

ANÁLISE DE SÉRIES TEMPORAIS

Ralph dos Santos Silva

<http://www.ralphsilva.org/seriestemporais.html>

Instituto de Matemática
Centro de Ciências Matemáticas e da Natureza
Universidade Federal do Rio de Janeiro

Referências

- Brockwell, P. J. e Davis, R. A. (2010). *Introduction to Time Series and Forecasting*, 2nd Edition, Springer.
- Morettin, P. A. e Toloi, C. M. C. (2004). *Análise de Séries Temporais*, Editora Edgar Blücher.
- West, M. e Harrison, J. (1997). Bayesian Forecasting and Dynamic Models.

Ementa

- Processos estacionários;
- Modelos ARMA e ARIMA;
- Modelagem e previsão com processos ARMA;
- Modelos de séries temporais não estacionários e sazonal; e
- Modelos em espaço de estados (modelos dinâmicos).

Introdução

Uma **Série Temporal** é um conjunto de observações y_t , cada uma observada no tempo específico t .

Uma série temporal é tempo-discreta se o conjunto de índices de tempo for um conjunto discreto, por exemplo $T = \{1, 2, 3, \dots\}$.

Uma série temporal é tempo-contínua se as observações ocorrem continuamente em um intervalo de tempo, por exemplo $T \in [0, 1]$.

Neste curso trataremos de séries temporais a tempo-discreto.

Utilizaremos o programa R e vários de seus pacotes.

- R: <https://www.r-project.org/>

Outros programas são o OpenBUGS e o JAGS (via pacote `rjags` do R).

- OpenBUGS: the-bugs-project-openbugs/
- JAGS: <http://mcmc-jags.sourceforge.net/>

Revisão: alguns conceitos e propriedades

Seja X uma variável aleatória contínua com função de densidade de probabilidade $f(x)$.

Momentos ordinários:

$$E(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x) dx.$$

Definimos $\mu_x = E(X)$ como a **média**.

Momentos centrais:

$$E([X - \mu_x]^\ell) = \int_{-\infty}^{\infty} (x - \mu_x)^\ell f(x) dx.$$

Definimos $\sigma_x^2 = E([X - \mu_x]^2)$ como a **variância**.

Definimos $\sigma_x = \sqrt{\sigma_x^2}$ como **desvio padrão**.

O **coeficiente de assimetria** é definido por

$$S(X) = E \left(\left[\frac{X - \mu_X}{\sigma_X} \right]^3 \right) = \frac{E([X - \mu_X]^3)}{\sigma_X^3}$$

e o **coeficiente de curtose** por

$$K(X) = E \left(\left[\frac{X - \mu_X}{\sigma_X} \right]^4 \right) = \frac{E([X - \mu_X]^4)}{\sigma_X^4}.$$

A quantidade $K(X) - 3$ é chamada de **excesso de curtose** porque $K(X) = 3$ para a distribuição normal¹.

Uma distribuição com excesso de curtose positivo é dita ter **caudas pesadas** (leptocúrtica).

Uma distribuição com excesso de curtose negativo é dita ter **caudas leves** (platicúrtica).

¹Se $X \sim \mathcal{N}(\mu, \sigma^2)$, então $K(X) = 3$.

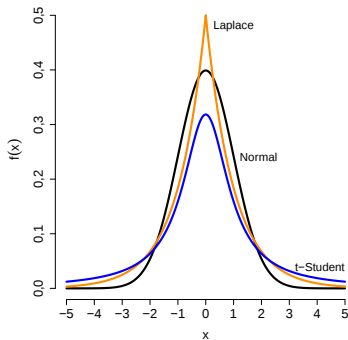


Figura 1: Comparação das funções de densidades de probabilidades da normal, Laplace e t -Student com 1 grau de liberdade.

- A curtose da Laplace é maior que a curtose da normal.
- A curtose da t -Student² é maior que a curtose da Laplace.

²Distribuição t -Student com 1 grau de liberdade ou distribuição de Cauchy.

Momentos amostrais

Suponha que $\{X_1, X_2, \dots, X_n\}$ seja uma amostra aleatória de X com n observações.

- Média amostral:

$$\hat{\mu}_x = \bar{X} = \frac{1}{n} \sum_{t=1}^n X_t.$$

- Variância amostral:

$$\hat{\sigma}_x^2 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})^2.$$

- O coeficiente de assimetria amostral:

$$\hat{S}(X) = \frac{M^3}{\hat{\sigma}_x^3} \quad \text{com} \quad M^3 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})^3.$$

- O coeficiente de curtose amostral:

$$\hat{K}(X) = \frac{M^4}{\hat{\sigma}_x^4} \quad \text{com} \quad M^4 = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})^4.$$

M^ℓ , $\ell = 1, 2, \dots$, podem ser chamados de **momentos centrais amostrais**.

Testes de Assimetria e Curtose

Sob a hipótese de normalidade (**a amostra aleatória é proveniente de uma distribuição normal**), temos que

$$\hat{S}(X) \approx \mathcal{N}(0, 6/n) \quad \text{e} \quad \hat{K}(X) \approx \mathcal{N}(3, 24/n)$$

para n “suficientemente grande”.

Dado uma amostra aleatória $\{X_1, X_2, \dots, X_n\}$, para testar a assimetria da distribuição dos dados, consideramos

$$\begin{cases} H_0 : & S(X) = 0 \\ H_1 : & S(X) \neq 0. \end{cases}$$

A estatística da razão t -Student da assimetria é

$$T = \frac{\hat{S}(X)}{\sqrt{6/n}} \approx \mathcal{N}(0, 1).$$

Rejeitamos H_0 ao nível α de significância se $|T_{obs}| > z_{1-\alpha/2}$, em que $z_{1-\alpha/2}$ é o percentil $100(1 - \alpha/2)$ da distribuição normal padrão.

Para testar o excesso de curtose da distribuição dos dados, consideramos

$$\begin{cases} H_0 : & K(X) - 3 = 0 \\ H_1 : & K(X) - 3 \neq 0. \end{cases}$$

A estatística da razão t -Student do excesso de curtose é

$$T = \frac{\hat{K}(X) - 3}{\sqrt{24/n}} \approx \mathcal{N}(0, 1).$$

Rejeitamos H_0 ao nível α de significância se $|T_{obs}| > z_{1-\alpha/2}$, em que $z_{1-\alpha/2}$ é o percentil $100(1 - \alpha/2)$ da distribuição normal padrão.

Teste de Normalidade de Jarque-Bera

Considere o seguinte teste de hipóteses:

$$\begin{cases} H_0 : & \text{A amostra aleatória é proveniente de uma dist. normal;} \\ H_1 : & \text{A amostra aleatória não é proveniente de uma dist. normal.} \end{cases}$$

A estatística do teste de normalidade de **Jarque e Bera** é dada por

$$JB = \frac{\widehat{S}^2(X)}{6/n} + \frac{(\widehat{K}(X) - 3)^2}{24/n} \approx \chi_2^2,$$

para n “suficientemente grande”.

Rejeitamos H_0 (normalidade) se $JB > \chi_{1-\alpha}^*$, em que $\chi_{1-\alpha}^*$ é o percentil $100(1 - \alpha)$ da distribuição χ_2^2 .

Utilizaremos estes testes - assimetria, curtose e normalidade - ao longo da disciplina.

Séries temporais

Definição

Uma **série temporal** para dados observados $\{y_t\}$ é uma especificação da distribuição conjunta (ou possivelmente somente de médias e covariâncias) de uma sequência de variáveis aleatórias $\{Y_t\}$ na qual $\{y_t\}$ é uma realização.

A distribuição conjunta de uma sequência de variáveis aleatórias $(Y_1, Y_2, \dots, Y_n)$ pode ser expressa por

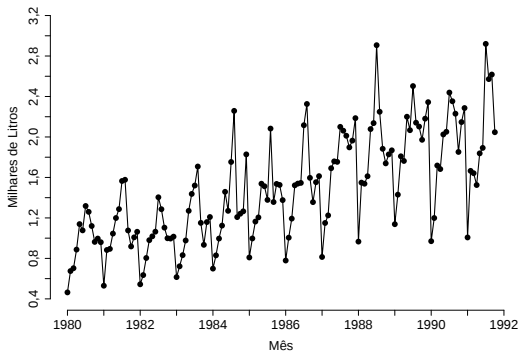
$$\Pr(Y_1 \leq y_1; Y_2 \leq y_2; \dots; Y_n \leq y_n), \quad \text{para} \quad -\infty < y_1, y_2, \dots, y_n < \infty.$$

Observação: a distribuição conjunta pode conter muitos parâmetros para serem estimados!

Uma alternativa é trabalhar com o primeiro e o segundo momentos: $E(Y_t)$ e $E(Y_t Y_{t+h})$, $t = 1, 2, \dots$ e $h = 0, 1, 2, \dots$

A abordagem acima é conhecida como **propriedades de segunda ordem**.

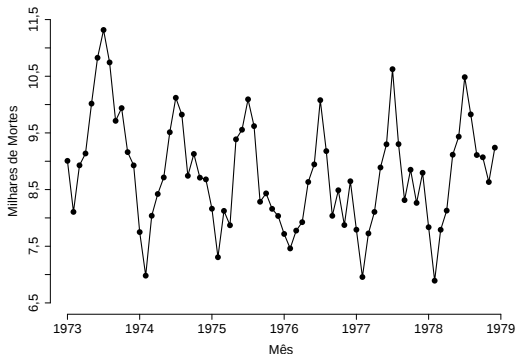
Exemplo: venda mensal de vinho tinto australiano



- Vendas em quilolitros de jan/1980 a out/1991 ($n = 142$ observações).
- Tendência crescente; sazonalidade; altas em julho (inverno); baixas em janeiro (verão).

Ver arquivo exemplo_01.r

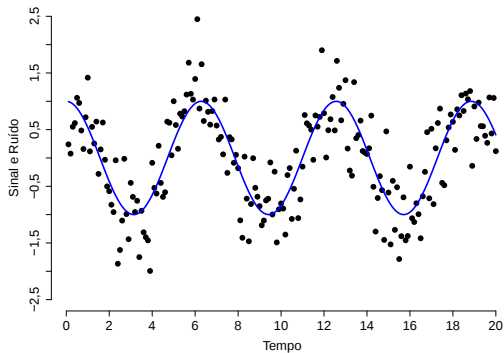
Exemplo: número de mortes mensais por acidente nos EUA



- Milhares de mortes de jan/1973 a dez/1978 ($n = 72$ observações).
- Sazonalidade; altas em julho; baixas em fevereiro.

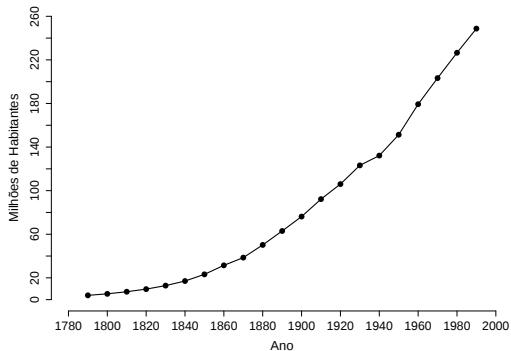
Ver arquivo exemplo_02.r

Exemplo: problema de detecção de sinal



Ver arquivo exemplo_03.r

Exemplo: população dos EUA, 1790-1990



Ver arquivo exemplo_04.r

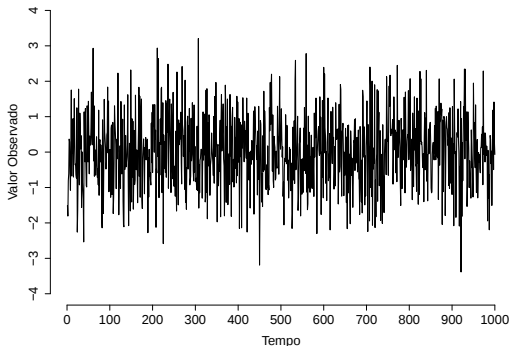
Exemplo: número de greves por ano nos EUA, 1951-1980



Ver arquivo exemplo_05.r

Exemplo: modelo ruído branco

Sequência de variáveis aleatórias Y_t com $E(Y_t) = 0$ (média zero), $\text{Var}(Y_t) = \sigma^2 < \infty$ (variância constante) e $\text{Cov}(Y_t, Y_{t+h}) = 0$ (não correlacionadas) para $t = 1, 2, \dots$ e $h \geq 1$. **Notação:** $Y_t \sim \mathcal{RB}(0, \sigma^2)$.

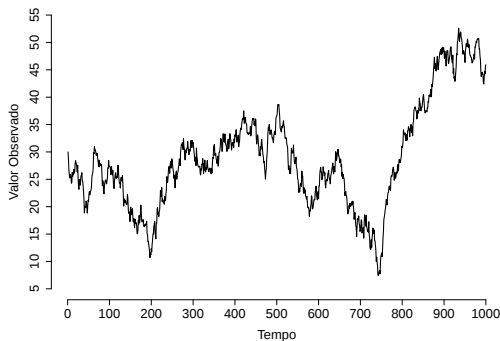


Ver arquivo exemplo_06.r

Exemplo: modelo passeio aleatório ($\mu = 30$)

Soma acumulada de variáveis aleatórias $X_t \sim \mathcal{RB}(0, \sigma^2)$ com $\sigma^2 < \infty$.

$$Y_0 = \mu \quad \text{e} \quad Y_t = \mu + X_1 + X_2 + \cdots + X_t = Y_{t-1} + X_t, \quad \text{para } t = 1, 2, \dots$$



Ver arquivo exemplo_07.r

Modelos com tendência

Recorremos ao exemplo da população dos EUA, 1790-1990.

Apresenta uma tendência quadrática (crescente) e não existe sazonalidade.

Podemos sugerir um modelo da forma:

$$Y_t = m_t + \varepsilon_t,$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e m_t é uma função (contínua) com mudanças suaves.

Para o exemplo acima, podemos sugerir que

$$m_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

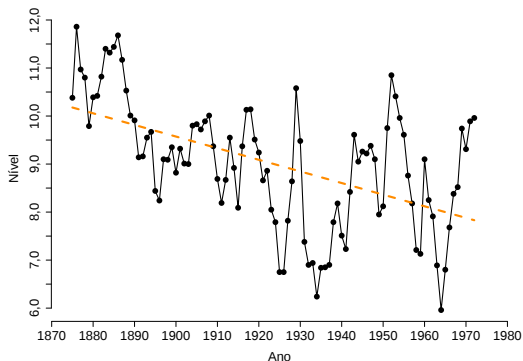
e estimar os parâmetros β_0 , β_1 e β_2 por mínimos quadrados.

Ver arquivo [exemplo_08.r](#)

Tabela 1: Estimativas de mínimos quadrados.

Parâmetro	Estimativa	Erro Padrão	Valor t	Valor p
β_0	6,958	1,9985	3,482	0,00266
β_1	-2,160	0,4184	-5,162	0,00005
β_2	0,651	0,0185	35,223	10^{-16}

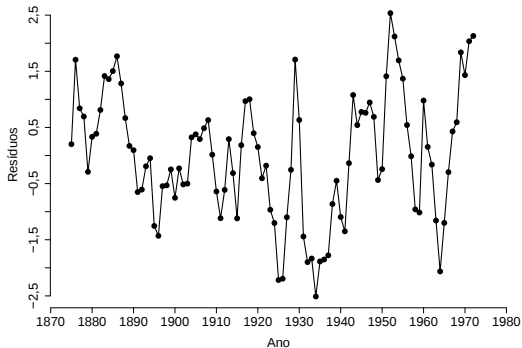
Exemplo: nível do lago Huron



- Possível tendência linear decrescente.
- Sugere-se o modelo $Y_t = m_t + \varepsilon_t$ com $m_t = \beta_0 + \beta_1 t$.

Ver arquivo exemplo_09.r

Exemplo: nível do lago Huron - resíduos da regressão linear



- Verifica-se a ausência de tendência nos resíduos.
- Observa-se certa suavidade dos resíduos - indicada por valores consecutivos somente negativos e outros somente positivos.

(Revise o comportamento de um ruído branco e compare com a série de resíduos acima.)

Modelos com sazonalidade via regressão com harmônicos

Suponha que Y_t seja uma série temporal sem tendência e com sazonalidade.

Um modelo possível é dado por:

$$Y_t = s_t + \varepsilon_t,$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e s_t é uma função periódica de t com período d ($s_{t-d} = s_t$).

Uma escolha conveniente é dada pela regressão harmônica:

$$s_t = \beta_0 + \sum_{j=1}^J [\alpha_j \cos(\omega_j t) + \gamma_j \text{sen}(\omega_j t)]$$

em que

- β_0 , α_j e γ_j , para $j = 1, \dots, J$, são parâmetros desconhecidos; e
- ω_j , para $j = 1, \dots, J$, são frequências fixas, cada uma sendo um múltiplo inteiro de $2\pi/d$.

Exemplo: número de mortes mensais por acidente nos EUA

- Número de observações: $n = 72$;
- Observações mensais;
- Número de harmônicos: $J = 2$; e
- Período: $d = 12$.

Assim, $\omega_1 = \frac{2\pi}{12} = \frac{\pi}{6}$ e $\omega_2 = \frac{4\pi}{12} = \frac{\pi}{3}$. Portanto, fazemos

- $x_{1t} = \cos(\omega_1 t)$;
- $x_{2t} = \sin(\omega_1 t)$;
- $x_{3t} = \cos(\omega_2 t)$; e
- $x_{4t} = \sin(\omega_2 t)$

para $t = 1, \dots, 72$.

Finalmente, ajustamos um modelo de regressão por mínimos quadrados.

Ver arquivo exemplo_10.r

Modelos com sazonalidade via regressão com variáveis binárias

Novamente, temos que

$$Y_t = s_t + \varepsilon_t, \quad \text{em que} \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

Agora, seja d a frequência sazonal (período).

Sejam $D_{1t}, D_{2t}, \dots, D_{dt}$ as variáveis binárias sazonais tal que

$$\left\{ \begin{array}{ll} D_{jt} = 1, & \text{se } t \text{ for o } j\text{-ésimo período} \\ D_{jt} = 0, & \text{caso contrário,} \end{array} \right\} \quad \text{para } j = 1, 2, \dots, d.$$

Em qualquer tempo t somente uma das variáveis binárias $D_{1t}, D_{2t}, \dots, D_{dt}$ assume o valor 1 (as restantes 0).

Suponha que $d = 12$ para dados mensais. Então,

$$s_t = \left\{ \begin{array}{ll} \beta_1, & \text{se } t = \text{janeiro;} \\ \beta_2, & \text{se } t = \text{fevereiro;} \\ \vdots & \vdots \\ \beta_{12}, & \text{se } t = \text{dezembro.} \end{array} \right\} = \sum_{j=1}^{12} \beta_j D_{jt}.$$

Exemplo: número de mortes mensais por acidente nos EUA

- Número de observações: $n = 72$;
- Observações mensais;
- Período: $d = 12$.

Finalmente, ajustamos um modelo de regressão por mínimos quadrados.

Ver arquivo `exemplo_11.r`

Modelagem de séries temporais

Examinar as principais características de uma série temporal através de seu gráfico:

- tendência;
- sazonalidade;
- qualquer mudança brusca no comportamento; e
- qualquer valor aberrante (*outlier*).

Transformar os dados se necessário ($\log(y)$, $\sqrt{y}$) ou Box-Cox.

Escolher um (ou vários) modelo(s) para ajustar a série temporal.

Comparar modelos e escolher o melhor ajuste de acordo com um ou mais critérios (de ajuste e de previsão).

Previsão para as séries temporais transformada e original.

Modelos estacionários e a função de autocorrelação

Definição

Uma série temporal $\{Y_t\}$ é **estritamente estacionária** se a distribuição conjunta de $(Y_t, Y_{t+1}, \dots, Y_{t+k})$ for idêntica a $(Y_{t+h}, Y_{t+1+h}, \dots, Y_{t+k+h})$, para todo t, k e h inteiros positivos.

Agora, focaremos nos momentos de primeira e segunda ordens de $\{Y_t\}$.

Definição

Seja $\{Y_t\}$ uma série temporal tal que $E(Y_t) < \infty$. A função média de $\{Y_t\}$ é

$$\mu_Y(t) = E(Y_t).$$

A função de covariância de $\{Y_t\}$ é

$$\gamma_Y(r, s) = \text{Cov}(Y_r, Y_s) = E((Y_r - \mu_Y(r))(Y_s - \mu_Y(s)))$$

para todos os inteiros r e s .

Definição

Dizemos que $\{Y_t\}$ é uma **série temporal (fracamente) estacionária** se

1. $\mu_Y(t)$ é independente de t , e
2. $\gamma_Y(t+h, t)$ é independente de t para cada h .

Observações

- Uma série temporal $\{Y_t, t = 0, \pm 1, \pm 2, \dots\}$ é estritamente estacionária se $(Y_1, \dots, Y_n)$ e $(Y_{1+h}, \dots, Y_{n+h})$ têm a mesma distribuição conjunta para todos os inteiros h e $n > 0$.
- Se $\{Y_t\}$ é estritamente estacionária com $E(Y_t) < \infty$ para todo t , então $\{Y_t\}$ também é fracamente estacionária.
- Sempre que usarmos o termo função de covariância com referência a uma série temporal estacionária $\{Y_t\}$, estaremos nos referindo na verdade a função de γ_Y , de uma variável, definida por

$$\gamma_Y(h) = \gamma_Y(0, h) = \gamma_Y(t+h, t).$$

$\gamma_Y(\cdot)$ é chamada de **função de autocovariância** e $\gamma_Y(h)$ como seu **valor na defasagem h** .

Definição

Seja $\{Y_t\}$ uma série temporal estacionária.

A **função de autocovariância** (ACVF) de $\{Y_t\}$ na defasagem h é

$$\gamma_Y(h) = \text{Cov}(Y_{t+h}, Y_t).$$

A **função de autocorrelação** (ACF) de $\{Y_t\}$ na defasagem h é

$$\rho_Y(h) = \frac{\gamma_Y(h)}{\gamma_Y(0)} = \text{Cor}(Y_{t+h}, Y_t).$$

Exemplo: ruído i.i.d.

Se $\{Y_t\}$ é uma série temporal ruído i.i.d. tal que $E(Y_t^2) = \sigma^2 < \infty$, então

1. $E(Y_t) = 0$ para todo t ; e
2. $\gamma_Y(h) = \begin{cases} \sigma^2, & \text{se } h = 0; \\ 0, & \text{se } h \neq 0, \end{cases}$

não dependem de t . Logo, $\{Y_t\}$ é estacionária.

Notação: $Y_t \sim IID(0, \sigma^2)$.

Exemplo: ruído branco

Se $\{Y_t\}$ é uma sequência de variáveis aleatórias não correlacionadas, cada uma com média 0 e variância σ^2 , então claramente $\{Y_t\}$ é estacionária com

$$\gamma_Y(h) = \begin{cases} \sigma^2, & \text{se } h = 0; \\ 0, & \text{se } h \neq 0. \end{cases}$$

Esta sequência é chamada de ruído branco.

Notação: $Y_t \sim \mathcal{RB}(0, \sigma^2)$.

Exemplo: passeio aleatório

Se $\{Y_t\}$ é um passeio aleatório com $Y_0 = 0$, então $E(Y_t) = 0$, isto é,

$$E(Y_t) = E(Y_0) + E(X_1) + \cdots + E(X_t) = 0 + 0 + \cdots + 0 = 0.$$

Além disso, $E(Y_t^2) = t\sigma^2$ e $\gamma_Y(h) = t\sigma^2$ para todo t e para $h \geq 0$, isto é,

$$\gamma_Y(h) = \text{Cov}(Y_{t+h}, Y_t) = \text{Cov}(Y_t + X_{t+1} + \cdots + X_{t+h}, Y_t) = \text{Cov}(Y_t, Y_t) = t\sigma^2.$$

Como $\gamma_Y(h)$ depende de t , a série temporal $\{Y_t\}$ não é estacionária.

Exemplo: processo médias móveis de ordem 1

Considere o modelo de médias móveis de ordem 1, MA(1), definido por

$$Y_t = \theta \varepsilon_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e θ é um valor real. Então,

$$\begin{aligned} E(Y_t) &= E(\theta \varepsilon_{t-1} + \varepsilon_t) = \theta E(\varepsilon_{t-1}) + E(\varepsilon_t) = 0; \\ E(Y_t^2) &= E([\theta \varepsilon_{t-1} + \varepsilon_t]^2) \\ &= \theta^2 E(\varepsilon_{t-1}^2) + 2\theta E(\varepsilon_{t-1} \varepsilon_t) + E(\varepsilon_t^2) = \sigma^2(1 + \theta^2); \\ \gamma_Y(h) &= \begin{cases} \sigma^2(1 + \theta^2), & \text{se } h = 0; \\ \sigma^2\theta, & \text{se } h = \pm 1; \\ 0, & \text{se } |h| > 1, \end{cases} \text{ pois} \\ \gamma_Y(1) &= \text{Cov}(Y_t, Y_{t+1}) = \text{Cov}(\theta \varepsilon_{t-1} + \varepsilon_t, \theta \varepsilon_t + \varepsilon_{t+1}) = \theta \sigma^2. \end{aligned}$$

Portanto, $\{Y_t\}$ é estacionária.

A função de autocorrelação é $\rho_Y(h) = \begin{cases} 1, & \text{se } h = 0; \\ \theta/(1 + \theta^2), & \text{se } h = \pm 1; \\ 0, & \text{se } |h| > 1. \end{cases}$

Exemplo: processo autoregressivo de ordem 1

Considere o modelo autoregressivo de ordem 1, AR(1), definido por

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$, $|\phi| < 1$ e ε_t é não correlacionado com Y_s para cada $s < t$, então

$$E(Y_t) = E(\phi Y_{t-1} + \varepsilon_t) = \phi E(Y_{t-1}) + E(\varepsilon_t) = 0 \text{ se } E(Y_t) = E(Y_{t-1}).$$

$$\begin{aligned} \gamma_Y(h) &= \text{Cov}(Y_t, Y_{t-h}) = \text{Cov}(\phi Y_{t-1}, Y_{t-h}) + \text{Cov}(\varepsilon_t, Y_{t-h}) \\ &= \phi \gamma_Y(h-1) + 0 = \phi^h \gamma_Y(0). \end{aligned}$$

$$\rho_Y(h) = \frac{\gamma_Y(h)}{\gamma_Y(0)} = \phi^{|h|}, \text{ para } h = 0, \pm 1, \pm 2, \dots \text{ pois } \gamma_Y(h) = \gamma_Y(-h).$$

$$\begin{aligned} \gamma_Y(0) &= \text{Cov}(Y_t, Y_t) = \text{Cov}(\phi Y_{t-1} + \varepsilon_t, \phi Y_{t-1} + \varepsilon_t) = \phi^2 \gamma_Y(0) + \sigma^2 \\ \Rightarrow \gamma_Y(0) &= \frac{\sigma^2}{1 - \phi^2} \text{ com } |\phi| < 1. \end{aligned}$$

Portanto, $\{Y_t\}$ é estacionária.

Função de autocorrelação amostral

Definição

Sejam $\{y_1, y_2, \dots, y_n\}$ observações de uma série temporal.

A **média amostral** é dada por

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t.$$

A **função de autocovariância amostral** é dada por

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (y_{t+|h|} - \bar{y})(y_t - \bar{y}), \quad -n < h < n.$$

A **função de autocorrelação amostral** é dada por

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad -n < h < n.$$

Observações

1. O divisor n (não $n - |h|$) garante que a matriz de autocovariâncias amostral $\hat{\Gamma}_n = [\hat{\gamma}(i - j)]_{i,j=1}^n$ seja não negativa definida.
2. Assim como a matriz de autocovariâncias amostral, a matriz de autocorrelações amostral $\hat{R}_n = [\hat{\rho}(i - j)]_{i,j=1}^n$ é não negativa definida.
3. As funções de autocovariância e autocorrelação amostrais podem ser calculadas para qualquer conjunto de dados $\{y_1, y_2, \dots, y_n\}$ e não estão restritas as séries temporais estacionárias.
4. O estimador $\hat{\rho}(h)$ é tendencioso para $\rho(h)$. A tendência é da ordem de $1/n$, o que pode ser significativo quando o tamanho amostral n é pequeno.
5. Se $\{Y_t\}$ for uma sequência i.i.d. com $E(Y_t^2) < \infty$, então $\hat{\rho}(h)$ é assintoticamente normal com média 0 e variância $1/n$ para todo inteiro positivo e fixo h , isto é, para n suficientemente grande, temos

$$\hat{\rho}(h)\sqrt{n} \approx \mathcal{N}(0, 1).$$

6. Podemos testar $H_0 : \rho(h) = 0$ contra $H_1 : \rho(h) \neq 0$ para h fixo.

Teste de Portmanteau

Queremos testar $H_0 : \rho(1) = \rho(2) = \dots = \rho(m) = 0$ contra $H_1 : \rho(j) \neq 0$ para algum $j \in \{1, 2, \dots, m\}$.

A estatística do teste é

$$Q^*(m) = n \sum_{j=1}^m [\hat{\rho}(j)]^2, \quad (\text{Box e Pierce}).$$

Sob a hipótese que $\{Y_t\}$ é uma sequência i.i.d. com certas condições nos momentos, $Q^*(m)$ tem distribuição assintótica qui-quadrado com m graus de liberdade.

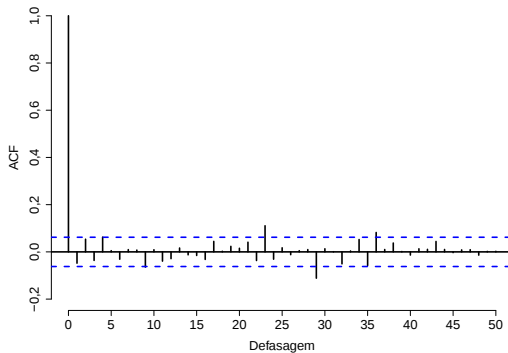
Podemos aumentar o poder do teste ao considerar

$$Q(m) = n(n+2) \sum_{j=1}^m \frac{[\hat{\rho}(j)]^2}{n-j}, \quad (\text{Ljung e Box}).$$

Rejeitamos H_0 se $Q(m) > \chi_\alpha^m$, em que χ_α^m é o percentil $100(1 - \alpha)$ da distribuição qui-quadrado com m graus de liberdade.

Na prática, sugere-se que $m \simeq \ln(n)$.

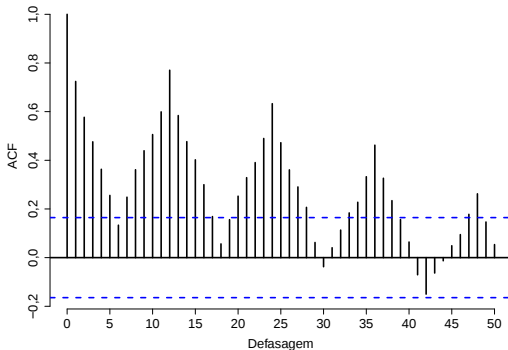
Exemplo: ACF do ruído branco



- Estimativa da função de autocorrelação para dados simulados de um processo ruído branco.

Ver arquivo exemplo_12.r

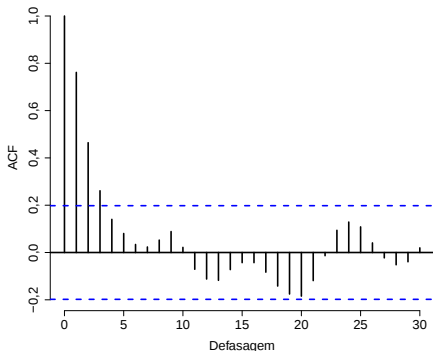
Exemplo: ACF das vendas de vinho tinto australiano



- Estimativa da ACF para os dados de vendas de vinho tinto australiano.
- Com presença de tendência, $|\hat{\rho}(h)|$ decrescerá devagar a medida que h cresce.
- Com presença de sazonalidade, $|\hat{\rho}(h)|$ terá uma periodicidade igual da série.

Ver arquivo `exemplo_13.r`

Exemplo: ACF dos resíduos da regressão do nível do lago Huron



- Uma função constante parece ser um modelo inadequado.
- Autocorrelações significativas em três defasagens.
- Exibe um decaimento geométrico. Temos $\hat{\rho}(1)/\hat{\rho}(0) \simeq 0,76$ e $\hat{\rho}(2)/\hat{\rho}(1) \simeq 0,61$.

Ver arquivo exemplo_14.r

Estimação das componentes de tendência e sazonalidade

O modelo de decomposição clássica é dado por

$$Y_t = m_t + s_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

Trataremos de duas abordagens para remover tendência e sazonalidade:

1. estimando m_t e s_t ; e
2. aplicando diferenças como, por exemplo, $Z_t = Y_t - Y_{t-1}$.

Modelo de tendência sem sazonalidade:

$$Y_t = m_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

Método 1: estimação da tendência

- (a) **Suavização com um filtro de médias móveis finito.** Seja q um inteiro não negativo e considere a média móvel de dois lados

$$Z_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}$$

do processo $\{Y_t\}$ supondo que $Y_t = m_t + \varepsilon_t$.

Então, para $q \leq t \leq n - q$,

$$Z_t = \frac{1}{2q+1} \sum_{j=-q}^q m_{t-j} + \frac{1}{2q+1} \sum_{j=-q}^q \varepsilon_{t-j} \simeq m_t,$$

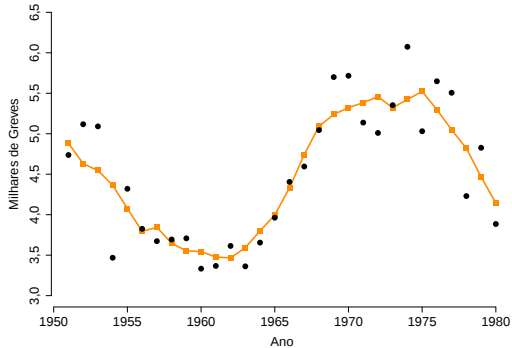
se m_t for aproximadamente linear no intervalo $[t - q, t + q]$ e se a média dos erros neste intervalo estiver próxima de zero.

A média móvel portanto nos fornece estimativas

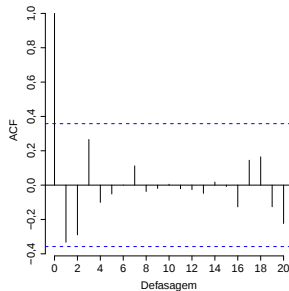
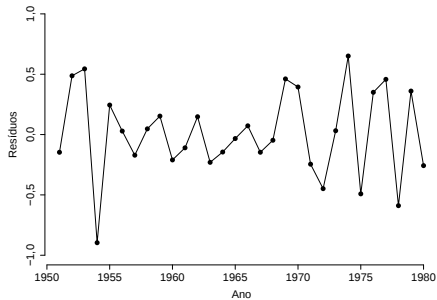
$$\hat{m}_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t-j}.$$

Definimos $Y_t = Y_1$ para $t < 1$ e $Y_t = Y_n$ para $t > n$ para poder aplicar a média móvel em todos os dados.

Exemplo: número de greves por ano nos EUA



- Resultado das médias móveis para $q = 2$.



- Série de resíduos ($e_t = y_t - \hat{m}_t$) via médias móveis.
- ACF dos resíduos (sem dependência significativa).

Ver arquivo exemplo_15.r

- (b) **Suavização exponencial.** Para qualquer valor fixo $\alpha \in [0, 1]$, a média móvel unilateral $\hat{m}_t$, $t = 1, \dots, n$ é definida pelas recursões

$$\begin{aligned}\hat{m}_t &= \alpha Y_t + (1 - \alpha)\hat{m}_{t-1} \\ \hat{m}_1 &= Y_1.\end{aligned}$$

Ver arquivo exemplo_16.r

- (c) **Ajuste polinomial.** Definimos m_t como um polinômio em t da forma

$$m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_k t^k.$$

O ajuste pode ser feito por mínimos quadrados. A série temporal sem tendência é dada pelos resíduos da regressão.

Método 2: eliminação da tendência por diferenciação

O **operador diferença** ∇ de primeira ordem (ou defasagem 1) é definido por

$$\nabla Y_t = Y_t - Y_{t-1} = (1 - B)Y_t,$$

em que **B é o operador de defasagem** (para trás), $BY_t = Y_{t-1}$.

As potências dos operadores acima são definidas de maneira similar:

- $B^j Y_t = Y_{t-j}$, para $j \geq 1$; e
- $\nabla^j Y_t = \nabla(\nabla^{j-1} Y_t)$, para $j \geq 1$ com $\nabla^0 Y_t = Y_t$.

Polinômios em B e ∇ são manipulados de maneira similar as funções polinomiais de variáveis reais. Por exemplo,

$$\begin{aligned}\nabla^2 Y_t &= \nabla(\nabla Y_t) = (1 - B)(1 - B)Y_t \\ &= (1 - 2B + B^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}.\end{aligned}$$

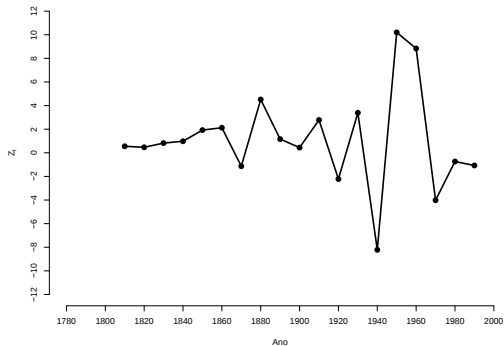
Se $m_t = \beta_0 + \beta_1 t$, então $\nabla m_t = m_t - m_{t-1} = \beta_0 + \beta_1 t - (\beta_0 + \beta_1(t-1)) = \beta_1$.

Se $Y_t = m_t + \varepsilon_t$ em que $m_t = \sum_{j=1}^k \beta_j t^j$, então a aplicação de ∇^k resulta em

$$\nabla^k Y_t = k! \beta_k + \nabla^k \varepsilon_t,$$

um processo estacionário com média $k! \beta_k$.

Exemplo: população dos EUA, 1790-1990



- Resultado da aplicação de duas diferenças sucessivas ($Z_t = \nabla^2 Y_t$).

Ver arquivo exemplo_17.r

Modelos com tendência e sazonalidade

$$Y_t = m_t + s_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$$

- $t = 1, \dots, n$;
- período d ;
- $s_{t+d} = s_t$; e
- $\sum_{j=1}^d s_j = 0$.

Método 1: eliminação via médias móveis

Dado a amostra $(y_1, \dots, y_n)$, a **componente de tendência** é estimada ao aplicar um filtro de médias móveis especialmente escolhido para eliminar a componente de sazonalidade e diminuir o ruído.

Se o período d for par ($d = 2q$), então usamos

$$\hat{m}_t = (0, 5y_{t-q} + y_{t-q+1} + \dots + y_{t+q-1} + 0, 5y_{t+q})/d, \quad q < t \leq n - q.$$

Se o período d for ímpar, então usamos o filtro de médias móveis simples.

O segundo passo é estimar a **componente sazonal**. Para cada $k = 1, \dots, d$, calculamos as médias ω_k dos desvios

$$\{(y_{k+jd} - \hat{m}_{k+jd}), \quad q < k + jd \leq n - q\}.$$

Visto que estes desvios não necessariamente somam zero, estimamos a componente sazonal s_k como

$$\hat{s}_k = \omega_k - \frac{1}{d} \sum_{i=1}^d \omega_i, \quad k = 1, \dots, d$$

e $\hat{s}_k = \hat{s}_{k-d}$, $k > d$.

A série com a **sazonalidade removida** é dada por

$$z_t = y_t - \hat{s}_t, \quad t = 1, \dots, n.$$

Finalmente, reestimamos a componente de tendência por um dos métodos já apresentados usando a série z_t . A série de ruídos é dada por

$$\hat{\varepsilon}_t = y_t - \hat{m}_t - \hat{s}_t, \quad t = 1, \dots, n.$$

Ver arquivo [exemplo_18.r](#)

Método 2: eliminação por diferenciação

Defina o **operador diferença de defasagem** d como

$$\nabla_d Y_t = Y_t - Y_{t-d} = (1 - B^d) Y_t.$$

Não confundir com $\nabla^d = (1 - B)^d$.

Aplicando o operador ∇_d ao modelo

$$Y_t = m_t + s_t + \varepsilon_t,$$

em que $\{s_t\}$ tem período d , obtemos

$$\nabla_d Y_t = m_t - m_{t-d} + \varepsilon_t - \varepsilon_{t-d},$$

que resulta em uma nova série $\nabla_d Y_t$ com tendência $(m_t - m_{t-d})$ e ruído $(\varepsilon_t - \varepsilon_{t-d})$.

A componente de tendência pode, então, ser removida pelos métodos descritos anteriormente, em particular pela aplicação da série de potências de ∇ .

Ver arquivo exemplo_19.r

Processos Estacionários

- Propriedades;
- Processos lineares;
- Introdução aos processos ARMA; e
- Previsão para séries temporais estacionárias.

Propriedades

Já definimos a função de autocovariância como

$$\gamma(h) = \text{cov}(Y_{t+h}, Y_t), \quad \text{para } h = 0, \pm 1, \pm 2, \dots$$

e também a função de autocorrelação como

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}, \quad \text{para } h = 0, \pm 1, \pm 2, \dots$$

A função de autocorrelação é uma medida útil para verificar o grau de dependência (linear) entre valores de uma série temporal.

Por isso, a função de autocorrelação desempenha um papel importante na previsão de valores futuros baseado nos valores passados.

Suponha que $\{Y_t\}$ seja uma série temporal gaussiana e estacionária, e que tenhamos observado Y_n .

Gostaríamos de encontrar uma função de Y_n que nos forneça o melhor preditor, segundo o **erro quadrático médio (EQM)**, para Y_{n+h} .

Então, de acordo com algumas propriedades, temos

$$(Y_{n+h}|Y_n) \sim \mathcal{N}(\mu + \rho(h)(y_n - \mu), \sigma^2(1 - \rho(h)^2)),$$

em que μ e σ^2 são a média e a variância de $\{Y_t\}$, respectivamente.

Sabemos que a constante $c = E(Y_{n+h})$ minimiza $E([Y_{n+h} - c]^2)$.

Além disso, a função $m(Y_n)$ que minimiza $E([Y_{n+h} - m(Y_n)]^2)$ é dada por

$$m(Y_n) = E(Y_{n+h}|Y_n) = \mu + \rho(h)(Y_n - \mu).$$

Logo, o EQM correspondente é dado por

$$E([Y_{n+h} - m(Y_n)]^2) = \sigma^2[1 - \rho(h)^2].$$

Este cálculo mostra que pelo menos para séries temporais gaussianas, a previsão de Y_{n+h} em termos de Y_n é mais precisa a medida que $|\rho(h)|$ se aproxima de 1.

No cálculo acima, a hipótese de normalidade conjunta para Y_n e Y_{n+h} tem um papel importante.

Para séries temporais com outras distribuições conjuntas os cálculos são mais complicados.

Entretanto, ao invés de procurar pela melhor função de Y_n para prever Y_{n+h} , procuraremos pelo melhor **preditor linear**:

$$\ell(Y_n) = a + bY_n.$$

Então, o problema é encontrar os valores de a e b que minimizam

$$E([Y_{n+h} - a - bY_n]^2).$$

Assim, após os cálculos, o melhor preditor linear é dado por

$$\ell(Y_n) = \mu + \rho(h)(Y_n - \mu)$$

com EQM correspondente igual a

$$E([Y_{n+h} - \ell(Y_n)]^2) = \sigma^2(1 - \rho(h)^2).$$

Para processos gaussianos, $m(Y_n)$ e $\ell(Y_n)$ são os mesmos.

Em geral, $m(Y_n)$ resulta em um menor EQM do que $\ell(Y_n)$, pois este é o melhor de uma grande classe de preditores.

O fato do melhor preditor linear só depender da média e da função de autocorrelação da série $\{Y_t\}$ significa que esta pode ser calculada sem maiores conhecimentos da distribuição conjunta.

Isto é extremamente importante na prática dada a dificuldade de se estimar toda a distribuição conjunta e porque é difícil o cálculo dos valores esperados condicionais mesmo quando as distribuições forem conhecidas.

Propriedades de $\gamma(\cdot)$

- $\gamma(0) \geq 0$,
- $|\gamma(h)| \leq \gamma(0)$ para todo h , e
- $\gamma(\cdot)$ é par, isto é, $\gamma(h) = \gamma(-h)$ para todo h .

Definição

Uma função real κ definida nos inteiros é **definida não negativa** se

$$\sum_{i,j=1}^n a_i \kappa(i-j) a_j \geq 0, \quad (1)$$

para todos os inteiros positivos n e vetores $\mathbf{a} = (a_1, \dots, a_n)^\top$ com valores reais a_i .

Teorema

Uma função real definida nos inteiros é uma função de autocovariância de uma série temporal estacionária se, e somente se, ela for par e definida não negativa.

- (a) Uma função de autocorrelação tem todas as propriedades de uma função de autocovariância e satisfaz a condição adicional $\rho(0) = 1$.

Em particular, podemos dizer que $\rho(\cdot)$ é a função de autocorrelação de um processo estacionário se, e somente se, $\rho(\cdot)$ é uma função de autocovariância com $\rho(0) = 1$.

- (b) Para verificar que uma dada função é definida não negativa, em geral é mais simples encontrar um processo estacionário que tenha a dada função como sua função de autocovariância do que verificar as condições da Equação (1) diretamente.

Por exemplo, a função $\kappa(h) = \cos(\omega h)$ é definida não negativa visto que esta é a função de autocovariância do processo estacionário

$$Y_t = A \cos(\omega t) + B \sin(\omega t),$$

em que A e B são variáveis aleatórias não correlacionadas, ambas com média 0 e variância 1.

Definição

$\{Y_t\}$ é um série temporal estritamente estacionária se

$$(Y_1, \dots, Y_n)^\top \approx (Y_{1+h}, \dots, Y_{n+h})^\top$$

para todos os inteiros h e $n \geq 1$.

O símbolo $\approx$ é usado para indicar que dois vetores aleatórios têm a mesma função de distribuição conjunta.

Propriedades elementares

- a. As variáveis aleatórias Y_t são identicamente distribuídas.
- b. $(Y_t, Y_{t+h})^\top \approx (Y_1, Y_{1+h})^\top$ para todos os inteiros t e h .
- c. $\{Y_t\}$ é fracamente estacionária se $E(Y_t^2) < \infty$ para todo t .
- d. **Estacionariedade fraca não implica estacionariedade estrita.**
- e. Uma sequência i.i.d. é estritamente estacionária.

Uma das maneiras mais simples de se construir uma série temporal $\{Y_t\}$ que seja estritamente estacionária (e estacionária se $E(Y_t^2) < \infty$) é “filtrar” uma sequência i.i.d. de variáveis aleatórias.

Seja $\{\varepsilon_t\}$ uma sequência i.i.d. e defina

$$Y_t = g(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q})$$

para alguma função real $g(., \dots, .)$. Então, Y_t é estritamente estacionária visto que

$$(\varepsilon_{t+h}, \dots, \varepsilon_{t+h-q})^\top \approx (\varepsilon_t, \dots, \varepsilon_{t-q})^\top$$

para todo inteiro h .

Segue-se $\{Y_t\}$ é também **dependente de ordem q** , isto é, Y_s e Y_t são independentes sempre que $|t - s| > q$.

De maneira similar, adotando-se um ponto de vista de segunda ordem, dizemos que uma série temporal estacionária é **correlacionada de ordem q** se $\rho(h) = 0$ sempre que $|h| > q$.

O processo médias móveis

Modelo

$\{Y_t\}$ é um processo médias móveis de ordem q se

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2)$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e $\theta_1, \dots, \theta_q$ constantes.

Notação: $Y_t \sim \text{MA}(q)$.

É fácil verificar que a Equação (2) define uma série temporal estritamente estacionária se ε_t for uma sequência i.i.d.

A importância do processo $\text{MA}(q)$ vem do fato de que todo processo correlacionado de ordem q é um processo $\text{MA}(q)$.

Proposição

Se $\{Y_t\}$ é uma série temporal estacionária correlacionada de ordem q com média 0, então esta pode ser representada como um processo $\text{MA}(q)$ na Equação (2).

Processos lineares

A classe de modelos de séries temporais lineares fornece uma abordagem geral para estudar processos estacionários.

Na verdade, cada processo estacionário de segunda ordem é um processo linear ou pode ser transformado em um processo linear ao se remover uma componente determinística.

Definição

A série temporal $\{Y_t\}$ é um **processo linear** se ela tiver a representação

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \quad \text{para todo } t, \quad (3)$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e $\{\psi_j\}$ é uma sequência de constantes com

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Em termos do operador de defasagem B , temos

$$Y_t = \psi(B)\varepsilon_t,$$

em que

$$\psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j.$$

Um processo linear é chamado de médias móveis (infinito) ou $MA(\infty)$ se $\psi_j = 0$ para todo $j < 0$, ou seja,

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}.$$

Observação:

A condição $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ garante que a soma infinita na Equação (3) converge (com probabilidade um), visto que $E(|\varepsilon_t|) \leq \sigma$ e

$$E(|Y_t|) \leq \sum_{j=-\infty}^{\infty} (|\psi_j| E(|\varepsilon_{t-j}|)) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right) \sigma < \infty.$$

A mesma condição garante que $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ e portanto a série converge em média quadrática.

Proposição

Seja ε_t uma série temporal estacionária com média 0 e função de autocovariância γ_ε . Se $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, então a série temporal

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B)\varepsilon_t \quad (4)$$

é estacionária com média 0 e função de autocovariância

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_\varepsilon(h + j - k).$$

No caso especial em que $\{\varepsilon_t\}$ seja um processo linear,

$$\gamma_Y(h) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h},$$

pois $k = h + j$ resulta em $\gamma_\varepsilon(0) = \sigma^2$.

Observação:

A convergência absoluta da Equação (4) implica que os filtros da forma $\alpha(B) = \sum_{j=-\infty}^{\infty} \alpha_j B^j$ e $\beta(B) = \sum_{j=-\infty}^{\infty} \beta_j B^j$ com coeficientes absolutamente somáveis podem ser aplicados sucessivamente a uma série temporal estacionária $\{\varepsilon_t\}$ para gerar uma nova série temporal estacionária

$$W_t = \sum_{j=-\infty}^{\infty} \tilde{\psi}_j \varepsilon_{t-j},$$

em que

$$\tilde{\psi}_j = \sum_{k=-\infty}^{\infty} \alpha_k \beta_{j-k} = \sum_{j=-\infty}^{\infty} \beta_k \alpha_{j-k}.$$

Estas relações podem ser expressas de forma equivalente como

$$W_t = \tilde{\psi}(B) \varepsilon_t,$$

em que

$$\tilde{\psi}(B) = \alpha(B)\beta(B) = \beta(B)\alpha(B).$$

Um processo AR(1)

Um processo AR(1) foi definido como uma solução estacionária $\{Y_t\}$ das equações

$$Y_t - \phi Y_{t-1} = \varepsilon_t, \quad (5)$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$, $|\phi| < 1$ e ε_t não correlacionados com Y_s para cada $s < t$.

Para mostrar que a solução existe e que é a única solução estacionária da Equação (5), consideraremos o processo linear definido por

$$Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \quad (6)$$

É fácil verificar que o processo (6) é a solução da Equação (5), que também é estacionário com média 0 e função de autocovariância

$$\gamma_Y(h) = \sum_{j=0}^{\infty} \phi^j \phi^{j+h} \sigma^2 = \frac{\sigma^2 \phi^h}{1 - \phi^2},$$

para $h \geq 0$.

Para mostrar que a Equação (6) é a única solução estacionária da Equação (5), consideraremos qualquer solução estacionária $\{Y_t\}$. Então, iterando a Equação (5), obtemos

$$\begin{aligned} Y_t &= \phi Y_{t-1} + \varepsilon_t \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 Y_{t-2} \\ &\vdots \\ &= \varepsilon_t + \phi \varepsilon_{t-1} + \cdots + \phi^k \varepsilon_{t-k} + \phi^{k+1} Y_{t-k-1}. \end{aligned}$$

Se $\{Y_t\}$ é estacionária, então $E(Y_t^2)$ é finito e independente de t , de tal forma que

$$E \left(\left[Y_t - \sum_{j=0}^k \phi^j \varepsilon_{t-j} \right]^2 \right) = \phi^{2k+2} E(Y_{t-k-1}^2) \rightarrow 0 \quad \text{quando } k \rightarrow \infty.$$

Isto implica que $\{Y_t\}$ é igual ao limite médio quadrático de $\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$ e portanto o processo definido pela Equação (6) é a única solução estacionária da Equação (5).

Introdução aos processos ARMA

Os processos ARMA são definidos por equações diferenças lineares com coeficientes constantes.

Definição

Uma série temporal $\{Y_t\}$ é um processo ARMA(1,1) se ela for estacionária e satisfazer (para todo t)

$$Y_t - \phi Y_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}, \quad (7)$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e $\phi + \theta \neq 0$.

Utilizando o operador defasagem B , podemos reescrever o processo ARMA(1,1) como

$$\phi(B)Y_t = \theta(B)\varepsilon_t,$$

em que $\phi(B)$ e $\theta(B)$ são filtros lineares

$$\phi(B) = 1 - \phi B \quad \text{e} \quad \theta(B) = 1 + \theta B.$$

Investigaremos os valores de ϕ e θ para que uma solução estacionária da Equação (7) exista.

Se $|\phi| < 1$, denotaremos por $\chi(z)$ a expansão em série de potências de $1/\phi(z)$, isto é, $\sum_{j=0}^{\infty} \phi^j z^j$, que tem coeficientes absolutamente somáveis.

Então, $\chi(B)\phi(B) = 1$ e

$$Y_t = \chi(B)\theta(B)\varepsilon_t = \psi(B)\varepsilon_t,$$

em que

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j = (1 + \phi B + \phi^2 B^2 + \cdots)(1 + \theta B).$$

Além disso, $\psi_0 = 1$ e $\psi_j = (\theta + \phi)\phi^{j-1}$ para $j \geq 1$.

Portanto, concluímos que o processo $MA(\infty)$

$$Y_t = \varepsilon_t + (\theta + \phi) \sum_{j=1}^{\infty} \phi^{j-1} \varepsilon_{t-j}$$

é a única solução estacionária da Equação (7).

Agora, suponha que $|\phi| > 1$. Representamos $1/\phi(z)$ como a série de potências de z com coeficientes absolutamente somáveis pela expansão da potência z^{-1}

$$\frac{1}{\phi(z)} = \frac{1}{1 - \phi z} = -\frac{\frac{1}{\phi z}}{1 - \frac{1}{\phi z}} = -\sum_{j=1}^{\infty} \phi^{-j} z^{-j}.$$

Então, podemos aplicar o mesmo argumento no caso $|\phi| < 1$ para obter uma única solução estacionária da Equação (7).

Seja $\chi(B) = -\sum_{j=1}^{\infty} \phi^{-j} B^{-j}$. Então,

$$Y_t = \chi(B)\theta(B)\varepsilon_t = -\theta\phi^{-1}\varepsilon_t - (\theta + \phi)\sum_{j=1}^{\infty} \phi^{-j-1}\varepsilon_{t+j}.$$

Se $\phi = \pm 1$, então não existe uma solução estacionária da Equação (7).

Logo, não existe um processo ARMA(1,1) com $\phi = \pm 1$ de acordo com a definição.

Uma solução estacionária das equações ARMA(1,1) existe se, e somente se, $|\phi| \neq 1$.

Se $|\phi| < 1$, então a solução estacionária é única. Neste caso, dizemos que $\{Y_t\}$ é **causadora (causal)** visto que Y_t pode ser expressa em termos dos **valores corrente e passados de ε_s , $s \leq t$** .

Se $|\phi| > 1$, então a solução estacionária é única. A solução é **não causadora** visto que Y_t é uma função de ε_s **para $s \geq t$** .

Assim, como a causalidade significa que Y_t pode ser expressa em termos de ε_s para $s \leq t$, a **invertibilidade** significa que ε_t pode ser expressa em termos de Y_s para $s \leq t$.

Agora, mostraremos que o processo ARMA(1,1) será inversível se $|\theta| < 1$.
Seja

$$\xi(z) = \frac{1}{\theta(z)} = \sum_{j=0}^{\infty} (-\theta)^j z^j$$

que tem coeficientes absolutamente somáveis. Logo,

$$\varepsilon_t = \xi(B)\phi(B)Y_t = \eta(B)Y_t,$$

em que

$$\eta(B) = \sum_{j=0}^{\infty} \eta_j B^j = (1 - \theta B + (-\theta)^2 B^2 + (-\theta)^3 B^3 + \dots)(1 - \phi B).$$

Os coeficientes η_j podem ser facilmente calculados e, portanto,

$$\varepsilon_t = Y_t - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{j-1} Y_{t-j}.$$

Assim, o processo ARMA(1,1) é inversível visto que ε_t pode ser expresso em termos dos valores presente e passados do processo Y_s , $s \leq t$.

De maneira similar, podemos mostrar que o processo ARMA(1,1) é **não inversível** se $|\theta| > 1$ visto que

$$\varepsilon_t = -\phi\theta^{-1} Y_t - (\phi + \theta) \sum_{j=1}^{\infty} (-\theta)^{-j-1} Y_{t+j}.$$

Propriedades da média amostral e da função de autocorrelação

Veremos a seguir as propriedades dos estimadores $\bar{Y}$ de μ e $\hat{\gamma}(h)$ de $\gamma(h)$, e consequentemente $\hat{\rho}(h)$ de $\rho(h)$.

Estimação de μ

O estimador dos momentos da média μ de um processo estacionário é a média amostral

$$\bar{Y}_n = \frac{Y_1 + Y_2 + \cdots + Y_n}{n}.$$

Este é um estimador não tendencioso de μ visto que

$$\begin{aligned} E(\bar{Y}) &= \frac{1}{n} E(Y_1 + Y_2 + \cdots + Y_n) \\ &= \frac{E(Y_1) + E(Y_2) + \cdots + E(Y_n)}{n} \\ &= \frac{\mu + \mu + \cdots + \mu}{n} = \mu. \end{aligned}$$

O erro quadrático médio (EQM) de $\bar{Y}_n$ é

$$\begin{aligned} E([\bar{Y}_n - \mu]^2) &= \text{Var}(\bar{Y}_n) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(Y_i, Y_j) \\ &= \frac{1}{n^2} \sum_{i-j=-n}^n (n - |i - j|) \gamma(i - j) \\ &= \frac{1}{n} \sum_{h=-n}^n \left(1 - \frac{|h|}{n}\right) \gamma(h). \end{aligned}$$

Proposição

Se $\{Y_t\}$ é uma série temporal estacionária com média μ e função de autocovariância $\gamma(\cdot)$, então quando $n \rightarrow \infty$

1. $\text{Var}(\bar{Y}_n) = E([\bar{Y}_n - \mu]^2) \rightarrow 0$ se $\gamma(n) \rightarrow 0$; e
2. $nE([\bar{Y}_n - \mu]^2) \rightarrow \sum_{|h| < \infty} \gamma(h)$ se $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$.

Para fazer inferência sobre o parâmetro μ usando a média amostral $\bar{Y}_n$, é necessário conhecer a distribuição ou uma aproximação da distribuição de $\bar{Y}_n$.

Se a série temporal for gaussiana, então temos

$$\sqrt{n}(\bar{Y}_n - \mu) \sim \mathcal{N}\left(0, \sum_{|h| < n} \left(1 - \frac{|h|}{n}\right) \gamma(h)\right).$$

É fácil construir intervalos de confiança para μ usando este resultado se $\gamma(h)$ for conhecida.

Intervalos de confiança aproximados podem ser obtidos se tivermos que estimar $\gamma(h)$ usando as observações.

Para várias séries temporais, em particular para modelos lineares e ARMA, $\bar{Y}_n$ é aproximadamente normal com média μ e variância $n^{-1} \sum_{|h| < \infty} \gamma(h)$ para n grande.

Um intervalo de confiança aproximado para μ de 95% é dado por

$$\left(\bar{y}_n - 1,96 \frac{\nu^{1/2}}{\sqrt{n}}; \bar{y}_n + 1,96 \frac{\nu^{1/2}}{\sqrt{n}} \right),$$

em que $\nu = \sum_{|h| < \infty} \gamma(h)$.

Em geral, ν precisa ser estimado. Podemos fazer isso da seguinte forma:

$$\hat{\nu} = \sum_{|h| < \sqrt{n}} \left(1 - \frac{|h|}{n} \right) \hat{\gamma}(h).$$

Para processos ARMA, a aproximação acima para ν é muito boa para n grande.

Exemplo

Seja $\{Y_t\}$ um processo AR(1) com média μ , definido por

$$Y_t = \mu + \phi(Y_{t-1} - \mu) + \varepsilon_t,$$

em que $|\phi| < 1$ e $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$.

Sabemos que $\gamma(h) = \frac{\phi^{|h|}\sigma^2}{1 - \phi^2}$, portanto

$$\nu = \left(1 + 2\sum_{h=1}^{\infty} \phi^h\right) \frac{\sigma^2}{1 - \phi^2} = \frac{\sigma^2}{(1 - \phi)^2}.$$

Então, outro intervalo de confiança aproximado para μ é dado por

$$\left(\bar{y}_n - 1,96 \frac{\sigma}{\sqrt{n}(1 - \phi)}; \bar{y}_n + 1,96 \frac{\sigma}{\sqrt{n}(1 - \phi)}\right).$$

Estimação de $\gamma(\cdot)$ e $\rho(\cdot)$

Como já vimos, a função de autocovariância amostral é dada por

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (Y_{t+|h|} - \bar{Y}_n)(Y_t - \bar{Y}_n)$$

e a função de autocorrelação amostral por

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

Os estimadores $\hat{\gamma}(h)$ e $\hat{\rho}(h)$ são tendenciosos para $\gamma(h)$ e $\rho(h)$ respectivamente. Isto ocorre mesmo se utilizarmos o divisor $(n - h)$.

Entretanto, o viés tende rapidamente a zero a medida que o tamanho da amostra cresce.

Vale lembrar que o divisor n garante que as matrizes $\hat{\Gamma}_k$ e $\hat{R}_k$ ($k \geq 1$) são não negativas definidas.

A função de autocorrelação tem um papel importante na escolha de modelos apropriados para os dados.

Para inferir sobre o valor de $\rho(h)$, precisamos conhecer a distribuição de $\hat{\rho}(h)$. É possível mostrar que

$$\hat{\rho}_k \approx \mathcal{N}_k(\rho_k, n^{-1} W),$$

em que $\rho_k = (\rho(1), \dots, \rho(k))^T$ e W é a matriz de covariâncias cujo elemento (i, j) é dado pela fórmula de Bartlett,

$$w_{ij} = \sum_{k=-\infty}^{\infty} \{ \rho(k+i)\rho(k+j) + \rho(k-i)\rho(k+j) + 2\rho(i)\rho(j)\rho^2(k) \\ - 2\rho(i)\rho(k)\rho(k+j) - 2\rho(j)\rho(k)\rho(k+i) \}.$$

Esta pode ser reescrita como

$$w_{ij} = \sum_{k=1}^{\infty} [\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)][\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)],$$

que é uma forma mais conveniente para os cálculos computacionais.

Exemplo

Se $\{Y_t\} \sim IID(0, \sigma^2)$, então $\rho(h) = 0$ para $|h| > 0$. Logo,

$$w_{ij} = \begin{cases} 1, & \text{se } i = j; \\ 0, & \text{caso contrário.} \end{cases}$$

Para n grande, $\hat{\rho}(1), \dots, \hat{\rho}(h)$ são aproximadamente normais independentes e identicamente distribuídas com média 0 e variância $1/n$.

Este resultado é a base dos testes que os dados são gerados por um processo i.i.d. utilizando a função de autocorrelação.

Exemplo

Se $\{Y_t\}$ é um processo MA(1), isto é,

$$Y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad t = 0, \pm 1, \pm 2, \dots$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$, então

$$w_{ii} = \begin{cases} 1 - 3\rho^2(1) + 4\rho^4(1), & \text{se } i = 1; \\ 1 + 2\rho^2(1), & \text{se } i > 1. \end{cases}$$

é a variância aproximada de $\sqrt{n}(\hat{\rho}(i) - \rho(i))$ para n grande.

A função de autocorrelação é

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0; \\ \theta/(1 + \theta^2), & \text{se } h = \pm 1; \\ 0, & \text{se } |h| > 1. \end{cases}$$

Se $\theta = -0,8$, então $\rho(1) = -0,8/(1 + (-0,8)^2) \simeq -0,4878049$.

Para uma determinada série artificial de tamanho $n = 200$, obtemos $\hat{\rho}(1) = -0,4258$. Logo, além do teste de hipótese para uma sequência i.i.d., podemos testar a hipótese de que os dados seguem um processo MA(1).

Por exemplo,

$$\begin{aligned} \text{IC}_{95\%}(\rho(1)) &= (\hat{\rho}(1) \pm 1,96n^{-1/2}(1 - 3\hat{\rho}^2(1) + 4\hat{\rho}^4(1))^{1/2}) \\ &= (-0,5321; -0,3196). \\ \text{IC}_{95\%}(\rho(2)) &= (\hat{\rho}(2) \pm 1,96n^{-1/2}(1 + 2\hat{\rho}^2(1)) \\ &= (-0,1915; 0,1321). \end{aligned}$$

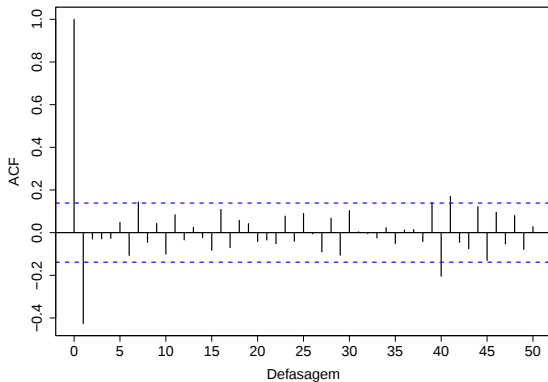


Figura 2: Função de autocorrelação amostral para 200 observações de um processo MA(1).

Ver arquivo exemplo_20.r

Exemplo

Para um processo AR(1),

$$Y_t = \phi Y_{t-1} + \varepsilon_t$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e $|\phi| < 1$, temos

$$\begin{aligned} w_{ii} &= \sum_{k=1}^i \phi^{2i} (\phi^{-k} - \phi^k)^2 + \sum_{k=i+1}^{\infty} \phi^{2k} (\phi^{-i} - \phi^i)^2 \\ &= (1 - \phi^{2i})(1 + \phi^2)(1 - \phi^2)^{-1} - 2i\phi^{2i}, \quad i = 1, 2, \dots \end{aligned}$$

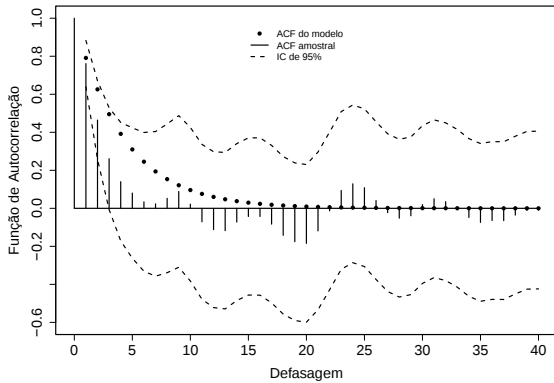
Ao estimar um modelo AR(1) para os dados do lago Huron (série sem tendência) obtemos $\hat{\phi} = 0,791$.

Então, **suponha que o modelo correto** seja dado por

$$Y_t = 0,791 Y_{t-1} + \varepsilon_t.$$

O intervalo de confiança é calculado como

$$\hat{\rho}(i) \pm 1,96n^{-1/2} w_{ii}^{1/2}, \quad \text{para } \phi = 0,791.$$



Ver arquivo exemplo_21.r

Previsão para séries temporais estacionárias

Considere o problema de previsão dos valores Y_{n+h} , $h > 0$, de uma série temporal estacionária com média μ e função de autocovariância $\gamma(\cdot)$ em termos dos valores $\{Y_n, \dots, Y_1\}$.

O objetivo é encontrar uma **combinação linear** de $1, Y_1, \dots, Y_n$ que preveja Y_{n+h} com erro quadrático médio mínimo.

O melhor preditor linear será denotado por

$$P_n Y_{n+h} = a_0 + a_1 Y_n + \dots + a_n Y_1.$$

Precisamos encontrar os valores de $a_0, a_1, \dots, a_n$ que minimize

$$\begin{aligned} S(a_0, \dots, a_n) &= E \left([Y_{n+h} - a_0 - a_1 Y_n - \dots - a_n Y_1]^2 \right) \\ &= E \left(\left[Y_{n+h} - a_0 - \sum_{i=1}^n a_i Y_{n+1-i} \right]^2 \right). \end{aligned}$$

Primeiro, calculamos

$$\frac{\partial S(a_0, \dots, a_n)}{\partial a_j} = 0, \quad \text{para } j = 0, 1, \dots, n.$$

O que resulta em

$$\begin{aligned} E \left(Y_{n+h} - a_0 - \sum_{i=1}^n a_i Y_{n+1-i} \right) &= 0, \quad \text{para } j = 0; \\ E \left(\left[Y_{n+h} - a_0 - \sum_{i=1}^n a_i Y_{n+1-i} \right] Y_{n+1-j} \right) &= 0, \quad \text{para } j = 1, \dots, n. \end{aligned}$$

Segue-se que

$$a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right) \quad \text{e} \quad \mathbf{\Gamma}_n \mathbf{a}_n = \gamma_n(h), \quad (8)$$

em que

$$\begin{aligned} \mathbf{a}_n &= (a_1, \dots, a_n)^\top, \\ \mathbf{\Gamma}_n &= [\gamma(i-j)]_{i,j=1}^n; \quad \text{e} \\ \gamma_n(h) &= (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))^\top. \end{aligned}$$

Assim,

$$P_n Y_{n+h} = \mu + \sum_{i=1}^n a_i (Y_{n+1-i} - \mu)$$

com $\mathbf{a}_n$ satisfazendo a Equação (8).

Finalmente, o erro quadrático médio de previsão é dada por

$$\begin{aligned} E \left([Y_{n+h} - P_n Y_{n+h}]^2 \right) &= \gamma(0) - 2 \sum_{i=1}^n a_i \gamma(h+i-1) \\ &\quad + \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(i-j) \\ &= \gamma(0) - \mathbf{a}_n^\top \boldsymbol{\gamma}_n(h). \end{aligned}$$

Propriedades de $P_n Y_{n+h}$:

1. $P_n Y_{n+h} = \mu + \sum_{i=1}^n a_i (Y_{n+1-i} - \mu).$
2. $E([Y_{n+h} - P_n Y_{n+h}]^2) = \gamma(0) - \mathbf{a}_n^\top \boldsymbol{\gamma}_n(h).$
3. $E(Y_{n+h} - P_n Y_{n+h}) = 0.$
4. $E([Y_{n+h} - P_n Y_{n+h}] Y_j) = 0, j = 1, \dots, n.$

Exemplo

Considere a previsão um passo à frente de um modelo AR(1) definido por

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots$$

em que $|\phi| < 1$ e $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$.

O melhor preditor linear de Y_{n+1} em termos de $\{1, Y_n, \dots, Y_1\}$, para $n \geq 1$ é

$$P_n Y_{n+1} = \mathbf{a}_n^\top \mathbf{Y}_n$$

em que $\mathbf{Y}_n = (Y_n, \dots, Y_1)^\top$ e

$$\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}.$$

A solução da equação acima é $\mathbf{a}_n = (\phi, 0, \dots, 0)^\top$.

Assim, o melhor preditor linear de Y_{n+1} em termos de $\{Y_1, \dots, Y_n\}$ é

$$P_n Y_{n+1} = \phi Y_n,$$

com erro médio quadrático

$$E([Y_{n+1} - P_n Y_{n+1}]^2) = \gamma(0) - \mathbf{a}_n^\top \gamma_n(1) = \frac{\sigma^2}{1 - \phi^2} - \phi \gamma(1) = \sigma^2.$$

Definição

$\{Y_t\}$ é um processo ARMA(p, q) se $\{Y_t\}$ é estacionário e se para cada t

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (9)$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$ e os polinômios $(1 - \phi_1 z - \cdots - \phi_p z^p)$ e $(1 + \theta_1 z + \cdots + \theta_q z^q)$ não tenham fatores comuns.

Dizemos que $\{Y_t\}$ é um processo ARMA(p, q) com média μ se $\{Y_t - \mu\}$ for um processo ARMA(p, q).

Outra forma de escrever o modelo é

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t$$

em que

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{e} \quad \Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Lembre-se que B é o operador defasagem, isto é, $B^j Y_t = Y_{t-j}$ para $j = 0, \pm 1, \dots$

O processo $AR(p)$ é um caso particular do $ARMA(p, q)$ quando $q = 0$ tal que $\Theta(z) = 1$.

O processo $MA(q)$ é um caso particular do $ARMA(p, q)$ quando $p = 0$ tal que $\Phi(z) = 1$.

Como a definição exige que o processo seja estacionário, restrições são impostas aos valores dos parâmetros.

Uma solução estacionária da Equação (9) existe (e é também uma solução estacionária única) se, e somente se,

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0, \quad \text{para todo } |z| = 1.$$

Se $z = a + bi$ é um número complexo, então $|z| = \sqrt{a^2 + b^2}$.

A região definida pelo conjunto de números complexos z tal que $|z| = 1$ é chamada de **círculo unitário**.

Um processo ARMA(p, q) $\{Y_t\}$ é **causador**, ou uma **função causadora** de $\{\varepsilon_t\}$, se existe constantes $\{\psi_j\}$ tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ e

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{para todo } t.$$

Causalidade é equivalente a condição

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0, \quad \text{para todo } |z| \leq 1.$$

Um processo ARMA(p, q) é **inversível** se existem constantes $\{\varphi_j\}$ tal que $\sum_{j=0}^{\infty} |\varphi_j| < \infty$ e

$$\varepsilon_t = \sum_{j=0}^{\infty} \varphi_j Y_{t-j} \quad \text{para todo } t.$$

Invertibilidade é equivalente a condição

$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0, \quad \text{para todo } |z| \leq 1.$$

Exemplo

Considere o processo ARMA(1,1) $\{Y_t\}$ satisfazendo as equações

$$Y_t - 0,5Y_{t-1} = \varepsilon_t + 0,4\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

A raiz de $\Phi(z) = 1 - 0,5z$ é 2 (**fora do círculo unitário**). Portanto o processo ARMA acima é único.

Como $\Phi(z) \neq 0$ para todo $|z| \leq 1$, o processo é também **causador**.

Então, podemos escrever $\{Y_t\}$ como um processo MA(∞). Quais os coeficientes?

$$\psi_0 = 1 \text{ e } \psi_j = (0,4 + 0,5)(0,5)^{j-1}, \quad j = 1, 2, \dots$$

O polinômio MA, $\Theta(z) = 1 + 0,4z$, tem raiz $-2,5$. Logo, $\{Y_t\}$ é inversível. Quais os coeficientes?

$$\varphi_0 = 1 \text{ e } \varphi_j = -(0,4 + 0,5)(-0,4)^{j-1}, \quad j = 1, 2, \dots$$

Exemplo

Seja $\{Y_t\}$ um processo AR(2),

$$Y_t = 0,7Y_{t-1} - 0,1Y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

Então, temos $\Phi(z) = 1 - 0,7z + 0,1z^2 = (1 - 0,5z)(1 - 0,2z)$. Como as raízes (2 e 5) estão fora do círculo unitário o processo é **causador**.

Exemplo

Considere o processo ARMA(2,1) $\{Y_t\}$ satisfazendo as equações

$$Y_t - 0,75Y_{t-1} + 0,5625Y_{t-2} = \varepsilon_t + 1,25\varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

O polinômio $\Phi(z) = 1 - 0,75z + 0,5625z^2$ tem zeros em $z = (2/3 \pm i2/\sqrt{3})$, que ficam fora do círculo unitário. Portanto, o processo é **causador**.

O polinômio MA, $\Theta(z) = 1 + 1,25z$ tem zero em -0,8. Logo, $\{Y_t\}$ **não é inversível**.

Cálculo da função de autocorrelação

Determinaremos a função de autocovariância de um processo causador ARMA(p, q) definido por

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2),$$

em que $\Phi(z) = (1 - \phi_1 z - \dots - \phi_p z^p)$ e $\Theta(z) = (1 + \theta_1 z + \dots + \theta_q z^q)$.

A série é causadora por hipótese, logo

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

em que

$$\sum_{j=0}^{\infty} \psi_j z^j = \frac{\Theta(z)}{\Phi(z)}, \quad |z| \leq 1.$$

Primeiro método

$$\gamma(h) = E(Y_{t+h} Y_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}.$$

Exemplo

O processo ARMA(1,1) é dado por

$$Y_t - \phi Y_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$$

com $|\phi| < 1$.

A função de autocorrelação é dada por

$$\gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 = \sigma^2 \left[1 + (\theta + \phi)^2 \sum_{j=0}^{\infty} \phi^{2j} \right] = \sigma^2 \left[1 + \frac{(\theta + \phi)^2}{1 - \phi^2} \right],$$

$$\gamma(1) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1} = \sigma^2 \left[\theta + \phi + (\theta + \phi)^2 \phi \sum_{j=0}^{\infty} \phi^{2j} \right]$$

$$= \sigma^2 \left[\theta + \phi + \frac{(\theta + \phi)^2 \phi}{1 - \phi^2} \right] \quad \text{e}$$

$$\gamma(h) = \phi^{h-1} \gamma(1), \quad h \geq 2.$$

Exemplo

Para o processo $MA(q)$

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2),$$

temos

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & \text{se } |h| \leq q, \\ 0, & \text{se } |h| > q. \end{cases}$$

em que $\theta_0 = 1$ por definição.

Logo, a função de autocovariância de um processo $MA(q)$ é zero para defasagens maiores do que q .

Lembre-se que um resultado anterior garante que todo processo estacionário de média zero com correlações zero para defasagens maiores do que q pode ser representado por um processo médias móveis.

Segundo método

Se multiplicarmos ambos os lados da equação

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

por Y_{t-k} , $k = 0, 1, 2, \dots$ e calcular os valores esperados de ambos os lados, encontraremos que

$$\gamma(k) - \phi_1 \gamma(k-1) - \cdots - \phi_p \gamma(k-p) = \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j} \psi_j, \quad 0 \leq k < m$$

e

$$\gamma(k) - \phi_1 \gamma(k-1) - \cdots - \phi_p \gamma(k-p) = 0, \quad k \geq m, \quad (10)$$

em que $m = \max(p, q+1)$, $\psi_j = 0$ para $j < 0$, $\theta_0 = 1$ e $\theta_j = 0$ para $j \notin \{0, \dots, q\}$.

Exemplo

O processo ARMA(1,1) é dado por

$$Y_t - \phi Y_{t-1} = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$$

com $|\phi| < 1$.

Como $m = \max(p, q + 1) = \max(1, 2) = 2$, temos

$$\gamma(0) - \phi\gamma(1) = \sigma^2(1 + \theta(\theta + \phi)), \quad k = 0$$

$$\gamma(1) - \phi\gamma(0) = \sigma^2\theta, \quad k = 1$$

$$\gamma(k) - \phi\gamma(k-1) = 0, \quad k \geq 2.$$

A solução do sistema acima é a mesma da primeira abordagem.

Terceiro método

As autocovariâncias podem ser encontradas resolvendo as $p + 1$ primeiras equações

$$\gamma(k) - \phi_1\gamma(k-1) - \dots - \phi_p\gamma(k-p) = \sigma^2 \sum_{j=0}^{\infty} \theta_{k+j}\psi_j, \quad 0 \leq k < m$$

e

$$\gamma(k) - \phi_1\gamma(k-1) - \dots - \phi_p\gamma(k-p) = 0, \quad k \geq m,$$

para $\gamma(0), \gamma(1), \dots, \gamma(p)$ e então usar as equações subsequentes para encontrar $\gamma(p+1), \gamma(p+2), \dots$

Este é um método conveniente para o cálculo numérico das autocovariâncias $\gamma(h)$.

Função de autocorrelação

A função de autocorrelação de um processo $\text{ARMA}(p, q)$ pode ser encontrada diretamente de sua função de autocovariância, isto é,

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

De maneira análoga, dado um conjunto de observações $\{y_1, y_2, \dots, y_n\}$, a função de autocorrelação amostral é dada por

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

A função de autocorrelação parcial

A **função de autocorrelação parcial** (PACF, abreviação do inglês) de um processo ARMA $\{Y_t\}$ é a função definida pelas equações

$$\begin{aligned}\alpha(0) &= 1, \quad \text{e} \\ \alpha(h) &= \phi_{hh}, \quad h \geq 1,\end{aligned}$$

em que ϕ_{hh} é a última componente de

$$\phi_h = \Gamma_h^{-1} \gamma_h,$$

$$\Gamma_h = [\gamma(i-j)]_{i,j=1}^h \quad \text{e} \quad \gamma_h = [\gamma(1), \gamma(2), \dots, \gamma(h)]^T.$$

Considere os seguintes modelos autoregressivos:

$$Y_t = \mu_{0,1} + \phi_{1,1} Y_{t-1} + \varepsilon_{1t}$$

$$Y_t = \mu_{0,2} + \phi_{1,2} Y_{t-1} + \phi_{2,2} Y_{t-2} + \varepsilon_{2t}$$

$$Y_t = \mu_{0,3} + \phi_{1,3} Y_{t-1} + \phi_{2,3} Y_{t-2} + \phi_{3,3} Y_{t-3} + \varepsilon_{3t}$$

$$\vdots \quad \vdots \quad \vdots$$

em que $\mu_{0,j}$, $\phi_{i,j}$ e ε_{jt} são, respectivamente, o termo constante, o coeficiente de Y_{t-i} e o termo de erro de um modelo AR(j).

Podemos estimar tudo por mínimos quadrados (múltiplo) e aplicar um teste F parcial.

A estimativa de $\hat{\phi}_{1,1}$ da primeira equação é chamada de PACF amostral de defasagem 1 de Y_t .

A estimativa de $\hat{\phi}_{2,2}$ da segunda equação é chamada de PACF amostral de defasagem 2 de Y_t .

A estimativa de $\hat{\phi}_{3,3}$ da terceira equação é chamada de PACF amostral de defasagem 3 de Y_t . (E assim por diante.)

A PACF de defasagem 2 $\hat{\phi}_{2,2}$ mostra a contribuição de Y_{t-2} adicionada a Y_t sobre o modelo AR(1) $Y_t = \mu + \phi_1 Y_{t-1} + \varepsilon_t$.

A PACF de defasagem 3 $\hat{\phi}_{3,3}$ mostra a contribuição de Y_{t-3} adicionada a Y_t sobre o modelo AR(2) $Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$. (E assim por diante.)

Portanto, para um modelo AR(p), a PACF amostral de defasagem p não deve ser zero.

Mas $\hat{\phi}_{j,j}$ deve estar perto de zero para todo $j > p$.

Utilizamos esta propriedade para determinar a ordem p .

Para um modelo AR(p) gaussiano estacionário, pode ser mostrado que a PACF amostral tem as seguintes propriedades:

- $\hat{\phi}_{p,p}$ converge para ϕ_p quando o tamanho da amostra n tende ao infinito.
- $\hat{\phi}_{k,k}$ converge para zero para todo $k > p$.
- A variância assintótica de $\hat{\phi}_{k,k}$ é $1/n$ para $k > p$.

Para qualquer conjunto de observações $\{y_1, \dots, y_n\}$ com $y_i \neq y_j$ para algum i e j , a PACF da amostral é dada por

$$\begin{aligned}\hat{\alpha}(0) &= 1, \quad \text{e} \\ \hat{\alpha}(h) &= \hat{\phi}_{hh}, \quad h \geq 1,\end{aligned}$$

em que $\hat{\phi}_{hh}$ é a última componente de

$$\hat{\phi}_h = \hat{\Gamma}_h^{-1} \hat{\gamma}_h.$$

Pode ser mostrado que ϕ_{hh} é a correlação entre os erros de previsão

$$Y_h - P(Y_h | Y_1, \dots, Y_{h-1}) \quad \text{e} \quad Y_0 - P(Y_0 | Y_1, \dots, Y_{h-1}).$$

Exemplo

Para um processo $AR(p)$ causador definido por

$$Y_t - \phi_1 Y_{t-1} - \cdots - \phi_p Y_{t-p} = \varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2),$$

sabemos que para $h \geq p$ o melhor preditor linear de Y_{h+1} em termos de $1, Y_1, \dots, Y_h$ é

$$\hat{Y}_{h+1} = \phi_0 + \phi_1 Y_h + \phi_2 Y_{h-1} + \cdots + \phi_p Y_{h+1-p}.$$

Como o coeficiente ϕ_{hh} de Y_1 é ϕ_p se $h = p$ e 0 se $h > p$, concluímos que a PACF $\alpha(\cdot)$ do processo $\{Y_t\}$ tem a propriedade

$$\begin{aligned} \alpha(p) &= \phi_p, \quad \text{e} \\ \alpha(h) &= 0, \quad \text{para } h > p. \end{aligned}$$

Exemplo

Para o processo $MA(1)$, pode ser mostrado que a PACF na defasagem h é

$$\alpha(h) = \phi_{hh} = -\frac{(-\theta)^h}{(1 + \theta^2 + \cdots + \theta^{2h})}.$$

Ver arquivo [exemplo_22.r](#)

Previsão: o algoritmo das inovações

Este algoritmo é aplicável a todas as séries temporais com segundo momento finito sendo estas estacionárias ou não.

Suponha que $\{Y_t\}$ seja uma série temporal com média zero e $E|Y_t|^2 < \infty$ para cada t e

$$E(Y_i Y_j) = \kappa(i, j).$$

Introduzimos aqui uma notação para os melhores previsores um passo à frente e seus erros quadráticos médios:

$$\hat{Y}_n = \begin{cases} 0, & \text{se } n = 1; \\ P_{n-1} Y_n, & \text{se } n = 2, 3, \dots, \end{cases} \quad \text{e} \quad \nu_n = E \left([Y_{n+1} - P_n Y_{n+1}]^2 \right).$$

Introduzimos também as inovações ou erros de previsão um passo à frente,

$$U_n = Y_n - \hat{Y}_n.$$

Em termos dos vetores $\mathbf{U}_n = (U_1, \dots, U_n)^\top$ e $\mathbf{Y}_n = (Y_1, \dots, Y_n)^\top$ as últimas equações podem ser escritas como

$$\mathbf{U}_n = \mathbf{A}_n \mathbf{Y}_n$$

$$\mathbf{A}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ a_{11} & 1 & 0 & \cdots & 0 \\ a_{22} & a_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \cdots & 1 \end{bmatrix}.$$

Isto implica que $\mathbf{A}_n$ é não singular com inversa $\mathbf{C}_n$ da forma

$$\mathbf{C}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 1 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 1 \end{bmatrix}.$$

O vetor de previsores um passo à frente $\hat{\mathbf{Y}}_n = (Y_1, P_1 Y_2, \dots, P_{n-1} Y_n)^\top$ pode ser expresso por

$$\hat{\mathbf{Y}}_n = \mathbf{Y}_n - \mathbf{U}_n = \mathbf{C}_n \mathbf{U}_n - \mathbf{U}_n = \mathbf{\Theta}_n (\mathbf{Y}_n - \hat{\mathbf{Y}}_n) \quad (11)$$

em que

$$\mathbf{\Theta}_n = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ \theta_{11} & 0 & 0 & \cdots & 0 \\ \theta_{22} & \theta_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \cdots & 0 \end{bmatrix}$$

(com zeros na diagonal tal que $\mathbf{C}_n \neq \mathbf{\Theta}_n$) e

$$\mathbf{Y}_n = \mathbf{C}_n (\mathbf{Y}_n - \hat{\mathbf{Y}}_n).$$

A Equação (11) pode ser reescrita como

$$\hat{\mathbf{Y}}_{n+1} = \begin{cases} 0, & \text{se } n = 0; \\ \sum_{j=1}^n \theta_{nj} (Y_{n+1-j} - \hat{Y}_{n+1-j}), & \text{se } n = 1, 2, \dots, \end{cases}$$

da qual os previsores um passo à frente $\hat{Y}_1, \hat{Y}_2, \dots$ podem ser calculados recursivamente uma vez definidos os valores dos coeficientes θ_{ij} .

O algoritmo a seguir gera estes coeficientes e os erros quadráticos médios $\nu_i = E([Y_{i+1} - \hat{Y}_{i+1}]^2)$, começando com as covariâncias $\kappa(i, j)$.

O algoritmo das inovações

Os coeficientes $\theta_{n1}, \dots, \theta_{nn}$ podem ser calculados recursivamente das equações

$$\begin{aligned} \nu_0 &= \kappa(1, 1), \\ \theta_{n, n-k} &= \nu_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} \nu_j \right), \quad 0 \leq k < n, \\ \nu_n &= \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n, n-j}^2 \nu_j. \end{aligned}$$

Previsão de processos ARMA

O algoritmo das inovações nos fornece um método para previsão em processos ARMA(p, q)

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2),$$

sendo possível uma simplificação do mesmo.

Para isso, aplicaremos o algoritmo a série transformada W_t dada por

$$\begin{cases} W_t = \sigma^{-1} Y_t, & \text{se } t = 1, \dots, m; \\ W_t = \sigma^{-1} \Phi(B) Y_t, & \text{se } t > m \end{cases}$$

em que $m = \max(p, q)$.

Para facilitar a notação, façamos $\theta_0 = 1$ e $\theta_j = 0$ para $j > q$.

A função de autocovariância pode ser calculada com um dos métodos descritos anteriormente.

Assim, as autocovariâncias $\kappa(i, j) = E(W_i W_j)$, $i, j > 1$, podem ser calculadas da seguinte forma:

$$\kappa(i, j) = \begin{cases} \sigma^{-2} \gamma_Y(i - j), & 1 \leq i, j \leq m, \\ \sigma^{-2} \left[\gamma_Y(i - j) - \sum_{r=1}^p \phi_r \gamma_Y(r - |i - j|) \right], & \min(i, j) \leq m < \max(i, j) \leq 2m, \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|}, & \min(i, j) > m, \\ 0, & \text{caso contrário.} \end{cases}$$

Aplicando os algoritmo das inovações ao processo W_t , obtemos

$$\begin{cases} \widehat{W}_{n+1} = \sum_{j=1}^n \theta_{nj} (W_{n+1-j} - \widehat{W}_{n+1-j}), & 1 \leq n < m, \\ \widehat{W}_{n+1} = \sum_{j=1}^q \theta_{nj} (W_{n+1-j} - \widehat{W}_{n+1-j}), & n \geq m, \end{cases}$$

em que os coeficientes θ_{nj} e os erros quadráticos médios $r_n = E([W_{n+1} - \widehat{W}_{n+1}]^2)$ são encontrados recursivamente utilizando o algoritmo das inovações.

Os previsores um passo à frente de W_{n+1} e Y_{n+1} são dados por

$$\widehat{W}_{n+1} = P_n W_{n+1} \quad \text{e} \quad \widehat{Y}_{n+1} = P_n Y_{n+1}.$$

Usando a linearidade de P_t temos que

$$\begin{cases} \widehat{W}_t = \sigma^{-1} \widehat{Y}_t, & \text{se } t = 1, \dots, m; \\ \widehat{W}_t = \sigma^{-1} [\widehat{Y}_t - \phi_1 Y_{t-1} - \dots - \phi_p Y_{t-p}], & \text{se } t > m \end{cases}$$

e, consequentemente,

$$Y_t - \widehat{Y}_t = \sigma(W_t - \widehat{W}_t), \quad \text{para todo } t \geq 1.$$

Substituindo $(W_t - \widehat{W}_t)$ por $\sigma^{-1}(Y_t - \widehat{Y}_t)$ nas equações de $\kappa(i, j)$ e de $\widehat{W}_{n+1}$, finalmente obtemos

$$\widehat{Y}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj} (Y_{n+1-j} - \widehat{Y}_{n+1-j}), & 1 \leq n < m \\ \phi_1 Y_n + \dots + \phi_p Y_{n+1-p} + \sum_{j=1}^q \theta_{nj} (Y_{n+1-j} - \widehat{Y}_{n+1-j}), & n \geq m, \end{cases}$$

e

$$E[(Y_{n+1} - \widehat{Y}_{n+1})^2] = \sigma^2 E[(W_{n+1} - \widehat{W}_{n+1})^2] = \sigma^2 r_n$$

em que θ_{nj} e r_t são obtidos através do algoritmo das inovações.

ARMA(p, q): Análise e Previsão

Estimação dos coeficientes autoregressivos $\phi_1, \phi_2, \dots, \phi_p$.

Estimação dos coeficientes médias móveis $\theta_1, \theta_2, \dots, \theta_q$.

Estimação da variância (σ^2) do ruído branco.

Escolher a ordem p e q do modelo ARMA utilizando critérios de seleção baseados na função de verossimilhança mais uma penalização: AIC, BIC e AICC.

Em geral, podemos trabalhar com processos ARMA com média zero:

$$\Phi(B)Y_t = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

Suponha que $\{Y_t\}$ seja um série temporal gaussiana com média zero e função de autocovariância

$$\kappa(i, j) = E(Y_i Y_j).$$

Seja $\mathbf{Y}_n = (Y_1, \dots, Y_n)^\top$ e seja $\hat{\mathbf{Y}}_n = (\hat{Y}_1, \dots, \hat{Y}_n)^\top$ em que $\hat{Y}_1 = 0$ e

$$\hat{Y}_j = E(Y_j | Y_1, \dots, Y_{j-1}) = P_{j-1} Y_j, \quad j \geq 2.$$

Denotaremos por $\mathbf{\Gamma}_n$ a matriz de covariância $\mathbf{\Gamma}_n = E(\mathbf{Y}_n \mathbf{Y}_n^\top)$.

Suponha que $\mathbf{\Gamma}_n$ seja não singular. Então, a função de verossimilhança é dada por

$$L(\mathbf{\Gamma}_n) = (2\pi)^{-n/2} |\mathbf{\Gamma}_n|^{-1/2} \exp \left(-\frac{1}{2} \mathbf{Y}_n^\top \mathbf{\Gamma}_n^{-1} \mathbf{Y}_n \right). \quad (12)$$

O cálculo direto de $|\mathbf{\Gamma}_n|$ e $\mathbf{\Gamma}_n^{-1}$ podem ser evitados.

Para isto se utiliza os erros de previsão um passo à frente $Y_j - \hat{Y}_j$ e suas variâncias ν_{j-1} , $j = 1, \dots, n$. Ambos podem ser calculados através do algoritmo das inovações.

Do algoritmo das inovações, temos

$$\mathbf{Y}_n = \mathbf{C}_n(\mathbf{Y}_n - \hat{\mathbf{Y}}_n).$$

Pode ser mostrado que as componentes de $(\mathbf{Y}_n - \hat{\mathbf{Y}}_n)$ são não correlacionadas.

Logo, $(\mathbf{Y}_n - \hat{\mathbf{Y}}_n)$ tem matriz de covariância diagonal dada por

$$\mathbf{D}_n = \text{diag}\{\nu_0, \dots, \nu_{n-1}\}.$$

Logo, $\mathbf{\Gamma}_n = \mathbf{C}_n \mathbf{D}_n \mathbf{C}_n^\top$,

$$\mathbf{Y}_n^\top \mathbf{\Gamma}_n^{-1} \mathbf{Y}_n = (\mathbf{Y}_n - \hat{\mathbf{Y}}_n)^\top \mathbf{C}_n^\top \mathbf{\Gamma}_n^{-1} \mathbf{C}_n (\mathbf{Y}_n - \hat{\mathbf{Y}}_n) = \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 / \nu_{j-1}$$

e

$$|\mathbf{\Gamma}_n| = |\mathbf{C}_n|^2 |\mathbf{D}_n| = \nu_0 \nu_1 \cdots \nu_{n-1}.$$

A função de verossimilhança dada na Equação (12) do vetor $\mathbf{Y}_n$ se reduz a

$$L(\mathbf{\Gamma}_n) = (2\pi)^{-n/2} \left(\prod_{j=1}^n \nu_{j-1} \right)^{-1/2} \exp \left(-\frac{1}{2} \sum_{j=1}^n \frac{(Y_j - \hat{Y}_j)^2}{\nu_{j-1}} \right). \quad (13)$$

Teremos $\{Y_t\}$ como um processo ARMA(p, q) e, portanto, Γ_n pode ser expresso em termos de um número finito de parâmetros desconhecidos.

Logo, o estimador de máxima verossimilhança dos parâmetros são aqueles valores que maximizam a função $L(\Gamma_n)$ para uma série observada.

Recorremos a métodos numéricos para encontrar as estimativas de máxima verossimilhança.

No programa R utilizaremos a função `arima`.

Seleção de modelos

$$\begin{aligned}\text{AIC} &= -2 \log L(\Gamma_n) + 2k \\ &= -2 \log L(\Phi, \Theta, \sigma^2) + 2(p + q + 1)\end{aligned}$$

$$\begin{aligned}\text{BIC} &= -2 \log L(\Gamma_n) + k \log(n) \\ &= -2 \log L(\Phi, \Theta, \sigma^2) + (p + q + 1) \log(n)\end{aligned}$$

$$\text{AICC} = -2 \log L(\Phi, \Theta, \sigma^2) + 2 \frac{(p + q + 1)n}{(n - p - q - 2)}$$

Devemos escolher, em geral, p , q , Φ_p and Θ_q que minimizam o valor do AIC. O mesmo vale se usarmos o BIC e o AICC.

Intervalo de confiança para os coeficientes

Para um tamanho de amostra grande o estimador de máxima verossimilhança $\hat{\beta}$ de

$$\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$$

tem distribuição aproximadamente normal com média β e matriz de covariância $n^{-1} \mathbf{V}(\beta)$.

A matriz de covariância $n^{-1} \mathbf{V}(\beta)$ pode ser aproximada por $2\mathbf{H}^{-1}(\beta)$, em que $\mathbf{H}$ é a matriz hessiana:

$$\left[\frac{\partial^2 \log L(\beta, \sigma^2)}{\partial \beta_i \partial \beta_j} \right]_{i,j=1}^{p+q}.$$

As variâncias e covariâncias podem ser usadas para construir intervalos de confiança.

Então, para uma amostra grande de um processo ARMA(p, q) temos

$$\hat{\beta} \approx \mathcal{N}(\beta, n^{-1} \mathbf{V}(\beta)).$$

Ver arquivo exemplo_23.r

ARIMA(p, d, q) para séries não estacionárias

Os processos ARIMA se reduzem aos processos ARMA quando diferenças finitas são aplicadas várias vezes.

Definição

Se d é um inteiro não negativo, então $\{Y_t\}$ é um processo ARIMA(p, d, q) se $W_t = (1 - B)^d Y_t$ for um processo ARMA(p, q) causador.

Uma outra maneira de representar o processo ARIMA(p, d, q) é

$$\Phi(B)(1 - B)^d Y_t = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2).$$

SARIMA(p, d, q)(P, D, Q): o ARIMA com sazonalidade estocástica

Definição

Se d e D são inteiros não negativos, então $\{Y_t\}$ é um processo ARIMA(p, d, q)(P, D, Q) sazonal com período s se a série diferenciada $W_t = (1 - B)^d(1 - B^s)^D Y_t$ é um processo ARMA causador definido por

$$\Phi(B)\tilde{\Phi}(B^s)W_t = \Theta(B)\tilde{\Theta}(B^s)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2),$$

em que $\Phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$, $\tilde{\Phi}(z) = 1 - \tilde{\phi}_1 z - \dots - \tilde{\phi}_p z^p$, $\Theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q$ e $\tilde{\Theta}(z) = 1 + \tilde{\theta}_1 z + \dots + \tilde{\theta}_q z^q$.

Por exemplo, se $s = 12$ e $P = 2$, então

$$\tilde{\Phi}(B^{12}) = 1 - \tilde{\phi}_1 B^{12} - \tilde{\phi}_2 B^{24} \quad \text{e} \quad \tilde{\Phi}(B^{12})W_t = W_t - \tilde{\phi}_1 W_{t-12} - \tilde{\phi}_2 W_{t-24}.$$

Os modelos ARIMA sazonal permitem aleatoriedade no padrão sazonal da série temporal de um ciclo para o outro. Isto é diferente do modelo aditivo $Y_t = m_t + s_t + \varepsilon_t$.

Regressão com erros ARMA (ARMAX)

O modelo de regressão com erros ARMA é definido por

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_k X_{kt} + W_t, \quad W_t \sim \text{ARMA}(p, q),$$

com média zero e para $t = 1, 2, \dots, n$.

Podemos reescrever o modelo como

$$\Phi(B)Y_t - \mathbf{X}_t^\top \boldsymbol{\beta} = \Theta(B)\varepsilon_t, \quad \varepsilon_t \sim \mathcal{RB}(0, \sigma^2),$$

em que $\mathbf{X}_t = (1, X_{1t}, X_{2t}, \dots, X_{kt})^\top$ e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$.

Ver arquivo exemplo_24.r

Modelos em Espaço de Estados

Outra abordagem para modelagem de séries temporais.

As séries temporais não precisam ser estacionárias!

Modelos

As formas básicas que englobam parte dos modelos são as seguintes:

- modelos de **tendência**;
- modelos de **sazonalidade** (ou ciclos sistemáticos); e
- modelos de **variáveis causais** (ou regressores).

Combinações dessas formas fornecem modelos para as seguintes séries:

- **financeiras** (vendas, estoque);
- **industriais** (produção, capacidade operativa);
- **agrícolas** (produção de leite, mercado de carnes);
- **médicas** (monitoração de órgãos); e
- **sociais** (acidentes, nascimentos).

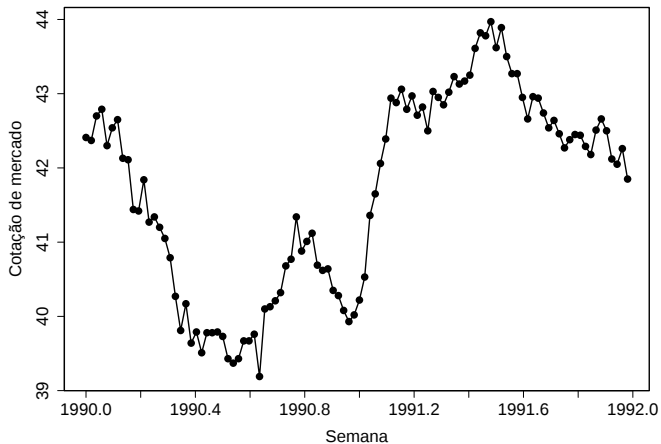


Figura 3: Série de cotação do mercado (1990:1-1991:52).

Ver arquivo exemplo_25.r

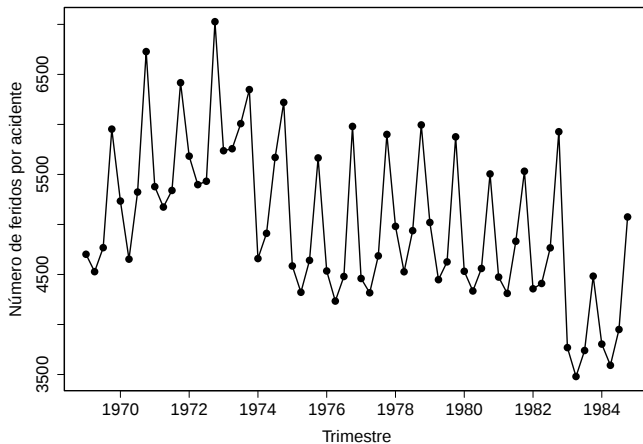


Figura 4: Número de feridos por accidente (1969:1-1984:4).

Ver archivo [ejemplo_25.r](#)

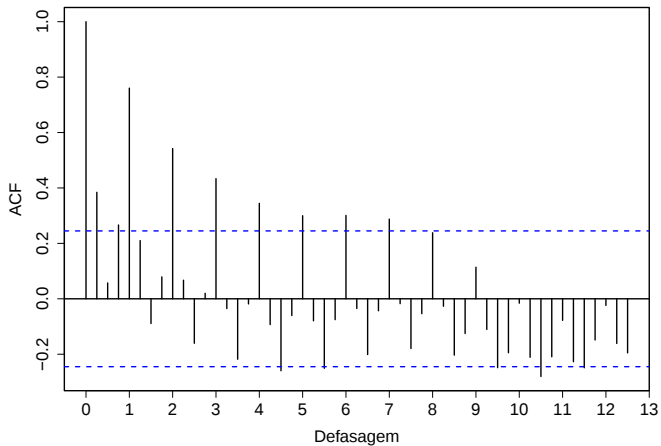


Figura 5: ACF do número de feridos por acidente (1969:1-1984:4).

Ver arquivo exemplo_25.r

Modelos estáticos

Os **modelos estáticos** têm uma descrição fixa (através de parâmetros fixos) ao longo das unidades de observação.

- Análise de regressão,

$$Y_t = \beta_0 + \beta_1 z_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

- Modelos ARMA,

$$Y_t = \alpha + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2).$$

Em séries temporais, essa hipótese muitas vezes é violada:

as estruturas mudam com a passagem do tempo.

Nesta parte do curso, faz-se uma introdução aos **modelos em espaço de estados**.

- **A descrição (os parâmetros) muda(m) com a passagem do tempo.**
- Os modelos em espaço de estados nos permitem a modelagem da tendência (média) e variância (volatilidade).
- Os modelos em espaço de estados incluem como casos particulares os modelos estáticos.

A passagem do tempo traz observações e aumenta o nosso conhecimento.

Em modelos em espaço de estados temos também perda de informação devido à passagem do tempo.

Por exemplo, o nível de vendas mês passado é mais relevante hoje que o nível de vendas em janeiro passado.

Abordagem bayesiana

Um **modelo** é uma representação de uma realidade.

- Será tão adequado quanto à sua capacidade de alcançar os objetivos a que ele se destina.
- A construção do modelo traz inerente em si um caráter subjetivo.

Previsão é uma afirmação sobre um futuro incerto.

A **incerteza** será sempre representada através de probabilidade.

- A previsão será sempre formulada em termos de probabilidade condicionada ao nosso estado de conhecimento.
- Se o estado de conhecimento muda, nossa previsão mudará.

Nosso conhecimento provém de duas fontes:

- a série histórica ou dados; e
- outros conhecimentos (subjetivos).

Exemplos: entrada em vigor de leis, falência de competidor, etc.

Ambas as fontes são importantes, podem e devem ser utilizadas.

Abordagem bayesiana incorpora esses elementos naturalmente e coerentemente.

Aplicando à previsão, significa que o modelo padrão é posto para funcionar.

Se acontecimentos não rotineiros intervêm, o modelo os incorpora:

- preparando para mudança; ou
- alterando o que for necessário.

Exemplo: se o competidor falirá, precisa-se utilizar o conhecimento sobre a divisão do mercado para formular a mudança que ele espera que aconteça.

O modelo de regressão linear múltiplo

Em um modelo de regressão linear múltiplo temos uma variável resposta y_t que é explicada por um conjunto de variáveis explicativas $x_{1t}, \dots, x_{pt}$ através da relação

$$y_t = \theta_0 + \theta_1 x_{1t} + \dots + \theta_p x_{pt} + \varepsilon_t,$$

em que $\varepsilon_t \sim \mathcal{RB}(0, \sigma^2)$.

A equação acima pode ser reescrita como:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\theta} + \varepsilon_t,$$

em que

$$\mathbf{x}_t = \begin{bmatrix} 1 \\ x_{1t} \\ x_{2t} \\ \vdots \\ x_{pt} \end{bmatrix} \quad \text{e} \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix}.$$

A natureza das variáveis explicativas é bastante ampla. Podendo assim, utilizar-se qualquer variável quantificável.

Os coeficientes da regressão linear $\theta_1, \dots, \theta_p$ informam sobre a influência que os regressores têm sobre a resposta y_t .

Na prática, seus valores são desconhecidos e estimados a partir de uma coleção de observações feitas sobre o modelo de regressão linear múltiplo.

Assim, observamos respostas $y_1, \dots, y_n$ com seus respectivos regressores $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. Novamente, temos que

$$y_t = \mathbf{x}_t^\top \boldsymbol{\theta} + \varepsilon_t, \quad t = 1, \dots, n.$$

Modelo em espaço de estados

Em modelos em espaço de estados os parâmetros mudam com o passar do tempo.

O modelo de regressão é estendido para

$$Y_t = \mathbf{x}_t^\top \boldsymbol{\theta}_t + \varepsilon_t, \quad t = 1, \dots, T.$$

Note a **indexação de $\boldsymbol{\theta}$** .

Esta formulação cria uma profusão de parâmetros a serem estimados.

O modelo necessita de mais informação. Essa informação vem do fato que os parâmetros sucessivos estão intimamente relacionados.

Por exemplo, um parâmetro é igual ao seu antecessor mais uma pequena perturbação causada pelas mudanças às quais o sistema está sujeito.

Se o sistema é estático, como no modelo de regressão, temos que

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}.$$

Em modelos em espaço de estados, admitimos a forma mais geral dada por

$$\theta_t = \mathbf{G}_t \theta_{t-1} + \mathbf{w}_t,$$

em que

- $\mathbf{G}_t$ contém valores conhecidos (ou parâmetros fixos); e
- $\mathbf{w}_t$ é uma perturbação aleatória.

A equação acima é conhecida como **equação do sistema**.

A **matriz de evolução** $\mathbf{G}_t$ controla a parte determinística da evolução do sistema e estabelece a propagação do sistema ao longo do tempo.

A **perturbação** $\mathbf{w}_t$ é responsável pela introdução de incertezas devidas à passagem do tempo e consequente perda de informação.

Note que se $\mathbf{G}_t = \mathbf{I}$ e $\mathbf{w}_t = \mathbf{0}$, o modelo se reduz ao caso estático.

Exemplo

Suponha que uma série de **vendas** possa ser explicada pela respectiva série de **preços** através da seguinte relação:

$$vendas_t = \mu_t + \beta_t \times preço_t + v_t$$

$$\mu_t = \mu_{t-1} + w_{1t}$$

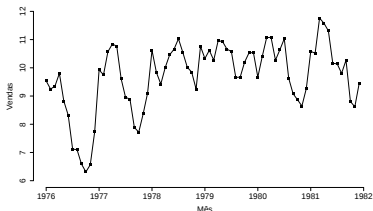
$$\beta_t = \beta_{t-1} + w_{2t}.$$

As **vendas** são explicadas pelo **preço** em uma regressão dinâmica.

A maior (ou menor) estabilidade dessa relação será controlada pela magnitude dos incrementos (erros) w_{1t} e w_{2t} .

Exemplo

Analisaremos a série de vendas mensais de um certo tipo doce no Reino Unido de janeiro de 1976 a dezembro de 1981. Também utilizaremos a série de preços, na escala do logaritmo, deste doce como variável explicativa.



(a) Vendas



(b) Preços

Figura 6: Vendas e logaritmo dos preços mensais de um certo tipo doce no Reino Unido de janeiro de 1976 a dezembro de 1981.

Ver arquivo [exemplo_26.r](#)

Modelo em espaço de estados linear e gaussiano

O **modelo linear** pode ser definido como

- **Equação das observações:**

$$Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + v_t, \quad v_t \sim \mathcal{N}(0, V_t)$$

- **Equação do sistema:**

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$$

- **Priori:**

$$\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0).$$

Em geral, supomos que $\{\mathbf{F}_t, V_t, \mathbf{G}_t, \mathbf{W}_t\}$ são conhecidos para todo instante de tempo t e podem depender de Ψ . Obviamente, precisamos que $\mathbf{m}_0$ e $\mathbf{C}_0$ sejam conhecidos também.

No exemplo anterior, $Y_t = \text{venda}_t$, $\mathbf{F}_t = (1, \text{preço}_t)^\top$.

$$\boldsymbol{\theta}_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, \quad \mathbf{G}_t = \mathbf{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{e} \quad \mathbf{w}_t = \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}$$

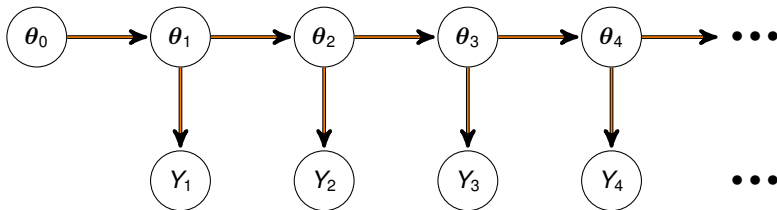


Figura 7: Um DAG de um modelo em espaço de estados.

Defina $\theta_{1:t} = (\theta_1, \theta_2, \dots, \theta_t)^\top$ como uma coleção de variáveis aleatórias.

Considera-se que

- $Y_{1:t}$ são variáveis observáveis; e
- $\theta_{1:t}$ são variáveis não observáveis (latentes).

Nestes modelos, temos as seguintes distribuições:

- $f(\theta_0|\Psi)$, priori para o instante inicial;
- $f(\theta_t|\theta_{t-1}, \Psi)$, $t = 1, 2, \dots$, para a evolução dos estados; e
- $f(y_t|\theta_t, \Psi)$, $t = 1, 2, \dots$, para as observações.

Análise sequencial

A natureza sequencial de séries temporais é devida à obtenção sequencial de informação.

A **distribuição do estado** nos informa como a partir da **posteriori de ontem** podemos chegar à **priori de hoje**.

O **método bayesiano** nos ensina como combinar a **priori de hoje** com a informação que acabamos de obter para chegar à **posteriori de hoje**.

Para amanhã e dias futuros, o ciclo se repete.

Em termos probabilísticos, temos que

$$\cdots \rightarrow f(\theta_{t-1} \mid y_{1:t-1}, \Psi) \rightarrow f(\theta_t \mid y_{1:t-1}, \Psi) \rightarrow f(\theta_t \mid y_{1:t}, \Psi) \rightarrow \cdots$$

e também $f(y_t \mid y_{1:t-1}, \Psi)$ como a distribuição preditiva.

Suponha que a posteriori, $f(\theta_{t-1} \mid y_{1:t-1}, \Psi)$, no instante $t - 1$ seja conhecida. Então, a priori para o instante t é dada por

$$f(\theta_t \mid y_{1:t-1}, \Psi) = \int_{\Theta_{t-1}} f(\theta_t \mid \theta_{t-1}, \Psi) f(\theta_{t-1} \mid y_{1:t-1}, \Psi) d\theta_{t-1}.$$

Note que utilizamos a **distribuição do estado**.

Agora, a posteriori para o instante t é dada via **teorema de Bayes** por

$$f(\theta_t \mid y_{1:t}, \Psi) = \frac{f(y_t \mid \theta_t, \Psi) f(\theta_t \mid y_{1:t-1}, \Psi)}{f(y_t \mid y_{1:t-1}, \Psi)},$$

em que

$$f(y_t \mid y_{1:t-1}, \Psi) = \int_{\Theta_t} f(y_t \mid \theta_t, \Psi) f(\theta_t \mid y_{1:t-1}, \Psi) d\theta_t.$$

É possível descrever a previsão k passos à frente via $f(y_{t+k} \mid y_{1:t-1}, \Psi)$.

Se o modelo em espaço de estado for linear e gaussiano, então é possível obter uma solução de forma analítica (condicional a Ψ) da sequência acima.

Previsão

Previsões no modelo dinâmico são obtidas pela **combinação da informação a priori com a equação das observações**.

A combinação de $y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + v_t$ com a priori $f(\boldsymbol{\theta}_t | y_{1:t-1})$ permite a obtenção da **distribuição preditiva** $f(y_t | y_{1:t-1})$ baseado na qual as previsões serão feitas.

Em particular, se quisermos

- uma **previsão pontual**, podemos tomar a média dessa distribuição, $E(y_t | y_{1:t-1})$.
- um **intervalo de predição** de 90% de probabilidade, basta tomar A de forma que

$$0,90 = \Pr(y_t \in A | y_{1:t-1}) = \int_A f(y_t | y_{1:t-1}) dy_t$$

A **previsão vários passos à frente** é feita de forma similar.

Se temos interesse em prever y_{t+k} no tempo $t - 1$, precisamos **utilizar a equação do sistema sucessivamente** até podermos escrever θ_{t+k} como função de θ_{t-1} .

Por exemplo, combinando

$$\theta_{t+1} = \mathbf{G}_{t+1}\theta_t + \mathbf{w}_{t+1} \quad \text{e} \quad \theta_t = \mathbf{G}_t\theta_{t-1} + \mathbf{w}_t,$$

pode-se obter

$$\theta_{t+1} = \mathbf{G}_{t+1}[\mathbf{G}_t\theta_{t-1} + \mathbf{w}_t] + \mathbf{w}_{t+1} = \mathbf{G}_{t+1}\mathbf{G}_t\theta_{t-1} + \mathbf{G}_{t+1}\mathbf{w}_t + \mathbf{w}_{t+1}.$$

A partir daí, combina-se com a equação das observações no tempo $t + k$ para obter a distribuição preditiva $f(y_{t+k} \mid y_{1:t-1})$.

Previsões cumulativas para os próximos k períodos, isto é, $y_t + y_{t+1} + \dots + y_{t+k-1}$ também podem ser obtidas pelo mesmo método.

Filtro de Kalman: $\{\mathbf{F}_t, \mathbf{V}_t, \mathbf{G}_t, \mathbf{W}_t\}$ conhecidos.

Priori:	$(\boldsymbol{\theta}_0 \mid D_0) \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$	
Observação:	$y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + v_t$	$v_t \sim \mathcal{N}(0, V_t)$
Sistema:	$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t$	$\mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$
Informação:	$(\boldsymbol{\theta}_{t-1} \mid y_{1:t-1}) \sim \mathcal{N}(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$ $(\boldsymbol{\theta}_t \mid y_{1:t-1}) \sim \mathcal{N}(\mathbf{a}_t, \mathbf{R}_t)$	$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^\top + \mathbf{W}_t$
Previsão:	$(y_t \mid y_{1:t-1}) \sim \mathcal{N}(f_t, Q_t)$	$f_t = \mathbf{F}_t^\top \mathbf{a}_t$ $Q_t = \mathbf{F}_t^\top \mathbf{R}_t \mathbf{F}_t + V_t$
Posteriori:	$(\boldsymbol{\theta}_t \mid y_{1:t}) \sim \mathcal{N}(\mathbf{m}_t, \mathbf{C}_t)$	$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t$ $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t^\top Q_t$ $\mathbf{e}_t = y_t - f_t$ $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / Q_t$

Distribuições preditivas para $k \geq 1$

$$(\boldsymbol{\theta}_{t+k} \mid D_t) \sim \mathcal{N}(\mathbf{a}_t(k), \mathbf{R}_t(k))$$

$$(y_{t+k} \mid D_t) \sim \mathcal{N}(f_t(k), Q_t(k))$$

$$\mathbf{a}_t(k) = \mathbf{G}_{t+k} \mathbf{a}_t(k-1)$$

$$\mathbf{R}_t(k) = \mathbf{G}_{t+k} \mathbf{R}_t(k-1) \mathbf{G}_{t+k}^\top + \mathbf{W}_{t+k}$$

$$f_t(k) = \mathbf{F}_{t+k}^\top \mathbf{a}_t(k)$$

$$Q_t(k) = \mathbf{F}_{t+k}^\top \mathbf{R}_t(k) \mathbf{F}_{t+k} + V_{t+k}$$

$$\mathbf{a}_t(0) = \mathbf{m}_t$$

$$\mathbf{R}_t(0) = \mathbf{C}_t$$

Modelo de tendência estável

É composto apenas de um nível que varia segundo um passeio aleatório.

O modelo de tendência estável é dado por

$$\begin{aligned}Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W_t).\end{aligned}$$

O nível permanece localmente constante, mas varia quando se considera longos períodos de tempo.

Usualmente, a variação das observações em torno dos níveis (medida por V_t) é bem maior que as variações temporais do nível ao longo do tempo (medidas por W_t).

O modelo de tendência estável é obtido ao particularizar o modelo dinâmico com $\mathbf{F}_t = 1$ e $\mathbf{G}_t = 1$.

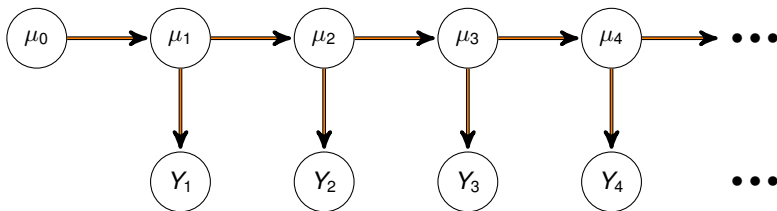


Figura 8: Um DAG do modelo de tendência estável (condicional a $\Psi = (V, W)$).

Aplicando a série de vendas, vendas_t e $\{\mathbf{F}_t = 1, V_t = V, \mathbf{G}_t = 1, \mathbf{W}_t = W\}$, temos que

$$\begin{aligned}\text{vendas}_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W)\end{aligned}$$

Para a série temporal de vendas de doce faremos $V_t = V$ e $W_t = W$ para todo t (homocedasticidade; variâncias constantes ao longo do tempo).

Modelo de tendência estável

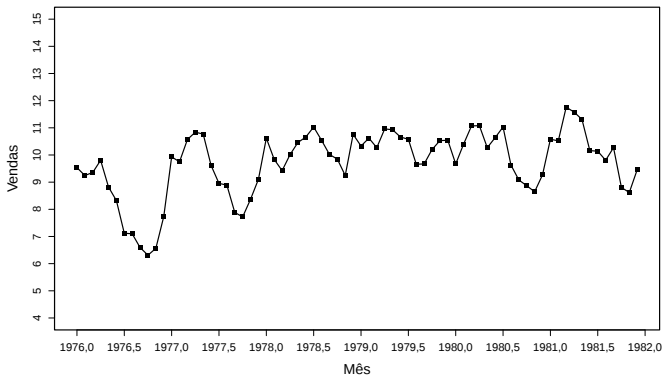


Figura 9: Vendas mensais de um tipo de doces de janeiro de 1976 a dezembro de 1981.

Ver arquivo exemplo_27.r

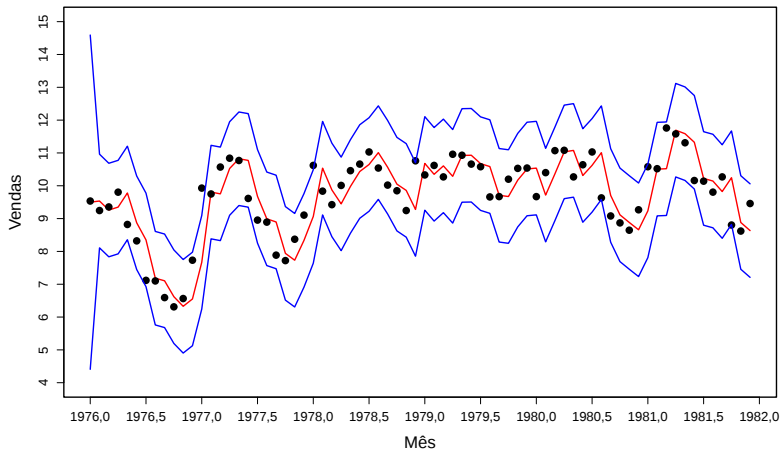


Figura 10: Previsão da média de vendas, $(\mu_t|y_{1:t-1})$. A distribuição a priori em t .

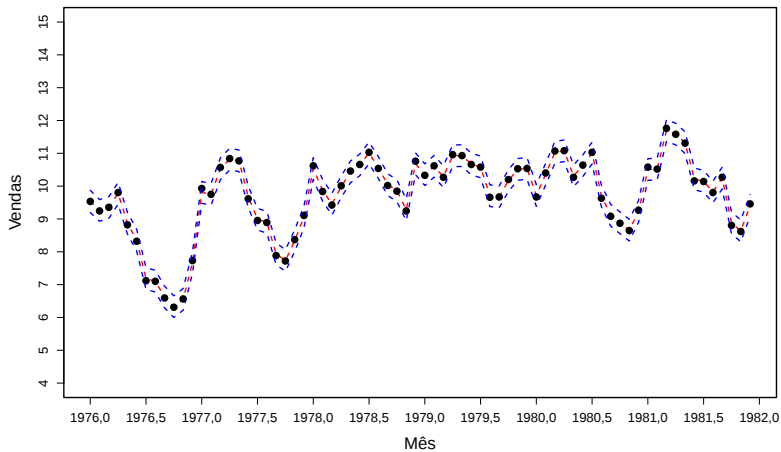


Figura 11: Estimativa da média de vendas, $(\mu_t|y_{1:t})$. A distribuição a posteriori em t .

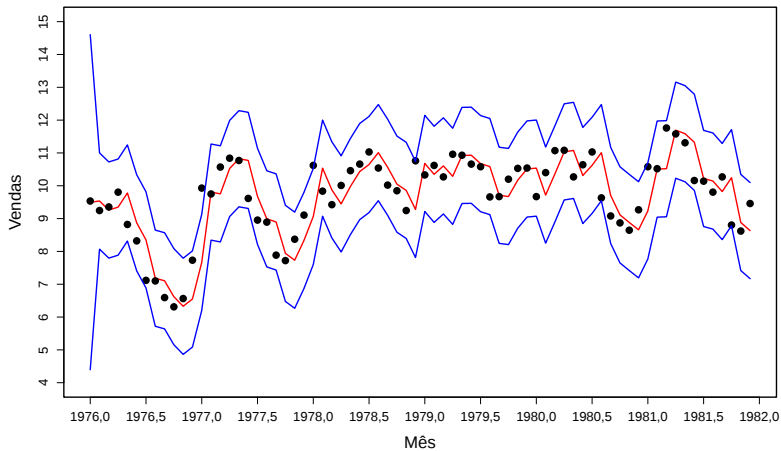


Figura 12: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.

Modelo de tendência linear

A tendência linear é quantificada através de um parâmetro adicional.

O modelo de tendência linear é dado por

$$\begin{aligned}Y_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1t} \\ \beta_t &= \beta_{t-1} + w_{2t},\end{aligned}$$

em que $\mathbf{w}_t = (w_{1t}, w_{2t})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$.

Esse modelo é obtido com $\mathbf{F}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $\mathbf{G}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

O nível permanece localmente linear, mas a forma da reta pode variar com o tempo.

Para a série de vendas de doce faremos $V_t = V$ e $\mathbf{W}_t = \mathbf{W} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ para todo t (independência e variâncias constantes ao longo do tempo).

Modelo de tendência linear

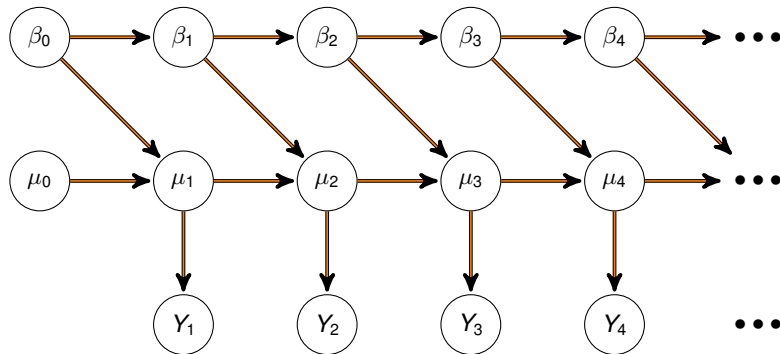


Figura 13: Um DAG do modelo de tendência linear (condicional a $\Psi = (V, \sigma_1^2, \sigma_2^2)$).

Ver arquivo exemplo_28.r

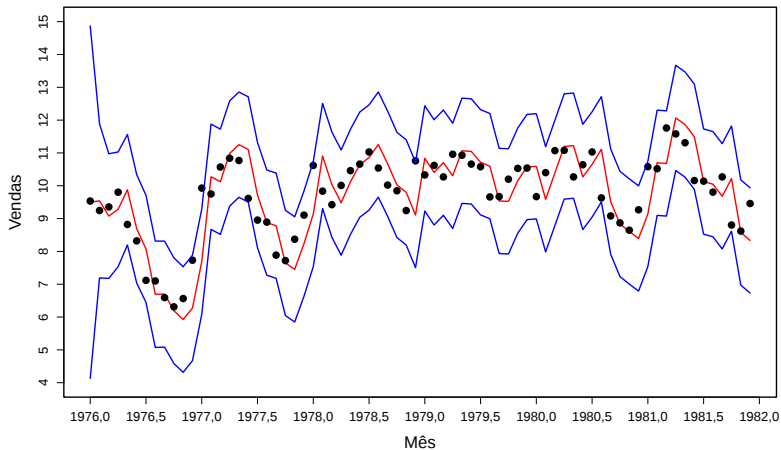


Figura 14: Previsão da média de vendas, $(\mu_t|y_{1:t-1})$. A distribuição a priori em t .

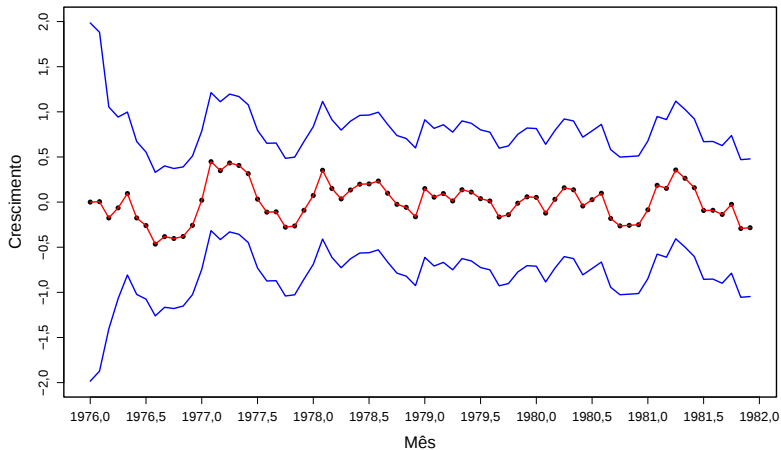


Figura 15: Previsão do crescimento das vendas, $(\beta_t | y_{1:t-1})$. A distribuição a priori em t .

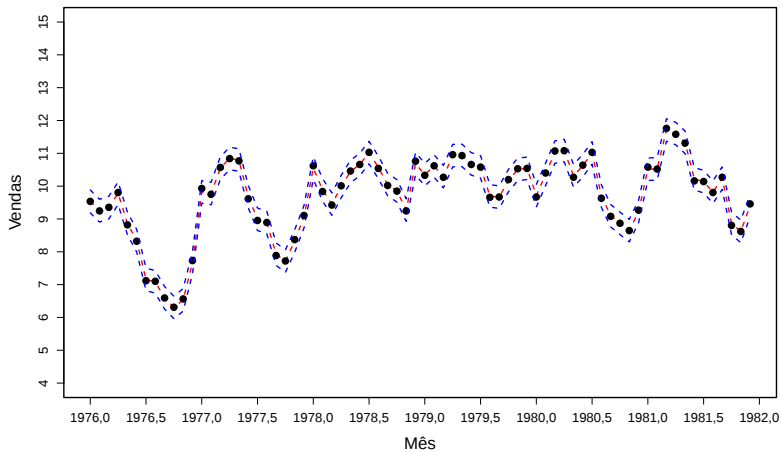


Figura 16: Estimativa da média de vendas, $(\mu_t|y_{1:t})$. A distribuição a posteriori em t .

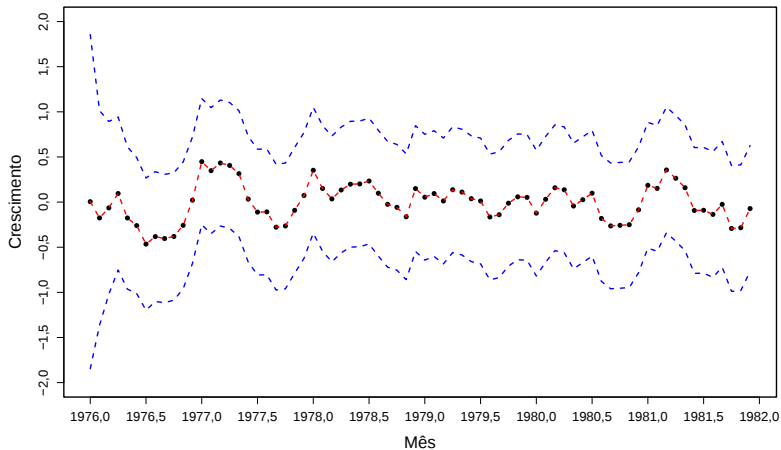


Figura 17: Estimativa do crescimento das vendas, $(\beta_t|y_{1:t})$. A distribuição a posteriori em t .

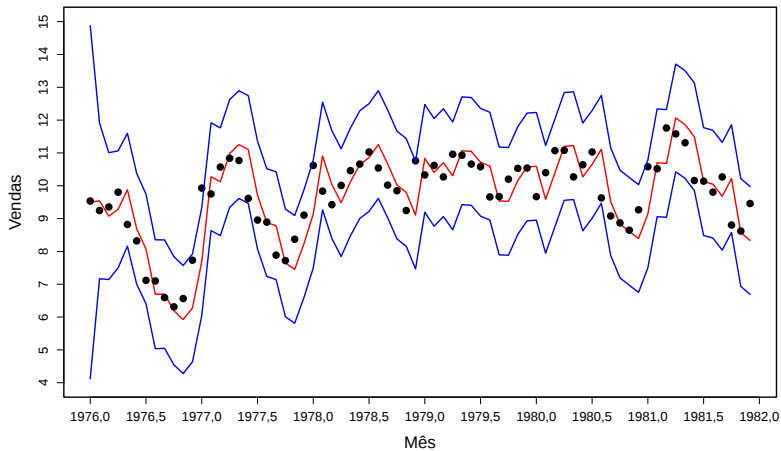


Figura 18: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.

Modelo de regressão dinâmica

Introduzimos variáveis explicativas com coeficientes variando no tempo.

A regressão dinâmica, com $Y_t = \text{vendas}_t$ e $x_t = \text{preço}_t$, é dado por

$$Y_t = \mu_t + \beta_t x_t + v_t, \quad v_t \sim \mathcal{N}(0, V_t)$$

$$\mu_t = \mu_{t-1} + w_{1t}$$

$$\beta_t = \beta_{t-1} + w_{2t},$$

em que $\mathbf{w}_t = (w_{1t}, w_{2t})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$.

Esse modelo é obtido com $\boldsymbol{\theta}_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}$, $\mathbf{F}_t = \begin{pmatrix} 1 \\ x_t \end{pmatrix}$ e $\mathbf{G}_t = \mathbf{I}$.

Para a série de vendas de doce faremos $V_t = V$ e $\mathbf{W}_t = \mathbf{W} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$
para todo t (independência e variâncias constantes ao longo do tempo).

Modelo de regressão dinâmica

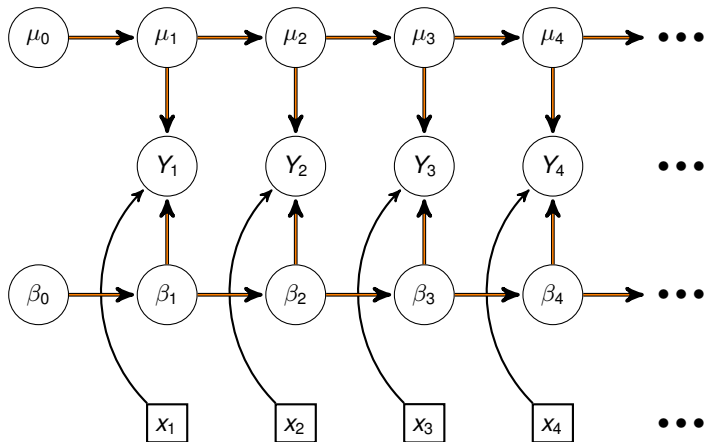


Figura 19: Um DAG do modelo de regressão dinâmica (condicional a $\Psi = (V, \sigma_1^2, \sigma_2^2)$).

Ver arquivo exemplo_29.r

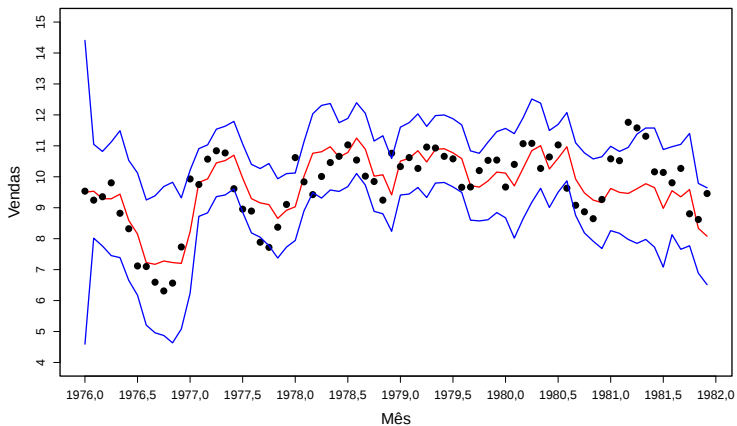


Figura 20: Previsão do nível de vendas, $(\mu_t | y_{1:t-1})$. A distribuição a priori em t .

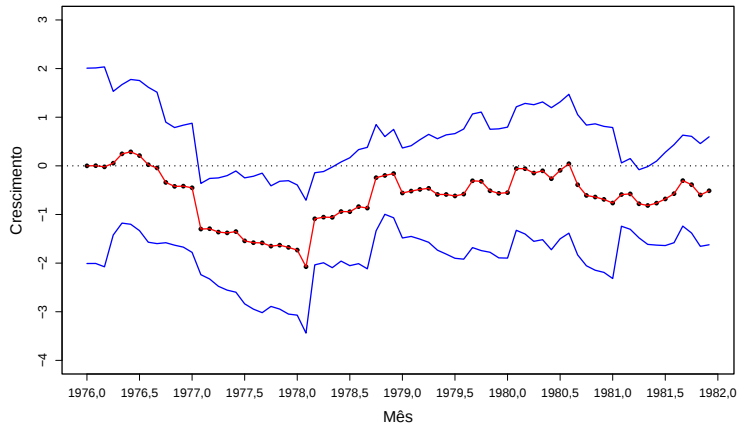


Figura 21: Previsão do crescimento das vendas, $(\beta_t|y_{1:t-1})$. A distribuição a priori em t .

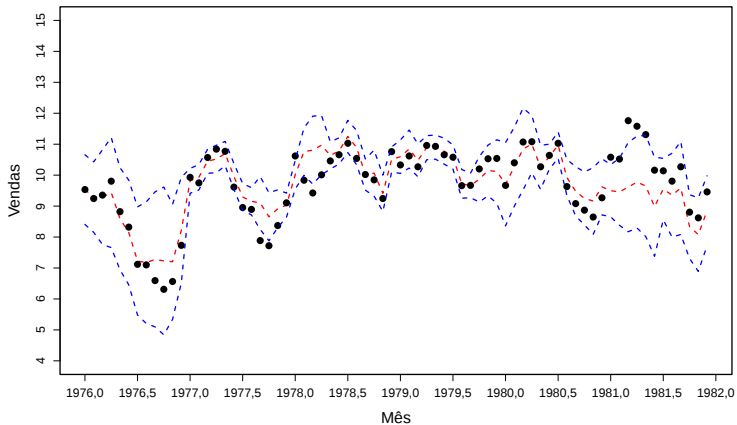


Figura 22: Estimativa do nível de vendas, $(\mu_t | y_{1:t})$. A distribuição a posteriori em t .

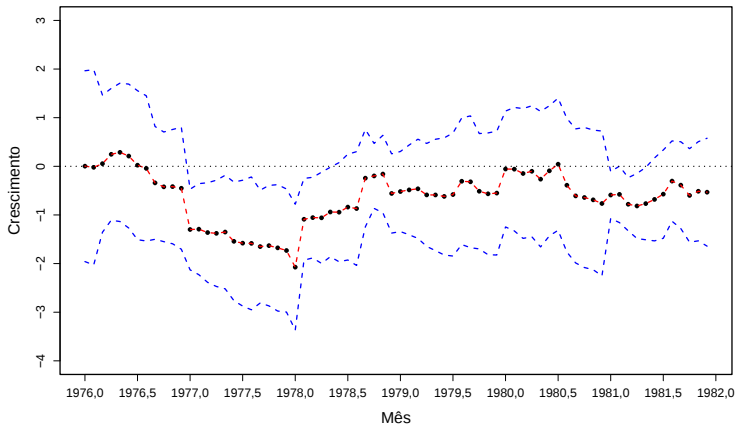


Figura 23: Estimativa do crescimento das vendas, $(\beta_t|y_{1:t})$. A dist. a posteriori em t .

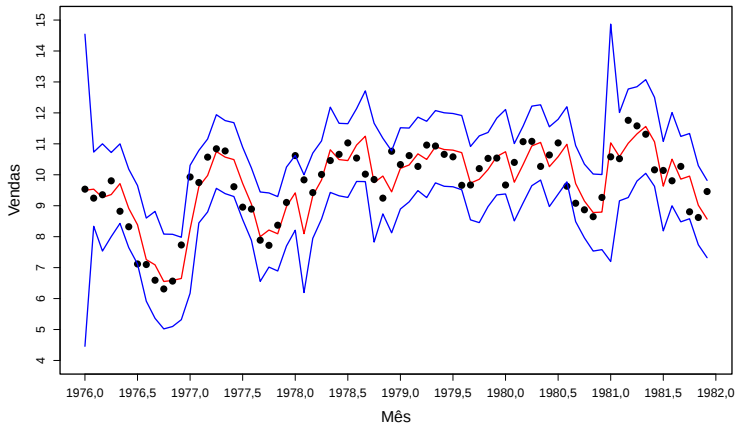


Figura 24: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.

Modelo de efeitos sazonais

Considere o **modelo linear** (um caso particular do modelo linear geral):

$$\begin{aligned}Y_t &= \mathbf{F}^\top \boldsymbol{\theta}_t + v_t, & v_t &\sim \mathcal{N}(0, \mathbf{V}) \\ \boldsymbol{\theta}_t &= \mathbf{G} \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, & \mathbf{w}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{W}) \\ \boldsymbol{\theta}_0 &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0).\end{aligned}$$

Seja s o número de períodos nos dados. Por exemplo, $s = 12$ para dados mensais. Então, defina

$$\mathbf{F} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{G} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Além disso, defina $\mathbf{W} = \text{diag}(\tau^2, 0, 0, \dots, 0)$.

Os efeitos sazonais devem somar zero, assim impomos que $\mathbf{m}_0^\top \mathbf{1} = 0$.

Exemplo

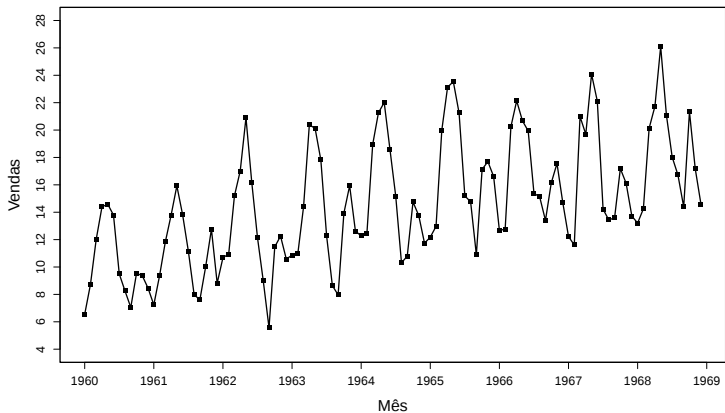


Figura 25: Vendas de carros nos EUA de janeiro de 1960 a dezembro de 1968.

Ver arquivo exemplo_30.r

Exemplo

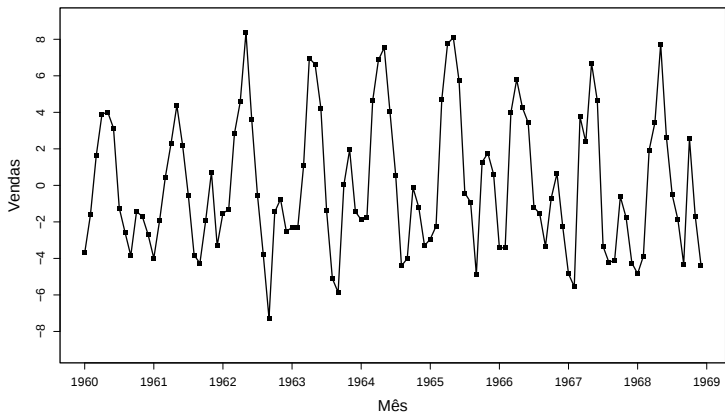


Figura 26: Vendas de carros sem tendência (removida por regressão linear).

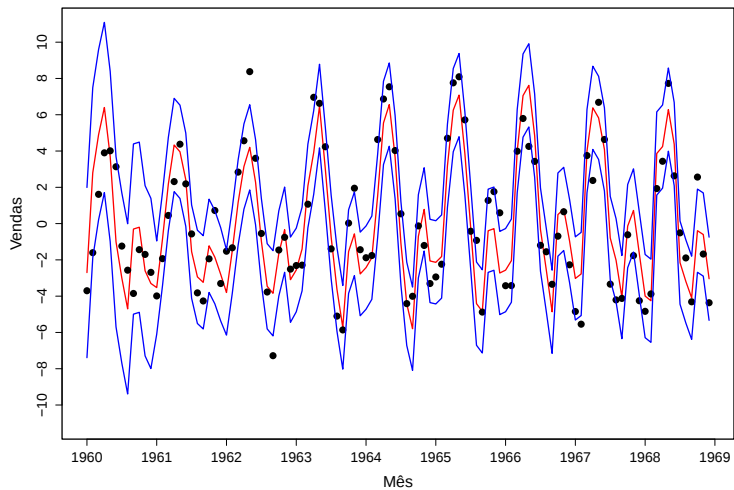


Figura 27: Previsão da sazonalidade das vendas, $(\theta_{1t}|y_{1:t-1})$.

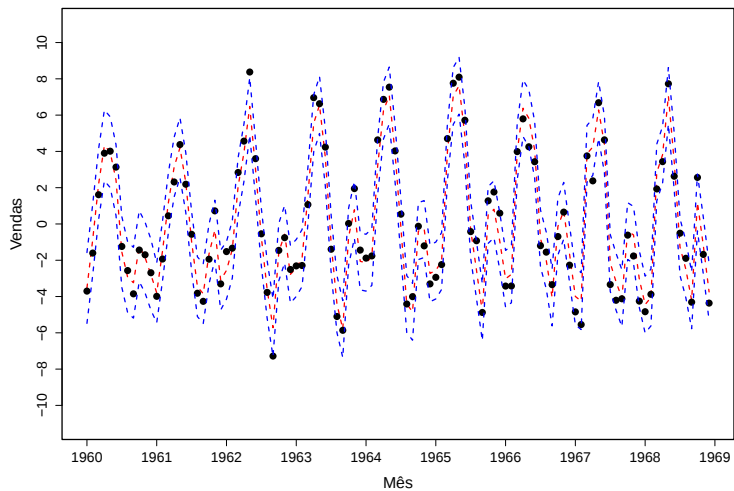


Figura 28: Sazonalidade das vendas, $(\theta_{1t}|y_{1:t})$. A distribuição a posteriori em t .

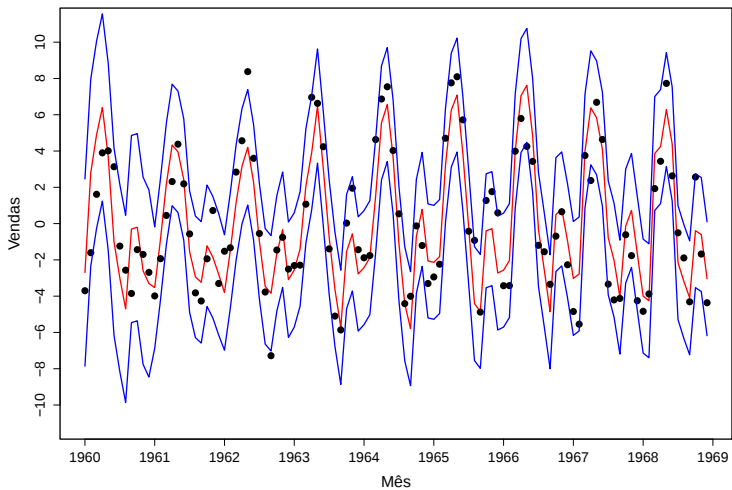


Figura 29: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.

Superposição de modelos

Estamos interessados em uma forma geral para estruturar e acomodar as várias componentes intervenientes em um modelo em espaço de estado. Por exemplo, variáveis causais e sazonalidade.

Muitas séries temporais exibem um comportamento bastante complexo.

Ao identificarmos as características mais marcantes, estamos caminhando na direção de formular um modelo.

A série de vendas de carros é um exemplo típico.

A tendência global parece ser de uma **variação suave do nível**.

Se agora nos concentramos na variação em torno desse nível, podemos detectar um **comportamento cíclico**.

Essa inspeção permitiu identificar duas componentes do modelo: uma componente para a **tendência** e outro para a **sazonalidade**.

A estrutura dos **modelos em espaço de estados** é apropriada para isto, pois permite que as componentes sejam modeladas separadamente e depois integradas em um modelo.

Em um exemplo com duas componentes - **tendência e sazonalidade** - estruturamos a equação das observações com dois termos:

$$Y_t = Y_{Nt} + Y_{St} + v_t.$$

Cada um dos termos é descrito na seguinte dinâmica:

$$\begin{aligned} Y_{Nt} &= \mathbf{F}_{Nt}^\top \boldsymbol{\theta}_{Nt} & \text{e} & & Y_{St} &= \mathbf{F}_{St}^\top \boldsymbol{\theta}_{St} \\ \boldsymbol{\theta}_{Nt} &= \mathbf{G}_{Nt} \boldsymbol{\theta}_{Nt-1} + \mathbf{w}_{Nt} & & & \boldsymbol{\theta}_{St} &= \mathbf{G}_{St} \boldsymbol{\theta}_{St-1} + \mathbf{w}_{St}. \end{aligned}$$

Se juntarmos esses termos, obtemos a **equação das observações**

$$Y_t = \mathbf{F}_t^\top \boldsymbol{\theta}_t + v_t$$

em que

$$\mathbf{F}_t = \begin{pmatrix} \mathbf{F}_{Nt} \\ \mathbf{F}_{St} \end{pmatrix} \quad \text{e} \quad \boldsymbol{\theta}_t = \begin{pmatrix} \boldsymbol{\theta}_{Nt} \\ \boldsymbol{\theta}_{St} \end{pmatrix}.$$

Similarmente, a **equação do sistema (integrado)** fica

$$\theta_t = \mathbf{G}_t \theta_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$$

em que

$$\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{Nt} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{St} \end{pmatrix} \quad \text{e} \quad \mathbf{W}_t = \begin{pmatrix} \mathbf{W}_{Nt} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{St} \end{pmatrix}.$$

Modelos com mais componentes são construídos da mesma forma: cada termo contribui para a equação das observações e com um bloco de parâmetros para a equação do sistema.

Modelo de tendência estável e sazonalidade

O modelo de tendência estável e sazonalidade ($s = 12$) é dado por

$$\begin{aligned}Y_t &= \mu_t + \mathbf{F}_S^\top \boldsymbol{\theta}_{St} + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, \delta^2) \\ \boldsymbol{\theta}_{St} &= \mathbf{G} \boldsymbol{\theta}_{S,t-1} + \mathbf{u}_t, & \mathbf{u}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{U}),\end{aligned}$$

tal que $\mu_0 \sim \mathcal{N}(m_{N0}, C_{N0})$, $\boldsymbol{\theta}_{S0} \sim \mathcal{N}(\mathbf{m}_{S0}, \mathbf{C}_{S0})$, $\mathbf{F}_{Nt} = \mathbf{1}$, $\mathbf{G}_{Nt} = \mathbf{1}$,

$$\mathbf{F}_S = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{G}_S = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Além disso, defina $\mathbf{U} = \text{diag}(\tau^2, 0, 0, \dots, 0)$.

Os efeitos sazonais devem somar zero, assim impomos que $\mathbf{m}_{S,0}^\top \mathbf{1} = \mathbf{0}$.

Modelo de tendência estável e sazonalidade

Podemos descrever o modelo de forma conjunta como

$$\begin{aligned}Y_t &= \mathbf{F}^\top \boldsymbol{\theta}_t + v_t, & v_t &\sim \mathcal{N}(0, V) \\ \boldsymbol{\theta}_t &= \mathbf{G}\boldsymbol{\theta}_{t-1} + \mathbf{w}_t, & \mathbf{w}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{W}) \\ \boldsymbol{\theta}_0 &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0),\end{aligned}$$

com $\boldsymbol{\theta}_t = (\mu_t, \boldsymbol{\theta}_{St}^\top)^\top$, $\mathbf{W} = \text{diag}(\delta^2, \tau^2, 0, \dots, 0)$,

$$\mathbf{F} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 1 & 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

$$\mathbf{m}_0 = (m_{N0}, \mathbf{m}_{S0}^\top)^\top \mathbf{e}$$

$$\mathbf{C}_0 = \begin{pmatrix} C_{N0} & 0 \\ 0 & \mathbf{C}_{S0} \end{pmatrix}.$$

Exemplo

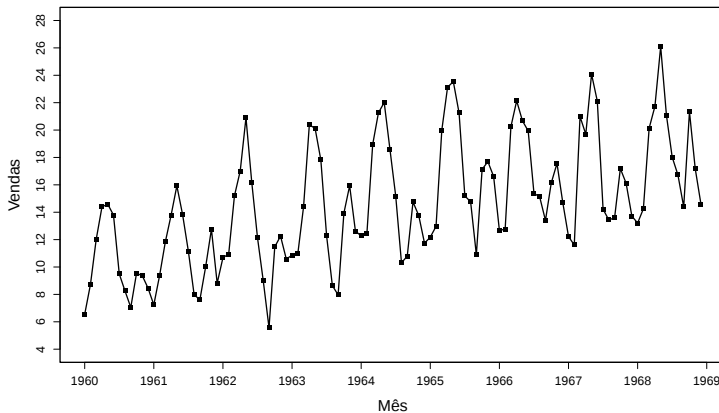


Figura 30: Vendas de carros nos EUA de janeiro de 1960 a dezembro de 1968.

Ver arquivo exemplo_31.r

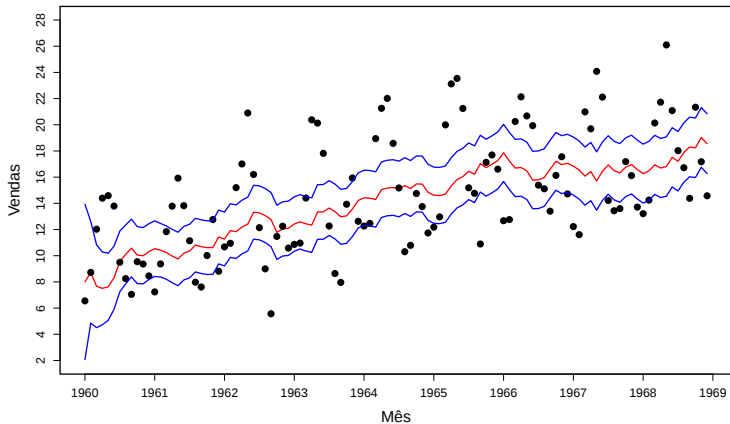


Figura 31: Previsão do nível de vendas, $(\mu_t | y_{1:t-1})$. A distribuição a priori em t .

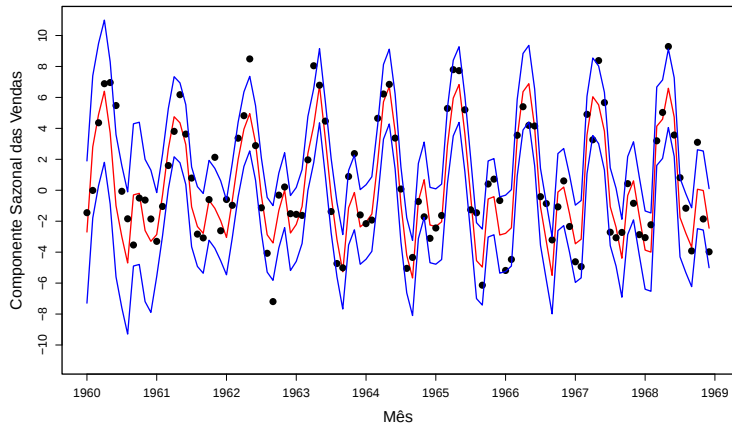


Figura 32: Previsão da sazonalidade das vendas, $(\theta_{2t} | y_{1:t-1})$.

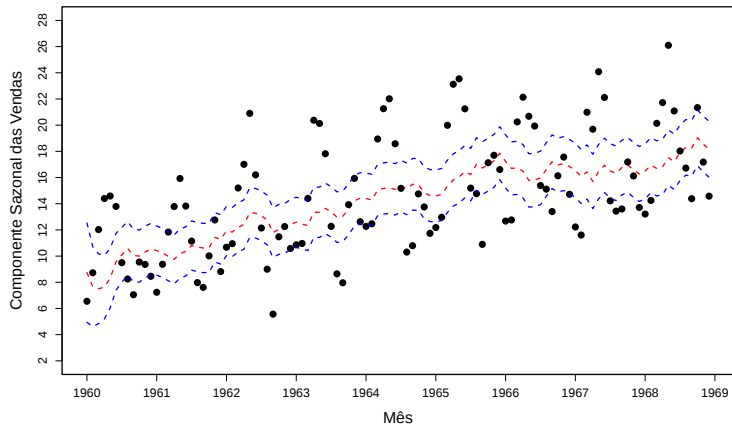


Figura 33: Estimativa do nível de vendas, $(\mu_t|y_{1:t})$. A distribuição a posteriori em t .

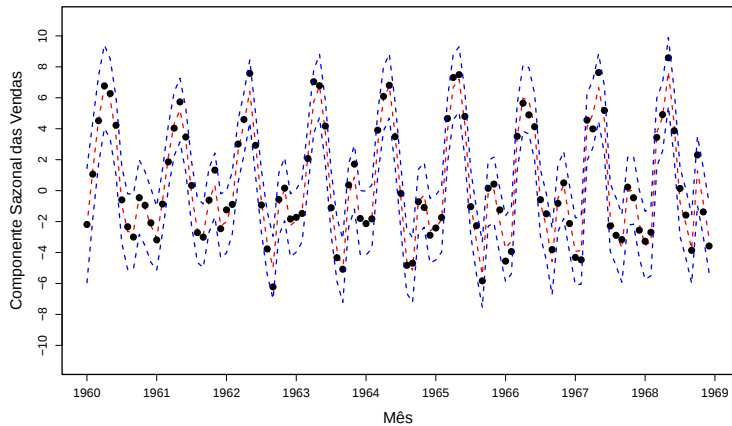


Figura 34: Sazonalidade das vendas, $(\theta_{2t} | y_{1:t})$. A distribuição a posteriori em t .

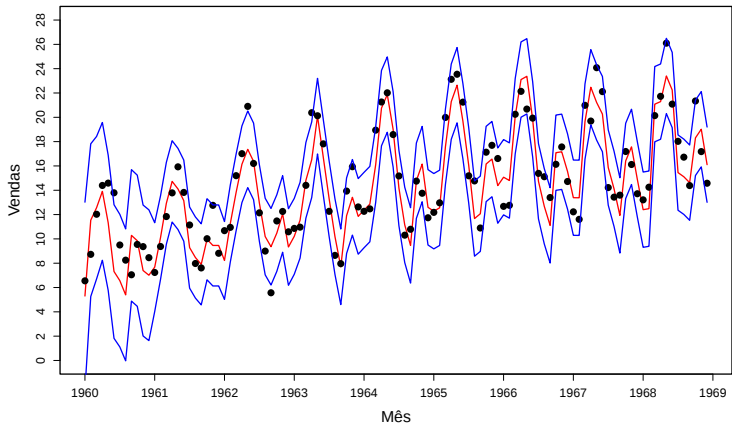


Figura 35: Previsão (um passo à frente) das vendas, $(y_t|y_{1:t-1})$.

Modelos em Espaço de Estados

Análise suavizada via OpenBUGS

Modelo de tendência estável (revisitando)

É composto apenas de um nível que varia segundo um passeio aleatório.

O modelo de tendência estável é dado por

$$\begin{aligned} \text{vendas}_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \mu_t &= \mu_{t-1} + w_t, & w_t &\sim \mathcal{N}(0, W_t). \end{aligned}$$

O nível permanece localmente constante, mas varia quando se considera longos períodos de tempo.

O modelo de tendência estável é obtido ao particularizar o modelo dinâmico com $\mathbf{F}_t = 1$ e $\mathbf{G}_t = 1$.

Para a série temporal de vendas de doce faremos $V_t = V$ e $W_t = W$ para todo t (homocedasticidade; variâncias constantes ao longo do tempo).

Observações importantes

Consideraremos a análise somente após observar toda a série. Isto se chama **análise suavizada**.

Nesta parte da aula, consideraremos somente a análise bayesiana.

Descreveremos o modelo e a distribuição priori dos parâmetros.

Utilizaremos o programa **OpenBUGS** para fazer as análises.

Calcularemos o **Critério de Informação do Desvio (DIC)** para comparar modelos.

Ver arquivo [doce_modelo_1.odc](#)

Operações básicas no OpenBUGS

1. *Model*→ *Specification...*
2. *check model*
Verificar por *model is syntactically correct*
3. *load data*
Verificar por *data loaded*
4. *compile*
Verificar por *model compiled*
5. *load inits* ou *gen inits*
Verificar por *initial values generated, model initialized*
6. *Inference*→ *Samples...*
7. *Model*→ *Update...* (até a convergência da cadeia)
8. *Inference*→ *DIC...*
9. *Model*→ *Update...* (para obter a amostra da posteriori)
10. *Inference*→ *Samples...* (para a análise dos valores gerados)
11. *Inference*→ *DIC...* (para conferir o valor do DIC)

Modelo de tendência linear (revisitando)

A tendência linear é quantificada através de um parâmetro adicional.

O modelo de tendência linear é dado por

$$\begin{aligned} vendas_t &= \mu_t + v_t, & v_t &\sim \mathcal{N}(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{1t} \\ \beta_t &= \beta_{t-1} + w_{2t}, \end{aligned}$$

em que $\mathbf{w}_t = (w_{1t}, w_{2t})' \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$.

Esse modelo é obtido com $\mathbf{F}_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $\mathbf{G}_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

O nível permanece localmente linear, mas a forma da reta pode variar com o tempo.

Para a série de vendas de doce faremos $V_t = V$ e $\mathbf{W}_t = \mathbf{W} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$ para todo t (independência e variâncias constantes ao longo do tempo).

Ver arquivo doce_modelo_2.odc

Modelo de regressão dinâmica (revisitando)

Introduzimos variáveis explicativas com coeficientes variando no tempo.

O modelo de regressão dinâmica é dado por

$$vendas_t = \mu_t + \beta_t \text{preço}_t + v_t, \quad v_t \sim \mathcal{N}(0, V_t)$$

$$\mu_t = \mu_{t-1} + w_{1t}$$

$$\beta_t = \beta_{t-1} + w_{2t},$$

em que $\mathbf{w}_t = (w_{1t}, w_{2t})^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t)$.

Esse modelo é obtido com $\boldsymbol{\theta}_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}$, $\mathbf{F}_t = \begin{pmatrix} 1 \\ \text{preço}_t \end{pmatrix}$ e $\mathbf{G}_t = \mathbf{I}$.

Para a série de vendas de doce faremos $V_t = V$ e $\mathbf{W}_t = \mathbf{W} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$
para todo t (independência e variâncias constantes ao longo do tempo).

Ver arquivo doce_modelo_3.odc

Previsão

Estamos interessados na previsão de horizonte h na origem t , y_{t+h} .

Consideraremos o conjunto $\{y_{t+1}, y_{t+2}, \dots, y_{t+h}\}$ como não observados (dados faltantes).

Do ponto de vista bayesiano, estimaremos estas quantidades condicional aos dados.

Este procedimento é facilmente implementado no **OpenBUGS**.

Em geral, tomamos a média a posterior de y_{t+h} como a previsão (pontual).

Temos também o intervalo de previsão (intervalo de credibilidade) de y_{t+h} .

Ver arquivos [doce_modelo_1_b.odc](#) e [doce_modelo_2_b.odc](#)

Modelo de efeitos sazonais (outra forma)

Considere o modelo $y_t = \gamma_t + v_t$, em que γ_t é a componente de sazonalidade.

Seja s o número de períodos nos dados. Por exemplo,

- $s = 12$ para dados mensais;
- $s = 4$ para dados trimestrais; e
- $s = 7$ para dados diários.

Uma forma de modelar efeitos sazonais é através de variáveis binárias.

A soma dos efeitos do período devem ser igual a zero,

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j}.$$

Para permitir que o padrão mude ao longo do tempo, introduzimos um novo termo de erro,

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + w_t, \quad w_t \sim \mathcal{N}(0, \sigma^2).$$

O valor esperado da soma dos efeitos sazonais é zero.

Exemplo de Superposição: Tendência Linear e Sazonalidade

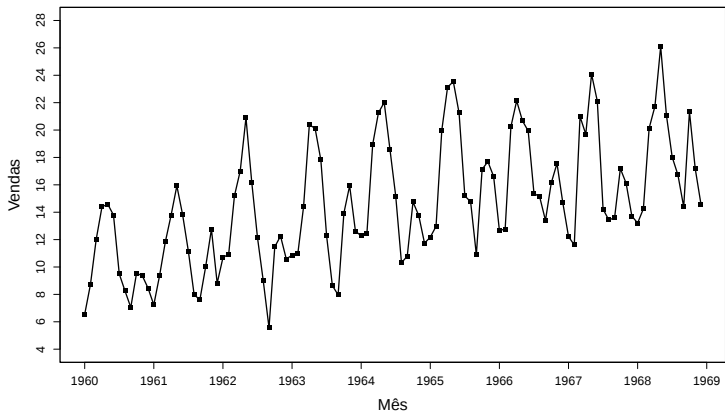


Figura 36: Vendas de carros nos EUA de janeiro de 1960 a dezembro de 1968.

Ver arquivo [carro_modelo_1.odc](#)

Motivação



Figura 37: Índice BOVESPA de 3 de janeiro de 2011 até 10 de setembro de 2021.

Ver arquivo exemplo_32.r

Motivação

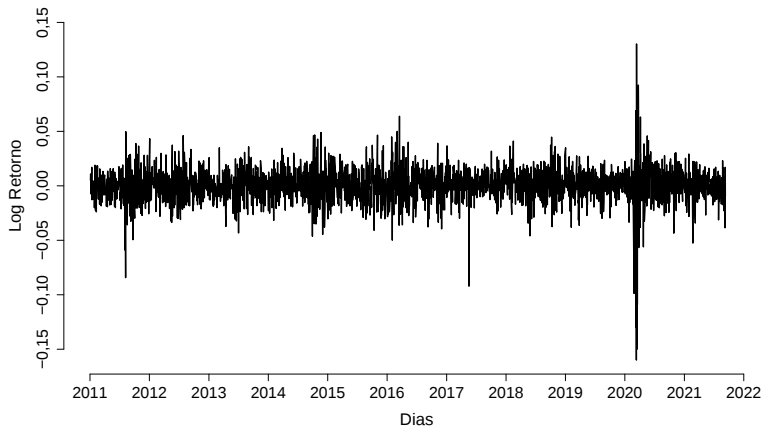


Figura 38: Log retornos diários do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021.

Ver arquivo [exemplo_32.r](#)

Retorno de um período:

Seja P_t o preço do ativo no tempo t .

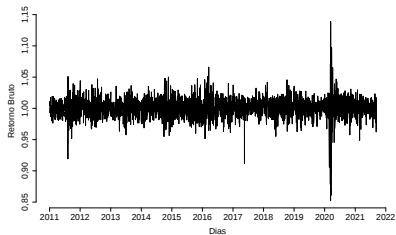
Fixando o ativo pelo período de tempo de $t - 1$ a t , definimos:

- **retorno bruto** por

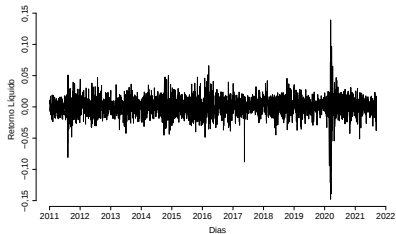
$$1 + R_t = \frac{P_t}{P_{t-1}};$$

- **retorno líquido** por

$$R_t = \frac{P_t}{P_{t-1}} - 1 = \frac{P_t - P_{t-1}}{P_{t-1}}.$$



(a) Retorno bruto



(b) Retorno líquido

Figura 39: Retornos do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021.

Retorno de vários períodos:

Fixando o ativo pelos períodos de tempo entre $t - k$ a t , temos o **retorno bruto de k períodos** (ou **o retorno composto**) dado por

$$1 + R_t[k] = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+1}}{P_{t-k}} = \prod_{j=0}^{k-1} (1 + R_{t-j}).$$

O **retorno líquido para k períodos** é dado por

$$R_t[k] = \frac{P_t}{P_{t-k}} - 1 = \frac{P_t - P_{t-k}}{P_{t-k}}.$$

Se um ativo é mantido por k períodos, então o retorno (líquido) médio é dado por

$$\begin{aligned}\tilde{R}_t[k] &= \overbrace{\left[\prod_{j=0}^{k-1} (1 + R_{t-j}) \right]}^{\text{média geométrica}}^{1/k} - 1 \\ &= \exp \left\{ \frac{1}{k} \sum_{j=0}^{k-1} \ln(1 + R_{t-j}) \right\} - 1.\end{aligned}$$

Uma aproximação de Taylor de primeira ordem para o retorno médio é dada por

$$\tilde{R}_t[k] \simeq \frac{1}{k} \sum_{j=0}^{k-1} R_{t-j},$$

pois

$$g(x) = \ln(1 + x) \simeq x \quad \text{e} \quad h(x) = \exp(x) \simeq 1 + x,$$

para x próximo de zero (tomando $x_0 = 0$).

Esta aproximação pode ser ruim em alguns casos.

Capitalização contínua

- Suponha que a taxa de juros de um depósito bancário seja 10% ao ano e que o depósito inicial seja de \$1,00.
- Se o banco paga os juros uma vez ao ano, então o valor líquido do depósito se torna $\$1,00 \times (1 + 0,1) = \$1,10$ depois de um ano.
- Se o banco paga os juros a cada 6 meses, então a taxa de juros é $10\%/2 = 5\%$ e o valor líquido do depósito é $\$1,00 \times (1 + 0,1/2)^2 = \$1,1025$ depois de um ano.
- Se o banco paga os juros m vezes por ano, então a taxa de juros é $10\%/m$ e o valor líquido do depósito é $\$1,00 \times (1 + 0,1/m)^m$ depois de um ano.
- A tabela a seguir mostra resultados para alguns intervalos de tempo sobre um depósito de \$1,00 com juros de 10% ao ano.
- O valor líquido se aproxima de \$1,1052, que é obtido por $\exp(0,1)$ e é referido como o resultado da **capitalização contínua**.
- O efeito da capitalização pode ser claramente visto.

Tabela 2: Ilustração do efeito da capitalização^a.

Tipo	Número de Pagamentos	Taxa de Juros por Período	Valor Líquido
Anual	1	0,1	1,10000
Semestral	2	0,05	1,10250
Trimestral	4	0,025	1,10381
Mensal	12	0,0083	1,10471
Semanal	52	0,1/52	1,10506
Diário	365	0,1/365	1,10516
Contínuo	∞		1,10517

^a O intervalo de tempo é de 1 ano e a taxa de juros é 10%.

A função exponencial pode ser representada por

$$\exp(x) = \lim_{m \rightarrow \infty} \left(1 + \frac{x}{m}\right)^m.$$

Retorno de capitalização contínua

O logaritmo natural de um retorno bruto de um ativo é chamado de **retorno de capitalização contínua** ou **log retorno**:

$$r_t = \ln(1 + R_t) = \ln \frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1} = p_t - p_{t-1},$$

em que $p_t = \ln P_t$ (por definição).

Considerando log retornos de vários períodos, temos

$$\begin{aligned} r_t[k] &= \ln(1 + R_t[k]) = \ln((1 + R_t)(1 + R_{t-1}) \dots (1 + R_{t-k+1})) \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + \dots + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + \dots + r_{t-k+1} \end{aligned}$$

- O log retorno de vários períodos é simplesmente a soma dos log retornos de um período.
- Propriedades estatísticas dos log retornos são mais tratáveis.
- Note também que $r_t[k] = \ln P_t - \ln P_{t-k}$.

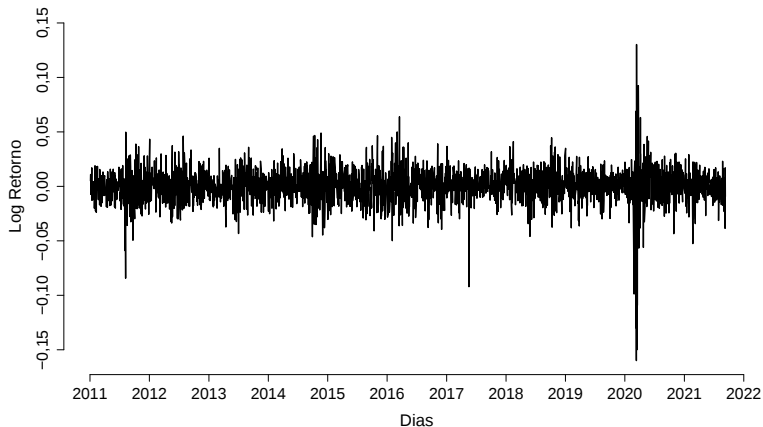


Figura 40: Log retornos do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021.

Ver arquivo exemplo_32.r

Modelos condicionalmente heterocedásticos

O objetivo é a modelagem da volatilidade do retorno de um ativo.

A volatilidade é um fator importante na negociação de opções (por exemplo, fórmula de Black-Scholes).

A volatilidade tem efeito sobre cálculo do valor em risco ($V@R$).

A volatilidade significa a variância (ou desvio padrão) condicional do retorno de um ativo.

A modelagem da volatilidade de uma série temporal pode melhorar a eficiência na estimação dos parâmetros e na precisão do intervalo de previsão.

Veremos os seguintes modelos:

- ARCH:** autoregressivo condicionalmente heterocedástico;
- GARCH:** ARCH generalizado; e
- SV:** volatilidade estocástica.

Características da volatilidade

A volatilidade de uma ação não é diretamente observável.

Assim, é difícil avaliar a performance de previsão nos modelos condicionalmente heterocedásticos.

Existem conglomerados de volatilidade (períodos de alta e de baixa).

A volatilidade muda de forma contínua através do tempo.

A volatilidade não diverge para o infinito. Logo, a volatilidade é geralmente estacionária.

A volatilidade parece reagir diferentemente para um aumento grande de preço ou uma queda grande de preço, conhecido como efeito de influência (*leverage effect*).

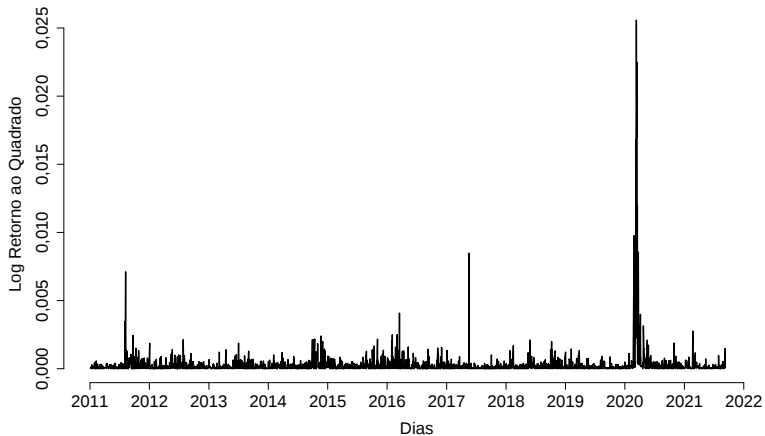
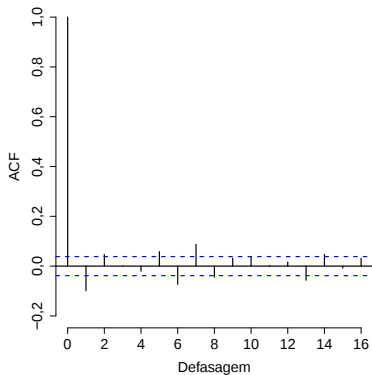
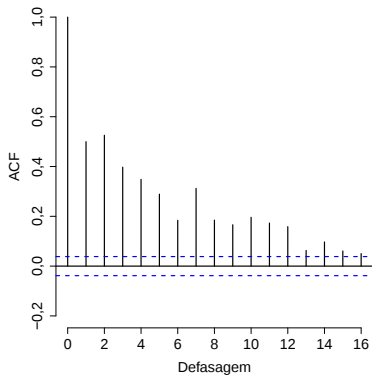


Figura 41: Log retornos diários ao quadrado do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021.

Ver arquivo [exemplo_32.r](#)



(a) Log retornos



(b) Log retornos ao quadrado

Figura 42: ACF dos log retornos e log retornos ao quadrado do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021.

O ACF dos retornos r_t sugere correlação serial fraca.

O ACF de r_t^2 indica que estes (retornos ao quadrado) são correlacionados.

Logo, os retornos r_t **são dependentes**.

Os modelos de volatilidade tentam capturar esta dependência na série de retornos.

Seja $\mathcal{F}_{t-1}$ a informação disponível até o tempo $t - 1$. Temos que

$$\mu_t = E(r_t | \mathcal{F}_{t-1}) \quad \text{e} \quad \sigma_t^2 = \text{Var}(r_t | \mathcal{F}_{t-1}).$$

A série de retornos tem uma estrutura de correlação baixa (ver ACF).

Podemos utilizar um modelo ARMA(p, q) e regressores. Temos $r_t = \mu_t + a_t$ e

$$\mu_t = \alpha + \sum_{j=1}^k \beta_j x_{jt} + \sum_{j=1}^p \phi_j r_{t-j} + \sum_{j=1}^q \theta_j a_{t-j},$$

para algum k inteiro e x_{jt} como regressores.

Para dados diários, x_{jt} poderia ser uma variável binária representando a segunda-feira para estudar o efeito do fim de semana no retorno de uma ação.

Modelos ARCH

Um modelo ARCH(m) é dado por

$$r_t = \mu_t + a_t, \quad a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_m a_{t-m}^2,$$

em que

- ε_t é uma sequência iid (ruído branco) com média 0 e variância 1 ; e
- $\alpha_0 > 0$ e $\alpha_j \geq 0$ para $j > 0$.

Os coeficientes α_j devem satisfazer a certas condições para garantir que a variância incondicional de a_t seja finita.

Em geral, ε_t segue distribuição normal ou t -Student. Podemos incluir outras distribuições e também permitir a modelagem da assimetria.

Note que valores de $\{a_{t-j}^2\}_{j=1}^m$ grandes, implicam que σ_t^2 será grande.

Se σ_t^2 é grande, então a_t poderá ser grande com maior probabilidade (se comparado a um σ_t^2 menor).

Construção do modelo

Construir um modelo de volatilidade para uma série de retornos de um ativo consiste em quatro passos:

1. Especifique a equação da média testando a dependência serial nos dados. Se necessário, construir um modelo econométrico (por exemplo um modelo ARMA) para a série de retornos para remover qualquer dependência linear.
2. Use os resíduos da equação da média para testar a existência de efeitos ARCH.
3. Especifique um modelo de volatilidade se os efeitos ARCH são estatisticamente significantes e faça a estimação conjunta das equações da média e da volatilidade.
4. Verifique o modelo ajustado cuidadosamente e refine-o se necessário.

Observações

A maioria das séries de retornos tem correlação serial fraca (talvez só precisemos remover a média).

Para algumas séries de retornos diários, um modelo AR (ou MA) pode ser necessário.

Em outros casos, uma variável indicadora pode ser necessária (indicadora do dia da semana ou do mês).

Teste para efeitos ARCH: Ljung-Box

Seja $\hat{a}_t = r_t - \hat{\mu}_t$ os resíduos da equação da média.

A série $\{\hat{a}_t^2\}$ é usada para verificar a heterocedasticidade condicional.

Teste de Ljung-Box com a estatística $Q(m^*)$ para a série $\{\hat{a}_t^2\}$.

H_0 : as primeiras m^* defasagens da ACF de $\hat{a}_t^2$ são zero.

Note que m^* , em geral, é diferente de m (ordem do modelo ARCH).

Ver arquivo exemplo_32.r

Fraquezas dos modelos ARCH

O modelo assume que choques positivos e negativos têm o mesmo efeito na volatilidade.

As restrições do modelo ARCH são bem complicadas. Por exemplo, para um ARCH(1) devemos ter $\alpha_1^2 \in [0, 1/3]$ se a série tiver o quarto momento finito.

Os modelos ARCH não nos ajudam a entender a fonte de variação de uma série temporal financeira (apenas descreve o comportamento).

Os modelos ARCH, em geral, fazem previsões acima para a volatilidade porque eles respondem devagar a choques grandes e isolados da série de retornos.

Construindo um modelo ARCH

Para determinar a ordem, podemos utilizar a PACF de $\hat{a}_t^2$.

No modelo, temos

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_m a_{t-m}^2.$$

Podemos substituir σ_t^2 por seu estimador (**tendencioso**) $\hat{a}_t^2$, e utilizar também $\hat{a}_{t-1}^2, \dots, \hat{a}_{t-m}^2$. Assim, teremos um modelo similar ao AR(m).

A estimação de modelos ARCH é feita através do método da máxima verossimilhança (condicional).

A função de verossimilhança é complicada e métodos numéricos são empregados para se obter as estimativas.

Podemos utilizar erros normais, t -Student, GED, e algumas distribuições assimétricas.

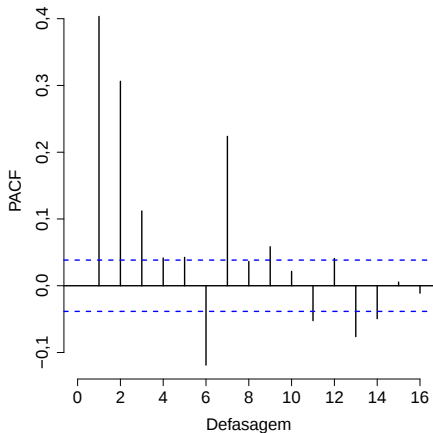


Figura 43: Autocorrelação parcial para os resíduos ao quadrado do modelo AR(10) ajustado aos log retornos diários do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021.

Verificação do modelo

Do modelo ARCH, temos que $\frac{a_t}{\sigma_t} = \varepsilon_t$, em que ε_t é um ruído branco.

Os resíduos padronizados $\frac{\hat{a}_t}{\hat{\sigma}_t} = e_t$ devem se comportar como ruído branco.

Teste de Ljung-Box para verificar a dependência.

Teste de Jarque e Bera (no caso de normalidade).

Histogramas, *boxplot* e *qq-plot* para analisar a hipótese da distribuição (normal, *t*-Student, etc.).

Previsão

A previsão do modelo ARCH pode ser obtida recursivamente como as de um modelo AR.

Ver arquivo [exemplo_33.r](#)

Modelos GARCH

Um modelo GARCH(m, s) é dado por

$$a_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \cdots + \alpha_m a_{t-m}^2 \\ + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \cdots + \beta_s \sigma_{t-s}^2,$$

em que

- ε_t é uma sequência iid com média 0 e variância 1;
- $\alpha_0 > 0$ e $\alpha_j \geq 0$ para $j > 0$, e $\beta_j \geq 0$ para $j \geq 1$; e
- $\sum_{i=j}^{\max(m,s)} (\alpha_j + \beta_j) < 1$ (para a variância incondicional de a_t ser finita).

Se $s = 0$, então o modelo se reduz a um ARCH(m).

Os α_j são chamados de parâmetros ARCH e os β_j de parâmetros GARCH.

O modelo pode resultar em uma forma mais simples de descrever a volatilidade de uma série.

Em geral, aplicamos o método de máxima verossimilhança condicional.

Os modelos GARCH, após alguns cálculos, podem ser vistos como um ARMA no quadrado dos retornos.

A previsão do modelo GARCH pode ser obtida recursivamente como as de um modelo ARMA.

Na prática, os melhores ajustes são dados por modelos **GARCH(1,1)**, GARCH(2,1) ou GARCH(1,2).

Ver arquivo exemplo_34.r

Modelo de volatilidade estocástica

O modelo de volatilidade estocástica é uma alternativa ao modelo GARCH. Este modelo também está na classe dos modelos dinâmicos (não lineares).

O modelo básico é dado por

$$\begin{aligned}r_t &= \exp\{h_t/2\}\varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0; 1) \quad (\text{equação das observações}) \\h_t &= \alpha + \phi(h_{t-1} - \alpha) + \omega_t, & \omega_t &\sim \mathcal{N}(0; \sigma^2) \quad (\text{equação dos estados}),\end{aligned}$$

para $t = 1, 2, \dots, n$. Temos que

- r_t é o log retorno, isto é, $r_t = \ln(P_t) - \ln(P_{t-1})$;
- h_t é a log volatilidade;
- α é o nível (constante) da log volatilidade;
- ϕ é o parâmetro de persistência; e
- σ^2 é a variância das log volatilidades.

Inferência bayesiana e função de verossimilhança

Utilizaremos a abordagem bayesiana para inferir sobre as quantidades desconhecidas do modelo.

Construímos a função de verossimilhança utilizando as equações das observações e do estado:

$$f(\mathbf{r}, \mathbf{h} | \alpha, \phi, \sigma^2) \propto \left[\prod_{t=1}^n f(r_t | h_t) \right] \left[\prod_{t=2}^n f(h_t | h_{t-1}, \alpha, \phi, \sigma^2) \right] f(h_1 | \alpha, \phi, \sigma^2),$$

em que

$$(h_1 | \alpha, \phi, \sigma^2) \sim \mathcal{N} \left(\alpha; \frac{\sigma^2}{1 - \phi^2} \right),$$

pois a log volatilidade é um AR(1) estacionário.

Além disso, $\mathbf{r} = (r_1, r_2, \dots, r_n)$ e $\mathbf{h} = (h_1, h_2, \dots, h_n)$.

Distribuição a priori

Assumimos que $f(\alpha, \phi, \sigma^2) = f(\alpha)f(\phi)f(\sigma^2)$.

Poderíamos ter a restrição que $|\phi| < 1$ por ser um AR(1) estacionário.

Contudo, aplicamos a restrição que $0 \leq \phi < 1$ utilizando uma priori uniforme.

Distribuição a priori é dada por

$$\begin{array}{lll} \alpha & \sim & \mathcal{N}(0; 10.000) & (\text{priori vaga}) \\ \phi & \sim & \mathcal{U}(0; 1) & (\text{priori vaga}) \\ \sigma^2 & \sim & \mathcal{IG}(0, 01; 0, 01) & (\text{priori vaga}). \end{array}$$

Distribuição a posteriori

A distribuição a posteriori é dada por

$$\begin{aligned} f(\alpha, \phi, \sigma^2, \mathbf{h}|\mathbf{r}) &\propto \left[\prod_{t=1}^n f(r_t|h_t) \right] \left[\prod_{t=2}^n f(h_t|h_{t-1}, \alpha, \phi, \sigma^2) \right] \\ &\times f(h_1|\alpha, \phi, \sigma^2) f(\alpha, \phi, \sigma^2), \end{aligned}$$

em que $f(\alpha, \phi, \sigma^2)$ é a distribuição a priori.

Lembrando que $\mathbf{r} = (r_1, r_2, \dots, r_n)$ e $\mathbf{h} = (h_1, h_2, \dots, h_n)$.

Recorreremos aos métodos de Monte Carlo via cadeias de Markov, via programa **OpenBUGS**, para fazer inferência sobre o $(h_1, h_2, \dots, h_n, \alpha, \phi, \sigma^2)$.

Exemplo

Ver arquivo exemplo_35.r



Figura 44: Índice BOVESPA semanal de 3 de janeiro de 2011 até 10 de setembro de 2021.

Exemplo (continuação)

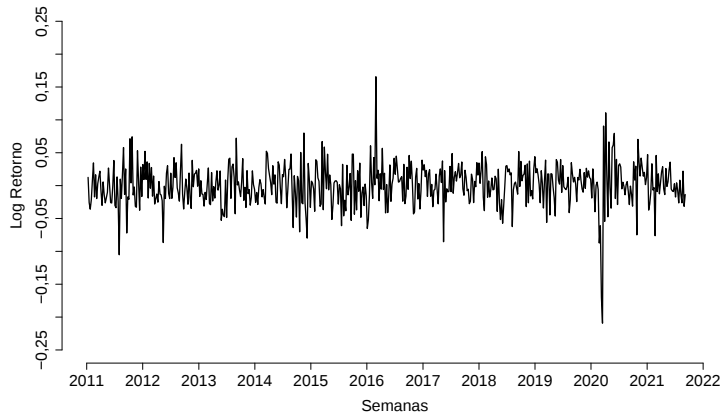


Figura 45: Log retornos semanais do IBOVESPA de 4 de janeiro de 2011 até 10 de setembro de 2021 ($n = 557$).

Previsão e resultados

A previsão segue o padrão mostrado para os modelos dinâmicos.

Ver arquivo volatilidade_1.odc

Parâmetro	Média	D.P.	2,5%	Mediana	97,5%
α	-7,141	0,1213	-7,386	-7,139	-6,903
ϕ	0,8511	0,0606	0,708	0,8599	0,941
σ^2	0,1111	0,0281	0,044	0,1014	0,230
r_{558}	0,0003	0,0265	-0,053	0,0002	0,053
r_{559}	0,0001	0,0281	-0,057	0,0006	0,057
r_{560}	-0,0001	0,0281	-0,057	-0,0005	0,057
h_{558}	-7,377	0,5615	-8,498	-7,372	-6,279
h_{559}	-7,347	0,5838	-8,511	-7,344	-6,213
h_{560}	-7,325	0,5989	-8,523	-7,316	-6,137

Superposição de modelos

A equação da média com estrutura de tendência estável

$$\begin{aligned}r_t &= \gamma_t + \exp\{h_t/2\}\varepsilon_t, & \varepsilon_t &\sim \mathcal{N}(0, 1) \\ \gamma_t &= \gamma_{t-1} + \xi_t, & \xi_t &\sim \mathcal{N}(0, \delta^2).\end{aligned}$$

Podemos incluir o retorno defasado como um AR(1), por exemplo

$$r_t = \gamma + \beta r_{t-1} + \exp\{h_t/2\}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1)$$

Considere o modelo

$$\begin{aligned}r_t &= \gamma + \beta r_{t-1} + \exp\{h_t/2\}\varepsilon_t \\ h_t &= \alpha + \phi(h_{t-1} - \alpha) + \omega_t.\end{aligned}$$

Resultados

Ver arquivo volatilidade_2.odc

Parâmetro	Média	D.P.	2,5%	Mediana	97,5%
α	-7,140	0,1226	-7,379	-7,1400	-6,899
ϕ	0,8498	0,0618	0,696	0,8595	0,942
σ^2	0,1120	0,0504	0,044	0,1022	0,239
γ	0,0015	0,0012	-0,001	0,0015	0,003
β	-0,0025	0,0189	-0,051	-0,0001	0,035
r_{558}	0,0001	0,0279	-0,057	0,0013	0,056
r_{559}	0,0012	0,0279	-0,053	0,0010	0,061
r_{560}	0,0007	0,0287	-0,056	0,0008	0,057
h_{558}	-7,327	0,5589	-8,404	-7,335	-6,235
h_{559}	-7,301	0,5803	-8,441	-7,307	-6,162
h_{560}	-7,287	0,5940	-8,464	-7,292	-6,106