

ANÁLISE DE SÉRIES TEMPORAIS

RICARDO S. EHLERS

Primeira publicação 2003

Segunda edição publicada em 2004

Terceira edição publicada em 2005

Quarta edição publicada em 2007

© RICARDO SANDES EHLERS 2003-2007

Sumário

1	Introdução	1
2	Técnicas Descritivas	6
2.1	Decomposição Clássica	6
2.2	Séries com Tendência	6
2.3	Séries Sazonais	11
2.4	Autocorrelação	11
2.4.1	O Correlograma	13
3	Modelos Probabilísticos	18
3.1	Introdução	18
3.2	Processos Estacionários	19
3.3	A Função de Autocorrelação	20
3.4	Alguns Processos Estocásticos	20
3.4.1	Sequência Aleatória	20
3.4.2	Passeio Aleatório	21
3.4.3	Processos de Média Móveis	22
3.4.4	Processos Autoregressivos	24
3.4.5	Modelos Mistos ARMA	29
3.4.6	Modelos ARMA Integrados	30
4	Estimação	33
4.1	Autocovariância e autocorrelação	34
4.2	Ajustando Processos Autoregressivos	35
4.3	Ajustando Processos Médias Móveis	39
4.4	Ajustando Processos ARMA	40
4.5	Modelos Sazonais	41
4.6	Adequação do Modelo	43
4.6.1	Análise dos Resíduos	43
4.6.2	Testes sobre os resíduos	44
5	Previsão	52
5.1	Métodos Univariados de Previsão	52
5.1.1	Alisamento Exponencial Simples	52

5.1.2	Método de Holt-Winters	56
5.2	Previsão em Modelos ARMA	58
5.3	Performance Preditiva	62
5.4	Critérios de Informação	64
5.5	Previsões Usando Todos os Modelos	67
5.6	Previsão Bayesiana	68
6	Modelando a Variância	73
6.1	Introdução	73
6.2	Modelos ARCH	74
6.3	Modelos GARCH	80
6.3.1	Estimação	81
6.3.2	Adequação	82
6.4	Volatilidade Estocástica	83
7	Modelos Lineares Dinâmicos	86
7.1	Introdução	86
7.2	Modelos Polinomiais	88
7.2.1	Análise Sequencial e Previsões	89
7.2.2	Variâncias de Evolução e das Observações	91
7.3	Modelo de Crescimento Linear	94
7.4	Modelos Sazonais	95
7.4.1	Modelos sem Crescimento	95
7.4.2	Modelos com Crescimento	96
7.5	Representação de Fourier	96
7.6	Ilustração	97
7.7	Modelos de Regressão	98
7.8	Monitoramento	99
A	Lista de Distribuições	105
A.1	Distribuição Normal	105
A.2	Distribuição Gama	105
A.3	Distribuição Wishart	106
A.4	Distribuição Gama Inversa	106
A.5	Distribuição Wishart Invertida	106
A.6	Distribuição Beta	106
A.7	Distribuição de Dirichlet	107
A.8	Distribuição t de Student	107
A.9	Distribuição F de Fisher	107
A.10	Distribuição Binomial	108
A.11	Distribuição Multinomial	108
A.12	Distribuição de Poisson	108
A.13	Distribuição Binomial Negativa	108

SUMÁRIO

iii

References

110

Capítulo 1

Introdução

Uma *série temporal* é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. Enquanto em modelos de regressão por exemplo a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial. Vale notar também que o tempo pode ser substituído por outra variável como espaço, profundidade, etc.

Como a maior parte dos procedimentos estatísticos foi desenvolvida para analisar observações independentes o estudo de séries temporais requer o uso de técnicas específicas. Dados de séries temporais surgem em vários campos do conhecimento como Economia (preços diários de ações, taxa mensal de desemprego, produção industrial), Medicina (eletrocardiograma, eletroencefalograma), Epidemiologia (número mensal de novos casos de meningite), Meteorologia (precipitação pluviométrica, temperatura diária, velocidade do vento), etc.

Algumas características são particulares a este tipo de dados, por exemplo,

- Observações correlacionadas são mais difíceis de analisar e requerem técnicas específicas.
- Precisamos levar em conta a *ordem temporal* das observações.
- Fatores complicadores como presença de tendências e variação sazonal ou cíclica podem ser difíceis de estimar ou remover.
- A seleção de modelos pode ser bastante complicada, e as ferramentas podem ser de difícil interpretação.
- É mais difícil de lidar com observações perdidas e dados discrepantes devido à natureza sequencial.

Terminologia

Uma série temporal é dita ser *contínua* quando as observações são feitas continuamente no tempo. Definindo o conjunto $T = \{t : t_1 < t < t_2\}$ a série temporal será denotada

por $\{X(t) : t \in T\}$. Uma série temporal é dita ser *discreta* quando as observações são feitas em tempos específicos, geralmente equiespaçados. Definindo o conjunto $T = \{t_1, \dots, t_n\}$ a série temporal será denotada por $\{X_t : t \in T\}$. Por simplicidade podemos fazer $T = \{1, 2, \dots, n\}$.

Note que estes termos não se referem à variável observada X , esta pode assumir valores discretos ou contínuos. Em muitas situações X pode ser discreta por definição (e.g. o número de casos notificados de AIDS) porém para efeito de análise estatística pode ser tratada como contínua se os seus valores observados não forem muito pequenos.

Por outro lado, séries temporais discretas podem surgir de várias formas. Séries contínuas podem ser *discretizadas*, i.e. seus valores são registrados a certos intervalos de tempo. Séries de valores agregados ou acumulados em intervalos de tempo, por exemplo exportações medidas mensalmente ou quantidade de chuva medida diariamente. Finalmente, algumas séries são inerentemente discretas, por exemplo dividendos pagos por uma empresa aos seus acionistas em anos sucessivos.

Uma série temporal também pode ser multivariada. Se k variáveis são observadas a cada tempo (por exemplo discreto) denota-se por $\{X_{1t}, \dots, X_{kt}, t \in T\}$. Neste caso várias séries correlacionadas devem ser analisadas conjuntamente, ou seja em cada tempo tem-se um vetor de observações.

Objetivos

Em algumas situações o objetivo pode ser fazer previsões de valores futuros enquanto em outras a estrutura da série ou sua relação com outras séries pode ser o interesse principal. De um modo geral, os principais objetivos em se estudar séries temporais podem ser os seguintes,

- Descrição. Descrever propriedades da série, e.g. o padrão de tendência, existência de variação sazonal ou cíclica, observações discrepantes (*outliers*), alterações estruturais (e.g. mudanças no padrão da tendência ou da sazonalidade), etc.
- Explicação. Usar a variação em uma série para explicar a variação em outra série.
- Predição: prever valores futuros com base em valores passados. Aqui assume-se que o futuro envolve incerteza, ou seja as previsões não são perfeitas. Porém devemos tentar reduzir os erros de previsão.
- Controle. Os valores da série temporal medem a “qualidade” de um processo de manufatura e o objetivo é o controle do processo. Um exemplo é o controle estatístico de qualidade aonde as observações são representadas em cartas de controle. Este tópico não será abordado nestas notas de aula.

Abordagens

- Técnicas Descritivas. Técnicas gráficos, identificação de padrões, etc.
- Modelos Probabilísticos. Seleção, comparação e adequação de modelos, estimação, predição. Ferramenta básica é a *função de autocorrelação*.
- Análise espectral.
- Métodos não paramétricos (alisamento ou suavização).
- Outras Abordagens. Modelos de espaço de estados, modelos não lineares, séries multivariadas, estudos longitudinais, processos de longa dependência, modelos para volatilidade, etc.

Sazonalidade

Muitas séries temporais exibem um comportamento que tende a se repetir a cada s períodos de tempo. Por exemplo, é natural esperar que as vendas mensais de brinquedos terão um pico no mês de dezembro e talvez um pico secundário em outubro. Este padrão possivelmente se repetirá ao longo de vários anos. Vejamos alguns possíveis modelos sazonais,

1. Sazonalidade determinística. Variáveis *dummies* (binárias). O coeficiente de cada variável *dummy* representa o fator sazonal do respectivo mês, trimestre, etc.
2. Funções trigonométricas.
3. Sazonalidade estocástica:
 - (a) Variável endógena com defasagem sazonal no modelo (modelos ARMA periódicos),
 - (b) modelo ARMA sazonal.

Tipos de Sazonalidade

- Aditiva. A série apresenta flutuações sazonais mais ou menos constantes não importando o nível global da série.
- Multiplicativa. O tamanho das flutuações sazonais varia dependendo do nível global da série.

No exemplo dos brinquedos, suponha que o aumento esperado nas vendas nos meses de dezembro é de 1 milhão de reais em relação à média anual. Então as previsões para os meses de dezembro dos próximos anos deve somar a quantia de 1 milhão de reais à uma média anual para levar em conta esta flutuação sazonal. Isto é o que se chama de *sazonalidade aditiva*.

Suponha agora que o aumento esperado nos meses de dezembro seja de 30%. Então o aumento esperado (em valor absoluto) de vendas em dezembro será pequeno ou grande dependendo da média anual de vendas ser baixa ou alta. Nas previsões para os próximos meses de dezembro deve-se multiplicar a média anual pelo fator 1,3. Isto é o que se chama de *sazonalidade multiplicativa*.

Tendência

Globalmente, uma série pode exibir tendência de crescimento (ou decrescimento) com vários possíveis padrões.

- Crescimento linear. Por exemplo, a cada ano o aumento esperado nas vendas de um certo brinquedo é de 1 milhão de reais.
- Crescimento exponencial. Por exemplo, a cada ano as vendas de um certo brinquedo aumentam de um fator 1,3.
- Crescimento amortecido. Por exemplo, as vendas de um certo brinquedo tem um aumento esperado de 70% sobre o ano anterior. Se o aumento esperado for de 1 milhão de reais no primeiro ano, no segundo ano será de 700 mil reais, no terceiro ano será de 490 mil reais e assim por diante.

Exemplos de Séries Temporais

Como primeira ilustração são apresentadas na Figura 1.1 quatro séries temporais disponíveis no pacote **R**. Nos eixos horizontais aparecem os anos de observação e nos eixos verticais os nomes das séries (mesmos nomes do **R**). A Figura 1.1a mostra totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960. Existe uma clara tendência de crescimento bem como um padrão sazonal ao longo dos anos. A Figura 1.1b mostra a série com o número anual de lincas capturados em armadilhas entre 1821 e 1934 no Canadá. Existe um padrão cíclico em torno de 10 ou 11 anos. A Figura 1.1c mostra a série com as medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970. Parece haver alguma alteração estrutural em torno do ano de 1900. Finalmente a Figura 1.1d mostra a série trimestral do consumo de gás no Reino Unido entre o primeiro trimestre de 1960 e o quarto trimestre de 1986. Há uma tendência de crescimento porém a amplitude do padrão sazonal aumenta bastante a partir de 1971.

Exercícios

1. Classifique as seguintes séries temporais quanto ao tempo e quanto a variável observada.
 - (a) Registros de maré durante 1 dia.
 - (b) Medidas de temperatura em uma estação meteorológica.

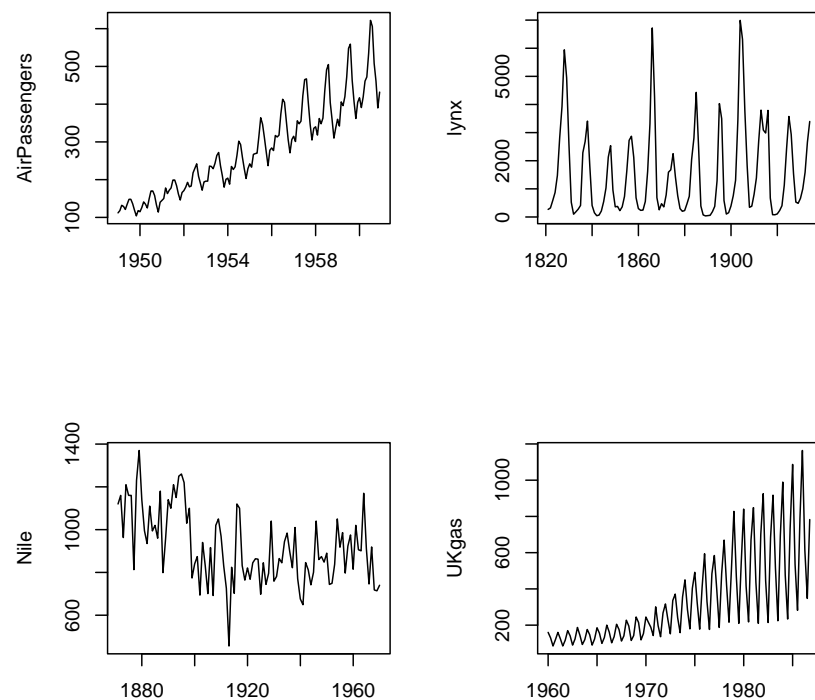


Figura 1.1: (a) Totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960, (b) número anual de linces capturados em armadilhas entre 1821 e 1934 no Canadá, (c) medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970, (d) consumo de gás no Reino Unido entre o primeiro trimestre de 1960 e o quarto trimestre de 1986.

- (c) O índice diário da bolsa de valores de São Paulo.
 - (d) A inflação mensal medida pelo índice de preços ao consumidor.
 - (e) Variação diária de um determinado índice financeiro, 1 para variação positiva, -1 para variação negativa ou zero se não ocorreu variação.
 - (f) Número mensal de novos casos de Dengue em uma determinada região.
2. Dê exemplos de séries temporais contínuas que poderiam ser discretizadas (e de que forma).

Capítulo 2

Técnicas Descritivas

Ao se analisar uma ou mais séries temporais a representação gráfica dos dados sequencialmente ao longo do tempo é fundamental e pode revelar padrões de comportamento importantes. Tendências de crescimento (ou decrescimento), padrões cíclicos, alterações estruturais, observações aberrantes, etc. são muitas vezes facilmente identificados. Sendo assim, o gráfico temporal deve ser sempre o primeiro passo e antecede qualquer análise. Outras ferramentas serão descritas ao longo deste capítulo.

2.1 Decomposição Clássica

Muitas das propriedades observadas em uma série temporal X_t podem ser captadas assumindo-se a seguinte forma de decomposição

$$X_t = T_t + C_t + R_t$$

onde T_t é uma componente de tendência, C_t é uma componente cíclica ou sazonal e R_t é uma componente aleatória ou ruído (a parte não explicada, que espera-se ser puramente aleatória). A componente cíclica se repete a cada intervalo fixo s , i.e.

$$\dots = C_{t-2s} = C_{t-s} = C_t = C_{t+s} = C_{t+2s} = \dots$$

Assim, variações periódicas podem ser captadas por esta componente.

2.2 Séries com Tendência

Não existe uma definição precisa de tendência e diferentes autores usam este termo de diferentes formas. Podemos pensar em tendência como uma mudança de longo prazo no nível médio da série. A dificuldade aqui é definir longo prazo.

A forma mais simples de tendência é

$$X_t = \alpha + \beta t + \epsilon_t \tag{2.1}$$

onde α e β são constantes a serem estimadas e ϵ_t denota um erro aleatório com média zero. O nível médio da série no tempo t é dado por $m_t = \alpha + \beta t$ que é algumas vezes

chamado de *termo de tendência*. Porém alguns autores preferem chamar a inclinação β de tendência, ou seja a mudança no nível da série por unidade de tempo já que $\beta = m_t - m_{t-1}$. Note que a tendência na equação (2.1) é uma função determinística do tempo e algumas vezes é chamada de *tendência global* (i.e. vale para toda a série), em oposição a *tendência local*.

De um modo geral, uma forma de se lidar com dados não sazonais que contenham uma tendência consiste em ajustar uma função polinomial,

$$X_t = \beta_0 + \beta_1 t + \cdots + \beta_p t^p + \epsilon_t.$$

Uma função linear ou quadrática seria apropriada no caso de uma tendência monotonicamente crescente ou decrescente. Caso contrário polinômios de ordem mais alta devem ser ajustados.

Outras possíveis formas de tendência são os crescimentos descritos por uma curva Gompertz,

$$\log x_t = a + br^t$$

onde a , b e r são parâmetros com $0 < r < 1$, ou uma curva Logística,

$$x_t = a/(1 + be^{-ct})$$

onde a , b e c são parâmetros. Estas duas últimas são chamadas curvas S e se aproximam de uma assíntota quando $t \rightarrow \infty$. Neste caso o ajuste pode levar a equações não lineares.

Seja qual for a curva utilizada, a função ajustada fornece uma medida da tendência da série, enquanto os resíduos (valores observados – valores ajustados) fornecem uma estimativa de flutuações locais.

Exemplo 2.1: A Figura 2.1 mostra as medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970 juntamente com polinômios de graus 3 e 6 superimpostos. Os polinômios foram ajustados por mínimos quadrados usando os comandos do R a seguir. A série original com as tendências estimadas aparecem na Figura (2.1).

```
> mypolytrend = function(y, degree = 1) {
+   n = length(y)
+   x = 1:n
+   X = matrix(NA, n, degree)
+   for (i in 1:degree) X[, i] = x^i
+   a = as.numeric(lm(y ~ X)$coeff)
+   out = cbind(rep(1, n), X) %*% a
+   return(ts(out, start = start(y), freq = frequency(y)))
+ }
> z3 = mypolytrend(Nile, 3)
> z6 = mypolytrend(Nile, 6)
```

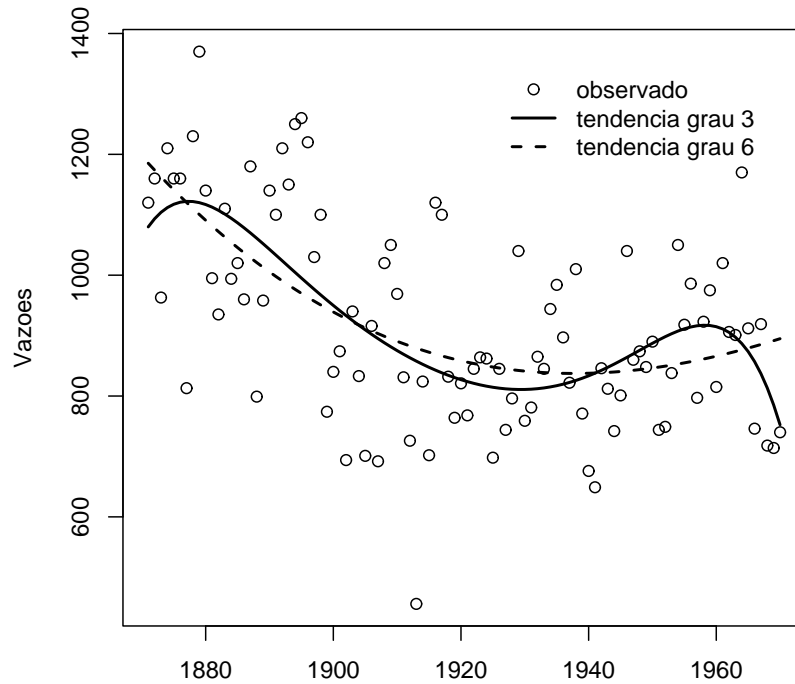


Figura 2.1: Medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970 (pontos), com polinômios de graus 3 e 6 ajustados por mínimos quadrados.

Regressão Local

A idéia aqui é estimar para cada t uma equação de regressão polinomial diferente, por exemplo

$$\hat{x}_t = \hat{\alpha}(t) + \hat{\beta}(t)t.$$

Note que as estimativas de α e β dependem do tempo o que dá o caráter local das retas de regressão.

O procedimento conhecido como *loess* é um procedimento iterativo que a cada passo aplica a regressão local anterior, calcula os resíduos $x_t - \hat{x}_t$ e aplica novamente a regressão local dando peso menor às observações com resíduos maiores. Este procedimento se repete até atingir convergência.

Exemplo 2.2: A Figura 2.2 apresenta os mesmos dados da Figura 2.1 sendo que as curvas superimpostas foram obtidas usando regressão local com os comandos do R a seguir.

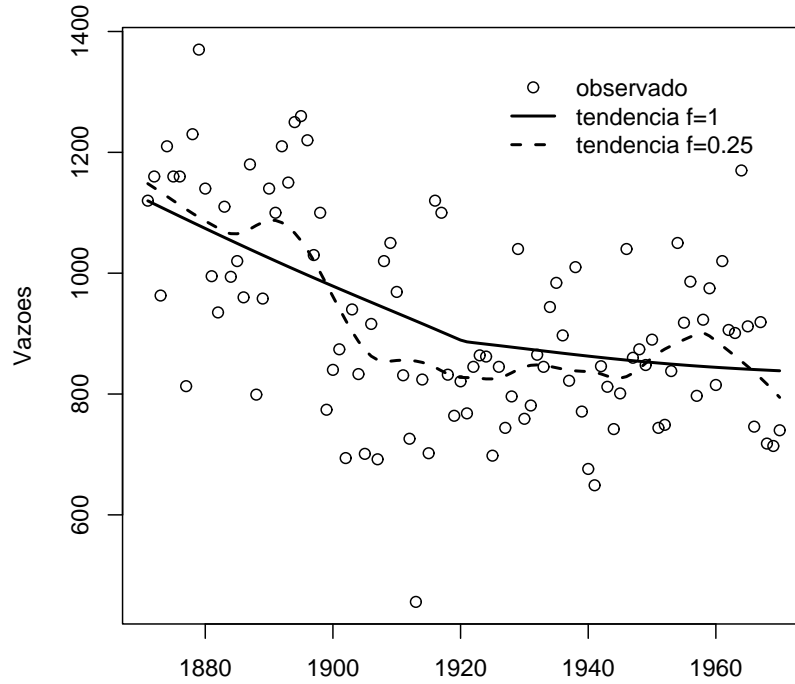


Figura 2.2: Medições anuais de vazões do Rio Nilo em Ashwan entre 1871 e 1970 (pontos), tendência estimada via função *lowess*.

Filtragem

Outro procedimento para analisar séries com tendência é através de *filtros lineares*. Um filtro linear converte uma série $\{x_t\}$ em outra $\{y_t\}$ através da seguinte operação linear

$$y_t = \sum_{j=-q}^s a_j x_{t+j}$$

onde $\{a_j\}$ é um conjunto de pesos. Além disso, como queremos estimar a média local os pesos devem ser tais que $\sum_{j=-q}^s a_j = 1$, garantindo assim que $\min\{x_t\} < y_t < \max\{x_t\}$. Neste caso a operação é chamada *média móvel*.

Médias móveis são em geral simétricas com $s = q$ e $a_{-r} = a_r$. Por exemplo, se $s = q = 2$ temos que

$$y_t = a_2 x_{t-2} + a_1 x_{t-1} + a_0 x_t + a_1 x_{t+1} + a_2 x_{t+2}.$$

O caso mais simples é quando todos os pesos a_j tem o mesmo valor e devido à restrição de soma 1 segue que $a_j = 1/(2q+1)$, para $j = -q, \dots, q$. Neste caso, o valor suavizado

de x_t é dado por

$$y_t = \frac{1}{2q+1} \sum_{j=-q}^q x_{t+j}.$$

Qualquer que seja o filtro utilizado, y_t é uma estimativa da tendência no tempo t e $x_t - y_t$ é uma série livre de tendência.

Exemplo 2.3 : A Figura 2.3 apresenta a série com os totais mensais de passageiros de linhas aéreas internacionais nos EUA, entre 1949 e 1960 (Box, Jenkins and Reinsel, 1976) juntamente com a tendência “estimada” superimposta. Foram aplicados filtros lineares com médias móveis aproximadamente trimestrais ($q = 2$) e médias móveis aproximadamente anuais ($q = 5$). Os seguintes comandos do R foram usados.

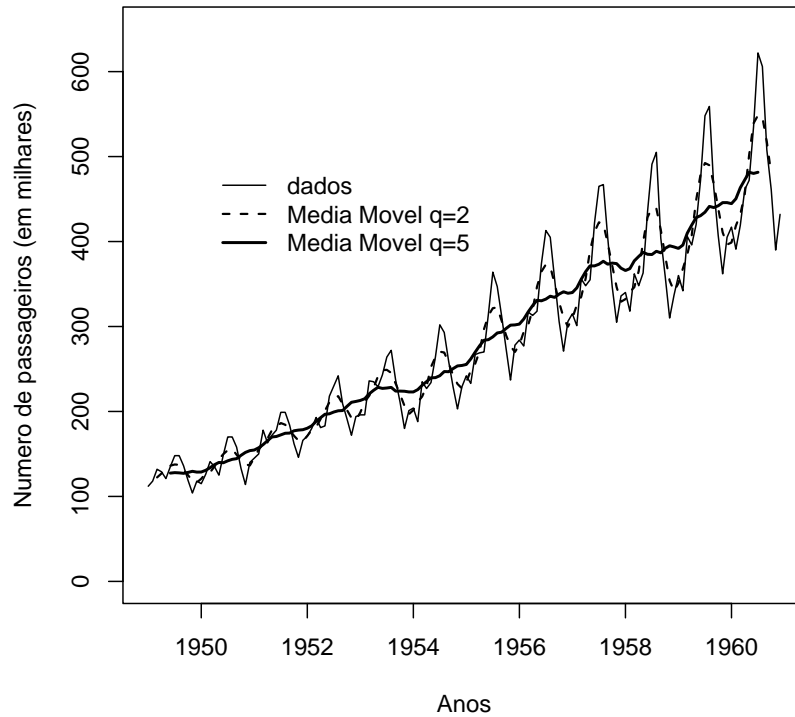


Figura 2.3: Totais mensais de passageiros de linhas aéreas internacionais nos EUA, com a tendência superimposta aplicando médias móveis aproximadamente trimestrais ($q = 2$) e médias móveis aproximadamente anuais ($q = 5$).

Note que, para a aplicação de qualquer filtro simétrico os valores suavizados só podem ser calculados para $t = q + 1, \dots, n - q$ e assim a série suavizada terá $n - 2q$ valores. Em algumas situações no entanto é importante obter valores suavizados até o período $t = n$ e uma alternativa é utilizar um filtro assimétrico que usa apenas os valores atual e passados de x_t . Por exemplo na técnica conhecida como *alisamento*

exponencial os valores suavizados são dados por

$$y_t = \sum_{j=0}^{\infty} \alpha(1-\alpha)^j x_{t-j}$$

onde $0 < \alpha < 1$. Note como, embora todas as observações passadas sejam usadas no filtro, os pesos $\alpha(1-\alpha)^j$ decaem geometricamente com j . Quanto mais próximo de 1 estiver α mais peso será dado às observações mais recentes e quanto mais próximo de zero mais os pesos estarão distribuídos ao longo da série. Por exemplo se $\alpha = 0,90$ a série filtrada fica $y_t = 0,9x_t + 0,09x_{t-1} + 0,009x_{t-2} + \dots$ enquanto que para $\alpha = 0,1$ temos que $y_t = 0,1x_t + 0,09x_{t-1} + 0,081x_{t-2} + \dots$

Este tipo de filtro pode ser utilizado para fazer previsões. Especificamente a previsão da série original em $t+1$ será o valor filtrado y_t (mais detalhes no Capítulo 5).

Diferenciação

Um tipo especial de filtro, muito útil para remover uma componente de tendência polinomial, consiste em *diferenciar* a série até que ela se torne *estacionária* (este conceito será formalizado no Capítulo 3). Para dados não sazonais, a primeira diferença é em geral suficiente para induzir estacionariedade aproximada. A nova série y_2, \dots, y_n é formada a partir da série original x_1, \dots, x_n como

$$y_t = x_t - x_{t-1} = \nabla x_t.$$

Note que isto nada mais é do que um filtro (assimétrico) com coeficientes 1 e -1.

Diferenciação de primeira ordem é a mais utilizada sendo que ocasionalmente uma diferenciação de segunda ordem pode ser requerida, i.e.

$$y_t = \nabla^2 x_t = \nabla(x_t - x_{t-1}) = x_t - 2x_{t-1} + x_{t-2}.$$

Além disso, independente do seu uso para induzir estacionariedade, a diferenciação pode ser muito útil como ferramenta exploratória. Observações discrepantes por exemplo podem ter um efeito dramático na série diferenciada e uma representação gráfica é em geral suficiente para identificar tais pontos.

2.3 Séries Sazonais

Uma forma bastante simples de eliminar o efeito sazonal é simplesmente tomar médias sazonais. Por exemplo, em dados mensais com sazonalidade anual, as médias anuais estarão livres do efeito sazonal. Embora este procedimento esteja correto muitos dados serão perdidos e ao invés disto pode-se recorrer mais uma vez às médias móveis.

2.4 Autocorrelação

Uma importante ferramenta para se identificar as propriedades de uma série temporal consiste de uma série de quantidades chamadas *coeficientes de autocorrelação*

amostral. A idéia é similar ao coeficiente de correlação usual, i.e. para n pares de observações das variáveis x e y o coeficiente de correlação amostral é dado por

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2.2)$$

Aqui no entanto queremos medir a correlação entre as observações de uma mesma variável em diferentes horizontes de tempo, i.e. correlações entre observações defasadas 1, 2, ... períodos de tempo. Assim, dadas n observações x_1, \dots, x_n de uma série temporal discreta podemos formar os pares $(x_1, x_2), \dots, (x_{n-1}, x_n)$. Considerando x_1, \dots, x_{n-1} e x_2, \dots, x_n como duas variáveis o coeficiente de correlação entre x_t e x_{t+1} é dado por

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)(x_{t+1} - \bar{x}_2)}{\sqrt{\sum_{t=1}^{n-1} (x_t - \bar{x}_1)^2 \sum_{t=1}^{n-1} (x_{t+1} - \bar{x}_2)^2}} \quad (2.3)$$

onde as médias amostrais são

$$\bar{x}_1 = \sum_{t=1}^{n-1} x_t / (n-1) \quad \text{e} \quad \bar{x}_2 = \sum_{t=2}^n x_t / (n-1).$$

Como o coeficiente r_1 mede as correlações entre observações sucessivas ele é chamado de coeficiente de autocorrelação ou coeficiente de correlação serial.

É usual simplificar a equação (2.3) utilizando-se a média de todas as observações, i.e. $\bar{x} = \sum_{t=1}^n x_t / n$ já que $\bar{x}_1 \approx \bar{x}_2$, e assumindo variância constante. Assim, a versão simplificada de (2.3) fica

$$r_1 = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(n-1) \sum_{t=1}^n (x_t - \bar{x})^2 / n} \quad (2.4)$$

sendo que alguns autores ainda retiram o termo $n/(n-1)$ que é próximo de 1 para n não muito pequeno. Esta última forma simplificada, sem o termo $n/(n-1)$ será utilizada neste texto.

A equação (2.4) pode ser generalizada para calcular a correlação entre observações defasadas de k períodos de tempo, i.e.

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (2.5)$$

fornece o coeficiente de correlação de ordem k . Assim como o coeficiente de correlação usual, as autocorrelações são adimensionais e $-1 < r_k < 1$.

Na prática é mais usual calcular primeiro os coeficientes de *autocovariância* $\{c_k\}$, definidos por analogia com a fórmula usual de covariância, i.e.

$$c_k = \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})/n.$$

Os coeficientes de autocorrelação são então obtidos como $r_k = c_k/c_0$.

2.4.1 O Correlograma

Um gráfico com os k primeiros coeficientes de autocorrelação como função de k é chamado de correlograma e pode ser uma ferramenta poderosa para identificar características da série temporal. Porém isto requer uma interpretação adequada do correlograma, i.e. devemos associar certos padrões do correlograma como determinadas características de uma série temporal. Esta nem sempre é uma tarefa simples e a seguir são dadas algumas indicações.

Séries aleatórias

A primeira questão que podemos tentar responder através do correlograma é se uma série temporal é aleatória ou não. Para uma série completamente aleatória os valores defasados são não correlacionados e portanto espera-se que $r_k \approx 0$, $k = 1, 2, \dots$.

Suponha que x_1, \dots, x_n sejam variáveis aleatórias independentes e identicamente distribuídas com média arbitrárias. Então, pode-se mostrar que o coeficiente de autocorrelação amostral r_k é assintoticamente normalmente distribuído, com média e variância dados por

$$E(r_k) \approx -1/n \quad \text{e} \quad \text{Var}(r_k) \approx 1/n.$$

(ver Kendall, Stuart, & Ord 1983, Capítulo 48). Portanto, limites de confiança aproximados de 95% são dados por $-1/n \pm 1,96/\sqrt{n}$, que são frequentemente ainda mais aproximados para $\pm 1,96/\sqrt{n}$.

Isto ilustra uma das dificuldades de interpretar o correlograma já que, mesmo para uma série completamente aleatória, espera-se que 1 em cada 20 coeficientes r_k esteja fora destes limites. Por outro lado, um valor muito grande de r_k tem menos chance de ter ocorrido ao acaso do que um valor que está apenas ligeiramente fora dos limites.

A Figura 2.4 mostra uma série temporal com 100 observações independentes e identicamente distribuídas geradas no computador juntamente com o seu correlograma. Neste caso os limites de confiança de 95% são aproximadamente $\pm 2/\sqrt{100} = \pm 0,2$ e podemos notar que 2 dentre as 20 primeiras autocorrelações estão ligeiramente fora destes limites. No entanto isto ocorre em defasagens aparentemente arbitrárias (12 e 18) e podemos concluir que não há evidência para rejeitar a hipótese de que as observações são independentes.

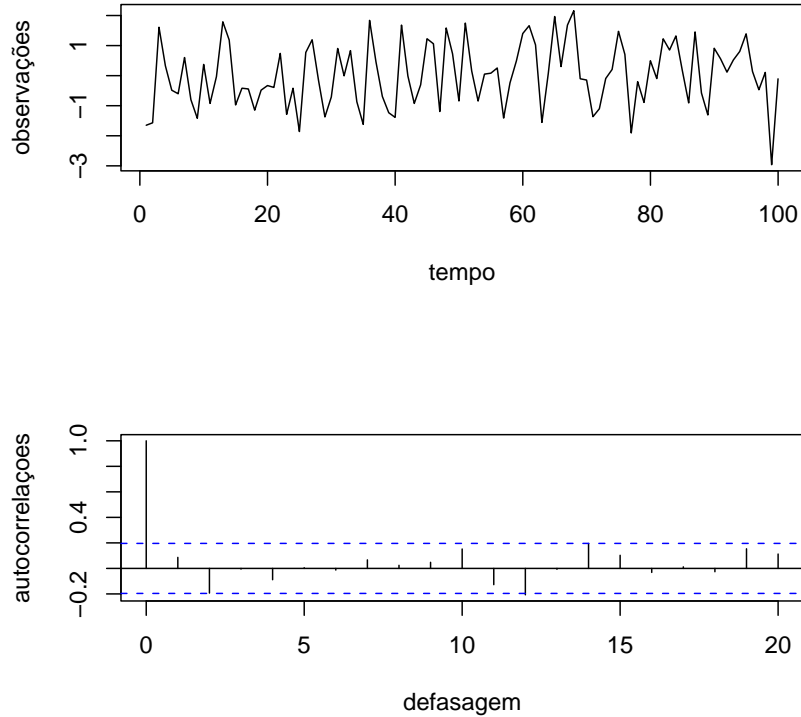


Figura 2.4: (a) 100 observações simuladas independentes e identicamente distribuídas. (b) 20 primeiras autocorrelações amostrais.

Correlação de curto-prazo

Uma série temporal na qual uma observação acima da média “tende” a ser seguida por uma ou mais observações acima da média, similarmente para observações abaixo da média, é dita ter correlação de curto-prazo. Um correlograma desta série deverá exibir um valor relativamente grande de r_1 seguido por valores que tendem a ficar sucessivamente menores. A partir de uma certa defasagem k os valores de r_k tendem a ser aproximadamente zero. Na Figura 2.5 temos 50 observações geradas de acordo com o processo $x_t = 0,7x_{t-1} + \epsilon_t$ juntamente com o seu correlograma.

Correlação negativa

Se os valores de uma série temporal tendem a se alternar acima e abaixo de um valor médio, o correlograma desta série também tende a se alternar. O valor de r_1 será negativo enquanto o valor de r_2 será positivo já que as observações defasadas de 2 períodos tendem a estar do mesmo lado da média. Esta característica está ilustrada na Figura 2.6 aonde temos 50 observações simuladas com autocorrelações negativas juntamente com as 14 primeiras autocorrelações amostrais.

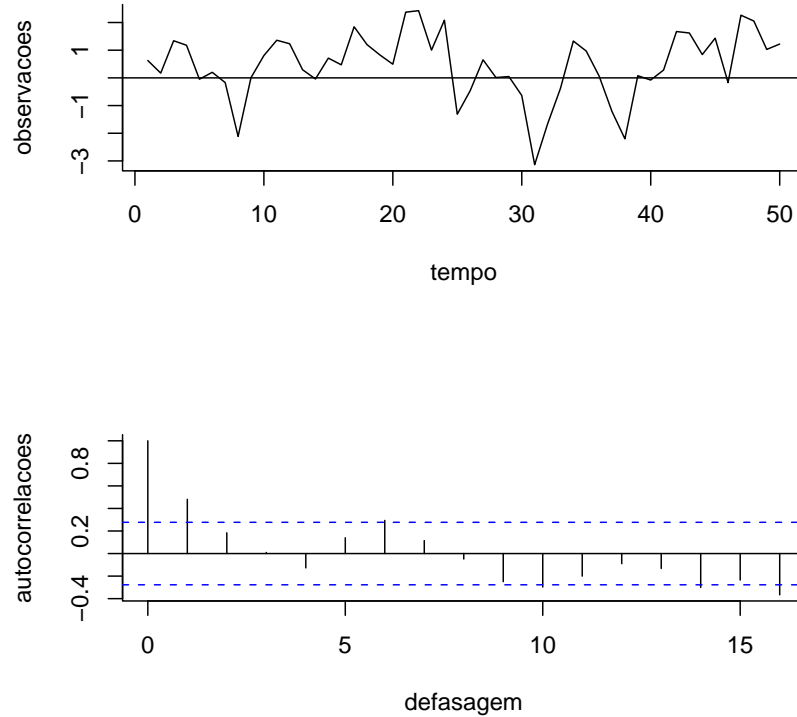


Figura 2.5: (a) 50 observações simuladas com autocorrelações de curto-prazo. (b) 16 primeiras autocorrelações amostrais.

Séries não estacionárias

Para uma série temporal com tendência os valores de r_k não decairão para zero a não ser em defasagens grandes. Intuitivamente, isto ocorre porque uma observação de um lado da média tende a ser seguida por um grande número de observações do mesmo lado (devido à tendência). Neste caso, pouca ou nenhuma informação pode ser extraída do correlograma já que a tendência dominará outras características. Na verdade, como veremos em outros capítulos a função de autocorrelação só tem um significado para séries estacionárias, sendo assim qualquer tendência deve ser removida antes do cálculo de $\{r_k\}$.

A Figura 2.7 mostra uma série temporal com 50 observações geradas segundo o modelo $x_t = x_{t-1} + \epsilon_t$, juntamente com o seu correlograma. Note que a não estacionariedade da série fica evidenciada no correlograma já que as autocorrelações amostrais decaem muito lentamente.

Variação sazonal

Um padrão sazonal é em geral facilmente identificado no correlograma. De fato, se uma série temporal contém flutuações sazonais o correlograma irá exibir oscilações na

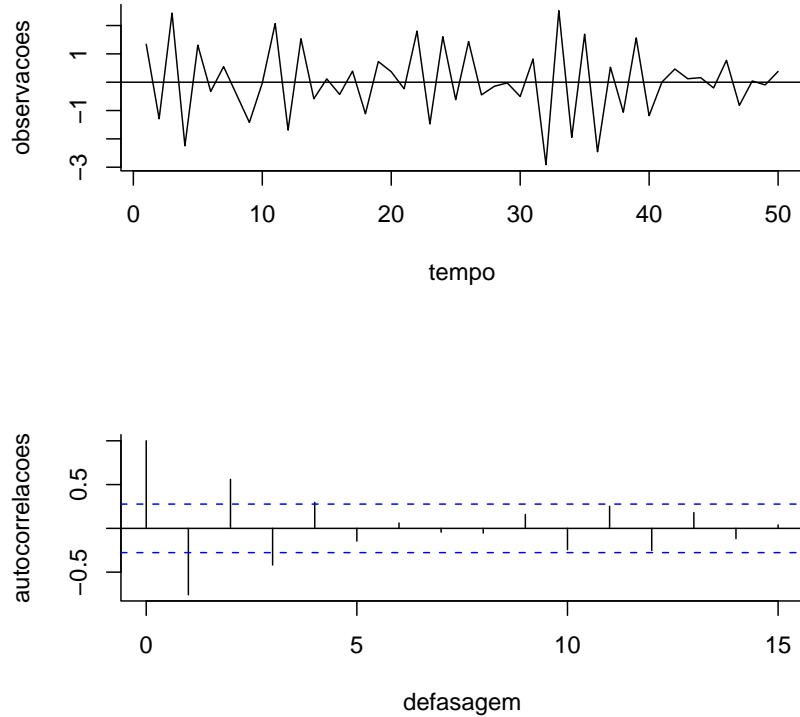


Figura 2.6: (a) 50 observações simuladas com autocorrelações negativas. (b) 15 primeiras autocorrelações amostrais.

mesma frequência. Por exemplo, com observações mensais r_6 será grande e negativo enquanto r_{12} será grande e positivo.

Na verdade, se o padrão sazonal já for evidente no gráfico da série original o correlograma trará pouca ou nenhuma informação adicional.

Observações discrepantes

Se uma série temporal contem uma ou mais observações discrepantes (“outliers”) o correlograma pode ser seriamente afetado. No caso de uma única observação discrepante o gráfico de x_t contra x_{t+k} terá pontos extremos o que pode viesar os coeficientes de correlação para zero. Com dois valores discrepantes o efeito pode ser ainda mais devastador, além de gerar uma correlação espúria quando k é igual à distância entre os valores.

Exercícios

1. Use o R para gerar uma série temporal $Y_t = b_0 + b_1t + \epsilon_t$, $t = 1, \dots, 100$, com $b_0, b_1 \neq 0$ e ϵ_t normais e independentes com média μ e variância σ_1^2 se $t \leq 70$ mas variância $\sigma_2^2 \neq \sigma_1^2$ se $t > 70$. Usando diferentes valores de α aplique o

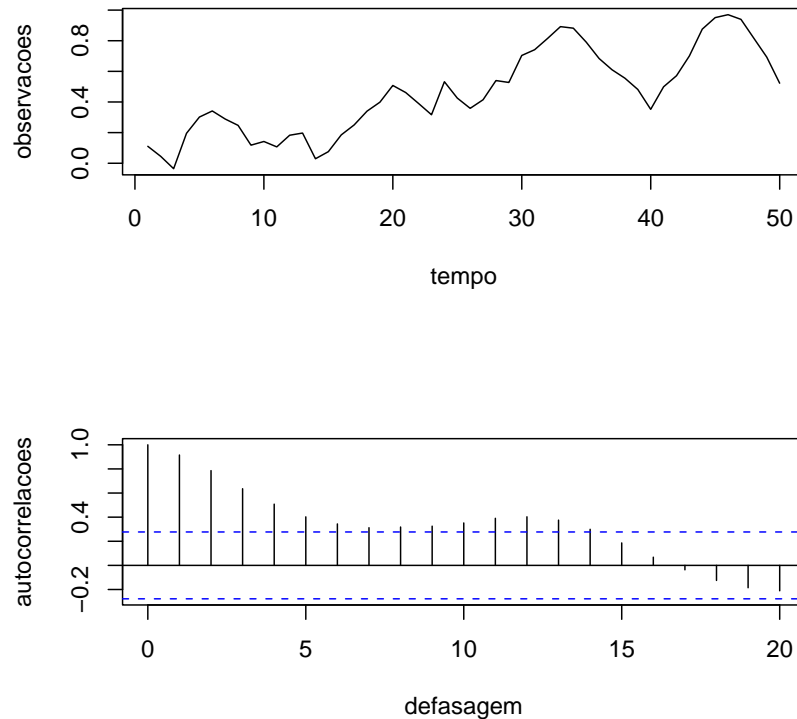


Figura 2.7: (a) 50 observações simuladas segundo um passeio aleatório. (b) 20 primeiras autocorrelações amostrais.

alisamento exponencial e faça um gráfico da série com os valores suavizados. Comente os resultados.

2. Para cada um dos processos abaixo gere 200 observações. Faça um gráfico da série e do correlograma.

- (a) Serie aleatória, observações iid da distribuição $N(0,1)$.
- (b) Serie com tendência estocástica, $x_t = x_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0, (0,1)^2)$
- (c) Outra serie com tendencia estocastica, $x_t = x_{t-1} + \epsilon_t$, $\epsilon_t \sim N(1, 5^2)$
- (d) Serie com correlação de curto-prazo, $x_t = 0,7x_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$
- (e) Serie com correlações negativas, $x_t = -0,8x_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$
- (f) Medias moveis, $x_t = \epsilon_t + 0,6\epsilon_{t-1}$, $\epsilon_t \sim N(0,1)$
- (g) passeio aleatorio com desvio $X_t = 1 + X_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$.

Capítulo 3

Modelos Probabilísticos

3.1 Introdução

Neste capítulo serão descritos vários modelos adequados para dados de séries temporais. Tais modelos são chamados de *processos estocásticos*.

Matematicamente um processo estocástico pode ser definido como uma coleção de variáveis aleatórias ordenadas no tempo e definidas em um conjunto de pontos T , que pode ser contínuo ou discreto. Iremos denotar a variável aleatória no tempo t por $X(t)$ no caso contínuo (usualmente $-\infty < t < \infty$), e por X_t no caso discreto (usualmente $t = 0, \pm 1, \pm 2, \dots$). O conjunto de possíveis valores do processo é chamado de *espaço de estados* que pode ser discreto (e.g. o número de chamadas que chegam a uma central telefônica a cada 2 horas) ou contínuo (e.g. a temperatura do ar em uma localidade observada em intervalos de 1 hora).

Em análise de séries temporais a situação é bem diferente da maioria dos problemas estatísticos. Embora seja possível variar o tamanho da série observada, usualmente será impossível fazer mais do que uma observação em cada tempo. Assim, tem-se apenas uma realização do processo estocástico e uma única observação da variável aleatória no tempo t denotada por $x(t)$ no caso contínuo e x_t , para $t = 1, \dots, N$ no caso discreto.

Uma maneira de descrever um processo estocástico é através da distribuição de probabilidade conjunta de $X(t_1), \dots, X(t_k)$ para qualquer conjunto de tempos t_1, \dots, t_k e qualquer valor de k . Esta é uma tarefa extremamente complicada e na prática costuma-se descrever um processo estocástico através das funções média, variância e autocovariância. Estas funções são definidas a seguir para o caso contínuo sendo que definições similares se aplicam ao caso discreto.

$$\begin{aligned} \text{média} \quad \mu(t) &= E[X(t)] \\ \text{variância} \quad \sigma^2(t) &= \text{Var}[X(t)] \\ \text{autocovariância} \quad \gamma(t_1, t_2) &= E[X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)] \end{aligned}$$

Note que a função de variância é um caso especial da função de autocovariância quando $t_1 = t_2$. Momentos de ordem mais alta do processo também ser definidos

mas são raramente utilizados na prática e as funções $\mu(t)$ e $\gamma(t_1, t_2)$ são em geral suficientes.

3.2 Processos Estacionários

Uma importante classe de processos estocásticos são os chamados processos estacionários. A idéia intuitiva de estacionariedade foi introduzida no capítulo anterior e aqui será apresentada a definição formal.

Uma série temporal é dita *estritamente estacionária* se a distribuição de probabilidade conjunta de $X(t_1), \dots, X(t_k)$ é a mesma de $X(t_1 + \tau), \dots, X(t_k + \tau)$. Ou seja, o deslocamento da origem dos tempos por uma quantidade τ não tem efeito na distribuição conjunta que portanto depende apenas dos intervalos entre t_1, \dots, t_k .

Em particular, para $k = 1$ a estacionariedade estrita implica que a distribuição de $X(t)$ é a mesma para todo t de modo que, se os dois primeiros momentos forem finitos, temos que

$$\mu(t) = \mu \quad \text{e} \quad \sigma^2(t) = \sigma^2$$

são constantes que não dependem de t .

Para $k = 2$ a distribuição conjunta de $X(t_1)$ e $X(t_2)$ depende apenas da distância $t_2 - t_1$, chamada *defasagem*. A função de autocovariância $\gamma(t_1, t_2)$ também depende apenas de $t_2 - t_1$ e pode ser escrita como $\gamma(\tau)$ onde

$$\gamma(\tau) = E[X(t) - \mu][X(t + \tau) - \mu] = \text{Cov}[X(t), X(t + \tau)]$$

é chamado de coeficiente de autocovariância na defasagem τ .

Note que o tamanho de $\gamma(\tau)$ depende da escala em que $X(t)$ é medida. Portanto, para efeito de interpretação, é mais útil padronizar a função de autocovariância dando origem a uma função de autocorrelação

$$\rho(\tau) = \gamma(\tau)/\gamma(0)$$

que mede a correlação entre $X(t)$ e $X(t + \tau)$. No capítulo anterior foi apresentado o seu equivalente empírico para séries discretas r_k . Note também que o argumento τ será discreto se a série temporal for discreta e contínuo se a série temporal for contínua.

Na prática é muito difícil usar a definição de estacionariedade estrita e costuma-se definir estacionariedade de uma forma menos restrita.

Definição 3.1. *Um processo estocástico $\{X(t), t \in T\}$ é dito ser estacionário de segunda ordem ou fracamente estacionário se a sua função média é constante e sua função de autocovariância depende apenas da defasagem, i.e.*

$$E[X(t)] = \mu \quad \text{e} \quad \text{Cov}[X(t), X(t + \tau)] = \gamma(\tau).$$

Nenhuma outra suposição é feita a respeito dos momentos de ordem mais alta. Além disso, fazendo $\tau = 0$ segue que $\text{Var}[X(t)] = \gamma(0)$, ou seja a variância do processo

assim como a média também é constante. Note também que tanto a média quanto a variância precisam ser finitos.

Esta definição mais fraca de estacionariedade será utilizada daqui em diante já que muitas propriedades dos processos estacionários dependem apenas da estrutura especificada pelo primeiro e segundo momentos. Uma classe importante de processos aonde isto se verifica é a classe de processos *normais* ou *Gaussianos* aonde a distribuição conjunta de $X(t_1), \dots, X(t_k)$ é normal multivariada para todo conjunto t_1, \dots, t_k . A distribuição normal multivariada fica completamente caracterizada pelo primeiro e segundo momentos, i.e. por $\mu(t)$ e $\gamma(t_1, t_2)$, assim estacionariedade fraca implica em estacionariedade estrita para processos normais. Por outro lado, μ e $\gamma(\tau)$ podem não descrever adequadamente processos que sejam muito “não-normais”.

3.3 A Função de Autocorrelação

Foi visto na Seção 2.4 que os coeficientes de autocorrelação amostral de uma série temporal observada são uma ferramenta importante para descrever a série. Analogamente, a função de autocorrelação teórica (fac) de um processo estocástico estacionário é uma ferramenta importante para avaliar suas propriedades. A seguir serão apresentadas propriedades gerais da função de autocorrelação.

Se um processo estocástico estacionário $X(t)$ tem média μ e variância σ^2 então

$$\rho(\tau) = \gamma(\tau)/\gamma(0) = \gamma(\tau)/\sigma^2$$

e portanto $\rho(0) = 1$. As seguintes propriedades são facilmente verificáveis.

1. A correlação entre $X(t)$ e $X(t + \tau)$ é a mesma que entre $X(t)$ e $X(t - \tau)$, ou seja $\rho(\tau) = \rho(-\tau)$.
2. $-1 < \rho(\tau) < 1$.
3. Embora um processo estocástico tenha uma estrutura de autocovariância única o contrário não é verdadeiro em geral. É possível encontrar vários processos com a mesma função de autocorrelação, o que dificulta ainda mais a interpretação do correlograma.

3.4 Alguns Processos Estocásticos

Nesta seção serão apresentados alguns processos estocásticos que são utilizados com frequência na especificação de modelos para séries temporais.

3.4.1 Sequência Aleatória

Um processo em tempo discreto é chamado *puramente aleatório* se consiste de uma sequência de variáveis aleatórias $\{\epsilon_t\}$ independentes e identicamente distribuídas. Isto implica nas seguintes propriedades

1. $E(\epsilon_t) = E(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = \mu$
2. $Var(\epsilon_t) = Var(\epsilon_t | \epsilon_{t-1}, \epsilon_{t-2}, \dots) = \sigma_\epsilon^2$
3. $\gamma(k) = Cov(\epsilon_t, \epsilon_{t+k}) = 0, \quad k = \pm 1, \pm 2, \dots$

Como a média e a função de autocovariância não dependem do tempo o processo é estacionário em segunda ordem. A função de autocorrelação é simplesmente

$$\rho(k) = \begin{cases} 1, & k = 0 \\ 0, & k = \pm 1, \pm 2, \dots \end{cases}$$

Um processo puramente aleatório é às vezes chamado de *ruído branco* e pode ser útil por exemplo na construção de processos mais complicados. As propriedades acima podem ser entendidas como ausência de *correlação serial* e *homocedasticidade condicional* (variância condicional constante).

3.4.2 Passeio Aleatório

Seja $\{\epsilon_t\}$ um processo discreto puramente aleatório com média μ e variância σ_ϵ^2 . Um processo $\{X_t\}$ é chamada de *passeio aleatório* se

$$X_t = X_{t-1} + \epsilon_t.$$

Fazendo-se substituições sucessivas obtém-se que

$$\begin{aligned} X_t &= X_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &= X_{t-3} + \epsilon_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &\vdots \\ &= X_0 + \sum_{j=1}^t \epsilon_j \end{aligned}$$

e iniciando o processo em $X_0 = 0$ não é difícil verificar que

$$\begin{aligned} E(X_t) &= \sum_{j=1}^t E(\epsilon_j) = t\mu \\ Var(X_t) &= \sum_{j=1}^t Var(\epsilon_j) = t\sigma_\epsilon^2. \end{aligned}$$

Além disso, a função de autocovariância é dada por

$$Cov(X_t, X_{t-k}) = Cov(\epsilon_1 + \dots + \epsilon_{t-k} + \dots + \epsilon_t, \epsilon_1 + \dots + \epsilon_{t-k}) = (t-k)\sigma_\epsilon^2$$

e portanto a função de autocorrelação fica

$$\rho_t(k) = \frac{t-k}{t}.$$

Como a média, a variância e as autocovariâncias dependem de t este processo é não estacionário. No entanto, é interessante notar que a primeira diferença de um passeio aleatório é estacionária já que

$$\nabla X_t = X_t - X_{t-1} = \epsilon_t.$$

Os exemplos mais conhecidos de séries temporais que se comportam como um passeio aleatório são os preços de ações em dias sucessivos (ver por exemplo Morettin e Toloi, 2004).

3.4.3 Processos de Média Móveis

Seja $\{\epsilon_t\}$ um processo discreto puramente aleatório com média zero e variância σ_ϵ^2 . Um processo $\{X_t\}$ é chamada de processo de médias móveis de ordem q , ou MA(q), se

$$X_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}, \quad (3.1)$$

sendo $\beta_i \in \mathbb{R}$, $i = 1, \dots, q$. Não é difícil verificar como ficam a média e a variância deste processo,

$$\begin{aligned} E(X_t) &= E(\epsilon_t) + \sum_{j=1}^q \beta_j E(\epsilon_{t-j}) = 0 \\ \text{Var}(X_t) &= \text{Var}(\epsilon_t) + \sum_{j=1}^q \beta_j^2 \text{Var}(\epsilon_{t-j}) = (1 + \beta_1^2 + \cdots + \beta_q^2) \sigma_\epsilon^2. \end{aligned}$$

Além disso, como $\text{Cov}(\epsilon_t, \epsilon_s) = \sigma_\epsilon^2$ para $t = s$ e $\text{Cov}(\epsilon_t, \epsilon_s) = 0$ para $t \neq s$, a função de autocovariância é dada por

$$\begin{aligned} \gamma(k) &= \text{Cov}(X_t, X_{t+k}) \\ &= \text{Cov}(\epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}, \epsilon_{t+k} + \beta_1 \epsilon_{t+k-1} + \cdots + \beta_q \epsilon_{t+k-q}) \\ &= \begin{cases} 0 & k > q \\ \sigma_\epsilon^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} & k = 0, \dots, q \\ \gamma(-k) & k < 0 \end{cases} \quad (3.2) \end{aligned}$$

com $\beta_0 = 1$. Como a média e a variância são constantes e $\gamma(k)$ não depende de t o processo é (fracamente) estacionário para todos os possíveis valores de β_1, \dots, β_q . Além disso, se os ϵ_t 's forem normalmente distribuídos os X_t 's também serão e portanto o processo será estritamente estacionário.

A função de autocorrelação pode ser facilmente obtida de (3.2) como

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \sum_{j=0}^{q-k} \beta_j \beta_{j+k} / \sum_{j=0}^q \beta_j^2 & k = 1, \dots, q \\ 0 & k > q \\ \rho(-k) & k < 0. \end{cases}$$

Note que a função tem um ponto de corte na defasagem q , i.e. $\rho(k) = 0$ para $k > q$. Esta é uma característica específica de processos médias móveis e será útil na especificação do valor de q na prática (Box & Jenkins 1970, p. 170).

```
> MAacf <- function(q, beta, lag.max) {
+   sig2x = 1 + sum(beta^2)
+   rho = rep(0, lag.max)
+   for (k in 1:q) {
+     rho[k] = beta[k]
+     if (q - k > 0) {
+       for (j in 1:(q - k)) rho[k] = rho[k] + beta[j] *
+         beta[j + k]
+     }
+     rho[k] = rho[k]/sig2x
+   }
+   return(rho)
+ }
> round(MAacf(q = 2, beta = c(0.5, 0.3), lag.max = 6), 4)

[1] 0.4851 0.2239 0.0000 0.0000 0.0000 0.0000
```

Vamos analisar agora com mais detalhes o caso particular do processo MA(1). A função de autocorrelação fica

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \beta_1/(1 + \beta_1^2) & k = \pm 1 \\ 0 & k > 1. \end{cases} \quad (3.3)$$

O processo é estacionário para qualquer valor de β_1 mas em geral é desejável impor restrições para que ele satisfaça uma condição chamada *inversibilidade*. Considere os seguintes processos MA(1)

$$\begin{aligned} X_t &= \epsilon_t + \theta \epsilon_{t-1} \\ X_t &= \epsilon_t + \frac{1}{\theta} \epsilon_{t-1}. \end{aligned}$$

Substituindo em (3.3) não é difícil verificar que estes dois processos diferentes têm exatamente a mesma função de autocorrelação. Assim, não é possível identificar um processo MA(1) único a partir da função de autocorrelação. Por outro lado, podemos fazer substituições sucessivas e reescrever estes dois processos colocando ϵ_t em função de X_t, X_{t-1}, \dots , i.e.

$$\begin{aligned} \epsilon_t &= X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \theta^3 X_{t-3} + \dots \\ \epsilon_t &= X_t - \frac{1}{\theta} X_{t-1} + \frac{1}{\theta^2} X_{t-2} - \frac{1}{\theta^3} X_{t-3} + \dots \end{aligned}$$

Se $|\theta| < 1$ a primeira série converge e o modelo é dito ser *inversível* mas a segunda não converge e o modelo é *não inversível*. Ou seja, a condição de inversibilidade (neste

caso $|\theta| < 1$) garante que existe um único processo MA(1) para uma dada função de autocorrelação. Outra consequência da inversibilidade é que o processo MA(1) pode ser reescrito como uma regressão de ordem infinita nos seus próprios valores defasados.

Para um processo MA(q) esta condição pode ser melhor expressa usando-se o *operador de retardo*, denotado por B e definido como

$$B^j X_t = X_{t-j}, \quad \text{para todo } j.$$

A equação (3.1) pode então ser reescrita como

$$X_t = (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q) \epsilon_t = \theta(B) \epsilon_t$$

onde $\theta(B)$ é um polinômio de ordem q em B . Um processo MA(q) é inversível se as raízes da equação

$$\theta(B) = 1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q = 0$$

estiverem fora do círculo unitário. Ou seja, se $\delta_1, \dots, \delta_q$ são q soluções de $\theta(B) = 0$ então o processo é inversível se $|\delta_i| > 1$, $i = 1, \dots, q$. Teremos então 2^q modelos com a mesma função de autocorrelação mas somente um deles será inversível.

Finalmente, vale notar que uma constante μ qualquer pode ser adicionada ao lado direito de (3.1) dando origem a um processo com média μ . O processo continuará sendo estacionário com $E(X_t) = \mu$ e em particular a função de autocorrelação não será afetada.

3.4.4 Processos Autoregressivos

Suponha que $\{\epsilon_t\}$ seja um processo puramente aleatório com média zero e variância σ_ϵ^2 . Um processo $\{X_t\}$ é chamada de processo autoregressivo de ordem p , ou AR(p), se

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t. \quad (3.4)$$

Note a similaridade com um modelo de regressão múltipla, onde os valores passados de X_t fazem o papel das regressoras. Assim, processos AR podem ser usados como modelos se for razoável assumir que o valor atual de uma série temporal depende do seu passado imediato mais um erro aleatório.

Por simplicidade vamos começar estudando em detalhes processos de primeira ordem, AR(1), i.e.

$$X_t = \alpha X_{t-1} + \epsilon_t. \quad (3.5)$$

Note que existe uma estrutura Markoviana no processo AR(1) no sentido de que, dado X_{t-1} , X_t não depende de X_{t-2}, X_{t-3}, \dots . Fazendo substituições sucessivas em (3.5)

obtemos que

$$\begin{aligned}
X_t &= \alpha(\alpha X_{t-2} + \epsilon_{t-1}) + \epsilon_t = \alpha^2 X_{t-2} + \alpha \epsilon_{t-1} + \epsilon_t \\
&= \alpha^2(\alpha X_{t-3} + \epsilon_{t-2}) + \alpha \epsilon_{t-1} + \epsilon_t \\
&= \alpha^3 X_{t-3} + \alpha^2 \epsilon_{t-2} + \alpha \epsilon_{t-1} + \epsilon_t \\
&\vdots \\
&= \alpha^{r+1} X_{t-r-1} + \sum_{j=0}^r \alpha^j \epsilon_{t-j}.
\end{aligned}$$

Se X_t for estacionário com variância finita σ_X^2 podemos escrever que

$$E[X_t - \sum_{j=0}^r \alpha^j \epsilon_{t-j}]^2 = \alpha^{2r+2} E(X_{t-r-1}^2) = \alpha^{2r+2} \sigma_X^2$$

e se $|\alpha| < 1$ temos que $\alpha^{2r+2} \rightarrow 0$ quando $r \rightarrow \infty$. Portanto, esta condição nos permite escrever X_t como o seguinte processo MA infinito,

$$X_t = \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \dots$$

e assim $|\alpha| < 1$ é uma condição suficiente para que X_t seja estacionário. Neste caso, reescrevendo o processo k períodos à frente, i.e.

$$X_{t+k} = \epsilon_{t+k} + \alpha \epsilon_{t+k-1} + \dots + \alpha^k \epsilon_t + \dots \quad (3.6)$$

note como o efeito de ϵ_t sobre X_{t+k} diminui a medida que k aumenta e por isso é chamado *efeito transitório*.

Podemos também usar o operador de retardo reescrevendo a equação (3.5) como

$$(1 - \alpha B)X_t = \epsilon_t$$

ou equivalentemente

$$X_t = \frac{1}{(1 - \alpha B)} \epsilon_t = (1 + \alpha B + \alpha^2 B^2 + \dots) \epsilon_t = \epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \dots$$

Escrevendo o processo AR(1) neste formato de MA infinito fica fácil ver que a sua média e variância são dados por

$$E(X_t) = 0 \quad \text{e} \quad \text{Var}(X_t) = \sigma_\epsilon^2 (1 + \alpha^2 + \alpha^4 + \dots) = \frac{\sigma_\epsilon^2}{1 - \alpha^2}.$$

A função de autocovariância pode ser obtida usando os resultados acima. Reescrevendo a equação (3.6) como

$$X_{t+k} = \epsilon_{t+k} + \dots + \alpha^{k-1} \epsilon_{t+1} + \alpha^k \epsilon_t + \alpha^{k+1} \epsilon_{t-1} + \alpha^{k+2} \epsilon_{t-2} + \dots$$

pode-se verificar que, para qualquer $k = 1, 2, \dots$,

$$\text{Cov}(\epsilon_t + \alpha \epsilon_{t-1} + \alpha^2 \epsilon_{t-2} + \dots, \epsilon_{t+k} + \dots + \alpha^{k-1} \epsilon_{t+1}) = 0.$$

Portanto,

$$\begin{aligned}
 E(X_t X_{t+k}) &= \text{Cov}(\epsilon_t + \alpha\epsilon_{t-1} + \alpha^2\epsilon_{t-2} + \dots, \alpha^k\epsilon_t + \alpha^{k+1}\epsilon_{t-1} + \alpha^{k+2}\epsilon_{t-2} + \dots) \\
 &= \alpha^k E(\epsilon_t^2) + \alpha^{k+2} E(\epsilon_{t-1}^2) + \alpha^{k+4} E(\epsilon_{t-2}^2) + \dots \\
 &= \alpha^k \sigma_\epsilon^2 (1 + \alpha^2 + \alpha^4 + \dots) = \alpha^k \frac{\sigma_\epsilon^2}{1 - \alpha^2} = \alpha^k \sigma_X^2 = \gamma(k).
 \end{aligned}$$

Assim, a função de autocorrelação é $\rho(k) = \alpha^k$ para $k = 0, 1, 2, \dots$. Assim, como a média e a variância são constantes e $\gamma(k)$ não depende de t o processo AR(1) com $|\alpha| < 1$ é estacionário.

Na Figura 3.1 são mostradas graficamente as autocorrelações teóricas de um processo AR(1) até a defasagem $k = 20$ para α igual a 0,8, -0,8, 0,2 e -0,2. Note como a função de autocorrelação decai rapidamente para zero quando $\alpha = 0,2$ e se alterna entre valores positivos e negativos quando $\alpha = -0,8$. Ou seja sempre há um decaimento exponencial para zero mas este decaimento depende do sinal e magnitude de α .

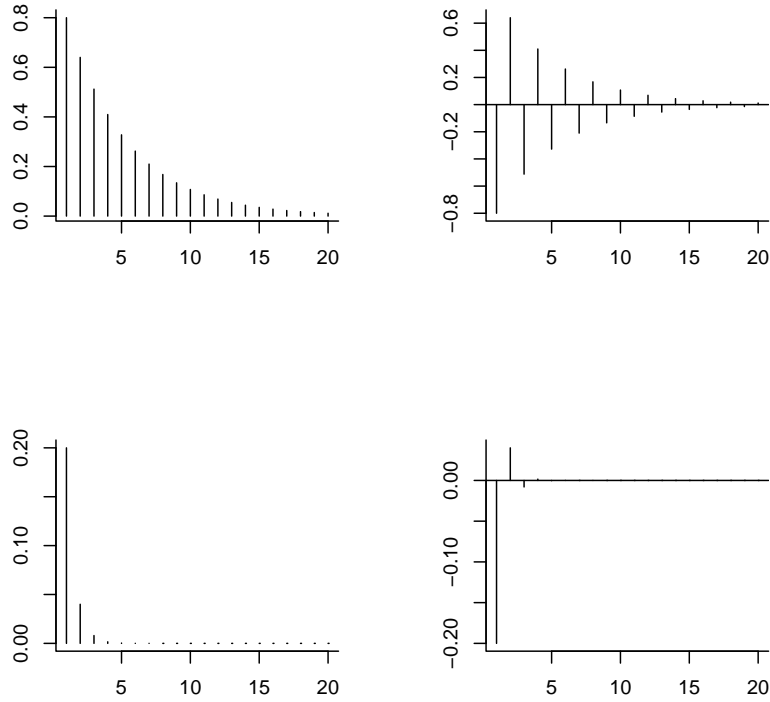


Figura 3.1: As 20 primeiras autocorrelações teóricas de um processo AR(1) com (a) $\alpha = 0,8$, (b) $\alpha = -0,8$, (c) $\alpha = 0,2$ e (d) $\alpha = -0,2$.

Generalizando os resultados acima para um processo AR(p) escrevemos novamente X_t como um processo MA infinito com coeficientes ψ_0, ψ_1, \dots , i.e.

$$X_t = \psi_0\epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \dots = (\psi_0 + \psi_1 B + \psi_2 B^2 + \dots)\epsilon_t = \psi(B)\epsilon_t.$$

e em analogia com o caso AR(1) segue que o processo será estacionário se $\sum_j \psi_j^2 < \infty$. Usando agora o operador de retardo a equação (3.4) fica

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p) X_t = \epsilon_t \quad \text{ou} \quad \phi(B) X_t = \epsilon_t$$

e portanto o processo AR(p) pode ser escrito como

$$X_t = \phi(B)^{-1} \epsilon_t = \psi(B) \epsilon_t.$$

Assim, os coeficientes ψ_j podem ser obtidos a partir dos coeficientes α_j fazendo-se

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)(\psi_0 + \psi_1 B + \psi_2 B^2 + \dots) = 1$$

Desenvolvendo-se esta expressão segue que

$$\begin{aligned} & \psi_0 + \psi_1 B + \psi_2 B^2 + \dots - \alpha_1 \psi_0 B - \alpha_1 \psi_1 B^2 - \alpha_1 \psi_2 B^3 - \dots \\ & - \alpha_2 \psi_0 B^2 - \alpha_2 \psi_1 B^3 - \alpha_2 \psi_2 B^4 - \dots \\ & \vdots \\ & - \alpha_p \psi_0 B^p - \alpha_p \psi_1 B^{p+1} - \dots = 1 + 0B + 0B^2 + \dots \end{aligned}$$

e agora agrupando em termos de B, B^2, \dots

$$\psi_0 + (\psi_1 - \alpha_1 \psi_0) B + (\psi_2 - \alpha_1 \psi_1 - \alpha_2 \psi_0) B^2 + \dots = 1 + 0B + 0B^2 + \dots$$

donde obtém-se os coeficientes MA recursivamente como

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \psi_0 \alpha_1 \\ \psi_2 &= \psi_1 \alpha_1 + \psi_0 \alpha_2 \\ \psi_3 &= \psi_2 \alpha_1 + \psi_1 \alpha_2 + \psi_0 \alpha_3 \\ &\vdots \\ \psi_i &= \sum_{j=1}^i \psi_{i-j} \alpha_j. \end{aligned}$$

O efeito de ϵ_t sobre X_{t+k} é dado por ψ_k , $k = 1, 2, \dots$

Pode-se mostrar que (ver por exemplo Box, Jenkins, & Reinsel 1994) a condição de estacionariedade do processo X_t é que todas as raízes de $\phi(B) = 0$ estejam fora do círculo unitário. Em particular, para $p = 1$ temos que $\phi(B) = 1 - \alpha B = 0$ implica que $B = 1/\alpha$ e a condição de estacionariedade fica $|\alpha| < 1$ conforme já havíamos verificado.

Para reescrever um processo AR(p) em forma vetorial, defina $\mathbf{Z}_t = (X_{t-1}, \dots, X_{t-p})'$ e portanto

$$\mathbf{Z}_t = \Phi \mathbf{Z}_{t-1} + \mathbf{u}_t$$

sendo a matriz Φ definida como

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

e $\mathbf{u}_t = (\epsilon_t, 0, \dots, 0)'$.

Para obter a função de autocorrelação de um processo $\text{AR}(p)$ é algebricamente mais simples assumir *a priori* que o processo é estacionário com $E(X_t) = 0$, $\text{Var}(X_t) = \sigma_X^2$ e $\text{Cov}(X_t, X_{t-k}) = \gamma(k)$. Neste caso, multiplicando a equação (3.4) por X_{t-k} , i.e

$$X_t X_{t-k} = \alpha_1 X_{t-1} X_{t-k} + \dots + \alpha_p X_{t-p} X_{t-k} + \epsilon_t X_{t-k}.$$

e tomando o valor esperado obtemos que

$$\begin{aligned} E(X_t X_{t-k}) = \gamma(k) &= \alpha_1 E(X_{t-1} X_{t-k}) + \dots + \alpha_p E(X_{t-p} X_{t-k}) \\ &= \alpha_1 \gamma(k-1) + \dots + \alpha_p \gamma(k-p), \quad k > 0. \end{aligned}$$

Dividindo-se ambos os lados pela variância constante σ_X^2 obtem-se que

$$\rho(k) = \alpha_1 \rho(k-1) + \dots + \alpha_p \rho(k-p), \quad k > 0$$

chamadas equações de Yule-Walker.

Por exemplo, para um processo $\text{AR}(1)$ com coeficiente α segue que $\rho(1) = \alpha$, $\rho(2) = \alpha \rho(1) = \alpha^2$, ..., $\rho(k) = \alpha^k$ como já havíamos verificado. Para um processo $\text{AR}(2)$ com coeficientes α_1 e α_2 segue que

$$\rho(1) = \alpha_1 \rho(0) + \alpha_2 \rho(1) \Rightarrow \rho(1) = \alpha_1 / (1 - \alpha_2)$$

e as outras autocorrelações são obtidas iterativamente como

$$\rho(k) = \alpha_1 \rho(k-1) + \alpha_2 \rho(k-2), \quad k \geq 2$$

Autocorrelações Parciais

Para um processo $\text{AR}(p)$, o último coeficiente α_p mede o “excesso de correlação” na defasagem p que não é levado em conta por um modelo $\text{AR}(p-1)$. Este é chamado de *p-ésimo coeficiente de autocorrelação parcial*. Assim, variando $k = 1, 2, \dots$ temos a chamada *função de autocorrelação parcial* (FACP).

Por outro lado, em um processo $\text{AR}(p)$ não existe correlação direta entre X_t e $X_{t-p-1}, X_{t-p-2}, \dots$ e substituindo $k = p+1, p+2, \dots$ nas equações de Yule-Walker obtem-se que todos os coeficientes de correlação parcial serão nulos para $k > p$. Por exemplo, substituindo-se $k = p+1$ segue que

$$\rho(p+1) = \alpha_1 \rho(p) + \dots + \alpha_p \rho(1) + \alpha_{p+1}.$$

O fato de que a facp é igual a zero para $k > p$ é sugerido em Box and Jenkins (1970, p. 170) como uma ferramenta para determinar a ordem p do processo autoregressivo para séries temporais observadas.

3.4.5 Modelos Mistos ARMA

Combinando-se modelos AR e MA pode-se obter uma representação adequada com um número menor de parâmetros. Processos autoregressivos médias móveis (ARMA) formam um classe de modelos muito úteis e parcimoniosos para descrever dados de séries temporais. O modelo $\text{ARMA}(p, q)$ é dado por

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$$

onde $\{\epsilon_t\}$ é um processo puramente aleatório com média zero e variância σ_ϵ^2 . Note que, modelos AR ou MA podem ser obtidos como casos especiais quando $p = 0$ ou $q = 0$. Usando o operador de retardo o modelo pode ser reescrito como

$$(1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p) X_t = (1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_q B^q) \epsilon_t$$

ou

$$\phi(B) X_t = \theta(B) \epsilon_t.$$

Os valores de $\alpha_1, \dots, \alpha_p$ que tornam o processo estacionário são tais que as raízes de $\phi(B) = 0$ estão fora do círculo unitário. Analogamente, os valores de β_1, \dots, β_q que tornam o processo inversível são tais que as raízes de $\theta(B) = 0$ estão fora do círculo unitário.

Vale notar que as funções de autocorrelação e autocorrelação parcial ficam consideravelmente mais complicadas em processos ARMA. De um modo geral, para um processo $\text{ARMA}(p, q)$ estacionário a função de autocorrelação tem um decaimento exponencial ou oscilatório após a defasagem q enquanto que a facp tem o mesmo comportamento após a defasagem p (Box & Jenkins 1970, p. 79). Em princípio este resultado pode ser utilizado para auxiliar na determinação da ordem (p, q) do processo mas na prática pode ser bastante difícil distinguir entre decaimentos exponenciais e oscilatórios através das estimativas destas funções.

A Tabela 3.1 mostra as propriedades teóricas das funções de autocorrelação e autocorrelação parcial para alguns processos estacionários como auxiliar na identificação do modelo.

Tabela 3.1: Propriedades teóricas da fac e facp.

Processo	FAC	FACP
série aleatória	0	0
AR(1), $\alpha > 0$	decaimento exponencial	0, $k \geq 2$
AR(1), $\alpha < 0$	decaimento oscilatório	idem
AR(p)	decaimento para zero	0, $k > p$
MA(1)	0, $k > 1$	decaimento oscilatório
ARMA(p, q)	decaimento a partir de q	decaimento a partir de p

3.4.6 Modelos ARMA Integrados

Os modelos discutidos até agora são apropriados para séries temporais estacionárias. Assim, para ajustar estes modelos a uma série temporal observada é necessário remover as fontes de variação não estacionárias. Por exemplo, se a série observada for não estacionária na média pode-se tentar remover a tendência tomando-se uma ou mais diferenças (esta abordagem é muito utilizada em Econometria).

Um modelo ARMA no qual X_t é substituído pela sua d -ésima diferença $\nabla^d X_t$ é capaz de descrever alguns tipos de séries não estacionárias. Denotando a série diferenciada por

$$W_t = \nabla^d X_t = (1 - B)^d X_t$$

o processo autoregressivo integrado médias móveis denotado ARIMA(p, d, q) é dado por

$$W_t = \alpha_1 W_{t-1} + \cdots + \alpha_p W_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$$

ou, equivalentemente

$$\phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t. \quad (3.7)$$

Da equação (3.7) acima pode-se notar que o modelo para X_t é claramente não estacionário já que o polinômio autoregressivo $\phi(B)(1 - B)^d$ tem exatamente d raízes sobre o círculo unitário, ou d *raízes unitárias*. Um processo que se torna estacionário após d diferenças é dito ser não estacionário *homogêneo*, ou *integrado* de ordem d , $I(d)$.

Na prática valores pequenos são em geral especificados para d , sendo $d = 1$ o valor mais frequentemente utilizado e excepcionalmente $d = 2$. Note também que o passeio aleatório pode ser considerado um processo ARIMA(0,1,0).

Vale notar que para dados reais um modelo ARIMA (e de fato qualquer modelo) é no máximo uma aproximação para o verdadeiro processo gerador dos dados. Na prática pode ser bem difícil distinguir entre um processo estacionário com memória longa (e.g. AR(1) com $\alpha \approx 1$) e um processo não estacionário homogêneo. Existe uma vasta literatura econométrica sobre testes de raiz unitária (ver por exemplo Hamilton 1994 e Bauwens, Lubrano, & Richard 1999). Mais recentemente, modelos da classe ARFIMA (ou ARIMA fracionários) tem sido utilizados para analisar séries com memória longa. Estes tópicos não serão abordados aqui e o leitor interessado pode consultar por exemplo Brockwell & Davis (1991) além das referências acima.

Exercícios

Nos exercícios a seguir $\{\epsilon_t\}$ é um processo discreto puramente aleatório com média zero e variância σ_ϵ^2 .

1. Encontre a fac do processo $X_t = \epsilon_t + 0,7\epsilon_{t-1} - 0,2\epsilon_{t-2}$.
2. Encontre a fac do processo $X_t - \mu = 0,7(X_{t-1} - \mu) + \epsilon_t$.
3. Encontre a fac do processo $X_t = \frac{1}{3}X_{t-1} + \frac{2}{9}X_{t-2} + \epsilon_t$.

4. Se $X_t = \mu + \epsilon_t + \beta\epsilon_{t-1}$ mostre que a fac do processo não depende de μ .
5. Reescreva cada um dos modelos abaixo em termos de operador de retardo B e verifique se o modelo é estacionário e/ou inversível:
 - (a) $X_t = 0,3X_{t-1} + \epsilon_t$.
 - (b) $X_t = \epsilon_t - 1,3\epsilon_{t-1} + 0,4\epsilon_{t-2}$.
 - (c) $X_t = 0,5X_{t-1} + \epsilon_t - 1,3\epsilon_{t-1} + 0,4\epsilon_{t-2}$.
 - (d) $\nabla X_t = 0,3\nabla X_{t-1} + \epsilon_t - 0,6\epsilon_{t-1}$
 - (e) $X_t = X_{t-1} + \epsilon_t - 1,5\epsilon_{t-1}$
6. Mostre que o processo $X_t = X_{t-1} + cX_{t-2} + \epsilon_t$ é estacionário se $-1 < c < 0$ e obtenha a fac para $c = -3/16$.
7. Mostre que o processo $X_t = X_{t-1} + cX_{t-2} - cX_{t-3} + \epsilon_t$ é não estacionário para qualquer valor de c .
8. Descreva como deve se comportar a função de autocorrelação teórica para os seguintes processos,
 - (a) AR(1) estacionário, para $\alpha = 0,1$, $\alpha = -0,75$ e $\alpha = 0,99$.
 - (b) Médias móveis de ordem q .
 - (c) Como deveriam ficar as funções de autocorrelação e autocorrelação parcial amostrais que identificam os processos acima?
9. Descreva como deveriam se comportar as funções de autocorrelação e autocorrelação parcial amostrais para processos AR, MA e ARMA não sazonais.
10. Para o modelo $(1 - B)(1 - 0,2B)X_t = (1 - 0,5B)\epsilon_t$, identifique os valores de p , q , e d e verifique se o processo é estacionário e inversível.
11. Mostre que a função de autocovariância de um processo AR(1) estacionário com variância σ_X^2 é dada por $\alpha^k \sigma_X^2$ (Sugestão: use a expressão (3.2) com $q \rightarrow \infty$)
12. Verifique se $X_t = \sum_{j=1}^t \epsilon_j$ é estacionário.
13. Mostre que a fac do processo $X_t = aX_{t-1} + \epsilon_t + b\epsilon_{t-1}$ é dada por

$$\begin{aligned}\rho(1) &= \frac{(1+ab)(a+b)}{1+b^2+2ab} \\ \rho(k) &= a\rho(k-1), \quad k = 2, 3, \dots\end{aligned}$$

14. Obtenha a função de autocovariância do processo

$$X_t = \epsilon_t + \frac{1}{a}\epsilon_{t-1} + \frac{1}{a^2}\epsilon_{t-2} + \dots + \frac{1}{a^m}\epsilon_{t-m}$$

sendo que $0 < a < 1$.

15. Se $\{X_t\}$ é um processo estacionário obtenha a função de autocovariância de $Y_t = X_t - X_{t-1}$.
16. Mostre que o processo $X_t = (\alpha + 1)X_{t-1} - \alpha X_{t-2} + \epsilon_t$ tem exatamente uma raiz unitária e reescreva-o como um processo ARIMA(1,1,0).
17. Obtenha a função de autocorrelação do passeio aleatório $X_t = X_{t-1} + \epsilon_t$ com $E(\epsilon_t) = \mu$, $Var(\epsilon_t) = \sigma_\epsilon^2$ e $Cov(\epsilon_t, \epsilon_s) = 0$, $\forall t \neq s$.
18. Verifique se o processo $\{Y_t\}$ tal que $P(Y_t = 1) = P(Y_t = -1) = 1/2$ é estacionário. Obtenha sua média, variância e covariância.
19. Sejam os processos $Y_t = \epsilon_t + \theta\epsilon_{t-1}$, $|\theta| > 1$ e $\{X_t\}$ tal que $X_t = 1$ se $Y_t \geq 0$ e $X_t = -1$ se $Y_t < 0$. Verifique se $\{X_t\}$ e $\{Y_t\}$ são estacionários. Calcule a função de autocorrelação de $\{X_t\}$.
20. Verifique que o processo $Y_t = (-1)^t \epsilon_t$ é estacionário e que $X_t = Y_t + \epsilon_t$ não é estacionário.
21. Se $\{X_t\}$ e $\{Y_t\}$ são independentes e estacionários verifique se $Z_t = \alpha X_t + \beta Y_t$, $\alpha, \beta \in \mathbb{R}$ também é estacionário.
22. Obtenha a representação MA(∞) de um processo AR(2) estacionário.
23. Obtenha a representação AR(∞) de um processo MA(1) inversível.

Capítulo 4

Estimação

No capítulo anterior foram estudados modelos probabilísticos que podem ser utilizados para descrever dados de séries temporais. Neste capítulo será discutido o problema de ajustar um modelo aos dados observados. A inferência será baseada na função de autocorrelação.

Para um processo estacionário $\{X_t\}$ ($t = 1, \dots, n$), a função de densidade de probabilidade conjunta de X_1, \dots, X_n pode ser sempre fatorada como

$$\begin{aligned} p(x_1, \dots, x_n) &= p(x_1)p(x_n, \dots, x_2|x_1) \\ &= p(x_1)p(x_2|x_1)p(x_n, \dots, x_3|x_2, x_1) \\ &\vdots \\ &= p(x_1) \prod_{t=2}^n p(x_t|x_{t-1}, \dots, x_1). \end{aligned}$$

Em particular para um modelo ARMA(p, q), denotando o vetor de parâmetros por $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \sigma_\epsilon^2)'$ e destacando-se a densidade conjunta das p primeiras realizações segue que

$$\begin{aligned} p(x_1, \dots, x_n|\boldsymbol{\theta}) &= p(x_1, \dots, x_p|\boldsymbol{\theta}) \prod_{t=p+1}^n p(x_t|x_{t-1}, \dots, x_1, \boldsymbol{\theta}) \\ &= p(x_1, \dots, x_p|\boldsymbol{\theta}) \prod_{t=p+1}^n p(x_t|x_{t-1}, \dots, x_p, \boldsymbol{\theta}). \end{aligned} \quad (4.1)$$

A última igualdade vem da estrutura Markoviana da componente autoregressiva. O segundo termo em (4.1) é a densidade condicional conjunta de x_{p+1}, \dots, x_n dados os valores iniciais x_1, \dots, x_p e define então uma função de verossimilhança condicional enquanto $p(x_1, \dots, x_n|\boldsymbol{\theta})$ define a função de verossimilhança exata.

Se for atribuída uma distribuição de probabilidades conjunta também para $\boldsymbol{\theta}$ então pelo Teorema de Bayes é possível obter sua distribuição atualizada após os dados serem observados (distribuição a posteriori),

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})} \propto p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

4.1 Autocovariância e autocorrelação

O coeficiente de autocovariância amostral de ordem k foi definido na Seção 2.4 como

$$c_k = \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})/n$$

que é o estimador usual do coeficiente de autocovariância teórico $\gamma(k)$. As propriedades deste estimador não serão detalhadas aqui mas podem ser encontradas por exemplo em Priestley (1981). Após obter as estimativas de $\gamma(k)$ os coeficientes de autocorrelação são então estimados como $r_k = c_k/c_0$, $k = 1, 2, \dots$.

Aqui serão consideradas apenas as propriedades de r_k quando a amostra vem de um processo puramente aleatório (propriedades gerais podem ser obtidas em Kendall et al. 1983, Capítulo 48). Vimos na Seção 2.4.1 que o coeficiente de autocorrelação amostral r_k é assintoticamente normalmente distribuído, com média e variância dados por

$$E(r_k) \approx -1/n \quad \text{e} \quad \text{Var}(r_k) \approx 1/n.$$

e os limites de confiança aproximados de 95% frequentemente utilizados são dados por $\pm 1,96/\sqrt{n}$. No caso geral, limites de $100(1-\alpha)\%$ podem ser construídos como $\pm q_{\alpha/2}/\sqrt{n}$ sendo $q_{\alpha/2}$ o percentil $\alpha/2$ da distribuição normal padrão.

Interpretando o correlograma

No Capítulo 2 foram vistos alguns exemplos de correlogramas associados a características de séries temporais observadas. O correlograma é útil também na identificação do tipo de modelo ARIMA que fornece a melhor representação de uma série observada. Um correlograma como o da Figura 2.7 por exemplo, aonde os valores de r_k decaem para zero de forma relativamente lenta, indica não estacionariedade e a série precisa ser diferenciada. Para séries estacionárias o correlograma é comparado com as autocorrelações teóricas de vários processos ARMA para auxiliar na identificação daquele mais apropriado. Por exemplo, se r_1 é significativamente diferente de zero e todos os valores subsequentes r_2, r_3, \dots são próximos de zero então um modelo MA(1) é indicado já que sua função de autocorrelação teórica se comporta assim. Por outro lado, se r_1, r_2, r_3, \dots parecem estar decaindo exponencialmente então um modelo AR(1) pode ser apropriado.

Vale notar entretanto que a interpretação de correlogramas é um dos aspectos mais difíceis da análise de séries temporais. A função de autocorrelação parcial é um importante coadjuvante nesta etapa de identificação se houver termos autoregressivos no modelo já que seus valores estimados tendem a ficar próximos de zero após a defasagem p .

Vimos no Capítulo 3 que para um processo ARMA(p, q) estacionário a função de autocorrelação teórica terá um decaimento exponencial ou oscilatório após a defasagem q enquanto que a função de autocorrelação parcial teórica terá o mesmo comportamento após a defasagem p . Mas na prática esta distinção entre decaimentos

exponenciais e oscilatórios através das estimativas destas funções pode ser bastante difícil.

4.2 Ajustando Processos Autoregressivos

Para um processo AR de ordem p com média μ dado por

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + \cdots + \alpha_p(X_{t-p} - \mu) + \epsilon_t,$$

e dadas n observações x_1, \dots, x_n , os parâmetros $\mu, \alpha_1, \dots, \alpha_p$ podem ser estimados pelo método de mínimos quadrados, i.e. minimizando-se a soma de quadrados

$$S = \sum_{t=p+1}^n [(x_t - \mu) - \alpha_1(x_{t-1} - \mu) - \cdots - \alpha_p(x_{t-p} - \mu)]^2$$

com respeito a $\mu, \alpha_1, \dots, \alpha_p$. Note que o somatório é de $t = p + 1$ em diante, mas esta pequena perda de informação não será importante se a série não for muito curta. Além disso, se o processo ϵ_t tiver distribuição normal então as estimativas de mínimos quadrado coincidem com as estimativas de máxima verossimilhança condicionada nas p primeiras observações.

Alternativamente, dois métodos aproximados podem ser utilizados tomando-se $\hat{\mu} = \bar{x}$. O primeiro ajusta os dados ao modelo

$$X_t - \bar{x} = \alpha_1(X_{t-1} - \bar{x}) + \cdots + \alpha_p(X_{t-p} - \bar{x}) + \epsilon_t,$$

como se fosse um modelo de regressão linear múltipla.

No segundo método os coeficientes de autocorrelação $\rho(k)$ são substituídos pelas suas estimativas r_k nas p primeiras equações de Yule-Walker. Ou seja, estamos usando o métodos dos momentos e por isto os estimadores resultantes são assintoticamente equivalentes aos estimadores de máxima verossimilhança. Assim, temos um sistema com p equações e p incógnitas $\alpha_1, \dots, \alpha_p$, i.e.

$$\begin{aligned} r_1 &= \alpha_1 + \alpha_2 r_1 + \cdots + \alpha_p r_{p-1} \\ r_2 &= \alpha_1 r_1 + \alpha_2 + \cdots + \alpha_p r_{p-2} \\ &\vdots \\ r_p &= \alpha_1 r_{p-1} + \alpha_2 r_{p-2} + \cdots + \alpha_p \end{aligned}$$

ou equivalentemente,

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$$

Exemplo 4.1: Usando os comandos do R abaixo vamos simular um processo AR(3) e usar as equações de Yule-Walker para estimar os coeficientes.

```

> x = arima.sim(n = 200, model = list(ar = c(0.6, -0.7, 0.2)))
> r = acf(x, plot = FALSE)$acf[2:4]
> R = diag(3)
> R[1, 2] = R[2, 1] = r[1]
> R[1, 3] = R[3, 1] = r[2]
> R[2, 3] = R[3, 2] = r[1]
> round(solve(R, r), 4)

[1] 0.6659 -0.7513 0.2678

```

Para estimação por mínimos quadrados basta escrever o $AR(p)$ como um modelo linear usual e resolver um sistema de equações lineares. Definindo-se

$$\mathbf{y} = \begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_p & \dots & x_1 \\ x_{p-1} & \dots & x_2 \\ \vdots & & \vdots \\ x_{n-1} & \dots & x_{n-p} \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_n \end{bmatrix} \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$$

podemos reescrever o modelo na forma matricial como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (4.2)$$

sendo $E(\boldsymbol{\epsilon}) = \mathbf{0}$, $Var(\boldsymbol{\epsilon}) = \sigma_\epsilon^2 \mathbf{I}_{n-p}$ e \mathbf{I}_{n-p} a matriz identidade de ordem $n-p$. A solução de mínimos quadrados para os coeficientes $\boldsymbol{\alpha}$ é obtida minimizando-se $\boldsymbol{\epsilon}'\boldsymbol{\epsilon}$ e é dada por $\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Usando o valor estimado de $\boldsymbol{\alpha}$ na equação do modelo calcula-se os resíduos como $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\alpha}}$, i.e.

$$e_t = x_t - \sum_{j=1}^p \hat{\alpha}_j x_{t-j}, \quad t = p+1, \dots, n$$

e a estimativa de mínimos quadrados de σ_ϵ^2 é dada por

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n-p} \sum_{t=p+1}^n e_t^2.$$

Note que os resíduos também foram calculados a partir de $t = p+1$.

Mantendo a representação (4.2) e adicionando a hipótese de normalidade dos erros, i.e. $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n-p})$ obtém-se uma função de verossimilhança aproximada dada por,

$$L(\boldsymbol{\alpha}, \sigma_\epsilon^2) \propto (\sigma_\epsilon^2)^{-(n-p)/2} \exp\{-\sigma_\epsilon^{-2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})/2\}.$$

Neste caso, os EMV de $\boldsymbol{\alpha}$ e σ_ϵ^2 coincidem com os estimadores de mínimos quadrados,

$$\begin{aligned} \frac{\partial \log(L(\boldsymbol{\alpha}, \sigma_\epsilon^2))}{\partial \boldsymbol{\alpha}} &= -\frac{\sigma_\epsilon^{-2}}{2} \frac{\partial (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\ &= -\frac{\sigma_\epsilon^{-2}}{2} \frac{\partial (-2\boldsymbol{\alpha}'\mathbf{X}'\mathbf{y} + \boldsymbol{\alpha}'\mathbf{X}'\mathbf{X}\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\ &= -\frac{\sigma_\epsilon^{-2}}{2} (-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\alpha}). \end{aligned}$$

$$\left. \frac{\partial \log(L(\boldsymbol{\alpha}, \sigma_\epsilon^2))}{\partial \boldsymbol{\alpha}} \right|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} = \mathbf{0} \iff \hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Lembrando que $(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}}) = \sum_{t=p+1}^n e_t^2$ segue que

$$\begin{aligned} \frac{\partial \log(L(\hat{\boldsymbol{\alpha}}, \sigma_\epsilon^2))}{\partial \sigma_\epsilon^2} &= -\frac{1}{2} \frac{\partial}{\partial \sigma_\epsilon^2} \left[(n-p) \log(\sigma_\epsilon^2) + \sigma_\epsilon^{-2} \sum_{t=p+1}^n e_t^2 \right] \\ &= -\frac{1}{2} \left[(n-p) \sigma_\epsilon^{-2} - \sigma_\epsilon^{-4} \sum_{t=p+1}^n e_t^2 \right] \\ \left. \frac{\partial \log(L(\hat{\boldsymbol{\alpha}}, \sigma_\epsilon^2))}{\partial \sigma_\epsilon^2} \right|_{\sigma_\epsilon^2=\hat{\sigma}_\epsilon^2} &= \mathbf{0} \iff \hat{\sigma}_\epsilon^2 = \frac{1}{n-p} \sum_{t=p+1}^n e_t^2. \end{aligned}$$

Exemplo 4.2: Para um modelo AR(1) com erros normais a matriz \mathbf{X} tem somente uma coluna e não é difícil verificar que

$$\mathbf{X}'\mathbf{X} = \sum_{t=2}^n x_{t-1}^2 \quad \text{e} \quad \mathbf{X}'\mathbf{y} = \sum_{t=2}^n x_t x_{t-1}.$$

Portanto, o EMV condicional é dado por

$$\hat{\alpha} = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2} \quad \text{e} \quad \hat{\sigma}_\epsilon^2 = \frac{1}{n-1} \sum_{t=2}^n (x_t - \hat{\alpha} x_{t-1})^2.$$

Exemplo 4.3: Novamente para o modelo AR(1) com erros normais o EMV incondicional é obtido maximizando-se da função de verossimilhança exata. A expressão (4.1) com $p = 1$ fica

$$p(x_1, \dots, x_n | \alpha, \sigma_\epsilon^2) = p(x_1 | \alpha, \sigma_\epsilon^2) \prod_{t=2}^n p(x_t | x_{t-1}, \alpha, \sigma_\epsilon^2).$$

Lembrando que $E(X_t) = 0$ e $Var(X_t) = \sigma_\epsilon^2 / (1 - \alpha^2)$ e razoável assumir que $X_1 \sim N(0, \sigma_\epsilon^2 / (1 - \alpha^2))$. Segue então que

$$\begin{aligned} L(\boldsymbol{\alpha}, \sigma_\epsilon^2) &\propto \left(\frac{\sigma_\epsilon^2}{1 - \alpha^2} \right)^{-1/2} \exp \left\{ -\frac{1 - \alpha^2}{2\sigma_\epsilon^2} x_1^2 \right\} \times \\ &\quad (\sigma_\epsilon^2)^{-(n-1)/2} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{t=2}^n (x_t - \alpha x_{t-1})^2 \right\} \\ &\propto (1 - \alpha^2)^{1/2} (\sigma_\epsilon^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \left((1 - \alpha^2) x_1^2 + \sum_{t=2}^n (x_t - \alpha x_{t-1})^2 \right) \right\}. \end{aligned}$$

Maximizar esta expressão (ou seu logaritmo) em relação a α requer algum algoritmo de otimização numérica (por exemplo métodos de Newton-Raphson). No R podemos usar a função `optim` como no Exemplo 4.4.

Exemplo 4.4: Foram gerados 200 valores de um processo AR(1) com parâmetros $\alpha = 0,8$ e $\sigma_\epsilon^2 = 1$. Os comandos abaixo podem ser usados para obter as estimativas de máxima verossimilhança (incondicional). Note que estamos maximizando o logaritmo da verossimilhança e verificando a condição de estacionariedade.

```
> fun = function(theta, x) {
+   s2 = theta[1]
+   alpha = theta[2]
+   if (abs(alpha) >= 1)
+     return(-Inf)
+   n = length(x)
+   e = x[2:n] - alpha * x[1:(n - 1)]
+   Q = (1 - alpha^2) * x[1]^2 + sum(e^2)
+   return(-0.5 * (n * log(s2) - log(1 - alpha^2) + Q/s2))
+ }

> x = arima.sim(n = 200, model = list(ar = 0.8))
> init = c(1, 0.5)
> out = optim(init, fn = fun, method = "BFGS", control = list(fnscale = -1),
+   hessian = T, x = x)
> out$par

[1] 1.0196881 0.8274672
```

Como o custo computacional de estimar modelos AR não é tão grande uma abordagem alternativa para determinação de p consiste em estimar modelos de ordem progressivamente mais alta e calcular a soma de quadrados residual para cada valor de p . Pode ser possível encontrar o valor de p para o qual a inclusão de termos extras não melhora sensivelmente o ajuste. Como vimos na Seção 3.4.4 este procedimento dá origem à função de autocorrelação parcial.

Suponha agora que vamos atribuir uma distribuição de probabilidades para o vetor de parâmetros $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_p, \sigma_\epsilon^2)$. Pelo Teorema de Bayes e usando a verossimilhança condicional segue que

$$p(\boldsymbol{\theta}|\mathbf{x}) \propto p(\boldsymbol{\theta}) (\sigma_\epsilon^2)^{-(n-p)/2} \exp\{-\sigma_\epsilon^{-2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})/2\}.$$

Para representar a informação a priori sobre $\boldsymbol{\theta}$ pode-se fazer por exemplo, $p(\boldsymbol{\theta}) = p(\boldsymbol{\alpha}|\sigma_\epsilon^2)p(\sigma_\epsilon^2)$ com $\boldsymbol{\alpha}|\sigma_\epsilon^2 \sim N(0, \sigma_\epsilon^2 \mathbf{I}_p)$ ou $p(\boldsymbol{\theta}) = p(\boldsymbol{\alpha})p(\sigma_\epsilon^2)$ com $\boldsymbol{\alpha} \sim N(0, \mathbf{I}_p)$. Nos dois casos comumente assume-se que σ_ϵ^2 tem distribuição Gama Inversa, i.e. $\sigma_\epsilon^2 \sim GI(a, b)$ (ver Apêndice A), ou equivalentemente $\sigma_\epsilon^{-2} \sim Gama(a, b)$.

Exemplo 4.5: No modelo AR(1) com erros normais vamos atribuir as seguintes distribuições a priori, $\alpha \sim N(0, 1)$ e $\sigma_\epsilon^2 \sim GI(1, 1)$. Portanto,

$$p(\alpha) \propto \exp(-\alpha^2/2) \quad \text{e} \quad p(\sigma_\epsilon^2) \propto (\sigma_\epsilon^2)^{-2} \exp(-1/\sigma_\epsilon^2)$$

e os comandos abaixo podem ser usados para obter a moda da distribuição a posteriori conjunta de σ_ϵ^2 e α .

```
> prior = function(theta) {
+   s2 = theta[1]
+   alpha = theta[2]
+   return(-alpha^2/2 - 1/s2 - 2 * log(s2))
+ }
> post = function(theta, x) fun(theta, x) + prior(theta)
> out = optim(init, fn = post, method = "BFGS", control = list(fnscale = -1),
+   hessian = T, x = x)
> out$par

[1] 1.0095348 0.8262561
```

Note que as estimativas pontuais nos Exemplos 4.4 e 4.5 são bastante similares. Nenhuma restrição de estacionariedade foi imposta na distribuição a priori, mas é possível fazer uma otimização restrita ou mesmo impor esta restrição a priori. No caso do AR(1) poderíamos atribuir uma distribuição normal truncada ou uma distribuição uniforme em $(-1,1)$ para o parâmetro α .

4.3 Ajustando Processos Médias Móveis

O problema de estimação dos parâmetros em modelos MA é bem mais complicado do que em modelos AR. Os erros ϵ_t são agora funções não lineares complicadas dos parâmetros β_1, \dots, β_q e expressões analíticas para os estimadores não podem ser obtidas. Assim, métodos computacionais iterativos precisam ser utilizados para minimizar a soma de quadrados residual.

Dado um modelo MA(q)

$$X_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1} + \dots + \beta_q \epsilon_{t-q}$$

e uma série observada x_1, \dots, x_n o procedimento iterativo consiste basicamente em fixar os valores de $\mu, \beta_1, \dots, \beta_q$ e calcular os resíduos

$$e_t = x_t - \mu - \beta_1 \epsilon_{t-1} - \dots - \beta_q \epsilon_{t-q}$$

sequencialmente para $t = 1, \dots, n$ assumindo que $\epsilon_0 = \epsilon_{-1} = \dots = \epsilon_{-q+1} = 0$ e substituindo $\epsilon_{t-1}, \dots, \epsilon_{t-q}$ pelos resíduos calculados. Assim,

$$\begin{aligned} e_1 &= x_1 - \mu \\ e_2 &= x_2 - \mu - \beta_1 e_1 = x_2 - \mu - \beta_1 x_1 + \beta_1 \mu \\ e_3 &= x_3 - \mu - \beta_1 e_2 - \beta_2 e_1 \\ &\vdots \end{aligned}$$

Dados estes resíduos pode-se calcular a soma de quadrados residual $S(\mu, \beta) = \sum_{t=1}^n e_t^2$. Repetindo este procedimento para $\mu, \beta_1, \dots, \beta_q$ variando em uma grade de pontos pode-se escolher os valores que minimizam a soma de quadrados. Este procedimento requer o uso de algoritmos eficientes de otimização numérica e nada garante a sua convergência para um mínimo global.

Além das estimativas pontuais, se o processo $\{\epsilon_t\}$ tem distribuição normal então Box & Jenkins (1970), p. 228 descrevem regiões de confiança para os parâmetros do modelo. Neste caso, se $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ a função de verossimilhança fica,

$$\begin{aligned} L(\mu, \beta, \sigma_\epsilon^2) &= \prod_{t=1}^n (2\pi\sigma_\epsilon^2)^{-1/2} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} e_t^2 \right\} \\ &\propto (\sigma_\epsilon^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^n e_t^2 \right\}. \end{aligned}$$

e os valores de e_t são calculados como anteriormente. Portanto $L(\mu, \beta, \sigma_\epsilon^2)$ é uma função não linear dos parâmetros.

Em termos práticos, se o procedimento de otimização utilizado levar muitas iterações para convergir ou mesmo não convergir deve-se “desconfiar” das estimativas. Neste caso as estimativas podem ser instáveis no sentido de que adicionando-se ou removendo-se uma ou duas observações pode-se obter valores muito diferentes. Nesta situação pode ser computacionalmente mais vantajoso ajustar um modelo AR aos dados mesmo que o modelo resultante tenha mais parâmetros do que o modelo MA sugerido pela função de autocorrelação.

4.4 Ajustando Processos ARMA

Os problemas de estimação para modelos ARMA são similares aqueles para modelos MA no sentido de que um procedimento iterativo precisa ser utilizado. Isto ocorre porque os erros $\{\epsilon_t\}$ são funções não lineares complicadas de todos os coeficientes $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$. Portanto os mesmos comentários da seção anterior são válidos para procedimentos que levam muitas iterações para convergir, i.e deve-se “desconfiar” das estimativas. Os resíduos são calculados de forma análoga ao modelo MA (ver Exercício 13).

Outra dificuldade, específica de modelos ARMA, é o problema de cancelamento de raízes. Por exemplo considere o modelo ARMA(2,1)

$$X_t = 2\theta X_{t-1} - \theta^2 X_{t-2} - \phi \epsilon_{t-1} + \epsilon_t$$

que pode ser reescrito em termos do operador de retardo como

$$(1 - \theta B)^2 X_t = (1 - \phi B) \epsilon_t.$$

Note como $\theta = \phi$ implica em um modelo AR(1) $X_t = \theta X_{t-1} + \epsilon_t$, ou seja ambos os modelos implicam exatamente no mesmo comportamento para a série temporal X_t .

Este é um problema de identificação que fica ainda mais complicado em modelos de ordem mais alta.

Em termos práticos é difícil identificar o problema de cancelamento de raízes a não ser, como já foi dito, que o procedimento iterativo deverá ter convergência lenta. No caso particular de um modelo ARMA(1,1) deve-se “desconfiar” quando as estimativas de α e β são muito similares. Para outros valores de p e q a única sugestão para tentar minimizar o problema é não incluir muitos parâmetros no modelo.

Exemplo 4.6: Vamos simular um processo ARMA(1,1) com raízes similares e verificar o problema de cancelamento de raízes.

```
> x = arima.sim(n = 100, list(ar = 0.7, ma = -0.75))
> arima(x, order = c(1, 0, 1), include.mean = F)
```

Call:

```
arima(x = x, order = c(1, 0, 1), include.mean = F)
```

Coefficients:

```
          ar1          ma1
          0.5992   -0.6835
s.e.    0.2946    0.2613
```

```
sigma^2 estimated as 1.020:  log likelihood = -142.88,  aic = 291.75
```

Note como as estimativas dos coeficientes estão muito diferentes dos valores verdadeiros e os erros padrões estão enormes!

4.5 Modelos Sazonais

Muitas séries temporais contém uma componente periódica sazonal que se repete a cada s observações ($s > 1$). Por exemplo, com dados mensais e $s = 12$ tipicamente espera-se que X_t dependa de X_{t-12} e talvez de X_{t-24} além de X_{t-1}, X_{t-2}, \dots

Neste caso, tomar a primeira diferença $x_t - x_{t-1}$ não é suficiente para tornar a série (aproximadamente) estacionária. A forma apropriada de diferenciar dados com padrão sazonal acentuado é tomar diferenças no período sazonal. Por exemplo, para dados mensais a primeira diferença sazonal é

$$\nabla_{12}x_t = (1 - B^{12})x_t = x_t - x_{t-12}$$

e terá variabilidade menor do que a primeira diferença não sazonal $\nabla x_t = x_t - x_{t-1}$, sendo portanto mais fácil de identificar e estimar.

Em geral, uma diferença sazonal é denotada por ∇_s onde s é o período sazonal. A D -ésima diferença sazonal é então denotada por ∇_s^D . Combinando-se os dois tipos de diferenciação obtém-se o operador $\nabla^d \nabla_s^D$. Por exemplo, tomando-se 1 diferença simples e 1 sazonal em uma série mensal tem-se que

$$\nabla \nabla_{12}x_t = x_t - x_{t-1} - x_{t-12} + x_{t-13}$$

Box & Jenkins (1970) generalizaram o modelo ARIMA para lidar com sazonalidade e definiram um modelo ARIMA sazonal multiplicativo, denominado SARIMA, dado por

$$\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)\epsilon_t \quad (4.3)$$

onde

$$\begin{aligned} \phi(B) &= (1 - \alpha_1 B - \dots - \alpha_p B^p) \\ \Phi(B^s) &= (1 - \phi_s B^s - \dots - \phi_P B^{Ps}) \\ W_t &= \nabla^d \nabla_s^D X_t \\ \theta(B) &= (1 + \beta_1 B + \dots + \beta_q B^q) \\ \Theta(B^s) &= (1 + \theta_s B^s + \dots + \theta_Q B^{Qs}). \end{aligned}$$

Este modelo é chamado SARIMA multiplicativo de ordem $(p, d, q) \times (P, D, Q)_s$ e parece extremamente complicado à primeira vista mas na prática os valores de d e D em geral não serão maiores do que 1 e um número pequeno de coeficientes será suficiente. Por exemplo, com $P = 1$ temos que

$$\Phi(B^s) = (1 - \alpha_s B^s)$$

o que significa simplesmente que W_t depende de W_{t-s} . A série W_t é formada a partir da série original tomando-se diferenças simples para remover a tendência e diferenças sazonais para remover a sazonalidade.

Para fixar idéias considere o modelo SARIMA(1,0,0) \times (0,1,1)₁₂ para dados mensais. Ou seja temos um termo autoregressivo e um termo média móvel sazonal modelando a primeira diferença sazonal. O modelo pode ser escrito como

$$(1 - \alpha B)(1 - B^{12})X_t = (1 - \theta B^{12})\epsilon_t$$

e desenvolvendo os produtos obtemos que

$$X_t = X_{t-12} + \alpha(X_{t-1} - X_{t-13}) + \epsilon_t + \theta\epsilon_{t-12}.$$

Assim, X_t depende de X_{t-1} , X_{t-12} e X_{t-13} além do erro no tempo $t - 12$.

Para finalizar, ao ajustar um modelo sazonal aos dados a primeira tarefa é especificar os valores de d e D que tornam a série (aproximadamente) estacionária e remove a maior parte da sazonalidade. Como já foi dito, estes valores raramente serão maiores do que 1. Posteriormente os valores de p , P , q e Q devem ser especificados com base nas funções de autocorrelação e autocorrelação parcial da série diferenciada. Os valores de P e Q são especificados basicamente a partir de r_k , $k = s, 2s, \dots$. Por exemplo, para dados mensais se r_{12} é grande mas r_{24} é pequeno isto sugere que um termo média móvel sazonal pode ser adequado.

Após ter identificado, por tentativa, o que parece ser um modelo SARIMA razoável os parâmetros serão estimados por algum procedimento iterativo similar àqueles propostos para modelos ARMA. Detalhes sobre as rotinas de estimação destes modelos não serão abordados aqui e podem ser obtidos em Box & Jenkins (1970).

4.6 Adequação do Modelo

Todos os modelos são errados mas alguns são úteis (George Box)

Após identificar a ordem e estimar eficientemente os parâmetros de um modelo é necessário verificar sua adequação antes de utilizá-lo por exemplo para fazer previsões. Pode-se fazer testes de sobreajustamento, que consistem em incluir parâmetros extras no modelo e verificar sua significância estatística. No caso de modelos ARMA deve-se incluir um parâmetro de cada vez para evitar o problema de cancelamento de raízes mencionado na Seção 4.4.

4.6.1 Análise dos Resíduos

Após um modelo ter sido ajustado a uma série temporal deve-se verificar se ele fornece uma descrição adequada dos dados. Assim como em outros modelos estatísticos a idéia é verificar o comportamento dos resíduos onde

$$\text{resíduo} = \text{observação} - \text{valor ajustado}.$$

Para os modelos vistos aqui o valor ajustado é a previsão 1 passo a frente de modo que o resíduo fica definido como o erro de previsão 1 passo a frente. Por exemplo, em um modelo AR(1) se $\hat{\alpha}$ é a estimativa do coeficiente autoregressivo então o valor ajustado no tempo t é $\hat{\alpha}x_{t-1}$ e o resíduo correspondente é $e_t = x_t - \hat{\alpha}x_{t-1}$.

Se o modelo tiver um “bom” ajuste espera-se que os resíduos se distribuam aleatoriamente em torno de zero com variância aproximadamente constante e sejam não correlacionados. Se a variância dos resíduos for crescente uma transformação logarítmica nos dados pode ser apropriada. O fenômeno de “não constância” na variância é denominado de volatilidade na literatura de séries temporais e pode ser tratado através de transformações nos dados (e.g. transformações de Box-Cox)¹.

Além disso, em modelos de séries temporais os resíduos estão ordenados no tempo e é portanto natural tratá-los também como uma série temporal. É particularmente importante que os resíduos de um modelo estimado sejam serialmente (i.e. ao longo do tempo) não correlacionados. Evidência de correlação serial nos resíduos é uma indicação de que uma ou mais características da série não foi adequadamente descrita pelo modelo.

Consequentemente, duas maneiras óbvias de verificar a adequação do modelo consistem em representar graficamente os resíduos e o seu correlograma. O gráfico temporal poderá revelar a presença de dados discrepantes, efeitos de autocorrelação ou padrões cíclicos enquanto que o correlograma permite uma análise mais detalhada da estrutura de autocorrelação indicando possíveis termos faltantes no modelo.

Ou seja, assim como em outros modelos estatísticos, a idéia é que os resíduos poderão identificar características que não foram adequadamente modeladas. Por exemplo, autocorrelações residuais significativas nas defasagens 1 ou 2, ou em defasagens sazonais (e.g. 12 para dados mensais) são uma indicação de que mais termos

¹Uma tendência mais recente no entanto consiste em tentar modelar simultaneamente a média e a variância ao invés de usar transformações.

médias móveis devem ser incluídos no modelo. Por outro lado, um valor de r_k ligeiramente fora dos limites de confiança em defasagens sem significado óbvio (e.g. $k=5$) não é indicação suficiente para se rejeitar o modelo. O mesmo comentário vale para as autocorrelações parciais dos resíduos no que diz respeito à inclusão de termos autoregressivos (sazonais e não sazonais).

4.6.2 Testes sobre os resíduos

Ao invés de olhar para as autocorrelações residuais individualmente pode-se testar se um grupo de autocorrelações é significativamente diferente de zero através das chamadas estatísticas Q . Para modelos ARMA Box & Jenkins (1970) sugeriram o uso do teste de Box-Pierce para as hipóteses

$$H_0 : \rho(1) = \dots = \rho(m) = 0$$

$$H_1 : \rho(k) \neq 0, \text{ para algum } k \in \{1, \dots, m\}.$$

sendo a estatística de teste dada por

$$Q = n \sum_{k=1}^m r_k^2.$$

Na prática o número m de autocorrelações amostrais é tipicamente escolhido entre 15 e 30. Se o modelo ajustado for apropriado então Q terá distribuição aproximadamente qui-quadrado com $m - p - q$ graus de liberdade. Assim, valores grandes de Q fornecem indicação contra a hipótese de que as autocorrelações são todas nulas, em favor da hipótese de que ao menos uma delas é diferente de zero.

O teste de Box-Pierce não tem bom desempenho em amostras pequenas ou moderadas no sentido de que a distribuição se afasta da qui-quadrado. Vários testes alternativos foram sugeridos na literatura e o mais conhecido é o teste de Ljung-Box, aonde a estatística de teste é dada por

$$Q = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k}.$$

Sua distribuição amostral também é aproximadamente qui-quadrado com $m - p - q$ graus de liberdade.

Exemplo 4.7: Considere novamente a série com os totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960 que aparece na Figura ???. Existe uma clara tendência de crescimento bem como um padrão sazonal ao longo dos anos. Foi feita uma transformação logaritmica nos dados (esta transformação é sugerida na literatura). Faça os gráficos da FAC amostral da série original, 1ª diferença e 1ª diferença sazonal. Os comandos abaixo podem ser utilizados e obtém-se os gráficos da Figura 4.1.

```
> y = log(AirPassengers)
> z = cbind(y, diff(y), diff(y, lag = 12))
> yl = c("No de passageiros", "Variacao mensal", "Variacao anual")
```


Os gráficos anteriores indicam que precisamos tomar 1 diferença simples mais 1 diferença sazonal para tentar induzir estacionariedade aproximada.

```
> z = diff(diff(y), lag = 12)
> m = acf(z, lag.max = 36, plot = F)
> m$lag = m$lag * 12
```

Note que há valores grandes nas defasagens 1, 3, 12 e 23 do último gráfico. Isto pode ser uma indicação de que termos MA sazonais e não sazonais devem ser incluídos no modelo. Um modelo candidato para o logaritmo da série é SARIMA(0,1,1)x(0,1,1) e foi estimado usando os comandos abaixo.

```
> m = arima(y, order = c(0, 1, 1), seasonal = list(order = c(0,
+      1, 1)))
> m
```

Call:

```
arima(x = y, order = c(0, 1, 1), seasonal = list(order = c(0, 1, 1)))
```

Coefficients:

	ma1	sma1
	-0.4018	-0.5569
s.e.	0.0896	0.0731

```
sigma^2 estimated as 0.001348: log likelihood = 244.7, aic = -483.4
```

Como primeira verificação da adequação do modelo vamos usar a função `tsdiag()` que retorna os gráficos dos resíduos padronizados, o correlograma e os p -valores do teste de Ljung-Box para autocorrelações de ordem 1, 2, ... O resultado está na Figura 4.3.

Compare estes p -valores com o resultado da função `Box.test()` que calcula as estatísticas de Box-Pierce e Ljung-Box para a hipótese nula de independência.

```
> for (i in 1:10) {
+   b = Box.test(m$residuals, i, type = "Ljung-Box")$p.value
+   print(b)
+ }
```

```
[1] 0.8610215
[1] 0.945251
[1] 0.4829255
[1] 0.3663102
[1] 0.4320234
[1] 0.488321
[1] 0.539204
[1] 0.6328112
[1] 0.5096084
[1] 0.5502512
```

Testando a Normalidade dos Resíduos

Para uma variável aleatória X tal que $E(X) = \mu$ e $Var(X) = \sigma^2$ define-se os coeficientes de assimetria e curtose como,

$$A(X) = E\left(\frac{(X - \mu)^3}{\sigma^3}\right) \quad \text{e} \quad K(X) = E\left(\frac{(X - \mu)^4}{\sigma^4}\right)$$

respectivamente. A distribuição normal tem assimetria 0 e curtose igual a 3. Substituindo os momentos teóricos de X pelos seus equivalentes amostrais

$$m_j = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X})^j$$

os estimadores da assimetria e curtose são dados por

$$\hat{A} = \frac{m_3}{\sqrt{m_2^3}} \quad \text{e} \quad \hat{K} = \frac{m_4}{m_2^2}$$

respectivamente. Sob a hipótese de normalidade as variáveis aleatórias $\sqrt{n/6}\hat{A}$ e $\sqrt{n/24}(\hat{K} - 3)$ são independentes e têm distribuição assintótica $N(0, 1)$ e assim a estatística

$$\frac{n\hat{A}^2}{6} + \frac{n(\hat{K} - 3)^2}{24}$$

tem distribuição assintótica χ^2 com 2 graus de liberdade e pode ser usada para testar a normalidade de X .

As outras verificações usuais sobre os resíduos também devem ser feitas. Por exemplo, um histograma com curva normal superposta, o gráfico de probabilidades normais e um teste de normalidade. Os comandos abaixo podem ser utilizados no R.

```
> z = m$residuals
> d = seq(range(z)[1] - 3 * sd(z), range(z)[2] + 3 * sd(z), 0.001)
> a = shapiro.test(z)
```

Exercícios

1. Calcule as autocorrelações teóricas de um processo $MA(Q)$ puramente sazonal.
2. Faça um esboço do correlograma para uma série com estrutura $MA(Q)$ puramente sazonal, i.e. não existe dependência dentro de um período sazonal.
3. Para uma série temporal observada foi identificado o modelo $ARIMA(1,1,1)$.
 - (a) Escreva o modelo em termos do operador de retardo.
 - (b) Descreva como deve ter sido o comportamento das funções de autocorrelação e autocorrelação parcial da série original e da série diferenciada.
4. Escreva o modelo $SARIMA(0,0,1) \times (1,1,0)_{12}$ em termos de operador de retardo.

5. Para uma série mensal observada foi identificado e estimado o modelo $\text{SARIMA}(1,1,0) \times (0,1,0)$.
 - (a) Escreva o modelo em termos de operador de retardo.
 - (b) Descreva como deve ter sido o comportamento das funções de autocorrelação e autocorrelação parcial da série original e da série diferenciada.
 - (c) Como deve ser o comportamento esperado dos resíduos em termos de suas autocorrelações para que o modelo seja adequado?
 - (d) O que se deve fazer no caso de autocorrelações residuais significativas nas defasagens 1, 8 e 12 ?
6. Para uma série observada trimestralmente foi identificado e estimado o modelo $\text{SARIMA}(1,1,0) \times (2,1,1)$.
 - (a) Escreva o modelo em termos de operador de retardo.
 - (b) Descreva como deve ter sido o comportamento das funções de autocorrelação e autocorrelação parcial da série original e da série diferenciada.
 - (c) O que se deve fazer se a autocorrelação residual na defasagem 4 for significativa ?
7. Explique como você estimaria os coeficientes em um modelo $\text{ARMA}(1,1)$ utilizando as duas primeiras autocorrelações amostrais?
8. Obtenha os estimadores de mínimos quadrados para os coeficientes em um modelo $\text{AR}(2)$.
9. Escreva as equações de mínimos quadrados para o modelo $\text{AR}(p)$. Como você estima a variância dos erros?
10. Em que condições as estimativas de mínimos quadrados de um modelo $\text{AR}(p)$ coincidirão com as de máxima verossimilhança?
11. Seja o modelo $\text{AR}(1)$ com erros normais.
 - (a) Obtenha os EMV usando a verossimilhança condicional.
 - (b) Obtenha os EMV usando a verossimilhança exata com
$$X_1 \sim N(0, \sigma_\epsilon^2 / (1 - \alpha^2)).$$
12. Usando as notas de aula e qualquer outra referência bibliográfica faça um resumo da análise de resíduos em séries temporais.
13. Explique como podem ser calculados os resíduos em um modelo $\text{ARMA}(p,q)$.

```

> par(mfrow = c(3, 2))
> for (i in 1:3) {
+   plot(z[, i], main = "", xlab = "Anos", ylab = yl[i])
+   m = acf(z[, i], lag.max = 36, plot = F, na.action = na.pass)
+   m$lag = m$lag * 12
+   plot(m, main = "", xlab = "defasagem", ylab = "FAC")
+ }

```

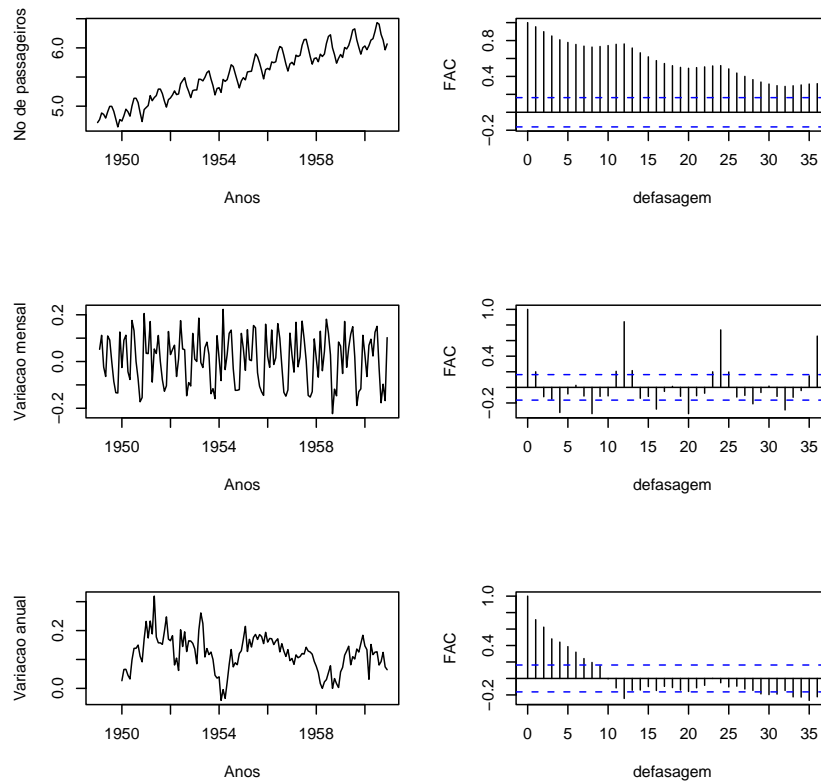


Figura 4.1:

```
> par(mfrow = c(2, 1))  
> plot(z, main = "serie com 1 diferenca simples e 1 sazonal", xlab = "Anos",  
+      ylab = "")  
> plot(m, main = "")
```

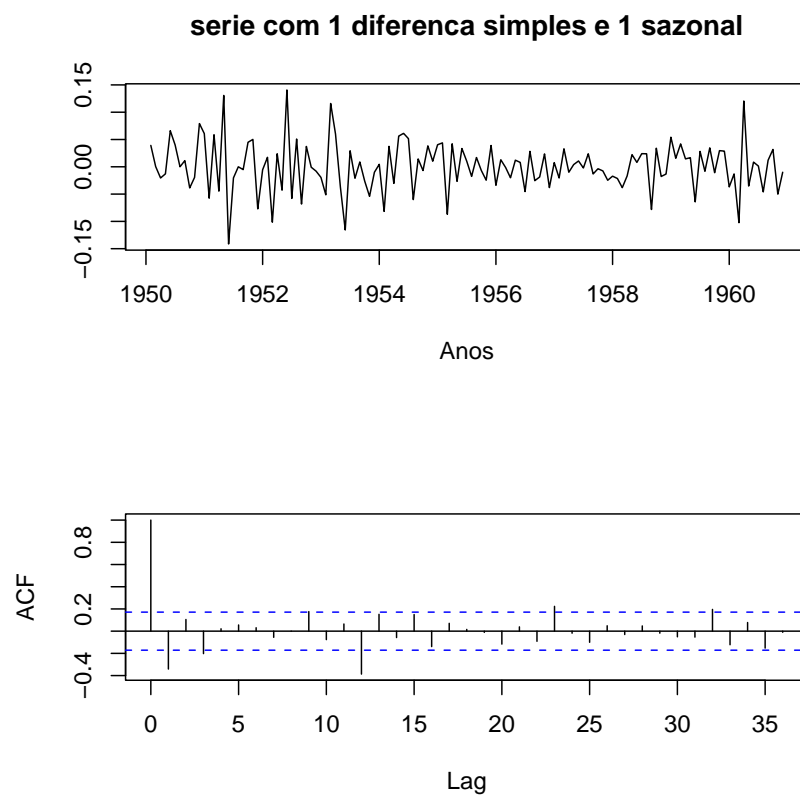


Figura 4.2:

```
> tsdiag(m)
```

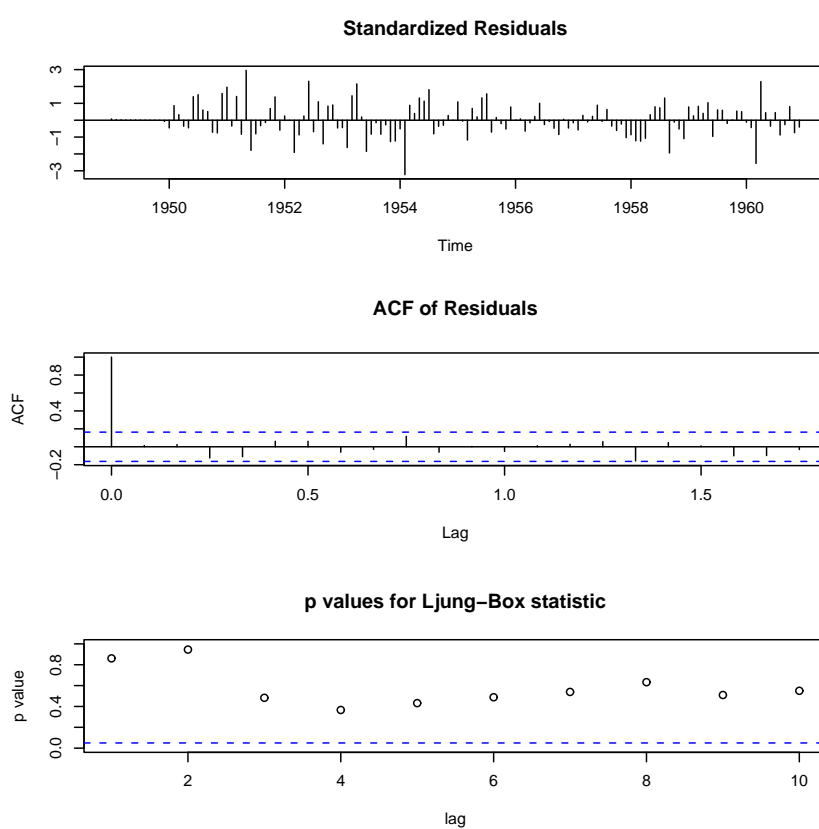


Figura 4.3:

```
> par(mfrow = c(2, 1))  
> hist(z, freq = F)  
> lines(d, dnorm(d, 0, sd(z)))  
> qqnorm(z)  
> qqline(z)  
> text(-1.5, 0.05, "Teste de Shapiro-Wilk")  
> text(-2, 0.01, paste("p-valor=", round(a$p.value, 4)))
```

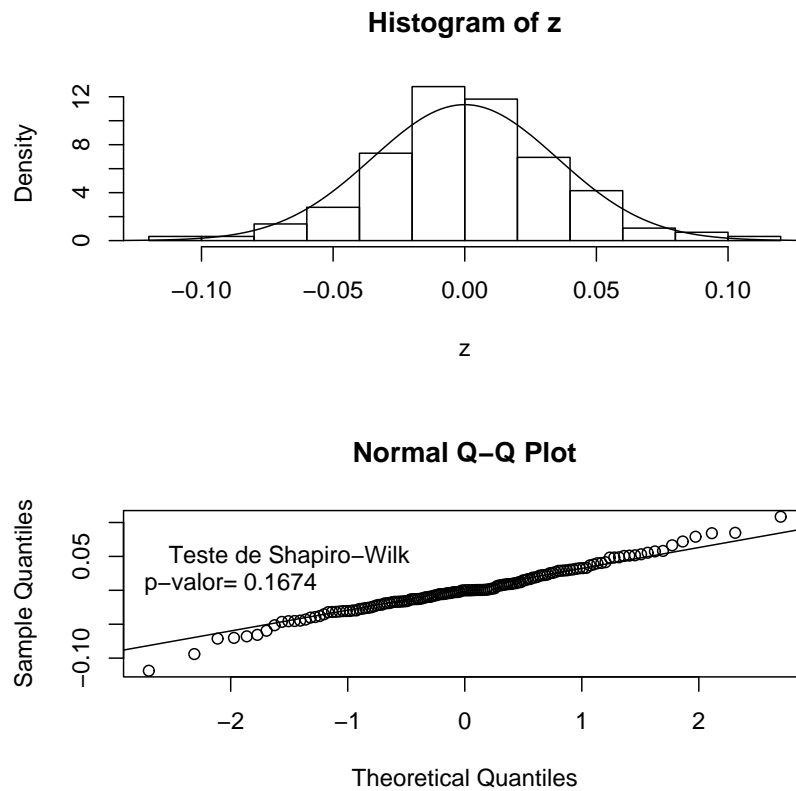


Figura 4.4:

Capítulo 5

Previsão

Uma das formas de utilização de um modelo ajustado é para fazer previsões de valores futuros. Assim, se t é o período corrente estamos interessados em prever os valores de X_{t+1}, X_{t+2}, \dots . A previsão de X_{t+k} , para $k = 1, 2, \dots$ será denotada por $\hat{x}_t(k)$ e é definida como a esperança condicional de X_{t+k} dados todos os valores passados, i.e.

$$\hat{x}_t(k) = E(X_{t+k} | x_t, x_{t-1}, \dots). \quad (5.1)$$

A equação acima é chamada de *função de previsão* e o inteiro k é chamado de *horizonte de previsão*. Pode-se mostrar que esta previsão tem o menor erro quadrático médio (EQM), $E(X_{t+k} - \hat{x}_t(k))^2$. Na prática temos um número finito de observações e obtemos então que $\hat{x}_t(k) = E(X_{t+k} | x_t, \dots, x_1)$ que não tem o EQM mínimo mas pode ser visto como uma aproximação de (5.1).

Note que se temos uma série temporal observada x_1, \dots, x_n as previsões podem ser feitas dentro do período amostral e comparadas com os valores observados. Esta é uma prática bastante comum para checar a performance preditiva do modelo. A diferença entre os valores previsto e observado, $\hat{x}_t(k) - x_{t+k}$, é chamada de *erro de previsão k passos à frente* e será denotado por e_{t+k} .

5.1 Métodos Univariados de Previsão

Os métodos descritos nesta seção têm um forte apelo intuitivo, decompondo uma série temporal em componentes de fácil interpretação. Dados os recursos computacionais disponíveis atualmente eles também têm a vantagem de serem extremamente simples de programar e sua utilização ter um custo computacional muito pequeno. Vamos começar com o caso mais simples, adequado para séries localmente constantes.

5.1.1 Alisamento Exponencial Simples

Dada uma série temporal x_1, \dots, x_n , não sazonal e sem tendência sistemática, é razoável tomar a estimativa de x_{n+1} como uma soma ponderada das observações passadas, i.e.

$$\hat{x}_n(1) = a_0 x_n + a_1 x_{n-1} + \dots$$

onde $\{a_j\}$ são os pesos. Parece razoável também dar um peso maior às observações mais recentes do que às observações mais distantes no passado, i.e. $a_0 > a_1 > \dots$. Neste procedimento são adotados pesos que decaem geometricamente a uma taxa constante dados por

$$a_j = \alpha(1 - \alpha)^j, \quad j = 0, 1, \dots$$

onde $0 < \alpha < 1$ é chamada de *constante de alisamento*. Assim, a previsão 1 passo à frente em $t = n$ fica

$$\hat{x}_n(1) = \alpha x_n + \alpha(1 - \alpha)x_{n-1} + \alpha(1 - \alpha)^2 x_{n-2} + \dots \quad (5.2)$$

Naturalmente que na prática haverá um número finito de observações passadas e a soma acima será também finita. A idéia de que o conteúdo informativo de uma observação decai com a sua “idade” é bastante intuitivo e o parâmetro α está controlando o grau de “envelhecimento” deste conteúdo.

A equação (5.2) costuma ser reescrita em forma de equação recursiva. Colocando-se $(1 - \alpha)$ em evidência obtém-se que

$$\begin{aligned} \hat{x}_n(1) &= \alpha x_n + (1 - \alpha)[\alpha x_{n-1} + \alpha(1 - \alpha)x_{n-2} + \alpha(1 - \alpha)^2 x_{n-3} + \dots] \\ &= \alpha x_n + (1 - \alpha)\hat{x}_{n-1}(1) \end{aligned} \quad (5.3)$$

i.e. uma média ponderada entre a observação mais recente e a previsão anterior. A equação (5.2) pode ainda ser reescrita na forma de *correção de erro*. Definindo $e_n = x_n - \hat{x}_{n-1}(1)$ o erro de previsão 1 passo à frente no tempo n então

$$\hat{x}_n(1) = \hat{x}_{n-1}(1) + \alpha e_n.$$

Ou seja, a previsão para $t = n + 1$ é igual à previsão para $t = n$ que foi feita em $t = n - 1$ mais uma proporção do erro cometido. A previsão k -passos a frente é a mesma, i.e. $\hat{x}_n(k) = \hat{x}_n(1)$, $k = 2, 3, \dots$.

Previsões Dentro da Amostra

Usando $\hat{x}_0(1) = x_1$ como previsão inicial em $t = 0$ e definindo $e_t = x_t - \hat{x}_{t-1}(1)$ os erros de previsão 1 passo à frente, a equação (5.3) pode ser usada recursivamente para obter as previsões, i.e.

$$\hat{x}_t(1) = \alpha x_t + (1 - \alpha)\hat{x}_{t-1}(1), \quad t = 1, 2, \dots$$

Na forma de correção de erro as recursões ficam

$$\hat{x}_t(1) = \hat{x}_{t-1}(1) + \alpha e_t, \quad t = 1, 2, \dots$$

Especificação de α

Vale notar que o valor de α não depende da escala em que as observações foram medidas, mas sim das propriedades da série temporal. O valor de α deve ser especificado de modo a refletir a influência das observações passadas nas previsões. Valores pequenos

produzem previsões que dependem de muitas observações passadas. Por outro lado, valores próximos de 1 levam a previsões que dependem das observações mais recentes e no caso extremo $\alpha = 1$ a previsão é simplesmente a última observação. O valor de α também pode ser estimado a partir dos dados e o critério utilizado é a minimização da soma de quadrados dos erros de previsão. Ou seja, dado um valor fixo de α e usando a equação (5.3), calcule

$$\begin{aligned}\hat{x}_0(1) &= x_1, \\ \hat{x}_1(1) &= \alpha x_1 + (1 - \alpha)\hat{x}_0(1), \quad e_2 = x_2 - \hat{x}_1(1) \\ \hat{x}_2(1) &= \alpha x_2 + (1 - \alpha)\hat{x}_1(1), \quad e_3 = x_3 - \hat{x}_2(1) \\ &\vdots \\ \hat{x}_{n-1}(1) &= \alpha x_{n-1} + (1 - \alpha)\hat{x}_{n-2}(1), \quad e_n = x_n - \hat{x}_{n-1}(1)\end{aligned}$$

e calcule $\sum_{t=2}^n e_t^2$. Repita o procedimento para valores de α variando entre 0 e 1 (digamos com incrementos de 0,1) e selecione o valor que minimiza esta soma de quadrados. Na prática, o valor mínimo pode ocorrer muito próximo de um dos extremos do intervalo de variação de α . Isto pode ocorrer quando a soma de quadrados varia muito pouco na região em torno do mínimo. Neste caso faz mais sentido utilizar valores não tão extremos.

Exemplo 5.1: No banco de dados do R, a série `lh` contém as quantidades de um tipo de hormônio em amostras de sangue coletadas a cada 10 minutos de uma pessoa do sexo feminino (Diggle 1990). Vamos aplicar o método de alisamento exponencial simples à esta série fazendo primeiro a seleção do valor de α que minimiza a soma dos quadrados dos erros de previsão 1 passo à frente. Na Figura 5.1 temos o gráfico desta soma de quadrados como função de α e o gráfico das previsões 1 passo à frente juntamente com a série observada.

O valor ótimo obtido foi $\alpha = 0,945$ com a soma de erros quadrados igual a 11,86 e os seguintes comandos do R podem ser utilizados para a seleção de α .

```
> AES = function(x, interval) {
+   e = NULL
+   for (alfa in interval) {
+     e2 = 0
+     prev = x[1]
+     for (i in 2:length(x)) {
+       prev = c(prev, alfa * x[i - 1] + (1 - alfa) * prev[i -
+         1])
+       e2 = e2 + (x[i] - prev[i])^2
+     }
+     e = c(e, e2)
+   }
+   plot(interval, e, type = "l", xlab = expression(alpha), ylab = "Soma de quadrados")
+   e.min = min(e)
```

```

+   alfa = interval[e == e.min]
+   prev = x[1]
+   for (i in 2:length(x)) prev = c(prev, alfa * x[i - 1] + (1 -
+     alfa) * prev[i - 1])
+   return(list(alfa = alfa, sq2 = e.min, prev = prev))
+ }

> par(mfrow = c(2, 1))
> m = AES(lh, seq(0.1, 0.99, 0.001))
> plot(1:48, m$prev, ylab = "Hormonio", xlab = "Amostras", type = "l")
> points(lh)

```

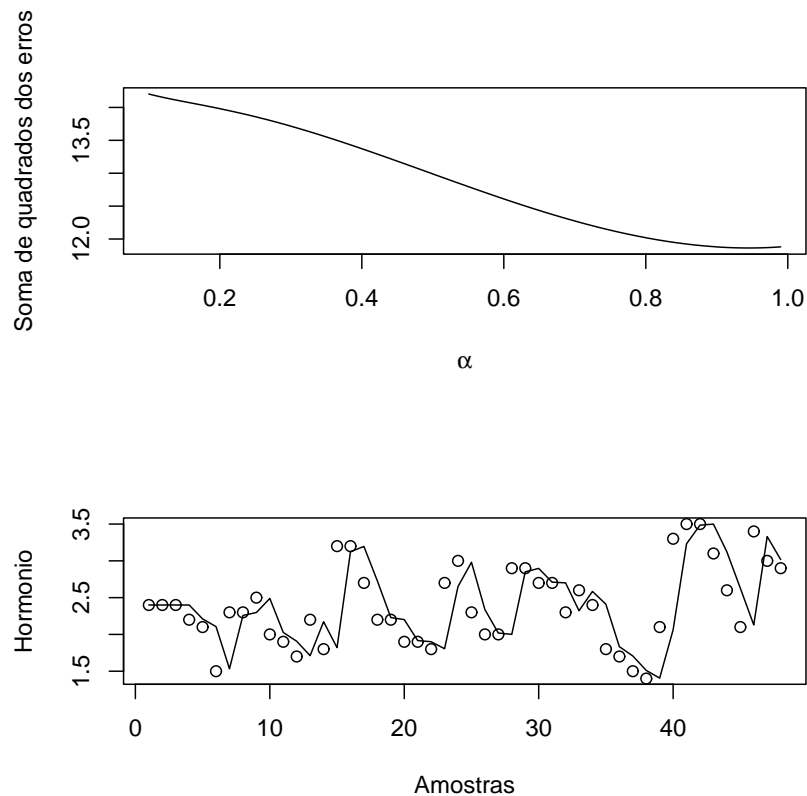


Figura 5.1: Soma dos quadrados dos erros de previsão 1 passo a frente em função de α . Valores observados (pontos) e previsões 1 passo a frente (linhas) usando o valor ótimo de α .

Exemplo 5.2: O procedimento do Exemplo 5.1 foi repetido para a série de medidas anuais de vazões do Rio Nilo entre 1871 e 1970, também do banco de dados do R. Os resultados estão na Figura 5.2.

```

> par(mfrow = c(2, 1))
> m = AES(Nile, seq(0.1, 0.99, 0.001))
> plot(1:length(Nile), m$prev, ylab = "", xlab = "", type = "l")
> points(1:length(Nile), Nile)

```

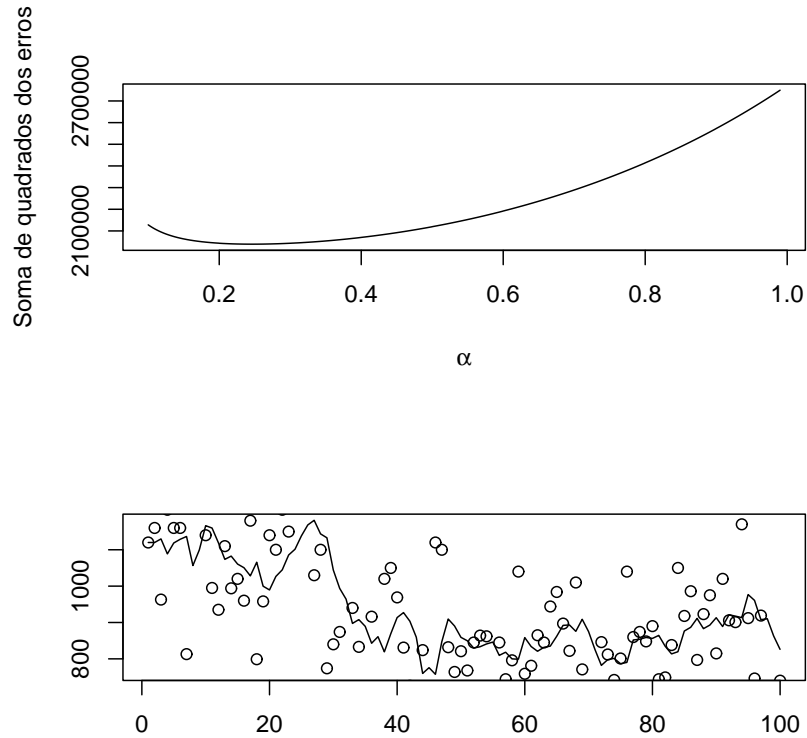


Figura 5.2: Soma dos quadrados dos erros de previsão 1 passo a frente em função de α . Valores observados (pontos) e previsões 1 passo a frente (linhas) usando o valor ótimo de α

5.1.2 Método de Holt-Winters

O procedimento de alisamento exponencial pode ser generalizado para séries que contenham tendência e variação sazonal. Suponha por exemplo que as observações são mensais e sejam L_t , T_t e I_t o nível, a tendência e o índice sazonal no tempo t . Assim, T_t é o aumento ou redução esperada por mês no nível atual da série.

Suponha que no tempo t os termos $(L_1, T_1, I_1), \dots, (L_{t-1}, T_{t-1}, I_{t-1})$ sejam conhecidos. Então, após observar x_t os termos L_t , T_t e I_t são atualizados via alisamento exponencial. Se a variação sazonal for multiplicativa, i.e. com amplitudes que tendem a crescer ao longo do tempo, as equações de atualização na forma de recorrência são

dadas por

$$\begin{aligned} L_t &= \alpha(x_t/I_{t-12}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \\ I_t &= \delta(x_t/L_t) + (1 - \delta)I_{t-12} \end{aligned}$$

e as previsões k períodos à frente são dadas por

$$\hat{x}_t(k) = (L_t + kT_t)I_{t-12+k}, \quad k = 1, 2, \dots$$

No caso de sazonalidade aditiva as equações de atualização para o nível e o índice sazonal são modificadas para

$$\begin{aligned} L_t &= \alpha(x_t - I_{t-12}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ I_t &= \delta(x_t - L_t) + (1 - \delta)I_{t-12} \end{aligned}$$

e as previsões k períodos à frente ficam

$$\hat{x}_t(k) = L_t + kT_t + I_{t-12+k}, \quad k = 1, 2, \dots$$

Aqui temos parâmetros de alisamento, α , γ e δ , para cada componente da série que são em geral escolhidos no intervalo (0,1) e podem ser estimados minimizando-se a soma de quadrados dos erros de previsão como na seção anterior. Aqui vale também o comentário sobre valores próximos aos extremos devido à soma de quadrados variar pouco nesta região. Além disso, estes parâmetros não dependem da escala das observações mas sim das propriedades temporais do nível, tendência e sazonalidade da série. Valem os mesmos comentários sobre estes valores refletindo a influência das observações passadas nas previsões de cada componente.

Para o caso particular de séries sem variação sazonal basta utilizar as equações para L_t e T_t acima (sem o índice I_{t-12}). Ou seja,

$$\begin{aligned} L_t &= \alpha x_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \\ T_t &= \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1} \end{aligned}$$

e a previsão k passos à frente no tempo t é simplesmente $L_t + kT_t$. Se a série também não tem uma tendência sistemática retorna-se à equação (5.3), ou seja

$$L_t = \alpha x_t + (1 - \alpha)L_{t-1}$$

e L_t é a previsão 1 passo à frente ($\hat{x}_t(1)$).

Exemplo 5.3: A variável `UKLungDeaths` contém os números mensais de mortes por doenças do pulmão (bronquite, efisema e asma) no Reino Unido entre janeiro de 1974 e dezembro de 1979. A variável é composta por 3 séries: ambos os sexos (`ldeaths`), sexo feminino (`fdeaths`) e sexo masculino (`mdeaths`). Aqui será utilizada a função `HoltWinters` do R que faz o alisamento exponencial de Holt-Winters com a série `ldeaths`. As constantes de alisamento (α , β e γ) são determinadas minimizando a soma

dos quadrados dos erro de previsão 1 passo à frente. Considere um modelo sazonal aditivo. O resultado são as constantes de alisamento calculadas e as Estimativas finais (em $t = n$) do nível, tendência e componentes sazonais. Pode-se também obter as previsões e intervalos de previsão (supondo normalidade) para modelos ajustados pelo método de Holt-Winters. No gráfico da Figura 5.3 temos a série original com a série suavizada mais as previsões para os anos de 1980, 1981 e 1982 da série ldeaths.

```
> data(UK LungDeaths)
> m = HoltWinters(ldeaths, seasonal = "addit")
> p = predict(m, n.ahead = 12, prediction.interval = T)

> plot(m, p)
```

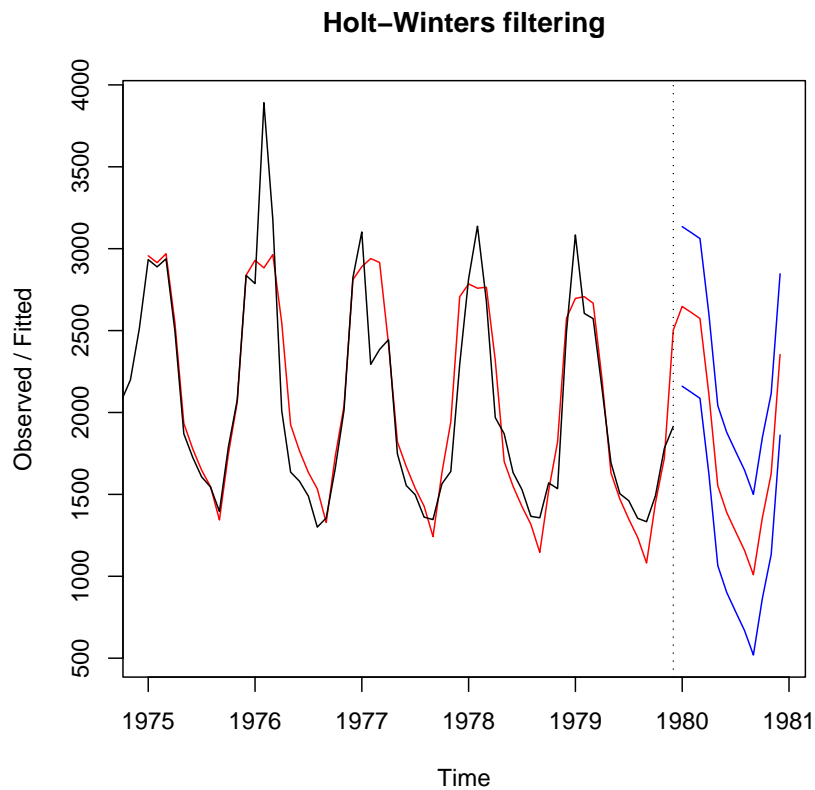


Figura 5.3: Série original, série suavizada e previsões para o ano de 1980 da série ldeaths via método de Holt-Winters.

5.2 Previsão em Modelos ARMA

Em modelos ARMA as previsões podem ser obtidas usando-se diretamente a equação do modelo. Assumindo que a equação do modelo seja conhecida a previsão $\hat{x}_n(k)$ é obtida substituindo valores futuros dos erros ϵ por zero, valores futuros da série

X_{n+1}, X_{n+2}, \dots pela sua esperança condicional, e valores passados de X e de ϵ pelos seus valores observados.

Tomemos como exemplo o modelo $\text{SARIMA}(1, 0, 0) \times (0, 1, 1)_{12}$. A equação do modelo é dada por

$$(1 - \alpha B)(1 - B^{12})X_t = (1 + \theta B^{12})\epsilon_t$$

ou equivalentemente

$$X_t = X_{t-12} + \alpha(X_{t-1} - X_{t-13}) + \epsilon_t + \theta\epsilon_{t-12}.$$

Neste caso, as previsões 1 e 2 passos à frente ficam

$$\begin{aligned}\hat{x}_n(1) &= x_{n-11} + \alpha(x_n - x_{n-12}) + \theta\epsilon_{n-11} \\ \hat{x}_n(2) &= x_{n-10} + \alpha(\hat{x}_n(1) - x_{n-11}) + \theta\epsilon_{n-10}.\end{aligned}$$

Note como o valor futuro X_{n+1} foi substituído na segunda equação pela sua esperança condicional $\hat{x}_n(1)$, i.e. a previsão feita em $t = n$ para $t = n + 1$. Previsões para horizontes maiores podem ser obtidas recursivamente.

No caso de modelos autoregressivos $\text{AR}(p)$ não é difícil verificar como fica a função de previsão.

$$\begin{aligned}\hat{x}_t(1) &= \alpha_1 x_t + \dots + \alpha_p x_{t-p+1} \\ \hat{x}_t(2) &= \alpha_1 \hat{x}_t(1) + \dots + \alpha_p x_{t-p+2} \\ &\vdots \\ \hat{x}_t(p+1) &= \alpha_1 \hat{x}_t(p) + \dots + \alpha_p \hat{x}_t(1)\end{aligned}$$

de modo que as previsões para horizontes maiores do que p usam apenas as previsões anteriores. Para $p = 1$ por exemplo segue que

$$\hat{x}_t(k) = \alpha \hat{x}_t(k-1) = \alpha^2 \hat{x}_t(k-2) = \dots = \alpha^k x_t$$

Para modelos médias móveis $\text{MA}(q)$ também não é difícil verificar que a equação de previsão fica

$$\begin{aligned}\hat{x}_t(1) &= \beta_1 \epsilon_t + \dots + \beta_q \epsilon_{t-q+1} \\ \hat{x}_t(2) &= \beta_2 \epsilon_t + \dots + \beta_q \epsilon_{t-q+2} \\ &\vdots \\ \hat{x}_t(q) &= \beta_q \epsilon_t \\ \hat{x}_t(q+j) &= 0, \quad j = 1, 2, \dots\end{aligned}$$

ou seja,

$$\hat{x}_t(k) = \begin{cases} \sum_{i=k}^q \beta_i \epsilon_{t+k-i}, & k = 1, \dots, q \\ 0, & k > q \end{cases}$$

Atualização das Previsões

É interessante notar também como as previsões podem ser atualizadas conforme novas observações da série forem obtidas. Suponha por exemplo que o valor x_{n+1} foi observado. Neste caso a previsão para $t = n + 2$ ficará condicionada em x_1, \dots, x_n, x_{n+1} e pode ser facilmente atualizada para a nova origem $n + 1$. Para o modelo SARIMA visto acima a previsão fica

$$\begin{aligned}\hat{x}_{n+1}(1) &= E(X_{n+2}|x_{n+1}, \dots, x_1) \\ &= x_{n-10} + \alpha(x_{n+1} - x_{n-11}) + \theta\epsilon_{n-10}.\end{aligned}\tag{5.4}$$

Somando e subtraindo $\alpha\hat{x}_n(1)$ no lado direito de (5.4) obtemos que

$$\begin{aligned}\hat{x}_{n+1}(1) &= x_{n-10} + \alpha(\hat{x}_n(1) - x_{n-11}) + \alpha(x_{n+1} - \hat{x}_n(1)) + \theta\epsilon_{n-10} \\ &= \hat{x}_n(2) + \alpha(x_{n+1} - \hat{x}_n(1))\end{aligned}$$

ou seja, a previsão atualizada é a previsão anterior mais uma proporção do erro de previsão 1 passo à frente em $t = n + 1$.

Previsões usando a forma MA

As previsões também podem ser obtidas reescrevendo-se o modelo como um processo médias móveis de ordem infinita. Neste caso temos que

$$X_{n+k} = \epsilon_{n+k} + \psi_1\epsilon_{n+k-1} + \dots + \psi_k\epsilon_n + \psi_{k+1}\epsilon_{n-1} + \dots$$

e fica claro que a previsão k passos à frente é dada por

$$\hat{x}_n(k) = \psi_k\epsilon_n + \psi_{k+1}\epsilon_{n-1} + \dots\tag{5.5}$$

Note que apenas os valores $\epsilon_n, \epsilon_{n-1}, \dots$ foram utilizados já que a esperança dos valores futuros é igual a zero. Esta forma é particularmente útil para o cálculo da variância do erro de previsão. Da equação (5.5) obtemos que o erro de previsão k passos à frente é dado por

$$x_{n+k} - \hat{x}_n(k) = \epsilon_{n+k} + \psi_1\epsilon_{n+k-1} + \dots + \psi_{k-1}\epsilon_{n+1}$$

e portanto a variância do erro de previsão é dada por

$$Var(e_{t+k}) = (1 + \psi_1^2 + \dots + \psi_{k-1}^2)\sigma_\epsilon^2.$$

O ponto importante a se notar aqui é que, para σ_ϵ^2 fixo, a variância do erro de previsão aumenta com o horizonte de previsão. Na prática, isto significa ter mais confiança em previsões de curto prazo do que em previsões de longo prazo.

Até agora não havíamos assumido nenhuma distribuição de probabilidade para os erros. Assumindo também que a sequência $\{\epsilon_t\}$ seja normalmente distribuída pode-se

construir intervalos de confiança para X_{t+k} simétricos em torno das previsões. Estes são chamados intervalos de previsão e são dados por

$$\hat{x}_t(k) \pm z_{\alpha/2} \sqrt{\left(1 + \sum_{j=1}^{k-1} \psi_j^2\right) \sigma_\epsilon^2}.$$

É claro que neste caso a hipótese de normalidade precisa ser checada.

Finalmente, vale notar que na prática os parâmetros do modelo não são conhecidos de forma exata e precisam ser estimados. Os valores passados dos erros ϵ_t também precisam ser estimados como erros de previsão um passo à frente. Assim, por exemplo para o modelo SARIMA(1,0,0) \times (0,1,1)₁₂ visto acima teremos que

$$\hat{x}_n(1) = x_{n-11} + \hat{\alpha}(x_n - x_{n-12}) + \hat{\theta}\hat{\epsilon}_{n-11}$$

onde o erro de previsão 1 passo à frente em $n - 11$ é dado por

$$\hat{\epsilon}_{n-11} = x_{n-11} - \hat{x}_{n-12}(1).$$

Além disso, os intervalos de previsão obtidos serão intervalos aproximados devido a esta substituição.

Exemplo 5.4: A Figura 5.4 mostra uma série temporal com os totais mensais de mortes por acidente nos Estados Unidos entre janeiro de 1973 e dezembro de 1978. Suponha que foi identificado o modelo SARIMA(0,1,1) \times (0,1,1). Após a estimação, análise de resíduos e verificação da adequação do modelo foram feitas previsões para o ano de 1979, i.e. previsões 1, 2, ..., 12 passos à frente. Em julho de 1979 os valores para os primeiros 6 meses daquele ano foram disponibilizados e aparecem na Figura 5.5 juntamente com as previsões. Note como os valores observados ficaram dentro dos intervalos de previsão fornecendo assim indicação de que o modelo teve uma boa performance preditiva. Sendo assim, uma estratégia inicial para o segundo semestre de 1979 consiste em simplesmente atualizar as previsões. Os comandos do R para este exemplo são dados a seguir.

Transformações

Em muitas aplicações a série modelada é na verdade uma transformação dos dados originais, sendo a transformação logarítmica a mais usual. Assim, tanto as previsões pontuais quanto os intervalos de previsão são obtidos para a série transformada e estes valores precisam ser transformados novamente para a escala original. A abordagem mais simples (e geralmente adotada) consiste simplesmente em tomar a transformação inversa, por exemplo se um modelo foi ajustado para a série $X_t = \log Y_t$ então $\hat{y}_n(k) = \exp(\hat{x}_n(k))$ é a previsão k passos a frente da série original. No entanto deve-se ter em mente que estas previsões via transformação inversa são em geral viesadas. Felizmente os intervalos de previsão tem boas propriedades e por exemplo quanto se toma o anti-logaritmo dos limites

$$\hat{x}_n(k) \pm z_{\alpha/2} \sqrt{\text{var}(e_{n+k})}$$

```
> data(USAccDeaths)
> plot(USAccDeaths, xlab = "Anos", ylab = "Numero de mortes por acidente")
```

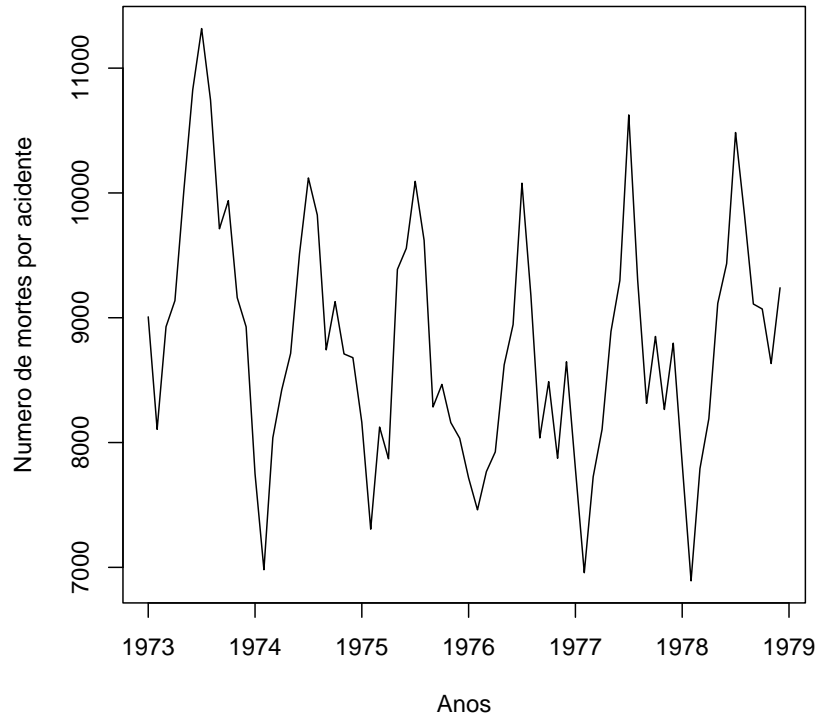


Figura 5.4: Totais mensais de mortes por acidente nos Estados Unidos entre janeiro de 1973 e dezembro de 1978.

obtém-se um intervalo (geralmente assimétrico) de $100(1 - \alpha)\%$ para a previsão pontual $\hat{y}_n(k)$.

Exemplo 5.5: Considere novamente a série *AirPassengers* e faça transformação logaritmica nos dados (conforme sugerido na literatura). Estime um modelo SARIMA(0,1,1)x(0,1,1) usando os dados até dezembro de 1960 e faça previsões de 1 até 12 meses à frente para o ano de 1961 nas 2 escalas. As previsões e intervalos de previsão na escala transformada são dados na Tabela 5.1, enquanto as previsões, intervalos de previsão e suas semi-amplitudes na escala original são dadas na Tabela 5.2.

5.3 Performance Preditiva

A idéia de verificar a adequação de um modelo em termos dos erros de previsão um passo à frente foi apresentada na Seção 4.6. Na prática é preciso verificar se os resíduos se comportam de maneira aleatória (ou imprevisível) em torno de zero e com variância

	previsão	li	ls
1961 Jan	6.11	6.04	6.18
1961 Feb	6.05	5.97	6.14
1961 Mar	6.17	6.08	6.27
1961 Apr	6.20	6.09	6.31
1961 May	6.23	6.12	6.35
1961 Jun	6.37	6.25	6.49
1961 Jul	6.51	6.38	6.64
1961 Aug	6.50	6.37	6.64
1961 Sep	6.32	6.18	6.47
1961 Oct	6.21	6.06	6.36
1961 Nov	6.06	5.91	6.22
1961 Dec	6.17	6.00	6.33

Tabela 5.1: Previsões e limites inferior (li) e superior (ls) dos intervalos de previsão.

aproximadamente constante, além de serem não correlacionados ao longo do tempo.

Além disso, dois ou mais modelos podem ser comparados segundo a sua performance preditiva, ou seja construindo-se medidas baseadas nos erros de previsão. A maioria dos métodos de previsão baseia-se na idéia de minimizar somas de quadrados ou de valores absolutos dos erros de previsão e esta é também uma medida usada para comparar a adequação de modelos alternativos. A idéia então é comparar o erro quadrático médio $\sum e_t^2/(n-m)$ ou erro absoluto médio $\sum |e_t|/(n-m)$ para diferentes modelos, onde m é o número de parâmetros a serem estimados.

Uma estratégia simples de se fazer previsões consiste em tomar a observação mais recente como a melhor previsão de um valor futuro da série, i.e. $\hat{x}_t(1) = x_t$. Note que esta é a previsão 1 passo à frente de um passeio aleatório. Assim, uma forma de medir a capacidade preditiva de um modelo consiste em comparar seus erros de previsão com aqueles do passeio aleatório. Isto pode ser feito através da chamada estatística U de Theil definida como

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} (x_{t+1} - \hat{x}_t(1))^2}{\sum_{t=1}^{n-1} (x_{t+1} - x_t)^2}}.$$

Note que valores maiores do que 1 são uma indicação de que globalmente os erros de previsão tendem a ser grandes em relação aos erros de um passeio aleatório. Esta não é uma boa característica e gostaríamos que o valor de U fosse sempre menor do que 1. Vale notar também que neste caso os erros de previsão estão sendo avaliados independente da escala dos dados.

Finalmente, vale notar que todas as medidas de capacidade preditiva citadas podem ser estendidas para erros de previsão k passos a frente.

Outra prática comum em séries temporais consiste em estimar o modelo excluindo algumas observações finais e usar o modelo estimado para fazer previsões. Neste caso as previsões podem ser comparadas com os valores observados. Por exemplo, para uma

	prev	li	ls	prev.li	ls.prev
1961 Jan	450.42	418.53	484.74	31.89	34.32
1961 Feb	425.72	390.81	463.75	34.91	38.03
1961 Mar	479.01	435.08	527.37	43.93	48.36
1961 Apr	492.40	443.00	547.32	49.41	54.92
1961 May	509.05	453.98	570.81	55.07	61.75
1961 Jun	583.34	516.02	659.45	67.33	76.11
1961 Jul	670.01	588.18	763.23	81.83	93.22
1961 Aug	667.08	581.40	765.38	85.68	98.30
1961 Sep	558.19	483.18	644.85	75.01	86.66
1961 Oct	497.21	427.59	578.17	69.62	80.96
1961 Nov	429.87	367.37	503.01	62.50	73.14
1961 Dec	477.24	405.40	561.81	71.84	84.57

Tabela 5.2: Previsões e limites inferior (li) e superior (ls) e semi-amplitudes dos intervalos de previsão.

série mensal observada ao longo de 5 anos poderia-se estimar o modelo identificado usando os primeiros 4 anos e meio (54 observações) e fazer previsões para os últimos 6 meses.

5.4 Critérios de Informação

Em muitas aplicações vários modelos podem ser julgados adequados em termos do comportamento dos resíduos. Uma forma de “discriminar” entre estes modelos competidores é utilizar os chamados critérios de informação que levam em conta não apenas a qualidade do ajuste mas também penalizam a inclusão de parâmetros extras. Assim, um modelo com mais parâmetros pode ter um ajuste melhor mas não necessariamente será preferível em termos de critério de informação. A regra básica consiste em selecionar o modelo cujo critério de informação calculado seja mínimo.

A regra mais utilizada em séries temporais é o chamado critério de informação de Akaike, denotado por AIC. A definição mais comumente utilizada é

$$AIC = -2 \log \text{verossimilhança maximizada} + 2m^1$$

onde m é o número de parâmetros (em modelos ARMA(p, q) $m = p + q + 1$). Para dados normalmente distribuídos e usando-se estimativas de máxima verossimilhança para os parâmetros pode-se mostrar que

$$AIC = n \log(\hat{\sigma}_\epsilon^2) + 2m$$

onde $\hat{\sigma}_\epsilon^2 = (1/n) \sum \hat{\epsilon}_t^2$.

¹O fator 2 é somente uma convenção e não irá alterar a seleção do modelo.

Existem outros critérios de informação que são basicamente modificações do AIC na forma de penalizar a inclusão de parâmetros extras. O mais famoso deles é o critério de informação Bayesiano, denotado por BIC e dado por

$$BIC = -2 \log \text{verossimilhança maximizada} + m \log n.$$

Note como este critério penaliza bem mais a inclusão de parâmetros do que o AIC e portanto tende a selecionar modelos mais parcimoniosos.

É sempre bom lembrar que estas medidas não têm nenhum significado quando olhadas individualmente, i.e. considerando-se um único modelo. Assim, tanto o AIC quanto o BIC podem assumir valores quaisquer, inclusive valores negativos, já que eles dependem da forma da função de verossimilhança.

Vale lembrar também que ao usar tais critérios para comparar modelos a estimação precisa ser feita no mesmo período amostral de modo que os modelos sejam comparáveis. Note também que aumentando-se o número de termos autoregressivos e/ou médias móveis, o valor de m aumenta. Assim se a inclusão de termos adicionais no modelo não melhorar sensivelmente o ajuste, então o AIC e o BIC (e qualquer outro critério de informação) serão maiores.

Para uma revisão geral destes e outros critérios de informação no contexto de séries temporais ver por exemplo Priestley (1981), Capítulo 5.

Identificação Revisitada

Vimos que as duas ferramentas básicas para identificação de modelos da classe ARIMA são as autocorrelações e autocorrelações parciais amostrais. Esta etapa envolve algum grau de arbitrariedade por parte do pesquisador ao interpretar estas funções, i.e. comparar subjetivamente seus valores amostrais com os correspondentes valores teóricos.

Uma abordagem alternativa consiste em usar os critérios de informação de um forma mais abrangente. Neste caso, um conjunto de possíveis modelos competidores é definido *a priori* e aquele que minimiza o AIC ou BIC é selecionado. Por exemplo, modelos ARMA(p, q) podem ser estimados sequencialmente variando os valores de p e q entre 0 e 3 digamos. Note que neste caso teremos 16 possíveis modelos sendo comparados e os critérios de informação são agora funções de p e q . Analogamente, para modelos AR(p) podemos variar o valor de p , digamos entre 1 e 10.

Na prática este procedimento pode ser aplicado de forma semi-automática já que muitos pacotes estatísticos fornecem estes valores. Porém após um modelo ser selecionado a análise residual ainda deve ser feita antes de se passar à etapa das previsões.

Outro problema de ordem prática é que pode haver dois ou mais modelos com AIC e/ou BIC muito similares de modo que não seja trivial discriminar entre eles. Nestas situações Burnham & Anderson (1998), Seção 4.2, sugerem o uso de pesos que são obtidos subtraindo-se o valor associado com o “melhor” modelo. Os pesos relativos ao AIC são dados por

$$w_k \propto \exp(-\Delta AIC(k)/2)$$

sendo $\Delta AIC(k) = AIC(k) - \min(AIC)$ e k é a ordem do modelo. Estes pesos são então normalizados para somarem 1 de modo que $0 < w_k < 1$ e a comparação entre os modelos fica mais fácil. Se M é o número total de modelos a comparação é então baseada em

$$w_i^* = \frac{w_i}{\sum_{j=1}^M w_j}, \quad i = 1, \dots, M.$$

Por exemplo, para modelos $AR(p)$ os pesos relativos ao AIC são dados por

$$w_p \propto \exp(-\Delta AIC(p)/2), \quad p = 1, \dots, p_{\max}$$

sendo $\Delta AIC(p) = AIC(p) - \min(AIC)$ e p_{\max} deve ser especificado.

Exemplo 5.6: Na Figura 5.6 é apresentada a série com os totais anuais de lincas canadenses capturados em armadilhas entre 1821 e 1934. Estes dados têm sido modelados na literatura após uma transformação que consiste em tomar o logaritmo na base 10 e subtrair a média dos dados transformados. Vamos ajustar modelos $AR(p)$ com p variando de 1 até 5 e calcular os critérios de informação e os respectivos pesos para cada modelo. Os resultados estão na Tabela 5.3. Note que há falta de concordância entre os critérios de informação quanto ao melhor modelo. Isto pode ser uma indicação de que na verdade há 2 modelos descrevendo bem os dados. Outro problema é que o AIC seleciona um modelo com o valor máximo de p e isto pode indicar a necessidade de considerar mais termos autoregressivos. Repetindo o exercício com p variando de 1 a 15 obteve-se a Tabela 5.4.

	p	AIC	pesos AIC	BIC	pesos BIC
1	1	-242.3913	0.0000	-234.9189	0.0000
2	2	-333.0988	0.1057	-321.8902	0.8137
3	3	-332.7283	0.0878	-317.7835	0.1044
4	4	-335.6596	0.3802	-316.9786	0.0698
5	5	-335.8881	0.4263	-313.4709	0.0121

Tabela 5.3: Critérios de informação AIC e BIC e respectivos pesos para modelos $AR(p)$ ajustados a série Lynx.

Os comandos do R utilizados no Exemplo 5.6 seguem abaixo.

```
> y = log10(lynx)
> x = y - mean(y)
> p = 1:15
> n = length(x)
> crit = matrix(0, nrow = length(p), ncol = 5)
> colnames(crit) = c("p", "AIC", "pesos AIC", "BIC", "pesos BIC")
> crit[, 1] = p
> for (k in p) {
+   ar = arima(x, order = c(k, 0, 0), include.mean = F)
```

```

+      crit[k, 2] = n * log(ar$sigma2) + 2 * (k + 1)
+      crit[k, 4] = n * log(ar$sigma2) + (k + 1) + (k + 1) * log(n)
+ }
> aicp = exp(-(crit[, 2] - min(crit[, 2]))/2)
> bicp = exp(-(crit[, 4] - min(crit[, 4]))/2)
> crit[, 3] = aicp/sum(aicp)
> crit[, 5] = bicp/sum(bicp)

```

	p	AIC	pesos AIC	BIC	pesos BIC
1	1	-242.3913	0.0000	-234.9189	0.0000
2	2	-333.0988	0.0000	-321.8902	0.8100
3	3	-332.7283	0.0000	-317.7835	0.1039
4	4	-335.6596	0.0000	-316.9786	0.0695
5	5	-335.8881	0.0000	-313.4709	0.0120
6	6	-334.4484	0.0000	-308.2950	0.0009
7	7	-338.8427	0.0001	-308.9531	0.0013
8	8	-338.8505	0.0001	-305.2247	0.0002
9	9	-338.3849	0.0001	-301.0229	0.0000
10	10	-341.8678	0.0006	-300.7696	0.0000
11	11	-354.5690	0.3581	-309.7346	0.0019
12	12	-354.7117	0.3846	-306.1411	0.0003
13	13	-353.0609	0.1685	-300.7541	0.0000
14	14	-351.0895	0.0629	-295.0465	0.0000
15	15	-349.2335	0.0249	-289.4543	0.0000

Tabela 5.4: Critérios de informação AIC e BIC e respectivos pesos para modelos AR(p) ajustados a série Lynx.

Finalmente vale notar que se o número de modelos candidatos for muito grande e a série analisada muito longa o custo computacional deste método pode ser muito alto. Por exemplo, em modelos SARIMA com $p_{\max} = q_{\max} = 5$, $P_{\max} = Q_{\max} = 2$ e $d_{\max} = D_{\max} = 2$ teremos mais de 500 modelos candidatos, sem contar possíveis transformações nos dados, diferentes distribuições dos erros, presença de dados discrepantes, alterações estruturais, etc.

5.5 Previsões Usando Todos os Modelos

Suponha que existem k modelos “candidatos” denotados por M_1, M_2, \dots, M_k e deseje-se fazer a previsão de X_{n+h} . Tratando tanto X_{n+h} quanto M_i como variáveis aleatórias então pelas regras de esperança condicional segue que

$$\hat{x}_n(h) = E(X_{n+h}|\mathbf{x}) = \sum_{i=1}^k E(X_{n+h}|\mathbf{x}, M_i)P(M_i|\mathbf{x}).$$

Ou seja, podemos escrever a previsão pontual como uma mistura discreta de previsões pontuais sob cada modelo considerado.

A mesma lógica se aplica a qualquer função de X_{n+h} , em particular

$$E(X_{n+h}^2|\mathbf{x}) = \sum_{i=1}^k E(X_{n+h}^2|\mathbf{x}, M_i)P(M_i|\mathbf{x}).$$

que pode ser usado para quantificar a incerteza sobre X_{n+h} , i.e.

$$\begin{aligned} Var(X_{n+h}|\mathbf{x}) &= E(X_{n+h}^2|\mathbf{x}) - [E(X_{n+h}|\mathbf{x})]^2 \\ &= \sum_{i=1}^k E(X_{n+h}^2|\mathbf{x}, M_i)P(M_i|\mathbf{x}) - [E(X_{n+h}|\mathbf{x})]^2 \\ &= \sum_{i=1}^k [Var(X_{n+h}|\mathbf{x}, M_i) + E^2(X_{n+h}|\mathbf{x}, M_i)]P(M_i|\mathbf{x}) - [\hat{x}_n(h)]^2 \end{aligned}$$

Um procedimento para fazer previsões usando todos os modelos estimados consiste em substituir as probabilidades $P(M_i|\mathbf{x})$ pelos pesos w_i padronizados. Por exemplo, para modelos autoregressivos se p_{\max} é o número máximo de defasagens então

$$E(X_{n+h}|\mathbf{x}) = \sum_{i=1}^{p_{\max}} E(X_{n+h}|\mathbf{x}, AR(i))w_i^*.$$

5.6 Previsão Bayesiana

Na prática, os métodos de previsão em modelos ARIMA são aplicados substituindo-se os parâmetros do modelo pelas suas estimativas pontuais. Porém o fato de não conhecermos os valores dos parâmetros é mais uma fonte de incerteza em relação as previsões e que em muitas situações pode ser muito grande para ser ignorada.

No contexto Bayesiano esta incerteza pode ser levada em conta já que a previsão de valores futuros é feita a partir da distribuição preditiva de X_{n+h} , que é dada por

$$p(x_{n+h}|\mathbf{x}) = \int p(x_{n+h}|\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.$$

Neste caso, todos os possíveis valores de $\boldsymbol{\theta}$ estão sendo levados em conta e não apenas a sua estimativa pontual.

Exercícios

1. No alisamento exponencial simples descreva a papel do parâmetro α .
2. No método de Holt-Winters descreva o papel dos parâmetros α , γ e δ .
3. Explique em que situações seriam usados os métodos de Holt-Winters aditivo ou multiplicativo.

4. Seja o modelo MA(1), $X_t = \epsilon_t + \theta\epsilon_{t-1}$.
 - (a) Obtenha a previsão 1 passo à frente em $t = n$ e mostre que as previsões k passos à frente para $k = 2, 3, \dots$ são iguais a zero.
 - (b) Mostre que a variância do erro de previsão k passos à frente é dada por σ_ϵ^2 para $k = 1$ e $(1 + \theta^2)\sigma_\epsilon^2$ para $k = 2, 3, \dots$.
5. Seja o modelo $X_t = 90 + \epsilon_t + 0,8\epsilon_{t-1} + 0,5\epsilon_{t-2}$.
 - (a) Obtenha as previsões k passos à frente em $t = n$.
 - (b) Obtenha a variância do erro de previsão k passos à frente.
6. Seja o modelo AR(1), $X_t = \alpha X_{t-1} + \epsilon_t$.
 - (a) Mostre que a previsão k passos à frente feita em $t = n$ é dada por $\alpha^k x_n$.
 - (b) Mostre que a variância do erro de previsão k passos à frente é dada por $\sigma_\epsilon^2(1 - \alpha^{2k})/(1 - \alpha^2)$.
7. Para o modelo SARIMA(0, 0, 1) \times (1, 1, 0)₁₂ obtenha as previsões no tempo $t = n$ para até 12 períodos à frente em termos das observações e resíduos até o tempo $t = n$.
8. Seja o modelo $(1 - B)(1 - 0,2B)X_t = (1 - 0,5B)\epsilon_t$.
 - (a) Obtenha as previsões 1 e 2 passos à frente.
 - (b) Mostre que as previsões 3 ou mais passos à frente são dadas pela equação recursiva $\hat{x}_n(k) = 1,2\hat{x}_n(k-1) - 0,2\hat{x}_n(k-2)$.
 - (c) Obtenha a variância dos erros de previsão 1, 2 e 3 passos à frente.
 - (d) Obtenha a previsão $\hat{x}_n(2)$ e o erro padrão do erro de previsão sabendo que $\epsilon_n = 1$, $x_n = 4$, $x_{n-1} = 3$ e $\sigma_\epsilon^2 = 2$.
9. Seja o modelo ARIMA(1,0,1) para uma série X_t com média zero.
 - (a) Reescreva o modelo na forma de choques aleatórios, i.e.

$$X_t = \epsilon_t + \psi_1\epsilon_{t-1} + \psi_2\epsilon_{t-2} + \dots$$

obtendo uma expressão geral para os coeficientes ψ_j .

- (b) Escreva a expressão da variância do erro de previsão $e_t(k) = x_{t+k} - \hat{x}_t(k)$.
 - (c) Obtenha as previsões $\hat{x}_t(k)$ para horizontes $k = 1$ e $k > 1$.
10. Sabe-se que se $Y \sim N(\mu, \sigma^2)$ então $X = \exp(Y)$ tem distribuição log-normal com $E(X) = \exp(\mu + \sigma^2/2)$ e $Var(X) = e^{2\mu+2\sigma^2}(e^{\sigma^2} - 1)$. Se foram obtidas as previsões k passos à frente de $Y_t = \log(X_t)$ e assumindo que Y_t é normal mostre que as previsões na escala original são dadas por

$$\hat{X}_t(k) = \exp(\hat{Y}_t(k) + V_y(k)/2)$$

com variância

$$\exp(2\hat{Y}_t(k) + V_y(k)) [\exp(V_y(k)) - 1].$$

11. Deseja-se ajustar um modelo ARMA a uma série temporal estacionária mas os gráficos das funções de autocorrelação e autocorrelação parcial são pouco informativos. Descreva um procedimento de identificação alternativo (você tem um pacote estatístico para fazer as contas).
12. Descreva um procedimento para obter previsões h passos à frente em modelos autoregressivos com número máximo de defasagens igual a k_{\max} utilizando todos os modelos estimados. Ilustre situações em que as previsões pontuais médias devem muito similares (ou muito diferentes) das previsões usando somente o melhor modelo.

```

> plot(ts(c(USAccDeaths, pacc$pred), frequency = 12, start = c(1973,
+ 1)), xlab = "Anos", ylab = "Numero de mortes por acidente",
+ ylim = c(6000, 12000))
> abline(v = 1979 - 1/12, lty = 2)
> lines(pacc$pred + 1.96 * pacc$se, lty = 2)
> lines(pacc$pred - 1.96 * pacc$se, lty = 2)
> obs79 = c(7798, 7406, 8363, 8460, 9217, 9316)
> points(1979 + (0:5)/12, obs79, pch = "*")

```

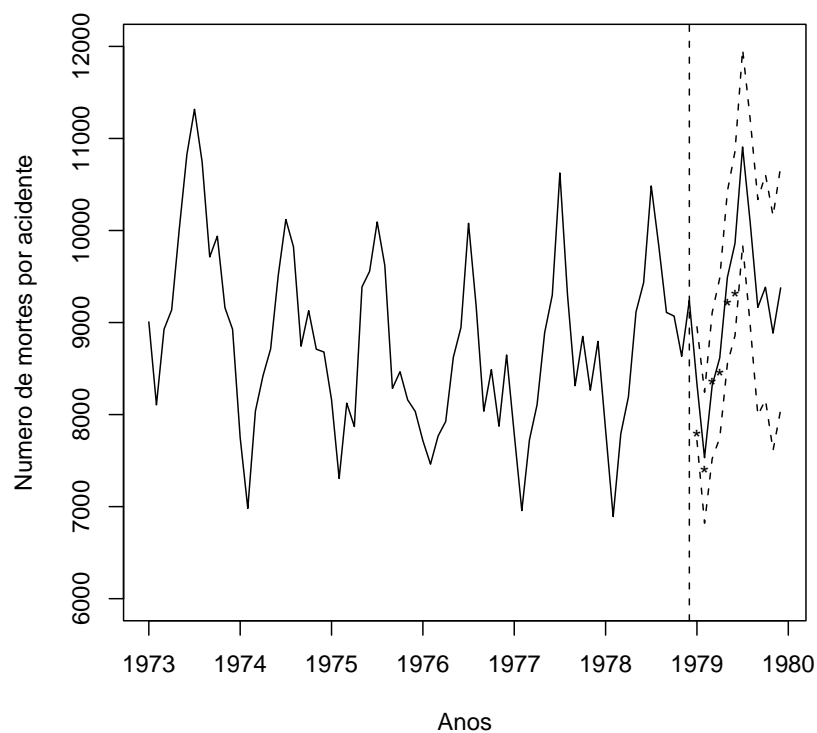


Figura 5.5: Previsões para 1979 com observações do primeiro semestre incluídas.

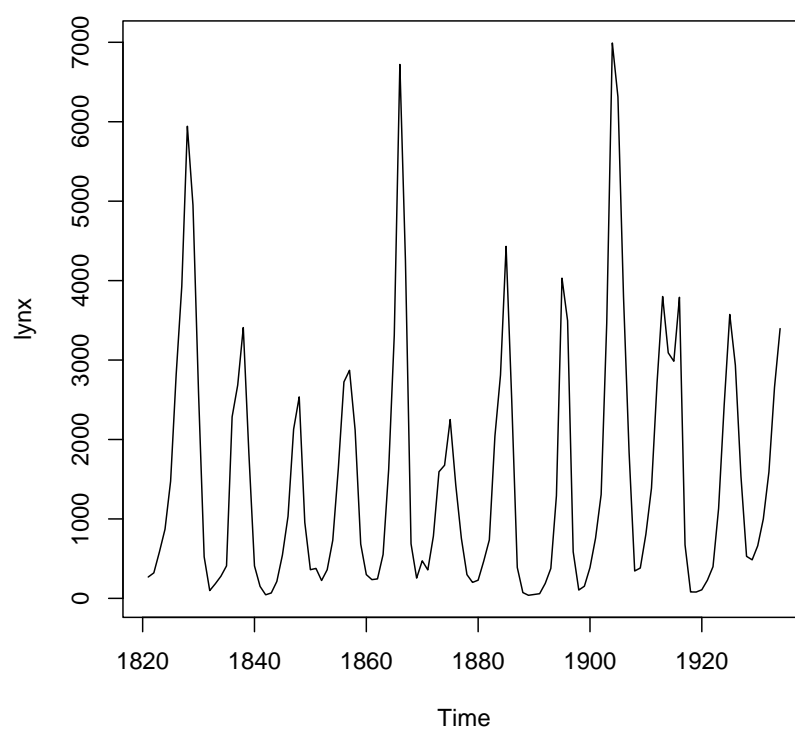


Figura 5.6: Totais anuais de lincas canadenses capturados em armadilhas entre 1821 e 1934.

Capítulo 6

Modelando a Variância

6.1 Introdução

Nos modelos vistos até aqui a variância dos erros foi assumida constante ao longo do tempo, i.e. $Var(\epsilon_t) = E(\epsilon_t^2) = \sigma_\epsilon^2$. Muitas séries temporais no entanto exibem períodos de grande volatilidade seguidos de períodos de relativa tranquilidade. Nestes casos, a suposição de variância constante (homocedasticidade) pode não ser apropriada. Na verdade, embora a variância incondicional dos erros ainda possa ser assumida constante, sua variância condicional pode estar mudando ao longo do tempo.

Além disso, em muitas situações práticas tem-se interesse em prever a variância condicional da série além da série propriamente dita. Por exemplo, no mercado de ações o interesse é não apenas prever a taxa de retorno mas também a sua variância ao longo de um certo período. Esta variância condicional é também chamada de *volatilidade*. Algumas referências para este capítulo são Taylor (1986), Franses (1998), e Tsay (2002).

Exemplo 6.1: Na Figura 6.1 os gráficos da esquerda apresentam as taxas de câmbio diárias da Libra Esterlina, Dolar Canadense, Marco Alemão e Iene Japones, em relação ao Dolar Americano, enquanto nos gráficos da direita estão os logaritmos das taxas de variação (retornos diários). O período amostral vai de janeiro de 1991 a dezembro de 1998. Uma característica comum nestes retornos é que embora as médias pareçam ser aproximadamente constantes as variâncias mudam ao longo do tempo. Na Figura 6.2 estão os histogramas com uma curva normal superimposta para os mesmos dados (retornos). Pode-se notar que muitos valores aparecem nas caudas das distribuições. Finalmente, na Figura 6.3 temos as autocorrelações amostrais dos retornos e dos retornos ao quadrado. Note como existe bastante autocorrelação entre os retornos ao quadrado. Todas estas características são em geral verificadas em séries reais de retornos e devem ser levadas em conta pelo modelo.

A idéia aqui é tentar modelar simultaneamente a média e a variância de uma série temporal. Para fixar idéias, suponha que um modelo AR(1), $X_t = \alpha X_{t-1} + \epsilon_t$ foi

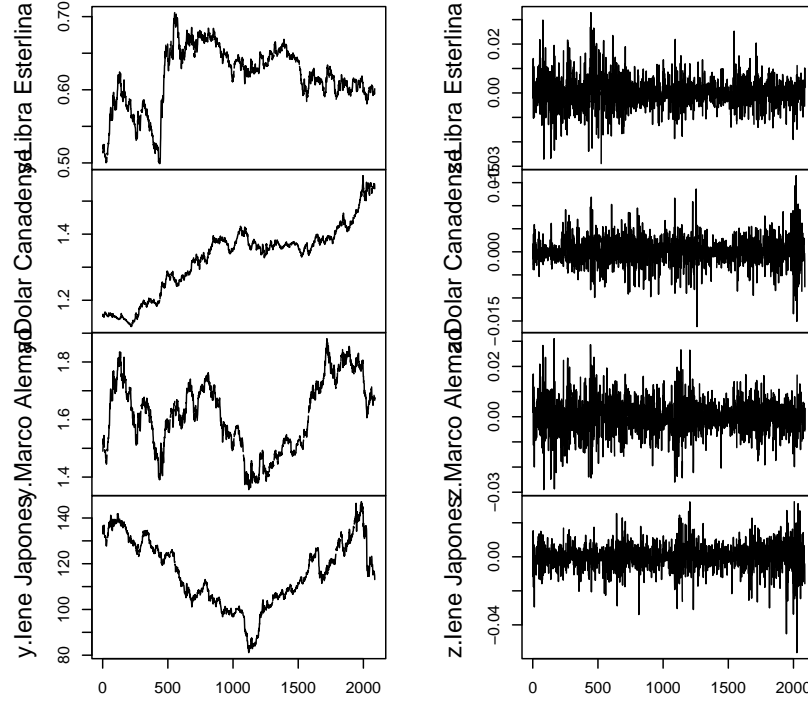


Figura 6.1: Taxas de câmbio e retornos diários em relação ao Dolar Americano da Libra Esterlina, Dolar Canadense, Marco Alemão e Iene Japones, entre janeiro de 1991 a dezembro de 1998.

estimado e deseja-se fazer previsões 1 passo à frente,

$$\hat{x}_t(1) = E(X_{t+1}|x_t) = \alpha x_t.$$

A variância condicional de X_{t+1} é dada por

$$Var(X_{t+1}|x_t) = Var(\epsilon_{t+1}|x_t) = E(\epsilon_{t+1}^2|x_t).$$

Até agora assumimos que $E(\epsilon_{t+1}^2|x_t) = \sigma_\epsilon^2$, mas suponha que a variância condicional não seja constante, i.e. $E(\epsilon_{t+1}^2|x_t) = \sigma_{t+1}^2$. Uma possível causa disto é que os dados se distribuem com caudas muito longas. Para facilitar a notação vamos denotar por $I_t = \{x_t, x_{t-1}, \dots, \epsilon_t, \epsilon_{t-1}, \dots\}$, ou seja $\sigma_t^2 = Var(\epsilon_t|I_{t-1})$.

6.2 Modelos ARCH

Existem várias formas de especificar como a variância condicional (volatilidade) varia com o tempo. Uma estratégia utilizada para modelar σ_t^2 , proposta em Engle (1982),

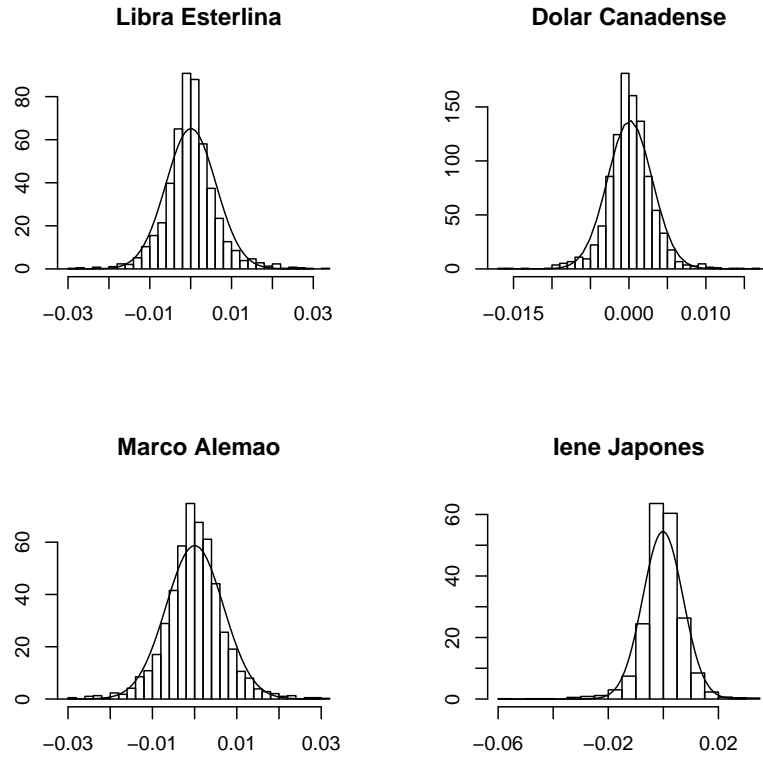


Figura 6.2: Histogramas dos retornos diários do Exemplo 6.1.

consiste em assumir que ela depende dos quadrados dos erros passados, $\epsilon_{t-1}, \epsilon_{t-2}, \dots$ através de uma autoregressão. No caso mais simples, faz-se

$$\epsilon_t = v_t \sqrt{c + \alpha \epsilon_{t-1}^2} \quad (6.1)$$

onde $\{v_t\}$ é uma série puramente aleatória com média zero e variância igual a 1 e v_t e ϵ_t são independentes. Segue que a esperança e a variância condicionais são dadas por,

$$\begin{aligned} E(\epsilon_t | I_{t-1}) &= E(v_t) \sqrt{c + \alpha \epsilon_{t-1}^2} = 0 \\ E(\epsilon_t^2 | I_{t-1}) &= \sigma_t^2 = c + \alpha \epsilon_{t-1}^2 \end{aligned} \quad (6.2)$$

Neste caso dizemos que a variância segue um processo *autoregressivo condicionalmente heterocedástico* de ordem 1, ARCH(1). Note que é necessário impor as restrições $c > 0$ e $\alpha \geq 0$ para que σ_t^2 seja sempre positiva. Quando $\alpha = 0$ a variância condicional é constante e ϵ_t é um processo condicionalmente homocedástico.

Além disso queremos garantir a estacionariedade da autoregressão de modo que a restrição imposta é $0 < \alpha < 1$. Note também que (6.2) não inclui um termo de erro e portanto não é um processo estocástico.

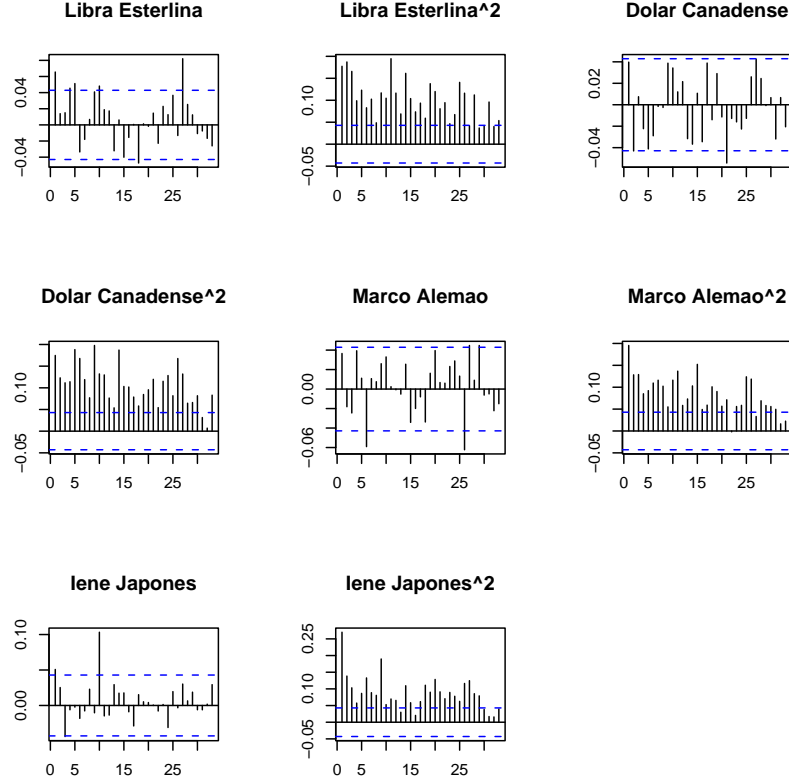


Figura 6.3: Correlogramas dos retornos e retornos ao quadrado no Exemplo 6.1

A esperança e variância incondicionais podem ser obtidas como,

$$\begin{aligned} E(\epsilon_t) &= E[E(\epsilon_t|I_{t-1})] = 0 \\ Var(\epsilon_t) &= E(\epsilon_t^2) = E[E(\epsilon_t^2|I_{t-1})] = c + \alpha E(\epsilon_{t-1}^2). \end{aligned}$$

Se o processo é estacionário então $E(\epsilon_t^2) = E(\epsilon_{t-1}^2) = Var(\epsilon_t)$ e portanto

$$Var(\epsilon_t) = \frac{c}{1 - \alpha}.$$

Além disso,

$$\begin{aligned} Cov(\epsilon_t, \epsilon_{t+k}) &= E(\epsilon_t \epsilon_{t+k}) = E[E(\epsilon_t \epsilon_{t+k}) | \epsilon_{t+k-1}, \dots, \epsilon_{t-1}] \\ &= E[\epsilon_t E(v_{t+k} \sqrt{c + \alpha \epsilon_{t+k-1}^2})] = 0, \text{ para } k > 0. \end{aligned}$$

Ou seja, ao postular o modelo ARCH(1) estamos assumindo que os $\{\epsilon_t\}$ são não correlacionados.

Exemplo 6.2 : Para ilustração a Figura 6.4 apresenta dois processos ARCH de ordem 1 simulados a partir de uma sequência $\{v_t\}$ de 200 números aleatórios i.i.d. gerados de uma distribuição $N(0, 1)$. A sequência $\{\epsilon_t\}$ foi construída usando a equação (6.1)

com $c = 1$ e $\alpha = 0,8$. Note como a sequência $\{\epsilon_t\}$ continua tendo média zero mas parece ter tido um aumento de volatilidade em alguns períodos. Em um modelo AR(1), a forma como esta estrutura nos erros afeta a série original depende do valor do parâmetro autoregressivo e duas possíveis situações são mostradas nos gráficos inferiores da figura. Na Figura 6.5 temos o histograma dos valores $\{\epsilon_t\}$ gerados, com uma curva normal superimosta, além do gráfico de probabilidades normais (QQ-plot normal). Note como há um excesso de valores nas caudas ocorrendo com uma frequência maior do que seria esperado na distribuição normal.

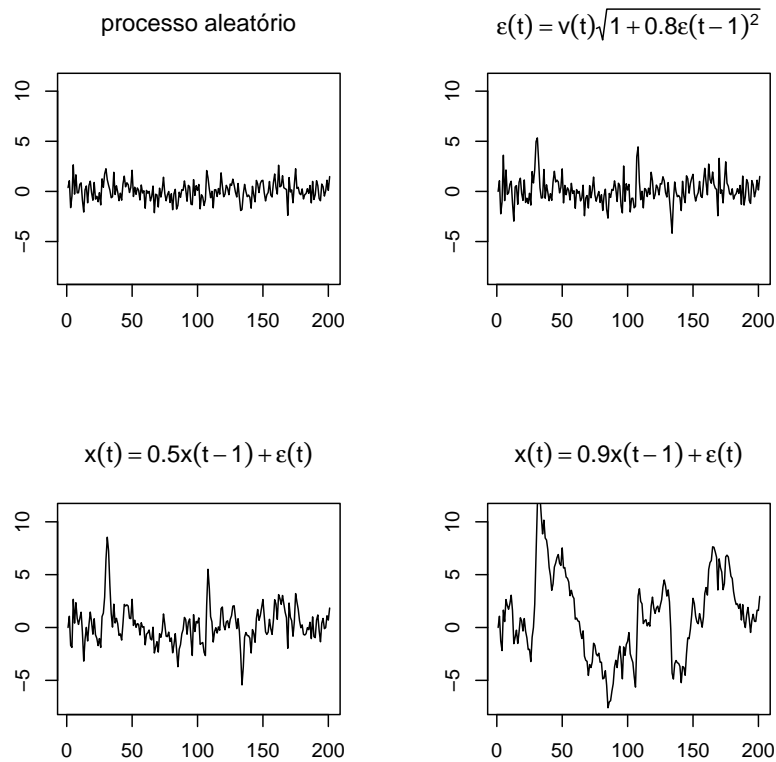


Figura 6.4: Processos autoregressivos com erros ARCH(1) simulados.

Basicamente a equação (6.2) nos diz que erros grandes (ou pequenos) em valor absoluto tendem a ser seguidos por erros grandes (ou pequenos) em valor absoluto. Portanto o modelo é adequado para descrever séries aonde a volatilidade ocorre em grupos. Além disso, na equação (6.2) somando ϵ_t^2 e subtraindo σ_t^2 de ambos os lados obtemos que

$$\epsilon_t^2 = c + \alpha \epsilon_{t-1}^2 + \nu_t$$

com $\nu_t = \epsilon_t^2 - \sigma_t^2 = \sigma_t^2(v_t^2 - 1)$. Ou seja, o modelo ARCH(1) pode ser reescrito como um AR(1) estacionário para ϵ_t^2 com erros não normais ($v_t^2 \sim \chi_1^2$ se $v_t \sim N(0, 1)$) que têm média zero e variância não constante. Portanto, a função de autocorrelação do processo $\{\epsilon_t^2\}$ é dada por $\rho(k) = \alpha^k$ e o correlograma amostral deve apresentar um

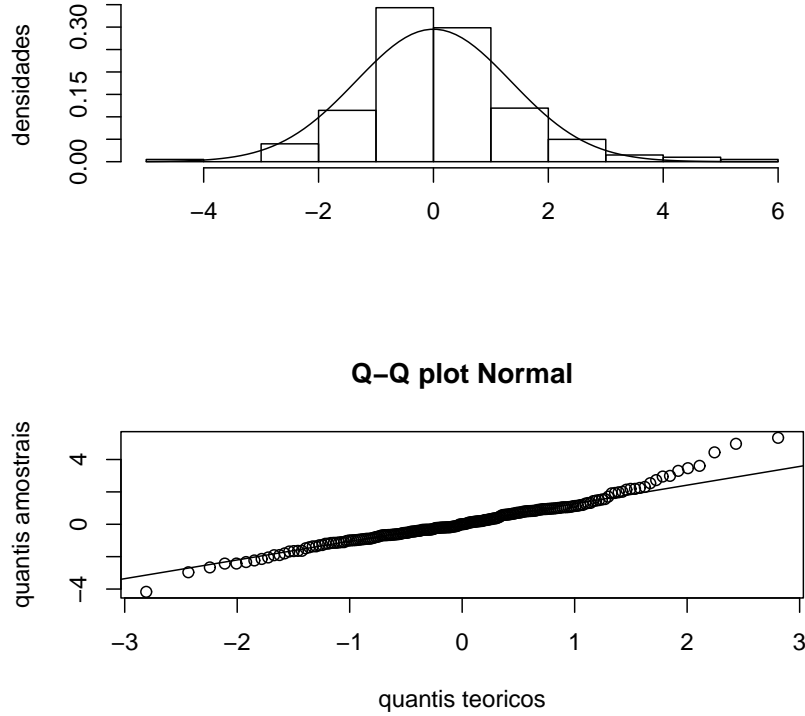


Figura 6.5: Características de um processo ARCH(1) simulado.

decaimento exponencial para zero.

Se o processo ARCH(1) for estacionário não é difícil calcular o seu coeficiente de curtose que é dado por

$$\kappa = \frac{E(\epsilon_t^4)}{[Var(\epsilon_t)]^2}.$$

Denotando por $E(v_t^4) = \lambda$ o quarto momento do erro segue que o quarto momento condicional é dado por

$$E(\epsilon_t^4 | I_{t-1}) = E(v_t^4 \sigma_t^4 | I_{t-1}) = \lambda E(\sigma_t^4 | I_{t-1}) = \lambda(c + \alpha \epsilon_{t-1}^2)^2.$$

(se assumirmos que $v_t \sim N(0, 1)$ então $\lambda = 3$). Portanto, o quarto momento incondicional fica,

$$E(\epsilon_t^4) = E[E(\epsilon_t^4 | I_{t-1})] = \lambda E(c^2 + \alpha^2 \epsilon_{t-1}^4 + 2c\alpha \epsilon_{t-1}^2).$$

Se o processo é estacionário de quarta ordem então podemos escrever $E(\epsilon_t^4) = E(\epsilon_{t-1}^4) = \mu_4$ e portanto,

$$\mu_4 = \lambda(c^2 + \alpha^2 \mu_4 + 2c\alpha \frac{c}{1-\alpha}) = \lambda c^2 \left(\frac{1+\alpha}{1-\alpha} \right) + \lambda \alpha^2 \mu_4$$

e finalmente,

$$\mu_4 = \frac{\lambda c^2 (1+\alpha)}{(1-\alpha)(1-\lambda \alpha^2)}.$$

O coeficiente de curtose então fica,

$$\kappa = \frac{\lambda c^2(1+\alpha)}{(1-\alpha)(1-\lambda\alpha^2)} \frac{(1-\alpha)^2}{c^2} = \frac{\lambda(1-\alpha^2)}{1-\lambda\alpha^2}, \quad \alpha^2 < 1/\lambda$$

que é sempre maior do que λ . Ou seja, qualquer que seja a distribuição de v_t o coeficiente de curtose será maior do que a curtose de v_t (desde que $\alpha > 0$ e $\lambda > 1$). Em particular, processos ARCH(1) têm caudas mais pesadas do que a distribuição normal e são portanto adequados para modelar séries temporais com esta característica. Séries de retornos, como as do Exemplo 6.1, frequentemente apresentam caudas mais pesadas do que a normal devido ao excesso de curtose.

Previsões da Volatilidade

Suponha que uma série temporal X_t segue um processo ARCH(1), i.e. $X_t = v_t\sqrt{h_t}$, $v_t \sim N(0,1)$. As previsões da volatilidade, k passos à frente, no tempo $t = n$ são obtidas como,

$$\hat{h}_n(k) = E(h_{n+k}|I_n) = c + \alpha E(X_{n+k-1}^2|I_n).$$

Para $k = 1$ segue que $E(X_{n+k-1}^2|I_n) = X_{n+k-1}^2$ e para $k > 1$ temos que

$$\begin{aligned} E(X_{n+k-1}^2|I_n) &= E(h_{n+k-1}v_{n+k-1}^2|I_n) \\ &= E(h_{n+k-1}|I_n)E(v_{n+k-1}^2|I_n) \\ &= E(h_{n+k-1}|I_n) = \hat{h}_n(k-1) \end{aligned}$$

pois h_{n+k-1} e v_{n+k-1} são independentes. As previsões então ficam,

$$\hat{h}_n(k) = \begin{cases} c + \alpha X_n^2, & k = 1 \\ c + \alpha \hat{h}_n(k-1), & k = 2, 3, \dots \end{cases}$$

O Modelo ARCH(p)

Estas idéias podem ser generalizadas para processos mais gerais ARCH(p) em que a variância condicional depende dos quadrados de p erros passados, i.e.

$$\epsilon_t = v_t \sqrt{c + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2} \quad (6.3)$$

e então a variância condicional é modelada como,

$$\sigma_t^2 = E(\epsilon_t^2|I_{t-1}) = c + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2.$$

Neste caso, para garantir que σ_t^2 seja sempre positiva é necessário impor as seguintes restrições $c > 0$ e $\alpha_1 \geq 0, \dots, \alpha_p \geq 0$ e para garantir estacionariedade é necessário também que as raízes de $1 - \alpha_1 B - \dots - \alpha_p B^p = 0$ estejam fora do círculo unitário. Juntando estas restrições equivale a impor a restrição $c > 0$ e $\sum_{i=1}^p \alpha_i < 1$.

Analogamente podemos reescrever o modelo ARCH(p) como um modelo AR(p) para ϵ_t^2 definindo os erros ν_t como anteriormente, i.e.

$$\epsilon_t^2 = c + \alpha_1 \epsilon_{t-1}^2 + \dots + \alpha_p \epsilon_{t-p}^2 + \nu_t.$$

com $\nu_t = \sigma_t^2(v_t^2 - 1)$.

Identificação

A característica chave dos modelos ARCH é que a variância condicional dos erros ϵ_t se comporta como um processo autoregressivo. Portanto deve-se esperar que os resíduos de um modelo ARMA ajustado a uma série temporal observada também sigam este padrão característico. Em particular, se o modelo ajustado for adequado então a FAC e a FACP dos resíduos devem indicar um processo puramente aleatório, no entanto se a FAC dos quadrados dos resíduos, ϵ_t^2 , tiver um decaimento característico de uma autoregressão isto é uma indicação de que um modelo ARCH pode ser apropriado. A ordem p do modelo pode ser identificada através da FACP dos quadrados dos resíduos.

Previsões da Volatilidade

Suponha que uma série temporal X_t segue um processo ARCH (p). As previsões da volatilidade, k passos à frente, no tempo $t = n$ são obtidas como,

$$\hat{h}_n(k) = E(h_{n+k}|I_n) = c + \sum_{i=j}^p \alpha_i E(X_{n+k-j}^2|I_n).$$

Para $k \leq j$ segue que $E(X_{n+k-j}^2|I_n) = X_{n+k-j}^2$ e para $k > j$ temos que

$$\begin{aligned} E(X_{n+k-j}^2|I_n) &= E(h_{n+k-j}v_{n+k-j}^2|I_n) \\ &= E(h_{n+k-j}|I_n)E(v_{n+k-j}^2|I_n) \\ &= E(h_{n+k-j}|I_n) = \hat{h}_n(k-j) \end{aligned}$$

já que h_{n+k-1} e v_{n+k-1} são independentes.

6.3 Modelos GARCH

Uma generalização natural dos modelos ARCH consiste em assumir que a variância condicional se comporta como um processo ARMA, i.e. depende também de seus valores passados. Fazendo $\epsilon_t = v_t\sqrt{h_t}$ onde

$$h_t = c + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}$$

segue que a esperança condicional de ϵ_t é zero e a variância condicional é $\sigma_t^2 = h_t$. Este modelo é chamado ARCH generalizado, ou GARCH, de ordem (p, q) . Aqui as restrições impostas sobre os parâmetros são dadas por $c > 0$ e $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$.

Embora a primeira vista pareça um modelo mais complexo, sua vantagem sobre os modelos ARCH é basicamente a parcimônia. Assim como um modelo ARMA pode ser mais parcimonioso no sentido de apresentar menos parâmetros a serem estimados do que modelos AR ou MA, um modelo GARCH pode ser usado para descrever a volatilidade com menos parâmetros do que modelos ARCH.

Em termos de identificação dos valores de p e q as ferramentas básicas são mais uma vez a FAC e a FACP dos quadrados dos resíduos. Assim, se o modelo ajustado for adequado a FAC e a FACP dos resíduos devem indicar um processo puramente aleatório, no entanto quando estas funções são aplicadas aos quadrados dos resíduos elas devem indicar um processo ARMA(p, q). A identificação pode não ser muito fácil em algumas aplicações embora na maioria dos casos um modelo GARCH(1,1) seja suficiente. Na prática recomenda-se também tentar outros modelos de ordem baixa como GARCH(1,2) e GARCH(2,1).

As previsões da volatilidade em modelos GARCH são obtidas de forma similar a de um modelo ARMA. Por exemplo, após estimar os parâmetros de um modelo GARCH(1,1) e assumindo-se que $\epsilon_0 = h_0 = 0$ pode-se construir as sequências $\epsilon_1, \dots, \epsilon_t$ e h_1, \dots, h_t e a previsão 1 passo à frente da volatilidade fica

$$\hat{\sigma}_t^2(1) = c + \alpha\epsilon_t^2 + \beta h_t.$$

6.3.1 Estimação

Para uma série x_1, \dots, x_n observada e um modelo GARCH(p, q), denotando-se o vetor de parâmetros por $\theta = (c, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ e destacando-se a densidade conjunta das p primeiras realizações segue que

$$p(x_1, \dots, x_n | \theta) = p(x_1, \dots, x_p | \theta) \prod_{t=p+1}^n p(x_t | x_{t-1}, \dots, x_{t-p}, \theta).$$

Assumindo normalidade segue que

$$X_t | x_{t-1}, \dots, x_{t-p} \sim N(0, h_t)$$

e portanto

$$p(x_1, \dots, x_n | \theta) = p(x_1, \dots, x_p | \theta) \prod_{t=p+1}^n (2\pi h_t)^{-1/2} \exp(-(1/2)x_t^2/h_t).$$

Em geral o número de observações será grande o suficiente para que o termo $p(x_1, \dots, x_p | \theta)$ possa ser desprezado. Por exemplo, para um modelo ARCH(1) a função log-verossimilhança fica

$$-0.5 \sum_{t=2}^n [\log(2\pi) + \log(c + \alpha x_{t-1}^2) + x_t^2 / (c + \alpha x_{t-1}^2)].$$

Note que algum algoritmo de otimização não linear deverá ser utilizado e nada garante sua convergência para um ótimo global. No R pode-se usar a função `garch` do pacote `tseries` para fazer a estimação por máxima verossimilhança.

6.3.2 Adequação

Se um modelo ARCH ou GARCH foi ajustado a uma série X_t não correlacionada então os resíduos padronizados são dados por

$$\tilde{X}_t = \frac{X_t}{\sqrt{h_t}}$$

e formam uma sequência *i.i.d.* com distribuição normal padrão. Assim, a adequação do modelo pode ser verificada aplicando os testes usuais de normalidade a estes resíduos padronizados e os testes de aleatoriedade (Box-Pierce e Ljung-Box) aos quadrados dos resíduos.

Exemplo 6.3 : Na parte superior da Figura 6.6 estão os preços diários no fechamento de um índice de mercado da Alemanha (DAX), entre 1991 e 1998. O interesse é em analisar os chamados retornos dados por $\log(x_t/x_{t-1})$ e estes estão no gráfico inferior da Figura 6.6. Existe evidência na literatura que modelos GARCH(1,1) conseguem captar bem os movimentos característicos dos retornos.

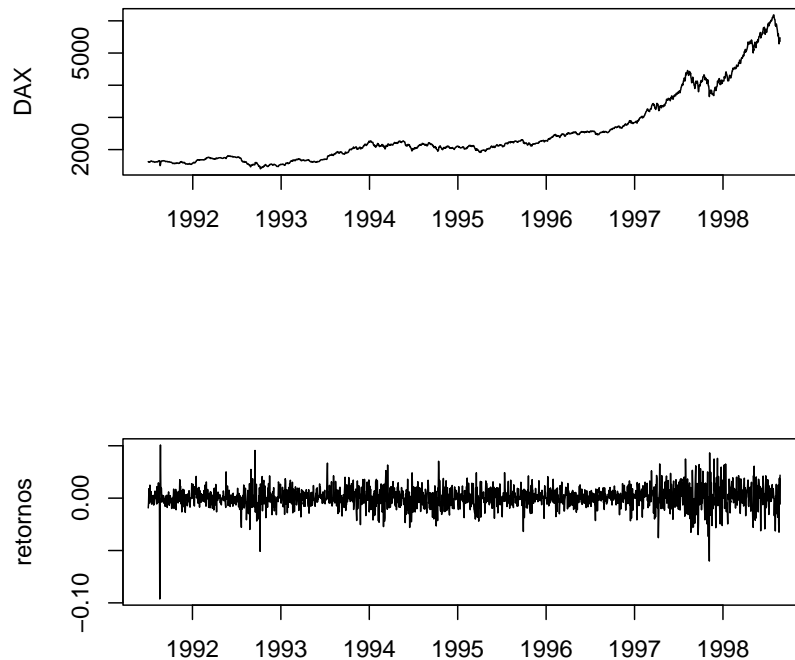


Figura 6.6: Preços diários no fechamento de um índice de mercado da Alemanha (DAX), entre 1991 e 1998 e respectivos retornos.

Usando a função `garch` no pacote `tseries` do R o modelo ajustado obtido foi

$$\begin{aligned} Y_t &= v_t \sqrt{h_t}, \quad v_t \sim N(0, 1) \\ h_t &= 0.000005 + 0.068Y_{t-1}^2 + 0.889h_{t-1} \end{aligned}$$

sendo todos os coeficientes significativos. O teste de Ljung-Box aplicado nos quadrados dos resíduos indicou aleatoriedade (p-valor = 0,71), no entanto o teste de normalidade de Jarque-Bera aplicado aos resíduos rejeitou a hipótese nula (p-valor < 0,001). Assim a hipótese de normalidade condicional parece estar sendo violada.

Na Figura 6.7 estão os histogramas, gráficos de probabilidades normais dos retornos e resíduos do modelo GARCH(1,1) estimado, além dos correlogramas dos quadrados dos retornos e resíduos. Use os comandos abaixo para estimar o modelo.

```
> library(tseries)
> data(EuStockMarkets)
> x = EuStockMarkets
> dax = diff(log(x))[, "DAX"]
> dax.garch = garch(dax, trace = F)
> r = dax.garch$residuals
> round(dax.garch$coef, 6)
```

```
      a0      a1      b1
0.000005 0.068329 0.889067
```

Um fato estilizado presente em séries temporais financeiras é que o mercado tem baixa volatilidade quando está em alta e alta volatilidade quando está em baixa. Tal assimetria não é levada em conta pelos modelos GARCH e para contornar esta limitação outros modelos foram propostos na literatura. Por exemplo, no modelo EGARCH (ou GARCH exponencial) modela-se o logaritmo da volatilidade como,

$$\log(\sigma_t^2) = c + \alpha \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta \sigma_{t-1}^2.$$

Em termos de estimação uma vantagem deste modelo é que os parâmetros c , α e β são irrestritos já que estamos modelando o logaritmo da volatilidade. A única restrição é $\gamma < 0$ pois assim a volatilidade aumenta quando $\epsilon_{t-1} < 0$.

6.4 Volatilidade Estocástica

As fórmulas para modelar σ_t^2 vistas até agora foram todas determinísticas, i.e. sem uma componente de erro aleatório. No entanto, pode ser mais razoável assumir que a variância condicional varia estocasticamente ao longo do tempo ao invés de deterministicamente, especialmente se existem mudanças abruptas na volatilidade (e.g. como resultado de greves, guerras, etc.).

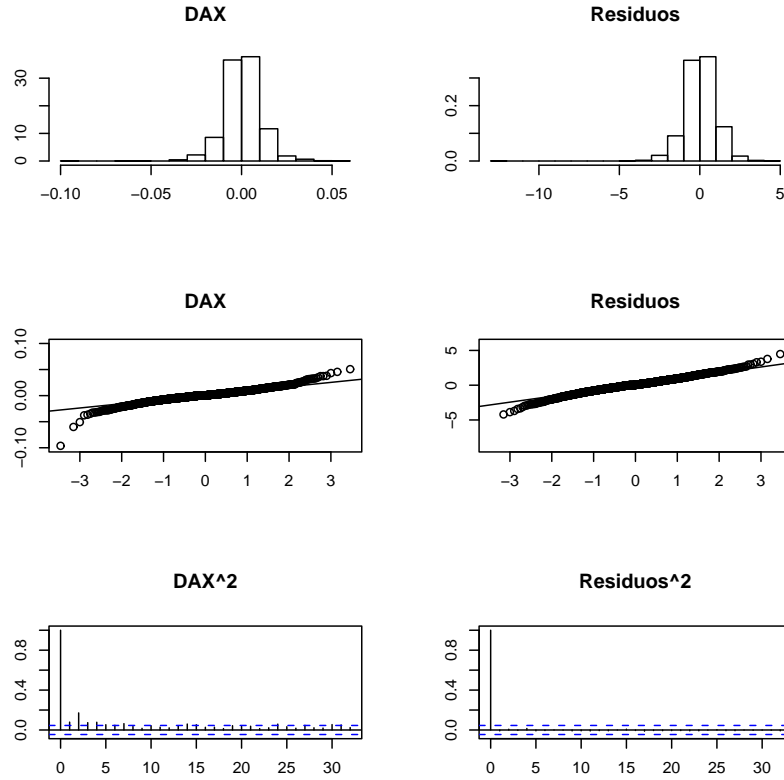


Figura 6.7: Histogramas e probabilidades normais dos retornos do índice de mercado da Alemanha (DAX) e resíduos dos modelos GARCH(1,1) e correlogramas dos seus quadrados.

Assim, uma alternativa aos modelos ARCH ou GARCH consiste em assumir que σ_t^2 segue um processo estocástico. Geralmente modela-se o logaritmo de σ_t^2 . Em sua forma mais simples um modelo de volatilidade estocástica (VE) é dado por

$$\begin{aligned} X_t &= v_t \exp(h_t/2), \quad v_t \sim N(0, 1) \\ h_t &= c + \alpha h_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \end{aligned}$$

com $|\alpha| < 1$ e $h_t = \log(\sigma_t^2)$. Note que não há necessidade de restrições de positividade nos parâmetros pois estamos modelando o logaritmo da volatilidade. O modelo pode ser estendido para uma estrutura AR(p) em h_t , ou seja

$$\begin{aligned} X_t &= v_t \exp(h_t/2), \quad v_t \sim N(0, 1) \\ h_t &= c + \sum_{i=1}^p \alpha_i h_{t-i} + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \end{aligned}$$

Propriedades

1. $E(X_t) = E(v_t e^{h_t/2}) = E(e^{h_t/2})E(v_t) = 0$, já que h_t e v_t são independentes.

2. $Var(X_t) = E(X_t^2) = E(e^{h_t} v_t^2) = E(e^{h_t})E(v_t^2) = E(e^{h_t})$. Mas, como estamos assumindo que h_t é estacionária segue que,

$$E(h_t) = \mu \quad \text{e} \quad Var(h_t) = \sigma_h^2 = \tau^2/(1 - \phi^2)$$

e a distribuição incondicional do log-volatilidade é $h_t \sim N(\mu, \sigma_h^2)$. Portanto, e^{h_t} é log-normal e segue que

$$Var(X_t) = E(e^{h_t}) = e^{\mu + \sigma_h^2/2}.$$

3. $E(X_t^4) = E(v_t^4 e^{2h_t}) = 3 e^{2\mu + 2\sigma_h^2}$ e a curtose é dada por

$$K = \frac{3 e^{2\mu + 2\sigma_h^2}}{e^{2\mu + \sigma_h^2}} = 3e^{\sigma_h^2} > 3.$$

Exercícios

1. Um modelo ARIMA foi identificado e estimado para uma série temporal observada mas há indicação de que a variância condicional deve ser modelada por um processo GARCH(1,1). Explique como se chegou a esta conclusão.
2. Refaça o exemplo da Figura 6.4 e estime um modelo AR(1) para a série X_t . Verifique se existe estrutura autoregressiva nos quadrados dos resíduos e identifique um modelo ARCH para os erros.
3. Obtenha as previsões 1, 2 e 3 passos a frente para um modelo GARCH(1,2).
4. Descreva duas vantagens de modelos EGARCH sobre modelos GARCH.

Capítulo 7

Modelos Lineares Dinâmicos

A classe de modelos lineares dinâmicos (MLD), também conhecidos como modelos de espaço de estados tem sido utilizada com sucesso em análise e previsão de séries temporais. Neste capítulo serão apresentadas as formas mais comumente utilizadas de MLD, maiores detalhes podem ser obtidos em West & Harrison (1997) e Pole, West, & Harrison (1994).

7.1 Introdução

Um modelo linear dinâmico pode ser caracterizado pelo seguinte par de equações

$$\begin{aligned}y_t &= \mathbf{F}_t' \boldsymbol{\theta}_t + \epsilon_t \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t\end{aligned}\tag{7.1}$$

chamadas equações de observação e evolução respectivamente, onde $\boldsymbol{\theta}_t$ denota o vetor de estados no tempo t , \mathbf{F} é um vetor de constantes conhecidas ou regressores, \mathbf{G} é uma matrix de evolução conhecida. Os erros ϵ_t e $\boldsymbol{\omega}_t$ são geralmente assumidos não correlacionados em todos os períodos de tempo e serialmente não correlacionados com média zero. Em muitas aplicações práticas pode-se assumir também que $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ e $\boldsymbol{\omega}_t$ tem distribuição normal multivariada com média zero e matriz de variância-covariância \mathbf{W}_t .

A idéia aqui é que a “idade” da informação que se tem sobre θ seja levada em conta no sentido de que nossa incerteza a respeito de θ deve aumentar com o passar do tempo. Neste sentido, a forma do modelo é apropriada apenas “localmente” no tempo e é necessário caracterizar algum tipo de evolução temporal de θ . O que se tem então é uma sequência de modelos ou um modelo dinâmico parametrizado por θ_t (o estado do processo no tempo t).

Considere um modelo em que uma variável y está relacionada a uma outra variável X de acordo com a seguinte forma paramétrica

$$y = X\theta + \epsilon.$$

Além disso, a incerteza do pesquisador em relação ao parâmetro θ é descrita em termos de uma distribuição de probabilidades $p(\theta)$.

Em um período t qualquer, D_t representa o conjunto de informações disponíveis sobre θ . Por simplicidade vamos assumir que $D_t = \{y_1, \dots, y_t\}$. Neste sentido, D_0 representa toda a informação inicial (antes de observar os dados) relevante sobre θ incluindo a própria definição do modelo.

No tempo $t - 1$, após observar y_1, \dots, y_{t-1} , toda a informação sobre o estado do processo está resumida probabilisticamente na distribuição *a posteriori* $p(\theta_{t-1}|D_{t-1})$. No tempo t , antes de observar y_t , toda a informação histórica D_{t-1} está resumida probabilisticamente na distribuição *a priori* de θ_t obtida como

$$p(\theta_t|D_{t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|D_{t-1})d\theta_{t-1}$$

que é atualizada após observar y_t para a posteriori θ_t , combinando-se com o modelo amostral $p(y_t|\theta_t)$ via teorema de Bayes

$$p(\theta_t|D_t) = \frac{p(y_t|\theta_t)p(\theta_t|D_{t-1})}{p(y_t|D_{t-1})}$$

sendo

$$p(y_t|D_{t-1}) = \int p(y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t$$

é a distribuição preditiva de y_t . Esquematicamente,

$$\begin{array}{ccccc} \theta_{t-1}|D_{t-1} & \longrightarrow & \theta_t|D_{t-1} & \longrightarrow & \theta_t|D_t \\ \text{Posteriori} & & \text{Priori} & & \text{Posteriori} \\ & & \downarrow & & \\ & & Y_t|D_{t-1} & & \\ & & \text{Previsão} & & \end{array}$$

Estas equações fornecem um sistema de aprendizado sequencial sobre os parâmetros do processo (não observáveis) e também uma sequência de distribuições preditivas (1 passo a frente) para as quantidades observáveis. Porém a sua implementação prática envolve a resolução de integrais que pode ser um problema de difícil solução em casos mais gerais. Um caso particular, onde as equações podem ser escritas em forma fechada, é o de modelos lineares dinâmicos (MLD) normais onde a distribuição amostral é definida pela

$$\text{equação das observações} \quad y_t = X_t\theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

e os parâmetros se relacionam em períodos sucessivos através da

$$\text{equação do sistema} \quad \theta_t = G\theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t)$$

onde as sequências ϵ_t e ω_t são independentes, mutuamente independentes e ambas são independentes da informação inicial $\theta_0|D_0 \sim N(m_0, C_0)$. A matriz G descreve a evolução (determinística) dos parâmetros. Modelos nesta classe serão analisados nas próximas seções.

7.2 Modelos Polinomiais

No MLD mais simples as observações são representadas por

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

onde μ_t é o nível da série no tempo t . A evolução do nível é modelada como um passeio aleatório simples, i.e.

$$\mu_t = \mu_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t).$$

Estas equações podem ser reescritas como

$$\begin{aligned} y_t | \mu_t &\sim N(\mu_t, V_t) \\ \mu_t | \mu_{t-1} &\sim N(\mu_{t-1}, W_t) \end{aligned}$$

e a informação inicial é $\mu_0 | D_0 \sim N(m_0, C_0)$. Vamos assumir por enquanto que as variâncias V_t e W_t são conhecidas. Este modelo pode ser pensado como uma aproximação de Taylor de 1ª ordem para uma função suave do tempo $\mu(t)$ de modo que

$$\mu(t + \delta t) = \mu(t) + \text{termos de ordem mais alta}$$

e o modelo descreve os termos de ordem mais alta simplesmente como ruídos de média zero. Como saber então se este modelo é adequado a uma particular aplicação?

No tempo t , o valor esperado da série k períodos a frente condicional ao nível atual é

$$E(Y_{t+k} | \mu_t) = E(\mu_{t+k} | \mu_t) = E(\mu_t + \sum_{i=1}^k \omega_{t+i} | \mu_t) = \mu_t$$

e denotando a média da distribuição a posteriori de μ_t por m_t então a *função de previsão* é constante

$$f_t(k) = E(Y_{t+k} | D_t) = E[E(Y_{t+k} | \mu_t, D_t)] = E(\mu_t | D_t) = m_t, \quad \forall k > 0.$$

Assim, este modelo é útil para previsões de curto prazo, particularmente quando a variação das observações (medida por V_t) é muito maior do que a variação do nível (medida por W_t).

Exemplo 7.1: Foram gerados 100 valores de um modelo polinomial de primeira ordem com variâncias constantes ($V_t = V$ e $W_t = W$). Na Figura 7.1 estão os valores gerados com as relações V/W iguais a 20, 2 e 0,2. Seguem os comandos do R para produção dos gráficos.

```
> mld.sim = function(n, V, W, mu0) {
+   mu = mu0 + cumsum(rnorm(n, sd = sqrt(W)))
+   obs = mu + rnorm(n, sd = sqrt(V))
+   ts(cbind(obs, mu))
+ }
```

7.2.1 Análise Sequencial e Previsões

A média inicial m_0 é uma estimativa pontual do nível da série e a variância inicial C_0 mede a incerteza associada. Assumindo que $\mu_{t-1}|D_{t-1} \sim N(m_{t-1}, C_{t-1})$, então condicionalmente a D_{t-1} , μ_t é a soma de 2 quantidades normais e independentes μ_{t-1} e ω_t e portanto é também normal com média e variância dadas por

$$\begin{aligned} E(\mu_t|D_{t-1}) &= E(\mu_{t-1}|D_{t-1}) + E(\omega_t|D_{t-1}) = m_{t-1} \\ \text{Var}(\mu_t|D_{t-1}) &= \text{Var}(\mu_{t-1}|D_{t-1}) + \text{Var}(\omega_t|D_{t-1}) = C_{t-1} + W_t = R_t \end{aligned}$$

$Y_t|D_{t-1}$ é também a soma de quantidades normais independentes e portanto tem distribuição normal com

$$\begin{aligned} E(Y_t|D_{t-1}) &= E(\mu_t|D_{t-1}) + E(\epsilon_t|D_{t-1}) = m_{t-1} \\ \text{Var}(Y_t|D_{t-1}) &= \text{Var}(\mu_t|D_{t-1}) + \text{Var}(\epsilon_t|D_{t-1}) = R_t + V_t = Q_t \end{aligned}$$

Após observar y_t , a distribuição atualizada do nível é obtida via teorema de Bayes combinando-se a verossimilhança

$$p(y_t|\mu_t, D_{t-1}) = (2\pi V_t)^{-1/2} \exp\{-(y_t - \mu_t)^2/2V_t\}$$

com a priori

$$p(\mu_t|D_{t-1}) = (2\pi R_t)^{-1/2} \exp\{-(\mu_t - m_{t-1})^2/2R_t\}$$

de modo que

$$\begin{aligned} p(\mu_t|D_t) &\propto \exp\left\{-\frac{1}{2}\left[\frac{(y_t - \mu_t)^2}{V_t} + \frac{(\mu_t - m_{t-1})^2}{R_t}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\mu_t^2(V_t^{-1} + R_t^{-1}) - 2\mu_t(V_t^{-1}y_t + R_t^{-1}m_{t-1})\right]\right\} \\ &\propto \exp\left\{-\frac{C_t^{-1}}{2}(\mu_t^2 - 2\mu_t m_t)\right\} \propto \exp\left\{-\frac{C_t^{-1}}{2}(\mu_t - m_t)^2\right\} \end{aligned}$$

onde

$$\begin{aligned} m_t &= C_t(V_t^{-1}y_t + R_t^{-1}m_{t-1}) \\ C_t^{-1} &= V_t^{-1} + R_t^{-1} \end{aligned}$$

e todos os termos que não dependem de μ_t foram colocados na constante de proporcionalidade. Portanto, $\mu_t|D_t \sim N(m_t, C_t)$.

A média a posteriori pode ser reescrita de 2 formas alternativas definindo-se o *coeficiente adaptativo* $A_t = C_t V_t^{-1} = R_t / Q_t \in (0, 1)$ e o erro de previsão 1 passo a frente $e_t = y_t - m_{t-1}$. Assim

$$m_t = (1 - A_t)m_{t-1} + A_t y_t = m_{t-1} + A_t e_t.$$

Note a similaridade com a equação de previsão do método de alisamento exponencial simples visto no Capítulo 5. Aqui A_t faz o papel da constante de alisamento porém

agora variando no tempo. A variância a posteriori também pode ser reescrita como função do coeficiente adaptativo como

$$C_t = R_t - A_t^2 Q_t < R_t.$$

Podemos utilizar as equações das observações e de evolução para obter a distribuição preditiva k passos a frente. Fazendo substituições sucessivas obtemos que

$$\begin{aligned}\mu_{t+k} &= \mu_t + \sum_{j=1}^k \omega_{t+j} \\ Y_{t+k} &= \mu_t + \sum_{j=1}^k \omega_{t+j} + \epsilon_{t+k}\end{aligned}$$

e como todos os termos são normais e independentes segue que Y_{t+k} é também normal com

$$\begin{aligned}E(Y_{t+k}|D_t) &= E(\mu_t|D_t) = m_t \\ \text{Var}(Y_{t+k}|D_t) &= C_t + \sum_{j=1}^k W_{t+j} + V_{t+k}\end{aligned}$$

A função abaixo estima um modelo com tendencia polinomial de 1a ordem fazendo a análise sequencial usando as equações dadas no texto com variâncias fixas e conhecidas.

```
> mld = function(Y, V, W, m0, C0) {
+   n = length(Y)
+   m = C = R = Q = f = A = e = ts(rep(NA, length = n), start = start(Y))
+   Y = ts(c(NA, Y), end = end(Y))
+   C[1] = C0
+   m[1] = m0
+   for (t in 2:n) {
+     R[t] = C[t - 1] + W[t]
+     f[t] = m[t - 1]
+     Q[t] = R[t] + V[t]
+     A[t] = R[t]/Q[t]
+     e[t] = Y[t] - f[t]
+     m[t] = m[t - 1] + A[t] * e[t]
+     C[t] = A[t] * V[t]
+   }
+   return(list(m = m, C = C, R = R, f = f, Q = Q))
+ }
```

Exemplo 7.2: A função `mld` pode ser usada para estimar sequencialmente o nível da serie de vazões do rio Nilo. Primeiro vamos permitir que o nível varie bastante ao

longo do tempo especificando um valor grande para W e depois reestimar com pouca variação temporal (W bem pequeno). Usaremos a variancia amostral da serie como estimativa de V . Como informação inicial usaremos uma estimativa do nível igual a 1000 mas com uma grande incerteza associada. O gráfico da série com os níveis superimpostos aparece na Figura 7.2.

7.2.2 Variâncias de Evolução e das Observações

Tipicamente, W_t é desconhecida. Sua estimação entretanto leva a uma intratabilidade analítica que pode ser evitada através de sua especificação subjetiva.

O *fator de desconto* é o parâmetro básico que controla o grau de “envelhecimento” da informação de uma observação. Por exemplo, podemos quantificar o envelhecimento da informação sobre o parâmetro μ_t como um aumento de 5% em sua variância a priori (no tempo t), i.e.

$$Var(\mu_t|D_{t-1}) = (1 + \delta)Var(\mu_{t-1}|D_{t-1}) \quad \text{ou} \quad R_t = (1 + \delta) C_{t-1}$$

com $\delta = 0.05$. Por outro lado, informação é em geral medida em termos de *precisão* (o inverso da variância) e podemos escrever

$$\text{Precisão}(\mu_t|D_{t-1}) = (1 + \delta)^{-1}\text{Precisão}(\mu_{t-1}|D_{t-1}) \quad \text{ou} \quad R_t^{-1} = (1 + \delta)^{-1}C_{t-1}^{-1}.$$

Nesta escala, o fator de desconto $\lambda = (1 + \delta)^{-1}$ varia entre 0 e 1 e $\delta = 5\%$ implica em $\lambda \approx 0.95$. Vale notar que o fator de desconto não depende da escala na qual as observações são medidas.

Se $\lambda = 1$ então não existe mudança ao longo do tempo no nível da série e quanto menor é o valor de λ maiores são as alterações esperadas e maior é a perda de informação contida em observações mais “antigas”.

Assim, para um valor fixo do fator de desconto λ temos que

$$R_t = C_{t-1}/\lambda = C_{t-1} + W_t$$

ou equivalentemente

$$W_t = C_{t-1} \left(\frac{1 - \lambda}{\lambda} \right) = \delta C_{t-1}.$$

Como $R_t = C_{t-1} + W_t$ podemos interpretar esta especificação intuitivamente como um aumento de incerteza, ao evoluir de $t-1$ para t , quantificado como uma proporção de C_{t-1} .

A sequência de variâncias V_t também é, em geral, desconhecida embora o pesquisador possa ter alguma informação a priori sobre características desta sequência. Por exemplo, $V_t = V$ (variância constante e desconhecida), $V_t = V k_t$ onde os pesos k_t são conhecidos, $V_t = V k(\mu_t)$ onde $k(\cdot)$ é uma função de variância do nível da série ou em particular $V_t = V \mu_t^p$.

Impondo-se uma particular estrutura para a sequência W_t e para a informação inicial obtem-se um procedimento de atualização sequencial para V além de μ_t . Para

isto redefine-se o modelo, agora condicionalmente em V , como

$$\begin{aligned} y_t &= \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, V), \\ \mu_t &= \mu_{t-1} + \omega_t, \quad \omega_t \sim N(0, VW_t^*), \\ \mu_0|V, D_0 &\sim N(m_0, VC_0^*) \\ V^{-1}|D_0 &\sim \text{Gama}\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right) \quad \text{ou} \quad n_0 S_0 V^{-1} \sim \chi_{n_0}^2 \end{aligned}$$

sendo que m_0 , C_0^* , n_0 e S_0 serão especificados. Surgiu assim mais um item na informação inicial com

$$E(V^{-1}|D_0) = \frac{n_0/2}{n_0 S_0/2} = \frac{1}{S_0}$$

e S_0 é a estimativa pontual a priori da variância V . Com esta definição pode-se mostrar que a distribuição inicial marginal de μ_0 é

$$\mu_0|D_0 \sim t_{n_0}(m_0, C_0)$$

com $C_0 = S_0 C_0^*$.

Se a distribuição a posteriori (incondicional) do nível em $t - 1$ é

$$\mu_{t-1}|D_{t-1} \sim t_{n_{t-1}}(m_{t-1}, C_{t-1})$$

então pode-se mostrar que as distribuições a priori, preditiva e a posteriori no tempo t são dadas por

$$\begin{aligned} \mu_t|D_{t-1} &\sim t_{n_{t-1}}(m_{t-1}, R_t) \\ Y_t|D_{t-1} &\sim t_{n_{t-1}}(m_{t-1}, Q_t) \\ \mu_t|D_t &\sim t_{n_t}(m_t, C_t) \end{aligned}$$

onde os parâmetros atualizados são

$$\begin{aligned} Q_t &= R_t + S_{t-1} \\ m_t &= m_{t-1} + A_t e_t \\ C_t &= (S_t/S_{t-1})(R_t - A_t^2 Q_t) \\ n_t &= n_{t-1} + 1 \\ n_t S_t &= n_{t-1} S_{t-1} + S_{t-1} e_t^2 / Q_t. \end{aligned}$$

A função `mld1` abaixo faz a análise sequencial com a variância das observações fixa e desconhecida. A especificação de W_t é feita via fator de desconto. Note que agora tanto o nível quanto a variância e os graus de liberdade são atualizados sequencialmente.

```
> mld1 = function(Y, delta, m0, C0, n0, S0) {
+   N = length(Y)
+   m = n = C = R = Q = S = f = A = e = rep(NA, length = N)
+   Y = c(NA, Y)
```



```

+      C[1] = C0
+      m[1] = m0
+      S[1] = S0
+      n[1] = n0
+      for (i in 2:N) {
+          n[i] = n[i - 1] + 1
+          R[i] = C[i - 1]/delta
+          f[i] = m[i - 1]
+          Q[i] = R[i] + S[i - 1]
+          A[i] = R[i]/Q[i]
+          e[i] = Y[i] - f[i]
+          S[i] = S[i - 1] + (S[i - 1]/n[i]) * (e[i]^2/Q[i] - 1)
+          m[i] = m[i - 1] + A[i] * e[i]
+          C[i] = A[i] * S[i]
+      }
+      return(list(m = m, C = C, R = R, f = f, Q = Q, n = n, S = S,
+                  e = e))
+ }

```

Exemplo 7.3: Novamente vamos examinar a série de vazões do rio Nilo, agora usando diferentes fatores de desconto na função `mld1`.

```

> res1 = mld1(y, delta = 0.98, m0 = 1000, C0 = 100, n0 = 1, S0 = 0.01)
> res2 = mld1(y, delta = 0.7, m0 = 1000, C0 = 100, n0 = 1, S0 = 0.01)

```

Os gráficos na Figura 7.3 mostram a série original, as estimativas do nível obtidas com descontos 0,70 e 0,98 e estas mesmas estimativas com um intervalo de $\pm 1,5\sqrt{C_t}$. Os gráficos foram feitos com os seguintes comandos do R,

Os modelos podem ser comparados calculando-se o erro quadrático médio e o desvio absoluto médio. Usando os comandos abaixo percebe-se que o modelo com fator de desconto 0,70 é melhor segundo estes critérios.

```

> eqm = dam = rep(0, 2)
> for (i in 2:length(y)) {
+     eqm[1] = eqm[1] + (y[i] - res1$m[i - 1])^2
+     dam[1] = dam[1] + abs(y[i] - res1$m[i - 1])
+     eqm[2] = eqm[2] + (y[i] - res2$m[i - 1])^2
+     dam[2] = dam[2] + abs(y[i] - res2$m[i - 1])
+ }
> eqm

[1] 2681716 2375484

> dam

[1] 13258.47 11904.16

```

7.3 Modelo de Crescimento Linear

Considere agora que a descrição “local” mais apropriada é uma tendência polinomial de 2ª ordem. Um modelo um pouco mais elaborado é então obtido criando-se um parâmetro extra para descrever o crescimento do nível do processo observado. A equação das observações fica inalterada, i.e.

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \sim N(0, V_t)$$

e a evolução do nível e do crescimento é modelada como

$$\begin{aligned} \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{1t} \\ \beta_t &= \beta_{t-1} + \omega_{2t}. \end{aligned}$$

Usando a representação matricial temos que o vetor de regressão e a matriz de evolução são dados por

$$X_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad \text{e} \quad G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Nesta notação, definindo $\theta_t = (\mu_t, \beta_t)'$ obtemos os momentos das distribuições a priori e preditiva como

$$\begin{aligned} E(\theta_t | D_{t-1}) &= a_t = GE(\theta_{t-1} | D_{t-1}) = Gm_{t-1} = (m_{t-1} + b_{t-1}, b_{t-1})' \\ \text{Var}(\theta_t | D_{t-1}) &= R_t = GC_{t-1}G' + W_t \\ E(Y_t | D_{t-1}) &= f_t = X_t a_t = m_{t-1} + b_{t-1} \\ \text{Var}(Y_t | D_{t-1}) &= Q_t = X_t R_t X_t' + S_{t-1}. \end{aligned}$$

Os momentos da distribuição a posteriori de θ_t são uma generalização matricial daqueles obtidos para o modelo anterior,

$$\begin{aligned} E(\theta_t | D_t) &= m_t = a_t + A_t e_t \\ \text{Var}(\theta_t | D_t) &= C_t = (S_t / S_{t-1})(R_t - A_t A_t' Q_t) \end{aligned}$$

Não é difícil verificar que a função de previsão é dada por

$$f_t(k) = X_t G^k m_t = m_t + k b_t$$

sendo que m_t e b_t são as estimativas pontuais do nível μ_t e do crescimento β_t . Portanto, assim como no caso anterior, este modelo também é apropriado para previsões de curto prazo.

As variâncias W_t são mais uma vez especificadas indiretamente através de um fator de desconto λ . Neste caso,

$$R_t = GC_{t-1}G' / \lambda$$

implica que

$$W_t = GC_{t-1}G'(\lambda^{-1} - 1).$$

7.4 Modelos Sazonais

Um comportamento periódico ou cíclico pode ser encontrado em várias séries temporais. É importante que se consiga descrever o padrão sazonal da série através de quantidades que possam ser estimadas incluindo-se assim este padrão na função previsão. Nos modelos aqui analisados define-se um componente sazonal descrevendo *desvios sazonais* em torno de um *nível dessazonalizado* ou *tendência*.

7.4.1 Modelos sem Crescimento

A idéia aqui é fazer a superposição de um modelo polinomial de 1ª ordem (para o nível dessazonalizado) com um modelo de efeitos sazonais. As equações das observações e de evolução são dadas por

$$\begin{aligned} y_t &= \mu_t + \phi_{t0} + \epsilon_t, \quad \epsilon_t \sim N(0, V_t) \\ \mu_t &= \mu_{t-1} + \omega_t \\ \phi_{tr} &= \phi_{t-1, r+1} + \omega_{t,r}, \quad r = 0, \dots, p-2 \\ \phi_{t, p-1} &= \phi_{t-1, 0} + \omega_{t, p-1} \end{aligned}$$

com a restrição $\sum_{r=0}^{p-1} \phi_{tr} = 0, \forall t$ e onde p é o período sazonal da série. Por exemplo, $p = 12$ para uma série com observações mensais e $p = 4$ para observações trimestrais.

Para fixar idéias, considere uma série trimestral e suponha que $t - 1$ é o segundo trimestre de um determinado ano. Então o vetor de parâmetros consiste de 4 efeitos sazonais, um para cada trimestre,

$$\phi_{t-1} = \begin{pmatrix} \phi_{t0} \\ \phi_{t1} \\ \phi_{t2} \\ \phi_{t3} \end{pmatrix} = \begin{pmatrix} \text{trim. 2} \\ \text{trim. 3} \\ \text{trim. 4} \\ \text{trim. 1} \end{pmatrix}$$

e ao passar de $t - 1$ para t ocorre simplesmente uma rotação nos elementos deste vetor,

$$\phi_t = \begin{pmatrix} \phi_{t0} \\ \phi_{t1} \\ \phi_{t2} \\ \phi_{t3} \end{pmatrix} = \begin{pmatrix} \text{trim. 3} \\ \text{trim. 4} \\ \text{trim. 1} \\ \text{trim. 2} \end{pmatrix}.$$

A função de previsão assume a forma $f_t(k) = m_t + h_{tj}$ onde m_t é o valor esperado do nível dessazonalizado no tempo $t + k$ e h_{tj} é o desvio sazonal esperado em torno deste nível. O desvio utilizado na função de previsão é tal que j é o resto da divisão k/p . Por exemplo, se $p = 12$, e ϕ_{t0} refere-se ao mês de janeiro então a previsão 1 passo a frente ($k = 1$) feita em dezembro é $m_t + E(\phi_{t0}|D_t)$, com $j = 1$. Se o horizonte de previsão for $k = 2$ então $j = 2$ e o desvio sazonal refere-se a fevereiro, i.e. $f_t(2) = m_t + E(\phi_{t1}|D_t)$.

7.4.2 Modelos com Crescimento

Novamente a idéia é fazer a superposição de um modelo para os efeitos sazonais mas agora com um modelo polinomial de 2ª ordem onde se tem um parâmetro que representa o crescimento do nível dessazonalizado.

O modelo pode ser escrito como

$$\begin{aligned} y_t &= \mu_t + \phi_{t0} + \epsilon_t, \quad \epsilon_t \sim N(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_t \\ \beta_t &= \beta_{t-1} + \omega_t^* \\ \phi_{tr} &= \phi_{t-1,r+1} + \omega_{t,r}, \quad r = 0, \dots, p-2 \\ \phi_{t,p-1} &= \phi_{t-1,0} + \omega_{t,p-1} \end{aligned}$$

com a restrição $\sum_{r=0}^{p-1} \phi_{tr} = 0, \forall t$. A função de previsão agora assume a seguinte forma

$$f_t(k) = m_t + kb_t + h_{tj}, \quad \text{com} \quad \sum_{j=0}^{p-1} h_{tj} = 0$$

onde h_{tj} tem a mesma interpretação anterior.

7.5 Representação de Fourier

Uma forma alternativa de se representar padrões cíclicos é através de combinações lineares de funções periódicas. Em particular a utilização de funções trigonométricas leva a representações de Fourier da sazonalidade.

O modelo (com crescimento) é representado pelas seguintes equações

$$\begin{aligned} y_t &= \mu_t + \sum_{j=1}^{p/2} \alpha_{j,t} + \epsilon_t, \quad \epsilon_t \sim N(0, V_t) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_t, \\ \beta_t &= \beta_{t-1} + \omega_t^*, \\ \begin{bmatrix} \alpha_{j,t} \\ \alpha_{j,t}^* \end{bmatrix} &= \begin{bmatrix} \cos 2\pi j/p & \sin 2\pi j/p \\ -\sin 2\pi j/p & \cos 2\pi j/p \end{bmatrix} \begin{bmatrix} \alpha_{j,t-1} \\ \alpha_{j,t-1}^* \end{bmatrix} + \begin{bmatrix} w_{j,t} \\ w_{j,t}^* \end{bmatrix}, j = 1, \dots, p/2 - 1 \end{aligned}$$

e $\alpha_{j,t} = -\alpha_{j,t-1} + \omega_{j,t}$ para $j = p/2$. A função de previsão é dada por

$$f_t(k) = \sum_{j=1}^{p/2} S_{jk} = \sum_{j=1}^{p/2} [a_{t,j} \cos(2\pi jk/p) + a_{t,j}^* \sin(2\pi jk/p)]$$

onde $a_{t,j}$ e $a_{t,j}^*$ são as estimativas pontuais de coeficientes de Fourier $\alpha_{t,j}$ e $\alpha_{t,j}^*$.

Como no capítulo anterior, as variâncias dos erros de evolução são especificadas indiretamente através de um fator de desconto. A estratégia recomendada em (Pole, West, & Harrison 1994) e West & Harrison (1997) consiste em especificar um fator de

desconto para cada componente do modelo. No modelo com uma tendência polinomial mais um componente sazonal teremos então 2 fatores de desconto.

Em geral, o fator de desconto do componente sazonal é maior do que o da tendência. Neste sentido estamos assumindo que o padrão sazonal da série, embora possa estar sujeito a alterações, é mais estável do que a sua tendência.

7.6 Ilustração

A Figura ?? apresenta o total de vendas trimestrais (em milhares) de perus na Irlanda entre o primeiro trimestre de 1974 e o terceiro trimestre de 1982. A série exibe um crescimento sistemático ao longo de todo o período juntamente com um padrão sazonal acentuado. Outra característica interessante é que a forma do padrão sazonal se alterou a partir de 1978. Vamos fazer a estimação sequencial de um modelo para os efeitos sazonais superpostos a uma tendência de crescimento linear e verificar o comportamento das previsões 1 passo a frente.

Suponha que a informação a priori foi acessada examinando-se as vendas dos anos anteriores a 1974. Esta informação está resumida na Tabela 7.1. Note a restrição de soma zero na especificação a priori dos efeitos sazonais e também que a especificação equivalente em termos de fatores sazonais seria 11, 19, 19 e 11 para os fatores e $(11+19+19+11)/4 = 15$ para o nível.

Tabela 7.1: Informação a priori.

Componente	Média (Desvio padrão)
Nível	15 (0.75)
Crescimento	0 (0.3)
Efeito sazonal 1	-4 (0.5)
Efeito sazonal 2	4 (0.5)
Efeito sazonal 3	4 (0.5)
Efeito sazonal 4	-4 (0.5)
D.P. das observações	1 com 1 g.l.

A performance preditiva do modelo foi investigada para fatores de desconto variando nos intervalos (0.9,1.0) para a tendência e (0.6,1.0) para os fatores sazonais. Estes intervalos estão coerentes com a idéia de que espera-se um padrão sazonal mais estável do que a tendência. Entretanto os valores encontrados após esta “busca” foram 0.90 para a tendência e 0.80 para os fatores sazonais. Uma idéia intuitiva é a alteração no padrão sazonal ocorrida em 1978 deve ter contribuído para este resultado atípico.

Os 2 gráficos a seguir apresentam as previsões pontuais (1 passo a frente) juntamente com intervalos de 90% de probabilidade e os valores observados da série. O primeiro gráfico refere-se ao modelo estático (ambos os fatores de desconto iguais a 1).

Note que a mudança no padrão sazonal ocorre muito lentamente no modelo estático e no final da série o padrão estimado é apenas ligeiramente diferente do padrão inicial. Já no modelo dinâmico o padrão sazonal evolui para uma forma completamente di-

Tabela 7.2:

Descontos	EQM	DAM	LLIK
0.90 e 0.80	3.11	1.34	-71.1
1.00 e 1.00	4.23	1.64	-77.6

ferente melhorando a performance preditiva. Este fato pode ser notado por inspeção visual e é confirmado pelos indicadores na Tabela 7.2.

A explicação intuitiva para este fato, lembrando da definição de fator de desconto, é que no modelo dinâmico um peso maior é dado para as observações mais recentes ao fazer previsões. Com isto a alteração no padrão sazonal é “incorporada” mais rapidamente do que no modelo estático. As previsões de vendas para o quarto trimestre de 1982 e para 1983 também levarão em conta os diferentes padrões sazonais do final da série.

7.7 Modelos de Regressão

Para completar o nosso modelo dinâmico podemos pensar em incluir na equação das observações *efeitos* de variáveis regressoras. Considere por exemplo a regressão linear da variável y_t em uma coleção de p variáveis independentes X_{1t}, \dots, X_{pt} . Se um termo constante for incluído no modelo então $X_{1t} = 1, \forall t$. Denotando o vetor de regressão e o vetor de coeficientes de regressão no tempo t por $X_t = (X_{1t}, \dots, X_{pt})$ e $\theta_t = (\theta_{1t}, \dots, \theta_{pt})'$ respectivamente então as equações do modelo são dadas por

$$\begin{aligned} y_t &= X_t \theta_t + \epsilon_t, & \epsilon_t &\sim N(0, V_t) \\ \theta_t &= \theta_{t-1} + \omega_t, & \omega_t &\sim N(0, W_t). \end{aligned}$$

Assim, os coeficientes da regressão evoluem segundo um passeio aleatório, como no modelo polinomial de 1ª ordem, i.e., a matriz de evolução $G = I_p$. O vetor de regressão é formado pelas próprias variáveis regressoras e note que a equação das observações pode ser reescrita como

$$y_t = \sum_{i=1}^p \theta_{it} X_{it} + \epsilon_t$$

de modo que o modelo pode ser visto como uma superposição de p regressões simples pela origem.

Todas as distribuições envolvidas são análogas aos casos anteriores e as equações dadas na Seção 2.3 podem ser utilizadas para obter os momentos das distribuições a

priori, preditiva e a posteriori fazendo-se $G = I_p$. Assim,

$$\begin{aligned} a_t &= m_{t-1} \\ R_t &= C_{t-1} + W_t \\ f_t &= X_t m_{t-1} \end{aligned}$$

e as outras equações permanecem inalteradas.

É interessante notar como fica a função de previsão $f_t(k)$ neste caso. Primeiro reescreva a equação de evolução para θ_{t+k} fazendo k substituições sucessivas obtendo

$$\theta_{t+k} = \theta_t + \sum_{j=1}^k \omega_{t+j}$$

de modo que

$$\begin{aligned} a_{t+k} &= m_t \\ R_{t+k} &= C_t + \sum_{j=1}^k W_{t+j}. \end{aligned}$$

Então, usando a equação das observações obtemos que

$$\begin{aligned} f_t(k) &= X_{t+k} m_t \\ Q_{t+k} &= X_{t+k} R_{t+k} X'_{t+k} + S_t. \end{aligned}$$

Assim, a previsão pontual k passos a frente é a própria função de regressão avaliada na estimativa dos coeficientes no tempo t e nos valores futuros dos regressores (que nem sempre estão disponíveis).

A sequência de variâncias W_t é mais uma vez estruturada usando um fator de desconto.

7.8 Monitoramento

Ao comparar sequencialmente as previsões com os valores observados pode-se julgar a adequação relativa de modelos alternativos com base em sua performance preditiva.

Observações ocorrendo nas caudas da distribuição preditiva são sempre possíveis por definição porém improváveis. Quanto mais afastada em uma das caudas mais improvável é a observação. É preciso então estabelecer um critério para julgar que tipo de inconsistência entre observação e previsão deve ser sinalizada pelo sistema. No entanto, sinalizar uma observação como improvável apenas indica uma possível deficiência geral do modelo. É preciso saber em que sentido o modelo é deficiente, i.e. verificar que modelos alternativos, com diferentes distribuições preditivas, teriam uma performance melhor. O *fator de Bayes*, definido a seguir, é a ferramenta utilizada para fazer esta comparação de modelos.

Se $p_A(y_t|D_{t-1})$ é a densidade preditiva 1 passo a frente de um modelo alternativo então o fator de Bayes é definido como

$$H_t = \frac{p(y_t|D_{t-1})}{p_A(y_t|D_{t-1})},$$

i.e. a razão das densidades preditivas avaliadas no valor observado y_t .

Outra forma de comparar a performance preditiva de dois modelos é considerar um grupo de observações ao invés de uma única e se basear no fator de Bayes acumulado

$$\begin{aligned} H_t(k) &= \frac{p(y_t, \dots, y_{t-k+1}|D_{t-k})}{p_A(y_t, \dots, y_{t-k+1}|D_{t-k})} = \frac{p(y_t|D_{t-1})p(y_{t-1}, \dots, y_{t-k+1}|D_{t-k})}{p_A(y_t|D_{t-1})p_A(y_{t-1}, \dots, y_{t-k+1}|D_{t-k})} \\ &= H_t H_{t-1}(k-1) = \prod_{j=0}^{k-1} H_{t-j}. \end{aligned}$$

Pode-se assim sinalizar evidências de alteração lenta na estrutura da série. A idéia é que, individualmente, estas evidências não são suficientes para se “questionar” as previsões do modelo em uso mas quando consideradas conjuntamente a evidência acumulada pode ser grande e deve ser sinalizada. A questão agora é como construir um sistema de monitoramento automático da série a partir destas idéias intuitivas.

Quando as observações estão cada vez mais afastadas das previsões então um fator de Bayes individual H_t pode não ser suficientemente pequeno e precisa ser acumulado para indicar alguma evidência contra o modelo padrão. Neste caso, o monitor identifica o grupo mais discrepante de observações consecutivas calculando V_t and l_t da seguinte forma,

$$V_t = \min_{1 \leq k \leq t} H_t(k) = H_t(l_t)$$

sendo calculado sequencialmente com as seguintes recursões,

$$V_t = H_t \min\{1, L_{t-1}\} \quad \text{e} \quad l_t = \begin{cases} l_{t-1} + 1, & \text{se } L_{t-1} < 1 \\ 1, & \text{se } L_{t-1} \geq 1 \end{cases}$$

conforme mostrado em West (1986).

O modelo padrão é aceito como sendo satisfatório até a ocorrência de um valor L_t menor do que um valor pré-especificado $\tau < 1$ (o limite inferior para aceitação de L_t) quando a ocorrência de uma descontinuidade na série é sinalizada. Se $l_t = 1$ então uma única observação discrepante é identificada como a causa mais provável de falha, embora o início de uma mudança também seja uma possibilidade. Por outro lado, $l_t > 1$ indica que uma mudança começou a ocorrer l_t periods atrás em $t - l_t + 1$. Além disso, se uma mudança estrutural lenta está ocorrendo na série as observações mais recentes indicarão evidência contra o modelo padrão que não será suficiente para fazer $L_t < \tau$. Assim, para aumentar a sensibilidade do monitor a estas mudanças uma descontinuidade deve ser sinalizada se $l_t > 3$ ou 4.

Para especificar o modelo alternativo assume-se que as densidades preditivas são normais com média comum f_t e variâncias Q_t e Q_t/ρ onde $0 < \rho < 1$, de modo que o fator de Bayes fica

$$H_t = \sqrt{\frac{1}{\rho}} \exp \left\{ -\frac{(y_t - f_t)^2}{2Q_t} (1 - \rho) \right\} = \sqrt{\frac{1}{\rho}} \exp \left\{ -\frac{1}{2} (1 - \rho) e_t^2 \right\}$$

onde e_t é o erro de previsão um passo a frente padronizado.

A escolha de ρ pode ser facilitada reescrevendo-se o fator de Bayes como

$$H_t = \exp(-0.5 \log \rho + (1 - \rho)e_t^2).$$

Claramente $H_t = 1$ ou equivalentemente $e_t^2 = -(\log \rho)/(1 - \rho)$ indica nenhuma evidência para discriminar entre os modelos. O valor de ρ , pode ser escolhido de modo a fornecer o valor máximo de $|e_t|$ que não indica evidence contra o modelo padrão. Por exemplo, $\rho \in (0.1, 0.3)$ implica que a evidência contra o modelo padrão deve ser acumulada para $1.3 < |e_t| < 1.6$ que são aproximadamente os percentil 0.90 e 0.95 distribuição normal padrão.

É claro que para ρ fixo, a evidência contra o modelo padrão aumenta com $|e_t|$. West & Harrison (1997) ilustraram como a escolha de ρ tem pouca influência quando o erro se torna muito grande em relação ao modelo alternativo. Este pode ser visto como um modelo geral no sentido de levar em conta vários tipos de mudanças além de observações discrepantes. Essencialmente, este procedimento pode ser visto como um método exploratório gerando informação sobre o tipo e o período mais provável de mudança estrutural.

```

> w = c(0.05, 0.5, 5)
> g = list(col = 1:2, xlab = "tempo", ylab = "y")
> par(mfrow = c(2, 2))
> for (i in w) {
+   ts.plot(mld.sim(100, 1, i, 25), gpars = g, main = paste("V/W=",
+   1/i))
+ }

```

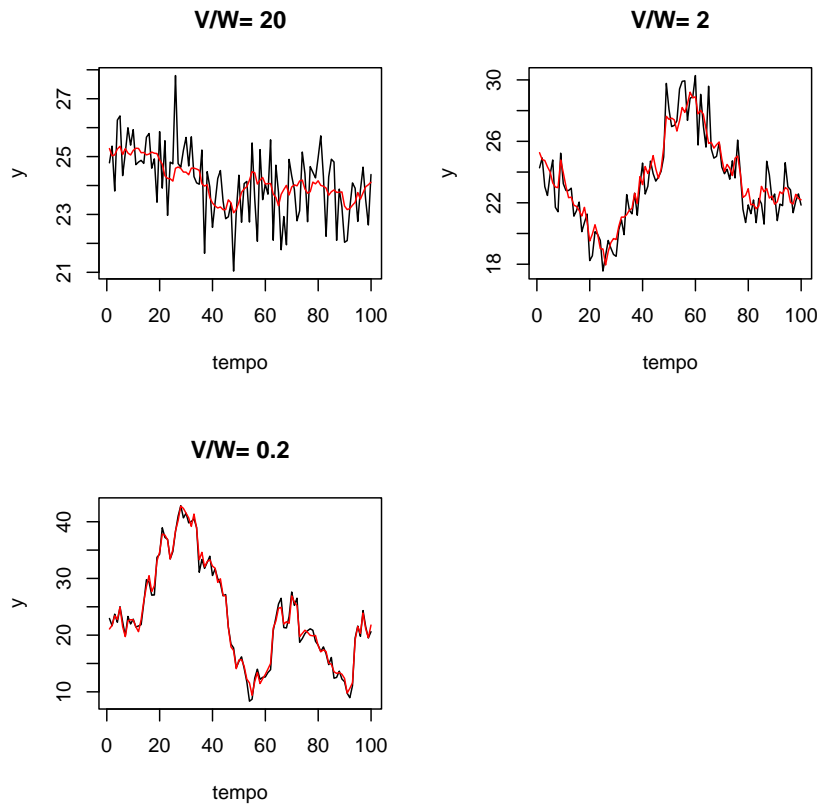


Figura 7.1: 100 valores simulados do modelo polinomial de 1^a ordem com (a) $V/W = 20$, (b) $V/W = 2$, (c) $V/W = 0,2$.

```

> y = Nile
> n = length(y)
> res = mld(y, V = rep(var(y), n), W = rep(50, n), m0 = 1000, C0 = 1000)
> plot(y, xlab = "Anos", ylab = "Medições", type = "p")
> lines(res$m, col = 2)
> lines(res$m - 2 * sqrt(res$C), col = 2, lty = 1)
> lines(res$m + 2 * sqrt(res$C), col = 2, lty = 1)
> res = mld(y, V = rep(var(y), n), W = rep(0.05, n), m0 = 1000,
+       C0 = 1000)
> lines(res$m, col = 4)
> legend(1940, 1350, c("obs", "W=50", "W=.05"), col = c(1, 2, 4),
+       bty = "n")

```

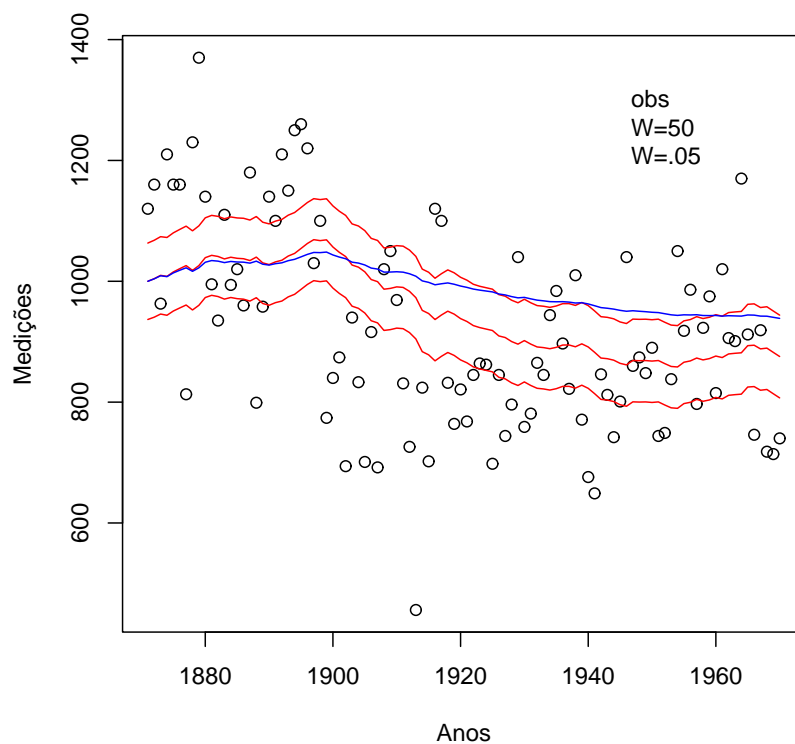


Figura 7.2:

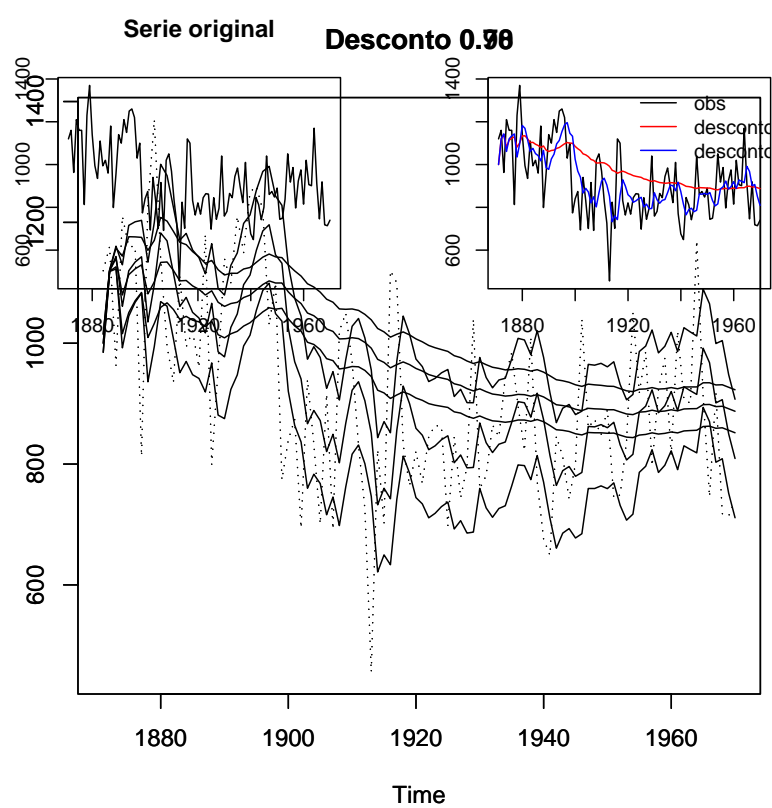


Figura 7.3:

Apêndice A

Lista de Distribuições

Neste apêndice são listadas as distribuições de probabilidade utilizadas no texto para facilidade de referência. São apresentadas suas funções de (densidade) de probabilidade além da média e variância. Uma revisão exaustiva de distribuições de probabilidades pode ser encontrada em Johnson et al. (1994), Johnson et al. (1995) e Johnson et al. (1992).

A.1 Distribuição Normal

X tem distribuição normal com parâmetros μ e σ^2 , denotando-se $X \sim N(\mu, \sigma^2)$, se sua função de densidade é dada por

$$p(x|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp[-(x - \mu)^2/2\sigma^2], \quad -\infty < x < \infty,$$

para $-\infty < \mu < \infty$ e $\sigma^2 > 0$. Quando $\mu = 0$ e $\sigma^2 = 1$ a distribuição é chamada normal padrão. A distribuição log-normal é definida como a distribuição de e^X .

No caso vetorial, $\mathbf{X} = (X_1, \dots, X_p)$ tem distribuição normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de variância-covariância Σ , denotando-se $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ se sua função de densidade é dada por

$$p(\mathbf{x}|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-p/2} |\Sigma|^{-1/2} \exp[-(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})/2]$$

para $\boldsymbol{\mu} \in \mathbb{R}^p$ e Σ positiva-definida.

A.2 Distribuição Gama

X tem distribuição Gama com parâmetros α e β , denotando-se $X \sim Ga(\alpha, \beta)$, se sua função de densidade é dada por

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,$$

para $\alpha, \beta > 0$.

$$E(X) = \alpha/\beta \quad \text{e} \quad V(X) = \alpha/\beta^2.$$

Casos particulares da distribuição Gama são a distribuição de Erlang, $Ga(\alpha, 1)$, a distribuição exponencial, $Ga(1, \beta)$, e a distribuição qui-quadrado com ν graus de liberdade, $Ga(\nu/2, 1/2)$.

A.3 Distribuição Wishart

Diz-se que uma matriz aleatória Ω ($n \times n$) segue uma distribuição Wishart com parâmetro Σ e ν graus de liberdade, denotando-se $\Omega \sim W(\Sigma, \nu)$, se sua função de densidade é dada por,

$$p(\Omega|\Sigma, \nu) \propto |\Omega|^{(\nu-n-1)/2} \exp(-(1/2)\text{tr}(\Sigma\Omega))$$

sendo $\nu \geq n$, Σ positiva-definida e $\text{tr}(A)$ indica o traço de uma matriz A . Uma propriedade útil é que $A\Omega A' \sim W(A\Sigma A', \nu)$.

A.4 Distribuição Gama Inversa

X tem distribuição Gama Inversa com parâmetros α e β , denotando-se $X \sim GI(\alpha, \beta)$, se sua função de densidade é dada por

$$p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\beta/x}, \quad x > 0,$$

para $\alpha, \beta > 0$.

$$E(X) = \frac{\beta}{\alpha - 1} \quad \text{e} \quad V(X) = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}.$$

Não é difícil verificar que esta é a distribuição de $1/X$ quando $X \sim Ga(\alpha, \beta)$.

A.5 Distribuição Wishart Invertida

Diz-se que uma matriz aleatória Ω ($n \times n$) segue uma distribuição Wishart-Invertida com parâmetro Σ e ν graus de liberdade, denotando-se $\Omega \sim WI(\Sigma, \nu)$ se sua função de densidade é dada por,

$$p(\Omega|\Sigma, \nu) \propto |\Omega|^{-(\nu+n+1)/2} \exp(-(1/2)\text{tr}(\Sigma\Omega))$$

sendo $\nu \geq n$, Σ positiva-definida e $\text{tr}(A)$ indica o traço de uma matriz A . Não é difícil verificar que $\Omega^{-1} \sim W(\Sigma, \nu)$. Outra propriedade é que $A\Omega A' \sim WI(A\Sigma A', \nu)$.

A.6 Distribuição Beta

X tem distribuição Beta com parâmetros α e β , denotando-se $X \sim Be(\alpha, \beta)$, se sua função de densidade é dada por

$$p(x|\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1,$$

para $\alpha, \beta > 0$.

$$E(X) = \frac{\alpha}{\alpha + \beta} \quad \text{e} \quad V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

A.7 Distribuição de Dirichlet

O vetor aleatório $\mathbf{X} = (X_1, \dots, X_k)$ tem distribuição de Dirichlet com parâmetros $\alpha_1, \dots, \alpha_k$, denotada por $D_k(\alpha_1, \dots, \alpha_k)$ se sua função de densidade conjunta é dada por

$$p(\mathbf{x}|\alpha_1, \dots, \alpha_k) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)\dots\Gamma(\alpha_k)} x_1^{\alpha_1-1} \dots x_k^{\alpha_k-1}, \quad \sum_{i=1}^k x_i = 1,$$

para $\alpha_1, \dots, \alpha_k > 0$ e $\alpha_0 = \sum_{i=1}^k \alpha_i$.

$$E(X_i) = \frac{\alpha_i}{\alpha_0}, \quad V(X_i) = \frac{(\alpha_0 - \alpha_i)\alpha_i}{\alpha_0^2(\alpha_0 + 1)}, \quad \text{e} \quad \text{Cov}(X_i, X_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0 + 1)}$$

Note que a distribuição Beta é obtida como caso particular para $k = 2$.

A.8 Distribuição t de Student

X tem distribuição t de Student (ou simplesmente t) com média μ , parâmetro de escala σ e ν graus de liberdade, denotando-se $X \sim t_\nu(\mu, \sigma^2)$, se sua função de densidade é dada por

$$p(x|\nu, \mu, \sigma^2) = \frac{\Gamma((\nu+1)/2)\nu^{\nu/2}}{\Gamma(\nu/2)\sqrt{\pi}\sigma} \left[\nu + \frac{(x-\mu)^2}{\sigma^2} \right]^{-(\nu+1)/2}, \quad x \in \mathbb{R},$$

para $\nu > 0$, $\mu \in \mathbb{R}$ e $\sigma^2 > 0$.

$$E(X) = \mu, \quad \text{para } \nu > 1 \quad \text{e} \quad V(X) = \frac{\nu}{\nu-2}, \quad \text{para } \nu > 2.$$

Um caso particular da distribuição t é a distribuição de Cauchy, denotada por $C(\mu, \sigma^2)$, que corresponde a $\nu = 1$.

A.9 Distribuição F de Fisher

X tem distribuição F com ν_1 e ν_2 graus de liberdade, denotando-se $X \sim F(\nu_1, \nu_2)$, se sua função de densidade é dada por

$$p(x|\nu_1, \nu_2) = \frac{\Gamma((\nu_1 + \nu_2)/2)}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} \nu_1^{\nu_1/2} \nu_2^{\nu_2/2} x^{\nu_1/2-1} (\nu_2 + \nu_1 x)^{-(\nu_1+\nu_2)/2}$$

$x > 0$, e para $\nu_1, \nu_2 > 0$.

$$E(X) = \frac{\nu_2}{\nu_2 - 2}, \quad \text{para } \nu_2 > 2 \quad \text{e} \quad V(X) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 4)(\nu_2 - 2)^2}, \quad \text{para } \nu_2 > 4.$$

A.10 Distribuição Binomial

X tem distribuição binomial com parâmetros n e p , denotando-se $X \sim \text{bin}(n, p)$, se sua função de probabilidade é dada por

$$p(x|n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, \dots, n$$

para $n \geq 1$ e $0 < p < 1$.

$$E(X) = np \quad \text{e} \quad V(X) = np(1-p)$$

e um caso particular é a distribuição de Bernoulli com $n = 1$.

A.11 Distribuição Multinomial

O vetor aleatório $\mathbf{X} = (X_1, \dots, X_k)$ tem distribuição multinomial com parâmetros n e probabilidades $\theta_1, \dots, \theta_k$, denotada por $M_k(n, \theta_1, \dots, \theta_k)$ se sua função de probabilidade conjunta é dada por

$$p(\mathbf{x}|\theta_1, \dots, \theta_k) = \frac{n!}{x_1! \dots x_k!} \theta_1^{x_1} \dots \theta_k^{x_k}, \quad x_i = 0, \dots, n, \quad \sum_{i=1}^k x_i = n,$$

para $0 < \theta_i < 1$ e $\sum_{i=1}^k \theta_i = 1$. Note que a distribuição binomial é um caso especial da multinomial quando $k = 2$. Além disso, a distribuição marginal de cada X_i é binomial com parâmetros n e θ_i e

$$E(X_i) = n\theta_i, \quad V(X_i) = n\theta_i(1-\theta_i), \quad \text{e} \quad \text{Cov}(X_i, X_j) = -n\theta_i\theta_j.$$

A.12 Distribuição de Poisson

X tem distribuição de Poisson com parâmetro θ , denotando-se $X \sim \text{Poisson}(\theta)$, se sua função de probabilidade é dada por

$$p(x|\theta) = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0, 1, \dots$$

para $\theta > 0$.

$$E(X) = V(X) = \theta.$$

A.13 Distribuição Binomial Negativa

X tem distribuição de binomial negativa com parâmetros r e p , denotando-se $X \sim \text{BN}(r, p)$, se sua função de probabilidade é dada por

$$p(x|r, p) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, \dots$$

para $r \geq 1$ e $0 < p < 1$.

$$E(X) = r(1 - p)/p \quad \text{e} \quad V(X) = r(1 - p)/p^2.$$

Um caso particular é quando $r = 1$ e neste caso diz-se que X tem distribuição geométrica com parâmetro p .

Referências

- Bauwens, L., Lubrano, M. & Richard, J. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.
- Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, California.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control* (Third ed.). Englewood Cliffs NJ: Prentice-Hall.
- Brockwell, P. & Davis, R. (1991). *Time Series: Theory and Methods* (2nd ed.). New York: Springer-Verlag.
- Burnham, K. P. & Anderson, D. R. (1998). *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer: New York.
- Diggle, P. (1990). *Time Series: A Biostatistical Introduction*. Oxford University Press: New York.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
- Franses, P. H. (1998). *Time Series Models for Business and Economic Forecasting*. Cambridge University Press.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions* (2nd ed.), Volume 1. John Wiley, New York.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions* (2nd ed.), Volume 2. John Wiley, New York.
- Johnson, N. L., Kotz, S. & Kemp, A. W. (1992). *Univariate Discrete Distributions* (2nd ed.). John Wiley, New York.
- Kendall, M. G., Stuart, A. & Ord, J. K. (1983). *Advanced theory of statistics* (4th ed.), Volume 3. Griffin: London.
- Pole, A., West, M. & Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Texts in Statistical Sciences. Chapman & Hall.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press.
- Taylor, S. (1986). *Modelling Financial Time Series*. Wiley.

Tsay, R. S. (2002). *Analysis of Financial Time Series*. Wiley.

West, M. & Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer Verlag, New York.