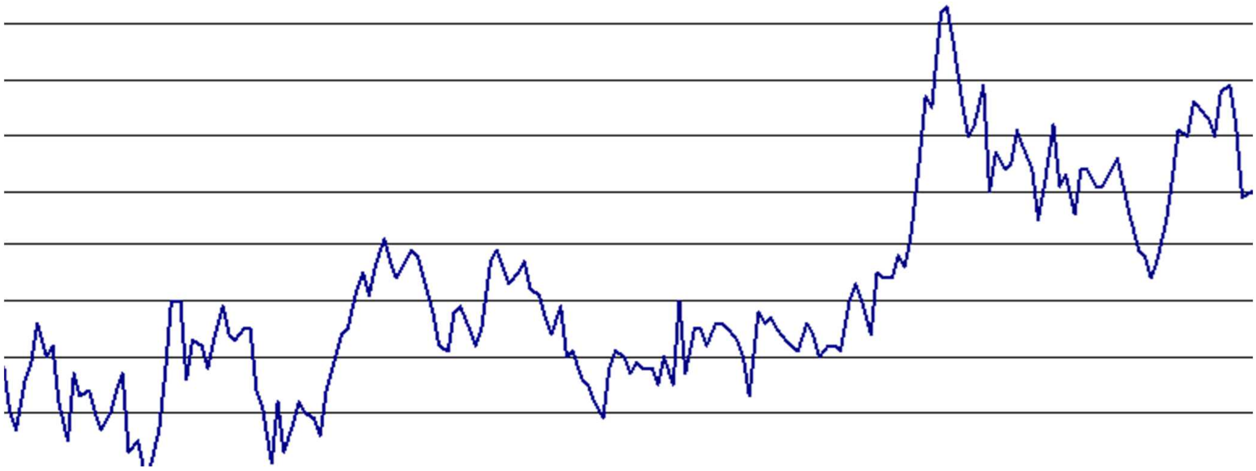


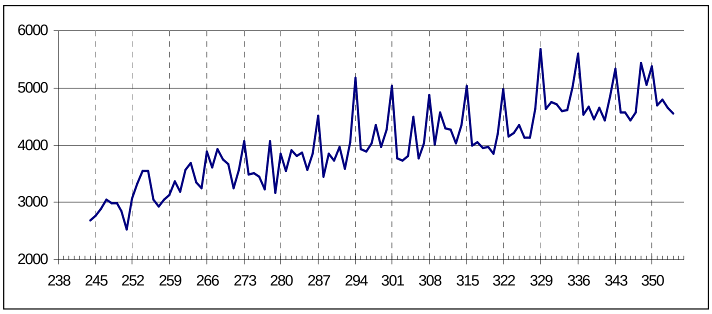
Séries Temporais

Nuno Fidalgo

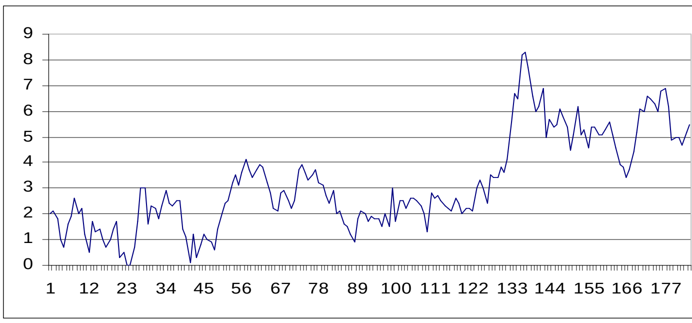
Introdução

- Metodologia clássica popular para a previsão a curto prazo.
- Previsão dos futuros valores da série temporal com base nos valores passados da própria variável e dos seus erros.
- A metodologia adotada em TPRE também se designa por **Box-Jenkings** (modelos ARIMA)
- ARIMA –
  - Auto-regressivos (AR),
  - integrados (I) e
  - média móvel (MA)

Consumos de gás em Lisboa

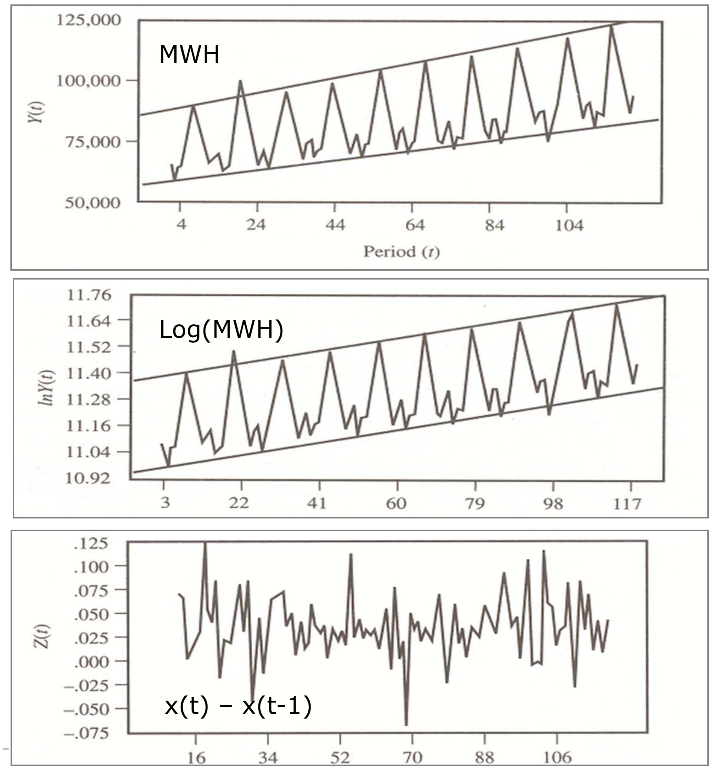


Produção de um parque eólico



Requisitos básicos

- Muitos modelos requerem estacionaridade. Série **estacionária** – média e variância constantes
- Se existir uma tendência ("trend") pode ajustar-se o desvio por uma curva, subtraindo o valor da curva à série (métodos de decomposição).
- Se a variância não for constante extrair o logaritmos ou uma potência da série
- Diferenciar a série pode levar a uma série estacionária
  - ver métodos integrativos ARIMA(0,d,0).



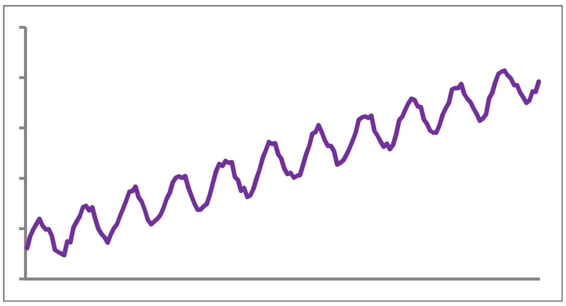
Série original:
x(t)

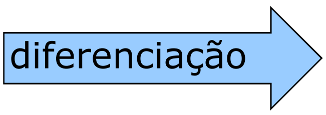
Série diferenciada:
x'(t) = x(t) – x(t-1)

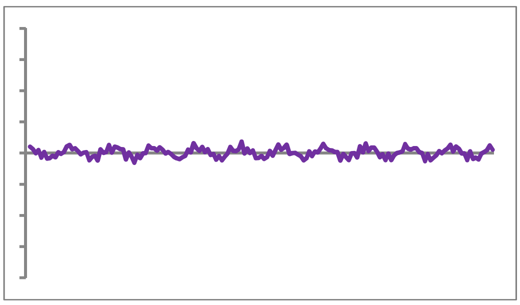
Diferenciação

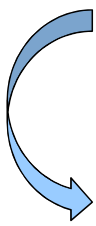
Série original:
x(t)

Série diferenciada:
x'(t) = x(t) – x(t-1)

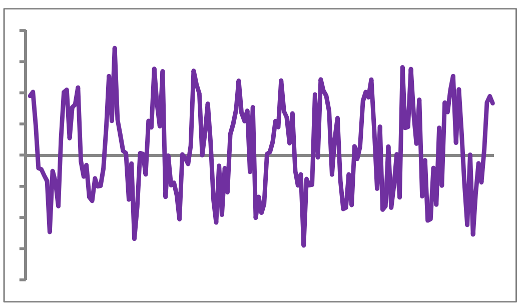








Aumento
de escala



Diferenciação – conceito e notação

- Diferenciação de primeira ordem para uma primeira diferença

$$X'_t = (1-B)X_t = X_t - X_{t-1}$$

- Diferenciação de primeira ordem para uma segunda diferença

$$X'_t = (1-B^2)X_t = X_t - X_{t-2}$$

- Diferenciação de segunda ordem

$$X'_t = (1-B)^2 X_t = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$

Modelos

- **Modelos auto-regressivos (AR) ou ARIMA(p,0,0)**
  - O valor presente $\mathbf{X_t}$ é uma função linear dos valores passados $\mathbf{X_{t-p}}$
  - A ordem da auto-regressão depende do valor mais antigo ***p***, que corresponde ao número de variáveis da regressão.
  - $\phi_{\mathbf{n}}$ são os coeficientes de regressão, constantes e reais. para encontrar estes valores podem ser usadas técnicas de mínimos quadrados.

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t$$

$$\delta + a_t = \left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) X_t$$

Modelos

- **Modelos de média móvel (MA) ou ARIMA(0,0,q)**
  - O valor presente $\mathbf{X_t}$ é uma função linear dos valores presente e passados dos resíduos (erros ou ruídos) $\mathbf{a_{t-q}}$
    - Diferente dos modelos AR que dependem dos valores passados da série $\mathbf{X_{t-p}}$
  - A ordem do modelo depende do valor mais antigo do erro ***q***, que é o número de variáveis do modelo.
  - $\boldsymbol{\theta_m}$ são os coeficientes de regressão. O sinal negativo é apenas uma questão de convenção.

$$X_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

$$X_t = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) a_t$$

Modelos

- Modelos mistos ARMA ou ARIMA(p,0,q)
  - O valor presente X_t é uma função linear dos valores passados da série $\mathbf{X_{t-1}}$ .. $\mathbf{X_{t-p}}$ e dos valores passados dos erros $\mathbf{a_{t-1}}$ .. $\mathbf{a_{t-q}}$
  - A ordem do modelo depende do valor mais antigo dos elementos da série $\mathbf{X_{t-p}}$ e do erro $\mathbf{a_{t-q}}$, ou seja, depende de ***p*** e de ***q***
  - $\boldsymbol{\phi_n}$ e $\boldsymbol{\theta_m}$ são os coeficientes de regressão

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) X_t = \delta + \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) a_t$$

Modelos

- Modelos mistos ARIMA ou ARIMA(p,d,q)
  - Quando a série não é estacionária recorre-se à diferenciação de ordem d…
  - A ordem da auto-regressão depende do valor mais antigo dos elementos da série *p* e do erro *q* e da ordem de diferenciação *d*

$$\underbrace{(1-B)^d}_{I(d)}\underbrace{\left(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p\right)}_{\text{autoregres siva AR(p)}}\underbrace{\left(1-\Phi_1 B^7\right)}_{\text{sasonal AR}}X_t=\delta+\underbrace{\left(1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q\right)}_{\text{médiamóvel MA(q)}}\underbrace{\left(1-\Theta_1 B^{30}\right)}_{\text{sasonal MA}}a_t$$

Modelos – exemplos

- ARIMA(1,0,1)

$$\boxed{(1-\phi_1 B)X_t=\delta+(1-\theta_1 B)a_t} \quad \Leftrightarrow \quad \boxed{X_t=\delta+\phi_1 X_{t-1}+a_t-\theta_1 a_{t-1}}$$

- ARIMA(1,1,1)

$$\boxed{\underbrace{(1-\phi_1 B)}_{\text{AR(1)}}\underbrace{(1-B)}_{\text{Diferenciação de 1ª ordem}}X_t=\delta+\underbrace{(1-\theta_1 B)\,a_t}_{\text{MA(1)}}} \quad \Leftrightarrow \quad \boxed{X_t=\delta+(1+\phi_1)X_{t-1}-\phi_1 X_{t-2}+a_t-\theta_1 a_{t-1}}$$

Cálculo dos parâmetros do modelo

Solver

- Como determinar os parâmetros da expressão

$$X_t = \delta + (1 + \phi_1) X_{t-1} - \phi_1 X_{t-2} + a_t - \theta_1 a_{t-1}$$

Pode ignorar-se ao estimar os parâmetros

- Organizar variáveis na folha de cálculo
  - Atenção aos atrasos X_{t-1}, X_{t-2},… a_{t-1}, a_{t-2}, …
- Inicializar parâmetros (variáveis do Solver): δ, ϕ_1, θ_1
- Calcular X_t (estimado)
  - Considerar $a_t = 0$
- Calcular erro quadrático e soma dos erros quadráticos
- Grandeza a minimizar: soma dos erros quadráticos

Cálculo dos parâmetros do modelo

- Exemplo

$$X_t = \delta + (1 + \phi_1) X_{t-1} - \phi_1 X_{t-2} + a_t - \theta_1 a_{t-1}$$

δ	φ1	θ1
0.5040	-0.0445	1.6528

Parâmetros de regressão (variáveis a alterar no Solver)

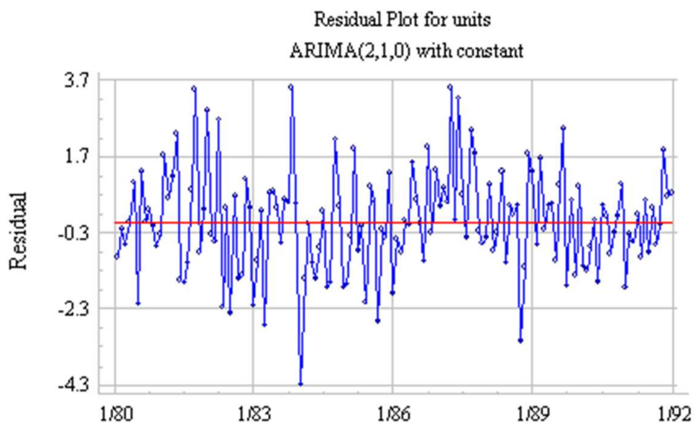
xt medido	xt-1	xt-2	at	at-1	xt estimado	at^2
100						
90	100					
87.02	90	100				
85.94	87.02	90	-1.72	0	87.66	2.95
86.26	85.94	87.02	-3.08	-1.72	89.33	9.46
94.93	86.26	85.94	3.10	-3.08	91.83	9.62
95.43	96.74	97.62	-0.83	0.62	96.26	0.69
97.64	95.43	96.74	0.28	-0.83	97.36	0.08
					soma (at^2)	47.23

Determinado com base na fórmula (ignorando at)

Grandeza a minimizar (Solver)

Identificação dos modelos ARIMA

- Os slides anteriores mostram como ajustar os parâmetros da regressão dos modelos ARIMA
- Apenas falta saber <u>como identificar os modelos</u>, isto é, como estabelecer a equação de base, identificando p, d e q de **ARIMA(p,d,q)**



Identificação dos modelos – conceitos base

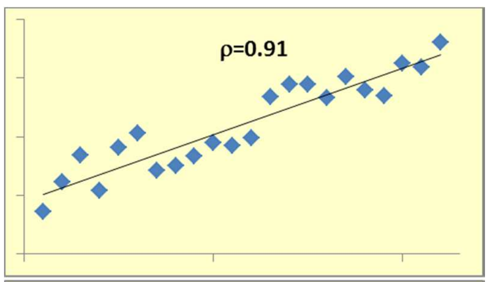
- Covariância
  - Até que ponto é que a variação de Y está associada à variação de Z?

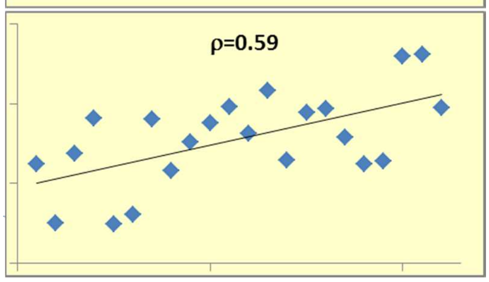
$$\mathrm{Cov}(Y,Z) = \frac{1}{N}\sum_{k=1}^{N}(Y_k - \overline{Y})(Z_k - \overline{Z})$$

$\overline{Y}$, $\overline{Z}$ - Valores médios de Y e de Z

- Correlação ρ

$$\rho = \frac{\mathrm{Cov}(Y,Z)}{\sqrt{V(Y)V(Z)}}$$



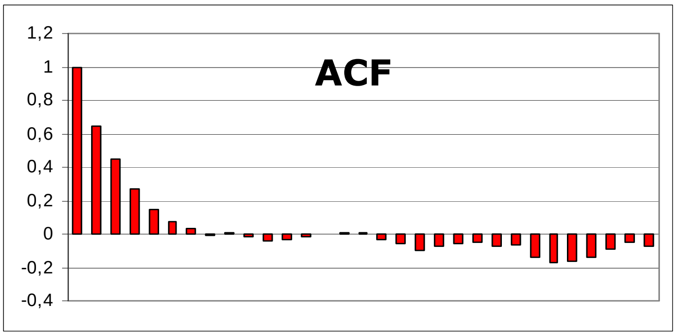


  - ρ = normalização da covariância para [–1; 1], dividindo pela média geométrica das variâncias de Y e Z

Identificação dos modelos – conceitos base

Auto-correlação (ACF)

$$\rho(X_t, X_{t-k}) = \rho_k = \frac{Cov(X_t, X_{t-k})}{V(X_t)}$$



- Correlação – é uma medida da dependência <u>linear</u> entre 2 variáveis.
- Se for relativa à própria variável com atraso k, X_{t-k}, diz-se **auto-correlação** de ordem k.

Identificação dos modelos – conceitos base

Auto-correlação Parcial (PACF)

- Apenas com ACF é difícil discriminar entre modelos AR(p). Por exemplo, entre AR(2) e AR(3) ou entre AR(3) e AR(4)
- Analisemos o seguinte caso, AR(1): $X_t = \phi X_{t-1} + a_t$
  - Dado que X_{t-1} tem toda a informação para determinar X_{t-1} então X_{t-2}, X_{t-3}, … não interessam. Uma regressão linear de X_t em X_{t-1}, converge para ϕ. Uma regressão linear de X_t em X_{t-2}, converge para ϕ^2. Contudo, se a regressão linear for de X_t em X_{t-1} e X_{t-2}, o coeficiente de X_{t-2} converge para zero, dado que X_{t-1} faz "todo o trabalho" na previsão de X_t. Ou seja, é como se a correlação entre X_t e X_{t-2} fosse zero.
- PACF ultrapassa esta dependência em cadeia, eliminando o efeito dos elementos intermédios (dentro do *time lag*)
  - A PACF de atraso k corresponde à autocorrelação entre X_t e X_{t-k} que não é explicada pelos atrasos de 1 a k.

Identificação dos modelos

Autoregressivas (AR) e média móvel (MA)

- Que diferenças?
  - O modelo AR inclui termos de atraso na própria série
  - O modelo MA inclui termos de atraso no ruído ou resíduos
- Como decidir qual usar?

ACF e PACF

Séries temporais

- Para identificar tipo e determinar p e q

	MA(q)	AR(p)	ARMA(p,q)
ACF	Desprezável após q	Decai*	Decai* após q
PACF	Decai*	Desprezável após p	Decai* após p

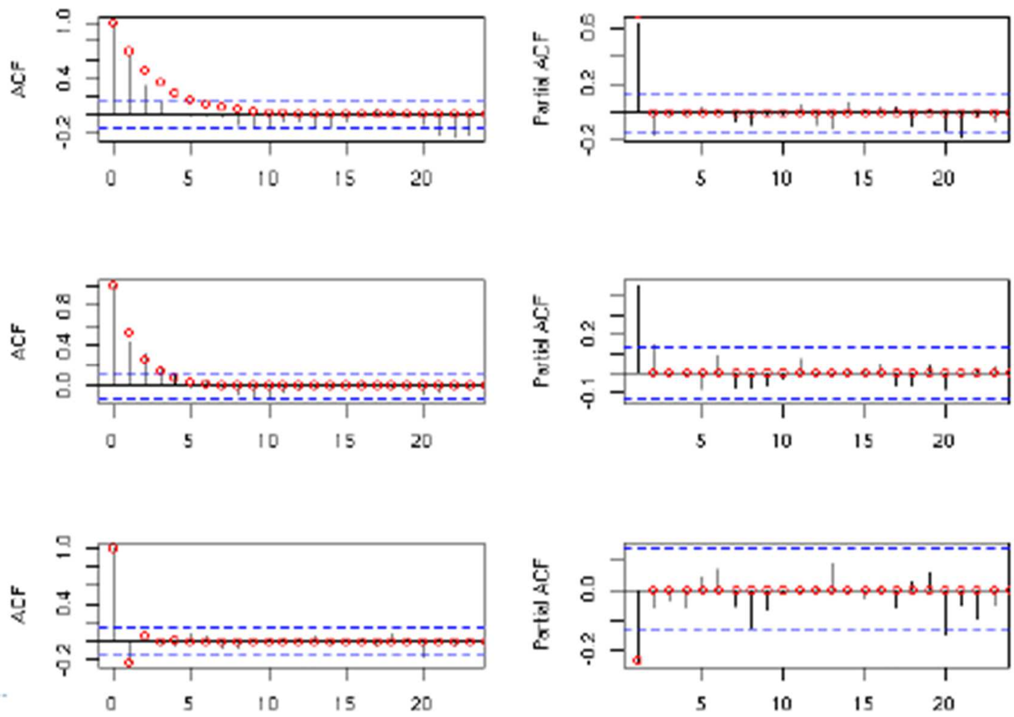
* Decaimento <u>acentuado</u> tipo exponencial ou sinusoide atenuada

Séries não-estacionárias: ACF permanece elevado para 6 ou mais lags $\Rightarrow$ diferenciar a série e reanalisar;

Exemplos

	MA(q)	AR(p)
ACF	Desprezável após q	Decai*
PACF	Decai*	Desprezável após p

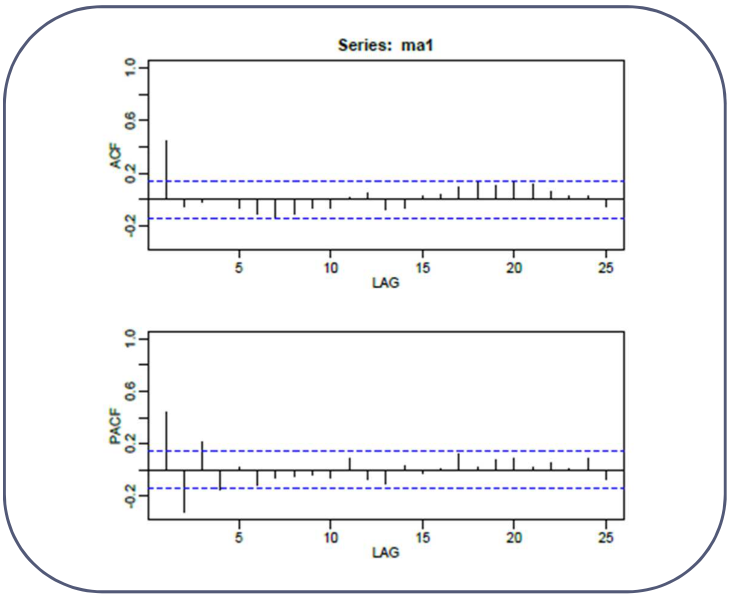
Modelos autoregressivos de 1ª ordem: AR(1) ou ARIMA(1,0,0)

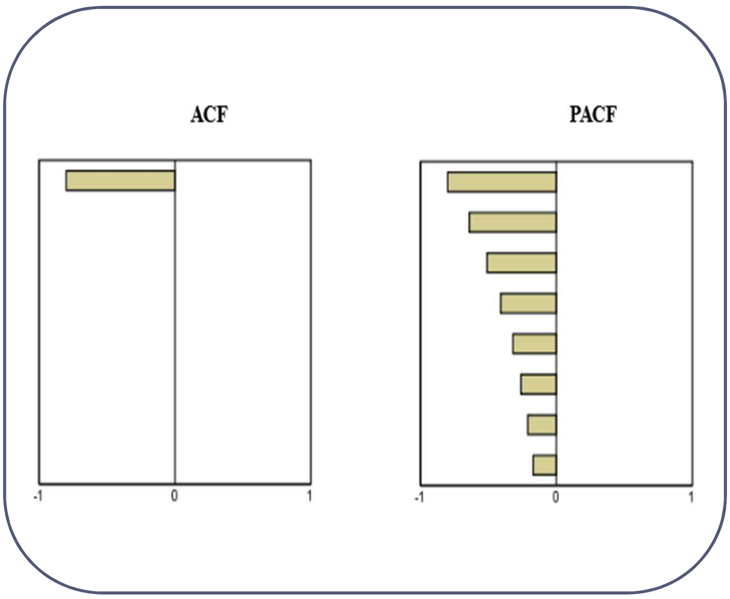


Exemplos

	MA(q)	AR(p)
ACF	Desprezável após q	Decai*
PACF	Decai*	Desprezável após p

Modelos média móvel de 1ª ordem: MA(1) ou ARIMA(0,0,1)

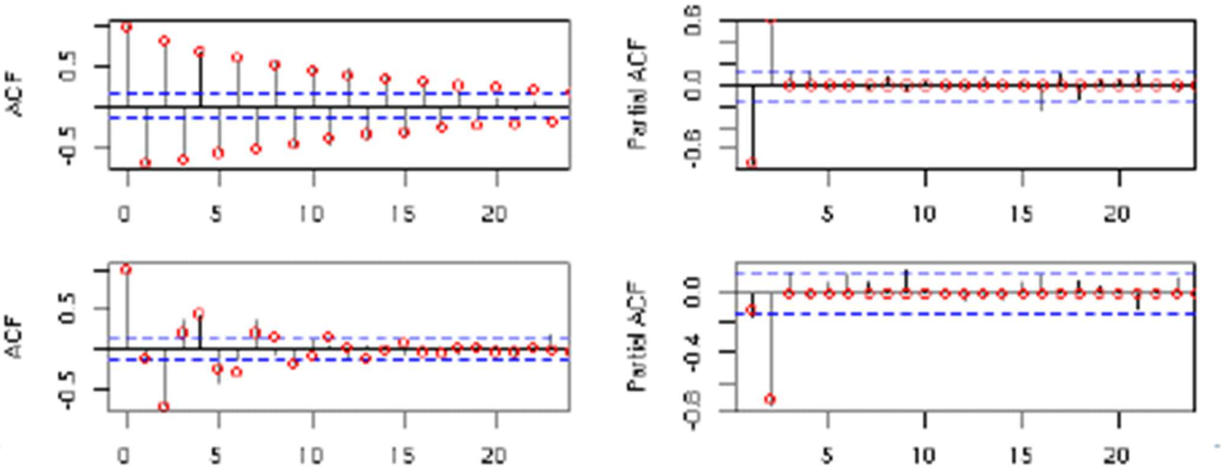




Exemplos

	MA(q)	AR(p)
ACF	Desprezável após q	Decai*
PACF	Decai*	Desprezável após p

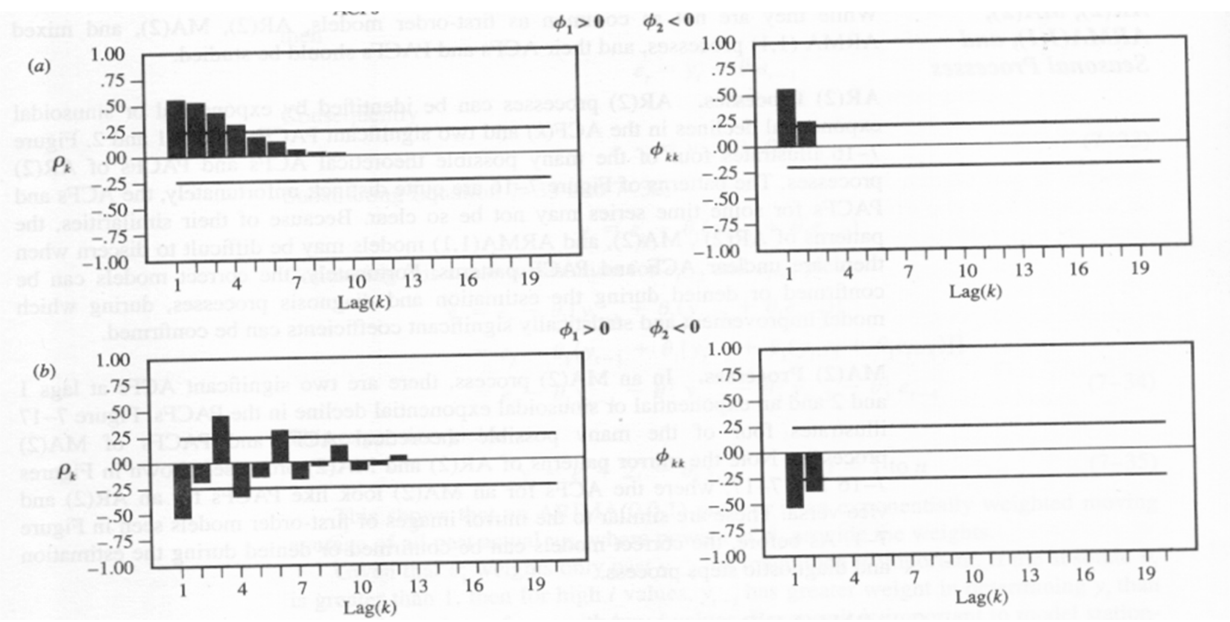
Modelos autoregressivos de 2ª ordem: AR(2) ou ARIMA(2,0,0)



Exemplos

	MA(q)	AR(p)
ACF	Desprezável após q	Decai*
PACF	Decai*	Desprezável após p

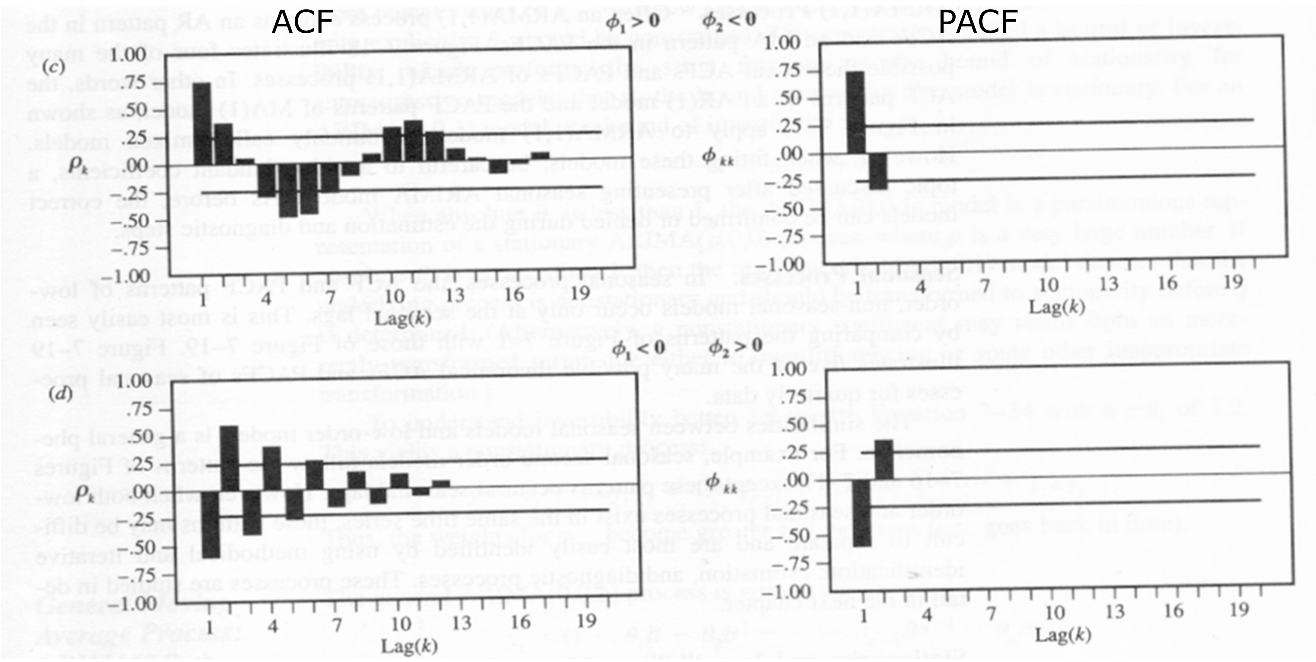
Modelos autoregressivos de 2ª ordem: AR(2) ou ARIMA(2,0,0)



	MA(q)	AR(p)
ACF	Desprezável após q	Decai*
PACF	Decai*	Desprezável após p

Exemplos

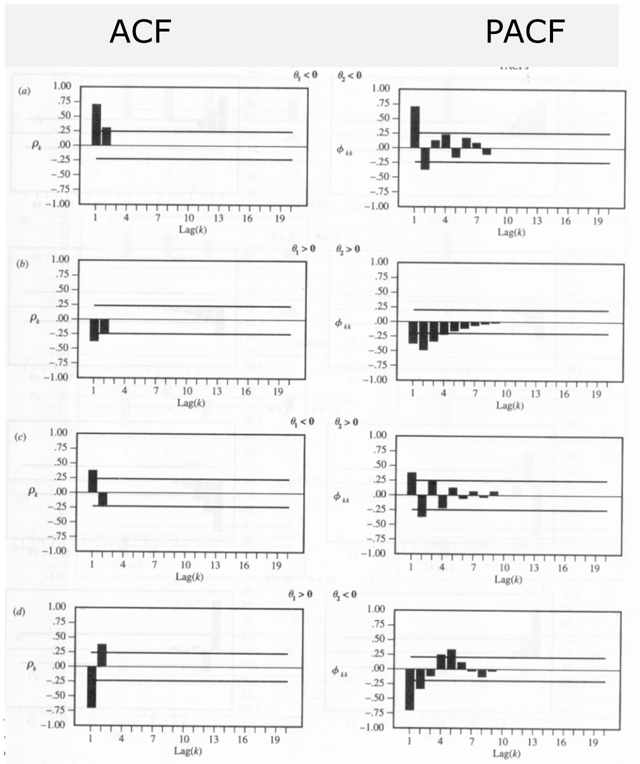
AR(2) ou ARIMA(2,0,0)

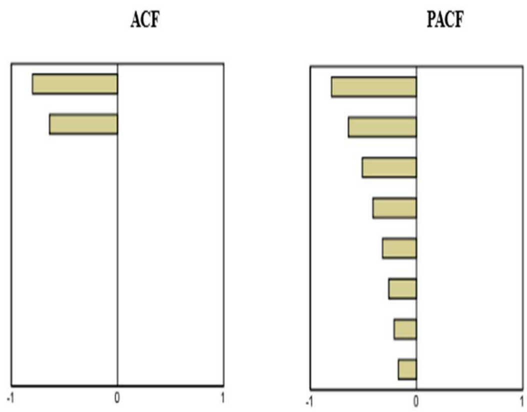


	MA(q)	AR(p)
ACF	Desprezável após q	Decai*
PACF	Decai*	Desprezável após p

Exemplos

MA(2) ou ARIMA(0,0,2)

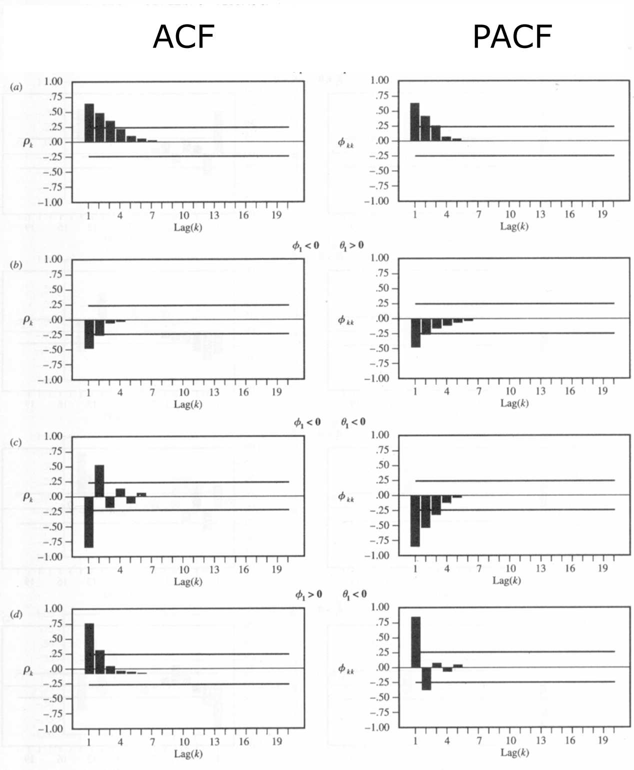


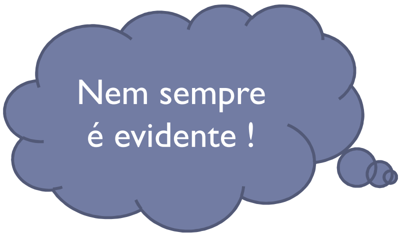


Exemplos

	MA(q)	AR(p)	ARMA(p,q)
ACF	Desprezável após q	Decai*	Decai* após q
PACF	Decai*	Desprezável após p	Decai* após p

Exemplos ARMA(1,1) ou ARIMA(1,0,1)

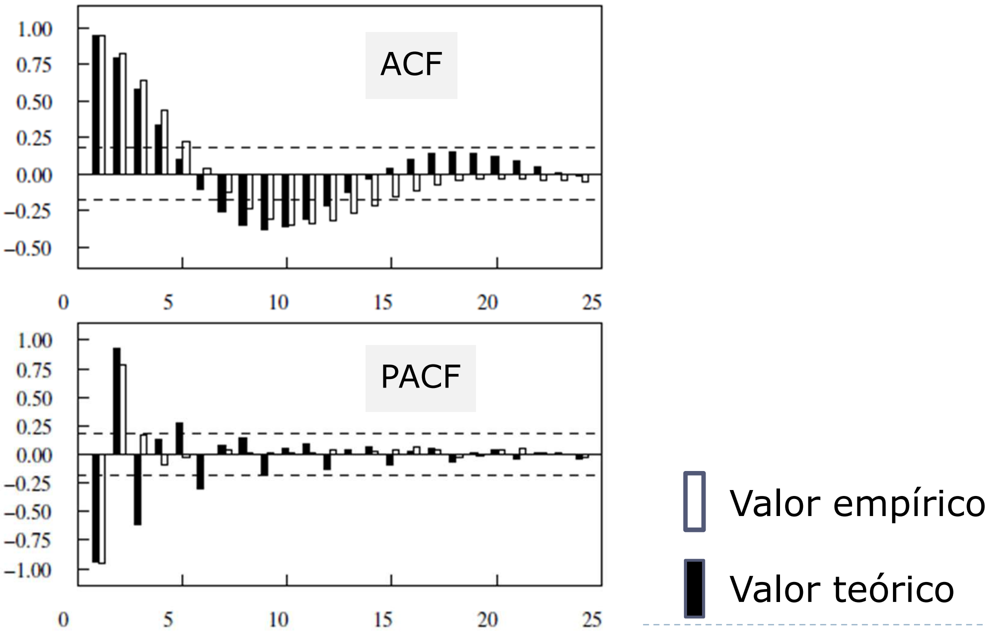




Exemplos

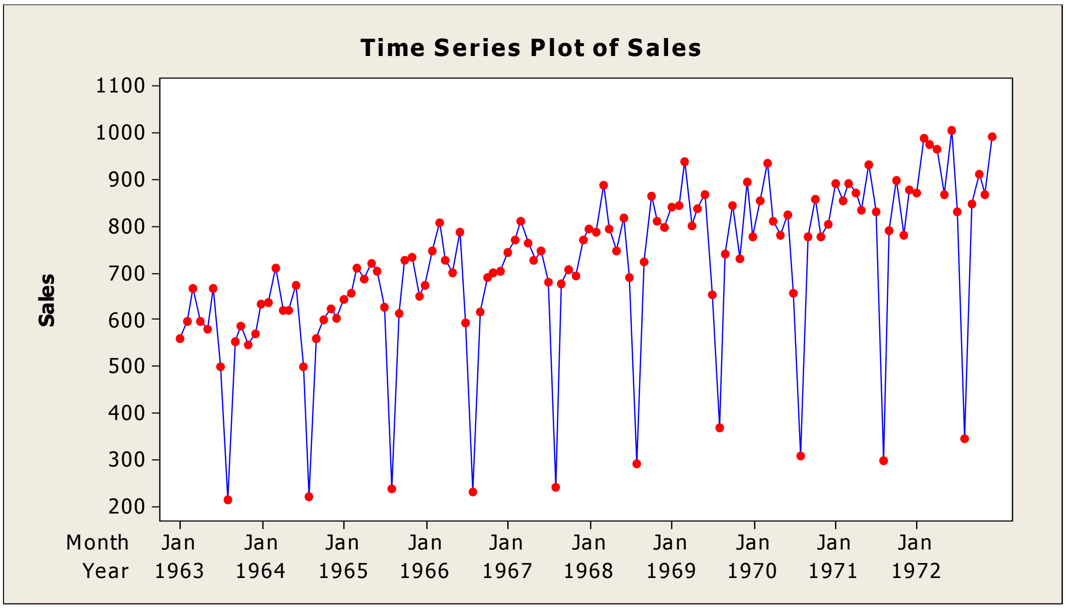
	MA(q)	AR(p)	ARMA(p,q)
ACF	Desprezável após q	Decai*	Decai* após q
PACF	Decai*	Desprezável após p	Decai* após p

ARMA(2,2) ou ARIMA(2,0,2)



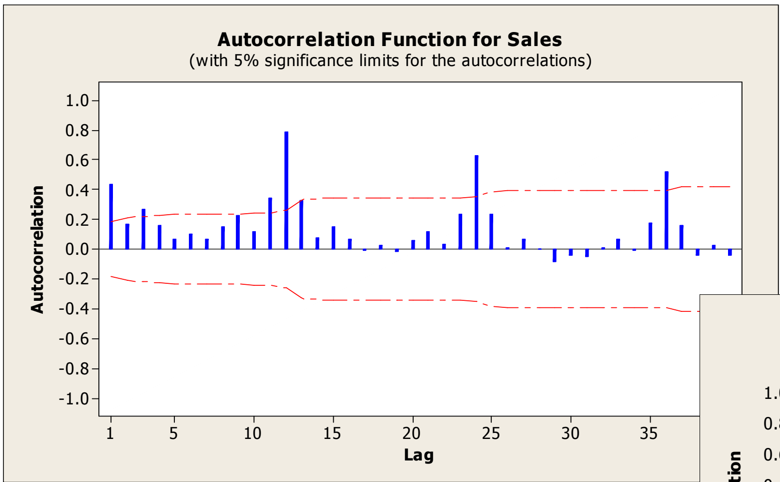
Sazonalidade

Sazonalidade



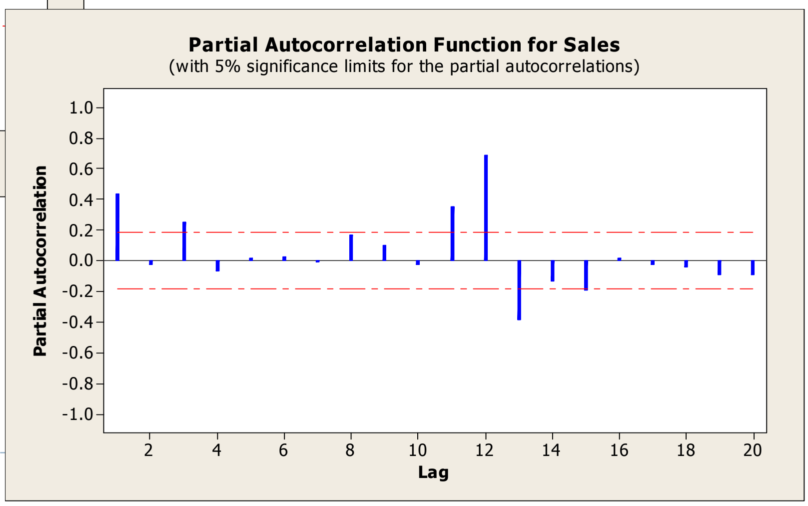
Sazonalidade

ACF



Sazonalidade nos *time lag* 12, 24 e 36

PACF

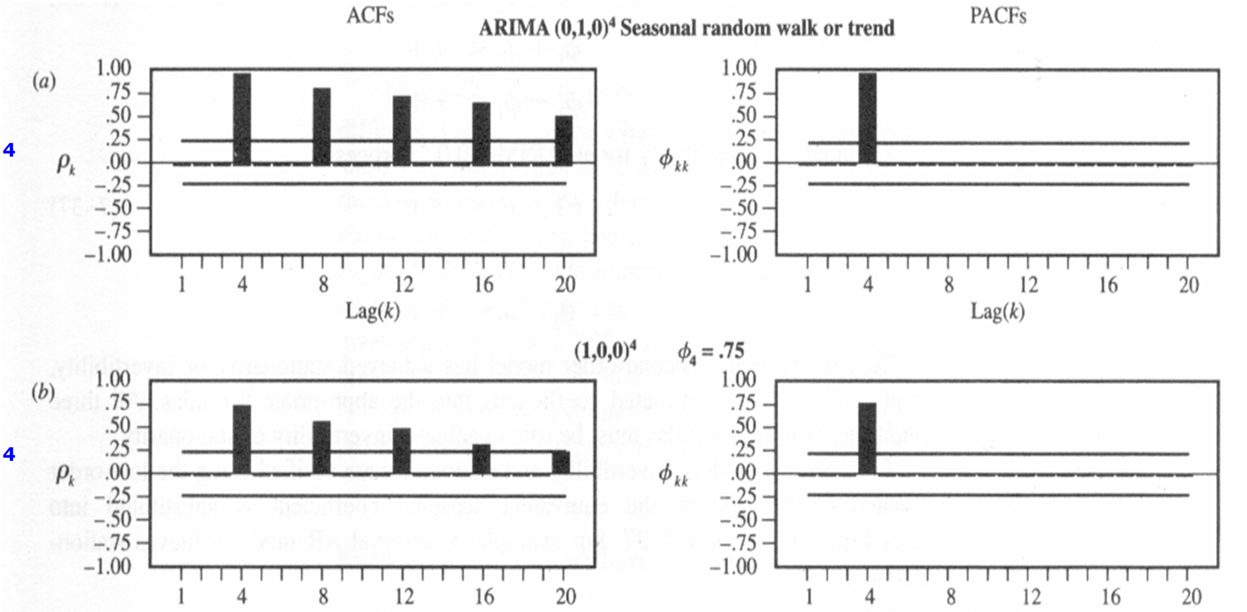


Sazonalidade

ACF e PACF para exemplos ARIMA(p,d,q)^4 sazonal

ARIMA(0,1,0)^4

ARIMA(1,0,0)^4



Nem sempre é fácil identificar os modelos… Nos integrativos, o decaimento é normalmente “lento e linear”.

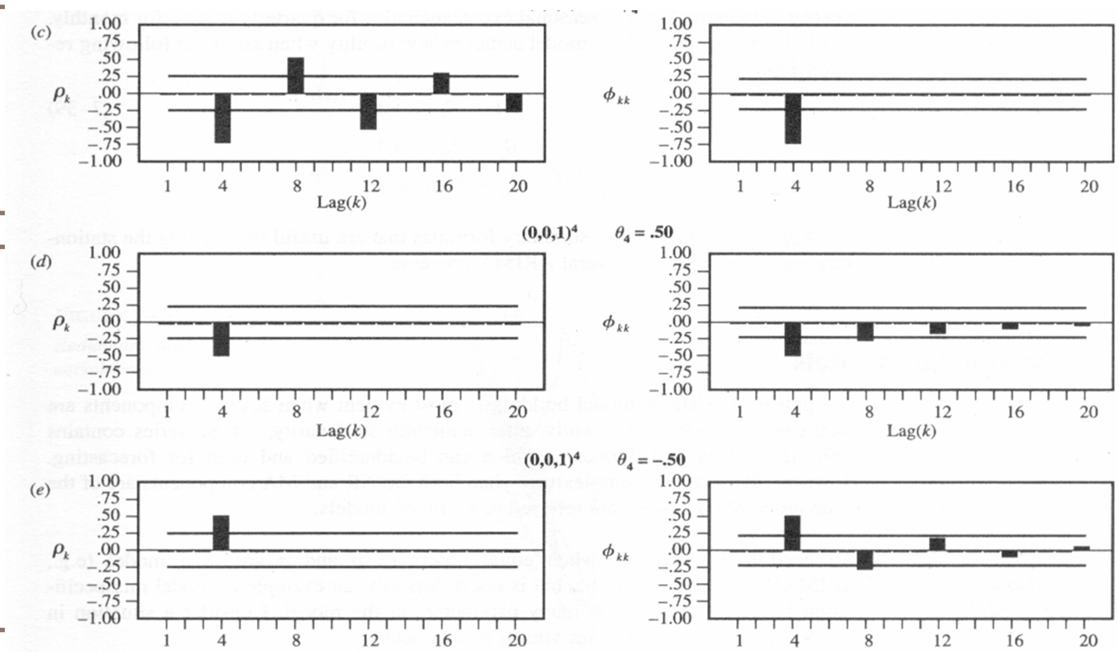
Sazonalidade

Exemplos ARIMA(p,d,q)^4 sazonal

ACF PACF

ARIMA(1,0,0)^4

ARIMA(0,0,1)^4



Identificação de modelos

- Para identificar tipo e determinar p e q

	MA(q)	AR(p)	ARMA(p,q)
ACF	Desprezável após q	Decai*	Decai* após q
PACF	Decai*	Desprezável após p	Decai* após p

* Decaimento acentuado tipo exponencial ou sinusoide atenuada

Séries não-estacionárias: ACF permanece elevado para 6 ou mais lags ⇒ diferenciar a série e reanalisar;

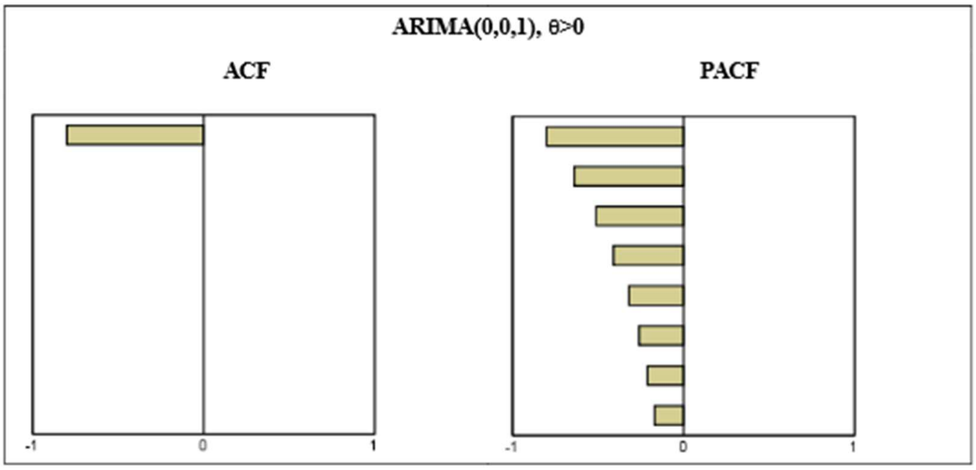
Identificação de modelos

- Para identificar tipo e determinar p e q (*cont.*)
  - Séries não-estacionárias: ACF permanece elevado para 6 ou mais lags ⇒ diferenciar a série e re-analisar;
  - Processos sazonais: picos em *lags* periódicos;
  - Nos modelos ARMA(p,q) contar os *lags* mais elevados que um dado limite
    - O SPSS atribui esse limite
  - ⇨ No final (após especificação do modelo), analisar sempre o ACF e PACF do resíduo para verificar se contém informação;
  - Podem ocorrer picos ocasionais (isolados) para grandes *lags*: à partida deverão ser ignorados.

Identificação de modelos

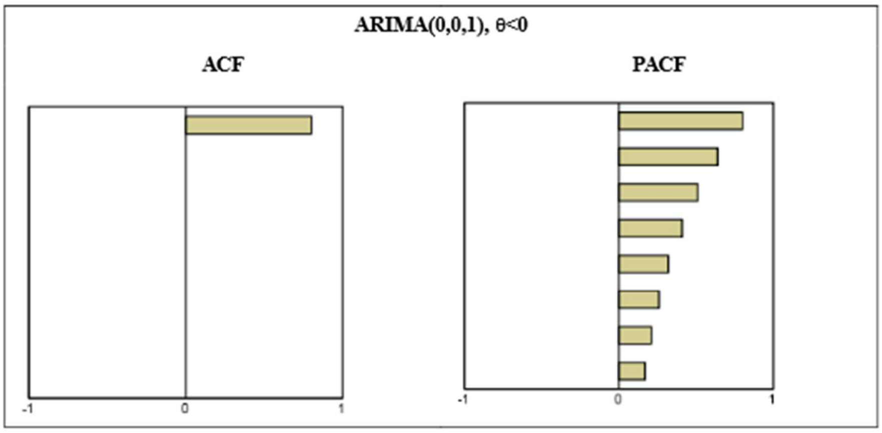
Mais exemplos…

Obs: Os exemplos seguintes são puros (teóricos). ACF e PACF de dados reais raramente são tão "limpos" como estes.

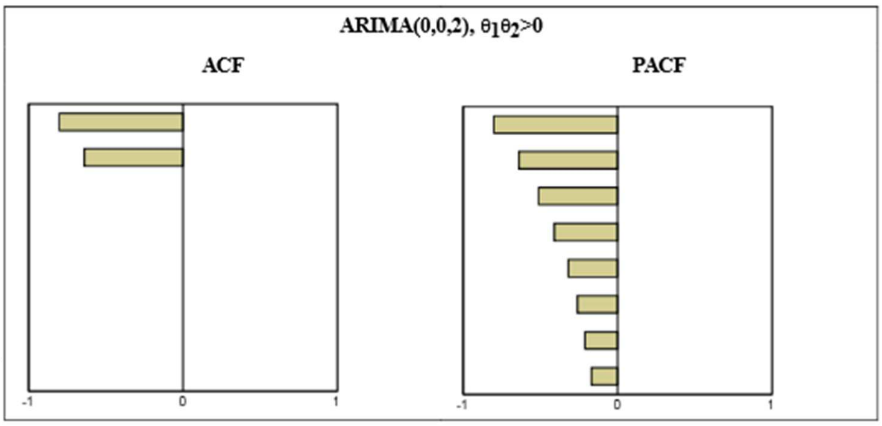


ARIMA(0,0,1)
ou
MA(1)

Identificação de modelos

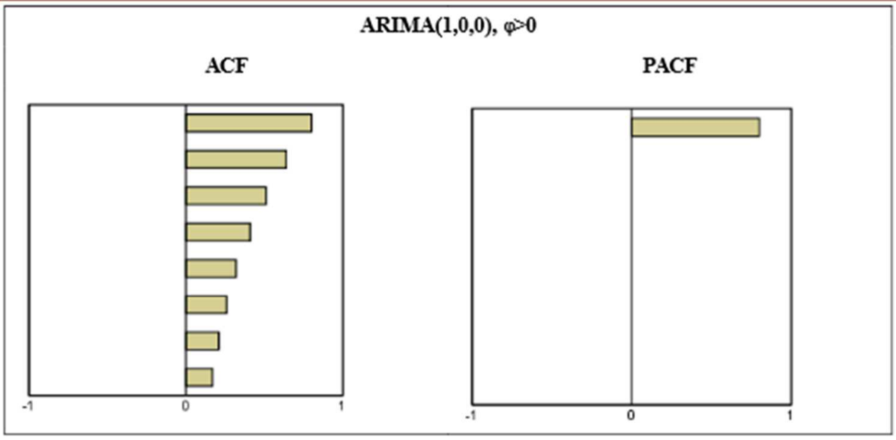


ARIMA(0,0,1)
ou
MA(1)

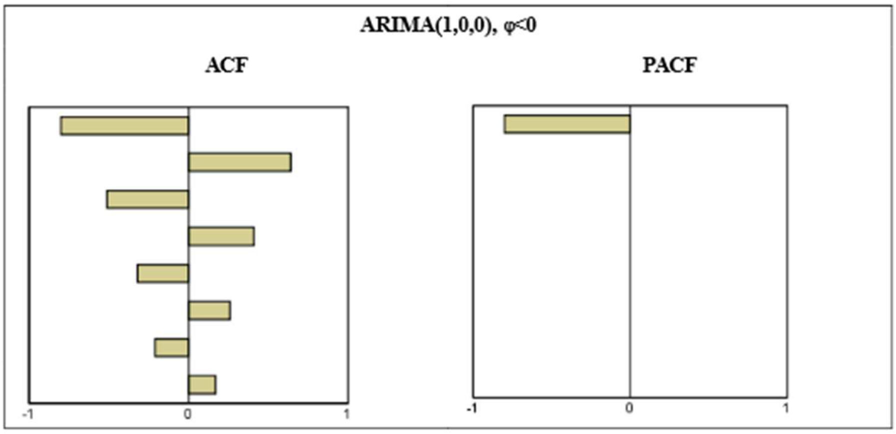


ARIMA(0,0,2)
ou
MA(2)

Identificação de modelos

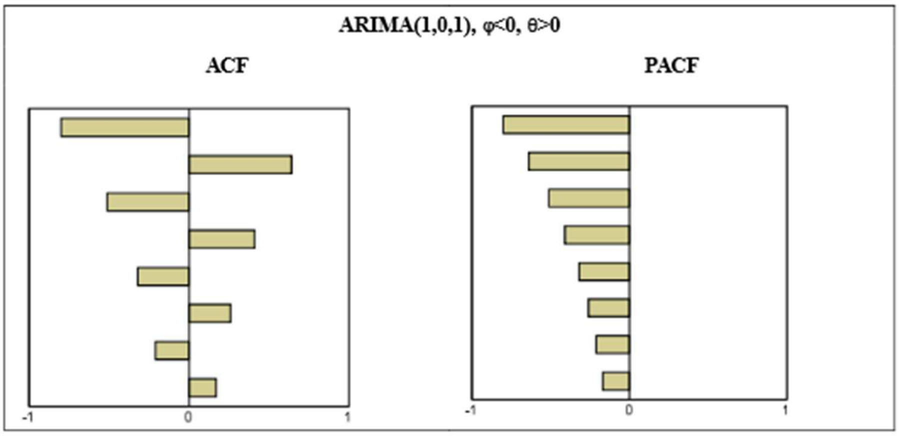


ARIMA(1,0,0)
ou
AR(1)

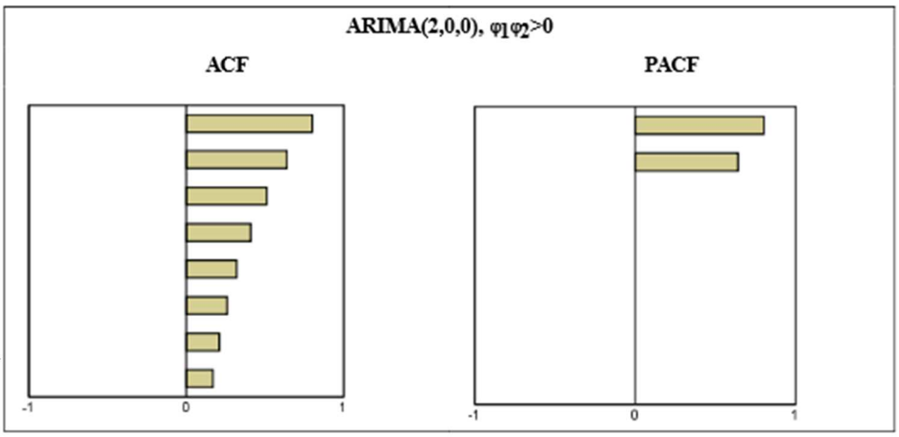


ARIMA(1,0,0
ou
AR(1)

Identificação de modelos



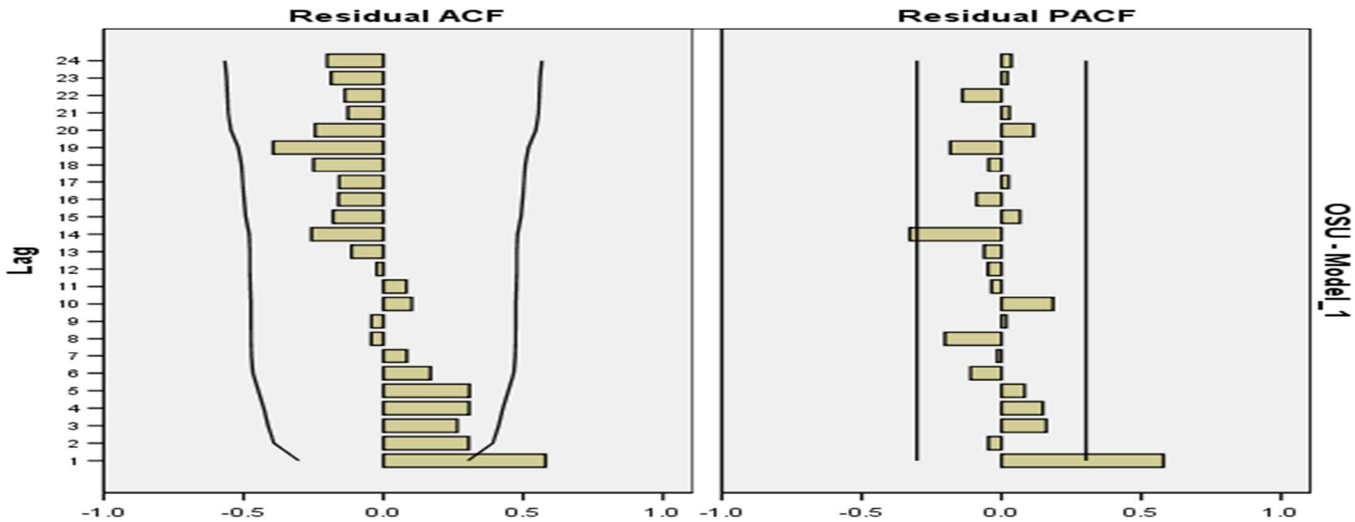
ARIMA(1,0,1)
Nem sempre é trivial...



ARIMA(2,0,0)
ou
AR(2)

Caso real

Série real: inscrições da universidade de Oregon 1962–2004



Talvez **ARIMA (1, 1, 0)**:

- o PACF tem um pico no lag 1 (e, a partir daí, decai rapidamente) $\Rightarrow$ AR(1);
- o ACF permanece elevado após bastantes lags $\Rightarrow$ integrativo.

Síntese

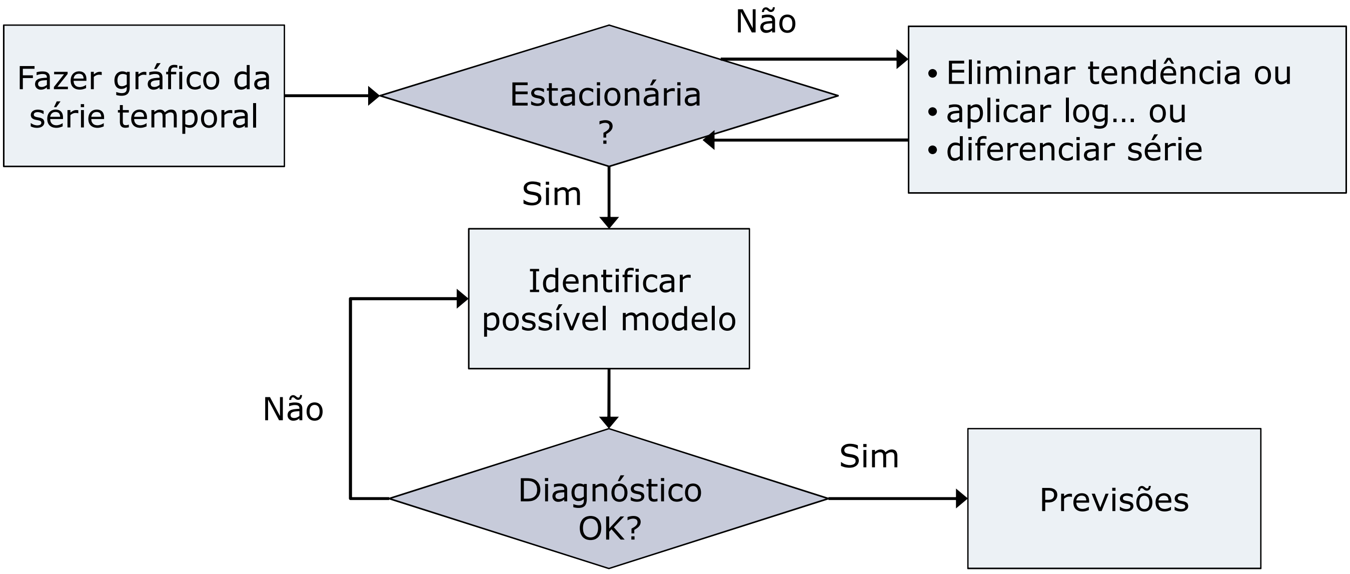
- Construção de um modelo ARIMA
  - Observar gráficos (linhas); identificar estacionaridade; identificar sazonalidade
  - Tornar a série estacionária:
    1. Subtrair média e dividir por desvio padrão **ou**
    2. Aplicar transformações logarítmicas ou potências **ou**
    3. Aplicar diferenciação para garantir estacionaridade e/ou para extrair sazonalidade
  - Observar ACF e PACF para identificar o tipo de modelo ARMA
  - Determinar os parâmetros do modelo ARMA
  - Construir o modelo completo; fazer a previsão; validar o modelo; avaliar o desempenho do modelo

Obs: A identificação de modelos baseada na análise de ACF e PACF não é infalível mas, normalmente, permite obter informações sobre modelos adequados. Nas séries reais ou com muito ruído, as conclusões poderão ser menos claras.

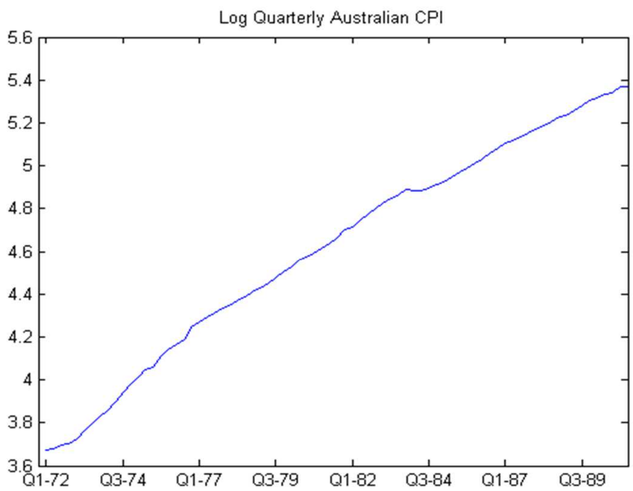
Outras medida para identificação de modelos: Akaike's information criterion (AIC), Bayesian Information Criterion (BIC), Hannan-Quinn Criteria (HQIC),…

Síntese

- Construção de um modelo ARIMA



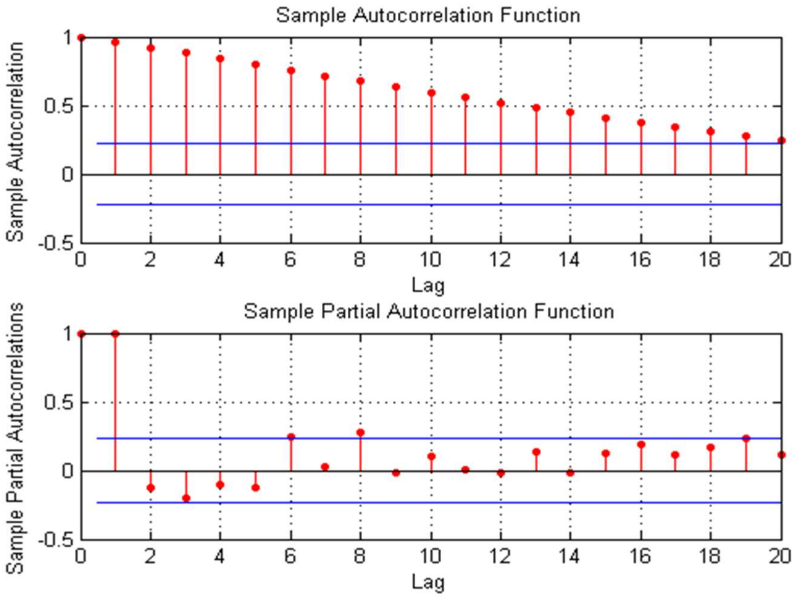
Exemplo



Série não estacionária com tendência crescente

Exemplo

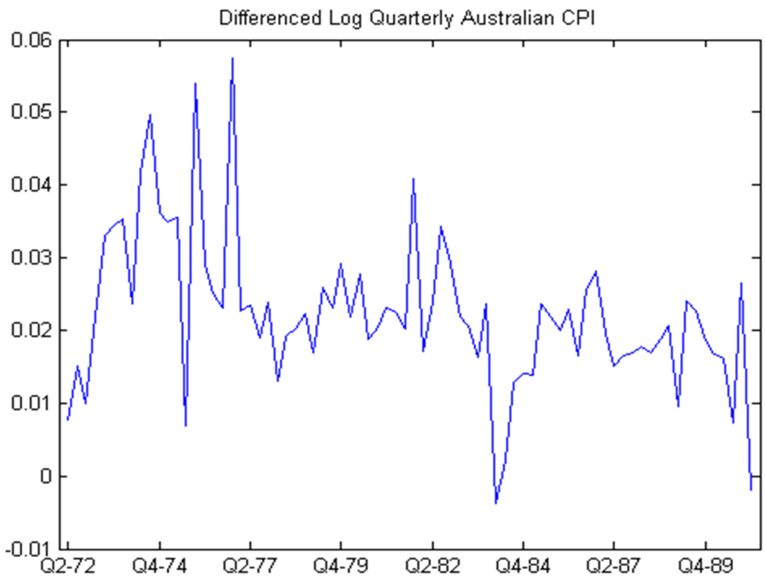
- ACF e PACF



O decaimento linear do ACF sugere processo não estacionário

Exemplo

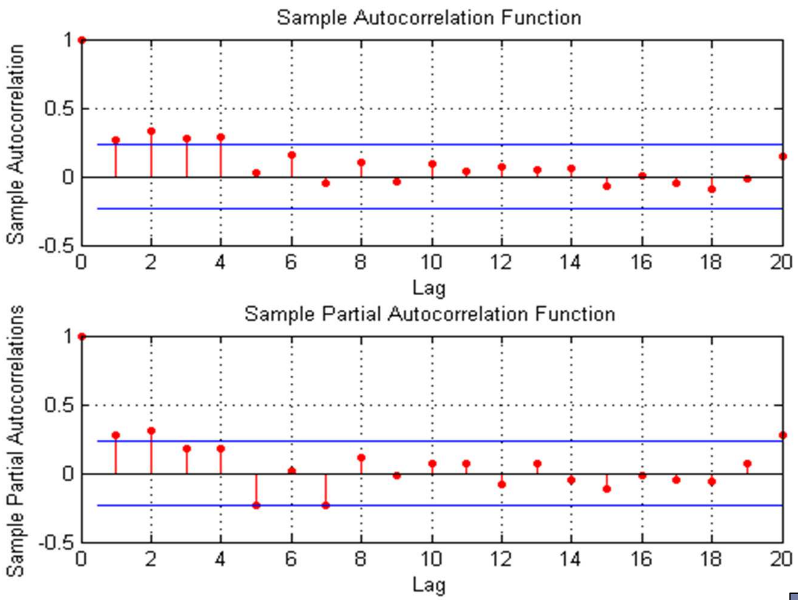
- Diferenciar série



A diferenciação tornou a série (mais) estacionária

Exemplo

- ACF e PACF (para a série diferenciada)



O ACF decai agora mais rapidamente.

O PACF é desprezável após o lag 2, comportamento típico do modelo autoregressivo de 2ª ordem AR(2).

ARIMA (2,1,0)

Modelo AR(2) da série diferenciada

Integrada (1 diferenciação)