AULA:

Introdução às Séries Temporais

Prof. Victor Hugo Lachos Davila

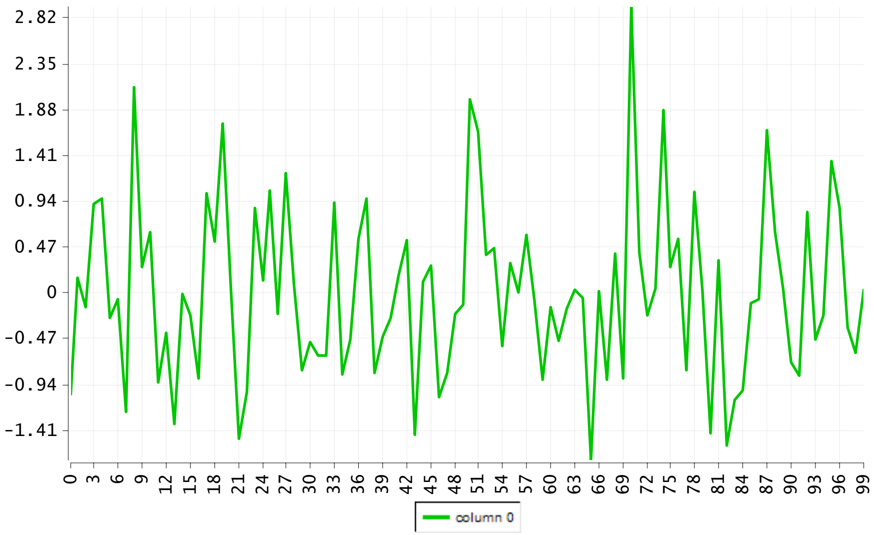
Natureza e Fonte de Dados

Existe 3 tipos de dados disponível para análise:

- **Séries Temporais**: quando os dados são observados em diferentes instantes do tempo, seja diariamente (preço de ações, relatórios meteorológicos), mensalmente (taxa de desemprego, IPC), trimestralmente (PIB).
- **Corte Transversal**: quando os dados observados foram coletados no mesmo ponto do tempo (pesquisas de opinião, dados de censos)
- **Dados em Painel**: Aqui uma unidade em corte transversal é pesquisada ao longo do tempo. (PIB de cada pais sul-americano para o período de 1990 a 2008)

Exemplos de séries temporais (1)

Ruído Branco Gaussiano



$$X_t \sim N(0, \sigma^2) \quad \text{i.i.d}$$

independente e identicamente distribuídos (a.a.)

- O valor esperado é constante
- A variância é constante
- Simétrica
- Não correlacionados entre instantes diferentes

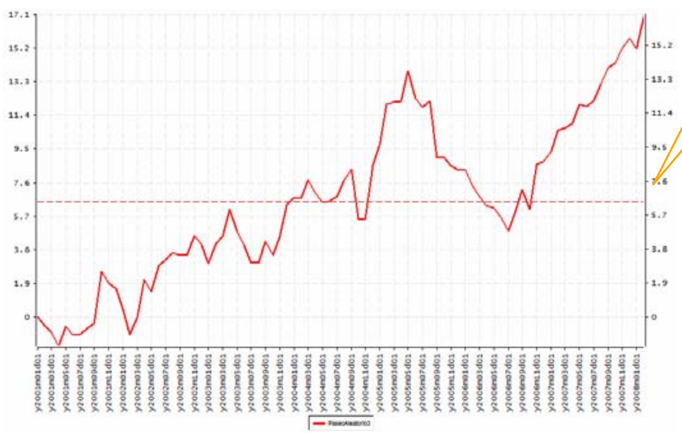
Movimento de uma partícula com relação a um ponto

μ=0

Passeio Aleatório

$$X_t = \mu + X_{t-1} + a_t$$

$$a_t \sim N(0, \sigma^2) \quad \text{i.i.d}$$

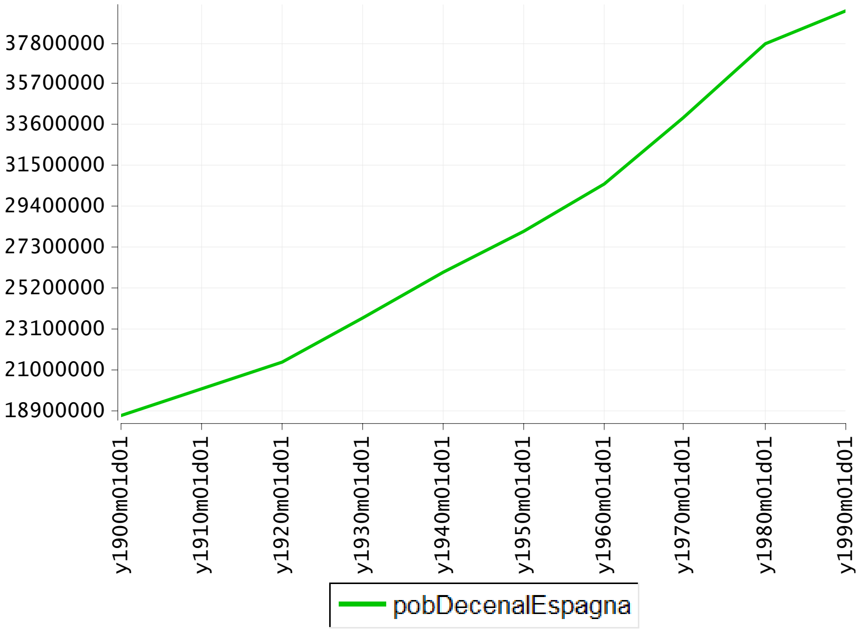


μ=0.2

- O valor esperado é constante
- A variância não é constante
- Simétrica
- Correlacionados entre instantes diferentes

Exemplos de séries temporais (3)

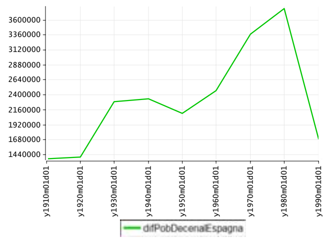
População da Espanha

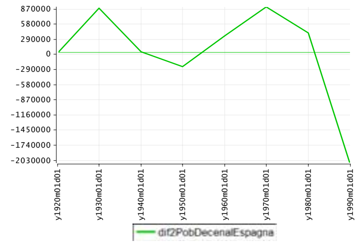


A média não é constante

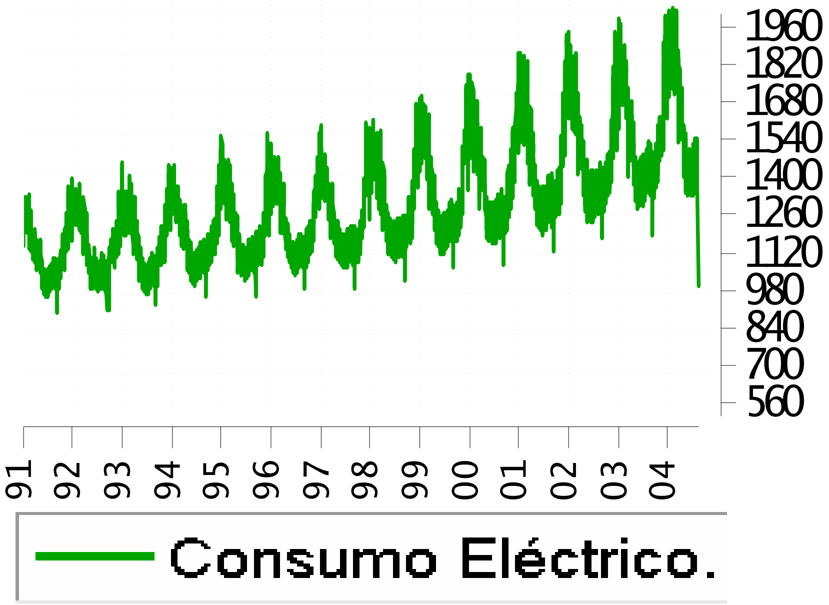
A população espanhola cresceu estritamente de década em década de maneira aparentemente lineal. Tendência crescente

Se tomamos a diferença $y_t - y_{t-1}$ pode-se observar as flutuações quanto à velocidade de crescimento.





Consumo elétrico



As séries de consumo de eletricidade apresentam uma clara tendência positiva que parece acelerar-se no final da série.

Por outro lado a série apresenta uma marcada sazonalidade com consumos muito elevados nos meses de inverno devido ao efeito da temperatura.

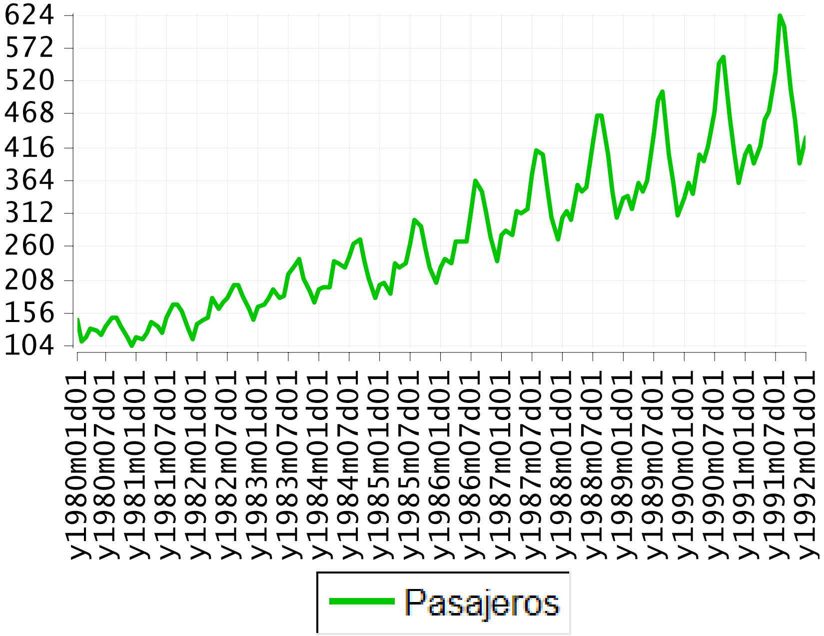
A série parece ter maior dispersão na medida que toma maiores valores.

Se a **sazonalidade** é estritamente periódica pode eliminar-se da série com um componente determinista.

Um truque poderia ser estudar separadamente as séries correspondentes em períodos equivalentes.

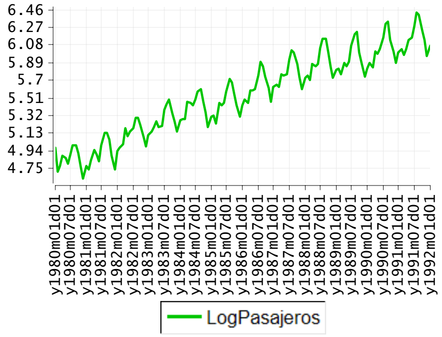
Outra alternativa é tomar diferenças de ordem apropriado.

Passageiros em linhas aéreas



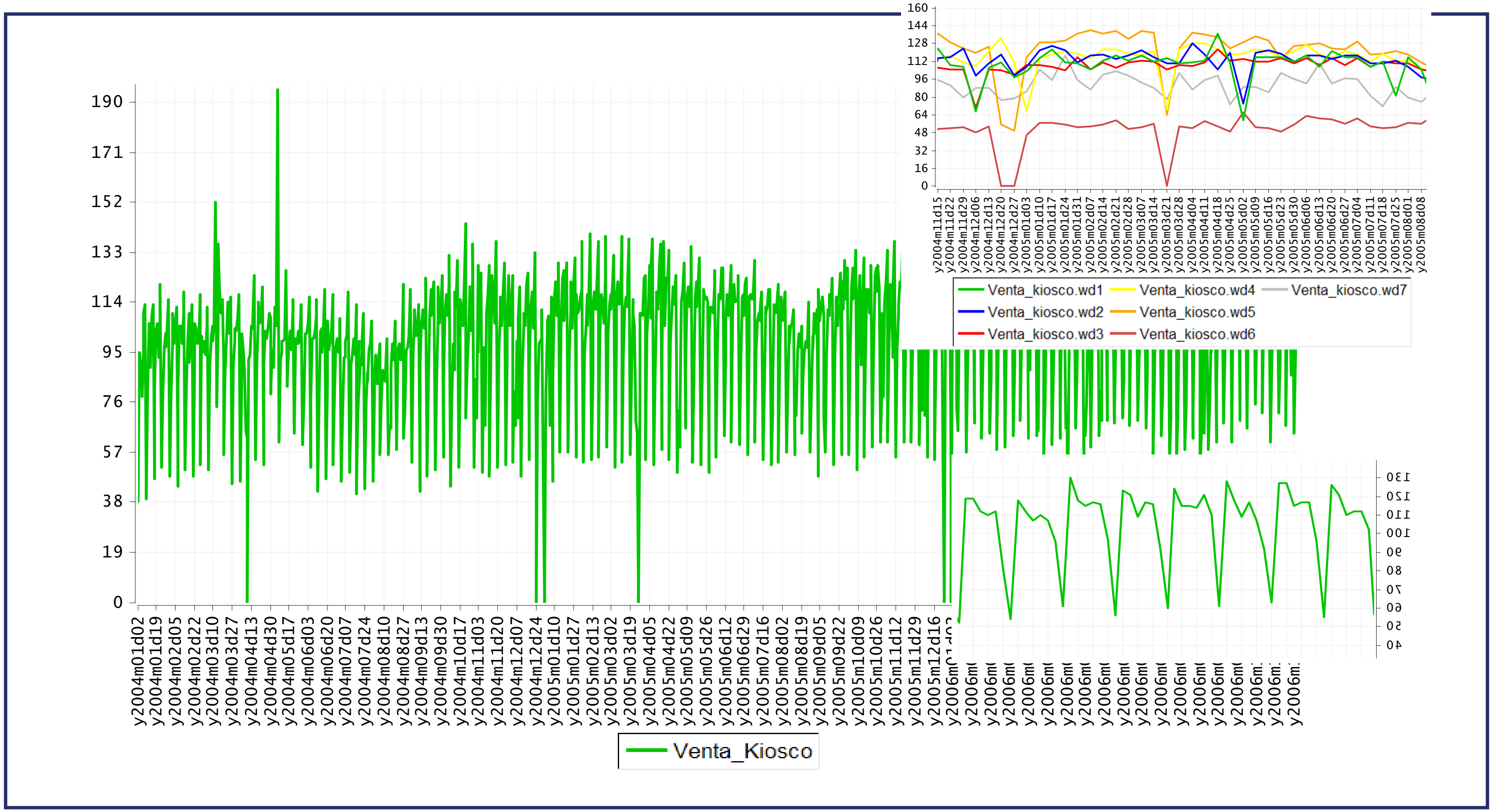
As séries de número de passageiros por mês em linhas aéreas apresenta sazonalidade com um marcado crescimento onde a variância aumenta na medida que toma maiores valores.

Heteroscedasticidade
Se a variância não é constante pode-se normalizar transformando apropriadamente os dados iniciais.



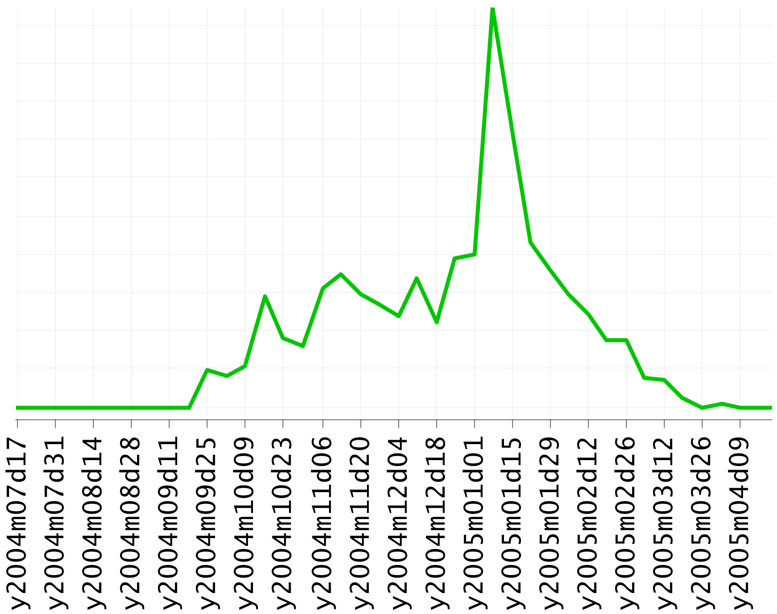
Exemplos de séries temporais (6)

Venda diária de um jornal em uma banca



Exemplos de séries temporais (7)

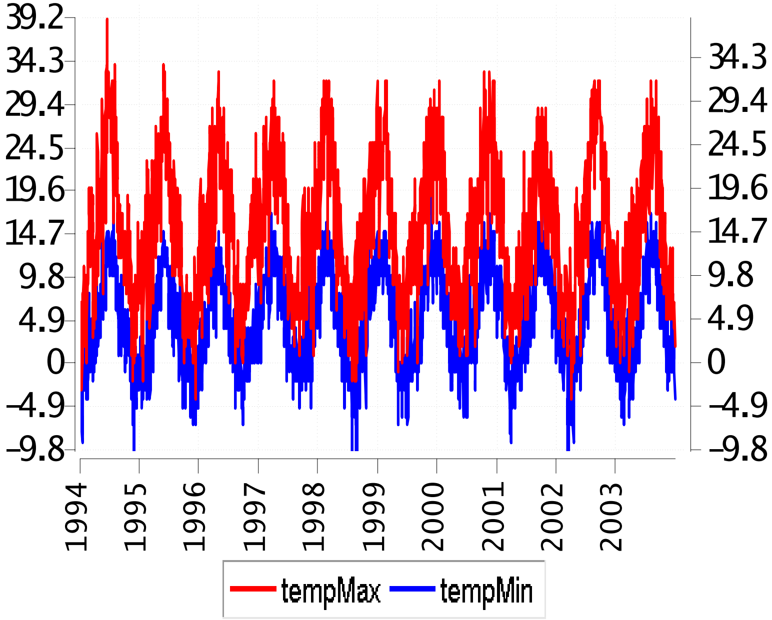
Venda de um tipo de produto em uma temporada



Efeitos como o começo das promoções produz um resultado em incremento instantâneo das vendas que se transmite com certo decaimento.

Exemplos de séries temporais (8)

Temperatura máxima e mínima diária

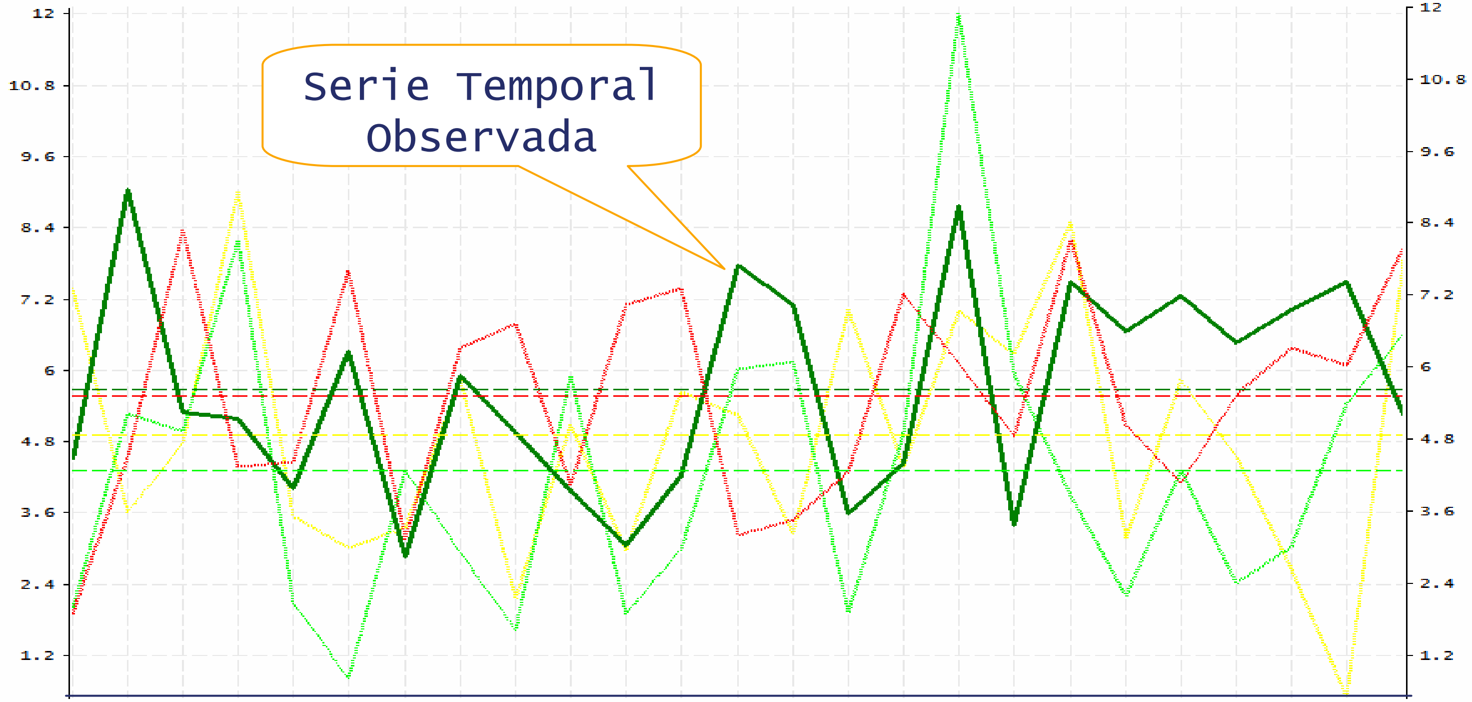


As séries de temperaturas oscilam em torno de um valor central, (15º máx e 5º mín) mas sistematicamente alguns meses tem valores mais altos que outros. Por exemplo os meses de verão tem valores mais altos que os meses de inverno.

Séries diferentes guardam altas correlações. Também pode-se considerar o problema de modelar ambas simultaneamente, como **Séries Temporais Multivariadas**.

Processos Estocásticos

Um **Processo Estocástico** pode ser definido como uma coleção de variáveis aleatórias ordenadas no tempo $\{x_t, t \, \varepsilon \, T\}$,onde T é um conjunto ordenado de índices.

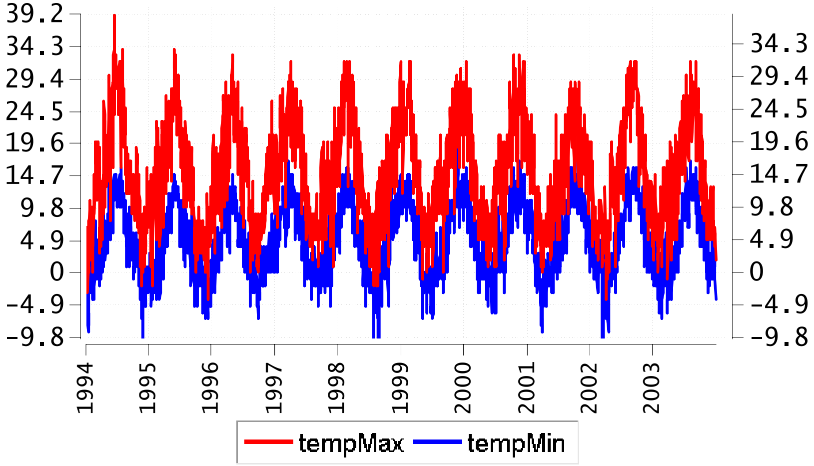


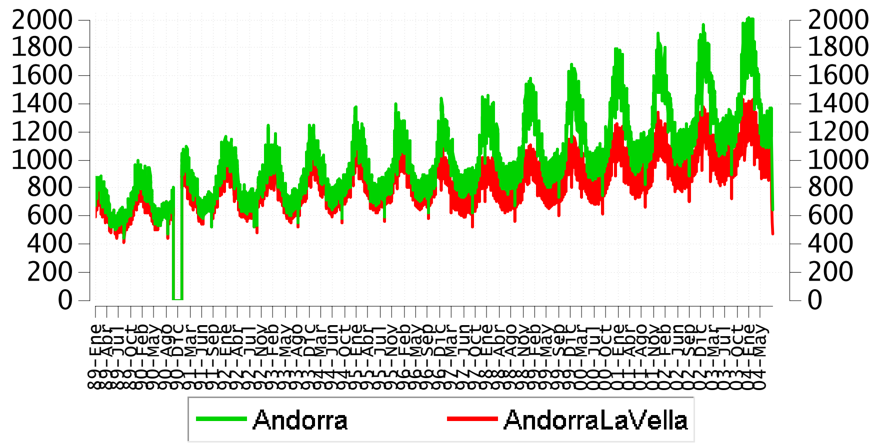
Z_t

Os $Z_{t,,}$, t=0,1,2,3,.....n, são realizações de x_t

Def.) **<u>Série temporal</u> :** Uma serie temporal se considera como a realização de um processo estocástico e que estão ordenadas em intervalos regulares de tempo (cada dia, cada mês, cada ano, etc)

$$\{z_t\} \qquad z_1, z_2, z_3, .. z_t$$





Os dados de séries temporais geralmente não são independentes, especialmente se os intervalos da amostra são curtos.
As observações próximas costumam ser mais parecidas que as mais distantes (p. ej. temperatura diária).

Caracterização de Processos Estocásticos

Um processo estocástico fica caracterizado se definimos a função da **distribuição conjunta** das variáveis aleatórias ($z_1,z_2,..\ z_T$) para qualquer valor de T. Dizemos que **conhecemos a estrutura probabilística** de um proceso estocástico quando conhecemos estas distribuições, sem embargo isto requer observar um grande número de realizações que não costumam estar disponíveis quando se trata de séries econômicas.

Def.) **Função de Médias :**proporciona as esperanças das distribuições marginais de z_t para cada instante

$$\mu_t = E[z_t]$$

Caracterizando o processo estocástico Zt

Def.) **Função de variâncias** do processo proporciona as variâncias em cada instante temporal.

$$\sigma_t^2 = Var[z_t]$$

Def.) **Função de AutoCovariâncias**: descreve a Covariâncias entre duas variáveis do processo em dois instantes quaisquer.

$$\gamma(t,t+j) = Cov(z_t, z_{t+j}) = E[(z_t - \mu_t)(z_{t+j} - \mu_{t+j})]$$

Def.) **Função de Autocorrelaçãon**: mede a dependência lineal entre os valores do processo no instante t e no instante t+j. Chamaremo coeficiente de correlação de orden (j) a:

$$\rho(t,t+j) = \frac{Cov(z_t, z_{t+j})}{\sigma_t \sigma_{t+j}}$$

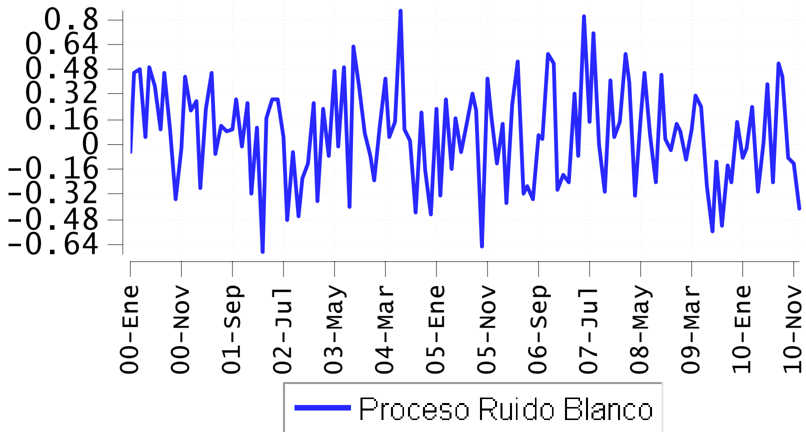
o **correlograma** não é mais que a representação dos coeficientes de autocorrelación em função do atraso j.

Tipos de processos estocásticos

Processos Estocásticos

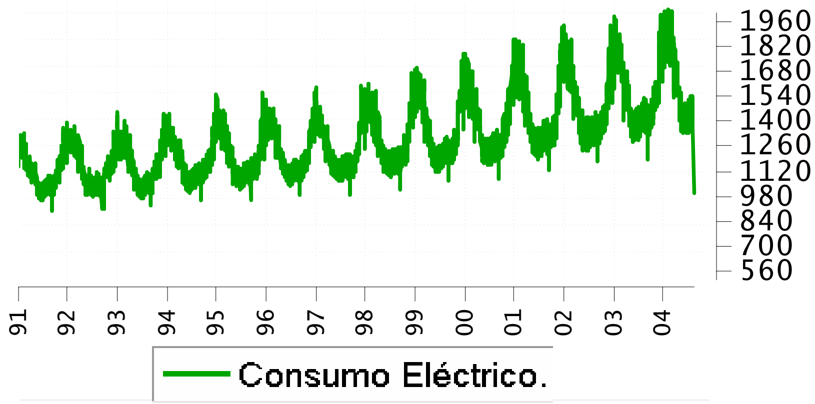
Processos Estacionários

Um processo estocástico Z_t é estacionário quando as **propriedades estatísticas de qualquer sequência finita** $z_1, z_2, \ldots z_k$ de componentes de Z_t são **semelhantes** às da sequência $z_{1+h}, z_{2+h}, \ldots z_{k+h}$ para **qualquer** número inteiro h



Processos Não Estacionários

Um processo estocástico Z_t é não estacionário quando as **propriedades estatísticas de ao menos uma sequência finita** $z_1, z_2, \ldots z_k$ de componentes de Z_t são **diferentes** das de sequência $z_{1+h}, z_{2+h}, \ldots z_{k+h}$ para **ao menos um** número inteiro h



Se existe a transformação:
Processos Homogêneos de Orden d

Processos Estocásticos Estacionários

Def.) **Processo Estocástico Estacionário em Sentido Estrito** : Um processo estocástico Zt é estacionário quando as propriedades estatísticas de qualquer sequência finita $z_1, z_2, .. z_k$ de componentes de Zt são **semelhantes** às da sequência $z_{1+h}, z_{2+h}, .. z_{k+h}$ para **qualquer** número inteiro h.
Um processo estocástico Zt é estacionário quando a distribuição conjunta de qualquer conjunto de variáveis não se modifica se transferimos as variáveis no tempo.

$$F(x_i, x_j, .. x_k) = F(x_{i+h}, x_{j+h}, ..., x_{k+h})$$

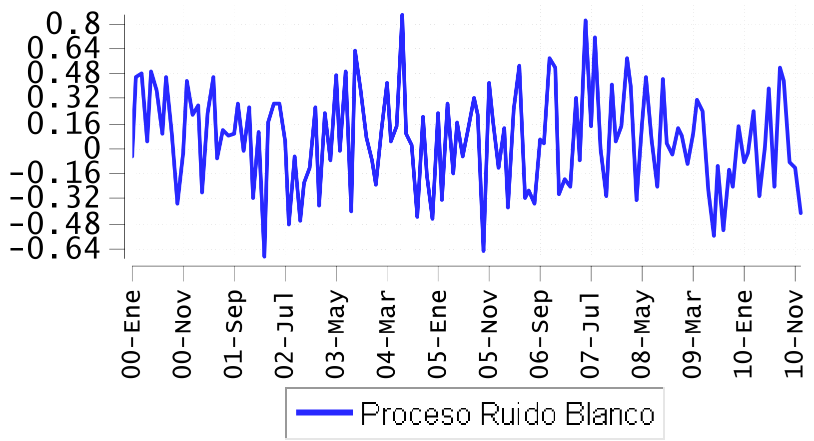
Def.) **Processo Estocástico Estacionário em Sentido Débil** : Quando os momentos de primeira e segunda ordem do processo são constantes e a Covariância entre duas variáveis depende somente de sua separação no tempo

$$E(x_t) = \mu$$

$$Var(x_t) = \sigma^2$$

$$\gamma(t, t-k) = \gamma(t, t+k) = \gamma_k \quad k = 0, \pm 1, \pm 2$$

Nota: Quando um processo é estacionário as propriedades estatísticas se simplificam notavelmente e sua caracterização é mais simples



Exemplos de Processos Estocásticos Estacionários

Processo Ruído Branco. ARIMA(0,0,0)

$$z_t = a_t$$

Processo Homogêneo de orden 1. ARIMA(p,0,0)

$$\nabla z_t = a_t$$

Processo Autoregressivo de orden (p).ARIMA(p,0,0)

$$z_t = \phi_1 z_t + .. + \phi_p z_{t-p} + a_t$$

Processo Média Móvel de ordem (q).ARIMA(0,0,q)

$$z_t = a_t + \theta_1 a_{t-1} + .. + \theta_q a_{t-q}$$

Processo ARIMA(p,0,q)

$$z_t = \phi_1 z_{t-1} + .. + \phi_p z_{t-p} + a_t + \theta_1 a_{t-1} + .. + \theta_q a_{t-q}$$

Caracterização

Média de Amostra

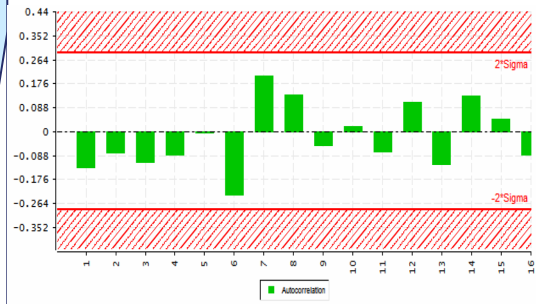
$$\hat{z}_t = \frac{\sum_{t=1}^{T} z_t}{T}$$

Matriz de Covariâncias

$$\hat{\gamma}_t = \frac{1}{T}\sum_{t=k+1}^{T}(z_t - \hat{z})(z_{t-k} - \hat{z})$$

Autocorrelações

$$r_k = \hat{\gamma}_k / \hat{\gamma}_o$$



Estimação dos momentos de Processos Estacionários

Dada uma série temporal Zt nosso objetivo é construir um modelo estatístico que capture toda a informação estatística sistemática contida em essa série. Se consideramos Zt como uma realização de um processo estocástico podemos obter **estimativas de seus momentos de amostras** tal que:

Caracterização

Def.) **Estimador Média de amostra**: é um estimador centrado da média populacional. $E(\hat{z}_t) = \mu$

$$\hat{z}_t = \frac{\sum_{t=1}^{T} z_t}{T}$$

Def.) **Estimador das Covariânçias de orden k quando a média populacional é desconhecida**. É um estimador enviesado da autocovariâncias populacionais mas têm menores erros quadráticos de estimação que o estimador com média conhecida. Garante que a matriz de covariâncias seja sempre definida positiva.

Média Desconhecida:

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} (z_t - \hat{z})(z_{t-k} - \hat{z})$$

Média Conhecida

$$\tilde{\gamma}_k = \frac{1}{T-k} \sum_{t=k+1}^{T} (z_t - \mu)(z_{t-k} - \mu)$$

Def.) **Estimador da matriz de covariâncias:**

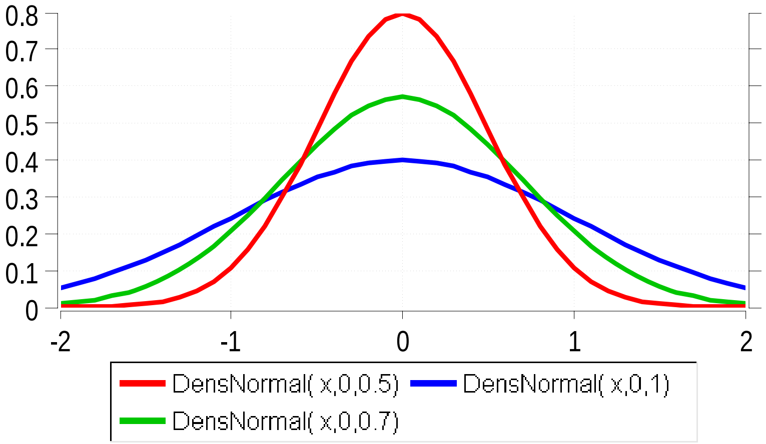
$$\tilde{\Gamma}_k = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{k-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{k-1} & \hat{\gamma}_{k-2} & \cdots & \hat{\gamma}_0 \end{bmatrix}$$

Def.) **Estimador das autocorrelações de orden k:**

$r_k = \hat{\gamma}_k / \hat{\gamma}_o$ Onde $\hat{\gamma}_o$ é a variância do proceso

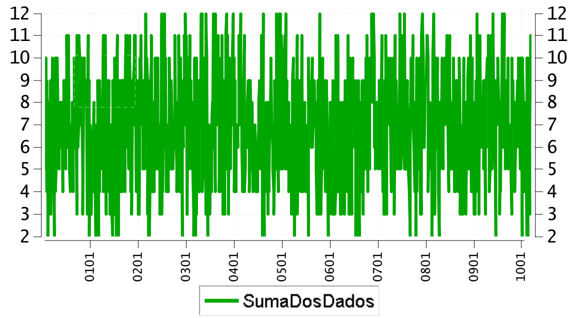
Conceito de Modelação Empírica

Def.) **Modelação Empírica**: Usando modelos estatísticos descrevem-se fenômenos estocásticos observáveis



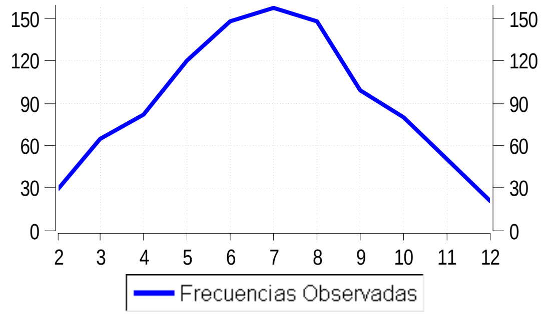
$$\Phi = \{ f(x,\theta) \,,\, \theta \in \Theta \,,\, x \in \Re_x \}$$

Série Temporal:
$z_1, z_2, z_3, .. z_t$



Padrão de regularidade:

Histograma



A ponte entre os padrões de regularidade e os conceitos probabilísticos é transformar o reconhecimento intuitivo de padrões em informação estatística sistemática para ser utilizada na modelação.

$$z_t = f(t,\beta) + a_t$$

onde **f(t,B)** é uma função conhecida determinista do tempo que depende de um vetor de parâmetros **B** e **at** é uma sequência de variáveis aleatórias independentes de média zero e variância constante.

$$z_t = f(t,\beta) + a_t = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + a_t$$

$$a_t -> RB(0, \sigma_a)$$

1. Modelo estocástico: A presença de término de error faz que a relação entre a variável endógena e a explicativa seja estocástica.
2. O modelo que relaciona as variáveis endógena e explicativa é lineal nos coeficientes beta.
3. Os coeficientes beta são constantes no tempo.
4. Existe uma relação causal desde as variáveis explicativas até as variáveis endógenas.
5. As variáveis x são linealmente independentes.
6. As variáveis x são deterministas.

Estimador Mínimos Quadrados Ordinários

$$y_t = X_t\beta + a_t \qquad a_t \therefore RB(0,\sigma_a)$$

Parâmetros desconhecidos do meu modelo beta e sigma

$$\hat{a}_t = y_t - \hat{y}_t = y_t - X_t\hat{\beta}$$

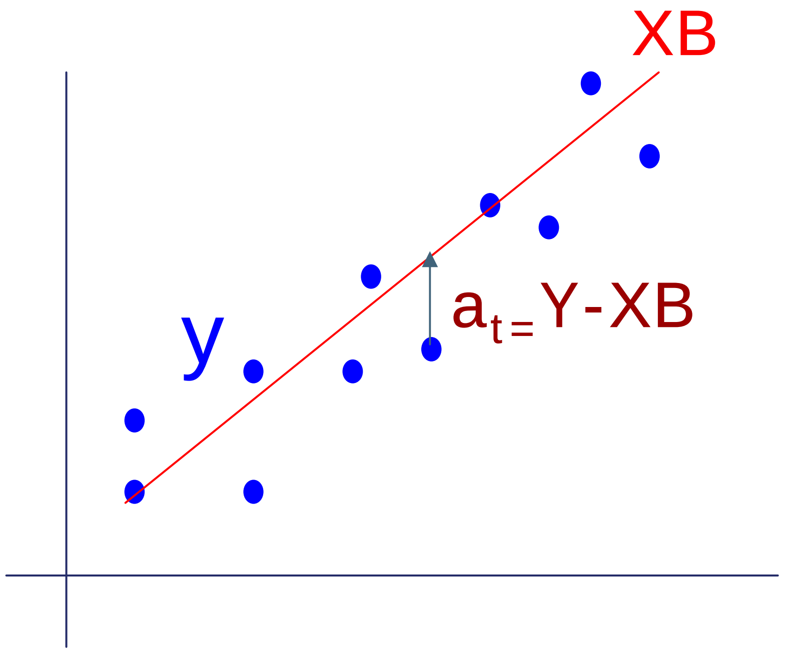
$$SR(\hat{\beta}) = \hat{a}_t'\hat{a}_t = y_t^{'}y_t - 2\hat{\beta}^{'}X^{'}y + \hat{\beta}^{'}X^{'}X\hat{\beta}$$

$$\min_{\beta} SR(\hat{\beta}) = \min_{\beta}(y_t^{'}y_t - 2\hat{\beta}^{'}X^{'}y + \hat{\beta}^{'}X^{'}X\hat{\beta})$$

$$\frac{\partial SR(\hat{\beta})}{\partial\hat{\beta}} = \frac{\partial\hat{a}_t'\hat{a}_t}{\partial\hat{\beta}} = -2\hat{\beta}^{'}X^{'}y + \hat{\beta}^{'}X^{'}X\hat{\beta}$$

$$\frac{\partial SR(\hat{\beta})}{\partial\hat{\beta}} = 0 \quad \Rightarrow \hat{\beta} = (X^{'}X)^{-1}X^{'}y$$

$$\frac{\partial^2 SR(\hat{\beta})}{\partial\hat{\beta}\partial\hat{\beta}^{'}} = X^{'}X \quad def\ +$$



Propriedades do Estimador MCO

Se $E(a_t)=0_T$ então o estimador MCO é insesgado e $E(\beta)=\beta$

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + a_t)$$

$$E(\hat{\beta}) = E[\beta + (X'X)^{-1}X'a_t] = \beta + (X'X)^{-1}X'E[a_t] = \beta$$

$$\hat{\beta} - \beta = (X'X)^{-1}X'a_t$$

$$Var(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)]$$

$$Var(\hat{\beta}) = E[(X'X)^{-1}X'a_t a'_t X(X'X)^{-1}] = (X'X)^{-1}X'E[a_t a'_t]X(X'X)^{-1}$$

$$Var(\hat{\beta}) = (X'X)^{-1}X'(\sigma_a^2 I_T)X(X'X)^{-1} = \sigma_a^2 (X'X)^{-1}$$

$$\beta \therefore N(\hat{\beta}, \sigma_a^2 (X'X)^{-1})$$

É necessário um estimador da variância, se demonstra que:

$$\hat{\sigma}_a^2 = \frac{a'_t a_t}{T-k} \qquad E(\hat{\sigma}_a^2) = E\left(\frac{a'_t a_t}{T-k}\right) = \sigma_a^2$$

Métodos Tradicionais

Os métodos atuais de análises de séries temporais entende-se melhor se connhecemos as limitações de outros métodos mais simples desenvolvidos para solucionar o problema da predição.

Modelo com Tendência Determinista

$$z_t = \mu + \beta_1 t + a_t$$

Modelo com Sazonalidade Cíclica Determinista

$$z_t = \mu + A.sen(wt) + B.\cos(wt) + a_t$$

Modelo de Ajuste de Múltiplos Ciclos Deterministas

$$z_t = \mu + \sum_{j=1}^{k} A_j sen(w_j t) + \sum_{j=1}^{k} B_j \cos(w_j t) + a_t$$

Modelo com Tendências Deterministas: $Z_t = b_o + b_1 t + a_t$

Dada a série de **consumo elétrico anual** procedemos a representar-lo mediante um modelo estatístico consistente em uma constante, uma tendência dependente do tempo e uma variável aleatória de média zero e variância constante. Nosso objetivo é realizar uma previsão de demanda de consumo para o ano seguinte.

$$z_t = f(t, \beta) + a_t = \beta_o + \beta_1 t + a_t$$

O modelo dá um grande ajuste !?!

Para isto realizamos uma regressão lineal com os seguintes resultados

Parâmetros Estimados

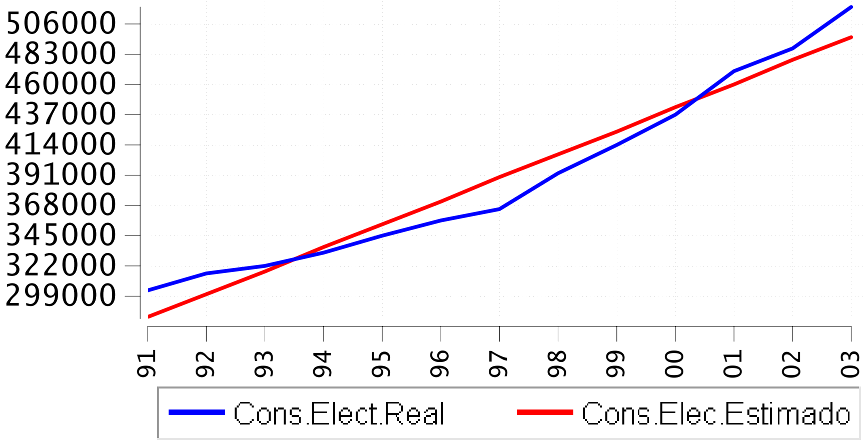
	Name	Value	StDs	TStudent	RefuseProb
B0	cte	265093.5	8403.6	31.5	6.49E-13
B1	trend	17745.6	1058.8	16.8	1.09E-09

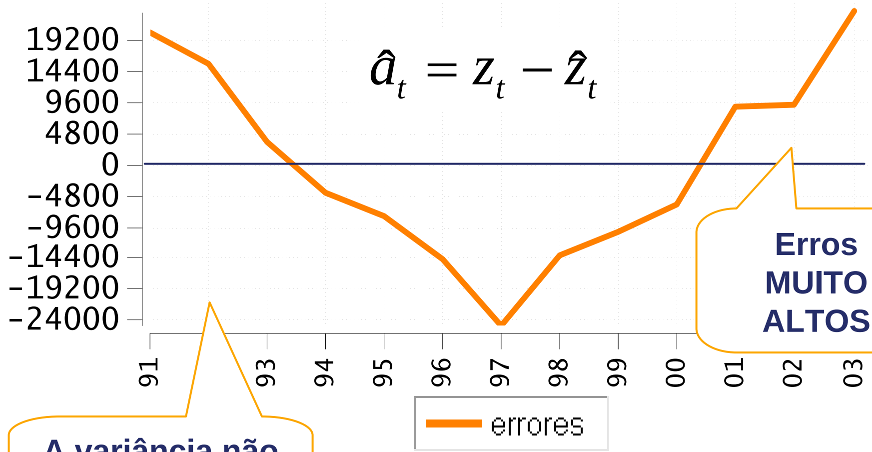
R2	**0.955771**
Desv. Est.	14283.3717

Os resíduos contém claramente informação, fato que nos leva a concluir que o modelo está mal especificado.

Modelo de Tendência Lineal

$$\hat{z}_t = 265093.5 + 17745.6\,t$$



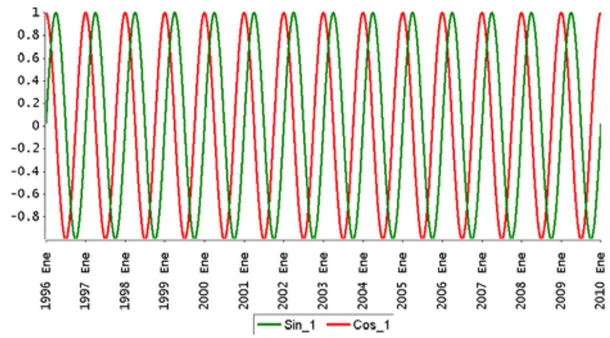


Erros MUITO ALTOS

A variância não é constante!

Um modelo para a **temperatura média diária** é uma estrutura harmônica no tempo da forma

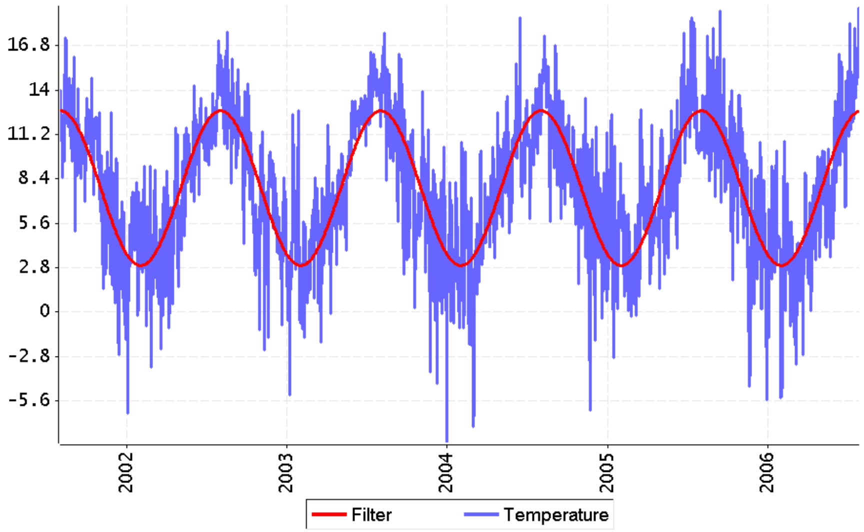
$$z_t = \mu + A sen(\omega t) + B\cos(\omega t) + a_t$$



onde

- μ é uma constante (**nível**)
- **A** e **B** dão os desvios com respeito a μ
- $\omega = \dfrac{2\pi}{T}$ é a frequência (T = 365 dias , T=12, T=52)

Realizamos uma regressão lineal com os seguintes resultados



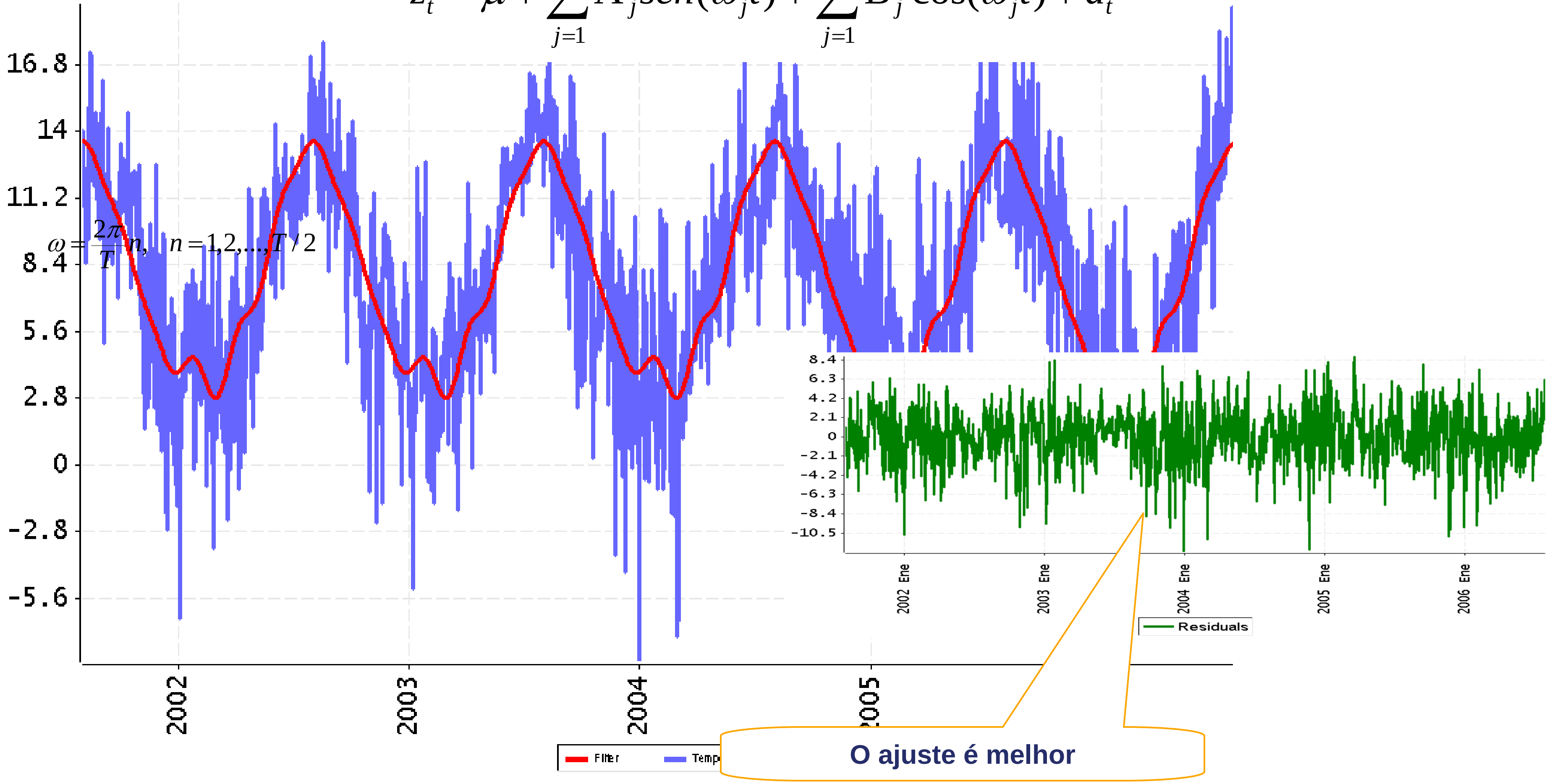
	Value	StDs	TStudent	RefuseProb
μ	7.82323835	0.06754142	115.828749	0
A	-2.57632028	0.09547423	-26.9844582	0
B	-4.14312321	0.09556174	-43.3554605	0

Modelo com múltiplos ciclos de Fourier

Uma função periódica pode descompor-se como uma superposição de funções harmônicas de distintas frequências e amplitudes

$$z_t = \mu + \sum_{j=1}^{k} A_j sen(\omega_j t) + \sum_{j=1}^{k} B_j \cos(\omega_j t) + a_t$$

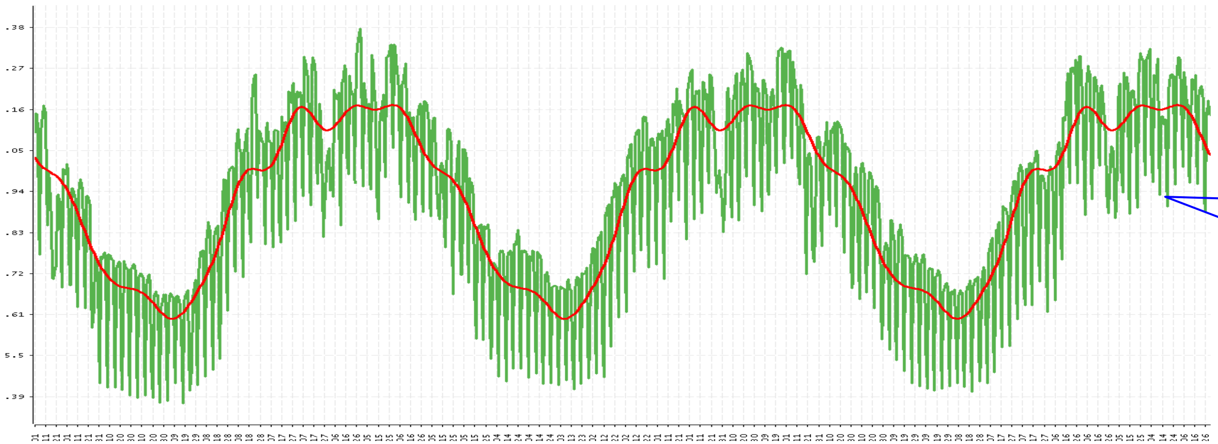
$$\omega = \frac{2\pi}{T} n, \quad n = 1,2,\ldots,T/2$$



O ajuste é melhor

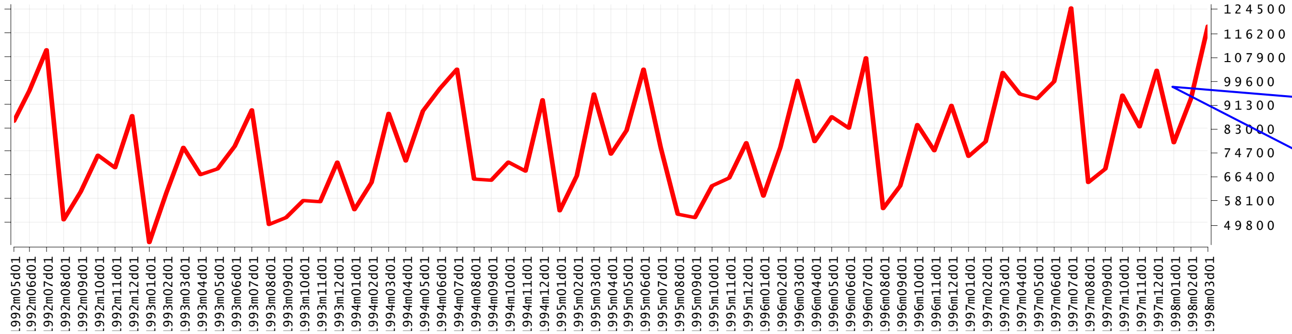
Limitações dos métodos deterministas

A Sazonalidade das séries econômicas pode a vezes ser representada com ciclos de Fourier



O padrão estacional é constante ao largo dos anos no consumo elétrico

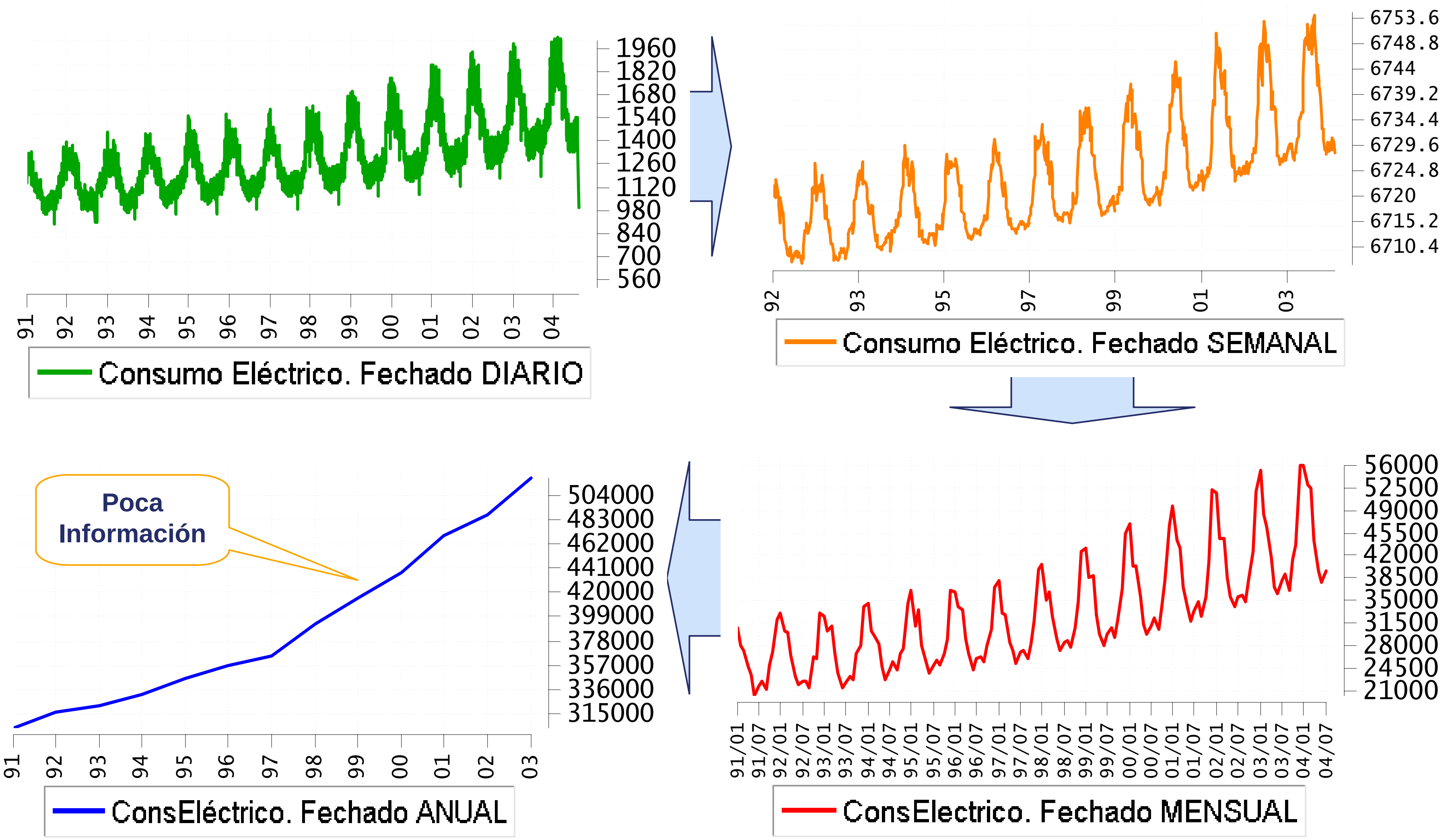
Normalmente as séries apresentam tendências não constantes e padrões estacionais irregulares:



Ainda que apresente estacionalidade não se ajusta bem aos ciclos de Fourier

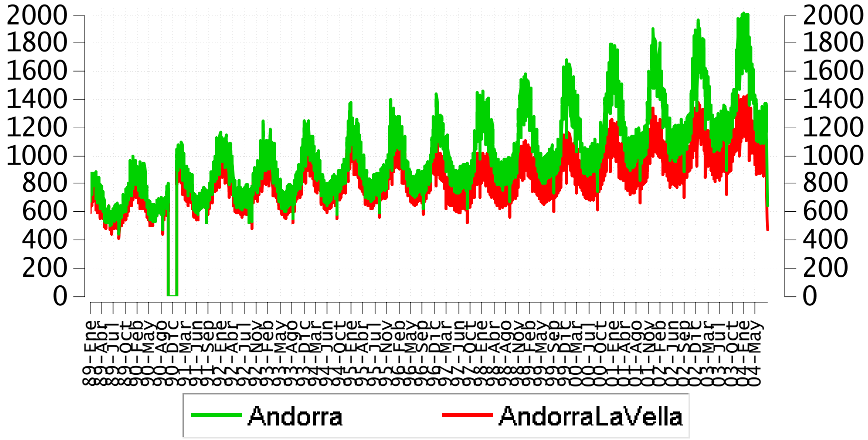
Tipos de períodos

A agregação nos dados sempre conduz a perda de informação. Sempre que seja possível se deve construir modelos sobre séries temporais com a maior profundidade de desagregação.



Objetivo da análise das Séries Temporais

O objetivo da análise de uma série temporal consiste em elaborar um **modelo estatístico** que descreva adequadamente a procedência de dita série, de maneira que as implicações teóricas do modelo resultem **compatíveis** com as **pautas de amostras** observadas nas séries temporais. Depois o modelo elaborado a partir da série temporal pode ser utilizado para **prever** a evolução futura da série ou explicar a relação entre os distintos componentes do modelo.

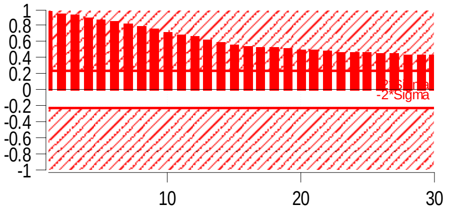


Pautas de Amostras

$$\mu_t = E[z_t]$$

$$\sigma_t^2 = Var[z_t]$$

$$\rho(t,t+j) = \frac{Cov(z_t, z_{t+j})}{\sigma_t \sigma_{t+j}}$$



Modelo Estatístico

$$z_t = \mu + a_t$$

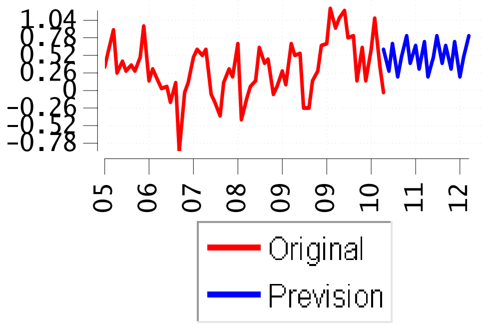
$$z_t = \beta_o + \beta_1 t + a_t$$

$$z_t = \mu + Rsen(wt + \theta) + a_t$$

$$z_t = (1 - \theta_1 B - \ldots - \theta_q B^q) a_t$$

$$(1 - \phi_1 B - \ldots - \phi_p B^p) z_t = a_t$$

Prever, Descrever



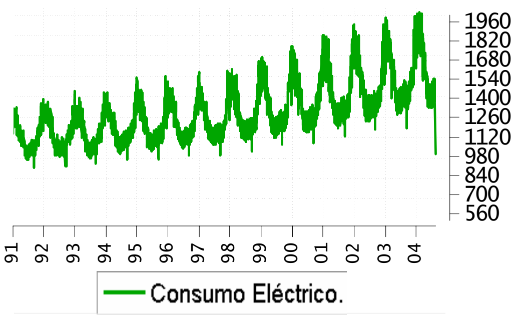
Construção do modelo

Modelo de séries temporais univariantes

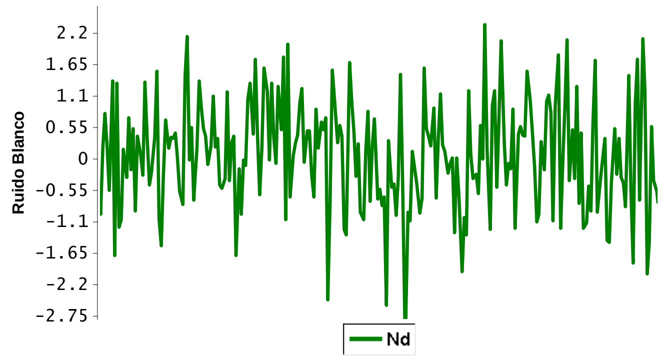
Observações = Filtro (fatores externos) + Ruído ou *Noise* (fatores estruturais)

$$T(X_t) = F_t + \frac{\theta(B)}{\varphi(B)} a_t$$

$$F_t = \sum \alpha_i (B) F_t^i$$



$$\frac{\varphi(B)}{\theta(B)} \left(T(X_t) - F_t \right) = a_t$$



Operadores: Diferença, Retardo, Inverso

Uma propriedade importante dos processos estacionários é que os processos obtidos através das combinações lineares dos processos estacionários são também estacionários.
Ou seja se Z_t é estacionário então o processo $w_t = z_t - z_{t-1}$ é também estacionário.

Operadores

Para poder operar com processos estacionários definimos os siguientes operadores

Def.) **Operador Retardo**: um operador linear que aplicado a uma função temporal proporciona a essa função retardada um período. Em particular se aplicamos o operador a uma série temporal obtenemos a mesma série retardada um período.

$$Bf(t) = f(t-1)$$

$$\boxed{B\,z_t = z_{t-1}}$$

$$B\mu = \mu$$

$$Baz_{t-1} = aBz_t = a\,z_{t-1}$$

$$B^k z_t = B..Bz_t = z_{t-k}$$

$$(1-\phi_1 B - \phi_2 B^2)z_t = z_t - \phi_1 z_{t-1} - \phi_2 z_{t-2}$$

Def.) **Operador Diferença**: é o operador polinômico (1-B). O resultado de aplicar o operador diferença a uma série Z_t com T observações é obter uma nova série com T-1 observações através

$$\nabla = (1-B)$$

$$\boxed{\nabla z_t = (1-B)z_t = z_t - z_{t-1}}$$

$$\nabla^2 z_t = (1-B)^2 z_t = (1-2B+B^2) = z_t - 2z_{t-1} + z_{t-2}$$

$$\nabla_s z_t = (1-B^s)\;\; z_t = z_t - z_{t-s}$$

Def.) **Operadores Inversos**: Verificam a propriedade que o produto por o operador inicial é a unidade. (pe. o inverso do operador retardo é o operador adiantado)

$$\boxed{F\;z_t = B^{-1}z_t = z_{t+1}}$$

$$F\,B = 1$$

$$(1+\phi B + \phi^2 B^2 \ldots.)\,z_t = \sum_{i=0}^{\infty} \phi^i B^i z_{t-i} = (1-\phi B)^{-1} z_t$$

$$(1-\phi B)^{-1}(1-\phi B) = \phi(B)^{-1}\phi(B) = 1$$

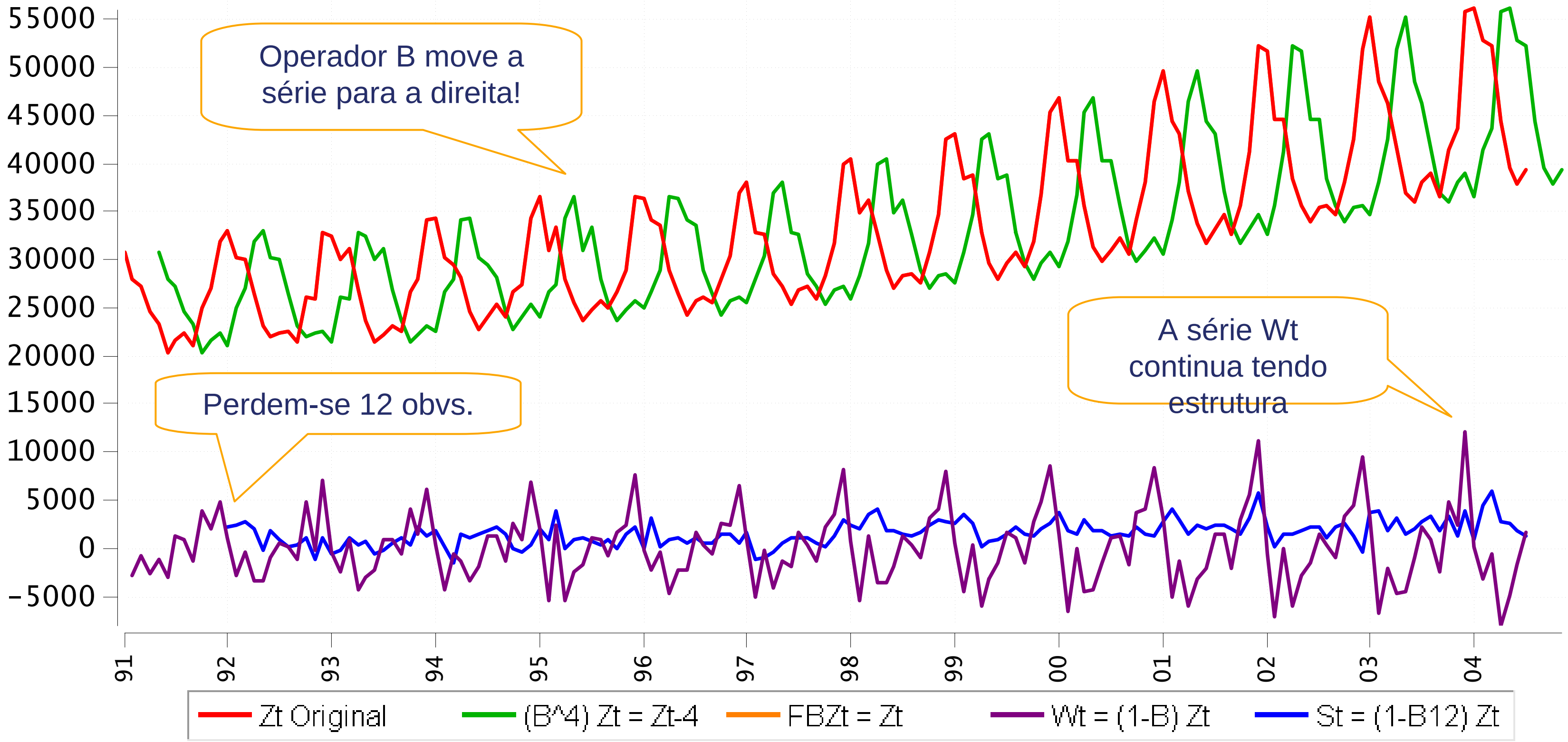
Aplicação de Operadores

Aplicando, à série de **consumo elétrico** em períodos **mensais,** os distintos operadores teremos que:

$B^4 z_t = z_{t-4}$ A série transformada têm a mesma tendência e estrutura estacional que a série original.

$w_t = (1-B)z_t$ A série diferença regular é estacionária em média mas apresenta estrutura estacional.

$S_t = (1-B^{12})z_t$ A série diferença estacional é estacionária em média e não apresenta estrutura estacional.



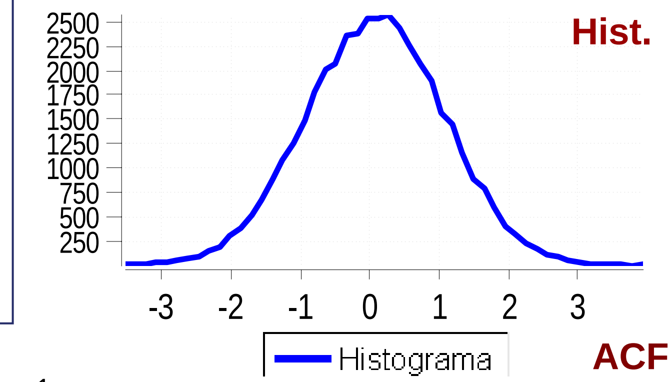
Processo Ruído Branco. ARIMA(0,0,0)

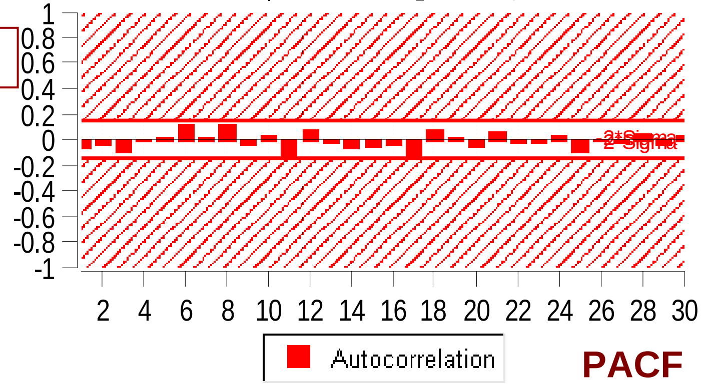
Processo

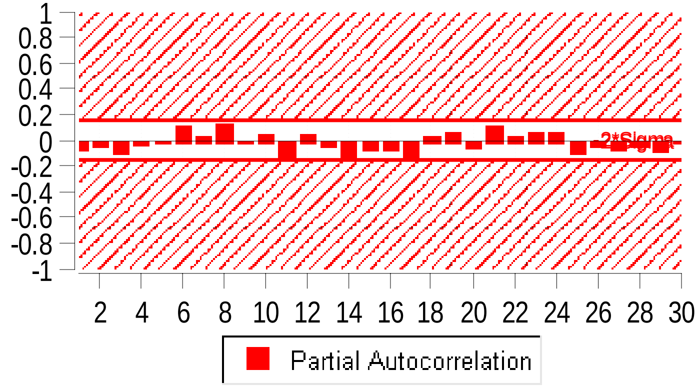
Def.) **Processo Ruído Branco**: é um processo estocástico onde todas as variáveis aleatórias seguem uma distribuição normal de média zero, variância constante e as covariâncias são nulas.

$$z_t = a_t$$

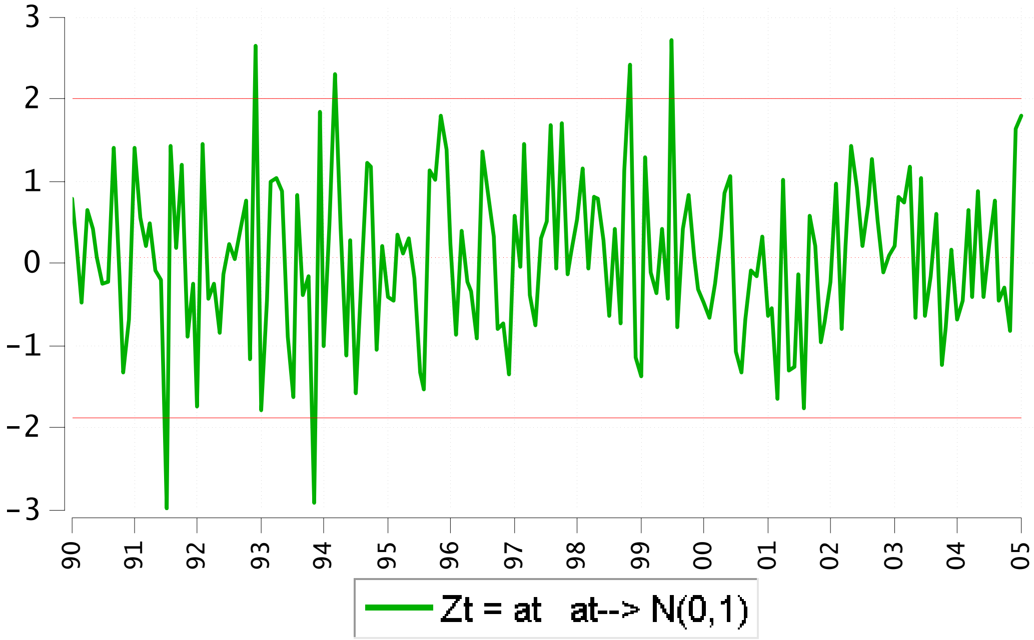
$$E(z_t) = 0$$
$$Var(z_t) = \sigma^2$$
$$Cov(z_t, z_{t-k}) = 0$$







Nota: Um processo ruído branco é estacionário em média e variância.



PROCESSOS ESTACIONÁRIOS

Autorregressivos AR(p)

Processos Autorregressivos AR(1). ARIMA(1,0,0)

Processo

Def.) **Proceso Autorregressivo AR(1) :** é um processo estocástico {zt} que estabelece uma dependência linear sobre o primeiro retardo da variável. Dizemos que uma série Zt segue um processo autorregressivo de primeira ordem se foi gerada por:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + a_t$$

$$= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \phi^3 a_{t-3} \ldots$$

$$(1-\phi_1 B)\tilde{z}_t = a_t$$

$$\tilde{z}_t = 1/(1-\phi_1) a_t$$

Onde –1< ϕ_1 <1 é uma constante a determinar, a_t é um processo ruído branco e $\tilde{z}_t = z_t - c$.

Caracterização

Def.) **Função de Médias de AR(1) :** Tomando esperanças de Z_t e dado que $E(Z_t) = E(Z_{t-1})$=cts posto que o processo é estacionário temos que

$\boxed{E(z_t) = c/(1-\phi_1)}$ já que $\mu = c + \phi_1 \mu$ onde $c = \mu(1-\phi_1)$

Def.) **Função de Autocovariâncias de AR(1) :** se obtém como

$$z_t = \phi_1 z_{t-1} + a_t$$

$E(z_{t-k} z_t) = \phi_1 E(z_{t-k} z_{t-1}) + E(z_{t-k} a_t)$ multiplicado Z_{t-k} e tomando esperanças onde $E(a_t)=0$

$\gamma_k = \phi_1 \gamma_{k-1} \quad k > 0$ onde $\boxed{\gamma_0 = \dfrac{\sigma_a^2}{(1-\phi_1^2)}}$ já que $\sigma_z^2 = \phi^2 \sigma_z^2 + \sigma_a^2$

Def.) **Função de Autocorrelação de AR(1) :** dividindo a função de autocovariâncias pela variância do processo

$$\rho_k = \phi_1 \rho_{k-1}$$

$\boxed{\rho_k = \phi_1^k \quad k \geq 1}$ que é a solução da equação em diferenças com $\rho_0 = 1$

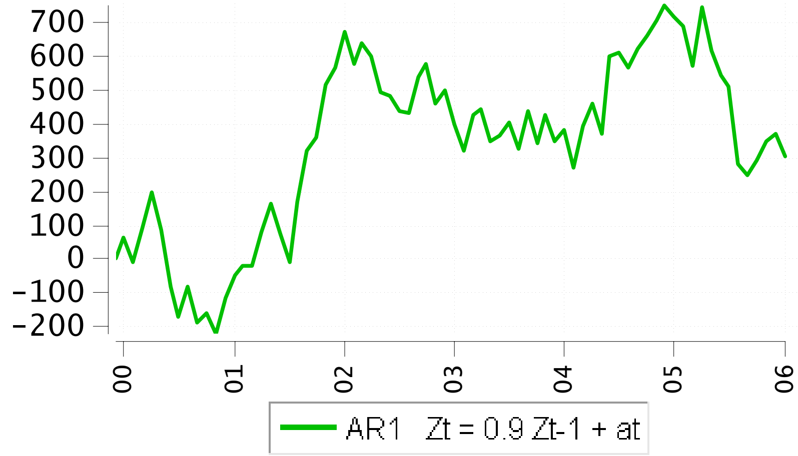
por tanto a função de autocorrelação tende a zero com uma rápidez que depende de ϕ_1 quanto maior seja, menor será o decrescimento, a condição de estacionariedade nos garante que a função de autocorrelação converge.

Processos Autorregressivos AR(1). ARIMA(1,0,0)

Vamos a representar duas realizações de um processo AR(1) com distintos valores do parâmetro e suas funções de autocorrelação teóricas.

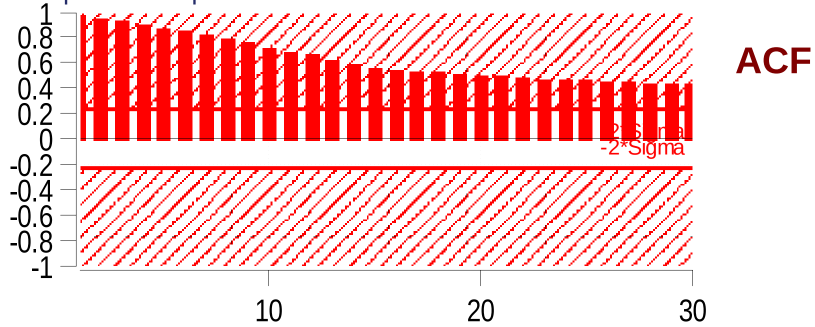
$$(1-0.9B)z_t = a_t$$

Valores próximos tem comportamentos similares e exibem acentuada tendência que se reflexa na ACF



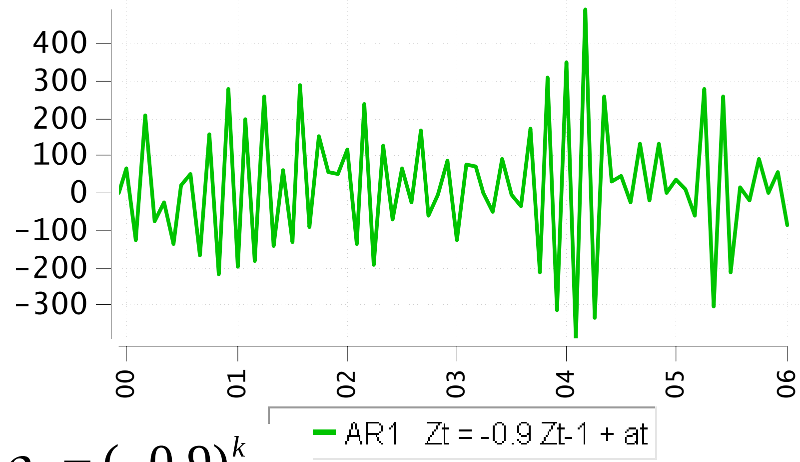
$$\rho_k = 0.9^k$$

Decrescimento lento exponencial em direção ao zero com parâmetro positivo



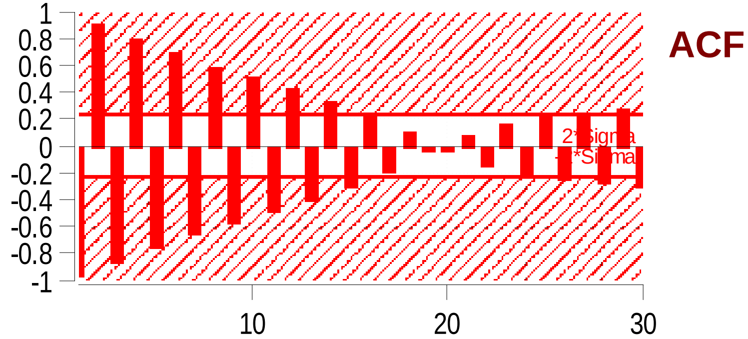
$$(1+0.9B)z_t = a_t$$

A série tende a oscilar e se reflexa na função de autocorrelação que muda de sinal.



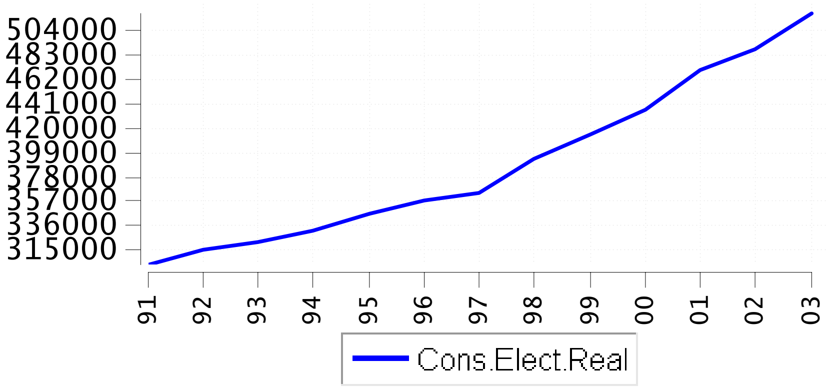
$$\rho_k = (-0.9)^k$$

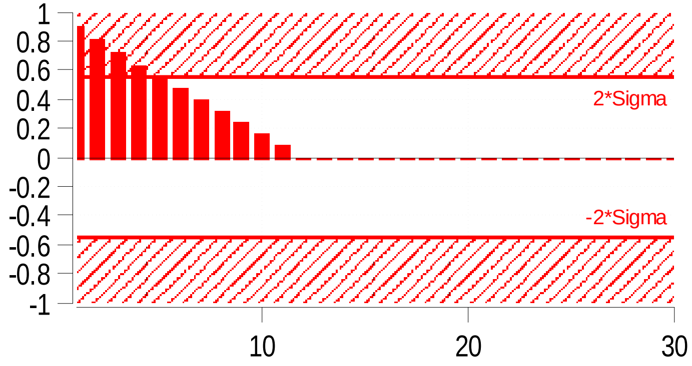
Alternância de sinal



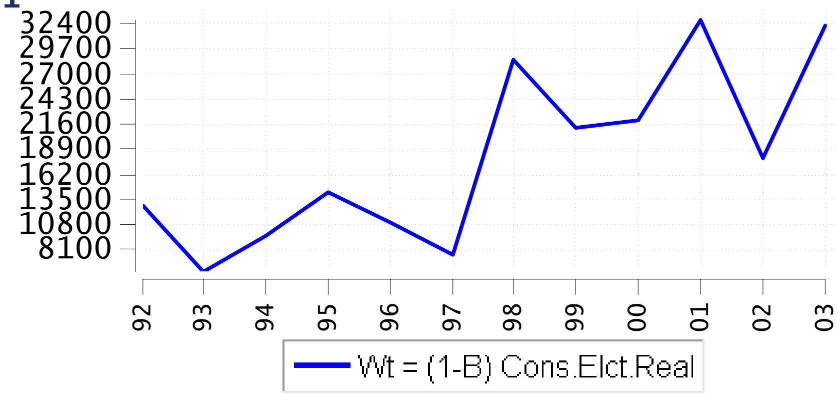
Consumo Elétrico Anual AR(1)

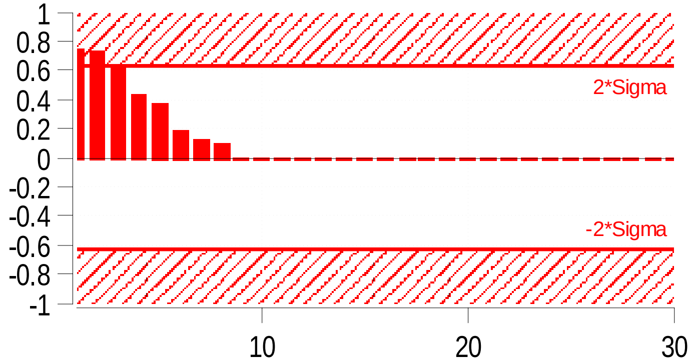
A série de **consumo elétrico** em períodos **anuais** é claramente não estacionária, com um primeiro valor do correlograma muito próximo que indica a necessidade de tomar uma primeira **diferença regular.**





A série de consumo uma vez diferenciada não apresenta uma tendência clara e o correlograma apresenta valores positivos que se amortizam com rapidez, o que nos sugere a existência de um **processo autorregressivo de ordem 1**





Por tanto o **modelo estatístico proposto** para representar o comportamento da série consumo elétrico em períodos anuais é

$$w_t = \phi_1 w_{t-1} + a_t \qquad w_t = (1-B)z_t$$

$$(1-\phi_1 B)\nabla z_t = a_t$$

Consumo Elétrico Anual AR(1)

Dada a série de **consumo elétrico** em períodos **mensais** procedemos a representar-la através de um modelo estatístico consistente em uma diferença regular e um processo autorregressivo de orden 1 tal que:

$$(1-\phi_1 B)\nabla z_t = a_t$$

Realizando uma estimativa por máxima verisimilitude

Name	Value	StDs	TStudent	RefuseProb
RegularAR	0.88149	0.17150	5.13991	0.00061

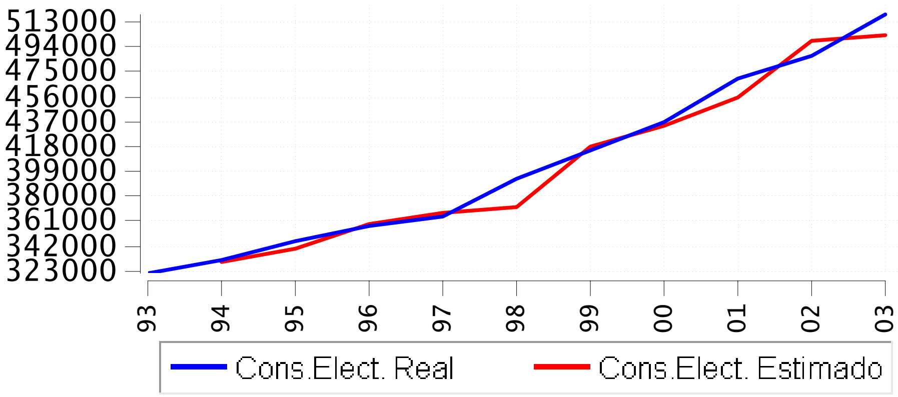
$$(1-0.88149B)(1-B)z_t = a_t$$

Pouco ajuste !?!

Predição de um período para frente é

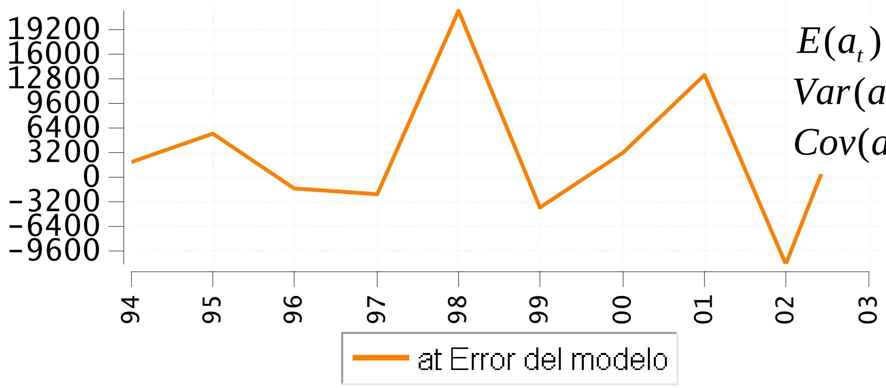
$$\hat{z}_{t+1} = 1.881490\, z_t - 0.881490\, z_{t-1} = 547794.515$$

R2Coeficient	**0.471662**
StandardError	11451.6037

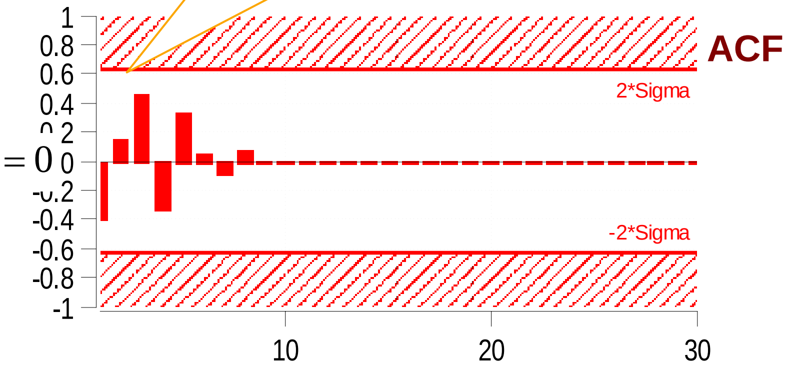


Nota: O R2 é baixo mas o erro padrão é menor que no modelo com tendência determinista.

Não apresenta nenhuma estrutura



$$E(a_t) = 0$$
$$Var(a_t) = \sigma^2$$
$$Cov(a_t, a_{t-k}) = 0$$



Função de Autocorrelação Parcial

Determinar a ordem de um processo autorregressivo a partir de sua função de autocorrelação é difícil ao ser uma mescla de decrescimentos exponenciais e sinusoidais, que se amortizam ao avançar o retardo e não apresenta rasgos fácilmente identificáveis para determinar a ordem do processo. Para resolver este problema introduz-se a função de autocorrelação parcial.

Processo

Def.) **Coeficiente de Autocorrelação Parcial de ordem k.** é o coeficiente de correlação entre observações separadas k períodos quando eliminamos a dependência produzida pelos valores intermediários.

Se elimina de Zt o efeito Zt-1, Zt-2,…, Zt-k+1

$$z_t = \beta_1 z_{t-1} + \ldots + \beta_{k-1} z_{t-k+1} + u_t$$

Se elimina de Zt o efeito Zt-1, Zt-2,…, Zt-k+1

$$z_{t-k} = \gamma_1 z_{t-1} + \ldots + \gamma_{k-1} z_{t-k+1} + v_t$$

Coeficiente de autocorrelação de ordem k

$$Corr(u_t, v_t)$$

Esta definição é análoga a de coeficiente de correlação parcial da regressão múltipla.

$$z_t = \alpha_{11}.z_{t-1} + \eta_{1,t}$$

$$z_t = \alpha_{21}.z_{t-1} + \alpha_{22}.z_{t-2} + \eta_{2.t}$$

$$\ldots.$$

$$z_t = \alpha_{k1} z_{t-1} + \ldots + \alpha_{kk} z_{t-k} + \eta_{k.t}$$

A sequência de coeficientes α_{ii} proporciona a função de autocorrelação parcial (PACF)

Em um processo AR(p) a PACF terá os p primeiros coeficientes distintos de zero.

PROCESSOS ESTACIONÁRIOS

Autorregressivos AR(2)

Processos Autorregressivos AR(2). ARIMA(1,0,0)

Processo

Def.) **Processo Autorregressivo AR(2) :** é um processo estocástico {zt} que estabelece uma dependência linear sobre o primeiro e segundo retardo da variável. Diremos que uma série Zt segue um processo autorregressivo de segunda ordem se foi gerada por:

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \phi_2 \tilde{z}_{t-2} + a_t \qquad (1 - \phi_1 B - \phi_2 B^2)\tilde{z}_t = a_t$$

Onde ϕ_1 ϕ_2 são duas constantes a determinar, a_t é um processo ruído branco e $\tilde{z}_t = z_t - c$.

Caracterização

Def.) **Função de Médias de AR(2) :** Tomando esperanças de Z_t e dado que $E(Z_t) = E(Z_{t-1})$=cts posto que o processo é estacionário temos que

$$\boxed{E(z_t) = c/(1 - \phi_1 - \phi_2)}$$

Def.) **Função de Autocovariâncias de AR(2) :** se obtém como

$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$ multiplicado Z_{t-k} e tomando esperanças onde $E(a_t)=0$ temos

$$E(z_{t-k} z_t) = \phi_1 E(z_{t-k} z_{t-1}) + \phi_2 E(z_{t-k} z_{t-2}) + E(z_{t-k} a_t)$$

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

onde

$$\gamma_1 = \phi_1 \gamma_0 + \phi_2 \gamma_1$$

$$\gamma_0 = \phi_1^2 \gamma_0 + \phi_2^2 \gamma_0 + 2\phi_2 \phi_1 \gamma_1 + \sigma_a^2$$

Variância do processo: para que seja positiva têm-se que cumprir que os parâmetros do processo estejam na região:

$$\gamma_0 = \frac{(1 - \phi_2)\sigma_a^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}$$

$$\boxed{\begin{array}{l} -1 < \phi_2 < 1 \\ \phi_1 - \phi_2 < 1 \\ \phi_1 + \phi_2 < 1 \end{array}}$$

Def.) **Função de Autocorrelação de AR(2) :** dividindo a função de autocovariâncias pela variância do processo

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \qquad k \geq 1 \qquad \textit{como} \quad \rho_{-i} = \rho_i$$

Resolver a Equação em diferenças

para $k = 1$ $\rho_1 = \dfrac{\phi_1}{1 - \phi_2}$, para $k = 2$ $\rho_2 = \dfrac{\phi_1^2}{1 - \phi_2} + \phi_2$,para $k \geq 3$ $\rho_k = A_1 G_1^k + A_2 G_2^k$

Equação em diferenças para um AR(2)

$$\boxed{(1-G_1B)(1-G_2B)z_t = a_t}$$

$$z_t = 1.2\,z_{t-1} - 0.32\,z_{t-2} + a_t$$

Resolver a Equação em diferenças

$$0.32X^2 - 1.2X + 1 = 0$$

$$X = \frac{1.2 \pm \sqrt{1.2^2 - 4\times 0.32}}{0.64} = \frac{1.2 \pm 0.4}{0.64}$$

$$G_1^{-1} = 2.5 \quad G_2^{-1} = 1.25$$

$$G_1 = 0.4 \quad G_2 = 0.8$$

$$(1-0.4X)(1-0.8X) = 0.32X^2 - 1.2X + 1$$

Função de Autocorrelação para o processo AR(2)

$$\boxed{\rho_k = A_1 0.4^k + A_2 0.8^k}$$

$$k = 0 \quad 1 = A_1 + A_2$$

$$k = 1 \quad 0.91 = 0.4A_1 + 0.8A_2$$

$$\boxed{\rho_k = -\frac{0.11}{0.4}0.4^k + \frac{0.51}{0.4}0.8^k}$$

Expressão do processo AR(2) como soma de inovações

$$(1-0.4B)(1-0.8B)z_t = a_t$$

$$z_t = 1.2\,z_{t-1} - 0.32\,z_{t-2} + a_t$$

$$z_t = \frac{1}{(1-0.4B)(1-0.8B)}a_t$$

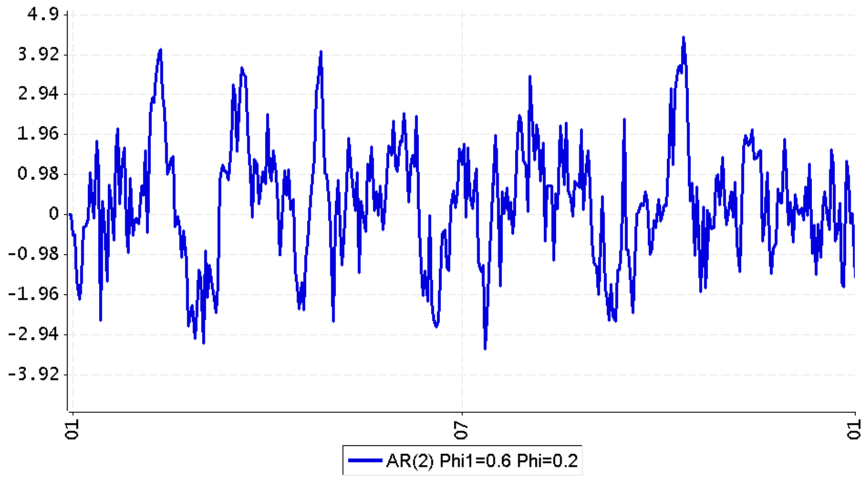
$$\frac{1}{(1-0.4B)} = 1 + 0.4B + 0.4^2B^2 + \ldots$$

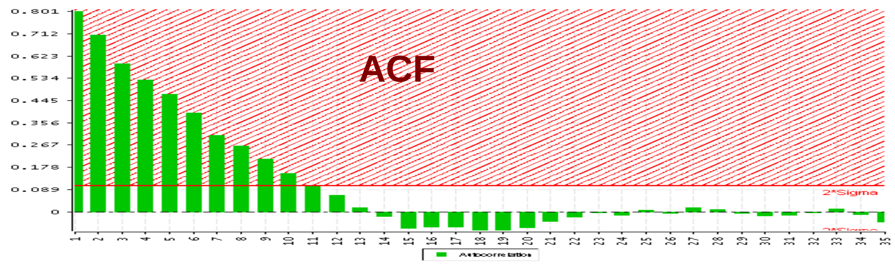
$$\frac{1}{(1-0.8B)} = 1 + 0.8B + 0.8^2B^2 + \ldots$$

$$z_t = (1 + 0.4B + 0.4^2B^2 + \ldots)$$

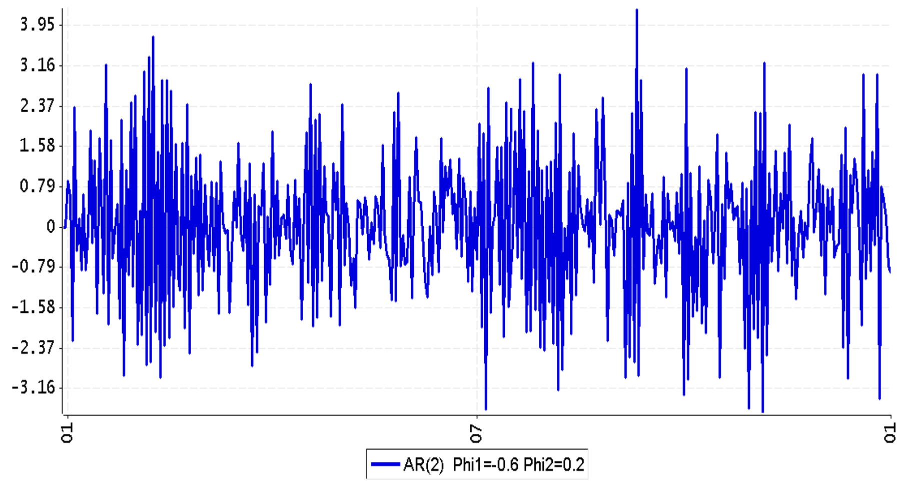
$$(1 + 0.8B + 0.8^2B^2 + \ldots)a_t$$

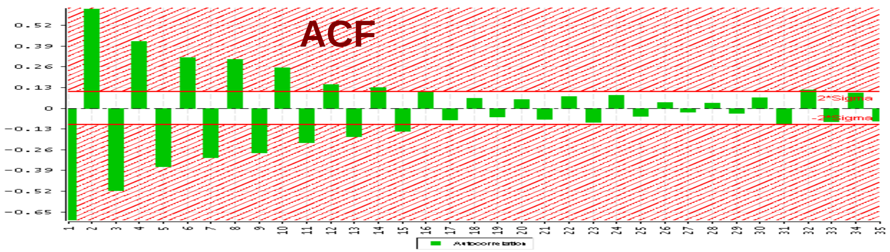
Processos Autorregressivos AR(2)

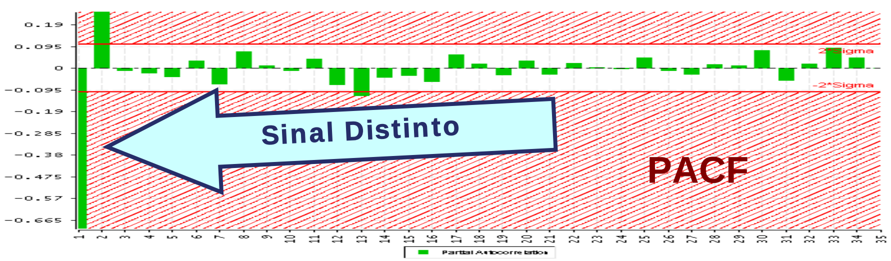




$$(1-0.6B-0.2B^2)z_t = a_t$$ Raízes Reais







$$(1+0.6B-0.2B^2)z_t = a_t$$ Raízes Reais

PROCESSOS ESTACIONÁRIOS

Média Móvel MA(q)

Processo Média Móvel MA(1). ARIMA(0,0,1)

Processo

Def.) **Processo Média Móvel MA(1):** é um processo estocástico {zt} gerado pela combinação linear das duas últimas inovações.

$$\tilde{z}_t = a_t - \theta_1 a_{t-1}$$

$$\tilde{z}_t = (1-\theta_1 B)a_t$$

$$\pi(B)z_t = \sum_{j=0}^{\infty} \theta_1^i B^i z_t = (1+\theta_1 B + \theta_1^2 B^2 + \ldots)\tilde{z}_t = a_t \quad AR(\infty)$$

Representação infinita com coeficientes que decrescem em progressão geométrica, somente é possível a representação se $|\theta_1| < 1$

onde –1< θ_1 <1 é uma constante a determinar, a_t é um processo ruído branco e $\tilde{z}_t = z_t - c$.

Caracterização

Def.) **Função de Autocovariâncias de MA(1) :** se obtém como

$$z_t = a_t - \theta_1 a_{t-1} = \psi(B)a_t$$

$$\gamma(B) = \sigma_a^2 \psi(B)\psi(B^{-1}) = \psi(B)\psi(F)$$

$$\gamma(B) = \sigma_a^2(1-\theta_1 B)(1-\theta_1 F) = \sigma_a^2\left\{-\theta_1 F + (1+\theta_1^2) - \theta_1 B\right\}$$

$$\boxed{\begin{cases} \gamma_0 = \sigma_a^2(1+\theta_1^2) \\ \gamma_1 = -\theta_1\sigma_a^2 \\ \gamma_k = 0 \quad k \geq 2 \end{cases}}$$

Def.) **Função de Autocorrelação de MA(1) :** dividindo a função de autocovariâncias pela variância do processo.

$$\boxed{\begin{cases} \rho_1 = -\theta_1 / (1+\theta_1^2) \\ \rho_k = 0 \quad k > 1 \end{cases}}$$

Por tanto o correlograma do processo terá unicamente um valor distinto de zero no primeiro retardo.

Def.) **Função de Autocorrelação Parcial de MA(1) :** utilizando as equações de Yule-Walker com $\rho_1 = -\theta_1 / (1+\theta_1^2)$ y $\rho_k = 0 \quad k > 1$ depois de manipulações algébricas obtemos que (Box-Cox 70)

$$\boxed{\phi_{kk} = -\theta_1^k \{1-\theta_1^2\} / \{1-\theta_1^{2k+1}\}}$$

Por tanto a função está dominada por um decrescimento exponencial que dependerá do valor de θ_1. Quando θ_1 é positivo a função é negativa e é negativo a função alterna de sinal.

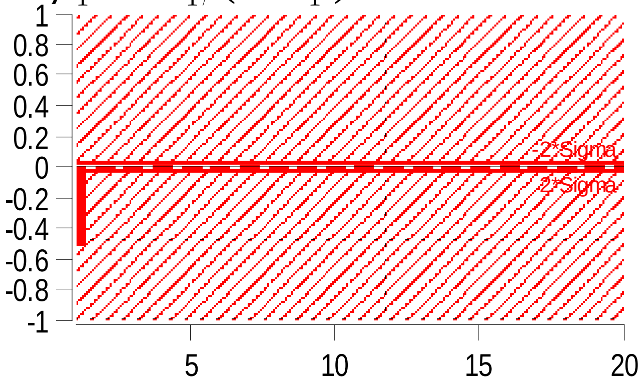
Processos Média Móvel MA(1). ARIMA(0,0,1)

Vamos representar duas realizações de um processo MA(1) com distintos valores de parâmetro e suas funções teóricas de autocorrelação simples e parcial.

$$z_t = a_t - \theta_1 a_{t-1}$$
$$z_t = (1 - 0.99B)a_t$$

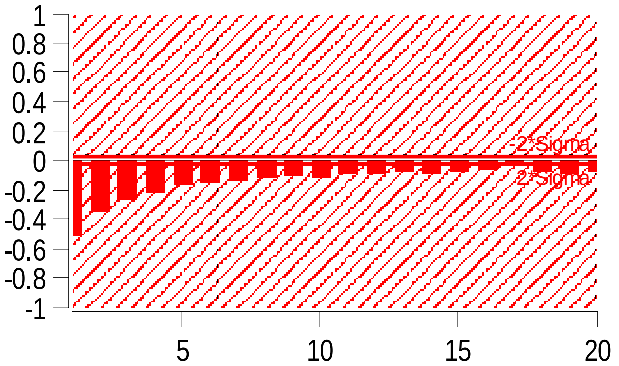
$$\rho_1 = -\theta_1 / (1+\theta_1^2) = -0.4999747$$

ACF



Decrescimento lento exponencial em direção a zero com parâmetro $\phi_{kk} = -\theta_1^k \{1-\theta_1^2\} / \{1-\theta_1^{2k+1}\}$

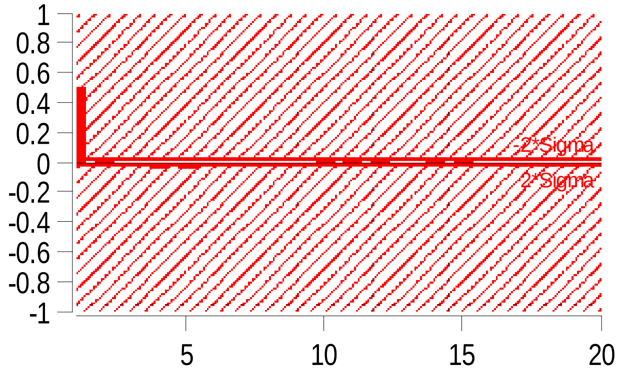
PACF



$$z_t = a_t - \theta_1 a_{t-1}$$
$$z_t = (1 + 0.99B)a_t$$

$$\rho_1 = -\theta_1 / (1+\theta_1^2) = 0.49997475$$

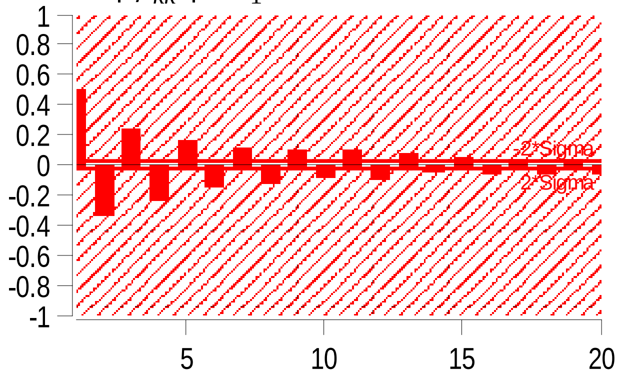
ACF



Alternância de sinal com parâmetro negativo

$$|\phi_{kk}| < \theta_1^k$$

PACF



Comportamento das ACF e PACF de um processo ARIMA(p,d,q)

Ordem	ARIMA(1,d,0)	ARIMA(0,d,1)
comportamento ACF	decai exponencialmente	$\rho_1 \neq 0$
comportamento PACF	$\phi_{11} \neq 0$	Decaimento exponencial
regiao de admisibilidade	$-1 < \phi < 1$	$-1 < \theta < 1$

Ordem	ARIMA(2,d,0)	ARIMA(0,d,2)
comportamento ACF	Misturas de exponenciais e ondas senóides amortecidas	$\rho_1 \neq 0$ $\rho_2 \neq 0$
comportamento PACF	$\phi_{11} \neq 0$ $\phi_{22} \neq 0$	Misturas de exponenciais e ondas senóides amortecidas
região de admisibilidade	$-1 < \phi_2 < 1$ $\phi_2 - \phi_1 < 1$ $\phi_2 + \phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 - \theta_1 < 1$ $\theta_2 + \theta_1 < 1$

Ordem	ARIMA(1,d,1)
comportamento ACF	Decai exponencialmente após o lag 1
comportamento PACF	Decai exponencialmente após o lag 1
região de admisibilidade	$-1 < \phi < 1$ $-1 < \theta < 1$

Processos Estocásticos Não Estacionários

Def.) **Processo Estocástico Não Estacionário** : Um processo estocástico Zt é não estacionário quando as propriedades estatísticas de ao menos uma sequência finita $z_1, z_2, .. z_k$ de componentes de Zt são **diferentes** das da secuencia $z_{1+h}, z_{2+h}, .. z_{k+h}$ para **ao menos um** número inteiro h.
Um processo estocástico Zt é não estacionário quando a distribuição conjunta de qualquier conjunto de variáveis se modifica se modificamos as variáveis no tempo.

$$F(x_i, x_j, .. x_k) \neq F(x_{i+h}, x_{j+h}, \ldots, x_{k+h})$$

$$E(x_t) \neq \mu$$

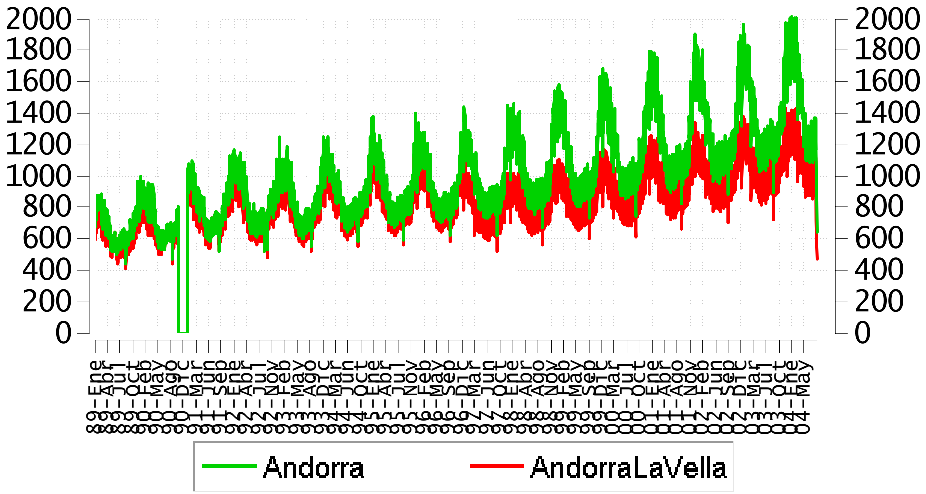
Se o nível da série não é estável no tempo podendo ter tendência crescente ou decrescente diremos que a série **não é estacionária em média**.

$$Var(x_t) \neq \sigma^2$$
$$\gamma(t, t-k) \neq \gamma(t, t+k)$$

Se a variância ou as covariâncias variam com o tempo diremos que a série **não é estacionária nas covariâncias.**

Def.) **Processo Integrado de Ordem h- I(h):** quando ao diferenciar-lo h vezes se obtém um processo estacionário. São processos não estacionários unicamente em média e têm a propriedade de converter-se em estacionários tomando uma diferença. Um processo estacionário é sempre I(0) **.**

Nota:A maioria das séries reais são não estacionárias e seu nível médio varia com o tempo, sem embargo podem converter-se em estacionárias tomando diferenças



Processos Estocásticos Não Estacionários

Passeio Aleatório. ARIMA(0,1,0)

$$(1-B)z_t = a_t$$

Processo Alisamento Exponencial Simples. IMA(1,1).

$$(1-B)z_t = (1-\theta_1 B)a_t$$

Processos Integrados ARIMA(p,d,q)

$$(1-\phi_1 B-\ldots-\phi_p B^p)(1-B)^d \tilde{z}_t = (1-\theta_1 B-\ldots-\theta_q B^q)a_t$$

Processos de Memória Longa

$$(1-B)^d z_t = a_t \qquad -0.5 < d < 0.5$$

Processo ARFIMA(p,d,q)

$$\theta_p(B)\nabla^d z_t = \phi_q(B)a_t \qquad -0.5 < d < 0.5$$

Passeio Aleatório. ARIMA(0,1,0)

Def.) **Passeio aleatório** :é um processo estocástico cujas primeiras diferenças formam um processo ruído branco.

Processo

$$\nabla z_t = a_t$$

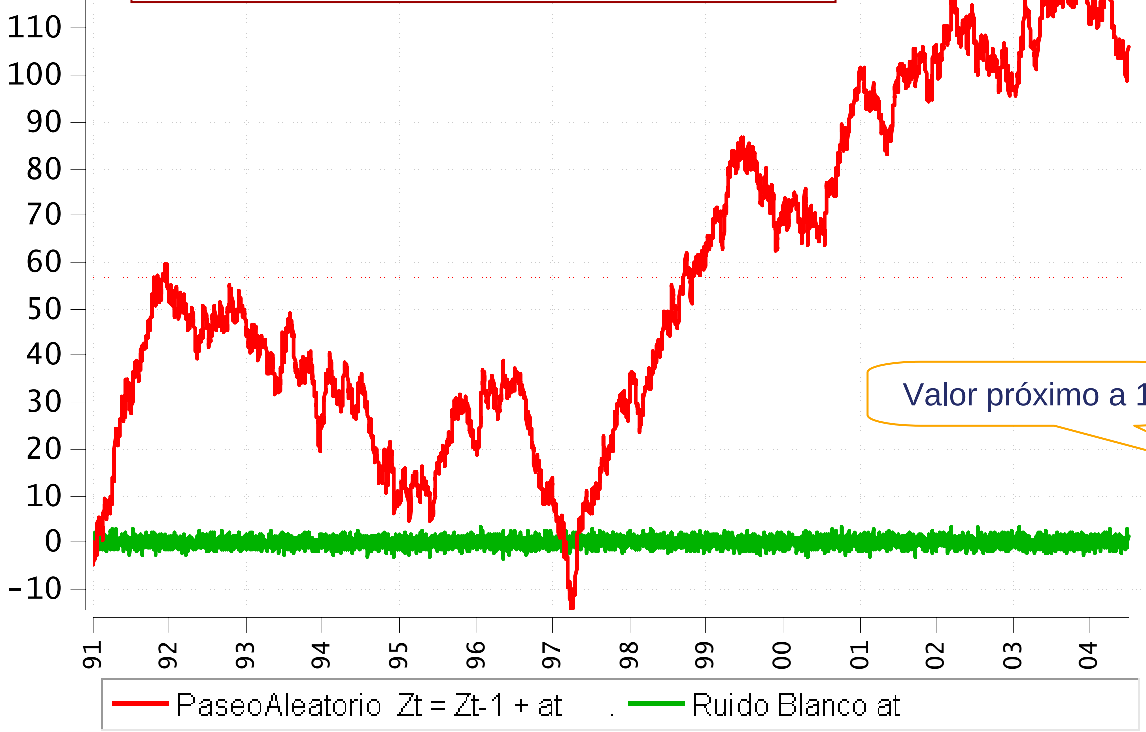
$$z_t = z_{t-1} + a_t$$

$$E(z_t) = 0$$

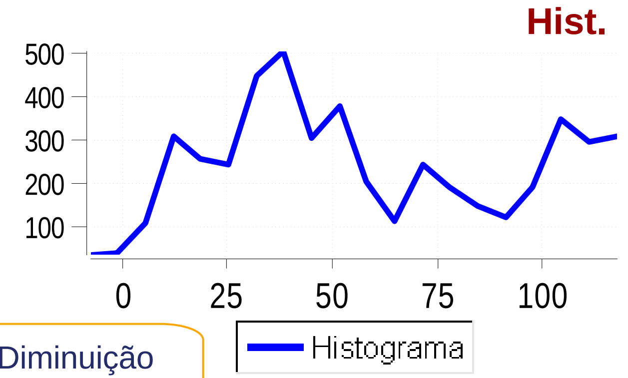
$$Var(z_t) = \sigma_a^2 t$$

$$Cov(z_t, z_{t-k}) = \sigma_a^2 (t-k)$$

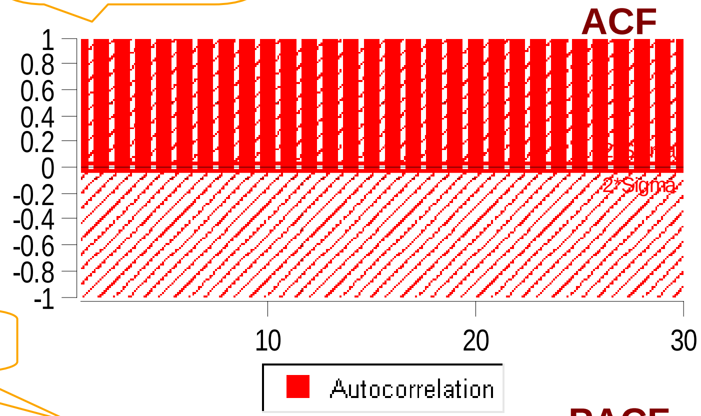
A variância cresce com o tempo



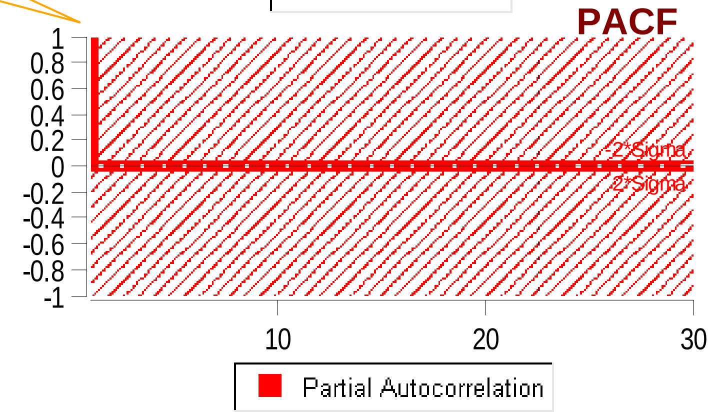
Para el gráfico se ha generado un paseo aleatorio mediante una serie aleatoria en fechado diario de media cero y varianza la unidad.



Diminuição muito lenta



Valor próximo a 1



Série do IBEX-35

A série de cotizações do IBEX_35 em períodos diários é um processo não estacionário em média e em variância. Sem embargo no correlograma se aprecia uma estrutura muito parecida ao correlograma de um passeio aleatório, o que nos leva a propor esta estrutura como modelo para representar a cotização da bolsa.

$$\nabla Ibex35_t = a_t$$

$$Ibex35_t = Ibex35_{t-1} + a_t$$

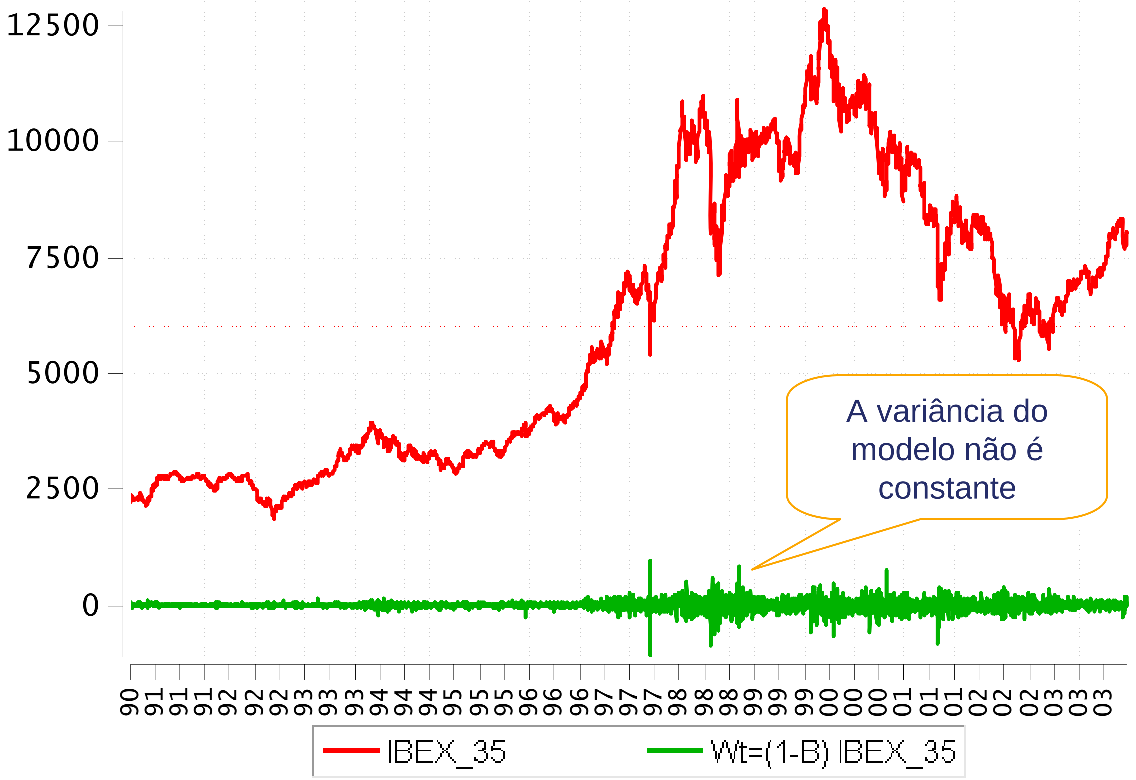
$$E(a_t) = 0$$

$$Var(a_t) \neq \sigma_a^{\ 2}$$

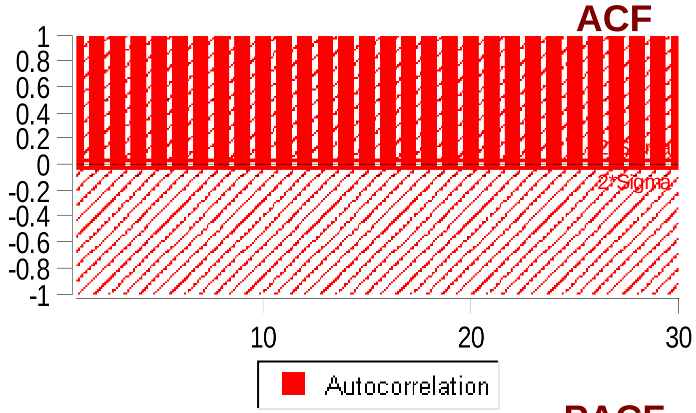
$$Cov(a_t, a_{t-k}) \neq 0$$

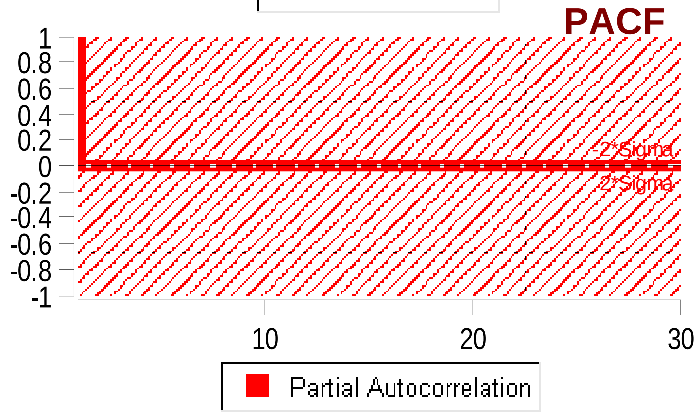
A_t não é um passeio aleatório

Uma vez tomada a diferença regular (1-B) observamos que a série é estacionária em média nu=0 mas não assim em variância que apresenta períodos de muita instabilidade. Uma representação mais adequada do processo deveria introduzir a variância do erro como variável explicativa (p.e.modelo **GARCH**), por tanto não podemos concluir que o IBEX35 seja um passeio aleatório.



A variância do modelo não é constante





Processo Integrado ARIMA(p,d,q)

Processo

Def.) **Processos autorregressivos integrados de média móvel ARIMA(pd,q)** : É um processo ARMA que possui uma ou várias raízes unidade no operador.

$$(1-\phi_1 B-\ldots-\phi_p B^p)(1-B)^d \tilde{z}_t=(1-\theta_1 B-\ldots-\theta_q B^q)a_t$$

$$\phi_p(B)\nabla^d \tilde{z}_t=\theta_q(B)a_t$$

onde a_t é um processo ruído branco com $\tilde{z}_t=z_t-\mu$

Sendo p a ordem da parte autorregressiva , q a ordem da parte média móvel e d o número de raízes unitárias. O processo chama-se integrado porque z_t obtém-se como soma infinita de w_t. Por exemplo se

$$w_t=(1-B)\tilde{z}_t$$

$$\tilde{z}_t=(1-B)^{-1}w_t=(1+B+B^2+B^3\ldots)w_t=\sum_{j=-\infty}^{t} w_t$$

Definimos $\phi_p(B)=(1-\phi_1 B-\ldots-\phi_p B^p)$ como a equação característica do processo autorregressivo

$$\theta_q(B)=(1-\theta_1 B-\ldots-\theta_q B^q)$$

Duas representações alternativas do processo ARIMA como

1.) **Soma de Inovações**:

$$z_t=a_t+\sum_{i=1}^{\infty}\psi_i a_{t-i}=\psi(B)a_t \qquad MA(\infty)$$

$$\varphi_{p+d}(B)=\phi_p(B)(1-B)^d=(1-\varphi_1 B-\ldots-\varphi_{p+d}B^{p+d})$$

$$z_t=a_t+\psi_1 a_{t-1}+\psi_2 a_{t-2}+\ldots=\psi(B)a_t \qquad MA(\infty)$$

$$\varphi_{p+d}(B)\tilde{z}_t=\theta_q(B)a_t$$

$$\varphi_{p+d}(B)\psi(B)=\theta_q(B)$$ Os ψ_i obtém-se igualando os coeficientes em B desta expressão

2.) **Soma de valores passados:**

$$\pi(B)z_t=z_t+\sum_{i=1}^{\infty}\pi_i z_{t-i}=a_t \qquad AR(\infty)$$

$$\varphi_{p+d}(B)\tilde{z}_t=\theta_q(B)a_t$$

$$\varphi_{p+d}(B)=\theta_q(B)\pi(B)$$ Os π_i obtém-se igualando os coeficientes em B desta expressão.