τ	VC a 5%
p_t (nível)	-1,42	-2,87
m_t (nível)	-1,81	-2,87
Δp_t	-8,34	-2,87
Δm_t	-7,11	-2,87
$\hat{u}_t$ (resíduos de $p_t = \mu + \eta m_t + u_t$)	-3,51	-3,34

Conclua: (i) as séries são $I(1)$? (ii) há cointegração?

Questão 11. Modelo de Correção de Erros – VECM.

Suponha que as séries C_t (consumo) e Y_t (renda), ambas $I(1)$, sejam cointegradas. Os dois estágios de Engle-Granger produzem:

Estágio 1	$\hat{C}_t = 0,15 + 0,82 Y_t + \hat{u}_t$
Estágio 2	$\Delta \hat{C}_t = 0,03 + 0,55 \Delta Y_t - 0,30 \hat{u}_{t-1} + \varepsilon_t$

- (a) Interprete os coeficientes do estágio 1: o que significa $\hat{\eta} = 0,82$?
- (b) Interprete os coeficientes do estágio 2:
- $\hat{\eta}_0 = 0,55$: interprete esse coeficiente.
 - $\hat{\lambda} = -0,30$: interprete esse coeficiente. Por que ele é negativo?
- (c) Suponha que no período $t - 1$, o consumo estava **acima** do equilíbrio de longo prazo em 2 bilhões de reais ($\hat{u}_{t-1} = 2$). Calcule a contribuição do termo de correção de erros para ΔC_t .
- (d) Mostre algebricamente que o VECM pode ser derivado do ARDL(1, 1) com coeficientes ϕ_1 , η_0 e η_1 . *Dica*: subtraia C_{t-1} dos dois lados e reescreva.