



# **ECONOMETRIA FINANCEIRA**

Um curso em séries temporais financeiras

2ª edição revista e ampliada

PEDRO A. MORETTIN

**Blucher**

---

# Conteúdo

<b>Prefácio</b>	<b>v</b>
<b>1 Preliminares</b>	<b>1</b>
1.1 Introdução . . . . .	1
1.2 Tipos de Dados . . . . .	2
1.3 Retornos . . . . .	8
1.4 Agregação de Retornos . . . . .	10
1.5 Distribuição de Retornos . . . . .	13
1.6 Assimetria e Curtose . . . . .	16
1.7 Fatos Estilizados Sobre os Retornos . . . . .	18
1.8 Volatilidade . . . . .	20
1.9 Aspectos Computacionais . . . . .	23
1.10 Problemas . . . . .	24
Apêndice 1.A: Distribuições Estáveis . . . . .	26
Apêndice 1.B: Teste de Normalidade . . . . .	28
<b>2 Processos Estocásticos</b>	<b>31</b>
2.1 Processos Estacionários . . . . .	31
2.2 Especificação de um Processo Estocástico . . . . .	33
2.3 Propriedades da Função de Auto-covariância . . . . .	37
2.4 Processos Estocásticos Complexos . . . . .	40
2.5 Processos Lineares Estacionários . . . . .	41

2.5.1	Processos Auto-regressivos . . . . .	44
2.5.2	Processos de Médias Móveis . . . . .	48
2.5.3	Processos Auto-regressivos e de Médias Móveis . . . . .	52
2.6	Processos Não-Estacionários . . . . .	56
2.7	Movimento Browniano . . . . .	62
2.8	Martingales . . . . .	69
2.9	Problemas . . . . .	72
<b>3</b>	<b>Modelagem ARIMA</b>	<b>75</b>
3.1	Introdução . . . . .	75
3.2	Identificação . . . . .	76
3.2.1	Procedimento de Identificação . . . . .	82
3.2.2	Formas alternativas de identificação . . . . .	89
3.3	Estimação . . . . .	94
3.3.1	Modelos Auto-regressivos . . . . .	94
3.3.2	Modelos de Médias Móveis . . . . .	96
3.3.3	Modelos Mistos . . . . .	97
3.3.4	Variâncias dos estimadores . . . . .	98
3.4	Diagnóstico . . . . .	100
3.4.1	Teste de auto-correlação residual . . . . .	101
3.4.2	Teste de Box-Pierce-Ljung . . . . .	102
3.5	Previsão com Modelos ARIMA . . . . .	104
3.6	Modelos Sazonais . . . . .	110
3.7	Problemas . . . . .	114
<b>4</b>	<b>Raízes Unitárias</b>	<b>123</b>
4.1	Introdução . . . . .	123
4.2	O Teste de Dickey-Fuller . . . . .	124
4.3	Extensões do Teste DF . . . . .	129
4.4	Comentários Finais . . . . .	136
4.5	Problemas . . . . .	141
	Apêndice 4: Provas dos Teoremas 4.1 e 4.2 . . . . .	143
<b>5</b>	<b>Modelagem da Volatilidade</b>	<b>147</b>
5.1	Introdução . . . . .	147
5.2	Alguns Modelos Não-Lineares . . . . .	148

5.3	Modelos ARCH . . . . .	150
5.4	Modelos GARCH . . . . .	160
5.5	Extensões do Modelo GARCH . . . . .	168
5.5.1	Modelos EGARCH . . . . .	168
5.5.2	Modelos TARCH . . . . .	173
5.6	Modelos de Volatilidade Estocástica . . . . .	176
5.7	Problemas . . . . .	182
	Apêndice 5. Algumas Distribuições Especiais . . . . .	186
<b>6</b>	<b>Valor em Risco</b>	<b>189</b>
6.1	Introdução . . . . .	189
6.2	Valor em Risco . . . . .	190
6.3	VaR Usando a Distribuição Normal . . . . .	192
6.4	VaR Usando Modelos ARMA e GARCH . . . . .	199
6.5	VaR Usando Quantis Empíricos . . . . .	203
6.6	VaR Usando a Teoria de Valores Extremos . . . . .	204
6.7	Tópicos Adicionais . . . . .	208
6.8	Problemas . . . . .	209
	Apêndice 6: Teoria de Valores Extremos . . . . .	211
<b>7</b>	<b>Modelos Lineares Multivariados</b>	<b>217</b>
7.1	Introdução . . . . .	217
7.2	Séries Estacionárias . . . . .	218
7.3	Estimação de Médias e Covariâncias . . . . .	222
7.4	Modelos Auto-regressivos Vetoriais . . . . .	224
7.5	Construção de Modelos VAR . . . . .	230
7.6	Modelos ARMA Vetoriais . . . . .	235
7.7	Causalidade de Granger . . . . .	241
7.8	Problemas . . . . .	247
	Apêndice 7.A: Alguns Resultados sobre Matrizes . . . . .	249
	Apêndice 7.B: Demonstração da Proposição 7.2 . . . . .	254
	Apêndice 7.C: Modelo VAR( $p$ ) na Forma VAR(1) . . . . .	255
	Apêndice 7.D: Modelos Estruturais . . . . .	256

<b>8</b>	<b>Processos Co-Integrados</b>	<b>259</b>
8.1	Introdução . . . . .	259
8.2	Tendências Comuns . . . . .	262
8.3	Modelo de Correção de Erros . . . . .	264
8.4	Testes para Co-integração . . . . .	270
8.4.1	Procedimento de Engle e Granger . . . . .	270
8.4.2	Procedimento de Johansen . . . . .	273
8.5	Comentários Finais . . . . .	282
8.5.1	Enfoque Bayesiano . . . . .	282
8.5.2	Termos Determinísticos no MCE . . . . .	283
8.5.3	Procedimento de Máxima Verossimilhança . . . . .	284
8.6	Problemas . . . . .	285
<b>9</b>	<b>Processos com Memória Longa</b>	<b>289</b>
9.1	Introdução . . . . .	289
9.2	Estimação e Testes para Memória Longa . . . . .	293
9.2.1	Estatística R/S . . . . .	294
9.2.2	Procedimento GPH . . . . .	296
9.3	Modelos ARFIMA . . . . .	299
9.4	Estimação de modelos ARFIMA . . . . .	305
9.5	Previsão de modelos ARFIMA . . . . .	308
9.6	Processos de Volatilidade com ML . . . . .	311
9.7	Problemas . . . . .	313
	Apêndice 9: Volatilidade de Garman-Klass . . . . .	317
	<b>Referências</b>	<b>321</b>
	<b>Séries Usadas no Texto</b>	<b>335</b>

---

# Prefácio

Este livro trata da aplicação de técnicas de séries temporais e econometria a dados financeiros. Séries temporais econômicas e financeiras apresentam certas características especiais. Uma delas é o chamado agrupamento de volatilidades, significando que a variância da série evolve no tempo e depende da informação passada. Além disso, podemos ter séries, como de taxas de câmbio e de preços de ações, que podem ser registradas em intervalos irregulares no decorrer do dia, dando origem aos chamados dados de alta frequência.

O texto é resultado de cursos ministrados pelo autor por vários anos no Instituto de Matemática e Estatística da Universidade de São Paulo. O conteúdo do livro pode ser ministrado para alunos do final da graduação, em áreas como Estatística, Economia e Finanças e para alunos de mestrado nestas áreas e outras afins. Seria desejável que os alunos tivessem noções provenientes de cursos básicos de probabilidade e estatística, bem como de séries temporais ou econometria.

Pressupõe-se que um curso com base neste livro tenha como boa parte de seu conteúdo dedicada a análises de séries reais com uso intensivo de pacotes computacionais apropriados. Utilizamos alguns pacotes disponíveis aos usuários, como o **S+FinMetrics**, **EViews**, **Minitab** e **SCA**. Outras possibilidades são o pacote **R**, acessível gratuitamente na Internet, **RATS**, **MatLab** etc.

Vários colegas e alunos contribuíram com comentários e sugestões no decorrer da elaboração deste livro. Em particular, Silvia R. C. Lopes e Beatriz V. M. Mendes leram versões prévias do livro e apresentaram correções, comentários

e sugestões que melhoraram consideravelmente o texto. Os erros remanescentes são, obviamente, de minha responsabilidade.

Gostaria, também, de agradecer à Diretoria da Associação Brasileira de Estatística pelo convite para ministrar um minicurso no 17<sup>o</sup> SINAPE, baseado no conteúdo deste livro.

Os dados usados e outros aspectos computacionais relacionados ao livro encontram-se em minha página <http://www.ime.usp.br/~pam>. Comentários e sugestões podem ser enviados ao meu endereço eletrônico, [pam@ime.usp.br](mailto:pam@ime.usp.br).

São Paulo, março de 2006

Pedro A. Morettin

Este livro é dedicado à  
memória de meus pais, Iole e  
Luiz, que agora estão num lugar melhor.



# Preliminares

## 1.1 Introdução

Este livro trata da análise de séries temporais financeiras. Em princípio não haveria diferenças entre a análise de tais séries e aquelas ocorrendo em outras áreas, como economia, oceanografia, meteorologia, etc. De fato, a maioria das técnicas de análise de séries temporais são aplicáveis em diversas áreas. Contudo, uma característica presente em séries de ativos financeiros é o que se convencionou chamar de *volatilidade*, que pode ser definida de várias maneiras, mas não é diretamente observável. Para levar em conta a presença de grupos (“clusters”) de volatilidade em uma série financeira é necessário recorrer a modelos ditos heteroscedásticos condicionais. Nestes modelos, a variância (volatilidade) de um retorno num dado instante de tempo, depende de retornos passados e de outras informações disponíveis até aquele instante, de modo que temos que definir uma *variância condicional*, que não sendo constante, não coincide com a variância global (“incondicional”) da série observada. Do mesmo modo, é possível que a média varie com o tempo, ou outros momentos da distribuição dos retornos variem com o tempo.

Uma característica marcante de séries financeiras é que elas são, em geral, não serialmente correlacionadas, mas dependentes. Deste modo, modelos lineares como aqueles pertencentes à família dos modelos ARMA (auto-regressivos e de médias móveis) podem não ser apropriados para descrever tais séries. Modelos da família ARCH (de “autoregressive conditional heteroscedasticity”) ou modelos de volatilidade estocástica são mais adequados. Contudo, diver-

sas séries apresentam alguma forma de auto-correlação, de modo que modelos ARMA podem ser inicialmente aplicados para remover esta correlação, antes de usar modelos heteroscedásticos. Por esse motivo, nos capítulos 2 e 3 apresentamos uma introdução aos modelos ARMA. Para uma análise mais detalhada o leitor poderá consultar Box, Jenkins e Reinsel (1994) ou Morettin e Toloi (2006).

A maior parte do livro analisará séries de retornos de ativos e não os preços desses ativos. Por essa razão, ainda neste primeiro capítulo faremos uma revisão de conceitos relativos a retornos. Além dos modelos ARMA, no Capítulo 2 apresentaremos outros processos estocásticos que são importantes na análise de séries temporais aparecendo em economia e finanças.

No Capítulo 4 tratamos do importante problema de raízes unitárias e como testar sua presença em dada série. No Capítulo 5 introduzimos brevemente alguns modelos não lineares e depois tratamos os modelos heteroscedásticos condicionais. Um dos problemas mais importantes atualmente em finanças é avaliar o risco de uma posição financeira, e o VaR (valor em risco) é um instrumento freqüentemente usado. No Capítulo 6 introduzimos várias formas de se calcular o valor em risco. No Capítulo 7 estendemos o estudo para englobar modelos lineares multivariados, analisando os modelos VAR (modelos auto-regressivos vetoriais) e VARMA (modelos auto-regressivos e de médias móveis vetoriais). O conceito de co-integração e o modelo de correção de erros são discutidos no Capítulo 8. Finalmente, no Capítulo 9, tratamos os processos de memória longa.

## 1.2 Tipos de Dados

Nesta seção descreveremos os diversos tipos de dados que são mais comuns na análise de séries financeiras. Numa primeira categoria, temos observações igualmente espaçadas: o intervalo  $\Delta t$  entre observações consecutivas é constante, por exemplo, um dia, uma semana, um mês. Quando analisamos dados diários, usualmente utilizamos o último valor observado no dia, como o preço de fechamento de uma ação numa bolsa de valores. Algumas vezes pode ser um valor agregado durante o período, como o volume (em moeda) negociado de dada ação na bolsa durante um dia.

Os dados podem ser observados em instantes de tempo irregularmente

espaçados, como os dados intradiários de ativos negociados em bolsas de valores ou de mercadorias, ou taxas de câmbio. Nestes casos, os intervalos entre observações são variáveis aleatórias (as chamadas “durações”) e podemos ter também várias observações (negócios) coincidindo num mesmo instante de tempo. Este tipo de dado é chamado de *alta frequência*.

As séries financeiras que serão usadas no texto estão listadas no final do livro e podem ser acessadas no site <http://www.ime.usp.br/~pam>.

**Exemplo 1.1.** Na Figura 1.1 (a) temos o gráfico dos índices diários da Bolsa de Valores de São Paulo (Ibovespa) no período de 3 de janeiro de 1995 a 27 de dezembro de 2000, num total de  $T = 1499$  observações. O arquivo d-ibv95.00.dat contém estes dados, que aparecem no arquivo na forma a seguir:

Date	IBOVESPA
2/jan/95	4319.0
3/jan/95	4097.9
4/jan/95	3967.9
5/jan/95	4036.6
6/jan/95	3827.8
...	...

Estes são os valores de fechamento do índice. Podemos ter outras informações, como os valores de abertura, mínimo e máximo, por exemplo.

**Exemplo 1.2.** Na Figura 1.2 (a) mostramos o gráfico dos índices diários do Dow Jones Industrial Average (fechamento), DJIA, no período de 3 de janeiro de 1995 a 26 de dezembro de 2002, com  $T = 1992$  observações. Estes dados encontram-se no arquivo d-dow95.02.dat. Na tabela a seguir temos parte do arquivo de dados, onde aparecem os índices de abertura, máximo, mínimo, fechamento e o volume negociado no dia.

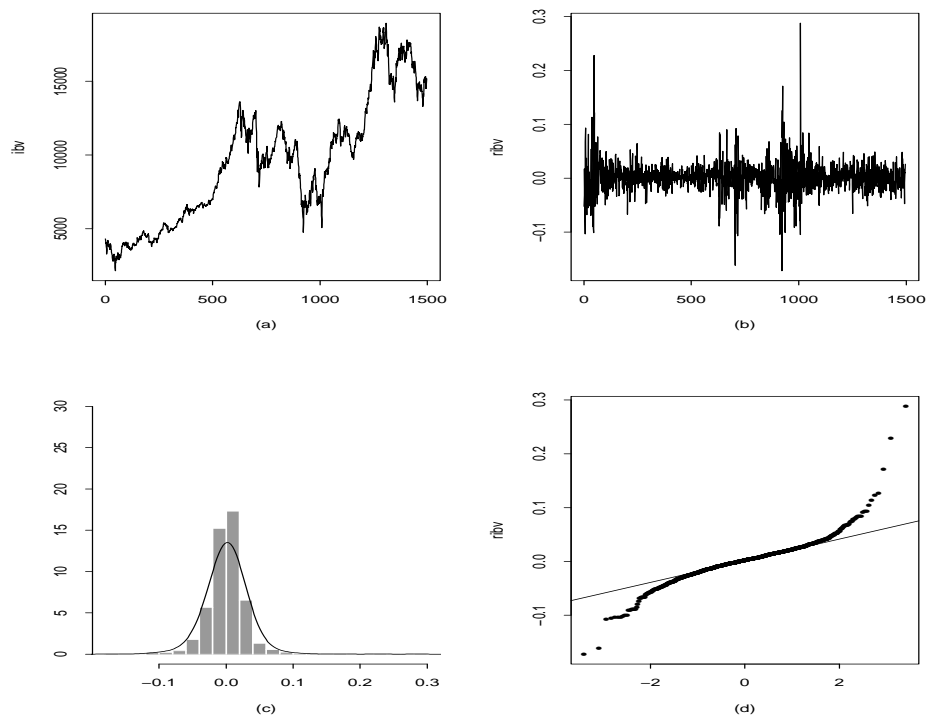


Figura 1.1: (a) Gráfico da série Ibovespa (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$

Date	Open	High	Low	Close	Volume
3/jan/95	3834.4	3864.7	3805.5	3838.5	2624500
4/jan/95	3838.5	3876.8	3815.3	3857.7	3195100
5/jan/95	3857.7	3876.8	3825.4	3850.9	3091400
6/jan/95	3850.9	3902.4	3823.7	3867.4	3080700
9/jan/95	3867.4	3889.3	3834.4	3861.4	2787100
...	...	...	...	...	...

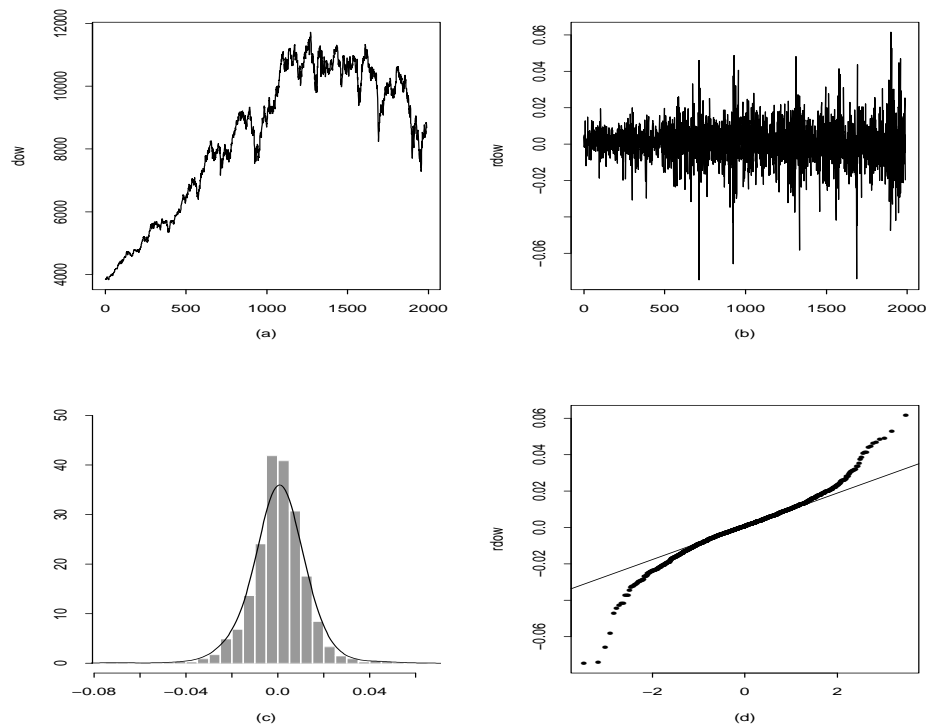


Figura 1.2: (a) Gráfico da série DJIA (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$

Dados de alta frequência podem ser registrados de diferentes formas, dependendo do tipo de ativo. Na tabela abaixo temos um “layout” típico de dados de ações negociadas na Bolsa de Valores de São Paulo.

Data	Código	Empresa	Tipo	Pr. Negócio	Hora	Nº Negócio
2003-02-03;	BBDC4;	BRABESCO;	PN*N1;	9.9900;	1101;	10;
2003-02-03;	BBDC4;	BRABESCO;	PN*N1;	10.0000;	1101;	20;
2003-02-03;	BBDC4;	BRABESCO;	PN*N1;	10.0000;	1101;	30;
2003-02-03;	BBDC4;	BRABESCO;	PN*N1;	10.0500;	1102;	40;
2003-02-03;	BBDC4;	BRABESCO;	PN*N1;	10.0500;	1102;	50;
...	...	...	...	...	...	...

Este arquivo de dados traz os preços das ações Bradesco PN, no dia 3 de fevereiro de 2003. As outras colunas trazem informação sobre o código do ativo, o preço e o número do negócio. Note que há três observações no mesmo instante de tempo, onze horas e um minuto. Para se ter uma quantidade razoável de dados intradiários para análise o ativo deve ter uma grande liquidez, o que não acontece com a maioria dos papéis negociados na Bolsa de Valores de São Paulo, por exemplo.

A seguir ilustramos os dados intradiários do Ibovespa e da Telemar PN, encontrados nos arquivos `id-ibv98.03.dat` e `id-tel02.05.dat`. Nestes arquivos os dados foram interpolados, de modo a se ter observações a cada quinze minutos.

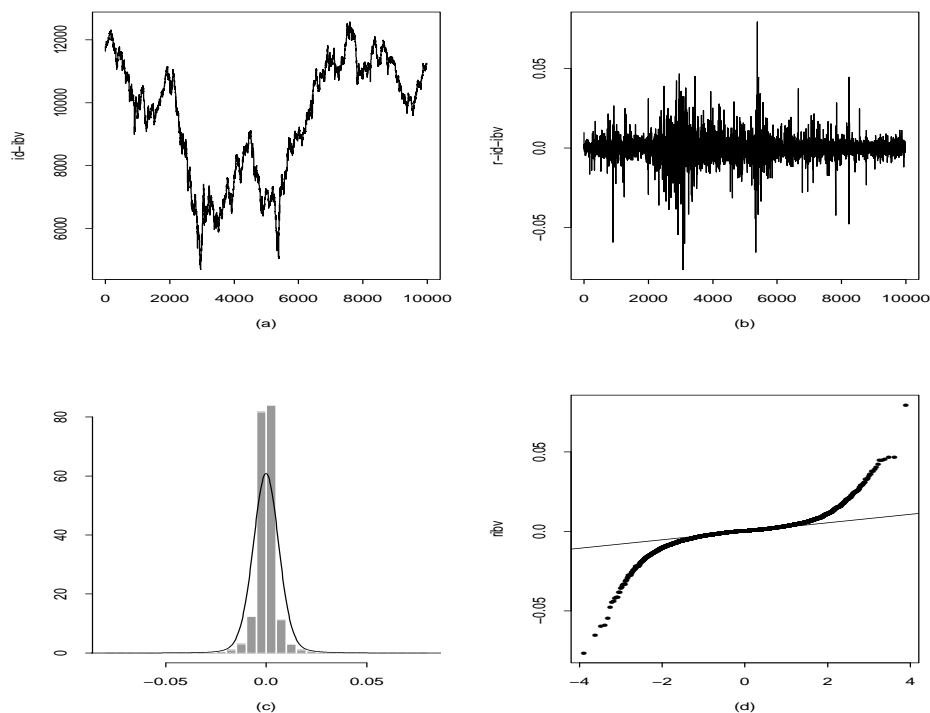


Figura 1.3: (a) Gráfico da série Ibovespa intradiária (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$

**Exemplo 1.3.** Na Figura 1.3 (a) temos o gráfico de parte dos dados do Ibovespa, observados a cada quinze minutos, de 6 de abril de 1998 a 13 de agosto de 2003, num total de 1309 dias e  $T = 37.961$  observações. Na Figura 1.4 (a) temos o gráfico de parte dos dados da Telemar PN, observados a cada 15 minutos, de 2 de janeiro de 2002 a 31 de março de 2005, com  $T = 21.429$  observações.

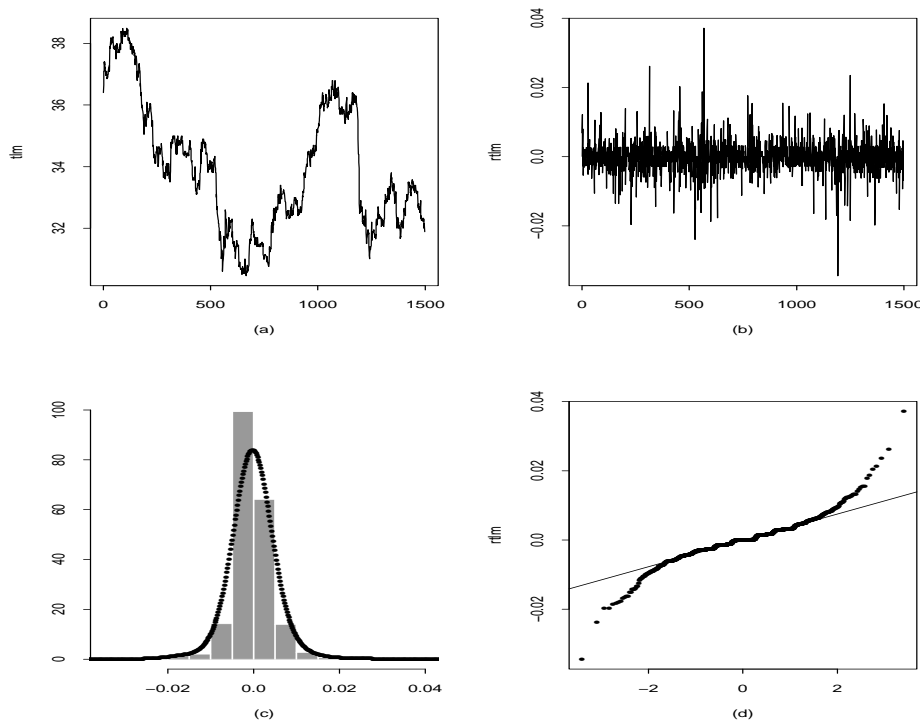


Figura 1.4: (a) Gráfico da série Telemar intradiária (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$

### 1.3 Retornos

Um dos objetivos em finanças é a avaliação de riscos de uma carteira de ativos (instrumentos) financeiros. O risco é freqüentemente medido em termos de variações de preços dos ativos.

Denotemos por  $P_t$  o preço de um ativo no instante  $t$ , normalmente um dia de negócio. Suponha, primeiramente, que não haja dividendos pagos no período. A variação de preços entre os instantes  $t - 1$  e  $t$  é dada por  $\Delta P_t = P_t - P_{t-1}$  e a variação relativa de preços ou *retorno líquido simples* deste ativo entre os mesmos instantes é definido por

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{\Delta P_t}{P_{t-1}}. \quad (1.1)$$

Note que  $R_t = P_t/P_{t-1} - 1$ . Chamamos  $1 + R_t = P_t/P_{t-1}$  de *retorno bruto simples*. Usualmente expressamos  $R_t$  em percentagem, relativamente ao período (um dia, um mês, um ano, etc). É também chamado de *taxa de retorno*.

Denotando  $p_t = \log P_t$  (sendo o logaritmo na base  $e$ ), definimos o *retorno composto continuamente* ou simplesmente *log-retorno* como

$$r_t = \log \frac{P_t}{P_{t-1}} = \log(1 + R_t) = p_t - p_{t-1}. \quad (1.2)$$

Esta definição será aquela comumente utilizada e, muitas vezes,  $r_t$  será chamado simplesmente de *retorno*. Note que, de (1.2), obtemos  $R_t = e^{r_t} - 1$ . Usaremos, no que segue, a notação  $\log(a)$  para designar o logaritmo de  $a$  na base  $e$ .

Na prática é preferível trabalhar com retornos, que são livres de escala, do que com preços, pois os primeiros têm propriedades estatísticas mais interessantes (como estacionariedade e ergodicidade). Um dos objetivos será, então, modelar retornos. Diversas classes de modelos podem ser utilizadas para esse fim, tais como os modelos ARMA, ARCH, GARCH, modelos de volatilidade estocástica, etc. Esses modelos serão estudados nos capítulos seguintes.

Note também que, para  $u$  pequeno,  $\log(1 + u) \approx u$ , do que segue que os retornos simples  $R_t$  e os log-retornos  $r_t$  serão em geral valores próximos.

Podemos definir também retornos multiperíodos. O retorno simples de período  $k$ , entre os instantes  $t - k$  e  $t$  é dado por



$$R_t[k] = \frac{P_t - P_{t-k}}{P_{t-k}}. \quad (1.3)$$

Em termos de retornos de um período podemos escrever (1.3) como

$$\begin{aligned} 1 + R_t[k] &= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}) \\ &= \frac{P_t}{P_{t-1}} \frac{P_{t-1}}{P_{t-2}} \cdots \frac{P_{t-k+1}}{P_{t-k}} = \frac{P_t}{P_{t-k}}, \end{aligned}$$

de modo que

$$R_t[k] = \frac{P_t}{P_{t-k}} - 1. \quad (1.4)$$

Para facilitar comparações em horizontes diferentes é comum “anualizar” os retornos simples, considerando

$$R_t[k]_{\text{anualizado}} = [\prod_{j=0}^{k-1} (1 + R_{t-j})]^{1/k} - 1,$$

que pode ser aproximado por  $(1/k) \sum_{j=0}^{k-1} R_{t-j}$ , usando uma expansão de Taylor até primeira ordem.

Por sua vez, o log-retorno de período  $k$  fica

$$r_t[k] = \log \frac{P_t}{P_{t-k}} = \log(1 + R_t[k]) = \sum_{j=0}^{k-1} \log(1 + R_{t-j}) = \sum_{j=0}^{k-1} r_{t-j}. \quad (1.5)$$

Por exemplo, um mês compreende normalmente cerca de 21 dias de transações, de modo que o log-retorno continuamente composto em um mês é dado por

$$r_t[21] = r_t + r_{t-1} + \cdots + r_{t-20},$$

para todo  $t$ .

A expressão (1.5) é interessante do ponto de vista estatístico, pois para  $k$  relativamente grande a soma pode ser aproximada por uma v.a. normal, usando o teorema limite central.

Se houver pagamento de dividendos  $D_t$  no período, então os retornos ficam, respectivamente,

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1, \quad (1.6)$$

$$r_t = \log(1 + R_t) = \log(P_t + D_t) - \log P_{t-1}. \quad (1.7)$$

Vemos que  $r_t$  é uma função não-linear de log-preços e log-dividendos.

**Exemplo 1.4.** Considere os índices diários do Ibovespa do exemplo 1.1 e sejam  $P_1 = 4319,0, \dots, P_5 = 3827,8$ . Então,

$$\begin{aligned} R_2 &= \frac{P_2 - P_1}{P_1} = -0,0512, \\ r_2 &= \log \frac{P_2}{P_1} = -0,0525, \\ R_5[3] &= \frac{P_5 - P_2}{P_2} = -0,0659, \\ r_5[3] &= r_5 + r_4 + r_3 = \log \frac{P_5}{P_2} = -0,0682. \end{aligned}$$

**Exemplo 1.1.** (continuação) Na Figura 1.1 (b) temos o gráfico dos retornos diários do Ibovespa. Note que estes retornos oscilam ao redor do zero, aparentam ser estacionários, mas apresentam uma variabilidade que depende do tempo (a chamada “volatilidade”), com períodos de alta variabilidade, períodos de baixa variabilidade e dias em que o retorno é um valor atípico (“outlier”), quando comparado com os demais.

**Exemplo 1.2.** (continuação). A Figura 1.2 (b) apresenta os retornos diários do DJIA, que tem características semelhantes aos retornos do Ibovespa.

## 1.4 Agregação de Retornos

Na equação (1.5) temos o que se chama de *agregação temporal* dos retornos. Podemos ter, também, uma agregação “cross-section”, para diversos ativos de

uma carteira de investimentos,  $c$ . Suponha que esta contenha  $N$  instrumentos  $A_1, \dots, A_N$ , com pesos  $w_1, \dots, w_N$ , com  $\sum_{i=1}^N w_i = 1$ . Denotemos por  $R_i$  os retornos simples e por  $r_i$  os log-retornos desses ativos,  $i = 1, \dots, N$ . Se  $P_0$  indicar o preço inicial da carteira, após um período teremos, para retornos continuamente compostos,

$$\frac{P_1}{P_0} = \sum_{i=1}^N w_i e^{r_i}. \quad (1.8)$$

O log-retorno da carteira é  $r_c = \log \frac{P_1}{P_0}$ , logo obtemos

$$r_c = \log \sum_{i=1}^N w_i e^{r_i}, \quad (1.9)$$

enquanto que o retorno simples da carteira é

$$\begin{aligned} R_c &= \frac{P_1}{P_0} - 1 \\ &= \sum_{i=1}^N w_i e^{r_i} - 1 \\ &= \sum_{i=1}^N w_i (1 + R_i) - 1, \end{aligned}$$

ou seja,

$$R_c = \sum_{i=1}^N w_i R_i. \quad (1.10)$$

No caso de composição discreta teremos

$$\frac{P_1}{P_0} = \sum_{i=1}^N w_i (1 + r_i), \quad (1.11)$$

de modo que o retorno simples da carteira é  $R_c = (P_1 - P_0)/P_0$ , ou seja,

$$R_c = \sum_{i=1}^N w_i(1 + r_i) - 1 = \sum_{i=1}^N w_i r_i. \quad (1.12)$$

Vemos, pois, de (1.10) e (1.12), que o retorno simples é uma soma ponderada de retornos simples, no caso de composição contínua, e uma soma ponderada de log-retornos, no caso de composição discreta.

De modo geral podemos ter:

(i) Agregação temporal: para  $i = 1, \dots, N$ ,

$$R_{i,t}[k] = \prod_{j=0}^{k-1} (1 + R_{i,t-j}) - 1, \quad (1.13)$$

$$r_{i,t} = \sum_{j=0}^{k-1} r_{i,t-j}, \quad (1.14)$$

para retornos simples e log-retornos, respectivamente.

(ii) Agregação “cross-section”: para a carteira  $c$  e período  $t$ ,

$$R_{ct} = \sum_{i=1}^N w_i R_{i,t}, \quad (1.15)$$

$$r_{ct} = \log \left( \sum_{i=1}^N w_i e^{r_{it}} \right). \quad (1.16)$$

Para agregação temporal é mais conveniente trabalhar com log-retornos, enquanto que para agregação de ativos os retornos simples são mais convenientes. Como já salientamos trabalha-se normalmente com log-retornos e a expressão (1.16) é aproximada por

$$r_{ct} \approx \sum_{i=1}^N w_i r_{i,t}.$$

**Exemplo 1.5.** Na Figura 1.5 (a) temos os retornos mensais do Ibovespa, no período de junho de 1994 a agosto de 2001, com  $T = 86$  dados (arquivo m-ibv94.01.dat). Estes retornos mensais são obtidos usando-se a fórmula (1.5),

ou seja, somando-se os retornos diários. Observe que obtemos uma série mais suave, ou seja, com menor variabilidade do que a série de retornos diários.

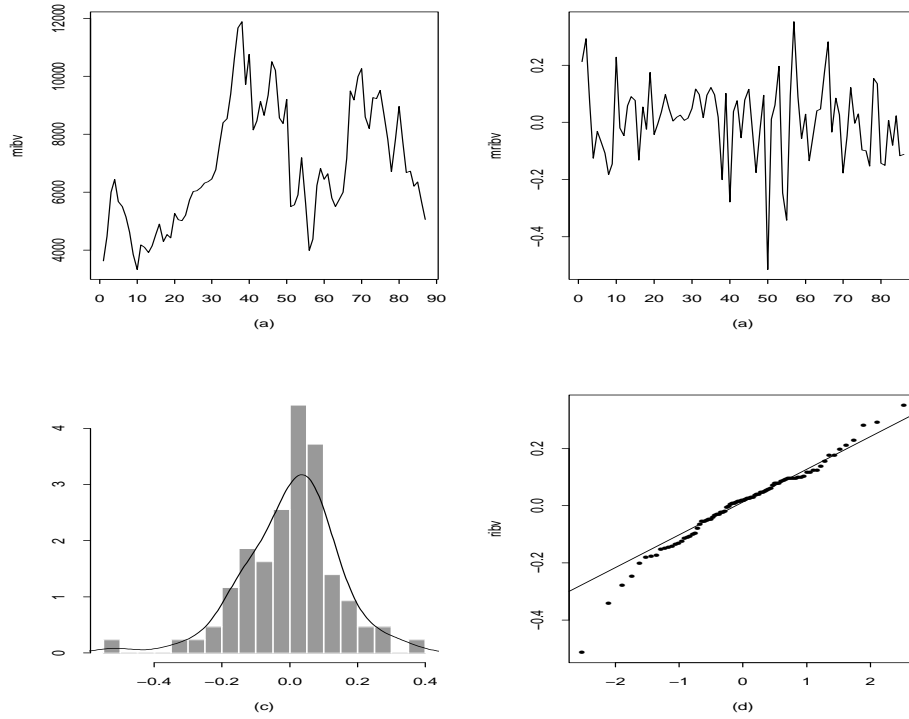


Figura 1.5: (a) Gráfico da série dos retornos mensais do Ibovespa (b) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$

## 1.5 Distribuição de Retornos

Considere, inicialmente, uma série de retornos  $\{r_t, t = 1, \dots, T\}$ , observados em instantes de tempo igualmente espaçados. Esta série pode ser considerada parte de uma realização de um processo estocástico  $\{r_t, t \in Z\}$ , onde  $Z = \{0, \pm 1, \pm 2, \dots\}$ . No Capítulo 2 trataremos da formalização deste conceito.

Veremos lá que o processo estará especificado completamente se conhecermos as distribuições finito-dimensionais

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P(r(t_1) \leq x_1, \dots, r(t_n) \leq x_n), \quad (1.17)$$

para quaisquer instantes de tempo  $t_1, \dots, t_n$  e qualquer  $n \geq 1$ . As distribuições (1.17) devem satisfazer certas condições. Contudo, na prática, é muito difícil (ou impossível) conhecer (1.17) e o que se faz é caracterizar o processo por momentos até determinada ordem, como a média

$$E(r_t) = \int_{-\infty}^{\infty} r dF(r; t) \quad (1.18)$$

ou a função de autocovariância

$$\gamma(t_1, t_2) = E(r_{t_1} r_{t_2}) - E(r_{t_1})E(r_{t_2}), \quad t_1, t_2 \in Z. \quad (1.19)$$

Outras suposições simplificadoras podem ser introduzidas, como condições de estacionariedade, ergodicidade ou normalidade do processo. Como vimos, os preços  $P_t$  em geral não são estacionários, ao passo que os log-retornos o são, donde o interesse nesses últimos. Todavia, a suposição de normalidade dos log-retornos em geral não é válida. Voltaremos a esse assunto mais tarde.

Por outro lado, se tivermos  $N$  ativos com retornos  $r_{it}$  em  $T$  instantes de tempo, teríamos que considerar as distribuições

$$F(r_{1,1}, \dots, r_{N,1}; \dots; r_{1,T}, \dots, r_{N,T}),$$

que usualmente podem depender de outras variáveis e parâmetros desconhecidos. Assim como no caso anterior, o estudo dessas distribuições é muito geral e há necessidade de introduzir restrições. Por exemplo, podemos supor que a distribuição é a mesma para todo instante de tempo (invariância temporal).

Podemos escrever (1.17) como (tomando-se  $t_i = i, i = 1, \dots, n$  e omitindo a dependência de  $F$  sobre estes tempos)

$$F(r_1, \dots, r_n) = F_1(r_1)F_2(r_2|r_1) \dots F_n(r_n|r_1, \dots, r_{n-1}). \quad (1.20)$$

No segundo membro de (1.20) temos as distribuições condicionais e podemos estar interessados em saber como essas evoluem no tempo. Uma hipótese

muitas vezes formulada é que os retornos são temporalmente independentes, ou seja, não são previsíveis usando retornos passados. Nesta situação, teremos que

$$F_t(r_t|r_1, \dots, r_{t-1}) = F_t(r_t).$$

Ergodicidade é uma propriedade mais difícil de estabelecer. Basicamente, um processo é ergódico se pudermos estimar características de interesse (média, autocovariância, etc) a partir de uma única trajetória do processo. Por exemplo, um processo é ergódico na média se a média amostral convergir, em probabilidade, para a média verdadeira do processo.

Uma outra suposição que às vezes é feita sobre a distribuição dos retornos é que esta seguiria uma distribuição estável. Veja o Apêndice 1.A para alguma informação sobre essas distribuições. Blattberg e Gonedes (1974) fazem uma comparação entre a distribuição  $t$  de Student e as distribuições estáveis como modelos para preços de ações. Veja também Mitnik et al. (1998) e Aparicio e Estrada (2001) para estudos similares.

A função de distribuição (1.20) depende, em geral, de co-variáveis  $\mathbf{Y}$  e de um vetor de parâmetros,  $\boldsymbol{\theta}$ , que a caracterizam. Supondo retornos com distribuição contínua, podemos obter de (1.20) a função de verossimilhança e a partir dela estimar  $\boldsymbol{\theta}$ . Por exemplo, supondo-se que as distribuições condicionais  $f_t(r_t|r_1, \dots, r_{t-1})$  sejam normais, com média  $\mu_t$  e variância  $\sigma_t^2$ , então  $\boldsymbol{\theta} = (\mu_t, \sigma_t^2)$  e a função de verossimilhança ficará

$$f(r_1, \dots, r_n; \boldsymbol{\theta}) = f_1(r_1; \boldsymbol{\theta}) \prod_{t=2}^n \frac{1}{\sigma_t \sqrt{2\pi}} \exp \left( -(r_t - \mu_t)^2 / 2\sigma_t^2 \right).$$

O estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  é obtido maximizando-se esta função ou o logaritmo dela.

Como vimos, podemos considerar  $N$  ativos ao longo do tempo,  $r_{1t}, r_{2t}, \dots, r_{Nt}$ , que podemos agrupar num vetor  $\mathbf{r}_t = (r_{1t}, r_{2t}, \dots, r_{Nt})'$ . Estaremos interessados em analisar a distribuição conjunta desses retornos e obteremos uma decomposição similar a (1.20). O interesse estará nas distribuições condicionais  $F_t(\mathbf{r}_t|\mathbf{r}_1, \dots, \mathbf{r}_{t-1}, \mathbf{Y}, \boldsymbol{\theta})$ .

**Exemplo 1.1.** (continuação) Na Figura 1.1 (c) temos o histograma dos retornos diários do Ibovespa, com uma densidade estimada a partir dos dados.

Vemos que o histograma tem a parte central mais alta do que uma normal e há a presença de valores bastante afastados da posição central dos dados. Estes fatos são característicos de retornos financeiros e são descritos pela chamada medida de curtose, a ser estudada na seção seguinte. Dizemos que os retornos são leptocúrticos, com caudas mais pesadas que a normal. Na Figura 1.1 (d) temos o gráfico  $Q \times Q$ , que é um gráfico que traz num dos eixos os quantis da normal padrão e no outro os quantis dos dados. Se os dados fossem aproximadamente normalmente distribuídos, os pontos estariam sobre uma reta, o que não acontece no caso em questão.

**Exemplo 1.2.** (continuação). As mesmas considerações do exemplo anterior aplicam-se aos retornos diários do DJIA, com os gráficos correspondentes sendo mostrados na Figura 1.2 (c) e Figura 1.2 (d).

**Exemplo 1.5.** (continuação) Nas figuras 1.5 (c) e 1.5(d) temos o histograma e gráfico  $Q \times Q$  para os retornos mensais do Ibovespa.

## 1.6 Assimetria e Curtose

Uma suposição muitas vezes utilizada é que os retornos  $r_t$  sejam independentes, identicamente distribuídos e normais (gaussianos). Contudo há argumentos contrários a essa suposição. Veja Campbell et al. (1997) para uma discussão mais elaborada. Se supusermos que os log-retornos  $r_t$  sejam normais, os retornos brutos serão *log-normais*, o que parece ser mais razoável.

De fato, se  $r_t \sim \mathcal{N}(\mu, \sigma^2)$ , então, como  $r_t = \log(1 + R_t)$ , segue-se que  $1 + R_t$  será log-normal, com

$$E(R_t) = e^{\mu + \sigma^2/2} - 1, \quad (1.21)$$

$$\text{Var}(R_t) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1). \quad (1.22)$$

Quando se considera a distribuição amostral dos retornos, nota-se que esta é aproximadamente simétrica, mas com *excesso de curtose*. Vamos discutir brevemente os conceitos de *assimetria* e *curtose*.

Seja  $X$  uma variável aleatória qualquer, com média  $\mu$  e variância  $\sigma^2$ . Então a *assimetria* de  $X$  é definida por



$$A(X) = E \left( \frac{(X - \mu)^3}{\sigma^3} \right), \quad (1.23)$$

enquanto que a *curtose* de  $X$  é definida por

$$K(X) = E \left( \frac{(X - \mu)^4}{\sigma^4} \right). \quad (1.24)$$

Para uma distribuição normal,  $A = 0$  e  $K = 3$ , donde a quantidade  $e(X) = K(X) - 3$  ser chamada *excesso de curtose*. Distribuições com caudas pesadas têm curtose maior do que 3 e esta pode mesmo ser infinita.

Com uma amostra  $X_1, \dots, X_T$  de  $X$ , considere o  $r$ -ésimo momento amostral

$$m_r = \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^r,$$

onde  $\hat{\mu} = \bar{X}$ . Substituindo os momentos verdadeiros de  $X$  pelos respectivos momentos amostrais, obtemos os estimadores

$$\hat{A}(X) = \frac{m_3}{m_2^{3/2}} = \frac{1}{T} \sum_{t=1}^T \left( \frac{X_t - \bar{X}}{\hat{\sigma}} \right)^3, \quad (1.25)$$

$$\hat{K}(X) = \frac{m_4}{m_2^2} = \frac{1}{T} \sum_{t=1}^T \left( \frac{X_t - \bar{X}}{\hat{\sigma}} \right)^4, \quad (1.26)$$

respectivamente, onde  $\hat{\sigma}^2 = \sum_{t=1}^T (X_t - \bar{X})^2 / T$ . Segue-se que  $\hat{e}(X) = \hat{K}(X) - 3$ .

Pode-se provar que, se tivermos uma amostra de uma distribuição normal e  $T$  for grande, então

$$\hat{A} \sim \mathcal{N}(0, 6/T), \quad \hat{K} \sim \mathcal{N}(3, 24/T). \quad (1.27)$$

Esses fatos podem ser utilizados para testar a normalidade de uma série. Veja o Apêndice 1.B.

Sabemos que os momentos amostrais são estimadores viesados dos respectivos momentos populacionais. Pode-se obter estimadores menos viesados, definindo-se os coeficientes de assimetria e curtose em termos de cumulantes. Veja Joanes e Gill (1998) para detalhes.

## 1.7 Fatos Estilizados Sobre os Retornos

Séries econômicas e financeiras apresentam algumas características que são comuns a outras séries temporais, como:

- (a) tendências;
- (b) sazonalidade;
- (c) pontos influentes (atípicos);
- (d) heteroscedasticidade condicional;
- (e) não-linearidade.

O leitor está, certamente, familiarizado com as características acima; para detalhes, veja Franses (1998). Dessas, a última talvez seja a mais complicada de definir. De um modo bastante geral, podemos dizer que uma série econômica ou financeira é não-linear quando responde de maneira diferente a choques grandes ou pequenos, ou ainda, a choques negativos ou positivos. Por exemplo, uma queda de um índice da Bolsa de Valores de São Paulo pode causar maior volatilidade no mercado do que uma alta.

Os retornos financeiros apresentam, por outro lado, outras características peculiares, que muitas séries não apresentam. Retornos raramente apresentam tendências ou sazonalidades, com exceção eventualmente de retornos intra-diários. Séries de taxas de câmbio e séries de taxas de juros podem apresentar tendências que variam no tempo.

Os principais *fatos estilizados* relativos a retornos financeiros podem ser resumidos como segue:

1. retornos são em geral não-auto-correlacionados;
2. os quadrados dos retornos são auto-correlacionados, apresentando uma correlação de lag um pequena e depois uma queda lenta das demais;
3. séries de retornos apresentam agrupamentos de volatilidades ao longo do tempo;

4. a distribuição (incondicional) dos retornos apresenta caudas mais pesadas do que uma distribuição normal; além disso, a distribuição, embora aproximadamente simétrica, é em geral leptocúrtica;
5. algumas séries de retornos são não-lineares, no sentido explicado acima.

**Exemplo 1.1.** (continuação) Na Figura 1.1 (b) temos a série de retornos do Ibovespa, na qual notamos os fatos estilizados apontadas antes, quais sejam, aparente estacionariedade, média ao redor de zero e agrupamentos de volatilidades. Períodos de alta volatilidade coincidem com épocas nas quais ocorreram crises em diversos países e no Brasil, que influenciaram o mercado financeiro brasileiro. Entre essas, destacamos a crise no México, em fevereiro e março de 1995, a crise na Ásia, em outubro de 1997, moratória na Rússia, em agosto de 1998, desvalorização do Real em janeiro de 1999, e queda da bolsa Nasdaq, em abril de 2000.

Na Tabela 1.1 apresentamos algumas estatísticas das séries Ibovespa e DJIA. Notamos que as curtoses são altas enquanto que os coeficientes de assimetria indicam distribuições aproximadamente simétricas. Já havíamos comentado que os dados não são normalmente distribuídos, fato apontado pelos gráficos  $Q \times Q$ .

Tabela 1.1: Estatísticas para as séries de retornos do Ibovespa e DJIA.

Estatística	Ibovespa	DJIA
Média	0,0008394	0,00040961
Mediana	0,0013533	0,00060610
Desvio padrão	0,02864553	0,01170508
Assimetria	0,7410184	-0,3029216
Curtose	12,91929	4,018403
Mínimo	-0,1722924	-0,07454077
Máximo	0,2881763	0,06155363

**Exemplo 1.6.** Na Figura 1.6 (a) temos a série de preços diários das ações da Petrobrás PN, no período de 3 de janeiro de 1995 a 27 de dezembro de 2000(arquivo d-petro95.00.dat), com  $T = 1499$  observações. Mostramos as mesmas

quantidades do exemplo 1.1 nas figuras 1.6 (b), 1.6 (c) e 1.6(d). Notam-se os mesmos fatos estilizados e o comportamento similar das duas séries, Ibovespa e Petrobrás; a correlação contemporânea entre elas é alta. Veja o capítulo 7.

**Exemplo 1.7.** Considere a série diária de taxas de câmbio USD/Real, de 30 de junho de 1994 a 1 de julho de 1998 (arquivo d-usre94.98.dat), contendo  $T = 997$  observações. A série, retornos, histograma e gráfico  $Q \times Q$  estão apresentados na Figura 1.7. Observe a grande variabilidade no início da série de retornos, comparada com a parte final.

## 1.8 Volatilidade

Um dos objetivos deste livro será o de modelar o que se chama de *volatilidade*, que é a variância condicional de uma variável, comumente um retorno. Embora não seja medida diretamente, a volatilidade manifesta-se de várias maneiras numa série financeira, como veremos a seguir.

Há três enfoques para o cálculo de volatilidades:

(i) uma maneira é equacionar um preço de mercado observado com o preço modelado de uma opção. Obtemos o que se chama de *volatilidade implícita*, que usualmente é baseada na fórmula de Black-Scholes para opções européias. Esta fórmula supõe normalidade dos preços e volatilidade constante;

(ii) outra maneira é modelar diretamente a volatilidade da série de retornos, usando alguma família, como a dos modelos ARCH; obtemos a chamada *volatilidade estatística*;

(iii) uma alternativa é modelar a volatilidade por meio de uma média de uma função dos últimos  $k$  retornos, digamos. Obtemos o que se chama de *volatilidade histórica*. Podemos considerar os quadrados dos retornos ou os valores absolutos dos retornos nesta média móvel. Uma definição geral calcula a volatilidade, para cada instante  $t$ , como uma média de  $k$  retornos passados, a saber,

$$v_t = \left[ \frac{1}{k} \sum_{j=0}^{k-1} |r_{t-j}|^p \right]^{\frac{1}{p}},$$

onde  $p > 0$ . Como dissemos acima, casos usuais são  $p = 2$  e  $p = 1$ .

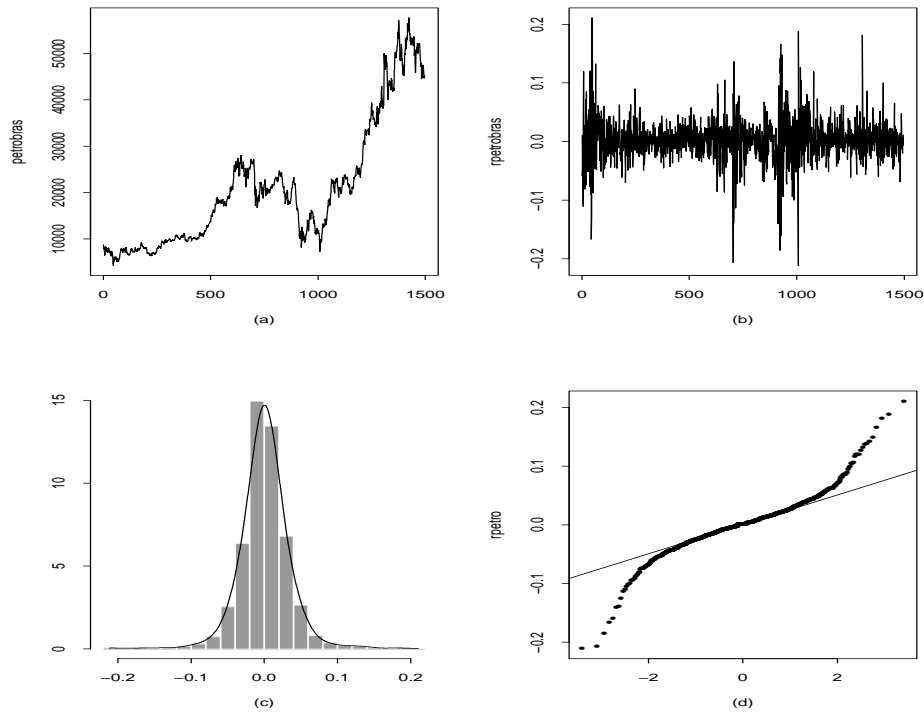


Figura 1.6: (a) Gráfico da série Petrobrás (b) Retornos diários da Petrobrás  
(c) Histograma com densidade ajustada (d) Gráfico  $Q \times Q$

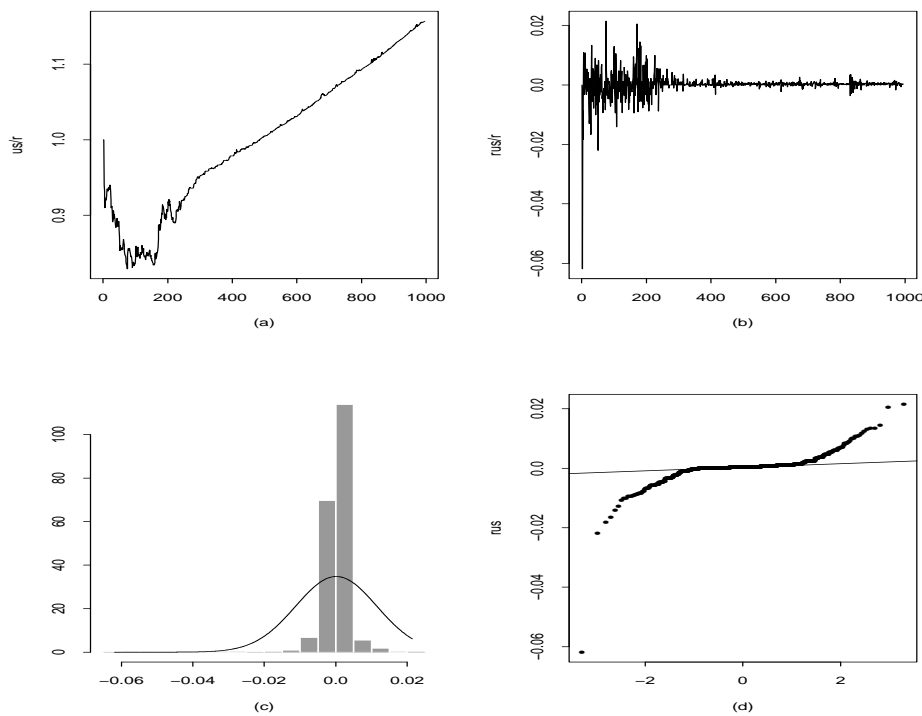


Figura 1.7: (a) Gráfico da série de taxas de câmbio us/real (b) Série dos retornos (c) Histograma com densidade ajustada (d) Gráfico  $Q \times Q$

No lugar de uma média podemos calcular a volatilidade por meio de um procedimento EWMA (“exponentially weighted moving average”), usado pelo RiskMetrics, desenvolvido pelo banco J.P. Morgan. Veja o Capítulo 6 para detalhes.

Para dados intra-diários, pode-se estimar a volatilidade diária por meio da *volatilidade realizada*, que é a soma dos quadrados dos retornos obtidos em intervalos regulares durante este dia, por exemplo, a cada 5 minutos.

Os valores obtidos pelas diversas abordagens acima descritas podem ser muito diferentes. De qualquer modo, a volatilidade é uma medida de variabilidade de preços de ativos e normalmente é difícil prever variações de preços.

Mas em toda atividade financeira (gestão de risco, precificação de derivativos e “hedging”, seleção de carteiras, etc) há a necessidade de se prever volatilidade. Por exemplo, um gestor de risco quer saber hoje a probabilidade de que uma carteira sua perca valor num futuro de curto prazo (um dia, por exemplo) ou razoavelmente longo (como 30 dias).

Vamos introduzir uma notação que será utilizada em capítulos seguintes. Seja  $r_t$  uma série de retornos. Defina

$$\mu_t = E(r_t | \mathcal{F}_{t-1}) = E_{t-1}(r_t), \quad (1.28)$$

$$h_t = E((r_t - \mu_t)^2 | \mathcal{F}_{t-1}) = E_{t-1}((r_t - \mu_t)^2), \quad (1.29)$$

a média e variância condicionais de  $r_t$ , dada a informação até o instante  $t - 1$ ,  $\mathcal{F}_{t-1}$ .

Um modelo típico para a volatilidade é da forma

$$r_t = \mu_t + \sqrt{h_t} \varepsilon_t, \quad (1.30)$$

onde  $E_{t-1}(\varepsilon_t) = 0$ ,  $\text{Var}_{t-1}(\varepsilon_t) = 1$  e tipicamente  $\varepsilon_t$  é i.i.d. com distribuição  $F$ . A média e variância incondicionais de  $r_t$  serão denotadas por  $\mu = E(r_t)$  e  $\sigma^2 = \text{Var}(r_t)$ , respectivamente, e seja  $G$  a distribuição de  $r_t$ . É claro que (1.28), (1.29) e  $F$  determinam  $\mu, \sigma^2$  e  $G$ , mas não o contrário.

## 1.9 Aspectos Computacionais

Vários programas computacionais (pacotes) podem ser utilizados para aplicações a dados reais ou simulados das técnicas desenvolvidas neste livro. Dentre estes citamos o **EViews**, o módulo **S+FinMetrics** do **SPlus**, o software livre **R**, o **MatLab** e o **STAMP**.

Utilizaremos, preferencialmente o **S+FinMetrics** e o **EViews** neste livro. Recomendamos aos leitores a consulta aos manuais destes pacotes ou a textos que os utilizam de maneira sistemática, como Zivot e Wang (2003) e Carmona (2004). Sempre que possível apresentaremos os principais comandos em algumas análises. Veja também Venables e Ripley (2001) para uma exposição geral do uso do **SPlus** em problemas estatísticos e Koopman et al. (2000) para mais detalhes sobre o **STAMP**.

Para ilustrar o uso do **S+FinMetrics**, vejamos os comandos para a construção da Figura 1.1. Designemos por *ibv* a série de índices do Ibovespa. Os gráficos (a)-(d) são obtidos por meio de:

```
> par(mfrow=c(2,2))
> tsplot(ibv)
> ribv = diff(log(ibv))
> tsplot(ribv)
> hist(ribv, probability=T, nclass=20)
> dens = density(ribv, n=200)
> points(dens, type="l")
> qqnorm(ribv)
> qqline(ribv)
```

O leitor poderá encontrar na página do livro, no sítio [www.ime.usp.br/~pam/](http://www.ime.usp.br/~pam/), roteiros de utilização do **EViews** e **S+FinMetrics** para análises de dados realizadas no livro.

## 1.10 Problemas

1. Suponha que os preços diários de fechamento de uma ação sejam:

dia	1	2	3	4	5	6	7	8	9	10
preço	47,9	46,0	45,8	48,9	49,4	50,7	50,6	51,2	50,1	51,3

- (a) Qual é o retorno simples do dia 1 para o dia 2? e do dia 1 para o dia 6?
- (b) Qual é o log-retorno do dia 4 para o dia 5? E do dia 4 para o dia 10?
- (c) Verifique que  $1 + R_5(3) = (1 + R_3)(1 + R_4)(1 + R_5)$ .
- (d) Verifique que  $r_{10}(5) = r_6 + \dots + r_{10}$ .
2. Note que, se os retornos são dados em *porcentagem*, teremos:



$$r_t = 100 \times \log(1 + R_t/100), \quad R_t = (e^{r_t/100} - 1) \times 100.$$

Se os log-retornos de um ativo nos primeiros quatro meses de um ano foram 5,2%, 3,8%, -0,5% e 2,6%:

- (a) calcule os correspondentes retornos simples;
  - (b) qual é o log-retorno no período?
  - (c) Qual é o retorno simples no período?
3. Dizemos que a variável  $Y$  tem distribuição *log-normal* se  $X = \log(Y)$  tiver distribuição normal. Verifique que, se  $X \sim \mathcal{N}(\mu, \sigma^2)$ , então  $Y = e^X$  é log-normal, com

$$E(Y) = e^{\mu + \sigma^2/2}, \quad \text{Var}(Y) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

4. Suponha que o log-retorno  $r_t \sim \mathcal{N}(0,025; (0,012)^2)$ . Pelo problema anterior,  $1 + R_t$  tem distribuição log-normal. Calcule a média e a variância de  $R_t$ .
5. Considere os log-retornos diários do Banespa de 3 de janeiro de 1995 a 27 de dezembro de 2000 (arquivo d-ban95.00.dat), com  $T = 1499$  observações:
- (a) Calcule as estatísticas : média, variância, coeficiente de assimetria e curtose, quartis, máximo e mínimo. Use algum programa, como o S+FinMetrics ou EViews.
  - (b) Obtenha um histograma dos dados e comente sobre a forma da distribuição. Compare com uma distribuição normal, com média e variância obtidas em (a).
  - (c) Qual é o log-retorno médio anual sobre o período dos dados?
  - (d) Se você investisse R\$ 10.000,00 em ações do Banespa, no começo de janeiro de 1995, qual seria o valor do investimento no final de dezembro de 2000? [Note que o montante líquido ao compor continuamente o capital inicial  $C$  por  $n$  anos, à taxa anual de juros  $r$  é dado por  $M = C \exp(r \times n)$ ].

6. Mesmo problema para os log-retornos diários da Petrobrás (arquivo d-petro95.00.dat).
7. Use a estatística (1.33) para testar se os log-retornos diários do Banespa têm uma distribuição normal.
8. Mesmo problema, para os log-retornos diários do Ibovespa (arquivo d-ibv95.00.dat).
9. Mesmo problema, para os log-retornos diários da IBM, de 1962 a 1999 (arquivo d-ibm62.99.dat).
10. Mostre que o retorno simples anualizado pode ser aproximado por uma média aritmética de retornos simples no mesmo período. [Note que  $e^x = 1 + x + x^2/2 + \dots$ ].

## Apêndice 1.A. Distribuições Estáveis

Sabemos que se  $X_1, X_2, \dots$  são v.a. independentes e identicamente distribuídas (i.i.d.), com média  $\mu$  e variância  $\sigma^2$ , então  $(X_1 + \dots + X_n - n\mu)/\sigma\sqrt{n}$  converge em distribuição para uma v.a com distribuição normal padrão. Este é um teorema limite da forma: se  $X_1, X_2, \dots$  são v.a. i.i.d., então  $(\sum_{i=1}^n X_i)/A_n - B_n$  converge em distribuição para uma v.a.  $X$ . Gostaríamos de descobrir todas as leis limites que aparecem dessa forma.

Suponha que  $X$  seja uma v.a. e que, para cada  $n$ , existam constantes  $a_n, b_n$  tais que

$$a_n X + b_n \approx X_1 + X_2 + \dots + X_n,$$

onde  $\approx$  significa “tem a mesma distribuição”, e onde  $X_1, X_2, \dots$  são i.i.d. e com a mesma distribuição que  $X$ . Então, dizemos que  $X$  é uma v.a. com *distribuição estável*. As distribuições normal e de Cauchy são exemplos. A primeira tem média e variância finita, ao passo que para a segunda esses momentos são infinitos. Na Figura 1.8 temos representadas a normal padrão e a Cauchy com densidade

$$f(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - \delta)^2},$$

com  $\gamma = 1$  e  $\delta = 0$ . Veja (1.31) abaixo para as definições de  $\gamma$  e  $\delta$ .

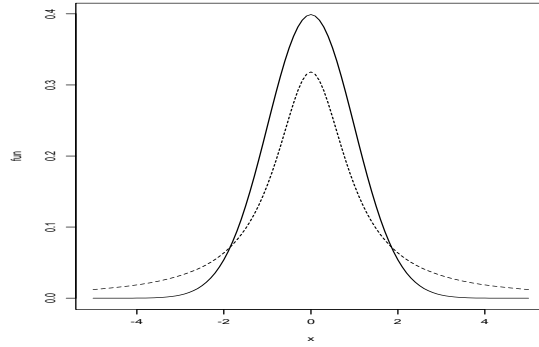


Figura 1.8: Distribuições estáveis: normal (linha cheia) e Cauchy (linha tracejada)

Um resultado fundamental diz que se o teorema limite acima vale, com  $X$  não-degenerada, então  $X$  é necessariamente uma v.a. com distribuição estável. Por outro lado, se  $X$  for estável, então  $X$  pode ser representada como um limite em distribuição de somas do tipo acima.

Outro fato importante é que se  $X$  é estável, então  $a_n = n^{1/\alpha}$ , com  $0 < \alpha \leq 2$ . O número  $\alpha$  é chamado o *índice* ou o *expoente* de  $X$ . Se  $\alpha = 2$  temos a normal. Ainda, o logaritmo da função característica de  $X$  é da forma

$$\log \varphi(t) = it\delta - \gamma|t|^\alpha [1 - i\beta \operatorname{sgn}(t) \operatorname{tg}(\pi\alpha/2)], \quad (1.31)$$

para  $0 < \alpha < 1$ .

Nesta expressão,  $\delta$  é um parâmetro de localização, real,  $\gamma > 0$  é um parâmetro de escala,  $\beta$  real é um índice de assimetria e  $\alpha$  é o expoente. Note que  $\operatorname{sgn}(t) = t/|t|$ . Se  $\alpha = 1$  teremos

$$\log \varphi(t) = i\delta t - \gamma|t|[1 - i\beta \operatorname{sgn}(t) \frac{2}{\pi} \log |t|]. \quad (1.32)$$

Se o expoente  $\alpha$  decresce de 2 até 0 as caudas de  $X$  tornam-se mais pesadas que a normal. Se  $1 < \alpha < 2$  a média de  $X$  é  $\gamma$ , finita, mas se  $0 < \alpha \leq 1$  a média é infinita. Se  $\beta = 0$ ,  $X$  é simétrica, ao passo que se  $\beta > 0$  ( $\beta < 0$ ) então  $X$  é assimétrica à direita (à esquerda).

Alguns outros fatos sobre distribuições estáveis:

- (i) Se  $X$  for estável, então  $X$  tem uma densidade limitada e contínua;
- (ii) Se  $X$  for estável, *simétrica*, então sua função característica é dada por  $\varphi(t) = e^{-c|t|^\alpha}$ .
- (iii) Se  $X$  for estável, então  $X$  é *infinitamente divisível*.

Uma v.a.  $X$  é infinitamente divisível se, para cada  $n$ , existem variáveis aleatórias  $X_{n1}, \dots, X_{nn}$ , i.i.d., tais que  $X$  tem a mesma distribuição que  $X_{n1} + \dots + X_{nn}$ .

Todavia, os retornos têm em geral, momentos de segunda ordem finitos, o que tornaria a utilização de distribuições estáveis problemática (com exceção da normal, eventualmente, que não parece ser adequada, como já vimos). Um procedimento mais adequado seria utilizar alguma distribuição com caudas mais pesadas, como a  $t$  de Student, ou então uma mistura de distribuições.

## Apêndice 1.B. Teste de Normalidade

Se uma série for considerada normal (gaussiana), seu comportamento poderá ser descrito por um modelo linear, tipo ARMA. Uma propriedade da distribuição normal é que todos os momentos ímpares maiores do que dois são nulos. Segue-se que o coeficiente de assimetria  $A$  de (1.23) deve ser igual a zero. Podemos usar, então, o resultado (1.27) para testar a hipótese  $H_0 : A = 0$ , ou seja, considerar a estatística teste  $\sqrt{T/6}\hat{A}$ , que terá distribuição limite  $\mathcal{N}(0, 1)$ .

Por outro lado, a medida de curtose  $K$ , dada por (1.24), será igual a 3 para distribuições normais e a hipótese  $H_0 : K = 3$  pode ser testada usando-se a estatística teste  $\sqrt{T/24}(\hat{K} - 3)$ , que terá também distribuição aproximada normal padrão, sob  $H_0$ .

Um teste largamente utilizado em econometria é o teste de Bera e Jarque (1981), que combina esses dois testes, usando a estatística

$$S = \left(\frac{T}{6}\right)\hat{A}^2 + \left(\frac{T}{24}\right)(\hat{K} - 3)^2, \quad (1.33)$$

que, sob  $H_0$  : a série é normal, tem distribuição qui-quadrado com dois graus

de liberdade.

Portanto, para testar a normalidade de uma série basta calcular as estimativas de  $A$  e  $K$ , calcular  $S$  por (1.33) e comparar o valor obtido com o valor tabelado de uma distribuição  $\chi^2(2)$ , com o nível de significância apropriado. Ou então, calcular o p-valor do teste, dado o valor obtido usando  $S$ . Os programas **S+FinMetrics** e **EViews**, ao calcularem várias estatísticas descritivas da série, calculam também (1.33) e fornecem o respectivo p-valor.



# Processos Estocásticos

## 2.1 Processos Estacionários

Uma das suposições básicas feitas na análise de séries temporais é que o processo estocástico gerador dos dados seja um processo estacionário. De modo bastante geral, um processo diz-se estacionário se ele oscila ao redor de uma média constante, com uma variância também constante. Formalmente, podemos distinguir duas formas de estacionariedade, forte e fraca. Vejamos, antes, a definição formal de processo estocástico.

**Definição 2.1.** *Seja  $\mathcal{T}$  um conjunto arbitrário. Um processo estocástico é uma família  $\{X(t), t \in \mathcal{T}\}$ , tal que, para cada  $t \in \mathcal{T}$ ,  $X(t)$  é uma variável aleatória.*

Nestas condições, um processo estocástico é uma família de variáveis aleatórias, que supomos definidas num mesmo espaço de probabilidades  $(\Omega, \mathcal{A}, P)$ . Normalmente supõe-se que as variáveis aleatórias (v.a.) envolvidas sejam reais, mas elas podem ser complexas.

O conjunto  $\mathcal{T}$  é normalmente tomado como o conjunto dos inteiros  $Z = \{0, \pm 1, \pm 2, \dots\}$ , ou o conjunto dos reais  $\mathbb{R}$ .

Como, para cada  $t \in \mathcal{T}$ ,  $X(t)$  é uma v.a. definida sobre  $\Omega$ , na realidade  $X(t)$  é uma função de dois argumentos,  $X(t, \omega), t \in \mathcal{T}, \omega \in \Omega$ . A Figura 2.1 ilustra esta interpretação de um processo estocástico. Vemos, na figura, que para cada  $t \in \mathcal{T}$ , temos uma v.a.  $X(t, \omega)$ , com uma função densidade de probabilidades  $f_t(x)$  (suposta existir, por simplicidade).

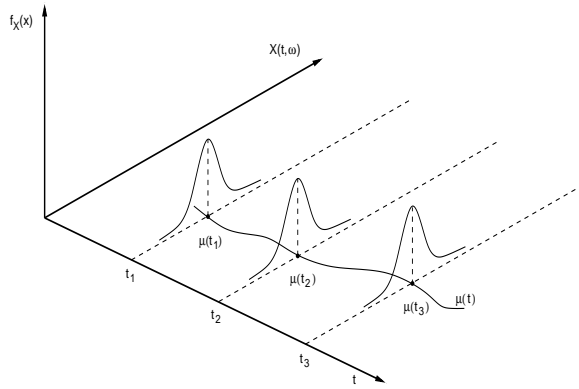


Figura 2.1: Processo estocástico como uma família de variáveis aleatórias

Por outro lado, para cada  $\omega \in \Omega$ , fixado, obteremos uma função de  $t$ , ou seja, uma realização ou trajetória do processo. Veja a Figura 2.2. Vamos designar as realizações do processo por  $X^{(1)}(t)$ ,  $X^{(2)}(t)$ , etc. O conjunto de todas as trajetórias é chamado o “ensemble”. Observamos que cada realização do processo é uma função de  $t$ , não aleatória, e para cada  $t$  fixo,  $X(t)$  é um número real ou complexo.

Uma maneira de visualizar a distribuição de probabilidades de  $X(t, \omega)$ , para  $t$  fixo, é considerar a proporção de trajetórias que passam por uma “janela” de amplitude  $\Delta$ , digamos. Tal proporção será  $f_t(x)\Delta$ . Esta é a mesma idéia para construir um histograma para a distribuição de valores de uma v.a..

O conjunto dos valores de  $\{X(t), t \in \mathcal{T}\}$  é chamado de *espaço dos estados*,  $S$ , do processo estocástico e os valores de  $X(t)$  podem ser chamados de estados. Se o conjunto  $\mathcal{T}$  for finito ou enumerável, como  $\mathcal{T} = \mathbb{Z}$ , o processo diz-se com *parâmetro discreto*. Se  $\mathcal{T}$  for um intervalo de  $\mathbb{R}$  teremos um processo com *parâmetro contínuo*. O espaço dos estados também pode ser discreto ou contínuo. No primeiro caso,  $X(t)$  pode representar uma contagem, como o número de transações de uma ação durante um dia, por exemplo. No segundo caso,  $X(t)$  representa uma medida que varia continuamente, como o retorno de um ativo ou o volume (em reais) negociado em cada dia de uma bolsa de valores.



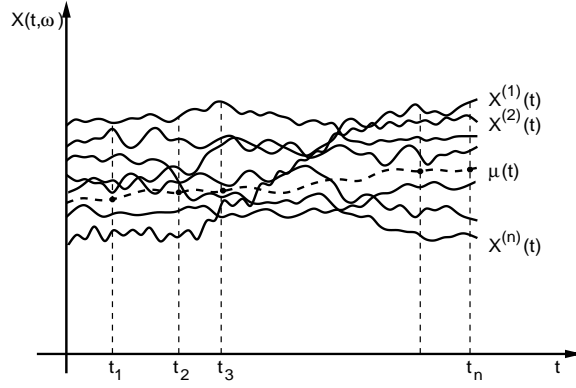


Figura 2.2: Processo estocástico como uma família de trajetórias

## 2.2 Especificação de um Processo Estocástico

Sejam  $t_1, t_2, \dots, t_n$  elementos quaisquer de  $\mathcal{T}$  e consideremos

$$F(x_1, \dots, x_n; t_1, \dots, t_n) = P\{X(t_1) \leq x_1, \dots, X(t_n) \leq x_n\}. \quad (2.1)$$

Então, o processo estocástico  $\{X(t), t \in \mathcal{T}\}$  estará especificado se conhecermos as *distribuições finito-dimensionais* (2.1), para todo  $n \geq 1$ . Isto significa que, para  $n = 1$  nós conhecemos as distribuições uni-dimensionais da v.a.  $X(t_1), t_1 \in \mathcal{T}$ , para  $n = 2$  nós conhecemos as distribuições bi-dimensionais da v.a.  $(X(t_1), X(t_2)), t_1, t_2 \in \mathcal{T}$ , e assim por diante. As funções de distribuição (2.1) devem satisfazer às duas condições seguintes:

(i) (*Condição de simetria*): para qualquer permutação  $j_1, \dots, j_n$ , dos índices  $1, 2, \dots, n$ , temos

$$F(x_{j_1}, \dots, x_{j_n}; t_{j_1}, \dots, t_{j_n}) = F(x_1, \dots, x_n; t_1, \dots, t_n). \quad (2.2)$$

(ii) (*Condição de compatibilidade*): para  $m < n$ ,

$$F(x_1, \dots, x_m, \infty, \dots, \infty; t_1, \dots, t_m, \dots, t_n) = F(x_1, \dots, x_m; t_1, \dots, t_m). \quad (2.3)$$

O lado esquerdo de (2.3) deve ser entendido como

$$\lim_{x_{m+1}, \dots, x_n \rightarrow \infty} F(x_1, \dots, x_m, x_{m+1}, \dots, x_n; t_1, \dots, t_n) .$$

Pode-se demonstrar que qualquer conjunto de funções de distribuição da forma (2.1), satisfazendo as condições (2.2) e (2.3) define um processo estocástico  $X(t)$  sobre  $\mathcal{T}$ . Este resultado é conhecido como o teorema da extensão de Kolmogorov.

Contudo, o conhecimento de todas estas distribuições finito-dimensionais é muito difícil de ocorrer na prática, senão impossível. O que se faz é estudar certas características associadas a (2.1) e que sejam simples de calcular e interpretar. Consideremos os momentos de ordem  $n$  das v.a.  $X(t_1), \dots, X(t_n)$ , para qualquer  $n \geq 1$ , ou seja,

$$\begin{aligned} \mu(r_1, \dots, r_n; t_1, \dots, t_n) &= E\{X^{r_1}(t_1) \cdots X^{r_n}(t_n)\} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_1^{r_1} \cdots x_n^{r_n} dF(x_1, \dots, x_n; t_1, \dots, t_n). \end{aligned} \quad (2.4)$$

Usualmente o que se faz é restringir o estudo a momentos de baixa ordem. Em particular, para a classe dos processos que vai nos interessar, os chamados processos estacionários, consideraremos momentos de primeira e segunda ordem.

A *função média*, ou simplesmente *média* de  $X(t)$  é dada por

$$\mu(1; t) = \mu(t) = E\{X(t)\} = \int_{-\infty}^{\infty} x dF(x; t), \quad (2.5)$$

enquanto a *função de auto-covariância* de  $X(t)$  é definida como

$$\begin{aligned} \mu(1, 1; t_1, t_2) - \mu(1; t_1)\mu(1; t_2) &= \gamma(t_1, t_2) \\ &= E\{X(t_1)X(t_2)\} - E\{X(t_1)\}E\{X(t_2)\}, \quad t_1, t_2 \in \mathcal{T}. \end{aligned} \quad (2.6)$$

Observe que  $\mu(t)$  é uma função de  $t \in \mathcal{T}$  e que  $\gamma(t_1, t_2)$  depende de dois argumentos,  $t_1$  e  $t_2$ . Em particular, se  $t_1 = t_2 = t$ , (2.6) nos fornece

$$\gamma(t, t) = \text{Var}\{X(t)\} = E\{X^2(t)\} - E^2\{X(t)\}, \quad (2.7)$$

que é a (função) *variância* do processo  $X(t)$ , e que será indicada por  $\sigma^2(t)$ .

Voltemos à Figura 2.1. Para cada  $t$  temos uma v.a.  $X(t)$ , que tem média  $\mu(t)$  e variância  $\sigma^2(t)$ . Na figura, estão indicadas as médias  $\mu(t_1)$ ,  $\mu(t_2)$  e  $\mu(t_3)$ . A função de auto-covariância  $\gamma(t_1, t_2)$  dá a covariância entre as duas variáveis aleatórias  $X(t_1)$  e  $X(t_2)$ , para quaisquer  $t_1, t_2 \in \mathcal{T}$ . A função  $\mu(t)$  é obtida variando-se  $t$  em  $\mathcal{T}$ .

Consideremos, agora, a Figura 2.2. Para cada  $t$ , temos um conjunto de valores  $X^{(1)}(t)$ ,  $X^{(2)}(t)$ , etc, correspondentes às várias realizações do processo. A função  $\mu(t)$  é obtida determinando-se, para cada  $t$ , a média dos valores  $X^{(j)}(t)$ , média esta calculada em relação a  $j$ .

Resumindo, os parâmetros mais importantes a serem considerados serão a média e a função de auto-covariância (f.a.c.v.),  $\mu(t)$  e  $\gamma(t_1, t_2)$ . Quando houver possibilidade de confusão, usaremos as notações  $\mu_X(t)$  e  $\gamma_X(t_1, t_2)$  para indicar a média e a f.a.c.v. de  $X(t)$ . Outra convenção: quando falarmos em “gráfico” de  $X(t)$ , na realidade estaremos falando em algo parecido com a Figura 2.2, onde estão “todas” as trajetórias de  $X(t)$ . Por isso, é usual representar apenas uma trajetória típica do processo.

Vimos, no Capítulo 1, que há outros parâmetros importantes em finanças, como o terceiro e quarto momentos, que são usados, por exemplo, para calcular os coeficientes de assimetria e curtose.

Observemos, também, que na prática, teremos que estimar as quantidades  $\mu(t)$ ,  $\sigma^2(t)$  e  $\gamma(t_1, t_2)$ . Observando a Figura 2.2, vemos que uma maneira de fazê-lo é considerar um número  $m$  de trajetórias  $X^{(1)}(t), \dots, X^{(m)}(t)$  e utilizá-las para estimar os parâmetros acima. Por exemplo, podemos estimar a média no instante  $t$  por

$$\hat{\mu}(t) = \frac{X^{(1)}(t) + \dots + X^{(m)}(t)}{m}.$$

O problema que surge é que usualmente temos uma só trajetória do processo, observada entre dois instantes de tempo.

**Definição 2.2.** *Um processo estocástico  $\{X(t), t \in \mathcal{T}\}$  diz-se estritamente estacionário se todas as distribuições finito dimensionais (2.1) permanecem as mesmas sob translações do tempo, ou seja,*

$$F(x_1, \dots, x_n; t_1 + \tau, \dots, t_n + \tau) = F(x_1, \dots, x_n; t_1, \dots, t_n), \quad (2.8)$$

para quaisquer  $t_1, \dots, t_n, \tau$  de  $\mathcal{T}$ .

Isto significa, em particular, que todas as distribuições uni-dimensionais são invariantes sob translações do tempo, logo a média  $\mu(t)$  e a variância  $\sigma^2(t)$  são constantes, isto é,

$$E\{X(t)\} = \mu(t) = \mu, \quad \text{para todo } t \in \mathcal{T}, \quad (2.9)$$

$$\text{Var}\{X(t)\} = \sigma^2(t) = \sigma^2, \quad \text{para todo } t \in \mathcal{T}. \quad (2.10)$$

Sem perda de generalidade, podemos supor que  $\mu = 0$ . Caso contrário, considere o processo  $X(t) - \mu$ .

Do mesmo modo, todas as distribuições bi-dimensionais dependem de diferenças de tempos. De fato, para  $t_1, t_2 \in \mathcal{T}$ ,  $\gamma(t_1, t_2) = \gamma(t_1 + t, t_2 + t)$  e fazendo  $t = -t_2$ , temos que

$$\gamma(t_1, t_2) = \gamma(t_1 - t_2, 0) = \text{Cov}\{X(t_1 - t_2), X(0)\}. \quad (2.11)$$

Na realidade, a covariância (2.11) é uma função de  $|t_1 - t_2|$  e para ver isto basta fazer  $t = -t_1$  acima.

Segue-se que podemos escrever a *função de auto-covariância* de um processo estacionário forte ou estrito como

$$\gamma(\tau) = \text{Cov}\{X(t), X(t + \tau)\} = \text{Cov}\{X(0), X(\tau)\} \quad (2.12)$$

para  $t, \tau \in \mathcal{T}$ .

Genericamente, os momentos de ordem  $n$  de  $X(t)$  dependem apenas das diferenças  $t_j - t_1$ , e são funções de  $n - 1$  argumentos.

Como dissemos anteriormente, estaremos interessados em caracterizar os processos estocásticos através de um número pequeno de funções de distribuição ou de momentos. Se nos restringirmos a momentos de primeira e segunda ordens, somos levados à seguinte

**Definição 2.3.** Um processo estocástico  $\{X(t), t \in \mathcal{T}\}$  diz-se *fracamente estacionário* (ou *estacionário de segunda ordem*) se e somente se

- (i)  $E\{X(t)\} = \mu(t) = \mu$ , constante, para todo  $t \in \mathcal{T}$ ;
- (ii)  $E\{X^2(t)\} < \infty$ , para todo  $t \in \mathcal{T}$ ;

(iii)  $\gamma(t_1, t_2) = \text{Cov}\{X(t_1), X(t_2)\}$  é uma função apenas de  $|t_1 - t_2|$ .

A partir de agora, estaremos interessados principalmente nesta classe de processo, que denominaremos simplesmente de *processos estacionários*. Note-se que, se  $X(t)$  for estritamente estacionário, ele não necessita ser fracamente estacionário, pois a condição (ii) da definição 2.2 pode não estar satisfeita. Um processo tal que (ii) esteja satisfeita diz-se um *processo de segunda ordem*.

**Definição 2.4.** Um processo estocástico real  $\{X(t), t \in \mathcal{T}\}$  diz-se *Gaussiano* se, para qualquer conjunto  $t_1, \dots, t_n$  de  $\mathcal{T}$ , as v.a.  $X(t_1), X(t_2), \dots, X(t_n)$  têm uma distribuição normal  $n$ -variada.

Como um processo Gaussiano, com variância finita, é determinado pelas médias e covariâncias, se ele for estacionário de segunda ordem, então ele será estritamente estacionário.

No que segue usaremos a seguinte notação: se o parâmetro  $t$  (tempo) for discreto, isto é,  $t \in Z = \{0, \pm 1, \pm 2, \dots\}$ , o processo será escrito  $\{X_t, t \in Z\}$ , ao passo que se  $t$  for contínuo, isto é,  $t \in \mathbb{R}$ , o processo será denotado por  $\{X(t), t \in \mathbb{R}\}$ . A mesma convenção aplica-se aos momentos. Por exemplo, a função de auto-covariância do processo estacionário discreto será denotada por  $\gamma_\tau$ , ao passo que a do processo contínuo será denotada por  $\gamma(\tau)$ .

## 2.3 Propriedades da Função de Auto-covariância

Seja  $\{X_t, t \in Z\}$  um processo estacionário real discreto, de média zero e f.a.c.v.  $\gamma_\tau = E\{X_t X_{t+\tau}\}$ .

**Proposição 2.1.** A f.a.c.v.  $\gamma_\tau$  satisfaz as seguintes propriedades:

- (i)  $\gamma_0 > 0$ ,
- (ii)  $\gamma_{-\tau} = \gamma_\tau$ ,
- (iii)  $|\gamma_\tau| \leq \gamma_0$ ,
- (iv)  $\gamma_\tau$  é não negativa definida, no sentido que

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} \geq 0, \quad (2.13)$$

para quaisquer números reais  $a_1, \dots, a_n$ , e  $\tau_1, \dots, \tau_n$  de  $Z$ .

**Prova.** As propriedades (i) e (ii) decorrem imediatamente da definição de  $\gamma_\tau$ . A propriedade (iii) segue do fato que

$$E\{X_{t+\tau} \pm X_t\}^2 = E\{X_{t+\tau}^2 \pm 2X_{t+\tau}X_t + X_t^2\} \geq 0.$$

Mas o segundo membro é igual a

$$\sigma^2 \pm 2\gamma_\tau + \sigma^2 \geq 0,$$

ou seja,

$$2\gamma_0 \pm 2\gamma_\tau \geq 0$$

e (iii) fica demonstrada. Quanto a (iv) temos que

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\tau_j - \tau_k} &= \sum_{j=1}^n \sum_{k=1}^n a_j a_k E\{X_{\tau_j} X_{\tau_k}\} \\ &= E\left\{\sum_{j=1}^n a_j X_{\tau_j}\right\}^2 \geq 0. \end{aligned}$$

**Observação.** A recíproca da propriedade (iv) também é verdadeira, isto é, dada uma função  $\gamma_\tau$  tendo a propriedade (2.13), existe um processo estocástico  $X_t$  tendo  $\gamma_\tau$  como f.a.c.v.. Na realidade,  $\gamma_\tau$  pode ser tomado como Gaussiano. Para a demonstração deste fato, ver Cramér e Leadbetter (1967, pag. 80).

Tipicamente, a f.a.c.v. de um processo estacionário tende a zero, para  $|\tau| \rightarrow \infty$ . A Figura 2.3 mostra este comportamento, além da verificação de (i)-(iii) acima. Todavia, um processo pode ser fracamente estacionário e sua f.a.c.v. pode não tender a zero. Veja o problema 2, por exemplo.

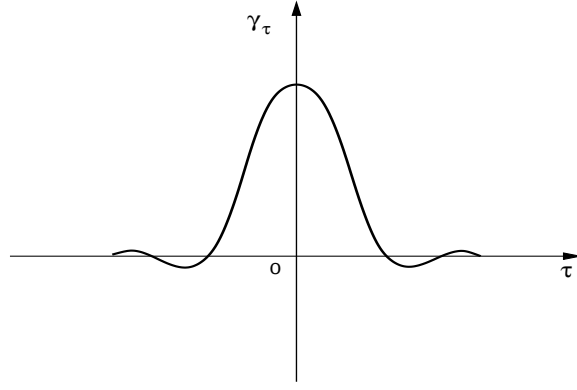


Figura 2.3: Função de auto-covariância

A *função de auto-correlação* (f.a.c.) do processo é definida por

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad \tau \in Z, \quad (2.14)$$

e tem as propriedades de  $\gamma_\tau$ , exceto que agora  $\rho_0 = 1$ .

Continuidade de um processo estocástico tem que ser definida de maneira apropriada.

**Definição 2.5.** Seja  $\{X(t), t \in \mathbb{R}\}$  um processo de segunda ordem. Dizemos que  $X(t)$  é contínuo em média quadrática no ponto  $t_0$  se e somente se

$$\lim_{t \rightarrow t_0} E\{|X(t) - X(t_0)|^2\} = 0. \quad (2.15)$$

Escreveremos  $X(t) \rightarrow X(t_0)$  mq.

Continuidade em mq de  $X(t)$  está relacionada com continuidade da f.a.c.v.  $\gamma(\tau)$ .

**Proposição 2.2.** Continuidade de  $\gamma(\tau)$  para  $\tau = 0$  implica em continuidade de  $\gamma(\tau)$  para todo  $\tau$ .

**Prova.** Usando a desigualdade de Schwarz para duas v.a. temos

$$|E\{[X(\tau + h) - X(\tau)][X(0)]\}|^2 \leq E\{|X(\tau + h) - X(\tau)|^2\}E\{|X(0)|^2\}$$

que desenvolvida resulta

$$|\gamma(\tau + h) - \gamma(\tau)|^2 \leq 2\gamma(0)[\gamma(0) - \gamma(h)]$$

e se  $\gamma(\tau)$  for contínua na origem vem que, para  $h \rightarrow 0$ , o primeiro termo tende a zero e  $\gamma(\tau)$  é contínua para todo  $\tau$ .

**Proposição 2.3.** *Se  $\gamma(\tau)$  for contínua, então  $X(t)$  é contínuo em média quadrática.*

**Prova.** Temos que

$$E\{|X(t+h) - X(t)|^2\} = 2\gamma(0) - 2\gamma(h)$$

e para  $h \rightarrow 0$ , obtemos o resultado.

**Observação.** Continuidade de um processo em mq não implica que as trajetórias do processo sejam contínuas. Um exemplo é o processo de Poisson.

Dadas observações  $X_1, \dots, X_T$ , a f.a.c.  $\rho_j$  é estimada por

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, T-1,$$

onde  $c_j$  é a estimativa da função de auto-covariância  $\gamma_j$ ,

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} [(X_t - \bar{X})(X_{t+j} - \bar{X})], \quad j = 0, 1, \dots, T-1,$$

sendo  $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$  a média amostral. Aqui, colocamos  $c_{-j} = c_j$  e  $r_{-j} = r_j$ .

Voltaremos a este assunto no Capítulo 3.

## 2.4 Processos Estocásticos Complexos

Em algumas situações é conveniente considerar processos estocásticos complexos, isto é, temos uma família  $\{X(t), t \in \mathcal{T}\}$ , onde para cada  $t \in \mathcal{T}$ ,  $X(t)$  é uma v.a. complexa. Ou seja, podemos escrever

$$X(t) = Y(t) + iZ(t),$$

onde  $Y(t)$  e  $Z(t)$  são processos estocásticos reais.



Neste caso,  $X(t)$  estará especificado se conhecermos as funções de distribuição das  $2n$  v.a. reais  $Y(t_1), \dots, Y(t_n), Z(t_1), \dots, Z(t_n)$ , para qualquer conjunto  $t_1, \dots, t_n$  de  $\mathcal{T}$ .

Definimos a média de  $X(t)$  por

$$E\{X(t)\} = E\{Y(t)\} + iE\{Z(t)\}, \quad (2.16)$$

e a variância por

$$\text{Var}\{X(t)\} = E\{|X(t) - E\{X(t)\}|^2\}. \quad (2.17)$$

Vemos, pois, que a média é um número complexo, mas a variância é um número real. A f.a.c.v. de  $X(t)$  é definida por

$$\gamma(t_1, t_2) = E\{[X(t_1) - E\{X(t_1)\}][\overline{X(t_2) - E\{X(t_2)\}}]\}, \quad (2.18)$$

para  $t_1, t_2 \in \mathcal{T}$ .

Se o processo complexo  $X(t)$  for estacionário, então (2.16) e (2.17) serão constantes (a primeira complexa e a segunda real) e a f.a.c.v. (2.18) dependerá apenas de  $|t_1 - t_2|$ , de modo que podemos escrever

$$\gamma(\tau) = E\{X(t + \tau)\overline{X(t)}\}, \quad (2.19)$$

supondo a média zero. As propriedades de  $\gamma(\tau)$ , dadas pela Proposição 2.1, no caso real, são facilmente adaptadas para o caso complexo.

## 2.5 Processos Lineares Estacionários

Apresentaremos, nestas seção, alguns exemplos de processos estocásticos estacionários que são utilizados com freqüência. Daremos ênfase aos processos auto-regressivos (AR), de médias móveis (MA) e combinação destes, os chamados processos ARMA, que serão amplamente utilizados neste livro. No Capítulo 3 veremos como construir modelos ARMA, ou seja, como a partir dos dados podemos identificar um particular modelo que os represente, como estimar este modelo e como verificar se o mesmo é adequado.

### Exemplo 2.1. Seqüência Aleatória

Consideremos  $\{X_n, n = 1, 2, \dots\}$  uma seqüência de v.a. definidas no

mesmo espaço amostral  $\Omega$ . Aqui,  $\mathcal{T} = \{1, 2, \dots\}$  e temos um processo com parâmetro discreto, ou uma seqüência aleatória. Para todo  $n \geq 1$ , podemos escrever

$$P\{X_1 = a_1, \dots, X_n = a_n\} = P\{X_1 = a_1\} \times P\{X_2 = a_2 | X_1 = a_1\} \\ \times \dots \times P\{X_n = a_n | X_1 = a_1, \dots, X_{n-1} = a_{n-1}\}. \quad (2.20)$$

Em (2.20), os  $a_j$ 's representam estados do processo e o espaço dos estados pode ser tomado como o conjunto dos reais. O caso mais simples é aquele em que temos uma seqüência  $\{X_n, n \geq 1\}$  de v.a. *mutuamente independentes* e neste caso (2.20) fica

$$P\{X_1 = a_1, \dots, X_n = a_n\} = P\{X_1 = a_1\} \dots P\{X_n = a_n\}. \quad (2.21)$$

Se as v.a.  $X_1, X_2, \dots$  tiverem todas a mesma distribuição, teremos, então, uma seqüência de v.a. independentes e identicamente distribuídas (i.i.d., brevemente). Neste caso, o processo  $X_n$  é estacionário. Se  $E\{X_n\} = \mu$ ,  $\text{Var}\{X_n\} = \sigma^2$ , para todo  $n \geq 1$ , então

$$\gamma_\tau = \text{Cov}\{X_n, X_{n+\tau}\} = \begin{cases} \sigma^2, & \text{se } \tau = 0 \\ 0, & \text{se } \tau \neq 0. \end{cases} \quad (2.22)$$

Segue-se que  $\rho_\tau = 1$ , para  $\tau = 0$  e  $\rho_\tau = 0$ , caso contrário.

**Definição 2.6.** Dizemos que  $\{\varepsilon_t, t \in Z\}$  é um ruído branco discreto se as v.a.  $\varepsilon_t$  são não correlacionadas, isto é,  $\text{Cov}\{\varepsilon_t, \varepsilon_s\} = 0, t \neq s$ .

Um tal processo será estacionário se  $E\{\varepsilon_t\} = \mu$  e  $\text{Var}\{\varepsilon_t\} = \sigma^2$ , para todo  $t$ . Segue-se que a f.a.c.v. de  $\varepsilon_t$  é dada por (2.22).

Obviamente, se as v.a.  $\varepsilon_t$  são independentes, elas também serão não correlacionadas. Uma seqüência de v.a. i.i.d., como definida acima, é chamada um *processo puramente aleatório*.

Ilustramos na Figura 2.4 a função de auto-correlação de um ruído branco. De agora em diante vamos reservar a notação  $\{\varepsilon_t, t \in Z\}$  para um ruído branco discreto e iremos supor que  $\mu = 0$ . Escreveremos, brevemente,

$$\varepsilon_t \sim \text{RB}(0, \sigma^2).$$

No caso de um processo puramente aleatório, escreveremos

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

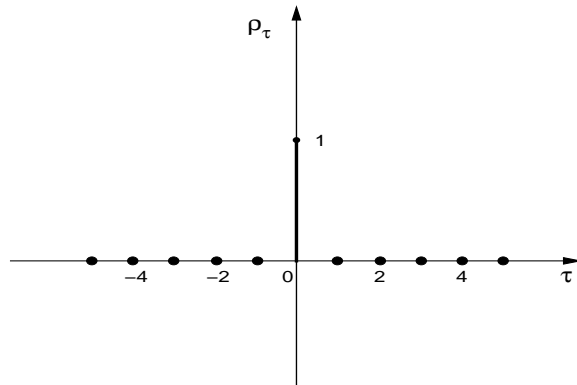


Figura 2.4: F.a.c. de um ruído branco

**Exemplo 2.2.** *Passeio Aleatório*

Considere uma seqüência aleatória  $\{\varepsilon_t, t \geq 1\}$ , de v.a. i.i.d.  $(\mu_\varepsilon, \sigma_\varepsilon^2)$ . Defina a seqüência

$$X_t = \varepsilon_1 + \dots + \varepsilon_t. \quad (2.23)$$

Segue-se que  $E(X_t) = t\mu_\varepsilon$  e  $\text{Var}(X_t) = t\sigma_\varepsilon^2$ , ou seja, ambas dependem de  $t$ . Não é difícil mostrar que

$$\gamma_X(t_1, t_2) = \sigma_\varepsilon^2 \min(t_1, t_2)$$

e portanto a auto-covariância de  $X_t$  depende de  $t_1$  e  $t_2$ . O processo (2.23) é chamado de *passeio aleatório* ou *casual* e à medida que o tempo passa  $X_t$  tende a oscilar ao redor de  $t\mu_\varepsilon$  com amplitude crescente. O processo é claramente não-estacionário.

Observemos que  $X_t = X_{t-1} + \varepsilon_t$ , logo dado o valor de  $X_{t-1}$ , o valor de  $X_t$  depende apenas de  $\varepsilon_t$ . Como  $\varepsilon_t = X_t - X_{t-1}$ , este processo tem *incrementos ortogonais* ou *não correlacionados*.

Passeios aleatórios têm grande importância em econometria e finanças. Uma hipótese célebre é que os preços de ativos financeiros seguem um passeio aleatório. Na seção 2.6 voltaremos a esse assunto.

### 2.5.1 Processos Auto-regressivos

Dizemos que  $\{X_t, t \in \mathbb{Z}\}$  é um processo auto-regressivo de ordem  $p$ , e escrevemos  $X_t \sim AR(p)$ , se satisfizer à equação de diferenças

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t, \quad (2.24)$$

onde  $\mu, \phi_1, \dots, \phi_p$  são parâmetros reais e  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$ . Segue-se que  $E(X_t) = \mu$  e se escrevermos o processo na forma

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t,$$

então

$$\mu = E(X_t) = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}.$$

Definamos o operador retroativo  $B$  através de  $B^s X_t = X_{t-s}$ ,  $s \geq 1$ . Então (2.24) pode ser escrita

$$\phi(B)\tilde{X}_t = \varepsilon_t, \quad (2.25)$$

onde  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  é o operador auto-regressivo de ordem  $p$  e  $\tilde{X}_t = X_t - \mu$ . Suponha  $\mu = 0$  no que segue.

Um caso particular importante é o processo  $AR(1)$ ,

$$X_t = \phi X_{t-1} + \varepsilon_t. \quad (2.26)$$

Aqui,  $\phi(B) = 1 - \phi B$ . Através de substituições sucessivas obtemos

$$X_t = \sum_{j=0}^r \phi^j \varepsilon_{t-j} + \phi^{r+1} X_{t-r-1}.$$

Se  $X_t$  for estacionário, com variância finita  $\sigma_X^2$ , então

$$E[X_t - \sum_{j=0}^r \phi^j \varepsilon_{t-j}]^2 = \phi^{2r+2} E[X_{t-r-1}^2] = \phi^{2r+2} \sigma_X^2.$$

Se  $|\phi| < 1$ ,  $\phi^{2(r+1)} \rightarrow 0$ , quando  $r \rightarrow \infty$ , portanto sob esta suposição, podemos escrever

$$X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, \quad (2.27)$$

onde a convergência é em média quadrática. Logo, a condição  $|\phi| < 1$  é suficiente para  $X_t$  ser estacionário. Multiplicando ambos os membros de (2.26) por  $X_{t-\tau}$  e tomando a esperança, obtemos

$$\gamma_\tau = \phi \gamma_{\tau-1} = \dots = \phi^\tau \gamma_0.$$

Mas de (2.27), obtemos

$$\gamma_0 = \sigma_X^2 = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1 - \phi^2}, \quad (2.28)$$

do que segue

$$\gamma_\tau = \frac{\sigma^2}{1 - \phi^2} \phi^\tau, \quad \tau \geq 0.$$

Como  $\gamma_\tau$  é simétrica, podemos escrever finalmente a f.a.c.v. de um processo AR(1) como

$$\gamma_\tau = \frac{\sigma^2}{1 - \phi^2} \phi^{|\tau|}, \quad \tau \in \mathbb{Z}. \quad (2.29)$$

A f.a.c. de  $X_t$  é obtida de (2.29), ou seja,

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \phi^{|\tau|}, \quad \tau \in \mathbb{Z}. \quad (2.30)$$

Na Figura 2.5 temos formas típicas de  $\rho_\tau$ . Em ambos os casos, esta decai exponencialmente para zero, e este decaimento depende do sinal de  $\phi$ .

A Figura 2.6 apresenta  $T = 100$  valores de um processo AR(1) com  $\phi = 0,8$  e  $\varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1)$ .

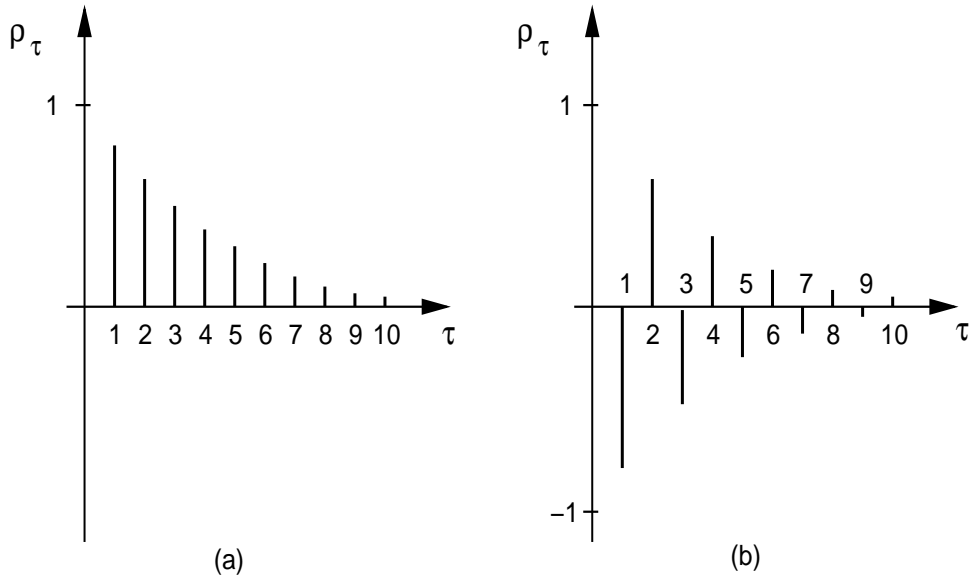


Figura 2.5: F.a.c. de um processo AR(1) (a)  $\phi = 0,8$  (b)  $\phi = -0,8$

Procuremos solução para (2.24) na forma (2.27), isto é,

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (2.31)$$

De (2.25), temos formalmente,

$$X_t = \phi(B)^{-1} \varepsilon_t = \psi(B) \varepsilon_t,$$

onde  $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$ . Em analogia com o caso AR(1), devemos ter  $\sum_j \psi_j^2 < \infty$  para que (2.31) seja uma solução estacionária. Como  $\phi(B)\psi(B) = 1$ , os coeficientes  $\psi_j$  podem ser obtidos desta identidade, em função dos  $\phi_j$ 's.

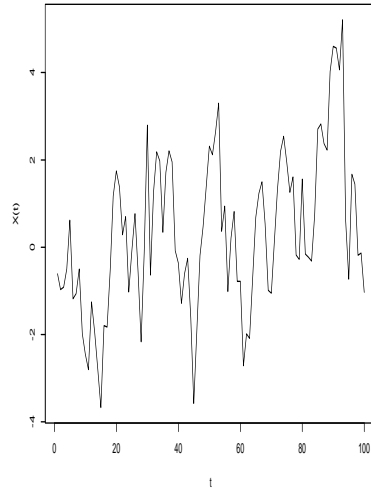


Figura 2.6: Processo AR(1) simulado,  $\phi = 0,8$

Pode-se demonstrar (ver Box, Jenkins e Reinsel, 1994) que a *condição para que  $X_t$  seja estacionário é que todas as raízes de  $\phi(B) = 0$  estejam fora do círculo unitário*. Em particular, para  $p = 1$ ,  $\phi(B) = 1 - \phi B = 0$  implica  $B = \phi^{-1}$  e a condição enunciada acarreta  $|\phi| < 1$ .

Supondo o processo estacionário, multiplicando-se ambos os membros de (2.24) por  $X_{t-\tau}$  e tomando valores esperados, obtemos, para  $\tau \geq 0$ ,

$$\sigma_X^2 = \frac{\sigma^2}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p}, \quad \text{para } \tau = 0, \quad (2.32)$$

$$\gamma_\tau = \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p}, \quad \text{para } \tau > 0. \quad (2.33)$$

A mesma equação de diferenças é satisfeita por  $\rho_\tau$ , bastando dividir todos os termos de (2.33) por  $\gamma_0$ .

A solução geral desta equação é dada por (Miller, 1969)

$$\gamma_\tau = A_1 G_1^\tau + A_2 G_2^\tau + \dots + A_p G_p^\tau, \quad (2.34)$$

onde os  $G_i$ 's satisfazem

$$\phi(B) = \prod_{i=1}^p (1 - G_i B).$$

Como as raízes de  $\phi(B) = 0$  devem estar fora do círculo unitário, devemos ter que  $|G_i| < 1$ , para todo  $i = 1, \dots, p$ .

Se fizermos  $\tau = 1, 2, \dots, p$  em (2.33), obtemos

$$\mathbf{\Gamma}_p \phi_p = \gamma_p, \quad (2.35)$$

onde  $\mathbf{\Gamma}_p = [\gamma_{ij}]$ , com  $\gamma_{ij} = \gamma_{|i-j|}$ ,  $i, j = 1, \dots, p$ ,  $\phi_p = (\phi_1, \dots, \phi_p)'$  e  $\gamma_p = (\gamma_1, \dots, \gamma_p)'$ .

A equação (2.35) pode ser utilizada para obter estimadores dos parâmetros  $\phi_j$ 's, substituindo-se as f.a.c.v.'s por suas estimativas. Estes estimadores são chamados estimadores de Yule-Walker.

Uma análise de (2.34) nos permite concluir que a f.a.c.v. de um processo auto-regressivo de ordem  $p$  é uma mistura de exponenciais (correspondentes às raízes  $G_i$  reais) e/ou senóides (correspondentes a pares de raízes complexas conjugadas) amortecidas.

Na Figura 2.7 temos as f.a.c.'s de dois processos AR(2), um com  $\phi_1 = 0,5$ ,  $\phi_2 = 0,3$  e outro com  $\phi_1 = 1,0$ ,  $\phi_2 = -0,89$ .

### 2.5.2 Processos de Médias Móveis

Dizemos que  $\{X_t, t \in Z\}$  é um processo de médias móveis de ordem  $q$ , denotado por MA( $q$ ), se satisfizer à equação de diferenças

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (2.36)$$



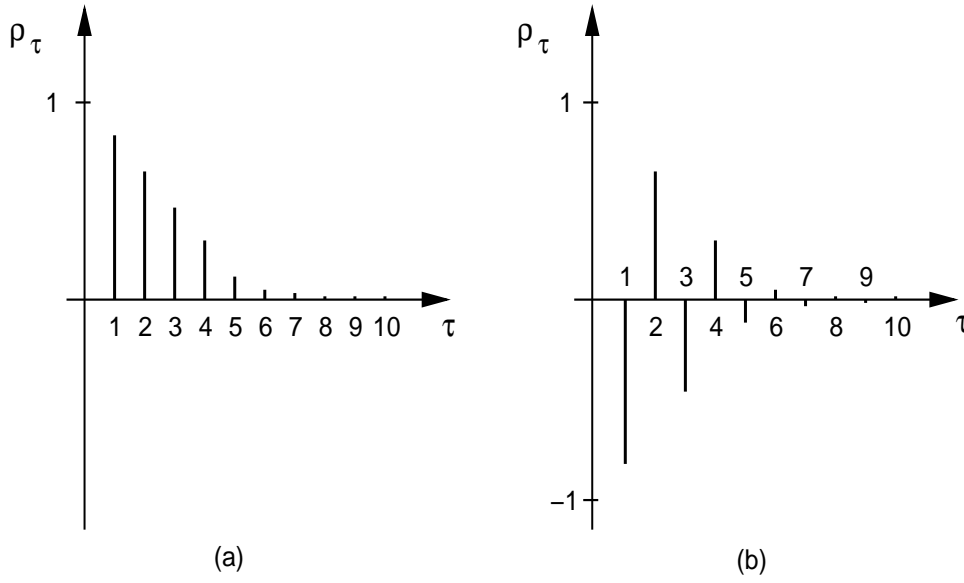


Figura 2.7: F.a.c. de dois processos AR(2) (a)  $\phi_1 = 0,5$ ,  $\phi_2 = 0,3$  (b)  $\phi_1 = 1,0$ ,  $\phi_2 = -0,89$

onde  $\mu, \theta_1, \dots, \theta_q$  são constantes reais e  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$ .

Segue-se que  $X_t$  é estacionário, de média  $\mu$  e como o  $\varepsilon_t$  são não correlacionados, podemos obter facilmente a variância do processo,

$$\sigma_X^2 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2). \quad (2.37)$$

Suponha  $\mu = 0$ . Quanto à f.a.c.v., temos

$$\begin{aligned} \gamma_\tau &= E\{X_t X_{t-\tau}\} = \gamma_\varepsilon(\tau) - \sum_{k=1}^q \theta_k \gamma_\varepsilon(k - \tau) \\ &\quad - \sum_{\ell=1}^q \theta_\ell \gamma_\varepsilon(\tau + \ell) + \sum_{k=1}^q \sum_{\ell=1}^q \theta_k \theta_\ell \gamma_\varepsilon(\tau + \ell - k), \end{aligned}$$

onde estamos denotando por  $\gamma_\varepsilon(\tau)$  a f.a.c.v. de  $\varepsilon_t$ . Resulta, então,

$$\gamma_\tau = \begin{cases} \sigma^2(-\theta_\tau + \theta_1\theta_{\tau+1} + \dots + \theta_q\theta_{q-\tau}), & \text{se } \tau = 1, \dots, q \\ 0, & \text{se } \tau > q \\ \gamma_{-\tau}, & \text{se } \tau < 0. \end{cases} \quad (2.38)$$

De (2.37) e (2.38) obtemos a f.a.c. do processo MA(q):

$$\rho_\tau = \begin{cases} \frac{-\theta_\tau + \theta_1\theta_{\tau+1} + \dots + \theta_q\theta_{q-\tau}}{1 + \theta_1^2 + \dots + \theta_q^2}, & \text{se } \tau = 1, \dots, q \\ 0, & \text{se } \tau > q \\ \rho_{-\tau}, & \text{se } \tau < 0. \end{cases} \quad (2.39)$$

Observamos, então, que a f.a.c.v. (ou a f.a.c.) de um processo MA(q) anula-se para  $|\tau| > q$ . Em particular, para um processo MA(1),

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1}, \quad (2.40)$$

obtemos

$$\begin{aligned} \text{Var}(X_t) &= \sigma_X^2 = \sigma^2(1 + \theta^2), \\ \rho_\tau &= \begin{cases} \frac{-\theta}{1 + \theta^2}, & \text{se } \tau = \pm 1 \\ 0, & \text{se } |\tau| > 1. \end{cases} \end{aligned} \quad (2.41)$$

Definindo-se o operador de médias móveis de ordem q por

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

o processo (2.36) pode ser escrito

$$X_t = \theta(B)\varepsilon_t. \quad (2.42)$$

Em particular, para o processo MA(1) temos  $\theta(B) = 1 - \theta B$ , de modo que podemos escrever

$$X_t = (1 - \theta B)\varepsilon_t$$

de onde, formalmente, segue

$$\varepsilon_t = (1 - \theta B)^{-1}X_t = (1 + \theta B + \theta^2 B^2 + \dots)X_t,$$

ou seja, temos

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} - \dots + \varepsilon_t, \quad (2.43)$$

se  $|\theta| < 1$ , para que a série do lado direito de (2.43) convirja. Nesta equação, temos  $X_t$  escrito como um processo auto-regressivo de ordem infinita. Dizemos que  $|\theta| < 1$  é uma *condição de invertibilidade* para o processo MA(1).

De modo geral, o processo (2.36) poderá ser escrito na forma

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t, \quad (2.44)$$

se a seguinte condição de invertibilidade estiver satisfeita: *todas as raízes de  $\theta(B) = 0$  devem estar fora do círculo unitário*. Ver Box, Jenkins e Reinsel (1994) para detalhes.

A relação (2.44) pode ser escrita

$$\pi(B)X_t = \varepsilon_t, \quad (2.45)$$

onde  $\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$ , de modo que  $\pi(B) = \theta(B)^{-1}$ . Portanto, os coeficientes  $\pi_j$  podem ser obtidos da identidade  $\theta(B)\pi(B) = 1$ .

A Figura 2.8 apresenta 100 observações de um processo MA(1), gerado segundo o modelo

$$X_t = \varepsilon_t - 0,8\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1). \quad (2.46)$$

Para este processo,  $\rho_1 = -0,49$ ,  $\rho_\tau = 0$ ,  $\tau \geq 2$  e  $\rho_{-\tau} = \rho_\tau$ . Temos também, na figura, o gráfico da f.a.c. de  $X_t$ . Note que, embora todas as auto-correlações sejam nulas, a partir do lag 2, na simulação aparecem valores distintos de zero, mas que estatisticamente são nulos, porque estão dentro do intervalo de confiança ao redor de zero (veja o Capítulo 3 para detalhes).

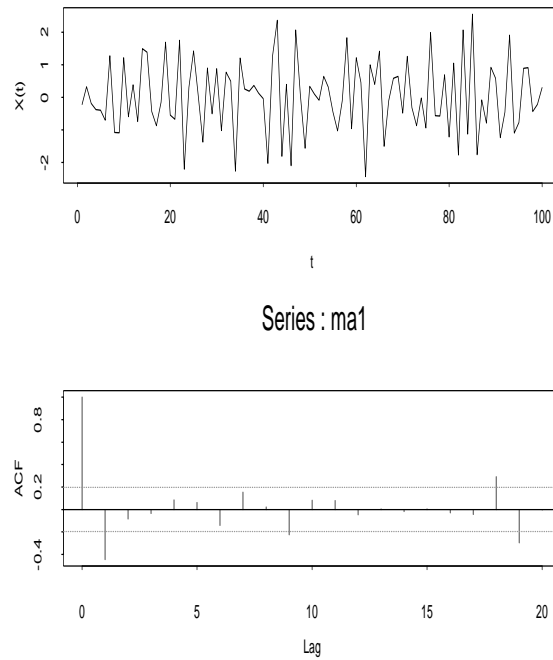


Figura 2.8: Processo MA(1) simulado,  $\theta = 0,8$  e f.a.c.

### 2.5.3 Processos Auto-regressivos e de Médias Móveis

Um processo auto-regressivo e de médias móveis, de ordem  $(p, q)$ , denotado por ARMA(p,q), é definido por

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}, \quad (2.47)$$

onde  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$ . Segue-se que a média do processo é  $\mu$ . Usando os operadores auto-regressivo e de médias móveis, definidos anteriormente, podemos escrever (2.47) na forma

$$\phi(B)\tilde{X}_t = \theta(B)\varepsilon_t, \quad (2.48)$$

onde  $\tilde{X}_t = X_t - \mu$ . Suponha que, a partir de agora,  $\mu = 0$ .

Um modelo freqüentemente usado é o ARMA(1,1), ou seja,

$$X_t = \phi X_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}. \quad (2.49)$$

É fácil ver, por substituições sucessivas, que podemos escrever

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

onde  $\psi_j = \phi^{j-1}(\phi - \theta)$ ,  $j \geq 1$ . A condição de estacionariedade é a mesma que para um processo AR(1), ou seja,  $|\phi| < 1$ . Do mesmo modo, a condição de invertibilidade  $|\theta| < 1$  vale aqui e implica que podemos escrever o processo na forma (2.44), com pesos  $\pi_j = \theta^{j-1}(\phi - \theta)$ ,  $j \geq 1$ .

Para um processo ARMA(p,q) genérico a condição de estacionariedade é a mesma que para processos AR(p), ou seja, as raízes de  $\phi(B) = 0$  devem estar fora do círculo unitário, e a condição de invertibilidade é a mesma que para processos MA(q), ou seja, as raízes de  $\theta(B) = 0$  devem estar fora do círculo unitário.

Multiplicando-se (2.47), com  $\mu = 0$ , por  $X_{t-\tau}$  e tomando-se esperanças, obtemos

$$\begin{aligned} \gamma_\tau &= \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p} + \gamma_{X\varepsilon}(\tau) \\ &\quad - \theta_1 \gamma_{X\varepsilon}(\tau-1) - \dots - \theta_q \gamma_{X\varepsilon}(\tau-q), \end{aligned} \quad (2.50)$$

onde  $\gamma_{X\varepsilon}(\tau)$  é a covariância cruzada entre  $X_t$  e  $\varepsilon_t$ , definida por

$$\gamma_{X\varepsilon}(\tau) = E(\varepsilon_t X_{t-\tau}).$$

Como  $X_{t-\tau}$  só depende de choques  $\varepsilon_t$  ocorridos até o instante  $t - \tau$ , temos que esta covariância cruzada só é diferente de zero para  $\tau \leq 0$ , logo

$$\gamma_\tau = \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p}, \quad \tau > q. \quad (2.51)$$

A conclusão é que as auto-covariâncias (e, portanto, as auto-correlações, que satisfazem equação similar) de lags  $1, 2, \dots, q$  serão afetadas pelos parâmetros de médias móveis, mas para  $\tau > q$ , as mesmas comportam-se como nos modelos auto-regressivos.

Para o caso do modelo (2.49), obtemos facilmente

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$$

e, para  $\tau > 1$ ,

$$\rho_\tau = \phi\rho_{\tau-1}.$$

A Figura 2.9 apresenta 100 observações geradas por um processo ARMA(1,1), com  $\phi = 0,8, \theta = 0,3$  e  $\varepsilon_t \sim \mathcal{N}(0,1)$ . Na Figura temos também o gráfico da f.a.c.

### Exemplo 2.3. Processo Linear Geral

Os processos AR, MA e ARMA são casos particulares do chamado processo linear geral (PLG), que pode ser expresso na forma

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (2.52)$$

onde  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$  e  $\psi_j$  são constantes satisfazendo  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ . Esta condição é necessária para que a variância do processo seja finita e neste caso,

$$\sigma_X^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2. \quad (2.53)$$

Também, de (2.52), vemos que  $E\{X_t\} = 0$  e

$$\gamma_\tau = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\tau}, \quad (2.54)$$

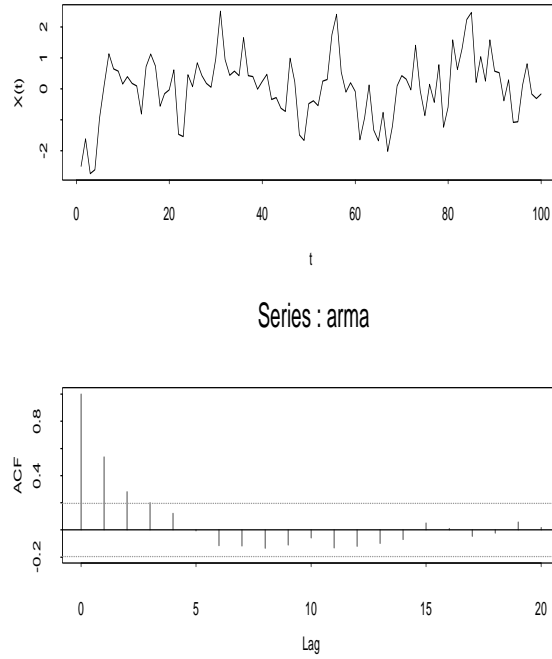


Figura 2.9: Processo ARMA(1,1) simulado,  $\phi = 0,8$ ,  $\theta = 0,3$  e f.a.c.

admitindo-se que a série do segundo membro de (2.54) convirja para um valor finito. Mas como

$$|E\{X_t X_{t-\tau}\}| \leq [E\{X_t^2\} E\{X_{t-\tau}^2\}]^{1/2} < \infty,$$

usando o fato que  $\sigma_X^2 < \infty$ , vemos que  $\gamma_\tau < \infty$  se  $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ . Logo, esta é a condição de estacionariedade para o PLG.

De (2.53) e (2.54) segue-se que a f.a.c. de um PLG é dada por

$$\rho_\tau = \frac{\sum_{j=0}^{\infty} \psi_j \psi_{j+\tau}}{\sum_{j=0}^{\infty} \psi_j^2}.$$

Uma média não nula  $\mu$  pode ser incluída no modelo (2.52), como no teorema de Wold abaixo.

É imediato verificar que um processo MA( $q$ ) é caso particular de (2.52), com  $\psi_j = 0, j > q$ . Também, o processo AR(1) é obtido de (2.52), colocando-se  $\psi_j = \phi^j$ . Não é difícil verificar que um processo AR( $p$ ) genérico é caso particular do PLG.

Pelo que vimos acima, um processo AR( $p$ ),  $p$  finito, pode ser escrito como um processo de médias móveis de ordem infinita e, reciprocamente, um processo MA( $q$ ),  $q$  finito, pode ser escrito como um processo auto-regressivo de ordem infinita, se as condições de estacionariedade e invertibilidade estiverem satisfeitas.

Na realidade, temos o seguinte resultado geral.

**Teorema 2.1.** (Wold) *Todo processo estacionário de segunda ordem, puramente não-determinístico, pode ser escrito como*

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \quad (2.55)$$

com  $\varepsilon_t \sim RB(0, \sigma^2)$ .

Um processo diz-se puramente não-determinístico se ele não puder ser previsto exatamente a partir de seu passado.

## 2.6 Processos Não-Estacionários

Muitas séries financeiras são não-estacionárias: exibem médias ou variâncias não-constantes, variando no tempo. No caso de não-estacionariedade na média, o nível médio não constante pode ser modelado de várias maneiras: polinômios no tempo, modelos ARIMA (auto-regressivos integrados e de médias móveis), etc. No caso de não-estacionariedade em variância, certas transformações (como a logarítmica) usualmente estabilizam a variância, mas outra possibilidade é considerar formas de modelar a variância.

Há, basicamente, duas formas de gerar processos não-estacionários e que



sejam não-explosivos.

(a) Incluir em (2.55) uma tendência determinística, como por exemplo

$$X_t = \beta_0 + \beta_1 t + \psi(B)\varepsilon_t, \quad (2.56)$$

obtendo-se um processo “trend-stationary”.

(b) Considerar um PLG com raiz unitária, da forma

$$(1 - B)X_t = \delta + \psi(B)\varepsilon_t, \quad (2.57)$$

com  $\psi(1) \neq 0$ . Este modelo, obviamente, descreve variações de  $X_t$  e como  $\psi(1) = \sum_{j=0}^{\infty} \psi_j \neq 0$ , o processo é não-estacionário.

**Exemplo 2.4.** Considere um caso especial de (2.56),

$$X_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad (2.58)$$

sendo  $\varepsilon_t$  um  $RB(0, \sigma^2)$ . Então temos que

- (i)  $E(X_t) = \mu_t = \beta_0 + \beta_1 t$ ;
- (ii) Tomando-se uma diferença,

$$X_t - X_{t-1} = \beta_1 + \varepsilon_t - \varepsilon_{t-1},$$

que é um modelo ARMA (1,1), com  $\phi = \theta = 1$ , portanto temos um modelo não-estacionário e não-invertível.

- (iii) Se  $W(t) = X_t - X_{t-1} = (1 - B)X_t = \Delta X_t$ ,

$$W(t) = \Delta X_t = \beta_1 + \Delta \varepsilon_t,$$

que é um modelo MA(1), estacionário, mas não-invertível.

- (iv) Extraíndo-se a tendência de (2.58) obtemos

$$Y_t = X_t - \beta_1 t = \beta_0 + \varepsilon_t, \quad (2.59)$$

que é estacionário.

De modo geral, se tivermos (2.56) com erro ARMA(p,q) e tendência  $\mu_t$  um polinômio determinístico de grau  $d$ , então

$$X_t = \sum_{j=0}^d \beta_j t^j + [\phi(B)]^{-1} \theta(B) a_t,$$

onde  $a_t$  é ruído branco. Segue-se que

$$\Delta^d X_t = (1 - B)^d X_t = \theta_0 + (1 - B)^d \psi(B) a_t, \quad (2.60)$$

na qual  $\psi(B) = [\phi(B)]^{-1} \theta(B)$  e  $\theta_0 = d! \beta_d$ . Este modelo é, novamente, estacionário, mas não invertível.

**Exemplo 2.5.** Considere, agora, o modelo

$$X_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad (2.61)$$

em que  $\varepsilon_t \sim \text{ARMA}(p, q)$ . Podemos, então, escrever

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

onde  $a_t$  é ruído branco, Segue-se que

$$\varepsilon_t = \frac{1 - \theta_1 B - \dots - \theta_q B^q}{(1 - G_1 B)(1 - G_2 B) \dots (1 - G_p B)} a_t = \psi(B) a_t. \quad (2.62)$$

Para que o processo seja estacionário as raízes de  $\phi(B)$  devem estar fora do círculo unitário, de modo que  $|G_i| < 1$ , para todo  $i = 1, \dots, p$ . Neste caso, o processo  $X_t$  é um caso especial de (a).

Suponha que  $G_1 = 1$  e  $|G_i| < 1, i = 2, \dots, p$ . Então, (2.62) fica

$$(1 - B) \varepsilon_t = \frac{\theta(B)}{(1 - G_2 B) \dots (1 - G_p B)} a_t = \psi^*(B) a_t, \quad (2.63)$$

ou seja,

$$(1 - B) X_t = \beta_1 + \psi^*(B) a_t, \quad (2.64)$$

que é da forma (b).

O nome *raiz unitária* vem da representação (2.63) acima. Uma das raízes do polinômio auto-regressivo  $\phi(B)$  é igual a um, todas as demais estão fora do círculo unitário. Podemos ter duas ou mais raízes iguais a um. Testes para raízes unitárias serão estudados no Capítulo 4.

Quando tivermos um processo da forma (2.57), dizemos que ele é *integrado de ordem 1*. Em geral, temos a definição seguinte.

**Definição 2.7.** Se  $\Delta^d X_t$  for estacionário, dizemos que  $X_t$  é *integrado de ordem d* e escrevemos  $X_t \sim I(d)$ .

Em particular, temos a

**Definição 2.8.** Se  $\Delta^d X_t \sim \text{ARMA}(p, q)$  dizemos que  $X_t$  segue um modelo  $\text{ARIMA}(p, d, q)$  : auto-regressivo integrado e de médias móveis de ordem  $(p, d, q)$ , ou seja,

$$\phi(B)\Delta^d X_t = \theta_0 + \theta(B)a_t. \quad (2.65)$$

Ou, de modo equivalente,

$$\phi(B)W(t) = \theta_0 + \theta(B)a_t, \quad \text{com } W(t) = \Delta^d X_t.$$

Observemos que

$$W(t) = \Delta^d X_t \iff X_t = S^d W(t),$$

onde  $S$  é o operador **soma** ou **integral**

$$S = (1 - B)^{-1} = \Delta^{-1}.$$

Ou seja,  $X_t$  pode ser obtido somando-se ou integrando-se o processo estacionário  $W(t)$   $d$  vezes, donde o nome *processo integrado* para  $X_t$ .

Processos explosivos podem ser obtidos considerando-se modelos ARMA cuja parte AR não satisfaz condições de estacionariedade. Por exemplo,

$$X_t = \phi X_{t-1} + a_t, \quad \phi > 1. \quad (2.66)$$

Se  $X_0 = x_0$ , pode-se mostrar que a equação de diferença (2.66) tem solução

$$X_t = x_0\phi^t + \sum_{i=0}^{t-1} \phi^i a_{t-i}. \quad (2.67)$$

Observe que

$$\text{Var}(X_t) = \sigma^2 \frac{\phi^{2(t+1)} - 1}{\phi^2 - 1},$$

que é crescente com  $t$ .

Se  $|\phi| < 1$  então sabemos que  $X_t$  é estacionário. No caso que  $\phi = 1$  teremos um passeio aleatório, estudado no exemplo 2.1,

$$X_t = X_{t-1} + a_t.$$

Incluindo-se uma constante, teremos um passeio aleatório com “drift”,

$$X_t = \theta_0 + X_{t-1} + a_t. \quad (2.68)$$

Se o processo começar em  $t = 0$ , com  $X_0 = x_0$ , podemos escrever

$$X_t = x_0 + t\theta_0 + \sum_{i=0}^{t-1} a_{t-i}, \quad (2.69)$$

de modo que obtemos (Ver Mills, 1999):

- (i)  $\mu_t = x_0 + t\theta_0$ ,
- (ii)  $\gamma_0(t) = \text{Var}(X_t) = t\sigma^2$ ,
- (iii)  $\gamma_k(t) = (t - k)\sigma^2$ ,
- (iv)  $\rho_k(t) = \frac{t-k}{t}$ .

Logo, se  $t$  grande,  $\rho_k(t) \approx 1$  e teremos uma seqüência suave mas não-estacionária.

Vimos acima que se removermos a tendência de um processo com tendência determinística, obteremos um processo estacionário, mas não invertível.

Vejamos, agora, o que acontece se tentarmos remover uma tendência de um processo que supostamente foi gerado por uma raiz unitária.

**Exemplo 2.6.** Considere (2.57), com  $\psi(B) = 1$ , ou seja

$$X_t = X_{t-1} + \delta + \varepsilon_t. \quad (2.70)$$

Considerando-se

$$Y_t = X_t - \delta t$$

e efetuando substituições sucessivas, obteremos

$$Y_t = X_0 + \sum_{j=1}^t \varepsilon_j, \quad (2.71)$$

de modo que efetivamente removemos a tendência, mas  $\text{Var}(Y_t) = t\sigma^2$ , que dependerá de  $t$ .

**Exemplo 2.7.** Uma suposição usual é que os preços de ativos financeiros sigam um passeio casual, ou seja,

$$P_t = \mu + P_{t-1} + \sigma\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1). \quad (2.72)$$

Note que a distribuição condicional de  $P_t$ , dado  $P_{t-1}$  é normal, com média  $\mu$  e variância  $\sigma^2$ . Este modelo é pouco realista, pois preços terão probabilidade não-nula de serem negativos, logo costuma-se modificá-lo e considerar que  $p_t = \log(P_t)$  é que segue o modelo (2.72), ou seja,

$$\log\left(\frac{P_t}{P_{t-1}}\right) = \mu + \sigma\varepsilon_t, \quad (2.73)$$

ou ainda, com a nomenclatura e notação do Capítulo 1,

$$r_t = \mu + \sigma\varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1). \quad (2.74)$$

Este modelo supõe que a variância seja constante. Uma suposição mais adequada é admitir que a variância (volatilidade) dos preços varie com o tempo. Além disso, parece ser razoável admitir que os log-retornos tenham

média zero, de modo que um modelo adotado por várias organizações financeiras é da forma

$$r_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1). \quad (2.75)$$

Na Figura 2.10 temos 500 valores simulados do modelo

$$p_t = 0,005 + p_{t-1} + \varepsilon_t,$$

sendo  $\varepsilon_t \sim \mathcal{N}(0, 1)$ ,  $p_0 = 0$ .

Um dos problemas importantes para avaliar, por exemplo, o VaR (valor em risco) de uma carteira de investimentos é estimar a volatilidade  $\sigma_t^2$ , para cada instante de tempo  $t$ . Veja o Capítulo 6.

## 2.7 Movimento Browniano

No Capítulo 4, quando tratarmos do problema de raízes unitárias em modelos ARMA, necessitaremos usar um processo não-estacionário particular, o movimento browniano.

**Definição 2.9.** Chamaremos de Movimento Browniano Padrão (ou processo de Wiener) ao processo contínuo  $W = \{W(t), t \geq 0\}$  tal que:

- (a)  $W(0) = 0$ ;
- (b) para quaisquer instantes  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k \leq 1$ , as v.a.  $W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_k) - W(t_{k-1})$  são independentes;
- (c) para quaisquer  $s, t$  e  $\tau$  não negativos, as v.a.  $W(t) - W(s)$  e  $W(t + \tau) - W(s + \tau)$  têm a mesma distribuição;
- (d) para todo  $t > 0$ ,  $W(t) \sim \mathcal{N}(0, t)$ ;
- (e) as trajetórias de  $W(t)$  são contínuas com probabilidade um.

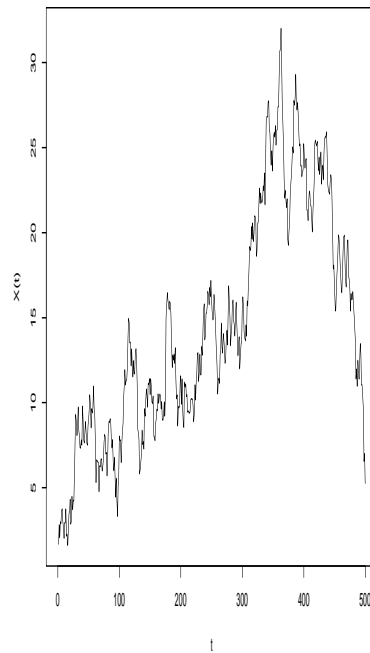


Figura 2.10: Passeio aleatório simulado

Quando as v.a.  $X$  e  $Y$  têm a mesma distribuição, escreveremos  $X \stackrel{d}{=} Y$ .

Segue-se de (b) que o movimento browniano padrão (MBP) tem *incrementos independentes* e de (c) que tem *incrementos estacionários*. Como as distribuições finito-dimensionais de  $W$  são normais multivariadas, o processo  $W$  é gaussiano. Como  $W(t) - W(s) \stackrel{d}{=} W(t-s) - W(0) = W(t-s)$ , temos que  $W(t) - W(s)$  e  $W(t-s)$  têm ambas distribuição  $\mathcal{N}(0, t-s)$ ,  $s < t$ .

O nome do processo deve-se ao biólogo Robert Brown, que estudou o movimento de partículas imersas em um líquido. Einstein (1905) fez estudo similar. Um trabalho importante é o de Wiener (1923).

Da definição do MBP temos  $E(W(t)) = 0$  e  $\text{Var}(W(t)) = t$ ,  $t \geq 0$ . Para

calcular a f.a.c.v. de  $W$ , considere  $s < t$ . Temos, então,

$$\begin{aligned}\gamma(t, s) &= E(W(t)W(s)) = E\{[(W(t) - W(s)) + W(s)]W(s)\} \\ &= E\{[(W(t) - W(s))W(s)]\} + E(W(s)^2).\end{aligned}$$

Como  $W(t) - W(s)$  e  $W(s) - W(0) = W(s)$  são independentes, vem que

$$\gamma(t, s) = E(W(t) - W(s))E(W(s)) + s = 0 + s = s, \quad 0 \leq s < t.$$

De modo análogo, se  $t < s$ , então  $\gamma(t, s) = t$ , logo

$$\gamma(t, s) = \min(t, s). \quad (2.76)$$

Como todo processo gaussiano é caracterizado por sua média e covariâncias, uma definição equivalente de MBP é a seguinte: um MBP  $W$  é um processo gaussiano com  $E(W(t)) = 0$  e  $\gamma(t, s) = \min(t, s)$ .

Embora as trajetórias de  $W$  sejam contínuas, o fato de ter incrementos independentes faz com que estas sejam bastante irregulares. A Figura 2.10 é um exemplo de tal trajetória. Na realidade, uma trajetória de um MBP é um ente matemático abstrato, que não possui uma representação gráfica real, devido aos seguintes fatos, que não serão provados aqui.

**Fato 1.** As trajetórias de um MBP não são deriváveis em qualquer ponto.

**Fato 2.** As trajetórias de um MBP não têm variação limitada em qualquer intervalo finito.

Um outro fato importante sobre MBP é que este processo é *auto-similar*, conforme a definição a seguir.

**Definição 2.10.** Um processo estocástico  $X = \{X(t), t \geq 0\}$  é *f*-auto-similar, par algum  $f > 0$ , se suas distribuições finito-dimensionais satisfazem

$$(\tau^f X(t_1), \dots, \tau^f X(t_n)) \stackrel{d}{=} (X(\tau t_1), \dots, X(\tau t_n)), \quad (2.77)$$

para todo  $\tau > 0$  e para quaisquer  $t_1, \dots, t_n \geq 0, n \geq 1$ .



Esta propriedade significa que partes escalonadas de uma trajetória em qualquer intervalo têm formas semelhantes (mas *não* idênticas). A Figura 2.11 mostra um exemplo.

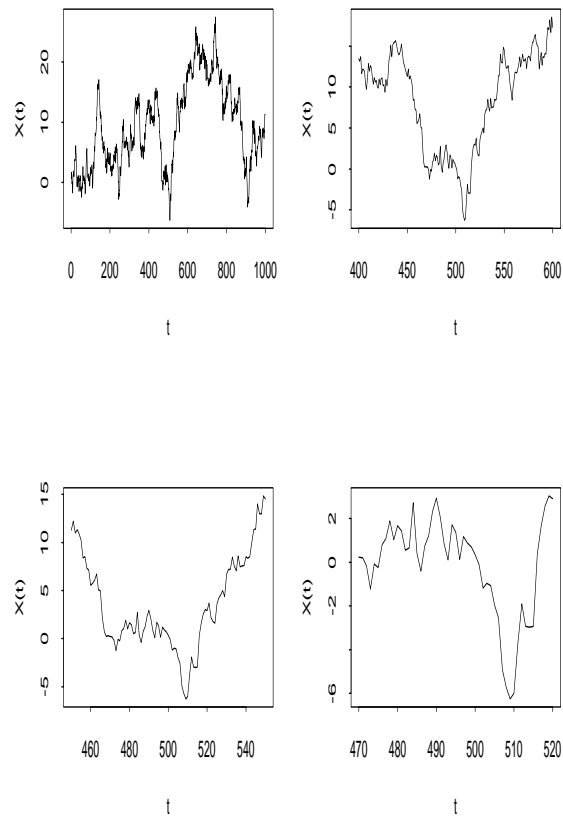


Figura 2.11: Um processo auto-similar

**Fato 3.** O MBP é 0,5-auto-similar, isto é,

$$(\tau^{1/2}W(t_1), \dots, \tau^{1/2}W(t_n)) \stackrel{d}{=} (W(\tau t_1), \dots, W(\tau t_n)), \quad (2.78)$$

para todo  $\tau > 0$  e para quaisquer  $t_1, \dots, t_n \geq 0, n \geq 1$ .

Portanto, basta considerar o processo no intervalo  $[0, 1]$  se estivermos interessados no processo no intervalo  $[0, \tau]$ .

Vejamos alguns exemplos de processos derivados do MBP.

**Exemplo 2.8.** Se considerarmos o processo  $X(t) = \sigma W(t)$ , então  $X(t)$  terá incrementos independentes e  $X(t) \sim \mathcal{N}(0, \sigma^2 t)$ . Podemos, também, considerar o movimento browniano “com drift”,

$$X(t) = \mu t + \sigma W(t), \quad t \geq 0, \sigma > 0, \mu \in \mathbb{R}.$$

Então,  $X(t)$  será gaussiano, com  $E(X(t)) = \mu t$  e  $\gamma_X(t, s) = \sigma^2 \min(t, s)$ , para  $s, t \geq 0$ .

**Exemplo 2.9.** *Movimento Browniano Geométrico.* Como vimos no exemplo 2.7, preços são não-negativos e o MB pode assumir valores negativos. Black e Scholes (1973) e Merton (1973) sugeriram outro processo para descrever preços num mercado especulativo, a saber

$$X(t) = e^{\mu t + \sigma W(t)}, \quad t \geq 0.$$

Não é difícil verificar que este processo não é gaussiano. A média de  $X(t)$  é dada por

$$E(X(t)) = e^{(\mu + 0,5\sigma^2)t},$$

e a covariância é

$$\gamma_X(t, s) = e^{(\mu + 0,5\sigma^2)(t+s)}(e^{\sigma^2 s} - 1).$$

Veja os problemas 12 e 13.

Um resultado importante e que também será usado posteriormente é o teorema limite central (TLC) funcional, que passamos a apresentar.

Se  $Y_1, Y_2, \dots$  é uma seqüência de v.a. i.i.d., com média  $\mu$  e variância  $\sigma^2$ , e considerarmos a média  $\bar{Y}_N = 1/N \sum_{t=1}^N Y_t$ , então o TLC usual nos diz que

$$\sqrt{N}(\bar{Y}_N - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2). \quad (2.79)$$

Passemos, agora, a tomar médias de uma proporção  $r$  dos dados,  $0 \leq r \leq 1$ . Por exemplo, com  $N$  observações, calculemos a média da primeira metade dos dados,

$$\bar{Y}_{[N/2]} = \frac{1}{[N/2]} \sum_{t=1}^{[N/2]} Y_t. \quad (2.80)$$

Então, mais uma vez, usando o TLC,

$$\sqrt{[N/2]}(\bar{Y}_{[N/2]} - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2). \quad (2.81)$$

De modo geral, seja

$$Y_N(r) = \frac{1}{N} \sum_{t=1}^{[Nr]} Y_t, \quad (2.82)$$

para  $0 \leq r \leq 1$ , que é proporcional à média das primeiras  $100r\%$  observações. É fácil verificar que

$$Y_N(r) = \begin{cases} 0, & 0 \leq r < 1/N, \\ Y_1/N, & 1/N \leq r < 2/N, \\ (Y_1 + Y_2)/N, & 2/N \leq r < 3/N, \\ \dots & \\ (Y_1 + \dots + Y_N)/N, & r = 1. \end{cases} \quad (2.83)$$

Podemos escrever

$$\sqrt{N}Y_N(r) = \frac{1}{\sqrt{N}} \sum_{t=1}^{[Nr]} Y_t = \frac{\sqrt{[Nr]}}{\sqrt{N}} \frac{1}{\sqrt{[Nr]}} \sum_{t=1}^{[Nr]} Y_t,$$

na qual

$$\sqrt{[Nr]} \frac{1}{[Nr]} \sum_{t=1}^{[Nr]} Y_t \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

pelo TLC e  $\sqrt{[Nr]}/\sqrt{N} \rightarrow \sqrt{r}$ , logo obtemos

$$\sqrt{N} Y_N(r) \xrightarrow{\mathcal{D}} \sqrt{r} \mathcal{N}(0, \sigma^2) = \mathcal{N}(0, r\sigma^2), \quad (2.84)$$

da qual segue, finalmente,

$$\sqrt{N} \frac{Y_N(r)}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0, r). \quad (2.85)$$

Observamos, também, que considerando-se médias baseadas em observações de  $[Nr_1]$  a  $[Nr_2]$ , com  $r_1 < r_2$ , teríamos

$$\sqrt{N} \left[ \frac{Y_N(r_2) - Y_N(r_1)}{\sigma} \right] \xrightarrow{\mathcal{D}} \mathcal{N}(0, r_2 - r_1),$$

independentemente de (2.85), se  $r < r_1$ , do que concluímos que a seqüência de funções aleatórias  $\{\frac{\sqrt{N}Y_N(\cdot)}{\sigma}, N = 1, 2, \dots\}$  tem uma distribuição limite que é o MBP:

$$\frac{\sqrt{N}Y_N(\cdot)}{\sigma} \xrightarrow{\mathcal{D}} W(\cdot). \quad (2.86)$$

Ou ainda, para cada  $0 \leq r \leq 1$ , a v.a.  $\{\frac{\sqrt{N}Y_N(r)}{\sigma}\}$  tem como distribuição limite uma v.a.  $\mathcal{N}(0, r)$ , como dado em (2.82).

Em (2.86) temos o TLC funcional. Se  $r = 1$ ,  $Y_N(1) = \frac{1}{N} \sum_{t=1}^N Y_t$ , e temos como resultado o TLC usual, a distribuição limite sendo a  $\mathcal{N}(0, 1)$ .

Um resultado importante em convergência de variáveis aleatórias diz que, se  $X_N \xrightarrow{\mathcal{D}} X$  e se  $g: \mathbb{R} \rightarrow \mathbb{R}$  for contínua, então  $g(X_N) \xrightarrow{\mathcal{D}} g(X)$ .

No caso de processos estocásticos, este resultado pode ser generalizado, considerando agora  $g(\cdot)$  um funcional contínuo. Para tanto, precisamos modificar a definição de convergência em distribuição para processos estocásticos.

**Definição 2.11.** Dizemos que  $S_N(\cdot) \xrightarrow{\mathcal{D}} S(\cdot)$  se:

(i) para quaisquer  $0 \leq r_1 \leq \dots \leq r_k \leq 1$ ,

$$y_N = \begin{bmatrix} S_N(r_1) \\ \cdots \\ S_N(r_k) \end{bmatrix} \xrightarrow{\mathcal{D}} y = \begin{bmatrix} S(r_1) \\ \cdots \\ S(r_k) \end{bmatrix};$$

(ii) para todo  $\varepsilon > 0$ ,  $P\{|S_N(r_1) - S(r_2)| > \varepsilon\} \rightarrow 0$ , uniformemente em  $N$ , para todo  $r_1, r_2$  tais que  $|r_1 - r_2| < \delta, \delta \rightarrow 0$ ;

(iii)  $P\{|S_N(0)| > \lambda\} \rightarrow 0$ , uniformemente em  $N$ , quando  $\lambda \rightarrow \infty$ .

Nestas condições, se  $S_N(\cdot) \xrightarrow{\mathcal{D}} S(\cdot)$  e  $g$  for um funcional contínuo, então  $g(S_N(\cdot)) \xrightarrow{\mathcal{D}} g(S(\cdot))$ .

Por exemplo, vimos que  $\sqrt{N}Y_N(\cdot) \xrightarrow{\mathcal{D}} \sigma W(\cdot)$ . Se considerarmos  $S_N(r) = [\sqrt{N}Y_N(r)]^2$ , então  $S_N(\cdot) \xrightarrow{\mathcal{D}} \sigma^2[W(\cdot)]^2$ .

## 2.8 Martingales

Um martingale descreve o que podemos chamar de “jogo justo”. Suponha que  $X_n$  represente a fortuna de um jogador após o  $n$ -ésimo jogo e  $\mathcal{F}_n$  represente a “história passada do jogo”, ou seja,  $\mathcal{F}_n = \{X_n, X_{n-1}, \dots, X_1\}$ . Se o ganho esperado no  $(n+1)$ -ésimo jogo, dada a informação passada, for igual à fortuna atual,  $X_n$ , dizemos que o processo estocástico  $\{X_n, n \geq 1\}$  é um martingale. Vamos formalizar este conceito.

Seja  $X = \{X_t, t \in \mathcal{T}\}$ , com  $\mathcal{T} \subset \mathbb{R}$ , um processo estocástico, ou seja, para cada  $t \in \mathcal{T}$ ,  $X_t$  é uma v.a. definida sobre um espaço de probabilidade  $(\Omega, \mathcal{F}, \mathcal{P})$ . Suponha, ainda, que  $\{\mathcal{F}_t, t \in \mathcal{T}\}$  seja uma família crescente de sub- $\sigma$ -álgebras de  $\mathcal{F}$ , isto é,  $\mathcal{F}_s \subset \mathcal{F}_t$ , se  $s \leq t$ , e que  $X$  seja *adaptado* a esta família, ou seja,  $X_t$  é  $\mathcal{F}_t$ -mensurável.

**Definição 2.12.** Um processo  $X = \{X_t, \mathcal{F}_t, t \in \mathcal{T}\}$  é um martingale se:

- (a)  $X_t$  é integrável, para cada  $t \in \mathcal{T}$ .
- (b) Se  $s \leq t$ , então

$$E(X_t | \mathcal{F}_s) = X_s. \quad (2.87)$$

Se em (b) tivermos  $E(X_t|\mathcal{F}_s) \geq X_s$ , o processo diz-se um *sub-martingale* e se tivermos  $E(X_t|\mathcal{F}_s) \leq X_s$ , temos um *super-martingale*.

A relação (2.87) nos diz que  $X_s$  é o melhor preditor de  $X_t$ , dada a informação  $\mathcal{F}_s$ .

Se  $\mathcal{T} = \{0, \pm 1, \pm 2, \dots\}$  ou  $\mathcal{T} = \{1, 2, \dots\}$ ,  $X_t$  é um martingale com parâmetro discreto e (b) acima fica

$$E(X_{n+1}|\mathcal{F}_n) = X_n, \quad n \geq 1. \quad (2.88)$$

**Exemplo 2.10.** (a) Se  $X_1, X_2, \dots$  são v.a. independentes, com  $E(X_i) = 0$ , para todo  $i$ , então  $Y_n = \sum_{i=1}^n X_i$  é um martingale.

(b) Com as variáveis de (a), mas  $E(X_i) = \mu_i \geq 0$ , segue-se que  $Y_n$  é um sub-martingale e  $Z_n = \sum_{i=1}^n (X_i - \mu_i)$  é um martingale.

(c) Se  $X_{n+1} = X_n + a_n$ , onde  $a_n \sim \text{RB}(0, \sigma^2)$ , então  $\{X_n, n \geq 1\}$  é um martingale. Ou seja, um passeio aleatório é um caso especial de um martingale.

Observe que, de (2.88) temos

$$E(X_{n+1} - X_n|\mathcal{F}_n) = 0, \quad n \geq 1. \quad (2.89)$$

Ou seja, podemos escrever

$$X_{n+1} = X_n + u_n, \quad (2.90)$$

onde agora  $u_n$  não precisa ser necessariamente um ruído branco, ter uma variância constante ou mesmo ser estacionário.

Quando escrevermos simplesmente que  $X = \{X_t, t \in \mathcal{T}\}$  é um martingale, então as  $\sigma$ -álgebras da definição são  $\mathcal{F}_t = \mathcal{F}\{X_s, s \leq t\}$ , ou seja, as  $\sigma$ -álgebras geradas pelos valores do processo até o instante  $t$ . Este será o caso usual a considerar, quando a história passada, ou o conjunto de informação passado, é dado por funções de valores passados do próprio processo. Em algumas situações, a informação passada contém valores de outros processos de

interesse, relacionados a  $X$ .

### Algumas propriedades:

1. Se  $X_t$  e  $Y_t$  forem martingales, então  $X_t + Y_t$  será um martingale se ambos os processos forem martingales com respeito à mesma seqüência de  $\sigma$ -álgebras.
2. Para verificar que  $\{X_n, \mathcal{F}_n, n \geq 1\}$  é um martingale, basta mostrar que (2.88) vale, se  $\{X_n, \mathcal{F}_n, n \geq 1\}$  for tal que  $\mathcal{F}_n$  é uma família crescente de sub- $\sigma$ -álgebras de  $\mathcal{F}$ ,  $X_n$  é adaptado a  $\mathcal{F}_n$  e  $X_n$  é integrável, para todo  $n \geq 1$ .
3. Se  $\{X_n, n \geq 1\}$  for um martingale, então  $E(X_n) = E(X_1)$ , para todo  $n$ .
4. Se  $\varphi$  for uma função convexa,  $\{X_n, \mathcal{F}_n, n \geq 1\}$  for um martingale e  $\varphi(X_n)$  for integrável, para todo  $n \geq 1$ , então  $\{\varphi(X_n), \mathcal{F}_n, n \geq 1\}$  será um sub-martingale. Se  $\varphi$  for ainda crescente e  $X_n$  um sub-martingale, então  $\varphi(X_n)$  será também um sub-martingale.
5. O resultado a seguir é um teorema de convergência para sub-martingales.

**Teorema 2.2.** (Doob) *Suponha que  $\{X_n, \mathcal{F}_n, n \geq 1\}$  seja um sub-martingale, com  $\sup_n E|X_n| < \infty$  (isto é,  $X_n$  é limitado em  $L_1$ ). Então  $\{X_n\}$  converge, com probabilidade um, para um limite  $X_\infty$ , que é integrável.*

O processo  $\{u_n\}$  em (2.90) é um caso particular de uma seqüência de variáveis que passamos a definir.

**Definição 2.13.** *Seja  $\{Y_n, n \geq 1\}$  um processo estocástico de média zero. Se*

$$E(Y_n | \mathcal{F}_{n-1}) = 0, \quad n \geq 2, \quad (2.91)$$

*então  $\{Y_n\}$  é chamado diferença martingale com respeito a  $\mathcal{F}_n$ .*

Esta definição pode ser estendida a processos vetoriais. A condição (2.91) é mais forte que a condição segundo a qual  $\{Y_n\}$  são não-correlacionadas, mas é mais fraca do que independência.

O resultado a seguir dá um teorema limite central para médias de um tal processo.

**Teorema 2.3.** (White) *Suponha que  $\{Y_n, n \geq 1\}$  seja uma diferença martingale,  $\bar{Y}_N = \sum_{n=1}^N Y_t / N$  e temos as seguintes condições válidas:*

- (a)  $E(Y_n^2) = \sigma_n^2 > 0$ , com  $\sum_{n=1}^N \sigma_n^2/N \rightarrow \sigma^2 > 0$ ;
- (b)  $E|Y_n|^r < \infty$ , para algum  $r > 2$  e para todo  $n$ ;
- (c)  $\sum_{n=1}^N Y_n^2/N \xrightarrow{P} \sigma^2$ .

Então,  $\sqrt{N}Y_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ .

## 2.9 Problemas

- Use a equação (2.8) para provar que, se  $X(t)$  é estacionário, então  $\mu(t)$  e  $\sigma^2(t)$  são constantes.
- Seja  $X(t) = \sum_{j=1}^n (A_j \cos \lambda_j t + B_j \sin \lambda_j t)$ , onde  $t = 0, \pm 1, \dots$  e  $\lambda_1, \dots, \lambda_n$  são constantes positivas e  $A_j, B_j$  são v.a. independentes, com médias 0 e variâncias  $\sigma_j^2 = \text{Var}(A_j) = \text{Var}(B_j)$ ,  $j = 1, \dots, n$ . Suponha, ainda, que os  $A_j$  sejam independentes dos  $B_j$ . O processo  $X(t)$  é fracamente estacionário? Encontre a média e a f.a.c.v. de  $X(t)$ . O processo é estritamente estacionário?
- Considere as observações:

$t$	1961	1962	1963	1964	1965	1966	1967
$Z_t$	15	19	13	17	22	18	20

Calcule  $c_k$  e  $r_k$ ,  $k = 0, 1, \dots, 6$ .

- Considere o processo estocástico  $Z_t = a_t$ , onde  $a_t$  é ruído branco, com  $t = 0, \pm 1, \pm 2, \dots$  e

$$a_t = \begin{cases} +1, & \text{com probabilidade } 1/2; \\ -1, & \text{com probabilidade } 1/2. \end{cases}$$

- Obtenha a média do processo  $Z_t$ ;
- Calcule  $\gamma_\tau$ ,  $\tau = 0, \pm 1, \pm 2, \dots$
- Calcule  $\rho_\tau$ ,  $\tau = 0, \pm 1, \dots$  e faça o seu gráfico.



5. Suponha  $\{a_t, t = 1, 2, \dots\}$  uma seqüência de v.a. independentes e identicamente distribuídas, com:

$$P(a_t = 0) = P(a_t = 1) = \frac{1}{2}.$$

- (a) O processo  $a_1 + a_2 \cos t$  é estacionário?
  - (b) O processo  $a_1 + a_2 \cos t + a_3 \cos t + \sin t$  é estacionário?
6. Se  $\{X_t, t \in \mathcal{T}\}$  e  $\{Y_t, t \in \mathcal{T}\}$  são estacionários,  $\{aX_t + bY_t, t \in \mathcal{T}\}$  será estacionário?
7. Seja  $\{Z_t\}$  um processo estacionário com média  $\mu_Z$  e função de auto-covariância  $\gamma_Z$ . Um novo processo é definido por  $Y_t = Z_t - Z_{t-1}$ . Obtenha a média e a função de auto-covariância de  $\{Y_t\}$  em termos de  $\mu_Z$  e  $\gamma_Z$ . Mostre que  $\{Y_t\}$  é um processo estacionário.
8. Prove que, se  $\{Z(t), t \in \mathbb{R}\}$  for Gaussiano e estacionário de segunda ordem, então ele será estritamente estacionário.
9. Use um programa computacional para calcular:
- (a) a média e a variância amostrais;
  - (b)  $c_k$  e  $r_k$ , para  $k = 1, \dots, 36$

da série de *índices* mensais do Ibovespa (arquivo m-ibv94.01.dat). Faça os gráficos da série e de  $r_k$ . Comente quanto à presença de tendências, sazonalidades, ciclos. Comente a natureza do gráfico de  $r_k$ .

10. Use um programa computacional para calcular:
- (a) média e variância amostrais, coeficientes de assimetria e curtose, máximo e mínimo, histograma;
  - (b) auto-correlações amostrais
- dos *log-retornos* mensais do Ibovespa do problema 9. Comente e estabeleça as diferenças com o problema anterior.
11. Prove as relações (i)–(iv) após a equação (2.69).

12. Mostre que, se  $Z \sim \mathcal{N}(0, 1)$ , então para  $\lambda$  real,

$$E(e^{\lambda Z}) = e^{\lambda^2/2}.$$

13. Use o resultado anterior para encontrar a média e covariância do movimento browniano geométrico do exemplo 2.9. Determine a variância do processo.

14. **Ponte browniana.** Considere o processo estocástico dado por

$$X(t) = W(t) - tW(1), \quad 0 \leq t \leq 1,$$

onde  $W(t)$  é o MBP no intervalo  $[0, 1]$ . Segue-se que  $X(0) = X(1) = 0$ . Mostre que  $X(t)$  é um processo gaussiano, com  $E(X(t)) = 0$  e covariância dada por  $\gamma_X(t, s) = \min(t, s) - ts$ ,  $t, s \in [0, 1]$ .

15. Mostre que o MB  $W = \{W(t), t \geq 0\}$  é um martingal com respeito à  $\sigma$ -álgebra natural  $\mathcal{F}_t = \sigma\{W(s), s \leq t\}$ .
16. Mesmo problema para o processo  $W(t)^2 - t$ .

# Modelagem ARIMA

## 3.1 Introdução

Neste capítulo trataremos de modelar séries temporais por meio de processos auto-regressivos integrados e de médias móveis, abreviadamente, ARIMA. Como vimos no Capítulo 2, um modelo ARMA (p,q) é dado pela equação de diferenças

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}, \quad (3.1)$$

onde  $\varepsilon_t \sim \text{RB}(0, \sigma_\varepsilon^2)$ . Com os operadores auto-regressivos e de média móveis definidos anteriormente (veja as expressões (2.25) e (2.42)), podemos escrever

$$\phi(B)\tilde{X}_t = \theta(B)\varepsilon_t, \quad (3.2)$$

com  $\tilde{X}_t = X_t - \mu$ . Quando  $\mu \neq 0$ , o modelo pode ser escrito

$$X_t = \theta_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$

onde  $\theta_0 = \mu(1 - \phi_1 - \dots - \phi_p)$ . No que segue, iremos supor que  $\mu = 0$ , a menos que se faça menção em contrário.

Um processo  $\{X_t, t \in \mathbb{Z}\}$  segue um modelo ARIMA(p,d,q) se  $\Delta^d X_t$  seguir um modelo ARMA(p,q), ou seja, temos

$$\phi(B)\Delta^d X_t = \theta(B)\varepsilon_t. \quad (3.3)$$

No estabelecimento de um modelo ARIMA para uma série temporal há três estágios a considerar:

- (i) Identificação
- (ii) Estimação
- (iii) Diagnóstico

Neste capítulo estudaremos com algum detalhe cada um desses estágios. Para mais detalhes o leitor pode consultar Morettin e Tolo (2006) e Box, Jenkins e Reinsel (1994).

O ciclo (i)-(iii) acima deve ser iterado se no estágio (iii) verificarmos que o modelo não é adequado para os fins desejados. Em muitas situações, mais de um modelo pode ser considerado adequado, e a decisão de qual adotar vai depender de algum critério. Por exemplo, escolhemos o modelo que minimiza o erro quadrático médio de previsão.

## 3.2 Identificação

A identificação do particular modelo ARIMA a ser ajustado aos dados é talvez a fase mais crítica do processo iterativo descrito acima.

Esta escolha é feita principalmente com base nas auto-correlações e auto-correlações parciais estimadas, que esperamos representem adequadamente as respectivas quantidades teóricas, que são desconhecidas.

Lembremos que a f.a.c.  $\rho_j$  é estimada por

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, T-1,$$

onde  $c_j$  é a estimativa da f.a.c.v  $\gamma_j$ ,

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} [(X_t - \bar{X})(X_{t+j} - \bar{X})], \quad j = 0, 1, \dots, T-1,$$

sendo  $\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$  a média amostral. Como as auto-correlações verdadeiras são funções pares, colocamos  $c_{-j} = c_j$  e  $r_{-j} = r_j$ .

Como veremos adiante, será necessário uma verificação mais ou menos grosseira para saber se  $\rho_j$  é nula além de um certo “lag”. Uma expressão aproximada para a variância de  $r_j$ , para um processo estacionário normal, é dada por

$$\text{Var}(r_j) \simeq \frac{1}{T} \sum_{v=-\infty}^{\infty} [\rho_v^2 + \rho_{v+j}\rho_{v-j} - 4\rho_j\rho_v\rho_{v-j} + 2\rho_v^2\rho_j^2]. \quad (3.4)$$

Para um processo em que as auto-correlações são nulas para  $v > q$ , todos os termos do lado direito de (3.4) anulam-se para  $j > q$ , exceto o primeiro, obtendo-se

$$\text{Var}(r_j) \simeq \frac{1}{T} \left[ 1 + 2 \sum_{v=1}^q \rho_v^2 \right], \quad j > q. \quad (3.5)$$

Como desconhecemos as auto-correlações  $\rho_v$ , substituímo-las por  $r_v$ , obtendo-se uma estimativa para (3.5),

$$\widehat{\sigma^2}(r_j) \simeq \frac{1}{T} \left[ 1 + 2 \sum_{v=1}^q r_v^2 \right], \quad j > q. \quad (3.6)$$

Para  $T$  suficientemente grande e sob a hipótese que  $\rho_j = 0$ , para  $j > q$ , a distribuição de  $r_j$  é aproximadamente normal, com média igual a zero e variância dada por (3.5) (Jenkins e Watts, 1968, p.187). Assim, pode-se construir um intervalo de confiança aproximado para as auto-correlações,

$$r_j \pm t_\gamma \cdot \hat{\sigma}(r_j), \quad (3.7)$$

onde  $t_\gamma$  é o valor da estatística  $t$  de Student com  $T - 1$  graus de liberdade, tal que  $P(-t_\gamma < t < t_\gamma) = \gamma$ . Na prática usa-se  $t_\gamma = 2$ , correspondendo a  $\gamma = 0,95$ , aproximadamente, de modo que podemos considerar  $\rho_j$  como sendo significativamente diferente de zero se

$$|r_j| > 2\hat{\sigma}(r_j), \quad j > q. \quad (3.8)$$

Como vimos no Capítulo 2, os processos  $\text{AR}(p)$ ,  $\text{MA}(q)$  e  $\text{ARMA}(p, q)$  apresentam f.a.c. com características especiais. Assim:

- (i) um processo  $\text{AR}(p)$  tem f.a.c. que decai de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão;

- (ii) um processo MA( $q$ ) tem f.a.c. finita, no sentido que ela apresenta um corte após o “lag”  $q$ ;
- (iii) um processo ARMA( $p, q$ ) tem f.a.c. infinita em extensão, que decai de acordo com exponenciais e/ou senóides amortecidas após o “lag”  $q - p$ .

Estas observações serão úteis no procedimento de identificação do modelo a ajustar aos dados observados; calculando-se as estimativas das f.a.c., que acreditamos reproduzir adequadamente as verdadeiras f.a.c. desconhecidas e comparando seu comportamento com o descrito acima, para cada modelo, tentaremos escolher um (ou mais) modelo (modelos, respectivamente) que descreva(m) a série observada.

Em particular, a f.a.c. é útil para identificar modelos MA, dada a característica (ii) acima, não sendo útil para identificar modelos ARMA, que têm f.a.c. complicada.

Box, Jenkins e Reinsel (1994) propõem a utilização de um outro instrumento para facilitar este procedimento de identificação: a *função de autocorrelação parcial* (f.a.c.p.).

Vamos denotar por  $\phi_{kj}$  o  $j$ -ésimo coeficiente de um modelo AR( $k$ ), de tal modo que  $\phi_{kk}$  seja o último coeficiente. Sabemos que

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}, \quad j = 1, \dots, k,$$

a partir das quais obtemos as equações de Yule-Walker

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_2 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (3.9)$$

Resolvendo estas equações sucessivamente para  $k = 1, 2, 3, \dots$  obtemos

$$\begin{aligned} \phi_{11} &= \rho_1, \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \end{aligned}$$

$$\phi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

e, em geral,

$$\phi_{kk} = \frac{|\mathbf{P}_k^*|}{|\mathbf{P}_k|},$$

onde  $\mathbf{P}_k$  é a matriz de auto-correlações e  $\mathbf{P}_k^*$  é a matriz  $\mathbf{P}_k$  com a última coluna substituída pelo vetor de auto-correlações.

A quantidade  $\phi_{kk}$ , encarada como função de  $k$ , é chamada *função de auto-correlação parcial*.

Pode-se demonstrar que  $\phi_{kk}$  é igual à correlação parcial entre as variáveis  $X_t$  e  $X_{t-k}$  ajustadas às variáveis intermediárias  $X_{t-1}, \dots, X_{t-k+1}$ . Ou seja,  $\phi_{kk}$  mede a correlação remanescente entre  $X_t$  e  $X_{t-k}$  depois de eliminada a influência de  $X_{t-1}, \dots, X_{t-k+1}$ .

**Exemplo 3.1.** A correlação entre os valores ajustados  $X_t - \phi_{11}X_{t-1}$  e  $X_{t-2} - \phi_{11}X_{t-1}$  é dada por

$$\begin{aligned} & \text{Corr}(X_t - \phi_{11}X_{t-1}, X_{t-2} - \phi_{11}X_{t-1}) \\ &= \text{Corr}(X_t - \rho_1X_{t-1}, X_{t-2} - \rho_1X_{t-1}) \\ &= \frac{\text{Cov}(X_t - \rho_1X_{t-1}, X_{t-2} - \rho_1X_{t-1})}{(\text{Var}(X_t - \rho_1X_{t-1})\text{Var}(X_{t-2} - \rho_1X_{t-1}))^{1/2}} \\ &= \frac{\gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0}{\gamma_0 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0} \\ &= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_{22}. \end{aligned}$$

Será necessário estimar a f.a.c.p. de um processo AR, MA ou ARMA. Uma maneira consiste em estimar, sucessivamente, modelos auto-regressivos de ordens  $p = 1, 2, 3, \dots$  por mínimos quadrados e tomar as estimativas do último coeficiente de cada ordem.

Outra maneira consiste em substituir nas equações de Yule-Walker as f.a.c.  $\rho_j$  por suas estimativas, ou seja, considerar

$$r_j = \hat{\phi}_{k1}r_{j-1} + \cdots + \hat{\phi}_{kk}r_{j-k}, \quad j = 1, \dots, k,$$

e resolver (3.9) com  $\rho_j$  substituída por  $r_j$  e  $\phi_{kj}$  substituída por  $\hat{\phi}_{kj}$ .

Quenouille (1949) mostra que, sob a suposição que o processo seja  $AR(p)$ , as f.a.c.p. estimadas de ordem  $p+1, p+2, \dots$  são, aproximadamente, independentemente distribuídas, com

$$\text{Var}(\hat{\phi}_{kk}) \approx \frac{1}{T}, \quad k \geq p+1.$$

Se o número de observações,  $T$ , for suficientemente grande,  $\hat{\phi}_{kk}$  tem distribuição aproximada normal, o que permite a construção de intervalos de confiança para  $\phi_{kk}$ .

Podemos resolver as equações (3.9) utilizando o algoritmo de Durbin-Levinson. Veja Morettin (1984) para detalhes.

Pode-se demonstrar (veja Box, Jenkins e Reinsel, 1994) que, para os processos estudados, temos:

- (i) um processo  $AR(p)$  tem f.a.c.p.  $\phi_{kk} \neq 0$ , para  $k \leq p$  e  $\phi_{kk} = 0$ , para  $k > p$ ;
- (ii) um processo  $MA(q)$  tem f.a.c.p. que se comporta de maneira similar à f.a.c. de um processo  $AR(p)$ : é dominada por exponenciais e/ou senóides amortecidas;
- (iii) um processo  $ARMA(p, q)$  tem f.a.c.p. que se comporta como a f.a.c.p. de um processo  $MA$  puro.

Segue-se que a f.a.c.p. é útil para identificar modelos  $AR$  puros, não sendo tão útil para identificar modelos  $MA$  e  $ARMA$ .

Vimos acima que sob a hipótese que o processo seja  $AR(p)$ ,

$$\text{Var}(\hat{\phi}_{jj}) \simeq \frac{1}{T}, \quad j > p, \tag{3.10}$$



de modo que

$$\hat{\sigma}(\hat{\phi}_{jj}) \simeq \frac{1}{\sqrt{T}}, \quad j > p. \quad (3.11)$$

Além disso, para  $T$  grande e sob a hipótese que o processo seja  $\text{AR}(p)$ ,  $\hat{\phi}_{jj}$  terá distribuição aproximadamente normal, com média zero e variância (3.10), de modo que consideraremos  $\phi_{jj}$  significativamente diferente de zero se

$$|\hat{\phi}_{jj}| > \frac{2}{\sqrt{T}}, \quad j > p. \quad (3.12)$$

**Exemplo 3.2.** Na Tabela 3.1 temos as estimativas das auto-correlações e de seus respectivos desvios padrões, bem como as auto-correlações parciais estimadas, para as séries simuladas das figuras 2.6, 2.8 e 2.9, a saber,

$$\begin{aligned} \text{AR}(1) : \quad & X_t = 0,8X_{t-1} + \varepsilon_t; \\ \text{MA}(1) : \quad & X_t = \varepsilon_t - 0,8\varepsilon_{t-1}; \\ \text{ARMA}(1,1) : \quad & X_t = 0,8X_{t-1} + \varepsilon_t - 0,3\varepsilon_{t-1}. \end{aligned}$$

Na tabela também estão indicadas as médias e os desvios padrões amostrais de cada série, bem como destacados com um asterisco (\*) os valores que caem fora do intervalo de dois desvios padrões.

Na Figura 3.1 temos os gráficos de  $r_j$ ,  $\hat{\phi}_{jj}$  para cada um dos modelos. Os intervalos de confiança para  $\phi_{jj}$  estão indicados em linha pontilhada.

Note que os valores estimados podem diferir bastante dos valores verdadeiros. Por exemplo, no caso do modelo  $\text{AR}(1)$ ,  $\phi = 0,8$  e  $\rho_1 = 0,8$ ,  $\rho_2 = 0,64$ , etc. Para o modelo  $\text{MA}(1)$ ,  $\theta = 0,8$  e  $\rho_1 = -\theta/(1+\theta^2) = -0,488$ , sendo as demais nulas. Para o  $\text{ARMA}(1,1)$ ,  $\rho_1 = 0,622$ ,  $\rho_2 = 0,498$ , etc. O valor de  $T$  aqui é 100. À medida que aumentarmos o tamanho da série, os valores estimados estarão mais próximos dos valores verdadeiros.

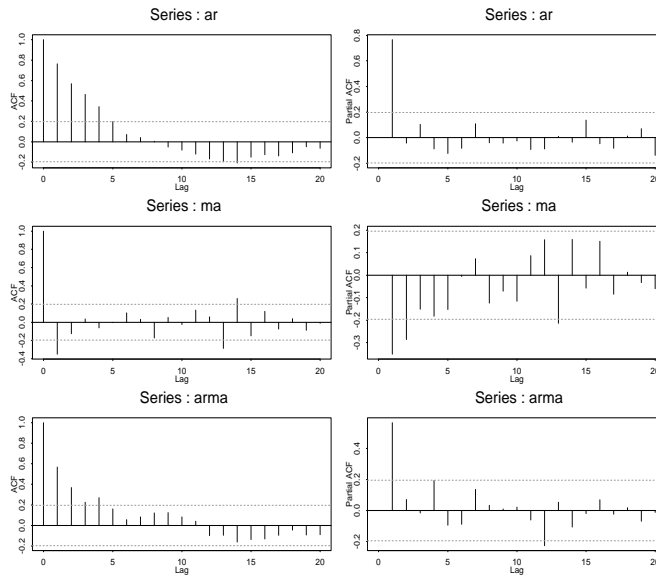


Figura 3.1: auto-correlações e auto-correlações parciais amostrais para as séries simuladas AR(1), MA(1) e ARMA(1,1).

### 3.2.1 Procedimento de Identificação

O objetivo da identificação é determinar os valores de  $p$ ,  $d$  e  $q$  do modelo  $\text{ARIMA}(p, d, q)$ . O procedimento de identificação consiste de três partes:

- (a) verificar se existe necessidade de uma transformação na série original, com o objetivo de estabilizar sua variância;
- (b) tomar diferenças da série, obtida no item (a), tantas vezes quantas necessárias para se obter uma série estacionária, de modo que o processo  $\Delta^d X_t$  seja reduzido a um  $\text{ARMA}(p, q)$ . O número de diferenças,  $d$ , necessárias para que o processo se torne estacionário, é alcançado quando a f.a.c. amostral de  $W_t = \Delta^d X_t$  decresce rapidamente para zero. Neste estágio, a utilização de um teste para verificar a existência de raízes unitárias no polinômio auto-regressivo, pode ser de grande utilidade. O teste de Dickey e Fuller (1979) será estudado no Capítulo 4;

Tabela 3.1: auto-correlações amostrais e respectivos desvios padrões e auto-correlações parciais amostrais para as séries simuladas AR(1), MA(1) e ARMA(1,1).

lag	AR(1)			MA(1)			ARMA(1)		
$j$	$r_j$	$\hat{\sigma}(r_j)$	$\hat{\phi}_{jj}$	$r_j$	$\hat{\sigma}(r_j)$	$\hat{\phi}_{jj}$	$r_j$	$\hat{\sigma}(r_j)$	$\hat{\phi}_{jj}$
1	0,765*	0,10	0,765*	-0,352*	0,10	-0,352*	0,567*	0,10	0,567*
2	0,568*	0,15	-0,041	-0,127	0,11	-0,286*	0,370*	0,13	0,071
3	0,464*	0,17	0,103	0,036	0,11	-0,152	0,225	0,14	-0,015
4	0,344	0,18	-0,088	-0,063	0,11	-0,182	0,269	0,14	0,193
5	0,197	0,19	-0,124	-0,004	0,11	-0,153	0,161	0,15	-0,096
6	0,071	0,19	-0,082	-0,102	0,11	-0,006	0,056	0,15	-0,090
7	0,041	0,19	0,109	0,031	0,11	0,073	0,084	0,15	0,136
8	0,005	0,19	-0,039	-0,173	0,11	-0,125	0,123	0,15	0,034
9	-0,052	0,19	-0,042	0,051	0,12	-0,071	0,125	0,15	0,010
10	-0,084	0,19	-0,024	-0,026	0,12	-0,116	0,084	0,15	0,022
$\bar{X}$		0,532			-0,042			1,077	
$S$		2,462			1,284			1,312	

- (c) identificar o processo ARMA( $p, q$ ) resultante por meio da análise das auto-correlações e auto-correlações parciais estimadas, cujos comportamentos devem imitar os comportamentos das respectivas quantidades teóricas. Estes comportamentos, para modelos AR, MA e ARMA, foram abordados no Capítulo 2 e um resumo das propriedades destes modelos encontra-se na Tabela 3.2.

A justificativa do item (b), do procedimento de identificação é o seguinte. Vimos que, para um modelo ARMA estacionário, as f.a.c. são dadas por

$$\rho_j = A_1 G_1^j + A_2 G_2^j + \cdots + A_p G_p^j, \quad j > q - p, \quad (3.13)$$

supondo raízes distintas. Como  $\phi(B) = \prod_{i=1}^p (1 - G_i B)$  e as raízes de  $\phi(B) = 0$  devem estar fora do círculo unitário, devemos ter  $|G_i| < 1$ . Segue-se de (3.13) que, se nenhuma raiz estiver muito próxima do círculo unitário, as auto-correlações  $\rho_j$  decairão para zero, para valores moderados de  $j$ .

Por outro lado, suponha que uma raiz real,  $G_1$ , esteja próxima de um, ou seja,  $G_1 = 1 - \varepsilon$ ,  $\varepsilon > 0$  pequeno. Como  $G_1^j = (1 - \varepsilon)^j \simeq 1 - j\varepsilon$ , vem que  $\rho_j \simeq A_1(1 - j\varepsilon)$ , o que mostra que a f.a.c. decairá lentamente para zero e de forma aproximadamente linear.

O maior problema, neste estágio do procedimento é evitar um excesso de diferenças. Na prática,  $d = 0, 1$  ou  $2$  e é suficiente inspecionar as primeiras 15 ou 20 auto-correlações da série e de suas diferenças.

Convém testar se  $E(W_t) = \mu_W$  é zero, comparando  $\bar{W}$  com seu desvio padrão estimado. A Tabela 3.3 fornece as variâncias de  $\bar{W}$  para alguns modelos usuais. Lembrar que se  $d = 0$ ,  $\bar{W} = \bar{X}$  e  $c_0$  é a variância amostral.

Como dissemos antes, a f.a.c. e f.a.c.p. são úteis para identificar modelos MA e AR puros, respectivamente. Para identificar modelos ARMA convém considerar modelos simples, como ARMA(1,1), ARMA(1,2), ARMA(2,1) ou ARMA(2,2) e depois da estimação, escolher aquele que melhor atende a alguns critérios. Veja a seção 3.2.2.

**Exemplo 3.3.** Suponha que temos os seguintes dados:

j	1	2	3	4	5	6	7	8
$r_j$	0,81	0,69	0,58	0,44	0,30	0,26	0,19	0,15
$\hat{\phi}_{jj}$	0,81	0,11	-0,03	-0,12	-0,13	0,17	-0,01	0,02

$$T = 50, \quad \bar{X} = 0,5327, \quad S^2 = 6,0579$$

Temos que

$$\hat{\sigma}(\hat{\phi}_{jj}) \simeq \frac{1}{\sqrt{T}} = \frac{1}{\sqrt{50}} \simeq 0,14, \quad \text{logo } 2\hat{\sigma}(\hat{\phi}_{jj}) = 0,28,$$

o que mostra que só  $\phi_{11} \neq 0$ ;  $\{r_j\}$  desqualifica a possibilidade de um processo MA e  $\{\hat{\phi}_{jj}\}$  sugere um processo AR(1).

Para um processo AR(1), usando a Tabela 3.3, temos que

$$\text{Var}(\bar{X}) = \frac{c_0(1 + r_1)}{n(1 - r_1)} = \frac{6,0579(1 + 0,81)}{50(1 - 0,81)} = 1,1542$$

e  $\hat{\sigma}(\bar{X}) = 1,0743$ ; como  $\bar{X} = 0,5327$ , a média pode ser considerada igual a zero e o modelo sugerido é

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{RB}(0, \sigma_\varepsilon^2).$$

**Exemplo 3.4.** Vamos agora identificar um ou mais modelos preliminares para a série ICV (arquivo m-icv70.79.dat), no período de janeiro de 1970 a junho de 1979, utilizando  $T = 114$  observações.

Inicialmente, vamos considerar a transformação logaritmica

$$Y_t = \ln(ICV), \quad t = 1, \dots, 114.$$

A Figura 3.2 apresenta a série original e a série transformada. Vemos que a transformação logaritmica tornou a série aproximadamente linear.

A Tabela 3.4 e a Figura 3.3 apresentam os valores e as correspondentes representações gráficas das funções de auto-correlação e auto-correlação parcial da série  $Y_t$  e de suas diferenças de ordens 1 e 2.

Analisando o comportamento das funções vemos que a f.a.c. de  $Y_t$  não decresce rapidamente para zero, indicando a não-estacionariedade da série. Para selecionarmos um valor apropriado para  $d$ , podemos verificar, graficamente, quantas diferenças são necessárias para que a f.a.c. convirja rapidamente para zero. Este fato parece ocorrer para  $d = 1$  ou, talvez,  $d = 2$ .

A Figura 3.4 apresenta os gráficos da primeira e segunda diferenças do  $\ln(ICV)$ .

Observando a f.a.c. amostral de  $(1 - B)^2 Y_t$ , verificamos que  $r_1 = -0,32$ , que além de ser negativo, assume um valor próximo de  $-0,5$ . Calculando as variâncias das diferenças das séries, temos que  $\text{Var}((1 - B)Y_t) = 0,00013$  e  $\text{Var}((1 - B)^2 Y_t) = 0,000162$  indicando um aumento de cerca de 25% da variância da série com duas diferenças em relação àquela com uma diferença. Assim, escolhemos  $d = 1$ .

Analisando o comportamento das f.a.c. e f.a.c.p da série  $(1 - B) \ln(ICV)$ , Tabela 3.4 e Figura 3.3, podemos notar que somente  $\phi_{11} \neq 0$ , pois  $\hat{\phi}_{11} = 0,49$  e  $\hat{\sigma}(\hat{\phi}_{11}) = 1/\sqrt{114} = 0,094$ . Assim, um modelo preliminar para  $\ln(ICV)$  é um ARIMA(1,1,0).

Para verificar se uma constante deve, ou não, ser adicionada ao modelo, observamos que

$$\overline{W} = 0,02116,$$

$$\text{Var}(\overline{W}) = \frac{c_0(1+r_1)}{n(1-r_1)} = \frac{0,00014(1+0,49)}{113(1-0,49)} = 3,62 \times 10^{-6},$$

ou seja,  $\hat{\sigma}(\overline{W}) = 0,0019$ . Assim, podemos concluir que a média da série  $(1-B)Y_t$  é diferente de zero e que uma constante deve ser incluída no modelo.

Tabela 3.2: Comportamento das f.a.c. e f.a.c.p. de um processo ARIMA( $p, d, q$ )

Ordem	$(1, d, 0)$	$(0, d, 1)$
comportamento de $\rho_k$	decai exponencialmente	somente $\rho_1 \neq 0$
comportamento de $\phi_{kk}$	somente $\phi_{11} \neq 0$	decaimento exponencial dominante
estimativas iniciais	$\phi = \rho_1$	$\rho_1 = -\frac{\theta}{1+\theta^2}$
região de admissibilidade	$-1 < \phi < 1$	$-1 < \theta < 1$
Ordem	$(2, d, 0)$	$(0, d, 2)$
comportamento de $\rho_k$	mistura de exponenciais ou ondas senóides amortecidas	somente $\rho_1 \neq 0$ e $\rho_2 \neq 0$
comportamento de $\phi_{kk}$	somente $\phi_{11} \neq 0$ e $\phi_{22} \neq 0$	dominada por mistura de exponenciais ou senóides amortecidas
estimativas iniciais	$\begin{cases} \phi_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}, \\ \phi_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \end{cases}$	$\begin{cases} \rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}, \\ \rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2} \end{cases}$
região de admissibilidade	$\begin{cases} -1 < \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ \phi_2 + \phi_1 < 1 \end{cases}$	$\begin{cases} -1 < \theta_2 < 1 \\ \theta_2 - \theta_1 < 1 \\ \theta_2 + \theta_1 < 1 \end{cases}$
Ordem	$(1, d, 1)$	
comportamento de $\rho_k$	decai exponencialmente após o lag 1	
comportamento de $\phi_{kk}$	dominada por decaimento exponencial após o lag 1	
estimativas iniciais	$\rho_1 = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta}, \quad \rho_2 = \rho_1\phi$	
região de admissibilidade	$-1 < \phi < 1, \quad -1 < \theta < 1$	

Tabela 3.3: Variâncias aproximadas para  $\bar{W}$ , onde  $W_t = \Delta^d X_t$ ,  $n = T - d$

AR(1)	MA(1)	ARMA(1,1)
$\frac{c_0(1+r_1)}{n(1-r_1)}$	$\frac{c_0(1+2r_1)}{n}$	$\frac{c_0}{n} \left[ 1 + \frac{2r_1^2}{r_1 - r_2} \right]$
AR(2)		MA(2)
$\frac{c_0(1+r_1)(1-2r_1^2+r_2)}{n(1-r_1)(1-r_2)}$		$\frac{c_0(1+2r_1+2r_2)}{n}$

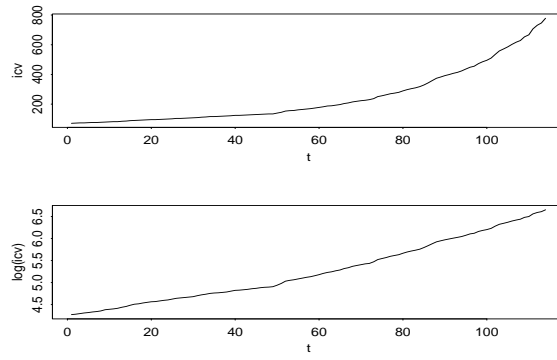


Figura 3.2: Gráfico das séries ICV e  $\ln(\text{ICV})$

Logo, o modelo preliminar para a série  $\ln(\text{ICV})$  é um  $\text{ARIMA}(1,1,0)$  com uma constante, isto é,

$$(1 - B)(1 - \phi_1 B)Y_t = \theta_0 + \varepsilon_t, \quad (3.14)$$

onde  $Y_t = \ln(\text{ICV})$  e  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$ .

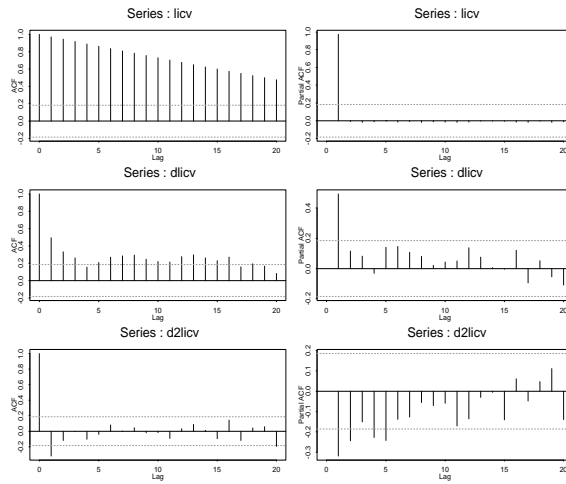


Figura 3.3: (a) auto-correlações e (b) auto-correlações parciais das séries  $\ln(\text{ICV})$ ,  $(1 - B)\ln(\text{ICV})$  e  $(1 - B)^2\ln(\text{ICV})$

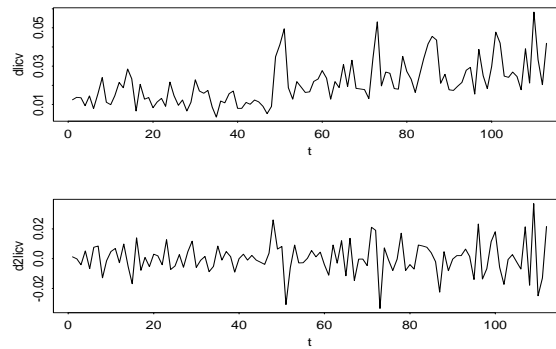


Figura 3.4: Gráfico das séries  $(1 - B)\ln(\text{ICV})$  e  $(1 - B)^2\ln(\text{ICV})$

**Exemplo 3.5.** Considere a série de retornos mensais do Ibovespa, de junho de 1994 a agosto de 2001 (arquivo m-ibv94.01.dat), com  $T = 86$  observações. Se denotarmos a série de índices por  $X_t$ , então os retornos são dados por



$$r_t = \log \left( \frac{X_t}{X_{t-1}} \right) = \log X_t - \log X_{t-1} = \Delta(\log X_t).$$

O gráfico da série, as f.a.c. e f.a.c.p. amostrais estão na Figura 3.5, mostrando que esta série pode ser considerada um ruído branco. Como a média amostral é 0,003872 e o desvio padrão amostral é 0,137428, podemos concluir que um modelo adequado para descrever esses retornos é

$$r_t = \Delta(\log X_t) = \varepsilon_t, \quad \text{com } \varepsilon_t \sim \text{RB}(0, \sigma^2),$$

e  $\hat{\sigma}^2 = 0,01889$ , ou ainda,

$$\log X_t = \log X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{RB}(0, \sigma^2),$$

ou seja, os logaritmos dos índices do Ibovespa seguem um passeio casual com média zero (sem “drift”).

Como veremos mais adiante, esta é, em geral, uma característica de retornos de ativos financeiros: são não-correlacionados. Mas podem não ser independentes, o que se pode verificar considerando-se os quadrados dos retornos. No Capítulo 5 voltaremos a tratar desse assunto.

### 3.2.2 Formas alternativas de identificação

A partir de 1970 foram propostos vários procedimentos para identificação de modelos ARMA. A idéia é escolher as ordens  $k$  e  $l$  que minimizem a quantidade

$$P(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{C(T)}{T}, \quad (3.15)$$

em que  $\hat{\sigma}_{k,l}^2$  é uma estimativa da variância residual obtida ajustando um modelo ARMA( $k, l$ ) às  $T$  observações da série e  $C(T)$  é uma função do tamanho da série.

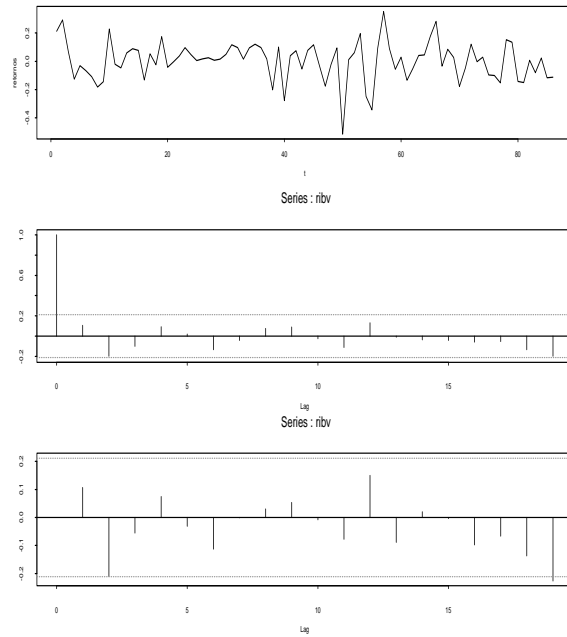


Figura 3.5: Retornos Ibovespa (a) gráfico da série  $r_t$  (b) f.a.c. amostral  
(c) f.a.c.p. amostral

A quantidade  $(k + l)C(T)/T$ , denominada termo penalizador, aumenta quando o número de parâmetros aumenta, enquanto que a variância residual  $\hat{\sigma}_{k,l}^2$  diminui. Assim, minimizar (3.15) corresponde a identificar as ordens  $k$  e  $l$  que equilibrem esse comportamento.

É natural supor que as ordens selecionadas aumentem quando  $T$  cresce. Hannan (1982) sugere limites superiores dados por  $(\ln T)^\alpha$ ,  $0 < \alpha < \infty$ .

Iremos citar agora alguns procedimentos de identificação que minimizam funções penalizadoras particulares.

#### A. Critério de Informação AIC

Akaike (1973, 1974) sugere escolher o modelo cujas ordens  $k$  e  $l$  minimizam o critério

$$AIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + \frac{2(k+l)}{T} \quad (3.16)$$

onde  $\hat{\sigma}_{k,l}^2$  é o estimador de máxima verossimilhança de  $\sigma^2$  para um modelo ARMA( $k, l$ ) (veja a seção 3.3).

Tabela 3.4: auto-correlações e auto-correlações parciais estimadas das séries (a)  $\ln(\text{ICV})$ , (b)  $(1-B)\ln(\text{ICV})$ , (c)  $(1-B)^2\ln(\text{ICV})$ .

lag	(a)		(b)		(c)	
	f.a.c.	f.a.c.p.	f.a.c.	f.a.c.p.	f.a.c.	f.a.c.p.
1	0,97	0,97	0,49	0,49	-0,32	-0,32
2	0,94	-0,01	0,33	0,12	-0,12	-0,24
3	0,92	-0,02	0,26	0,08	0,00	-0,15
4	0,89	-0,01	0,16	-0,03	-0,10	-0,23
5	0,86	0,00	0,21	0,14	-0,04	-0,24
6	0,84	-0,02	0,27	0,15	0,08	-0,14
7	0,81	-0,01	0,29	0,11	0,00	-0,13
8	0,78	-0,02	0,29	0,08	0,05	-0,06
9	0,76	-0,01	0,25	0,02	-0,02	-0,07
10	0,73	-0,02	0,22	0,04	-0,02	-0,06
11	0,70	-0,02	0,21	0,05	-0,09	-0,17
12	0,68	-0,02	0,28	0,14	0,03	-0,14
13	0,65	0,00	0,30	0,08	0,09	-0,03
14	0,62	0,00	0,26	0,01	0,01	-0,01
15	0,60	-0,01	0,23	0,00	-0,10	-0,14
16	0,57	-0,01	0,27	0,12	0,14	0,06
17	0,55	-0,02	0,16	-0,09	-0,12	-0,05
18	0,53	-0,01	0,19	0,05	0,04	0,05
19	0,50	-0,02	0,17	-0,06	0,06	0,11
20	0,48	-0,01	0,08	-0,11	-0,19	-0,14

O que se faz, então, é estipular valores limites superiores  $K$  e  $L$  para  $k$  e  $l$  e calcular (3.16) para todas as possíveis combinações  $(k, l)$  com  $0 \leq k \leq K$  e  $0 \leq l \leq L$ . Em geral,  $K$  e  $L$  são funções de  $T$ , por exemplo,  $K = L = \ln T$ .

Para o caso de modelos  $AR(p)$ , o critério AIC reduz-se a

$$AIC(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{T}, \quad k \leq K. \quad (3.17)$$

Shibata (1976) demonstra que a minimização do AIC fornece estimativas inconsistentes da verdadeira ordem do processo AR. Hannan (1980) generalizou o resultado de Shibata para o processo  $ARMA(p, q)$ .

Existem várias correções para melhorar o comportamento do AIC, no sentido de diminuir a probabilidade de selecionar uma ordem maior do que a verdadeira.

Hurvich e Tsai (1989) propõem uma correção para o AIC, que no caso de um processo AR é dada por

$$AIC_c(k) = AIC(k) + \frac{2(k+1)(k+2)}{T-k+2}, \quad k \leq K$$

e utilizando simulações, mostram que esta correção é útil quando  $T$  é pequeno ou quando  $K$  é uma fração “moderadamente grande” de  $T$ .

#### B. Critério de Informação Bayesiano BIC

Akaike (1977), Rissanen (1978) e Schwarz (1978) sugerem minimizar o Critério de Informação Bayesiano, dado por

$$BIC = -2 \ln(FV) + (\text{número de parâmetros}) \ln T,$$

na qual  $FV$  é a função de verossimilhança.

No caso de um processo  $ARMA$ , essa quantidade é dada por

$$BIC(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{\ln T}{T}. \quad (3.18)$$

Hannan (1980) mostra que, sob determinadas condições, as estimativas  $k$  e  $l$  que minimizam a expressão (3.18) são fortemente consistentes.

**Exemplo 3.6.** Na Figura 3.6 temos o gráfico da série de retornos  $r_t$  diários da Petrobrás PN, de 3/1/1995 a 27/12/2000, suas f.a.c. e f.a.c.p. amostrais; estas indicam que a série é auto-correlacionada e que um modelo auto-regressivo é

apropriado. Fixemos uma ordem máxima  $K = 10$ . Na Tabela 3.5 temos os valores de AIC, BIC e do logaritmo da verossimilhança para cada modelo estimado. Vemos que a ordem  $p = 9$  deve ser escolhida, usando qualquer um dos três critérios. Estimando este modelo, somente os coeficientes  $\phi_1$ ,  $\phi_3$  e  $\phi_9$  são significativos (veja a seção 3.3), ou seja, o modelo final é

$$r_t = \phi_1 r_{t-1} + \phi_3 r_{t-3} + \phi_9 r_{t-9} + \varepsilon_t.$$

Lembremos que o retorno é definido como a primeira diferença do logaritmo da série original de preços das ações.

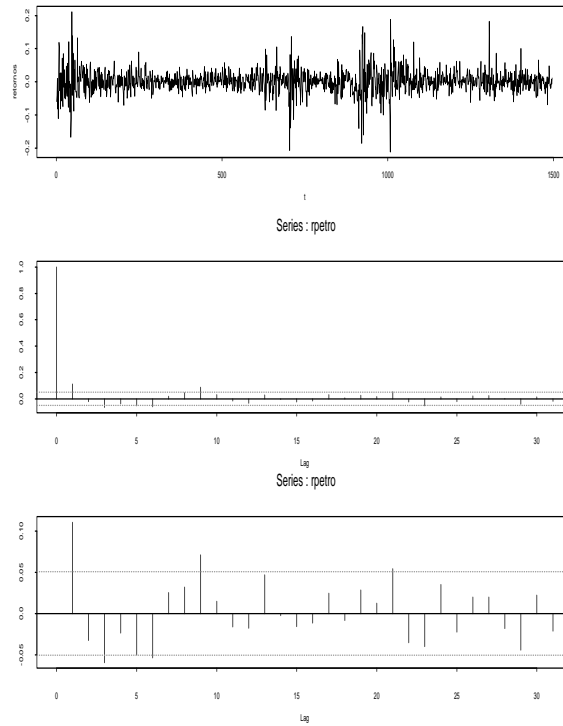


Figura 3.6: Retornos Petrobrás (a) gráfico da série (b) f.a.c. amostral  
(c) f.a.c.p. amostral

Tabela 3.5: Valores de AIC, BIC e log-verossimilhança para o exemplo 3.6. (\*) indica melhor modelo.

Modelo	AIC	BIC	log-verossimilhança
AR(1)	-3,90508	-3,89798	2924,95
AR(2)	-3,90486	-389421	2923,84
AR(3)	-3,90652	-3,89231	2924,12
AR(4)	-3,90953	-3,89176	2925,42
AR(5)	-3,91042	-3,88909	2925,13
AR(6)	-3,91898	-3,89408	2930,56
AR(7)	-3,91793	-3,88945	2928,81
AR(8)	-3,92376	-3,89170	2932,20
AR(9)	-3,93185*	-3,89978*	2936,26*
AR(10)	-3,93076	-3,89510	2934,48

### 3.3 Estimação

Tendo-se identificado um modelo provisório para a série temporal, o passo seguinte é estimar seus parâmetros. Os métodos dos momentos, mínimos quadrados (MQ) e máxima verossimilhança (MV) podem ser usados. Estimadores obtidos pelo método dos momentos não têm propriedades boas, quando comparadas com os demais. Por isso, esses estimadores em geral são usados como valores iniciais para procedimentos mais complexos, como MV.

Vamos analisar separadamente os modelos AR, MA e ARMA.

#### 3.3.1 Modelos Auto-regressivos

Consideremos o modelo AR(p)

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (3.19)$$

onde  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$ . Considere  $\phi = (\phi_0, \phi_1, \dots, \phi_p)'$  o vetor contendo os coeficientes e seja  $\xi = (\phi, \sigma^2)'$ . O objetivo é estimar  $\phi$  e  $\sigma^2$ .

Para ilustrar, considere o caso  $p = 1$ , ou seja,

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t.$$

Suponha que tenhamos os dados  $\mathbf{X} = (X_1, \dots, X_T)'$ . Então, supondo  $\varepsilon_t$  com distribuição normal e condicional ao valor  $X_0 = x_0$ , a função de verossimilhança é dada por

$$L(\boldsymbol{\xi}|\mathbf{X}, x_0) = (2\pi\sigma^2)^{-T/2} \exp\left\{\frac{-1}{2\sigma^2} \sum_{t=1}^T (X_t - \phi_0 - \phi_1 X_{t-1})^2\right\}$$

e considerando a log-verossimilhança temos

$$\begin{aligned} \ell(\boldsymbol{\xi}|\mathbf{X}, x_0) &\approx (-T/2) \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (X_t - \phi_0 - \phi_1 X_{t-1})^2 \\ &= (-T/2) \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\phi}|\mathbf{X}, x_0). \end{aligned} \quad (3.20)$$

onde  $S(\cdot|\cdot)$  é a soma de quadrados condicional. Segue-se que maximizar  $\ell$ , com  $\sigma^2$  fixo, é equivalente a minimizar  $S$  e estimadores de MV de  $\boldsymbol{\phi}$  serão equivalentes a estimadores de MQ.

Os  $\varepsilon_t$  são obtidos recursivamente:

$$\varepsilon_1 = X_1 - \phi_0 - \phi_1 X_0 = X_1 - \phi_0 - \phi_1 x_0,$$

$$\varepsilon_2 = X_2 - \phi_0 - \phi_1 X_1, \text{ etc}$$

Outra possibilidade é considerar apenas os dados efetivamente observados e começar a regressão de  $X_t$  sobre 1 e  $X_{t-1}$  a partir de  $t = 2, \dots, T$ .

Para o caso geral, podemos obter os estimadores dos coeficientes do modelo pela regressão em (3.19), para  $t = p+1, \dots, T$ . Os resíduos serão dados por  $\hat{\varepsilon}_t = X_t - \hat{X}_t$ , com

$$\hat{X}_t = \hat{\phi}_0 + \hat{\phi}_1 X_{t-1} + \dots + \hat{\phi}_p X_{t-p}.$$

Um estimador de  $\sigma^2$  é obtido de

$$\hat{\sigma}^2 = \frac{\sum_{t=p+1}^T \hat{\varepsilon}_t^2}{T - 2p - 1}. \quad (3.21)$$

O valor do denominador em (3.21) justifica-se pelo fato de termos  $T - p$  termos na soma do numerador e  $p + 1$  parâmetros a estimar.

Além de estimadores de MQ, podemos considerar também os estimadores de Yule-Walker, obtidos das equações de Yule-Walker amostrais, que têm boas propriedades e são fáceis de calcular usando-se o algoritmo de Durbin-Levinson. Veja Brockwell e Davis (1991), por exemplo.

### 3.3.2 Modelos de Médias Móveis

Para o modelo MA(q)

$$X_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (3.22)$$

onde novamente  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ , podemos considerar também estimadores de MV condicionais. Aqui, supomos que  $\varepsilon_0 = \varepsilon_{-1} = \dots = 0$  e obtemos recursivamente

$$\begin{aligned} \varepsilon_1 &= X_1 - \theta_0, \\ \varepsilon_2 &= X_2 - \theta_0 + \varepsilon_1 \theta_1, \text{ etc} \end{aligned}$$

e a partir desses valores formamos a função de verossimilhança.

**Exemplo 3.7.** Consideremos um processo ARIMA(0, 1, 1),

$$\Delta X_t = (1 - \theta B) \varepsilon_t,$$

e suponha que  $\theta = 0,8$ . Então, podemos escrever, com  $W_t = \Delta X_t$ ,

$$\varepsilon_t = W_t + 0,8 \varepsilon_{t-1}.$$

Suponha que utilizemos os dados (hipotéticos) da Tabela 3.6. Como

$$\varepsilon_1 = W_1 + 0,8 \varepsilon_0,$$

iniciamos  $\varepsilon_t$  especificando  $\varepsilon_0 = 0$  e  $X_0 = 150$ . Então,

$$\begin{aligned} \varepsilon_1 &= -3 + (0,8) \times 0 = -3, \\ \varepsilon_2 &= W_2 + 0,8 \varepsilon_1 = -4 + (0,8)(-3) = -6,4, \text{ etc.} \end{aligned}$$



Tabela 3.6: Cálculo recursivo de  $\varepsilon_t$ ,  $\theta = 0,8$ 

$t$	$X_t$	$W_t = \Delta X_t$	$\varepsilon_t = W_t + 0,8\varepsilon_{t-1}$
0	150		0
1	147	-3	-3,0
2	143	-4	-6,4
3	148	5	-0,12
4	153	5	4,9
5	149	-4	-0,08
6	155	6	5,9
7	162	7	11,7
8	170	8	17,4
9	172	2	15,9

Segue-se que a SQ condicional fica

$$S(0,8|a_0=0) = \sum_{t=1}^9 \varepsilon_t^2(0,8|a_0=0) = 801,26.$$

Calculando-se  $S$  para uma grade suficientemente fina de valores de  $\theta$  no intervalo  $(-1,1)$ , já que  $-1 < \theta < 1$ , podemos obter o mínimo aproximado dessa função.

Podemos obter estimadores de MV incondicionais; veja Morettin e Toloi (2004). Estimadores de MV *exatos* podem também ser calculados, tratando  $\varepsilon_0, \varepsilon_{-1}, \dots$  como parâmetros adicionais a serem estimados. Para  $T$  suficientemente grande as várias abordagens fornecem resultados próximos. Veja Box, Jenkins e Reinsel (1994) para detalhes.

### 3.3.3 Modelos Mistos

Estimadores condicionais para modelos ARMA (ou ARIMA) podem ser obtidos como na seção anterior. Por exemplo, considerando-se o modelo ARMA(1,1)

$$X_t - \phi X_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1},$$

condicionando-se a  $\varepsilon_0 = 0$  e  $X_0 = x_0$  obtemos, recursivamente,

$$\begin{aligned}\varepsilon_1 &= X_1 - \phi x_0, \\ \varepsilon_2 &= X_2 - \phi X_1 + \theta \varepsilon_1 = X_2 - (\phi - \theta)X_1 - \phi \theta x_0, \text{ etc}\end{aligned}$$

Também, estimadores de MV incondicionais e exatos podem ser obtidos para modelos ARIMA genéricos.

### 3.3.4 Variâncias dos estimadores

Para se ter uma idéia da precisão dos estimadores encontrados, devemos construir intervalos de confiança para os parâmetros. Seja  $\boldsymbol{\eta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$ , de ordem  $k \times 1$ , onde  $k = p + q$ . Para  $n$  grande, onde  $n = T - d$  ( $d$  é a ordem de diferenças) os EMV têm uma distribuição assintótica normal, de modo que podemos escrever

$$\begin{aligned}\hat{\boldsymbol{\eta}} &\xrightarrow{\mathcal{D}} \mathcal{N}_k(\boldsymbol{\eta}, \mathbf{V}), \\ \mathbf{V} &= 2\sigma^2 \begin{bmatrix} \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1^2} & \cdots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1 \partial \eta_k} \\ \vdots & & \vdots \\ \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k \partial \eta_1} & \cdots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k^2} \end{bmatrix},\end{aligned}\quad (3.23)$$

onde  $S(\boldsymbol{\eta})$  é a soma de quadrados que aparece em (3.20), por exemplo.

Pode-se também provar que o EMV de  $\sigma^2$  é

$$\hat{\sigma}^2 = \frac{S(\hat{\boldsymbol{\eta}})}{n}$$

e que, para  $n$  grande,  $\hat{\sigma}^2$  e  $\hat{\boldsymbol{\eta}}$  são não-correlacionados. Substituindo  $\sigma^2$  em (3.23) por  $\hat{\sigma}^2$  e calculando as derivadas  $\frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j}$  numericamente, obtemos estimativas das variâncias dos estimadores e estimativas das covariâncias entre os estimadores. A partir das estimativas das variâncias podemos obter intervalos de confiança para os parâmetros  $\eta_i$ ,  $i = 1, \dots, k$ .

Para os modelos mais comuns a Tabela 3.7 mostra as variâncias aproximadas dos estimadores.

Tabela 3.7: Variâncias aproximadas para os estimadores dos parâmetros dos modelos usuais;  $n = T - d$ .

Modelo	Variância
AR(1)	$\text{Var}(\hat{\phi}) \simeq \frac{1 - \phi^2}{n}$
AR(2)	$\text{Var}(\hat{\phi}_1) = \text{Var}(\hat{\phi}_2) \simeq \frac{1 - \phi_2^2}{n}$
MA(1)	$\text{Var}(\hat{\theta}) \simeq \frac{1 - \theta^2}{n}$
MA(2)	$\text{Var}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_2) \simeq \frac{1 - \theta_2^2}{n}$
ARMA(1,1)	$\text{Var}(\hat{\phi}) \simeq \frac{(1 - \phi^2)}{n} \frac{(1 - \phi\theta)^2}{(\phi - \theta)^2}$
	$\text{Var}(\hat{\theta}) \simeq \frac{(1 - \theta^2)}{n} \frac{(1 - \phi\theta)^2}{(\phi - \theta)^2}$

**Exemplo 3.8.** A estimação dos parâmetros das séries simuladas nos exemplo 2.3, 2.4 e 2.6, utilizando o programa Minitab, resulta:

(a) Modelo AR(1)

$$\begin{aligned}\hat{\phi} &= 0,8286, \\ \hat{\sigma}(\hat{\phi}) &= 0,0812, \\ P &= 0,0000.\end{aligned}$$

(b) Modelo MA(1)

$$\begin{aligned}\hat{\theta} &= 0,8241, \\ \hat{\sigma}(\hat{\theta}) &= 0,0804, \\ P &= 0,0000.\end{aligned}$$

(c) Modelo ARMA(1,1)

$$\begin{aligned}\hat{\phi} &= 0,9185, \quad \hat{\sigma}(\hat{\phi}) = 0,0714 \text{ e } P = 0,000, \\ \hat{\theta} &= 0,3911, \quad \hat{\sigma}(\hat{\theta}) = 0,1626 \text{ e } P = 0,020.\end{aligned}$$

Aqui,  $P$  é o p-valor, ou probabilidade de significância ou ainda nível descritivo associado a cada teste. Lembremos que um valor pequeno de  $P$  (usualmente menor do que 0,05) significa que devemos rejeitar a hipótese nula de

que o parâmetro correspondente é nulo.

**Exemplo 3.9.** Retornemos à série ICV, para a qual identificamos o modelo preliminar (3.14), ARIMA(1,1,0) com  $\theta_0$ , para a série  $\ln$  (ICV).

O programa Minitab forneceu as seguintes estimativas de máxima verossimilhança condicional

$$\begin{aligned}\hat{\phi} &= 0,5119, \quad \hat{\sigma}(\hat{\phi}) = 0,0833, \quad P = 0,000, \\ \hat{\theta}_0 &= 0,01036, \quad \hat{\sigma}(\hat{\theta}_0) = 0,0009, \quad P = 0,000.\end{aligned}$$

**Exemplo 3.10.** Para a série de retornos da Petrobrás, do exemplo 3.6, o programa EViews forneceu as seguintes estimativas:

$$\begin{aligned}\hat{\phi}_1 &= 0,09821, \quad \hat{\sigma}(\hat{\phi}_1) = 0,0256, \quad P = 0,0001, \\ \hat{\phi}_3 &= -0,05099, \quad \hat{\sigma}(\hat{\phi}_3) = 0,0255, \quad P = 0,046, \\ \hat{\phi}_9 &= 0,08017, \quad \hat{\sigma}(\hat{\phi}_9) = 0,0254, \quad P = 0,0016.\end{aligned}$$

### 3.4 Diagnóstico

Após estimar o modelo temos que verificar se ele representa, ou não, adequadamente, os dados. Qualquer insuficiência revelada pode sugerir um modelo alternativo como sendo adequado.

Uma técnica que pode ser utilizada, se suspeitarmos que um modelo mais elaborado (contendo mais parâmetros) é necessário, é o *superajustamento*. Estimamos um modelo com parâmetros extras e examinamos se estes são significativos e se sua inclusão diminui significativamente a variância residual. Este método é útil quando sabemos a priori em que direção pode estar ocorrendo a inadequação do modelo.

A verificação pode ser feita analisando os resíduos. Suponha que o modelo proposto seja

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

com

$$W_t = \Delta^d X_t.$$

Se este modelo for verdadeiro, então os “erros verdadeiros”  $\varepsilon_t$  constituirão um ruído branco.

### 3.4.1 Teste de auto-correlação residual

Estimados  $\phi$  e  $\theta$ , as quantidades

$$\hat{\varepsilon}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)W_t$$

são chamadas *resíduos estimados* ou simplesmente *resíduos*. Se o modelo for adequado, os  $\hat{\varepsilon}_t$  deverão estar próximos dos  $\varepsilon_t$  e, portanto, deverão ser aproximadamente não-correlacionados. Se indicarmos por  $\hat{r}_k$  as auto-correlações dos resíduos  $\hat{\varepsilon}_t$ , então deveríamos ter  $\hat{r}_k \simeq 0$ . Em particular, deveríamos ter, aproximadamente,

$$\hat{r}_k \sim \mathcal{N}(0, \frac{1}{n}),$$

onde  $n = T - d$  é o número efetivo de observações e sempre sob a suposição que o modelo ajustado seja apropriado. As auto-correlações  $\hat{r}_k$  são calculadas por

$$\hat{r}_k = \frac{\sum_{t=k+1}^n \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^n \hat{\varepsilon}_t^2}.$$

Contudo, o desvio padrão de  $\hat{r}_k$  pode ser consideravelmente menor que  $1/\sqrt{n}$ , especialmente para pequenos valores de  $k$ , como mostrou Durbin (1970). Ele provou que para um AR(1),  $\text{Var}(\hat{r}_k) \simeq \phi^2/n$ , que pode ser bem menor que  $1/n$ . Box, Jenkins e Reinsel (1994) provaram que, para um modelo AR(1), tem-se

$$\begin{aligned} \text{Var}(\hat{r}_k) &\simeq \frac{1}{n}[1 - \phi^{2(k-1)}(1 - \phi^2)] \\ \text{Cov}(\hat{r}_i, \hat{r}_j) &\simeq \frac{1}{n}\{\delta_{ij} - \phi^{i+j-2}(1 - \phi^2)\}, \end{aligned}$$

onde  $\delta_{ij}$  é o delta de Kronecker. Daqui, temos que, para  $k$  grande ou moderado, a variância de  $\hat{r}_k$  é, aproximadamente,  $1/n$ , e as auto-correlações são não-correlacionadas.

De qualquer modo, a comparação de  $\hat{r}_k$  com os limites  $\pm 2/\sqrt{n}$  fornece uma indicação geral de possível quebra de comportamento de ruído branco em  $\varepsilon_t$ , com a condição de que seja lembrado que, para pequenos valores de  $k$ , estes limites subestimarão a significância de qualquer discrepância.

### 3.4.2 Teste de Box-Pierce-Ljung

Box e Pierce (1970) sugeriram um teste para as auto-correlações dos resíduos estimados, que, apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos. Uma modificação deste teste foi proposta por Ljung e Box (1978), que é apresentada a seguir.

Se o modelo for apropriado, a estatística

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{(n-k)} \quad (3.24)$$

terá uma distribuição  $\chi^2$  com  $K - p - q$  graus de liberdade. A hipótese de ruído branco para os resíduos é rejeitada para valores grandes de  $Q(\cdot)$ . Em geral basta utilizar as 15 ou 20 primeiras  $\hat{r}_k$ .

**Exemplo 3.11.** Vamos agora testar se o modelo ARIMA(1,1,0) com  $\theta_0$ , proposto para a série  $\ln(ICV)$ , é adequado (veja os exemplos 3.4 e 3.9).

O Quadro 3.1 apresenta os valores dos parâmetros estimados e os respectivos valores da estatística  $t$ , além dos níveis descritivos ( $p$ -valores). Temos, também, os valores das estatísticas modificadas de Box-Pierce-Ljung. A Figura 3.7 apresenta as f.a.c. e f.a.c.p dos resíduos com os correspondentes intervalos de confiança. Analisando os valores dos  $p$ -valores, observamos que todos os parâmetros são significantemente diferentes de zero. Os valores das auto-correlações e auto-correlações parciais não indicam nenhuma quebra de comportamento de ruído branco nos resíduos. Os valores da estatística de Box-Pierce-Ljung, para diversos valores de  $K$ , também são indicadores do bom ajustamento do modelo ARIMA(1,1,0) com  $\theta_0$ .

Assim, o modelo ajustado à série  $\ln(ICV)$  é dada por

$$(1 - 0,5119B)(1 - B) \ln(ICV) = 0,01036 + \varepsilon_t, \quad (3.25)$$

com  $\hat{\sigma}^2 = 0,0000923$ .

**Exemplo 3.12.** Retomemos o exemplo 3.10, no qual obtivemos as estimativas dos parâmetros do modelo AR(9) ajustado à série de retornos da Petrobrás. Uma estimativa da variância residual é  $\hat{\sigma}^2 = 0,0011$ . Na Figura 3.8 temos os resíduos, f.a.c. e f.a.c.p. amostrais dos resíduos, mostrando que podemos aceitar a hipótese de que estes sejam um ruído branco. Além disso, a estatística  $Q$ , para diversos valores de  $K$ , por exemplo, 10 e 15, indicam o mesmo resultado.

---

```

ARIMA model for lnicv

Final Estimates of Parameters
Type      Coef      StDev      T      P
AR 1      0.5119      0.0833      6.14    0.000
Constant 0.0103600  0.0009040  11.46    0.000

Differencing: 1 regular difference
Number of observations: Original series 114, after differencing 113
Residuals:  SS = 0.0102405 (backforecasts excluded)
              MS = 0.0000923  DF = 111

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag        12        24        36        48
Chi-Square 10.5      34.1      43.3      67.1
DF          10       22       34       46
P-Value     0.398    0.048    0.132    0.023

Correlation matrix of the estimated parameters
              1
2  0.031

```

---

Quadro 3.1 - Ajustamento de um modelo ARIMA(1,1,0) à série  $\ln(ICV)$ .

No Capítulo 5 veremos que esta série apresenta, na realidade, uma variância condicional não-constante, de modo que será necessário ajustar um modelo da classe ARCH-GARCH para levar esse fato em conta. O gráfico da série, Figura 3.6 (a) mostra grupos de retornos com variabilidade (volatilidade) grande e outros com variabilidade menor.

### 3.5 Previsão com Modelos ARIMA

Queremos prever  $X_{T+h}$ , tendo-se observações até o instante  $T$ , usando um modelo ARIMA. Seja

$$\varphi(B) = \phi(B)\Delta^d = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_{p+d} B^{p+d})$$

Chamemos de  $\hat{X}_T(h)$  a previsão de  $X_{T+h}$  de origem  $T$  e horizonte  $h$ . Pode-se provar que a previsão de EQMM (erro quadrático médio mínimo) é dada pela esperança condicional de  $X_{T+h}$  dado o passado  $X_T, X_{T-1}, \dots$ , ou seja,

$$\begin{aligned} \hat{X}_T(h) = & E(\varphi_1 X_{T+h-1} + \dots + \varphi_{p+d} X_{T+h-p-d} \\ & + \theta_0 + a_{T+h} - \theta_1 a_{T+h-1} - \dots - \theta_q a_{T+h-q} \mid X_T, X_{T-1}, \dots) \end{aligned}$$

Para calcular as previsões usamos os fatos:

$$(a) \ E(X_{T+j} \mid X_T, X_{T-1}, \dots) = \begin{cases} X_{T+j}, & \text{se } j \leq 0 \\ \hat{X}_T(j), & \text{se } j > 0 \end{cases}$$

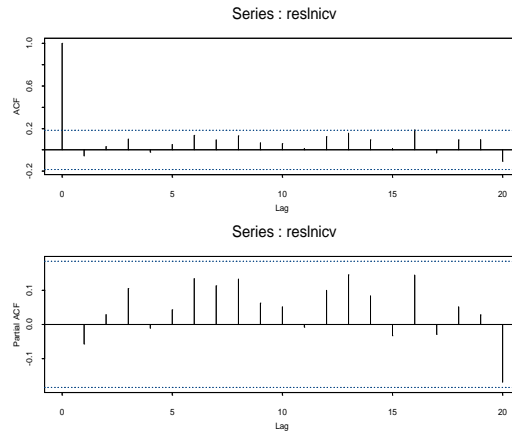


Figura 3.7: F.a.c. e f.a.c.p. dos resíduos do modelo ARIMA(1,1,0) ajustado à série  $\ln(ICV)$ .



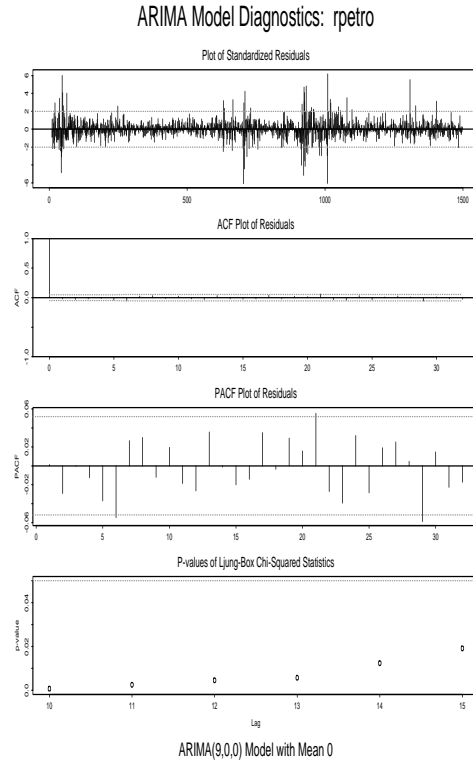


Figura 3.8: Diagnóstico do modelo do exemplo 3.10.

$$(b) \ E(\varepsilon_{T+j}|X_T, X_{T-1}, \dots) = \begin{cases} \varepsilon_{T+j}, & \text{se } j \leq 0 \\ 0, & \text{se } j > 0 \end{cases}$$

Logo, para calcular previsões temos que:

- (a) substituir esperanças passadas ( $j \leq 0$ ) por valores conhecidos,  $X_{T+j}$  e  $\varepsilon_{T+j}$ ;
- (b) substituir esperanças futuras ( $j > 0$ ) por previsões  $\hat{X}_T(j)$  e 0.

**Exemplo 3.13.** Suponha o modelo AR(2),

$$(1 - \phi_1 B - \phi_2 B^2)X_t = \phi_0 + \varepsilon_t.$$

Temos que

$$X_{T+h} = \phi_1 X_{T+h-1} + \phi_2 X_{T+h-2} + \phi_0 + \varepsilon_{T+h},$$

logo:

- (i) Para  $h = 1$ , temos  $\hat{X}_T(1) = \phi_1 X_T + \phi_2 X_{T-1} + \phi_0$ ;
- (ii) para  $h = 2$ , temos  $\hat{X}_T(2) = \phi_1 \hat{X}_T(1) + \phi_2 X_T + \phi_0$ ;
- (iii) para  $h > 2$ , temos  $\hat{X}_T(h) = \phi_1 \hat{X}_T(h-1) + \phi_2 \hat{X}_T(h-2) + \phi_0$ .

Escrevendo-se o modelo na forma de médias móveis (infinito), pode-se provar que o *erro de previsão* é dado por

$$e_T(h) = X_{T+h} - \hat{X}_T(h) = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \dots + \psi_{h-1} \varepsilon_{T+1}, \quad (3.26)$$

em que os pesos  $\psi_j$  vêm de  $\psi(B) = \varphi^{-1}(B)\theta(B)$ .

A variância do erro de previsão é dada por

$$V(h) = \text{Var}(e_T(h)) = \sigma^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2). \quad (3.27)$$

Observe que, como  $e_T(1) = X_{T+1} - \hat{X}_T(1) = \varepsilon_{T+1}$ , *os erros de previsões a um passo são não-correlacionados*.

No desenvolvimento acima supusemos os parâmetros do modelo conhecidos. Na prática, usamos o modelo estimado para fazer as previsões. Neste caso, (3.27) fica

$$\hat{V}(h) = \hat{\sigma}^2(1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2), \quad (3.28)$$

onde os  $\hat{\psi}_j$  são obtidos de  $\hat{\psi}(B) = (\hat{\varphi}(B))^{-1}\hat{\theta}(B)$ .

**Exemplo 3.14.** Considere o modelo MA(1)

$$X_t = \theta_0 + \varepsilon_t - \theta\varepsilon_{t-1},$$

onde  $\theta_0 = E(X_t)$ .

Para obtermos as previsões com origem  $T$  e  $h = 1$ , escrevamos o modelo na forma

$$X_{T+1} = \theta_0 + \varepsilon_{T+1} - \theta\varepsilon_T.$$

Tomando-se a esperança condicional ao passado, temos

$$\hat{X}_T(1) = \theta_0 + E(\varepsilon_{T+1}|X_T, X_{T-1}, \dots) - \theta E(\varepsilon_T|X_T, \dots),$$

ou seja,

$$\hat{X}_T(1) = \theta_0 - \theta\varepsilon_T.$$

O valor de  $\varepsilon_T$  é calculado recursivamente, como vimos na seção 3.3.2. O erro de previsão um passo à frente é dado por

$$e_T(1) = X_{T+1} - \hat{X}_T(1) = \varepsilon_{T+1}$$

e, portanto,  $\text{Var}(e_T(1)) = \sigma^2$ . Esta variância é estimada por  $\hat{\sigma}^2$ .

Para  $h = 2$ , de

$$X_{T+2} = \theta_0 + \varepsilon_{T+2} - \theta\varepsilon_{T+1},$$

obtemos

$$\hat{X}_T(2) = E(X_{T+2}|X_T, \dots) = \theta_0,$$

de modo que

$$e_T(2) = \varepsilon_{T+2} - \theta\varepsilon_{T+1}.$$

A variância do erro de previsão a dois passos será, então,

$$\text{Var}(e_T(2)) = (1 + \theta^2)\sigma^2,$$

que é maior do que a variância do erro de previsão a um passo.

Vemos que a previsão a dois passos é igual à média  $\theta_0$  do processo. De modo geral,  $\hat{X}_T(h) = \theta_0$ , para  $h \geq 2$ . Para um modelo MA(q), a previsão  $\hat{X}_T(h)$  será igual à média do processo após  $q$  passos.

**Exemplo 3.15.** Consideremos o  $\ln(ICV)$ , para o qual ajustamos um modelo ARIMA(1, 1, 0) com  $\theta_0$ ,

$$(1 - 0,5119B)(1 - B)\ln ICV = 0,01036 + \varepsilon_t$$

ou

$$(1 - 1,5119B + 0,5119B^2)Y_t = 0,01036 + \varepsilon_t.$$

Assim, para a origem de previsão  $T$ ,

$$Y_{T+h} = 1,5119Y_{T+h-1} - 0,5119Y_{T+h-2} + 0,01036 + \varepsilon_{T+h}$$

e

$$\begin{aligned}\hat{Y}_T(h) &= 1,5119E(Y_{T+h-1}|Y_T, Y_{T-1} \dots) - 0,5119E(Y_{T+h-2}|Y_T, Y_{T-1} \dots) \\ &\quad + 0,01036 + E(\varepsilon_{T+h}|Y_T, Y_{T-1} \dots).\end{aligned}$$

Segue-se que as previsões de origem  $T$  são dadas por

$$\hat{Y}_T(h) = 1,5119\hat{Y}_T(h-1) - 0,5119\hat{Y}_T(h-2) + 0,01036, \quad h \geq 1.$$

Na Tabela 3.8 temos as previsões feitas a partir da origem  $T = 114$  (junho de 1979) para  $h = 1, 2, \dots, 12$  (isto é, previsões para os restantes 6 meses de 1979 e os primeiros 6 meses de 1980).

Sabemos como calcular a variância do erro de previsão, que é dada por (3.27). Para podermos determinar um intervalo de confiança para  $X_{t+h}$ , será necessário fazer uma suposição adicional para os resíduos, ou seja, além de supor que  $E(\varepsilon_t) = 0$ ,  $\text{Var}(\varepsilon_t) = \sigma^2$  para todo  $t$  e  $E(\varepsilon_t \varepsilon_s) = 0$ ,  $t \neq s$ , iremos supor que  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ , para cada  $t$ .

Segue-se que, dados os valores passados e presente da série,  $X_T, X_{T-1}, \dots$ , a distribuição condicional de  $X_{T+h}$  será  $\mathcal{N}(\hat{X}_T(h), V(h))$ .

Logo,

$$U = \frac{X_{T+h} - \hat{X}_T(h)}{[V(h)]^{1/2}} \sim \mathcal{N}(0, 1)$$

e fixado o coeficiente de confiança  $\gamma$ , podemos encontrar um valor  $u_\gamma$  tal que  $P(-u_\gamma < U < u_\gamma) = \gamma$ . Ou seja, com probabilidade  $\gamma$ ,

$$\hat{X}_T(h) - u_\gamma[V(h)]^{1/2} \leq X_{T+h} \leq \hat{X}_T(h) + u_\gamma[V(h)]^{1/2}.$$

Substituindo  $V(h)$  por  $\hat{V}(h)$  obtemos

$$\hat{X}_T(h) - u_\gamma \hat{\sigma} \left[ 1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2 \right]^{1/2} \leq X_{T+h} \leq \hat{X}_T(h) + u_\gamma \hat{\sigma} \left[ 1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2 \right]^{1/2}. \quad (3.29)$$

**Exemplo 3.16.** Podemos determinar o intervalo de confiança para  $Y_{T+h}$  no Exemplo 3.15, para cada valor de  $h$ .

Assim, para  $h = 3$ ,

$$\hat{V}(3) = (1 + \psi_1^2 + \psi_2^2) \hat{\sigma}^2, \text{ onde } \hat{\sigma}^2 = 0,0000923.$$

Os pesos  $\hat{\psi}_j, j = 1, 2$  neste caso valem 1,5119 e 1,7739 respectivamente, logo

$$\hat{V}(3) = (1 + (1,5119)^2 + (1,7739)^2) \cdot 0,0000923 = 0,000594.$$

Para  $\gamma = 95\%$ , temos  $u_\gamma = 1,96$ , portanto (3.29) fica

$$\hat{Y}_T(3) - 1,96[\hat{V}(3)]^{1/2} \leq Y_{T+3} \leq \hat{Y}_T(3) + 1,96[\hat{V}(3)]^{1/2}$$

ou seja, o intervalo de confiança para  $Y_{117}$  é  $[6,6915; 6,7871]$ .

Da mesma maneira podem ser obtidos intervalos de confiança para  $X_{T+1}$ ,  $X_{T+2}$ , etc. Observe que a variância aumenta com  $h$ , logo as amplitudes destes intervalos aumentarão à medida que nos afastamos da origem  $t$ , caracterizando o aumento da incerteza das previsões para  $h$  passos à frente,  $h$  grande.

Obtidas as previsões e intervalos de confiança para  $Y_t$ , podemos obter previsões e intervalos de confiança para a série original  $X_t = \exp\{Y_t\}$ .

### 3.6 Modelos Sazonais

Tradicionalmente, procedimentos com médias móveis e regressão são usados para estimar a componente sazonal de uma série temporal. Não iremos nos deter nesse assunto aqui. Veja Morettin e Toloi (2006) para detalhes.

Tabela 3.8: Previsões para o logaritmo da série ICV, com origem  $T = 114$  (junho de 1979), usando o modelo ARIMA(1,1,0) com  $\hat{\theta}_0 = 0,01036$  e  $\hat{\phi} = 0,5119$ .

$t$	$Y_t$	$\hat{Y}_T(h)$
112	6,5944	
113	6,6147	
114	6,6567	
115	6,6995	6,6886
116	6,7334	6,7153
117	6,7441	6,7393
118	6,8416	6,7619
119	6,8876	6,7839
120	6,9556	6,8057
121	6,9994	6,8269
122	7,0326	6,8482
123	7,0750	6,8695
124	7,1204	6,8907
125	7,1770	6,9120
126	7,2255	6,9332

**Exemplo 3.17.** A Figura 3.9 (a) mostra a série de índices mensais de produção industrial do Brasil (IPI, arquivo m-ipi85.00.dat), setor de alimentação, no período de janeiro de 1985 a julho de 2000,  $T = 187$ . Observamos que a série apresenta uma componente sazonal de período 12 meses, além de uma tendência suave, mas crescente. Na Figura 3.9 (b) temos a f.a.c. da série, que denominaremos  $X_t$ , que reflete a aparente periodicidade existente na série.

Tomando-se uma diferença obtemos a série da Figura 3.9 (c), cuja f.a.c. está na Figura 3.9 (d). Notamos que a f.a.c. ainda mostra a forma típica da

de uma série com componente sazonal, com valores grandes nos “lags” 6, 12 e 18. Seja  $Y_t = \Delta X_t$ .

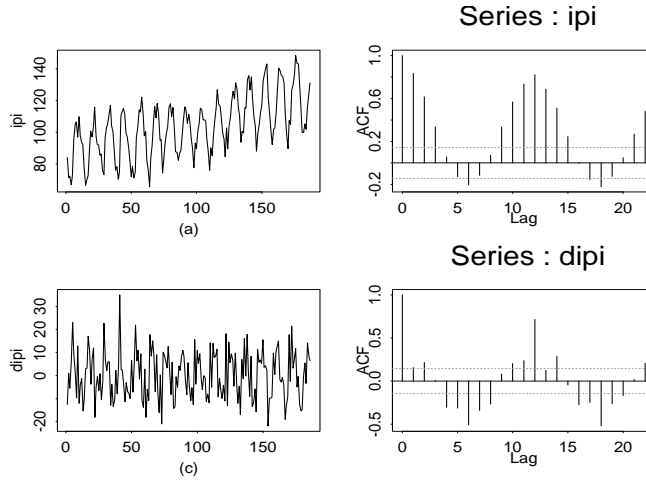


Figura 3.9: (a) Série IPI (b) f.a.c. da série (c) Primeira diferença da série IPI (d) f.a.c. da primeira diferença

Para obtermos uma série livre de componente sazonal tomamos uma *diferença sazonal de ordem 12*,

$$Z_t = \Delta_{12}Y_t = (1 - B^{12})Y_t, \quad (3.30)$$

ou seja,

$$Z_t = Y_t - Y_{t-12}.$$

Segue-se que podemos escrever

$$Z_t = (1 - B)(1 - B^{12})X_t. \quad (3.31)$$

Note que esta série tem 12 observações a menos que a série  $Y_t$ , que por sua vez tem uma observação a menos do que a série  $X_t$ . Logo, perdemos, no total, 13 observações neste processo de tomar primeiro, uma diferença simples (ou regular) e, depois, uma diferença sazonal (de ordem 12, no caso).

Na Figura 3.10 temos a f.a.c. e f.a.c.p. da série  $Z_t$ .

Uma análise da f.a.c. e f.a.c.p. de  $Z_t$  pode nos dar uma sugestão do modelo final para a série. Todavia, como já salientamos, estas funções têm, em geral, uma forma complicada, e pode ser difícil identificarmos um modelo a partir delas.

Um modelo sugerido por Box, Jenkins e Reinsel (1994) e bastante utilizado em aplicações práticas é o chamado “airline model”, que tem a forma

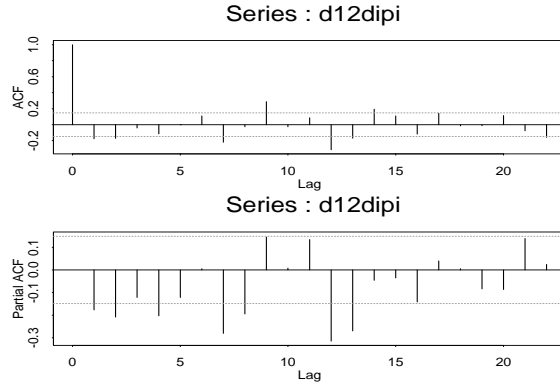


Figura 3.10: Série  $Z_t = \Delta\Delta_{12}X_t$  (a) f.a.c. (b) f.a.c.p.

$$(1 - B)(1 - B^s)X_t = (1 - \theta B)(1 - \Theta B^s)\varepsilon_t, \quad (3.32)$$

com  $|\theta| < 1$ ,  $|\Theta| < 1$ . Em (3.32),  $\theta$  é o parâmetro “regular” de médias móveis,  $\Theta$  é o parâmetro de médias móveis sazonal e  $s$  é o período sazonal.

O modelo (3.32) é um caso particular do chamado modelo ARIMA *sazonal multiplicativo* (SARIMA) de ordem  $(p, d, q) \times (P, D, Q)_s$ , dado por

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (3.33)$$

onde

$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  é o operador auto-regressivo estacionário de ordem  $p$ ;



$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  é o operador de médias móveis invertível de ordem  $q$ ;

$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{sP}$  é o operador auto-regressivo sazonal de ordem  $P$ , estacionário;

$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ}$  é o operador de médias móveis sazonal de ordem  $Q$ , invertível;

$\Delta = (1 - B)$  é o operador diferença;

$\Delta^d = (1 - B)^d$ ,  $d$  indicando o número de diferenças;

$\Delta_s = (1 - B^s)$  é o operador diferença sazonal;

$\Delta_s^D = (1 - B^s)^D$ ,  $D$  indicando o número de “diferenças sazonais”.

Estimando o modelo (3.32) para a série IPI obtemos o Quadro 3.2 e a Figura 3.11, esta contendo a f.a.c. e f.a.c.p. dos resíduos do modelo ajustado.

Todos os parâmetros são significativos, mas o comportamento da f.a.c. e f.a.c.p. mostram que o modelo é inadequado. Valores grandes de  $\hat{\phi}_{44}$  e  $\hat{\phi}_{77}$  sugerem a introdução de um polinômio auto-regressivo incompleto contendo os parâmetros  $\phi_4$  e  $\phi_7$ . Estimado este modelo, ainda há a necessidade de incluirmos um termo de ordem 5 neste polinômio AR, de modo que o modelo final proposto é

$$(1 - \phi_4 B^4 - \phi_5 B^5 - \phi_7 B^7)(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})\varepsilon_t, \quad (3.34)$$

com ajustamento e f.a.c. e f.a.c.p residual apresentados no Quadro 3.3 e Figura 3.12, respectivamente. Substituindo os valores estimados dos parâmetros em (3.34), temos que um modelo adequado à série IPI é dado por

$$\begin{aligned} & (1 + 0,2562B^4 + 0,1587B^5 + 0,2984B^7)(1 - B)(1 - B^{12})X_t \\ & = (1 - 0,5409B)(1 - 0,6584B^{12})\varepsilon_t, \end{aligned} \quad (3.35)$$

com  $\hat{\sigma}_a^2 = (5,51)^2$ .

As previsões para os meses de janeiro a julho de 2000, com origem em dezembro de 1999 ( $T = 180$ ), encontram-se na Tabela 3.9 e Figura 3.13. Na Tabela 3.9,  $EQMP_{180}$  denota o erro quadrático médio de previsão, com origem em  $T = 180$  e horizonte  $h$ , isto é,

$$EQMP_{180} = \frac{1}{7} \sum_{h=1}^7 (X_{T+h} - \hat{X}_T(h))^2.$$

Tabela 3.9: Previsões para a série IPI, utilizando o modelo (3.35), com origem em  $T = 180$  e  $h = 1, 2, \dots, 7$ .

$T + h$	$\hat{X}_T(h)$	Erro padrão	$X_{T+h}$
181	106,4046	5,5106	100,1300
182	98,2554	6,0637	99,9000
183	109,5588	6,5704	105,3800
184	109,7577	7,0407	101,9600
185	125,3837	7,1290	116,1900
186	131,6298	7,1998	124,6600
187	143,4999	7,3504	131,1000
$EQMP_{180} = 58,17$			

### 3.7 Problemas

1. Para um modelo AR(1), prove que, se  $\rho_j = \phi^{|j|}$ ,  $|\phi| < 1$ , então

$$\text{Var}(r_j) = \frac{1}{T} \left[ \frac{(1 + \phi^2)(1 - \phi^{2j})}{1 - \phi^2} - 2j\phi^{2j} \right];$$

em particular,  $\text{Var}(r_1) = \frac{1}{T}(1 - \phi^2)$ .

---

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	VALUE	STD ERROR	T VALUE
1	IPI	MA	1	1	.3842	.0725	5.30
2	IPI	MA	2	12	.6361	.0614	10.36
EFFECTIVE NUMBER OF OBSERVATIONS . .					167		
R-SQUARE . . . . .					0.895		
RESIDUAL STANDARD ERROR. . . . .					0.595373E+01		

---

Quadro 3.2: Estimação do modelo (3.32) para a série IPI

---

PARAMETER LABEL	VARIABLE NAME	NUM./ DENOM.	FACTOR	ORDER	VALUE	STD ERROR	T VALUE
1	IPI	MA	1	1	.5409	.0669	8.09
2	IPI	MA	2	12	.6584	.0586	11.23
3	IPI	AR	1	4	-.2562	.0729	-3.52
4	IPI	AR	1	5	-.1587	.0737	-2.15
5	IPI	AR	1	7	-.2984	.0727	-4.10
EFFECTIVE NUMBER OF OBSERVATIONS . .					160		
R-SQUARE . . . . .					0.910		
RESIDUAL STANDARD ERROR. . . . .					0.551065E+01		

---

Quadro 3.3: Ajustamento do modelo (3.34) à série IPI

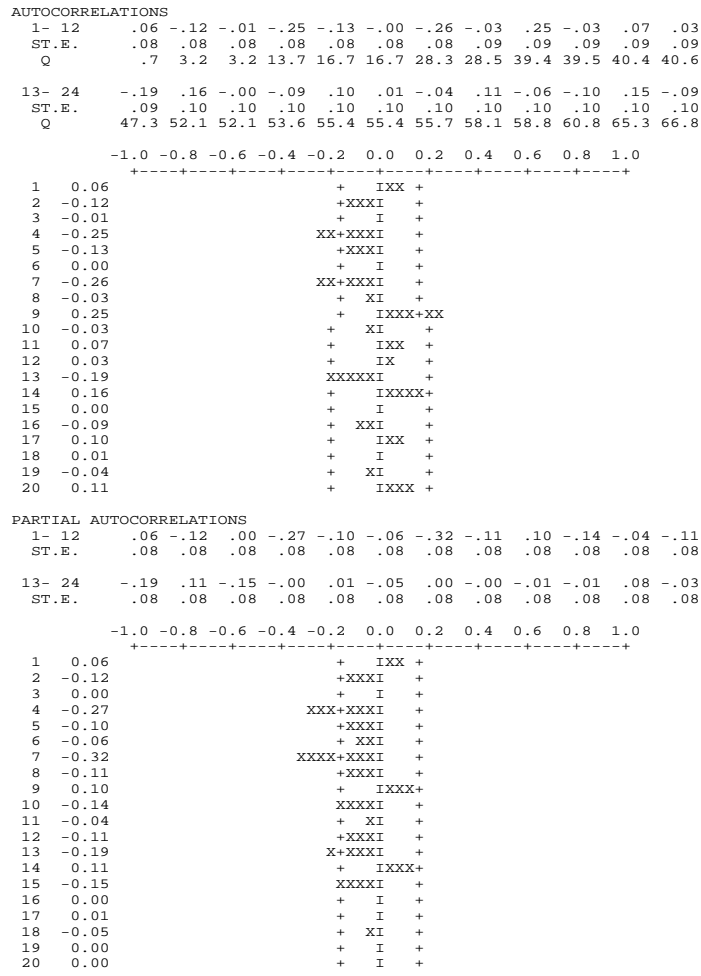


Figura 3.11: F.a.c. e f.a.c.p. dos resíduos do modelo (3.32)

2. Prove que  $\text{Var}(r_j)$ , dada no Problema 1, converge para  $\frac{1}{T} \left[ \frac{1+\phi^2}{1-\phi^2} \right]$  quando  $j \rightarrow +\infty$ , se  $|\phi| \ll 1$ .

3. Prove que, para um processo AR(1),  $\text{Var}(\bar{W}) = \frac{c_0(1+r_1)}{n(1-r_1)}$ .

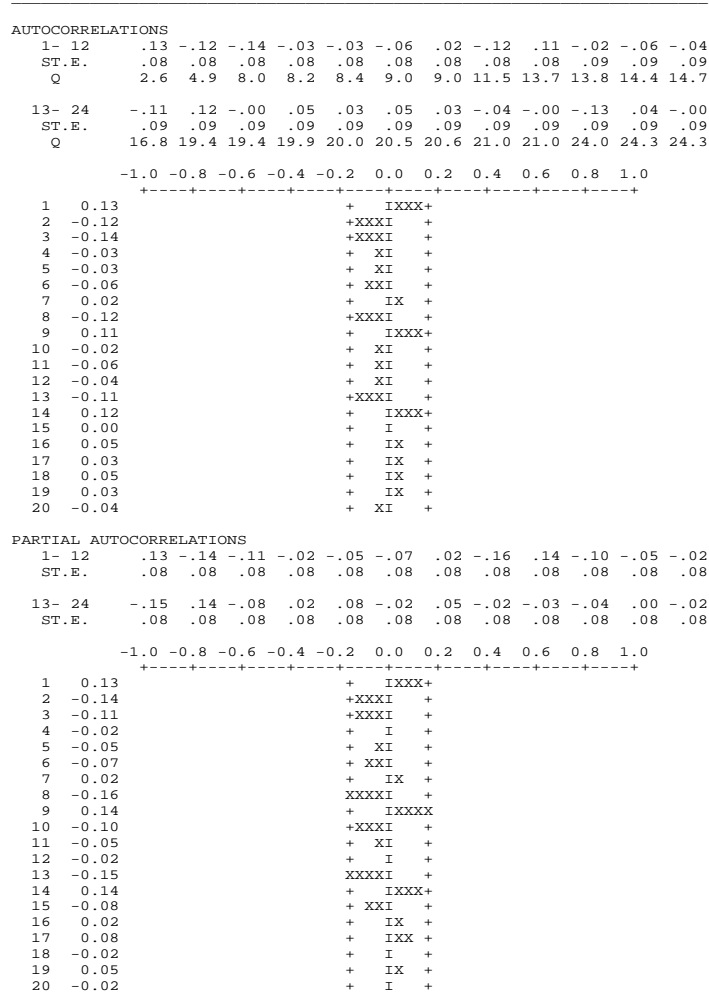


Figura 3.12: F.a.c e f.a.c.p dos resíduos do modelo (3.35)

4. Suponha que um programa de identificação forneceu os seguintes resultados:

$j$	1	2	3	4	5	6
$r_j$	-0,82	0,41	-0,12	0,08	-0,09	0,05
$\hat{\phi}_{jj}$	-0,82	-0,43	-0,05	0,25	0,20	0,12

$T = 100$ ,  $\bar{X} = -0,08$ ,  $S_X^2 = 2,40$ . Identifique um modelo para  $X_t$ .

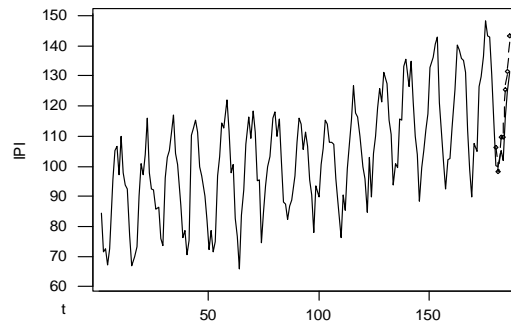


Figura 3.13: Série IPI com previsões utilizando o modelo (3.35), com origem em dezembro de 1999 e  $h = 1, 2, \dots, 7$

5. Considere o modelo  $(1 - \phi B)X_t = (1 - \theta B)\varepsilon_t$ . Mostre que:
- $X_t = (1 + \phi B)(1 - \theta B)\varepsilon_t$ , se  $\phi$  pequeno;
  - $X_t = \{1 + (\phi - \theta)B\}\varepsilon_t$ , se  $\phi$  e  $\theta$  são pequenos.
6. Suponha que a f.a.c. amostral de uma série com  $T = 100$  observações seja dada por

$j$	1	2	3	4	5	6	7	8	9	10
$r_j$	0,61	0,37	-0,05	0,06	-0,21	0,11	0,08	0,05	0,12	-0,01

Sugira um modelo ARMA que seja apropriado.

7. Suponha que 100 observações de uma série temporal forneçam as seguintes estimativas:  $c_0 = 250$ ,  $r_1 = 0,8$ ,  $r_2 = 0,7$ ,  $r_3 = 0,5$ . Use as estimativas de Yule-Walker para determinar se a série é adequadamente ajustada por um modelo AR(1) ou por um modelo AR(2).
8. Uma série com 400 observações apresentou os seguintes resultados:

$j$	1	2	3	4	5	6	7
$\hat{\phi}_{jj}$	0,8	-0,5	0,07	-0,02	-0,01	0,05	0,04

$$\bar{X} = 8,0; c_0 = 9,0.$$

- (a) Explique porque podemos ajustar à série um modelo AR(2).
  - (b) Obtenha as estimativas  $\hat{\phi}_1$  e  $\hat{\phi}_2$  do modelo AR(2) utilizando as equações de Yule-Walker; obtenha também estimativas do termo constante  $\theta_0$  e da  $\text{Var}(a_t)$ .
  - (c) Verifique se o modelo ajustado satisfaz as condições de estacionariedade.
  - (d) Usando  $\hat{\phi}_1$  e  $\hat{\phi}_2$  como sendo os verdadeiros valores de  $\phi_1$  e  $\phi_2$  do processo AR(2), determine os valores de  $\rho_1$ ,  $\rho_2$  e  $\rho_3$ . Descreva, também, o comportamento geral da f.a.c. desse processo.
9. Identifique um modelo ARIMA para a série de log-retornos diários do Banespa (arquivo d-ban95.00.dat).
  10. Idem, para a série de log-retornos mensais do S&P500, de janeiro de 1962 a dezembro de 1999 (arquivo m-sp62.99.dat),  $T = 456$ .
  11. Construa os intervalos de confiança com coeficientes de confiança igual a 0,95, para os parâmetros  $\phi$  e  $\theta_0$  do Exemplo 3.9.
  12. Suponha que para um modelo ARMA(1,1), com  $T = 152$ , obtemos  $\hat{\phi} = 0,85$ ,  $\hat{\theta} = -0,6$ ,  $\hat{\sigma}^2 = 0,086$ . Obtenha intervalos de confiança para  $\phi$  e  $\theta$ , com coeficiente de confiança 0,95.

13. Suponha que os resíduos obtidos, ajustando-se o modelo  $\Delta X_t = (1 - 0,6B)b_t$  a uma série com  $T = 127$  observações, forneceram as seguintes auto-correlações:

$k$	1	2	3	4	5	6	7	8	9	10
$\hat{r}_k(b)$	-0,40	0,02	-0,07	-0,01	-0,07	-0,02	0,15	-0,07	0,04	0,02

- (a) Verifique se há valores anormais.  
 (b) Use o teste de Box-Pierce-Ljung para verificar se o modelo é adequado.  
 (c) Os resíduos sugerem que o modelo deva ser modificado? Em caso afirmativo, qual modelo deveria ser considerado?
14. Suponha que os resíduos  $\hat{\varepsilon}_t$  do modelo  $(1-B)X_t = (1+0,6B)\varepsilon_t$ , ajustado a uma série de 80 observações, forneceram as seguintes auto-correlações

$j$	1	2	3	4	5	6	7	8	9	10
$r_{\varepsilon}(j)$	0,39	0,20	0,09	0,04	0,09	-0,13	-0,05	0,06	0,11	-0,02

Analise se o modelo ajustado é adequado e se existe alguma indicação de falta de ajustamento do modelo. Se isto ocorreu, sugira um modelo modificado e teste-o.

15. Obtenha a função de previsão  $\{\hat{X}_t(h), h = 1, 2, \dots\}$  para os seguintes modelos:
- (a)  $\Delta^2 X_t = (1 - 0,9B + 0,5B^2)\varepsilon_t$  ;  
 (b)  $(1 - 1,8B + 0,8B^2)X_t = \varepsilon_t$ ;  
 (c)  $(1 - 0,6B)\Delta X_t = (1 - 0,5B)\varepsilon_t$ .
16. Considere o modelo  $X_t = 0,8X_{t-1} + \varepsilon_t$ , com  $\varepsilon_t \sim \mathcal{N}(0, 1)$ .
- (a) Obtenha  $\hat{X}_t(h)$ ,  $h = 1, 2, 3, 100$ .  
 (b) Calcule  $V(h)$ ,  $h = 1, 2, 3, 100$ .  
 (c) Suponha os dados:



$t$	1	2	3	4	5	6	7
$X_t$	0,66	0,57	0,66	-1,47	-1,38	-1,9	-0,7

Calcule  $\hat{X}_7(h)$ ,  $h = 1, 2, 3, 100$ .

(d) Obtenha intervalos de confiança para  $X_8$  e  $X_9$ .

17. Considere o modelo  $\text{SARIMA}(0, 1, 2) \times (0, 1, 1)_{12}$ :

$$\Delta\Delta_{12}Z_t = (1 - \Theta B^{12})(1 - \theta_1 B - \theta_2 B^2)\varepsilon_t.$$

(a) Escreva o modelo na forma de um modelo ARMA.

(b) Qual a ordem do modelo ARMA resultante?

(c) Obtenha a f.a.c. do modelo.

18. Para o modelo  $\text{SARIMA}(0, 1, 1) \times (0, 1, 1)_{12}$ :

(a) escreva-o explicitamente;

(b) obtenha a região de invertibilidade;

(c) obtenha as auto-correlações do processo.

19. Obtenha as estimativas dos parâmetros dos modelos identificados nos Problema 9 e 10.

20. Suponha que o modelo

$$(1 - B^4)X_t = 3,0 + \varepsilon_t + 1,0\varepsilon_{t-1} - 0,5\varepsilon_{t-4}, \quad \sigma = 2,25,$$

foi ajustado às observações de uma série de dados trimestrais.

(a) Suponha que as observações e resíduos dos últimos quatro trimestres são dadas por

	I	II	III	IV
$X_t$	124	121	129	139
$\varepsilon_t$	2	-1	1	3

Encontre as previsões  $\hat{X}_t(h)$ ,  $h = 1, 2, 3, 4$ .

(b) Determine os pesos  $\psi_j$ ,  $j \geq 0$ .

- (c) Calcule as variâncias dos erros de previsão  $e_t(h) = X_{t+h} - \hat{X}_t(h)$ ,  $h = 1, 2, 3$  e  $4$  e utilize-as para construir intervalos de confiança para os futuros valores  $X_{t+h}$ ,  $h = 1, 2, 3$  e  $4$ .
21. Ajuste um modelo SARIMA apropriado à série de Produção Física Industrial Geral de janeiro de 1985 a julho de 2000 (arquivo m-pfi85.00.dat).

# Raízes Unitárias

## 4.1 Introdução

Como vimos na seção 2.6, o problema de raiz unitária em modelos ARMA aparece quando o polinômio auto-regressivo apresenta uma raiz sobre o círculo unitário. Isto implica que devemos tomar uma diferença da série original antes de ajustar o modelo. Podemos ter raízes unitárias também no polinômio de médias móveis, como no exemplo 2.4. Isto pode indicar que os dados foram super-diferenciados. Veja a seção 4.4 para mais detalhes.

Neste capítulo vamos considerar testes para raízes unitárias em modelos AR e ARMA. Para efeito de ilustração consideremos o modelo AR(1) estacionário,

$$X_t = \theta_0 + \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{RB}(0, \sigma^2), \quad (4.1)$$

no qual  $\theta_0 = (1 - \phi)\mu$ ,  $\mu = E(X_t)$ ,  $|\phi| < 1$ . Se  $\hat{\phi}_{MV}$  indica o EMV de  $\phi$ , então sabemos que, para  $T$  observações do processo,

$$\hat{\phi}_{MV} \overset{a}{\sim} \mathcal{N}(\phi, (1 - \phi^2)/T). \quad (4.2)$$

Se quisermos testar a hipótese  $H_0 : \phi = \phi_0$  contra a alternativa  $H_1 : \phi \neq \phi_0$ , usamos a estatística

$$\frac{\hat{\phi}_{MV} - \phi_0}{\widehat{\text{e.p.}}(\hat{\phi}_{MV})}, \quad (4.3)$$

em que o denominador indica o erro padrão estimado de  $\hat{\phi}_{MV}$ . Sob a hipótese nula, a estatística (4.3) tem uma distribuição  $t$  de Student. Observe que (4.2) pode ser escrita

$$\sqrt{T}(\hat{\phi}_{MV} - \phi) \stackrel{a}{\sim} \mathcal{N}(0, (1 - \phi^2)), \quad (4.4)$$

de modo que podemos dizer que  $\hat{\phi}_{MV} = O_p(T^{-1/2})$ , ou seja, a taxa de convergência do estimador é  $1/\sqrt{T}$ .

No caso de raízes unitárias, a aproximação normal (4.2) não se aplica, logo não podemos usar a distribuição  $t$  para testar

$$\begin{aligned} H_0 &: \phi = 1, \\ H_1 &: \phi < 1. \end{aligned} \quad (4.5)$$

Suponha  $\theta_0 = 0$  em (4.1). Sabemos que EMV são assintoticamente equivalentes a EMQ, de modo que supondo  $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ , teremos

$$\hat{\phi}_{MQ} = \frac{\sum_{t=1}^T X_{t-1} X_t}{\sum_{t=1}^T X_{t-1}^2}. \quad (4.6)$$

É fácil ver que

$$\hat{\phi}_{MQ} - \phi = \frac{\sum_{t=1}^T X_{t-1} \varepsilon_t}{\sum_{t=1}^T X_{t-1}^2}, \quad (4.7)$$

que entra no numerador de (4.3) com  $\phi = \phi_0$ .

Para testar (4.5) temos que estudar o comportamento de

$$\hat{\phi}_{MQ} - 1 = \frac{\sum_{t=1}^T X_{t-1} \varepsilon_t}{\sum_{t=1}^T X_{t-1}^2}. \quad (4.8)$$

## 4.2 O Teste de Dickey-Fuller

Consideremos o modelo (4.1) com média zero, isto é,

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{RBN}(0, \sigma^2). \quad (4.9)$$

Segue-se que

$$\Delta X_t = \phi^* X_{t-1} + \varepsilon_t, \quad (4.10)$$

na qual  $\phi^* = \phi - 1$ . Podemos obter o EMQ de  $\phi^*$  por meio da regressão de MQ de  $\Delta X_t$  sobre  $X_{t-1}$ . Logo (4.5) é equivalente a

$$\begin{aligned} H_0^* &: \phi^* = 0, \\ H_1^* &: \phi^* < 0. \end{aligned} \quad (4.11)$$

O seguinte resultado será demonstrado no Apêndice a este capítulo.

**Teorema 4.1.** *Considere o modelo (4.9) com  $X_0 = 0$  e suponha que  $\varepsilon_t \sim i.i.d.(0, \sigma^2)$ . Então,*

$$T(\hat{\phi}_{MQ} - 1) \xrightarrow{\mathcal{D}} \frac{\frac{1}{2}([W(1)]^2 - 1)}{\int_0^1 [W(r)]^2 dr}, \quad (4.12)$$

onde  $W(r)$  é o Movimento Browniano padrão.

Em particular,  $W(1)^2 \sim \chi^2(1)$  e como  $P(\chi^2(1) < 1) = 0,68$ , de (4.12) temos que a probabilidade de que o lado esquerdo de (4.12) seja negativo converge para 0,68, para  $T \rightarrow \infty$ . Ou seja, mesmo que tenhamos um passeio aleatório ( $\phi = 1$ ), simulando-se muitas amostras de tal processo, em aproximadamente 2/3 delas o estimador  $\hat{\phi}_{MQ}$  será menor que 1. De (4.12) vemos que a taxa de convergência do estimador é diferente do caso estacionário:  $\hat{\phi}_{MQ} = O_p(T^{-1})$ .

Para testar (4.5) ou (4.11) podemos usar a estatística

$$\hat{\tau} = \frac{\hat{\phi}_{MQ}^*}{\widehat{\text{e.p.}}(\hat{\phi}_{MQ}^*)}, \quad (4.13)$$

em que

$$\widehat{\text{e.p.}}(\hat{\phi}_{MQ}^*) = \frac{S}{\left(\sum_{t=1}^T X_{t-1}^2\right)^{1/2}}, \quad (4.14)$$

e

$$S^2 = \frac{1}{T-2} \sum_{t=2}^T (\Delta X_t - \hat{\phi}_{MQ}^* X_{t-1})^2 \quad (4.15)$$

é o estimador de  $\sigma^2$  na regressão acima. Segue-se que a estatística (4.13) é equivalente a

$$\hat{\tau} = \frac{\hat{\phi}_{MQ} - 1}{(S^2 / \sum X_{t-1}^2)^{1/2}}, \quad (4.16)$$

que pode ainda ser escrita na forma

$$\hat{\tau} = \frac{T^{-1} \sum X_{t-1} \varepsilon_t}{S(T^{-2} \sum X_{t-1}^2)^{1/2}}. \quad (4.17)$$

O resultado a seguir é consequência do Teorema 4.1.

**Teorema 4.2.** *Sob as mesmas suposições do teorema anterior,*

$$\hat{\tau} \xrightarrow{\mathcal{D}} \frac{1}{2} \frac{([W(1)]^2 - 1)}{\left( \int_0^1 [W(r)]^2 dr \right)^{1/2}}. \quad (4.18)$$

Os testes usando (4.12) ou (4.18) são chamados *testes de Dickey-Fuller*, abreviadamente DF. Veja Dickey e Fuller(1979). As distribuições das estatísticas correspondentes são tabuladas. Valores críticos de  $\hat{\tau}$  para níveis de significância 0,01, 0,05 e 0,10 são dados, respectivamente, por  $-2,60$ ,  $-1,95$  e  $-1,61$ , para amostras de tamanho  $n = 100$ . Para amostras grandes, maiores que 500, esses valores são, respectivamente,  $-2,58$ ,  $-1,95$  e  $-1,62$ . Observe que rejeitamos  $H_0$  se  $\hat{\tau}$  for menor que o valor crítico apropriado. As densidades simuladas de  $T(\hat{\phi}_{MQ} - 1)$  e  $\hat{\tau}$ , sob  $H_0$ , estão representadas na Figura 4.1.

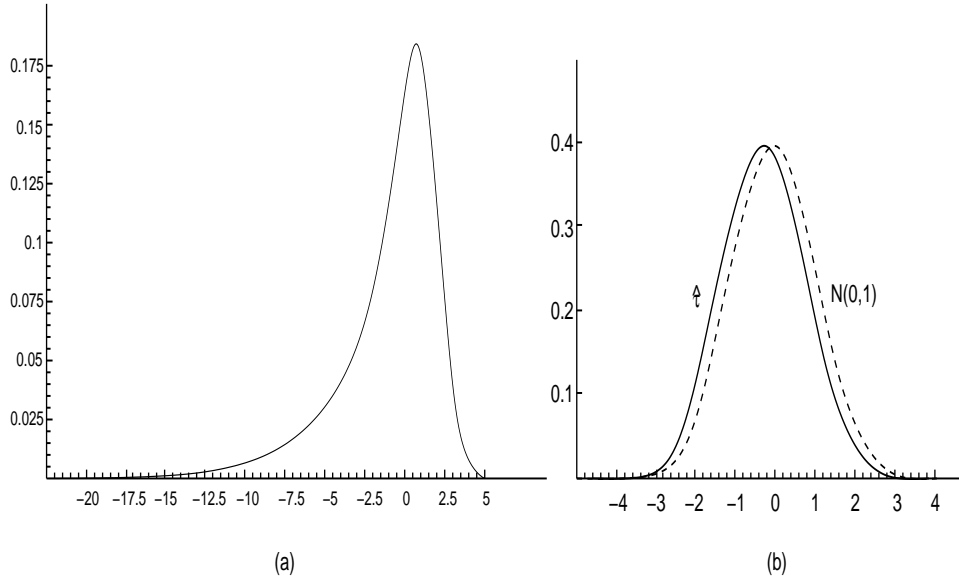


Figura 4.1: Distribuições, sob  $H_0 : \phi = 1$ , de (a)  $T(\hat{\phi}_{MQ} - 1)$  e (b)  $\hat{\tau}$ .

Suponha, agora, que a média não seja zero e temos o modelo (4.1). Neste caso,

$$\Delta X_t = \theta_0 + \phi^* X_{t-1} + \varepsilon_t, \quad (4.19)$$

onde  $\phi^* = \phi - 1$ . Novamente, teremos (4.5) e (4.11) como hipóteses equivalentes. O EMQ de  $\phi^*$  é obtido por meio da regressão de  $\Delta X_t$  sobre 1 e  $X_{t-1}$ . O denominador de (4.14) ficará, agora,  $(\sum (X_{t-1} - \bar{X})^2)^{1/2}$ . Embora  $\hat{\tau}$  ainda seja dada por (4.13), ou pelas expressões equivalentes (4.16) e (4.17), com os denominadores corrigidos, a presença de  $\theta_0$  altera a distribuição assintótica da estatística. Neste caso, a notação padrão utilizada para  $\hat{\tau}$  é  $\hat{\tau}_\mu$ , entendendo-se que o processo  $X_t$  tem média  $\mu = \theta_0/(1 - \phi)$ . No lugar de (4.12) e (4.18) teremos, respectivamente,

$$T(\hat{\phi}_{MQ} - 1) \xrightarrow{\mathcal{D}} \frac{\frac{1}{2}([W(1)]^2 - 1) - W(1) \int_0^1 W(r) dr}{\int_0^1 [W(r)]^2 dr - (\int_0^1 W(r) dr)^2}, \quad (4.20)$$

$$\hat{\tau}_\mu \xrightarrow{\mathcal{D}} \frac{1}{2} \frac{([W(1)]^2 - 1) - W(1) \int_0^1 W(r) dr}{[\int_0^1 [W(r)]^2 dr - (\int_0^1 W(r) dr)^2]^{1/2}}. \quad (4.21)$$

A distribuição de  $\hat{\tau}_\mu$  afasta-se mais da normal do que no caso  $\mu = 0$ . Veja a Figura 4.2. Valores críticos de  $\hat{\tau}_\mu$  para níveis de significância 0,05, 0,025 e 0,01 são dados por  $-2,86$ ,  $-3,12$  e  $-3,42$ , respectivamente, para amostras grandes.

O teste usando  $\hat{\tau}_\mu$  é chamado *teste de Dickey-Fuller aumentado* (“augmented Dickey-Fuller test”), abreviadamente, teste ADF. Veja Dickey e Fuller (1981).

Na realidade, as estatísticas  $\hat{\tau}$  e  $\hat{\tau}_\mu$  são usadas para testar  $H_0 : \phi = 1 \mid \theta_0 = 0$ . Poderíamos testar separadamente  $\theta_0 = 0$ , mas a estatística  $t$  novamente não é apropriada.

Para testar a hipótese

$$H_0 : \theta_0 = 0, \phi = 1, \quad (4.22)$$

podemos usar um teste do tipo Wald. Sob  $H_0$ , o processo é um passeio aleatório sem “drift”, de modo que podemos usar um teste da forma

$$\Phi_1 = \frac{\text{SQR}(\text{restrita}) - \text{SQR}(\text{irrestrita})/r}{\text{SQR}(\text{irrestrita})/(T - k)}, \quad (4.23)$$

onde  $r$  é o número de restrições sob  $H_0$  e  $k$  é o número de parâmetros do modelo irrestrito (de modo que  $T - k$  é o número de graus de liberdade do modelo irrestrito).

A estatística  $\Phi_1$  não tem distribuição  $F(r, T - k)$  e foi tabulada em Dickey e Fuller (1981).

Em nosso caso,  $r = k = 2$  e

$$\Phi_1 = \frac{(\sum \Delta X_t^2 - \sum \hat{\varepsilon}_t^2)/2}{\sum \hat{\varepsilon}_t^2/(T - 2)}. \quad (4.24)$$

Note que  $\sum \hat{\varepsilon}_t^2 = \sum (X_t - \hat{\theta}_0 - \hat{\phi}_{MQ} X_{t-1})^2$ . Valores críticos de  $\Phi_1$  para níveis de significância 0,05 e 0,01 são, respectivamente, 4,59 e 6,43.



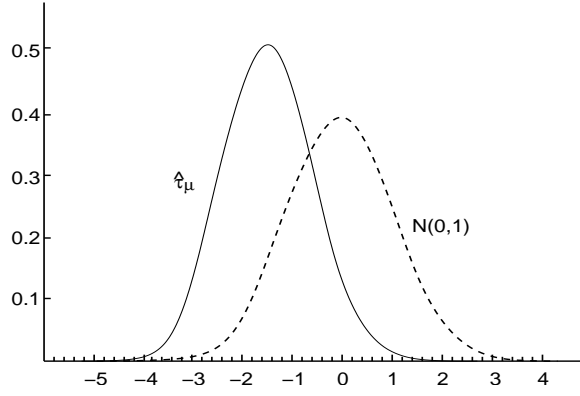


Figura 4.2: Distribuição de  $\hat{\tau}_\mu$  sob  $H_0 : \phi = 1$ .

### 4.3 Extensões do Teste DF

Suponha, agora, que a série possa ser representada por um processo AR(p):

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t, \quad (4.25)$$

onde  $\varepsilon_t$  como sempre é ruído branco de média zero e variância  $\sigma^2$ . O modelo pode, ainda, ser escrito na forma

$$X_t = \theta_0 + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t, \quad (4.26)$$

ou, ainda,

$$\Delta X_t = \theta_0 + \phi_1^* X_{t-1} + \phi_2^* \Delta X_{t-1} + \dots + \phi_p^* \Delta X_{t-p+1} + \varepsilon_t, \quad (4.27)$$

onde  $\phi_1^* = \sum_{i=1}^p \phi_i - 1$ ,  $\phi_j^* = -\sum_{i=j}^p \phi_i$ ,  $j = 2, \dots, p$ .

Se o polinômio auto-regressivo  $\phi(B)$  tiver uma raiz unitária, então  $\phi(1) = 0$ , ou seja  $1 - \phi_1 - \phi_2 - \dots - \phi_p = 0$ , ou ainda,  $\sum_{i=1}^p \phi_i = 1$  e portanto  $\phi_1^* = 0$ . Logo, testar a hipótese que o polinômio auto-regressivo tem uma raiz unitária é equivalente a testar a hipótese que  $\phi_1^* = 0$ .

Vemos que  $\phi_1^*$  pode ser estimado como o coeficiente de  $X_{t-1}$  na regressão de mínimos quadrados de  $\Delta X_t$  sobre  $1, X_{t-1}, \Delta X_{t-1}, \dots, \Delta X_{t-p+1}$ .

Para  $T$  grande, as estatísticas  $T(\hat{\phi}_1^* - 1)$  e  $\hat{\tau}_\mu = \hat{\phi}_1^* / \widehat{\text{e.p.}}(\hat{\phi}_1^*)$  têm as mesmas distribuições assintóticas dadas em (4.20) e (4.21).

Ng e Perron (1995) sugerem o seguinte procedimento para escolher  $p$ . Fixe um limite superior  $p_{\max}$  para  $p$  e estime a regressão (4.27) para  $p = p_{\max}$ . Se o valor absoluto da estatística  $t$  para testar a significância da última diferença  $\Delta X_t$  for maior do que 1,6, coloque  $p = p_{\max}$  e faça o teste de raiz unitária. Caso contrário, reduza o valor de  $p$  de uma unidade e repita o processo.

Schwert (1989) propõe tomar

$$p_{\max} = \left\lceil 12 \left( \frac{T}{100} \right)^{1/4} \right\rceil, \quad (4.28)$$

onde  $[x]$  denota o maior inteiro menor ou igual a  $x$ .

No caso de  $X_t \sim \text{ARMA}(p, q)$ , Said e Dickey (1985) provaram que  $\hat{\tau}_\mu$ , obtida do modelo

$$\Delta X_t = \theta_0 + \phi_1^* X_{t-1} + \sum_{i=1}^k \phi_{i+1}^* \Delta X_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (4.29)$$

com  $k = p - 1$ , tem a mesma distribuição assintótica que  $\hat{\tau}_\mu$  obtida de (4.27). Aqui supomos  $p$  e  $q$  conhecidos e o lag  $k$  usualmente é escolhido como em (4.28).

**Exemplo 4.1.** Suponha que  $X_t$  seja gerado por um processo ARIMA(1,1,0), com  $X_0 = 0$ ,  $\phi = 0,7$  e  $\varepsilon_t \sim \mathcal{N}(0, 1)$ . O gráfico está na Figura 4.3, juntamente com a f.a.c. e f.a.c.p.

Se não soubessemos como os dados  $X_t$  foram gerados, estas funções sugeririam que estamos diante de uma série não-estacionária. Na Figura 4.4 temos a série  $\Delta X_t$  e as respectivas f.a.c. e f.a.c.p.

Vemos, agora, que a hipótese de estacionariedade para a primeira diferença parece ser razoável. Em particular, a f.a.c.p. sugere um modelo AR(1) para  $\Delta X_t$ , que estimado pelo SPlus resulta

$$\hat{\phi} = 0,716, \hat{\sigma}^2 = 0,982, \text{d.p.}(\hat{\phi}) = 0,0313,$$

mostrando que rejeitamos  $H_0 : \phi = 0$ . O valor estimado está próximo do verdadeiro valor de  $\phi = 0,7$ . Ou seja, se estivéssemos de posse da série gerada

na Figura 4.3, ajustariamos o modelo

$$(1 - 0,716B)(1 - B)X_t = \varepsilon_t, \quad \varepsilon_t \sim \text{RB}(0; 0,982).$$

Novamente, não tendo conhecimento de como os dados foram gerados, vamos testar a presença de uma raiz unitária. A f.a.c.p. da série original sugere um modelo AR(3), de modo que podemos considerar

$$\Delta X_t = \theta_0 + \phi_1^* X_{t-1} + \phi_2^* \Delta X_{t-1} + \phi_3^* \Delta X_{t-2} + \varepsilon_t. \quad (4.30)$$

A regressão de  $\Delta X_t$  sobre 1,  $X_{t-1}$ ,  $\Delta X_{t-1}$  e  $\Delta X_{t-2}$  fornece o modelo ajustado

$$\Delta X_t = -0,094869 - 0,000275X_{t-1} + 0,707155\Delta X_{t-1} + 0,024766\Delta X_{t-2},$$

onde os desvios padrões dos coeficientes estimados são, respectivamente, 0,093498, 0,0012, 0,045049 e 0,045093.

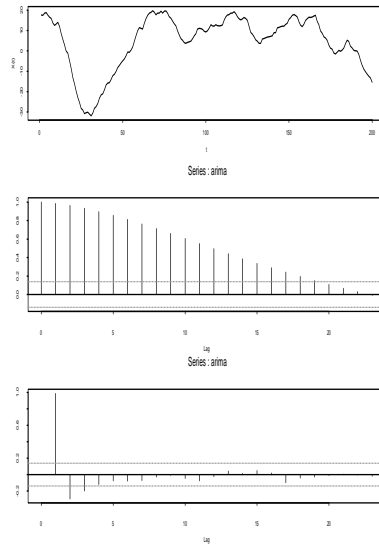


Figura 4.3: Série  $X_t$ , f.a.c. e f.a.c.p.

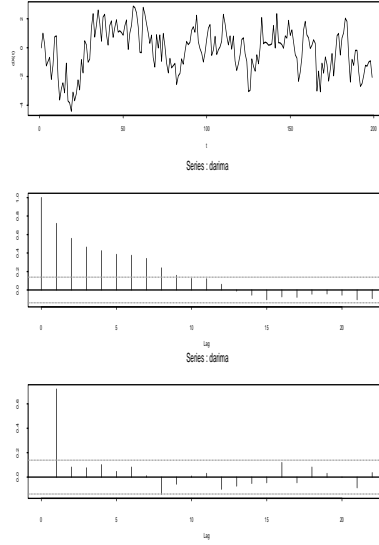


Figura 4.4: Série  $\Delta X_t$ , f.a.c. e f.a.c.p.

O valor da estatística do teste ADF é  $\hat{\tau}_\mu = (-0,000275/0,0012) = -0,2292$ , logo não rejeitamos a hipótese de que há uma raiz unitária com o nível de significância 0,01; o valor crítico é  $-3,45$ .

Na seção 4.2 consideramos o modelo (4.9) e o teste (4.5) ou (4.11). É claro que uma hipótese equivalente a (4.5) é

$$H_0 : \Delta X_t = \varepsilon_t, \quad (4.31)$$

onde  $\varepsilon_t \sim \text{RB}(0, \sigma^2)$ . Esta hipótese implica que a diferença de  $X_t$  é estacionária ( $X_t$  é “difference stationary”). A hipótese alternativa é  $\phi < 1$  ou  $X_t$  é estacionário.

Uma primeira extensão foi considerar adicionar ao modelo um termo constante, de modo que

$$H_0 : \Delta X_t = \theta_0 + \varepsilon_t. \quad (4.32)$$

Uma possível alternativa a esta hipótese é supor que

$$H_1 : X_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad (4.33)$$

ou seja,  $X_t$  apresenta uma tendência determinística (o processo é “trend stationary”).

Perron (1988) mostra que  $\hat{\tau}_\mu$  não é capaz de distinguir entre (4.33) e (4.32). Para testar  $H_0$  contra  $H_1$  acima, temos que estender o procedimento anterior, de modo a incluir uma tendência linear em (4.27):

$$\Delta X_t = \beta_0 + \beta_1 t + \phi_1^* X_{t-1} + \sum_{i=1}^k \phi_{i+1}^* \Delta X_{t-i} + \varepsilon_t, \quad (4.34)$$

com  $k = p - 1$ . A estatística para testar  $H_0 : \phi_1^* = 0$  é

$$\hat{\tau}_\tau = \frac{\hat{\phi}_{1MQ}^* - 1}{\widehat{\text{e.p.}}(\hat{\phi}_{1MQ}^*)}, \quad (4.35)$$

cujas distribuição limite é dada pelo resultado a seguir.

**Teorema 4.3.** *Sob a condição de que os erros sejam i.i.d., de média zero e variância  $\sigma^2$ ,*

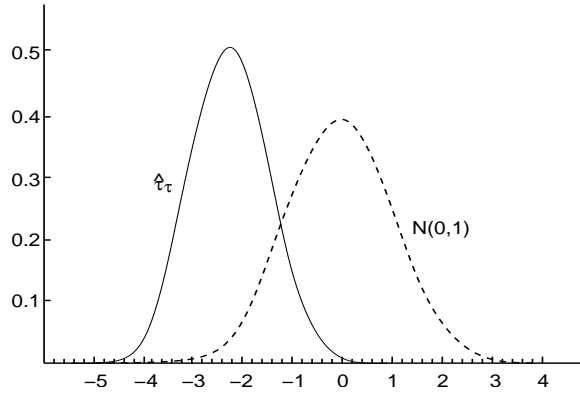
$$\hat{\tau}_\tau \xrightarrow{\mathcal{D}} \frac{\frac{1}{2}([W(1)]^2 - 1) - W(1) \int_0^1 W(r) dr + A}{[\int_0^1 [W(r)]^2 dr - (\int_0^1 W(r) dr)^2 + B]^{1/2}}, \quad (4.36)$$

em que

$$A = 12 \left[ \int_0^1 t W(t) dt - \frac{1}{2} \int_0^1 W(t) dt \right] \left[ \int_0^1 W(t) dt - \frac{1}{2} W(1) \right],$$

$$B = 12 \left[ \int_0^1 t W(t) dt \int_0^1 t W(t) dt - \left( \int_0^1 t W(t) dt \right)^2 \right] - 3 \left[ \int_0^1 W(t) dt \right]^2.$$

Na Figura 4.5. temos a densidade limite de  $\hat{\tau}_\tau$  sob  $H_0$ . Valores críticos da estatística para níveis 0,01, 0,025 e 0,05 são dados por  $-3,96$ ,  $-3,67$  e  $-3,41$ , respectivamente. Um teste conjunto  $H_0 : \phi_1^* = 0, \beta_1 = 0$  em (4.33) pode ser construído utilizando (4.23). Ver tabelas em Hamilton (1994).

Figura 4.5: Distribuição limite de  $\hat{\tau}_\tau$ .

Phillips (1987) e Phillips e Perron (1988) desenvolveram um teste que difere do teste ADF, supondo os erros correlacionados e possivelmente heteroscedásticos. Considera-se o modelo

$$\Delta X_t = \theta_0 + \phi_1^* X_{t-1} + u_t,$$

em que  $u_t \sim I(0)$ , satisfazendo determinadas condições de regularidade. Veja a seção 4.4. As estatísticas por eles consideradas são modificadas para levar em conta a auto-correlação e heteroscedasticidade, após estimar os modelos por MQ ordinários.

A estatística, correspondente a  $\hat{\tau}_\mu$ , usada no teste de Phillips e Perron (brevemente, PP) é dada por

$$Z(\tau_\mu) = \tau_\mu \left( \frac{\hat{\sigma}^2}{\hat{\lambda}^2} \right)^{1/2} - \frac{1}{2} \left( \frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2} \right) \left( \frac{T \widehat{\text{e.p.}}(\phi_1^*)}{\hat{\sigma}^2} \right), \quad (4.37)$$

onde  $\hat{\sigma}^2$  e  $\hat{\lambda}^2$  são estimadores consistentes de

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(u_t^2),$$

$$\lambda^2 = \lim_{T \rightarrow \infty} \sum_{t=1}^T E(T^{-1} S_T^2), \quad S_T = \sum_{t=1}^T u_t,$$

respectivamente. O estimador  $\hat{\sigma}^2$  é calculado como  $T^{-1} \sum_{t=1}^T \hat{u}_t^2$  e  $\hat{\lambda}^2$  é calculado como

$$\hat{\lambda}^2 = T^{-1} \sum_{t=1}^T \hat{u}_t^2 + \frac{2}{T} \sum_{j=1}^{\ell} \omega_j(\ell) \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j},$$

com  $\omega_j(\ell) = 1 - j/(\ell + 1)$ . Tal estimador é conhecido como estimador de Newey-West (Veja Newey e West, 1987). Na prática, Phillips e Perron (1988) sugerem usar  $\ell = [T^{1/4}]$ .

Sob  $H_0 : \phi_1^* = 0$ , a estatística (4.37) tem a mesma distribuição limite que  $\hat{\tau}_\mu$ . O teste PP é robusto a várias formas de heteroscedasticidade nos erros; além disso, não há necessidade de se especificar  $p$ . Para detalhes ver Hamilton (1994).

**Exemplo 4.2.** Considere a série Ibovespa, discutida no exemplo 1.1. O gráfico da série está na Figura 4.6 (a) e os gráficos da f.a.c. e f.a.c.p. estão na Figura 4.6 (b). A f.a.c.p. sugere um modelo AR(1) para a série, logo o modelo a considerar é (4.19). Usando o EViews obtemos o Quadro 4.1. O valor da estatística de teste é  $\hat{\tau}_\mu = -1,057312$ , e não rejeitamos a hipótese nula em (4.11), com o nível de 0,01 (valor crítico  $-3,4376$ ). Como a constante do modelo parece ser não significativa, podemos considerar o caso (4.10), e a estatística do teste tem valor  $\hat{\tau} = 0,610074$ , e novamente não rejeitamos a hipótese de existência de uma raiz unitária.

**Exemplo 4.3.** Vamos aplicar os testes ADF e PP para a série de índices de fechamento diários da Bolsa de New York, o Dow Jones Industrial Average (DJIA), de 3/1/95 a 26/11/02, com  $T = 1992$  observações. Na Figura 4.7 temos a série e sua f.a.c., bem como a série de retornos e respectiva f.a.c. Usando a formulação (4.34), o coeficiente  $\beta_1$  não foi significativo, o que nos leva a considerar o modelo com uma constante. Usando o S+FinMetrics, obtemos o Quadro 4.2 para o teste ADF, tomando-se  $p_{\max} = 6$ . Nota-se, contudo, que os coeficientes dos lags de 2 a 6 não são significativos. Com  $p = 1$  obtemos o Quadro 4.3, do qual segue que não rejeitamos a hipótese nula de raiz unitária. No Quadro 4.4 temos o resultado da aplicação do teste PP, que também detecta raiz unitária na série.

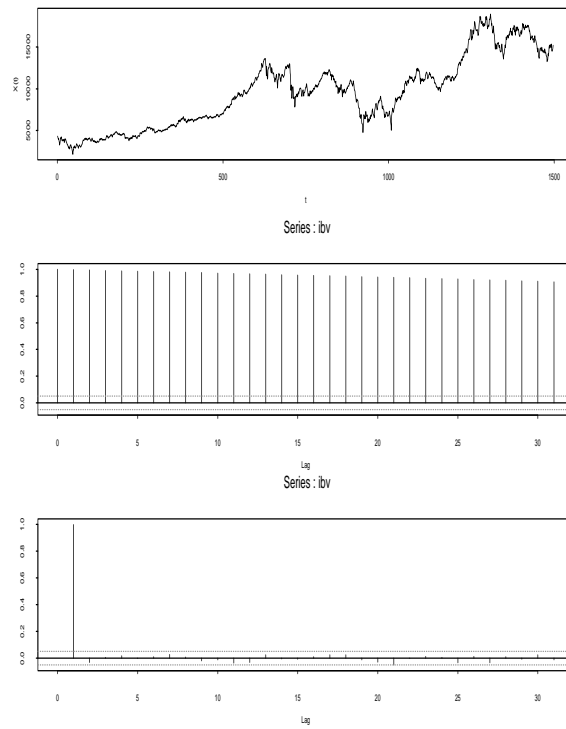


Figura 4.6: Série Ibovespa, f.a.c. e f.a.c.p.

## 4.4 Comentários Finais

Encerramos este capítulo com algumas observações sobre os testes de raízes unitárias.

[1] No início da seção 4.1 ressaltamos que podemos ter raízes unitárias também na parte de médias móveis de um modelo ARMA. Na realidade isto tem a ver com um teste de estacionariedade, onde a hipótese nula especifica que o processo é estacionário ao redor de uma tendência determinística (“trend-stationary”) e a hipótese alternativa especifica que o processo é  $I(1)$ .



```

=====
Null Hypothesis: IBV has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=23)

-----
Augmented Dickey-Fuller test statistic -1.057312 0.7343
Test critical values:                1% level -3.434508
                                   5% level -2.863264
                                   10% level -2.567736

-----
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(IBV)
Method: Least Squares
Date: 01/18/06   Time: 15:53
Sample (adjusted): 1/04/1995 9/29/2000
Included observations: 1498 after adjustments

-----
Variable   Coefficient Std. Error t-Statistic   Prob.
-----
IBV(-1)    -0.001604    0.001517    -1.057312    0.2905
C           22.42976    15.70572     1.428126    0.1535

-----
R-squared          0.000747    Mean dependent var    7.254339
Adjusted R-squared 0.000079    S.D. dependent var    246.8253
S.E. of regression 246.8155    Akaike info criterion 13.85649
Sum squared resid  91133196    Schwarz criterion     13.86359
Log likelihood     -10376.51    F-statistic           1.117909
Durbin-Watson stat 1.876427    Prob(F-statistic)     0.290540
=====

```

Quadro 4.1 - Teste ADF para a série Ibovespa

```

Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
Type of Test: t test
Test Statistic: -2.019
P-value: 0.2787

Coefficients:
      Value Std. Error  t value Pr(>|t|)
lag1  -0.0021   0.0011   -2.0190   0.0436
lag2   0.0091   0.0225    0.4038   0.6864
lag3  -0.0402   0.0225   -1.7905   0.0735
lag4  -0.0187   0.0225   -0.8319   0.4056
lag5   0.0055   0.0225    0.2440   0.8073
lag6  -0.0300   0.0225   -1.3350   0.1820
constant 20.5289  9.1876    2.2344   0.0256

Regression Diagnostics:

      R-Squared 0.0050
Adjusted R-Squared 0.0020
Durbin-Watson Stat 1.9993

Residual standard error: 105.2 on 1982 degrees
of freedom
F-statistic: 1.649 on 6 and 1979 degrees of freedom,
the p-value is 0.1298

```

#### Quadro 4.2 - Teste ADF, com $p = 6$ , para a série DJIA

```

Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
Type of Test: t test
Test Statistic: -2.021
P-value: 0.2777

Coefficients:
      Value Std. Error  t value Pr(>|t|)
lag1  -0.0021   0.0011   -2.0215   0.0434
constant 20.2425  9.1211    2.2193   0.0266

Regression Diagnostics:

      R-Squared 0.0021
Adjusted R-Squared 0.0015
Durbin-Watson Stat 1.9804

Residual standard error: 105.1 on 1987 degrees
of freedom
F-statistic: 4.086 on 1 and 1989 degrees of freedom,
the p-value is 0.04337

```

#### Quadro 4.3 - Teste ADF, com $p = 1$ , para a série DJIA

```

Test for Unit Root: Phillips-Perron Test

Null Hypothesis: there is a unit root
  Type of Test: t test
  Test Statistic: -2.02
    P-value: 0.2782
using bartlett window with bandwidth 8

Coefficients:
      Value Std. Error  t value Pr(>|t|)
lag1  -0.0021   0.0011  -2.0215   0.0434
constant 20.2425   9.1211   2.2193   0.0266

Regression Diagnostics:

      R-Squared 0.0021
Adjusted R-Squared 0.0015
Durbin-Watson Stat 1.9804

Residual standard error: 105.1 on 1987 degrees
of freedom
F-statistic: 4.086 on 1 and 1989 degrees of freedom,
the p-value is 0.04337

```

#### Quadro 4.4 - Teste de Phillips-Perron para a série DJIA

Para ilustrar, retomemos o modelo (2.58)

$$X_t = \beta_0 + \beta_1 t + \varepsilon_t,$$

onde agora  $\varepsilon_t = \phi \varepsilon_{t-1} + \eta_t$ , sendo  $\eta_t$  ruído branco. Se  $|\phi| < 1$ ,  $X_t$  será um processo “trend-stationary”, ao redor de  $\mu_t = \beta_0 + \beta_1 t$ . Por outro lado, se  $\phi = 1$ ,  $\varepsilon_t$  é passeio casual e  $X_t$  é um processo  $I(1)$  com “drift”. Como

$$\begin{aligned}\Delta X_t &= \beta_1 + \Delta \varepsilon_t, \\ \Delta \varepsilon_t &= \phi \Delta \varepsilon_{t-1} + \eta_t - \eta_{t-1},\end{aligned}$$

temos uma raiz unitária na representação ARMA de  $\Delta X_t$ . Para detalhes sobre este teste veja Kviatkowski et al. (1992).

[2] Phillips (1987) mostra que

$$\frac{1}{2}[W(1)^2 - 1] = \int_0^1 W(r)dW(r), \quad (4.38)$$

logo os numeradores das distribuições assintóticas de (4.12) e (4.18) podem ser substituídos pelo lado direito de (4.38), o mesmo valendo para (4.36).

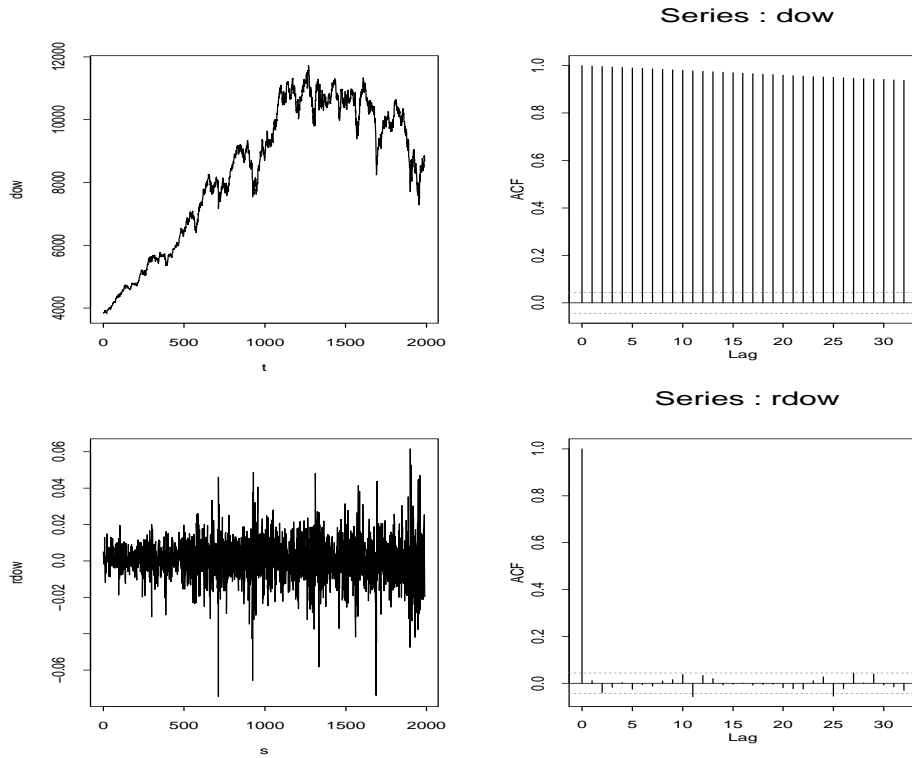


Figura 4.7: (a) Série DJIA (b) f.a.c. da série (c) Retornos do DJIA (d) f.a.c. dos retornos

[3] As suposições para o modelo considerado no teste de PP são:

- (i)  $E(u_t) = 0$ , para todo  $t$ .
- (ii)  $\sup_t E(|u_t|^\beta) < \infty$ , para algum  $\beta > 2$ .

- (iii)  $\lambda^2 = \lim_{T \rightarrow \infty} E(S_T^2/T)$  existe e é finito,  $S_T = \sum_{t=1}^T u_t$ .  
 (iv)  $u_t$  é fortemente “mixing”.

Ver Phillips e Perron (1988) e Hamilton (1994) para o conceito de “mixing” (independência assintótica) e discussão das implicações dessas suposições.

[4] Testes de raízes unitárias apresentam vários problemas, tais como: (a) baixo poder para discriminar processos estacionários persistentes ( $|\phi|$  próximo de um) de processos não estacionários; (b) o poder diminui com a introdução de termos determinísticos ao modelo AR(1) básico sem constante. Veja Perron e Ng (1996) e Elliot et al. (1996) para sugestões que aliviam estes problemas.

[5] Phillips e Shimotsu (2004) tratam do problema da estimação do parâmetro  $d$  de um processo de memória longa quando  $d > 1/2$ , ou seja, no caso não estacionário, incluindo o caso de raízes unitárias e quando o processo tem uma tendência polinomial. O estimador de  $d$  por eles utilizado é o estimador local de Whittle. Veja o Capítulo 9 para noções sobre processos de memória longa.

## 4.5 Problemas

1. Considere a série de preços diários de ações da Petrobrás (arquivo d-petro95.00.dat). Considere o modelo ajustado no Capítulo 3, exemplos 3.6 e 3.10. Teste se há raiz unitária presente na série, usando os testes ADF e PP.
2. Teste para raízes unitárias nas seguintes séries, usando os testes ADF e PP.
  - (a) preços diários de ações do Banespa (arquivo d-ban95.00.dat);
  - (b) índices mensais do Bovespa (arquivo m-ibv94.01.dat);
  - (c) dados mensais dos juros do C-Bond brasileiro (arquivo m-cbond94.01.dat), de julho de 1994 a agosto de 2001,  $T = 86$ .
3. Teste para mais de uma raiz unitária. O desenvolvimento apresentado no texto supõe que a série seja  $I(1)$ , ou seja, contém no máximo uma raiz unitária. Podemos testar se a série é  $I(2)$ , dado que não rejeitamos

$H_0 : \phi = 1$ . Os testes DF e ADF não podem ser usados para  $\Delta X_t$ , pois esses admitem que há, no máximo, uma raiz. Dickey e Pantula (1987) sugerem um procedimento para o caso da série conter no máximo duas raízes unitárias. Para testar  $H_0 : d = 2$  (a série tem duas raízes) contra  $H_1 : d = 1$  (a série contém uma raiz), considere (no caso AR(1)),

$$\Delta^2 X_t = \beta_0 + \beta_2 \Delta X_{t-1} + \varepsilon_t.$$

Compare o valor da razão  $t$  para  $\beta_2$  com valor crítico de  $\hat{\tau}_\mu$ . Rejeite  $H_0$  se  $t_{\beta_2} < \hat{\tau}_\mu$ . Se rejeitarmos  $H_0$ , podemos testar  $H'_0$  : há uma raiz unitária, contra  $H'_1$ : não há raiz unitária, comparando  $\hat{\tau}_\mu$  com  $t_{\beta'_1}$  e  $t_{\beta'_2}$  em

$$\Delta^2 X_t = \beta'_0 + \beta'_1 X_{t-1} + \beta'_2 \Delta X_{t-1} + \varepsilon_t.$$

Rejeite  $H'_0$  se as duas razões  $t$  forem menores do que o valor crítico de  $\hat{\tau}_\mu$ .

4. Teste se a série ICV (arquivo m-icv70.79.dat) tem no máximo duas raízes unitárias.
5. Mesmo problema para as séries do problema 2.
6. Considere (4.33) com  $k = 1$  ( ou  $p = 2$ ). Por meio de substituições sucessivas, mostre que se  $\phi = 1$  e  $\beta_1 \neq 0$ ,  $X_t$  conterà uma tendência quadrática.

*Observação.* É sempre possível que  $\beta_1 \neq 0$  e neste caso a tendência quadrática dominará a componente I(1). Dolado et al. (1990) e Banerjee et al. (1993) mostram que, nesse caso, podemos usar a distribuição normal padrão como aproximação para  $\hat{\tau}_\tau$ . O mesmo acontece para  $\hat{\tau}_\mu$  se  $\beta_1 = 0$  mas  $\beta_0 \neq 0$ . Dolado et al. (1990) propõem um procedimento para testar raízes unitárias quando há tendências presentes.

## Apêndice 4: Provas dos Teoremas 4.1 e 4.2

Neste apêndice vamos demonstrar os teoremas 4.1 e 4.2, baseados em resultados da seção 2.7. Os demais teoremas deste capítulo podem ser demonstrados de modo análogo e as provas podem ser vistas, por exemplo, em Hamilton (1994).

### Prova do Teorema 4.1.

De (4.8) temos (com  $\hat{\phi}_{MQ} \equiv \hat{\phi}$ )

$$T(\hat{\phi} - 1) = \frac{\frac{1}{T} \sum_{t=1}^T X_{t-1} \varepsilon_t}{\frac{1}{T^2} \sum_{t=1}^T X_{t-1}^2}. \quad (A.1)$$

Se  $\phi = 1$ , de (4.9), com  $X_0 = 0$ , temos que  $X_t = X_0 + \varepsilon_1 + \dots + \varepsilon_t = \sum_{j=1}^t \varepsilon_j$ , logo concluímos que  $X_t \sim \mathcal{N}(0, \sigma^2 t)$ . Também,

$$X_t^2 = (X_{t-1} + \varepsilon_t)^2 = X_{t-1}^2 + \varepsilon_t^2 + 2X_{t-1}\varepsilon_t,$$

de onde obtemos

$$\begin{aligned} \sum_{t=1}^T X_{t-1} \varepsilon_t &= \frac{1}{2} \sum_{t=1}^T \{(X_t^2 - X_{t-1}^2 - \varepsilon_t^2)\}, \\ &= \frac{1}{2} (X_T^2 - X_0^2) - \frac{1}{2} \sum_{t=1}^T \varepsilon_t^2. \end{aligned}$$

Dividindo-se ambos os membros por  $T\sigma^2$  teremos

$$\frac{1}{T\sigma^2} \sum_{t=1}^T X_{t-1} \varepsilon_t = \frac{1}{2} \left( \frac{X_T}{\sigma\sqrt{T}} \right)^2 - \frac{1}{2\sigma^2 T} \sum_{t=1}^T \varepsilon_t^2. \quad (A.2)$$

Mas  $X_T/\sigma\sqrt{T} \sim \mathcal{N}(0, 1)$ , logo  $(X_T/\sigma\sqrt{T})^2$  tem uma distribuição  $\chi^2(1)$ , e como os  $\varepsilon_t$  são v.a. i.i.d., com  $E(\varepsilon_t^2) = \sigma^2$ , então pela lei fraca dos grandes números temos que

$$\frac{1}{T} \sum_{t=1}^T \varepsilon_t^2 \xrightarrow{P} \sigma^2. \quad (A.3)$$

Conclui-se por (A.2) e (A.3) que

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T X_{t-1} \varepsilon_t \xrightarrow{\mathcal{D}} \frac{1}{2} (X - 1), \quad (\text{A.4})$$

onde  $X \sim \chi^2(1)$ .

Vejamos o que acontece com o denominador de (A.1). Como  $\varepsilon_t$  são i.i.d., de média zero e variância  $\sigma^2$ , podemos formar as somas (2.82), ou seja,

$$Y_T(r) = \frac{1}{T} \sum_{t=1}^{[Tr]} \varepsilon_t,$$

e então, de acordo com (2.83),

$$Y_T(r) = \begin{cases} 0, & \text{se } 0 \leq r < 1/T, \\ X_1/T, & \text{se } 1/T \leq r < 2/T, \\ \vdots & \\ X_T/T, & \text{se } r = 1. \end{cases}$$

Lembremos que  $Y_T(r)$  é uma função em escada, com valor  $X_j/T$ , para  $j/T \leq r < (j+1)/T$ .

Como consequência da definição 2.11 e do teorema da aplicação contínua, vimos que

$$S_T(r) = [\sqrt{T} Y_T(r)]^2 \xrightarrow{\mathcal{D}} \sigma^2 [W(r)]^2, \quad (\text{A.5})$$

onde  $W(\cdot)$  é o movimento Browniano padrão. Mas

$$S_T(r) = \begin{cases} 0, & \text{se } 0 \leq r < 1/T \\ X_1^2/T, & \text{se } 1/T \leq r < 2/T \\ \vdots & \\ X_T^2/T, & \text{se } r = 1. \end{cases}$$

Segue-se que

$$\int_0^1 S_T(r) dr = \frac{1}{T^2} \sum_{t=2}^T X_{t-1}^2,$$



e de (A.5) e novamente do teorema da aplicação contínua temos que

$$\frac{1}{T^2} \sum_{t=2}^T X_{t-1}^2 \xrightarrow{\mathcal{D}} \sigma^2 \int_0^1 [W(r)]^2 dr. \quad (\text{A.6})$$

Como  $T(\hat{\phi} - 1)$  é uma função contínua de (A.4) e (A.6) obtemos (4.12), lembrando que  $W(1)^2 = \chi^2(1)$ .

**Prova do Teorema 4.2.**

Como  $\hat{\tau}$  pode ser escrita como em (4.17) e  $S^2 \xrightarrow{P} \sigma^2$ , segue-se que

$$\hat{\tau} \xrightarrow{\mathcal{D}} \frac{\frac{1}{2}\sigma^2[W(1)^2 - 1]}{\left(\sigma^2 \int_0^1 W(r)^2 dr\right)^{1/2} (\sigma^2)^{1/2}},$$

que coincide com (4.18).



# Modelagem da Volatilidade

## 5.1 Introdução

Neste capítulo iremos estudar alguns modelos apropriados para séries financeiras que apresentam a variância condicional evoluindo no tempo. Vimos que os modelos lineares do tipo ARMA, por exemplo, admitem que as inovações sejam um ruído branco, com média zero e variância constante.

Há uma variedade muito grande de modelos não-lineares disponíveis na literatura, mas nós iremos nos concentrar na classe de modelos ARCH (de “auto-regressive conditional heterocedasticity”), introduzida por Engle (1982) e suas extensões. Esses modelos são não-lineares no que se refere à variância, como veremos na seção seguinte. Consideraremos, também, os chamados modelos de volatilidade estocástica, que também admitem que a volatilidade varie com o tempo, mas têm uma premissa diferente da dos modelos ARCH-GARCH. Faremos apenas uma descrição sumária de alguns modelos não-lineares na próxima seção. Para detalhes sobre outros modelos, veja Tong (1990) e Tsay (2002).

Como dissemos acima, o objetivo será modelar a volatilidade de um retorno. A notação a ser usada é aquela estabelecida na seção 1.7. Consideremos uma série de retornos,

$$X_t = \ln(P_t) - \ln(P_{t-1}),$$

e sejam

$$\mu_t = E(X_t|\mathcal{F}_{t-1}), \quad h_t = \text{Var}(X_t|\mathcal{F}_{t-1}) \quad (5.1)$$

a média e variância condicional de  $X_t$ , onde  $\mathcal{F}_t$  denotará a informação até o instante  $t$ , que consideraremos ser  $\{X_t, X_{t-1}, \dots, X_1\}$ .

Em algumas situações iremos supor que  $\mu_t = 0$ , de modo que neste caso  $h_t = E(X_t^2|\mathcal{F}_{t-1})$ .

## 5.2 Alguns Modelos Não-Lineares

Definimos no exemplo 2.3 o que seja um modelo linear; veja (2.52). Temos que o valor do processo no instante  $t$  depende de uma função linear de valores presente e passados do ruído branco  $\varepsilon_t$ .

Na análise de modelos não-lineares as inovações (choques aleatórios)  $\varepsilon_t$  são em geral supostas i.i.d. e o modelo tem a forma

$$X_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (5.2)$$

de modo que  $g(\cdot)$  representa a média condicional e  $h^2(\cdot)$  é a variância condicional. Se  $g(\cdot)$  for não-linear, o modelo diz-se *não-linear na média*, enquanto se  $h(\cdot)$  for não-linear, o modelo diz-se *não-linear na variância*.

O modelo

$$X_t = \varepsilon_t + \alpha \varepsilon_{t-1}^2,$$

é não-linear na média, pois  $g(\cdot) = \alpha \varepsilon_{t-1}^2$  e  $h(\cdot) = 1$ , ao passo que o modelo ARCH(1)

$$X_t = \varepsilon_t \sqrt{\alpha X_{t-1}^2}$$

é não-linear na variância, pois  $g(\cdot) = 0$  e  $h(\cdot) = \sqrt{\alpha X_{t-1}^2}$ , e  $X_{t-1}$  depende de  $\varepsilon_{t-1}$ .

Vejamos alguns modelos não-lineares.

### Exemplo 5.1. Modelos Polinomiais

Para este modelo, a função  $g(\cdot)$  fica

$$g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sum_{i=1}^{\infty} c_i \varepsilon_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} b_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \dots \quad (5.3)$$

Esta é uma expansão de Volterra, envolvendo termos lineares, bilineares, etc das inovações.

**Exemplo 5.2.** *Modelos Bilineares*

Estes modelos, extensivamente estudados por Subba Rao (1981), Subba Rao e Gabr (1984) e Granger e Andersen (1978), são tais que

$$g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i} + \sum_{i=1}^{\infty} \beta_i X_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \gamma_{ij} X_{t-i} \varepsilon_{t-j}. \quad (5.4)$$

**Exemplo 5.3.** *Modelos Lineares por Partes*

Dentre esses modelos destacamos os modelos TAR (“threshold autoregressive”) de Tong(1983, 1990). Por exemplo, um TAR(1) é da forma

$$X_t = \begin{cases} \alpha_1 + \beta_1 X_{t-1} + \varepsilon_t, & \text{se } X_{t-1} < r, \\ \alpha_2 + \beta_2 X_{t-1} + \varepsilon_t, & \text{se } X_{t-1} \geq r. \end{cases}$$

Aqui, o processo tem dois regimes, dependendo de  $X_{t-1}$  ser menor ou maior do que o limiar  $r$ . Em geral podemos ter um TAR(p), com  $k$  regimes. Podemos considerar modelos TARMA(p,q), ou seja, modelos ARMA com limiares. Veja Sáfadi e Morettin (2000) para detalhes. Modelos com quebras estruturais podem ser incluídos nessa classe de modelos e um dos objetivos é estimar os pontos de mudanças de regimes.

**Exemplo 5.4.** *“Switching Models”*

Esses modelos são semelhantes aos modelos lineares por partes, mas agora a troca de regimes é governada por uma variável de estado, não-observada, que tipicamente é modelada como uma cadeia de Markov. Por exemplo, considere

$$X_t = \begin{cases} \alpha_1 + \beta_1 X_{t-1} + \varepsilon_{1t}, & \text{se } S_t = 1 \\ \alpha_2 + \beta_2 X_{t-1} + \varepsilon_{2t}, & \text{se } S_t = 0. \end{cases} \quad (5.5)$$

Aqui,  $S_t$  é uma cadeia de Markov com dois estados, 0 e 1. Veja Hamilton (1994) e Kim e Nelson(1999) para mais informação sobre esses modelos.

### 5.3 Modelos ARCH

Os modelos ARCH, ou modelos auto-regressivos com heteroscedasticidade condicional, foram introduzidos por Engle (1982), com o objetivo de estimar a variância da inflação. A idéia básica aqui é que o retorno  $X_t$  é não-correlacionado serialmente, mas a volatilidade (variância condicional) depende de retornos passados por meio de uma função quadrática.

**Definição 5.1.** Um modelo ARCH( $r$ ) é definido por

$$X_t = \sqrt{h_t} \varepsilon_t, \quad (5.6)$$

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_r X_{t-r}^2, \quad (5.7)$$

onde  $\varepsilon_t$  i.i.d.  $(0, 1)$ ,  $\alpha_0 > 0, \alpha_i \geq 0, i > 0$ .

Na prática, usualmente supomos  $\varepsilon_t \sim \mathcal{N}(0, 1)$ ,  $\varepsilon_t \sim t_\nu$  ou uma distribuição que descreva melhor as caudas pesadas de séries financeiras. Veja o apêndice a este capítulo para algumas sugestões adicionais. Os coeficientes  $\alpha_i$  devem satisfazer certas condições, dependendo do tipo de imposição que colocarmos sobre o processo  $X_t$ .

Pela própria definição, valores grandes de  $X_t$  são seguidos por outros valores grandes da série.

Para investigar algumas propriedades dos modelos ARCH, consideremos o caso especial  $r = 1$ , ou seja, temos o modelo

$$X_t = \sqrt{h_t} \varepsilon_t, \quad (5.8)$$

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2, \quad (5.9)$$

com  $\alpha_0 > 0, \alpha_1 \geq 0$ .

Calculemos a média e variância incondicionais da série.

$$(i) E(X_t) = E\{E(X_t|\mathcal{F}_{t-1})\} = 0;$$

$$(ii) \text{Var}(X_t) = E(X_t^2) = E\{E(X_t^2|\mathcal{F}_{t-1})\}$$

$$= E(\alpha_0 + \alpha_1 X_{t-1}^2) = \alpha_0 + \alpha_1 E(X_{t-1}^2).$$

Se o processo  $\{X_t\}$  for estacionário de segunda ordem, então, para todo  $t$ ,  $E(X_t^2) = E(X_{t-1}^2) = \text{Var}(X_t)$ , do que decorre

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \alpha_1}. \quad (5.10)$$

Como  $\text{Var}(X_t) > 0$ , deveremos ter  $0 \leq \alpha_1 < 1$ .

$$\begin{aligned} \text{(iii)} \quad \text{Cov}(X_t, X_{t+k}) &= E(X_t X_{t+k}) \\ &= E[E(X_t X_{t+k} | \mathcal{F}_{t+k-1})] = E[X_t E(X_{t+k} | \mathcal{F}_{t+k-1})] \\ &= E[X_t E(\sqrt{h_{t+k}} \varepsilon_{t+k} | \mathcal{F}_{t+k-1})] = 0, \text{ para } k > 0, \text{ pois } X_t \text{ está em } \mathcal{F}_{t+k-1} \text{ e} \\ &E(\varepsilon_{t+k} | \mathcal{F}_{t+k-1}) = 0. \end{aligned}$$

Dessa forma,

$$\gamma_X(k) = 0, \quad k \geq 1,$$

indicando que  $X_t$  é uma seqüência de variáveis não-correlacionadas (ruído branco), com média zero e variância dada por (5.10).

Sabemos que os retornos apresentam geralmente caudas longas, de modo que a curtose é maior do que 3. Para calcular a curtose, supondo que  $X_t$  siga o modelo (5.8)-(5.9) é necessário calcular o momento de quarta ordem de  $X_t$ . Suponha que os  $\varepsilon_t$  sejam normais, para facilidade de cálculo. Então temos:

$$E(X_t^4 | \mathcal{F}_{t-1}) = E(h_t^2 \varepsilon_t^4 | \mathcal{F}_{t-1}) = 3(\alpha_0 + \alpha_1 X_{t-1}^2)^2, \quad (5.11)$$

pois  $E(\varepsilon_t^4) = 3$ , do que decorre

$$E(X_t^4) = 3E(\alpha_0 + \alpha_1 X_{t-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0\alpha_1 X_{t-1}^2 + \alpha_1^2 X_{t-1}^4).$$

Admitindo-se que o processo seja estacionário de quarta ordem, o momento de quarta ordem pode ser escrito  $\mu_4 = E(X_t^4)$  e teremos

$$\begin{aligned} \mu_4 &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(X_t) + \alpha_1^2 \mu_4) \\ &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 [\alpha_0/(1 - \alpha_1)] + \alpha_1^2 \mu_4) \\ &= 3\alpha_0^2(1 + 2\alpha_1/(1 - \alpha_1)) + 3\alpha_1^2 \mu_4. \end{aligned}$$

Daqui obtemos, finalmente,

$$\mu_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}. \quad (5.12)$$

Supondo-se que momentos de quarta ordem sejam finitos e positivos, de (5.12) devemos ter  $1 - 3\alpha_1^2 > 0$ , ou seja,  $0 \leq \alpha_1^2 < 1/3$ . Portanto, quanto mais restrições impusermos ao processo de retornos, mais restrições teremos para os coeficientes do modelo. Isto é verdade para o modelo geral ARCH( $r$ ).

A curtose de  $X_t$  será, então, dada por

$$K = \frac{\mu_4}{[\text{Var}(X_t)]^2} = 3 \frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3. \quad (5.13)$$

Vemos, pois, que se admitirmos que  $X_t$  siga um modelo ARCH, as caudas serão mais pesadas do que as da normal, o que é uma propriedade vantajosa do modelo. Por outro lado, uma desvantagem do modelo é que trata retornos positivos e negativos de forma similar, já que quadrados dos retornos entram na fórmula da volatilidade. Na prática, sabe-se que a volatilidade reage de modo diferente a retornos positivos e negativos. Também, devido ao fato de termos retornos ao quadrado, alguns retornos grandes e isolados podem conduzir a super-previsões.

Utilizando (5.8) e (5.9) e calculando  $X_t^2 - h_t$ , temos que

$$X_t^2 - (\alpha_0 + \alpha_1 X_{t-1}^2) = h_t(\varepsilon_t^2 - 1),$$

ou seja,

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + v_t, \quad (5.14)$$

na qual

$$v_t = h_t(\varepsilon_t^2 - 1) = h_t(X - 1), \quad (5.15)$$

onde  $X$  é uma v.a. com distribuição  $\chi^2(1)$ , o que mostra que temos um modelo AR(1) para  $X_t^2$ , mas com erros não-gaussianos. Ainda, é fácil ver que  $\{v_t\}$  é



uma sequência de v.a. de média zero, não-correlacionadas, mas com variância não-constante.

De (5.14) temos que a f.a.c. de  $X_t^2$  é dada por

$$\rho_{X^2}(k) = \alpha_1^k, \quad k > 0.$$

Para um modelo ARCH( $r$ ) teremos

$$X_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + v_t, \quad (5.16)$$

onde os  $v_t$  são como no caso  $r = 1$ . Ou seja, temos um modelo AR( $p$ ) para  $X_t^2$ , com inovações não-gaussianas. Além disso, pode-se demonstrar que os retornos  $X_t$  também formam um ruído branco, com variância dada por

$$\text{Var}(X_t) = \frac{\alpha_0}{1 - \sum_{i=1}^r \alpha_i}.$$

### Identificação

Um primeiro passo na construção de modelos ARCH é tentar ajustar modelos ARMA, para remover a correlação serial na série, se esta existir. Se este for o caso, teremos

$$\phi(B)X_t = \theta_0 + \theta(B)a_t,$$

sendo que  $a_t \sim \text{ARCH}(r)$ . No que segue, quando nos referirmos a  $X_t$ , estaremos supondo ou que a série é não-correlacionada, ou então ela é o resíduo da aplicação de um modelo ARMA à série original.

Para verificarmos se a série apresenta heteroscedasticidade condicional, podemos utilizar dois testes, examinando-se a série  $X_t^2$ .

(i) Teste de Box-Pierce-Ljung para  $X_t^2$ . Veja a seção 3.4.2.

(ii) Teste de multiplicadores de Lagrange (ML); veja Engle (1982). Queremos testar  $H_0 : \alpha_i = 0$ , para todo  $i = 1, \dots, r$ , na regressão

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_r X_{t-r}^2 + u_t,$$

para  $t = r + 1, \dots, T$ . A estatística do teste é  $S = TR^2$ , que tem distribuição assintótica  $\chi^2(r)$  sob  $H_0$ . Aqui,  $R^2$  é o quadrado do coeficiente de correlação múltipla da regressão acima. Um teste assintoticamente equivalente, que pode ter propriedades melhores para amostras pequenas, consiste em utilizar a estatística

$$F = \frac{(SQR_0 - SQR_1)/r}{SQR_1/(T - 2r - 1)} \sim F(r, T - 2r - 1), \quad (5.17)$$

na qual  $SQR_0 = \sum_{t=r+1}^T (X_t^2 - \bar{X})^2$  e  $SQR_1 = \sum_{t=r+1}^T \hat{u}_t^2$ , com  $\bar{X}$  a média amostral dos  $X_t^2$  e  $\hat{u}_t$  os resíduos de MQ da regressão acima. Se o valor de  $F$  for significativo, dizemos que há heteroscedasticidade condicional na série.

Dada a forma (5.7) de modelarmos a volatilidade e dado que  $X_t^2$  é um estimador (não-viesado) de  $h_t$ , o valor atual do quadrado do retorno depende de quadrados de retornos passados, comportamento similar a de um modelo auto-regressivo. Segue-se que a função de auto-correlação parcial de  $X_t^2$  pode ser usada para encontrar a ordem  $r$  de um modelo ARCH( $r$ ).

### Estimação

Os estimadores dos parâmetros do modelo são obtidos pelo método de máxima verossimilhança condicional. A função de verossimilhança é dada por

$$L(x_1, \dots, x_T | \alpha) = f(x_T | \mathcal{F}_{T-1}) f(x_{T-1} | \mathcal{F}_{T-2}) \cdots f(x_{r+1} | \mathcal{F}_r) f(x_1, \dots, x_r | \alpha),$$

e supondo normalidade dos  $\varepsilon_t$  podemos escrever

$$L(x_1, \dots, x_T | \alpha) = \prod_{t=r+1}^T (\sigma_t \sqrt{2\pi})^{-1} \exp\left\{\frac{-x_t^2}{2\sigma_t^2}\right\} f(x_1, \dots, x_r | \alpha).$$

Para  $T$  grande,  $f(x_1, \dots, x_r | \alpha)$  pode ser desprezado. Veja Engle (1982) para detalhes. Logo temos que maximizar a função de verossimilhança condicional

$$L(x_{r+1}, \dots, x_T | \alpha, x_1, \dots, x_r) = \prod_{t=r+1}^T (\sigma_t \sqrt{2\pi})^{-1} \exp\left\{\frac{-x_t^2}{2\sigma_t^2}\right\}, \quad (5.18)$$

onde a volatilidade  $\sigma_t^2 = h_t$  é obtida recursivamente.

No caso particular de um modelo ARCH(1), temos

$$L(x_2, \dots, x_T | \alpha_0, \alpha_1, x_1) = f(x_T | x_{T-1}) f(x_{T-1} | x_{T-2}) \cdots f(x_2 | x_1),$$

em que

$$(X_t | X_{t-1}) \sim \mathcal{N}(0, h_t),$$

e  $h_t = \alpha_0 + \alpha_1 X_{t-1}^2$ . Segue-se que

$$L(x_2, \dots, x_T | \alpha_0, \alpha_1, x_1) = (2\pi)^{-T/2} \prod_{t=2}^T (\alpha_0 + \alpha_1 x_{t-1}^2)^{-1/2} \exp\left\{\frac{-x_t^2}{2(\alpha_0 + \alpha_1 x_{t-1}^2)}\right\}.$$

A log-verossimilhança fica

$$\ell(x_2, \dots, x_T | \alpha_0, \alpha_1, x_1) \propto -\frac{1}{2} \sum_{t=2}^T \ln(\alpha_0 + \alpha_1 x_{t-1}^2) - \frac{1}{2} \sum_{t=2}^T \left( \frac{x_t^2}{\alpha_0 + \alpha_1 x_{t-1}^2} \right). \quad (5.19)$$

Se  $\varepsilon_t \sim t_\nu$  ( $t$  de Student com  $\nu$  graus de liberdade), a verossimilhança pode ser também escrita explicitamente. Em ambas as situações, algum procedimento não-linear de otimização numérica terá que ser utilizado.

Os programas S+FinMetrics, EViews, RATS e PcGIVE, dentre outros, podem ser usados para estimar modelos ARCH (e GARCH em geral).

#### Verificação

Para um modelo ARCH(r), com  $\varepsilon_t$  normal ou  $t$ -Student, os resíduos padroniza-

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

são v.a. i.i.d. com distribuição normal padrão ou  $t$ -Student. Logo, uma maneira de verificar se o modelo é adequado é calcular a estatística  $Q$  de Ljung-Box, dada por (3.24), para a sequência  $\tilde{X}_t$ . Além disso, podemos calcular os coeficientes de assimetria e curtose estimados e fazer um gráfico  $Q \times Q$  para avaliar a suposição de normalidade.

Para se verificar se ainda existe heteroscedasticidade condicional nos resíduos, pode-se aplicar o teste ML para a sequência  $\tilde{X}_t^2$ .

### Previsão

As previsões para a volatilidade utilizando o modelo ARCH(r) dado em (5.6)-(5.7) são obtidas recursivamente. Assim,

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 X_t^2 + \dots + \alpha_r X_{t-r+1}^2, \quad (5.20)$$

é a previsão de  $h_{t+1}$ , com origem fixada no instante  $t$ . As previsões  $\ell$  passos a frente, com origem em  $t$ , são dadas por

$$\hat{h}_t(\ell) = \alpha_0 + \sum_{i=1}^r \alpha_i \hat{h}_t(\ell - i), \quad (5.21)$$

em que  $\hat{h}_t(\ell - i) = X_{t+\ell-i}^2$ , se  $\ell - i \leq 0$ .

**Exemplo 5.5.** Vamos ajustar um modelo ARCH aos retornos diários da Petrobrás, série esta já estudada no exemplo 3.6. Lá ajustamos um modelo AR(9), com termos de primeira, terceira e nona ordens significativos. Veja os exemplos 3.11 e 3.13. As três primeiras auto-correlações parciais não nulas dos quadrados dos retornos, da Figura 5.1, sugerem um modelo ARCH(3).

Um primeiro modelo proposto, então, é

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_3 X_{t-3} + \phi_9 X_{t-9} + a_t, \\ a_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \alpha_3 a_{t-3}^2. \end{aligned}$$

(a) Assumindo  $\varepsilon_t \sim \mathcal{N}(0, 1)$  e utilizando o programa EViews, obtemos que  $\phi_3$  e  $\phi_9$  não são significativamente diferentes de zero. Re-estimando o modelo, somente com o termo auto-regressivo de primeira ordem para representar a parte linear do modelo, obtemos o seguinte modelo ajustado:

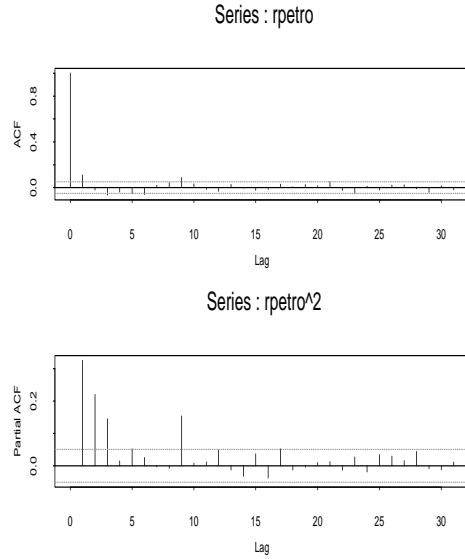


Figura 5.1: F.a.c. dos retornos e f.a.c.p. dos quadrados dos retornos da Petrobrás

$$\begin{aligned}
 X_t &= 0,1604X_{t-1} + a_t, \\
 a_t &= \sqrt{h_t}\varepsilon_t, \\
 h_t &= 0,0004 + 0,1938a_{t-1}^2 + 0,2373a_{t-2}^2 + 0,2708a_{t-3}^2,
 \end{aligned} \tag{5.22}$$

com todos os coeficientes significativamente diferentes de zero (p-valor, indicado por Prob= 0,0000). Os resultados estão no Quadro 5.1.

O Quadro 5.2 apresenta a f.a.c. e f.a.c.p. dos resíduos do modelo, com as estatísticas de Ljung-Box, com p-valores maiores do que 0,360, indicando

que o modelo é adequado para modelar a dependência linear entre sucessivos retornos. O Quadro 5.3 apresenta a f.a.c. e f.a.c.p. dos quadrados dos resíduos, mostrando que estes não apresentam mais heteroscedasticidade condicional. Note que os gráficos das auto-correlações e auto-correlações parciais não mostram os valores, por serem muito pequenos.

É fácil ver que a aplicação de um teste ML à seqüência dos quadrados dos resíduos do modelo (5.22) conduz à mesma conclusão sobre a validade do modelo ajustado.

	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	0.160398	0.022428	7.151633	0.0000
Variance Equation				
C	0.000391	2.05E-05	19.05547	0.0000
ARCH(1)	0.193785	0.031517	6.148615	0.0000
ARCH(2)	0.237373	0.021546	11.01706	0.0000
ARCH(3)	0.270795	0.033201	8.156154	0.0000
R-squared	0.009026	Mean dependent var		0.001134
Adjusted R-squared	0.006369	S.D. dependent var		0.034517
S.E. of regression	0.034407	Akaike info criterion		-4.205107
Sum squared resid	1.766285	Schwarz criterion		-4.187367
Log likelihood	3152.522	Durbin-Watson stat		2.088239
Inverted AR Roots	.16			

Quadro 5.1: Ajuste de um modelo AR(1)-ARCH(3) aos retornos da Petrobrás, erros normais.

A Figura 5.2 apresenta a estimativa da volatilidade dada pelo modelo,  $\sqrt{h_t}$ . Analisando essa figura, observamos que os cinco maiores picos (em ordem cronológica) na volatilidade estimada correspondem a: (a)  $t \approx 50$ , março de 1995, durante a crise do México; (b)  $t \approx 708$ , outubro de 1997, período final da crise na Ásia; (c)  $t \approx 920$ , setembro de 1998, após a moratória na Rússia; (d)  $t \approx 1010$ , janeiro de 1999, desvalorização do Real e (e)  $t \approx 1320$ , abril de 2000, queda da bolsa Nasdaq.

(b) Vamos supor, agora, que  $\varepsilon_t \sim t_\nu$ . Temos no Quadro 5.4, obtido pelo S+FinMetrics, os resultados do ajuste do mesmo modelo:

$$\begin{aligned} X_t &= 0,1443X_{t-1} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= 0,0004 + 0,2449a_{t-1}^2 + 0,2275a_{t-2}^2 + 0,2434a_{t-3}^2. \end{aligned} \quad (5.23)$$

As estatísticas usadas antes mostram que este modelo ajustado é adequado. Para comparar este modelo ajustado com o modelo (5.22), devemos comparar os valores AIC e BIC. No caso do Quadro 5.4, devemos dividir os valores apresentados pelo número efetivo de observações, 1498, para obter valores comparáveis com aqueles do Quadro 5.1. Os valores corrigidos são

$$\text{AIC}(6) = -4,2822, \quad \text{BIC}(6) = -4,2609,$$

que são menores do que os fornecidos quando consideramos os erros normais. A estimativa da volatilidade é mostrada na Figura 5.3, que é bastante semelhante à Figura 5.2, que considera inovações gaussianas.

Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob
				1 0.005	0.005	0.0358	
				2 0.007	0.007	0.1200	0.729
				3 -0.021	-0.021	0.7751	0.679
				4 -0.015	-0.015	1.1094	0.775
				5 -0.042	-0.042	3.7956	0.434
				6 -0.026	-0.026	4.7919	0.442
				7 0.014	0.014	5.0696	0.535
				8 0.027	0.025	6.1754	0.519
				9 0.027	0.024	7.2430	0.511
				10 0.042	0.040	9.8782	0.360
				11 0.008	0.007	9.9866	0.442
				12 0.001	0.003	9.9885	0.531
				13 0.046	0.052	13.213	0.354
				14 0.003	0.008	13.227	0.430
				15 -0.027	-0.023	14.299	0.428
				16 -0.010	-0.007	14.462	0.491
				17 0.013	0.013	14.712	0.546
				18 0.007	0.007	14.778	0.611
				19 0.017	0.017	15.241	0.645
				20 0.026	0.021	16.263	0.640
				21 0.042	0.038	19.003	0.522
				22 -0.017	-0.018	19.423	0.558
				23 -0.019	-0.019	19.955	0.586
				24 0.008	0.012	20.049	0.639

Quadro 5.2: F.a.c. e f.a.c.p. dos resíduos do modelo (5.22)

## 5.4 Modelos GARCH

Uma generalização dos modelos ARCH foi sugerida por Bollerslev (1986, 1987, 1988), o chamado modelo GARCH (“generalized ARCH”). Vimos que um modelo ARMA pode ser mais parcimonioso, no sentido de apresentar menos parâmetros do que um modelo AR ou MA puro. Do mesmo modo, um modelo GARCH pode ser usado para descrever a volatilidade com menos parâmetros do que um modelo ARCH.

**Definição 5.2.** Um modelo  $GARCH(r, s)$  é definido por

$$X_t = \sqrt{h_t} \varepsilon_t \quad (5.24)$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j}, \quad (5.25)$$

em que  $\varepsilon_t$  i.i.d.  $(0, 1)$ ,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$ ,  $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$ ,  $q = \max(r, s)$ .

Coefficientes positivos dão uma condição suficiente, mas não necessária, para que  $h_t > 0$ . Para condições gerais, veja Nelson e Cao (1992).

Como no caso de modelos ARCH, usualmente supomos que os  $\varepsilon_t$  são normais ou seguem uma distribuição  $t$  de Student, ou ainda, uma distribuição de erro generalizada.

Chamemos

$$\nu_t = X_t^2 - h_t, \quad (5.26)$$

de modo que, substituindo em (5.24) obtemos

$$X_t^2 = \alpha_0 + \sum_{i=1}^q (\alpha_i + \beta_i) X_{t-i}^2 + \nu_t - \sum_{j=1}^s \beta_j \nu_{t-j}, \quad (5.27)$$

ou seja, temos um modelo  $ARMA(q, s)$  para  $X_t^2$ , mas  $\nu_t$  não é, em geral, um processo i.i.d. Na realidade,  $\nu_t$  é uma diferença martingale, no sentido da definição 2.12, pois



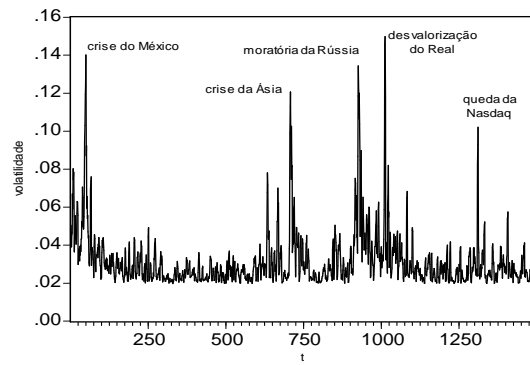


Figura 5.2: Estimativa da volatilidade dada pelo modelo (5.22)

Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob
				1	-0.005	-0.005	0.0318
				2	-0.008	-0.008	0.1293
				3	-0.028	-0.028	1.2954
				4	-0.018	-0.018	1.7829
				5	0.002	0.002	1.7922
				6	-0.002	-0.003	1.7995
				7	0.021	0.020	2.4826
				8	0.012	0.012	2.7085
				9	0.056	0.056	7.3632
				10	0.036	0.038	9.3409
				11	0.002	0.005	9.3459
				12	0.023	0.028	10.166
				13	0.038	0.043	12.396
				14	0.020	0.022	12.975
				15	0.049	0.052	16.661
				16	-0.022	-0.020	17.404
				17	0.009	0.009	17.521
				18	0.034	0.033	19.245
				19	0.021	0.017	19.894
				20	0.017	0.014	20.356
				21	0.049	0.048	24.018
				22	0.007	0.002	24.094
				23	0.016	0.014	24.500
				24	-0.009	-0.013	24.635

Quadro 5.3: F.a.c. e f.a.c.p. dos quadrados dos resíduos do modelo (5.22)

---

Mean Equation: petrobras ~ -1 + ar(1)  
 Conditional Variance Equation: ~ garch(3, 0)  
 Conditional Distribution: t with estimated parameter and standard error 0.755907

---

Estimated Coefficients:

---

Value		Std.Error	t value	Pr(> t )
AR(1)	0.14438	0.0262263	5.51	2.17e-008
A	0.00037	0.0000371	9.97	0.00e+000
ARCH(1)	0.24493	0.0557259	4.40	5.92e-006
ARCH(2)	0.22753	0.0490070	4.64	1.87e-006
ARCH(3)	0.24338	0.0512865	4.75	1.14e-006

---

AIC(6) = -6414.71 e BIC(6) = -6382.84

---

Ljung-Box test for standardized residuals:

---

Statistic	P-value	Chi^2-d.f.
10.3	0.591	12

---

Ljung-Box test for squared standardized residuals:

---

Statistic	P-value	Chi^2-d.f.
8.85	0.715	12

---

Lagrange multiplier test:

---

Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8
-0.533	-1.05	-0.736	0.0581	0.0143	0.734	0.529	2.05

---

Lag 9	Lag 10	Lag 11	Lag 12	C
1.32	0.205	1.02	-0.0887	-0.791

---

TR^2	P-value	F-stat	P-value
9.55	0.655	0.874	0.678

---

Quadro 5.4: Ajuste de um modelo AR (1) - ARCH (3) aos retornos da Petrobrás, erros  $t$ .

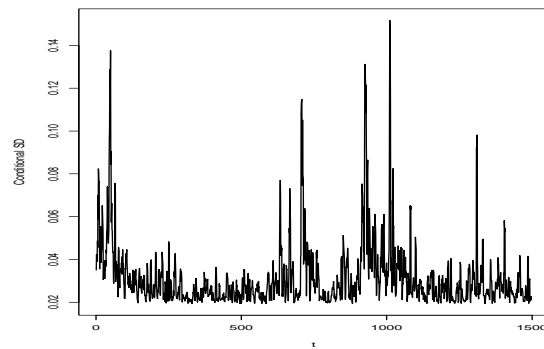


Figura 5.3: Estimativa da volatilidade dada pelo modelo (5.23)

$$E(\nu_t) = E(X_t^2 - h_t) = E(h_t \varepsilon_t^2 - h_t) = E(h_t)(E(\varepsilon_t^2) - 1) = 0,$$

para todo  $t$  e

$$E(\nu_t|\mathcal{F}_{t-1}) = E(X_t^2|\mathcal{F}_{t-1}) - E(h_t|\mathcal{F}_{t-1}) = h_t - h_t = 0,$$

Segue-se, em particular, que

$$E(X_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^q (\alpha_i + \beta_i)}.$$

A longo prazo, a volatilidade convergirá para esta média.

Um modelo bastante usado na prática é o GARCH(1,1), para o qual a volatilidade é expressa como

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 h_{t-1}, \quad (5.28)$$

com  $0 \leq \alpha_1, \beta_1 < 1, \alpha_1 + \beta_1 < 1$ .

Para os modelos GARCH temos as mesmas vantagens e desvantagens dos modelos ARCH. Volatilidades altas são precedidas de retornos ou volatilidades grandes, observando-se os grupos de volatilidades presentes em séries financeiras.

Para o modelo (5.28) obtemos facilmente

$$K = \frac{E(X_t^4)}{[E(X_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3, \quad (5.29)$$

dado que o denominador seja positivo, o que novamente mostra que se  $X_t$  segue um modelo GARCH, as caudas de  $X_t$  serão mais longas do que as da normal.

A identificação da ordem de um modelo GARCH a ser ajustado a uma série real usualmente é difícil. Recomenda-se o uso de modelos de ordem baixa, como (1,1), (1,2), (2,1) ou (2,2), e depois se escolha o modelo com base em vários critérios, como AIC ou BIC, valores da assimetria e curtose, da log-verossimilhança e de alguma função perda, como

$$\sum_{t=1}^N (X_t^2 - h_t)^2.$$

Veja Mills (1999), Pagan e Schwert (1990) e Bollerslev et al. (1994).

Os estimadores dos parâmetros do modelo (5.23)-(5.24) são obtidos pelo método de máxima verossimilhança condicional. Supondo normalidade dos  $\varepsilon_t$ , temos que a log-verossimilhança, condicional às primeiras  $r$  observações, é dada por

$$\ell(x_{r+1}, \dots, x_T | \alpha, \beta, x_1, \dots, x_r) \propto -\frac{1}{2} \sum_{t=r+1}^T \ln(h_t) - \frac{1}{2} \sum_{t=r+1}^T \frac{x_t^2}{h_t}. \quad (5.30)$$

Bollerslev (1986) utiliza em (5.30),  $h_t = \hat{\sigma}^2, t = 1, \dots, s$ , com  $\hat{\sigma}^2 = \sum_{t=1}^T X_t^2 / T$ .

As estimativas dos parâmetros são obtidas por meio de métodos numéricos de maximização.

Previsões da volatilidade, usando um modelo GARCH, podem ser calculadas de forma similar àquelas de modelo ARMA. As previsões, com origem  $t$ , considerando um modelo GARCH(1,1) da forma (5.28), são dadas por

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 X_t^2 + \beta_1 h_t,$$

e para  $\ell > 1$ ,

$$\begin{aligned} \hat{h}_t(\ell) &= \alpha_0 + \alpha_1 \hat{X}_t^2(\ell-1) + \beta_1 \hat{h}_t(\ell-1), \\ &= \alpha_0 + \alpha_1 \hat{h}_t(\ell-1) \hat{\varepsilon}_t^2(\ell-1) + \beta_1 \hat{h}_t(\ell-1), \end{aligned}$$

pois  $X_t = \sqrt{h_t} \varepsilon_t$ . Substituindo  $\hat{\varepsilon}_t^2(\ell-1)$  por  $E(\varepsilon_{t+\ell-1}^2) = 1$ , temos que

$$\hat{h}_t(\ell) = \alpha_0 + (\alpha_1 + \beta_1) \hat{h}_t(\ell-1), \quad \ell > 1. \quad (5.31)$$

Em muitas situações práticas podemos obter, por exemplo no GARCH(1,1),  $\alpha_1 + \beta_1$  próximo de um. Se a soma desses parâmetros for um, teremos o modelo IGARCH (“integrated GARCH”). Neste caso teremos

$$\begin{aligned} X_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= \alpha_0 + \beta_1 h_{t-1} + (1 - \beta_1) X_{t-1}^2, \end{aligned}$$

com  $0 < \beta_1 < 1$ . Mas nesse caso a variancia incondicional de  $X_t$  não estará definida.

**Exemplo 5.6.** Vamos ajustar um modelo GARCH à série de retornos diários do Ibovespa (veja o exemplo 1.1). A Figura 5.4 re-apresenta a f.a.c. e f.a.c.p. da série, que indicam uma dependência linear entre as observações, e que um modelo apropriado é um auto-regressivo, pois algumas auto-correlações parciais são significativas. Analisando a f.a.c. e f.a.c.p. dos quadrados dos retornos, Figura 5.5, vemos uma forte dependência, sem um padrão bem definido. Iremos, usar, então um modelo AR(10)-GARCH(1,1), com erros gaussianos. Eliminando os coeficientes não-significativos, o modelo ajustado foi (veja o Quadro 5.5)

$$\begin{aligned} X_t &= 0,0020 + 0,0530X_{t-1} - 0,0518X_{t-5} + 0,0439X_{t-10} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= 0,00002 + 0,1938a_{t-1}^2 + 0,7806h_{t-1}. \end{aligned} \quad (5.32)$$

Utilizando a equação da volatilidade, temos que a variância incondicional de  $\varepsilon_t$  é dada por

$$\widehat{\text{Var}}(a_t) = \frac{0,00002}{1 - 0,1938 - 0,7806} = 0,0008.$$

O ajustamento do modelo é apresentado no Quadro 5.5; verificamos que com o um nível de significância de 6% todos os parâmetros são significativos. Os valores da estatística de Ljung-Box para os resíduos padronizados,  $\tilde{a}_t$ , são dados por  $Q(10)=11,198$  ( $P=0,13$ ) e  $Q(20)=26,607$  ( $P=0,064$ ), enquanto que, para os quadrados dos resíduos padronizados, temos  $Q(10)=9,524$  ( $P=0,217$ ) e  $Q(20)=22,732$  ( $P=0,158$ ). Logo, podemos concluir que o modelo (5.32) é adequado para modelar a série.

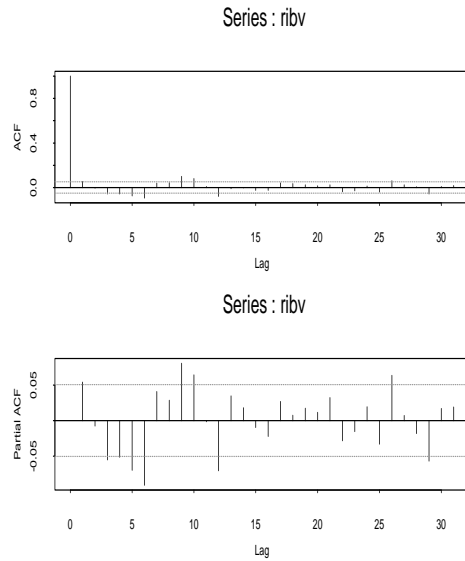


Figura 5.4: F.a.c. e f.a.c.p. dos retornos do Ibovespa

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.002026	0.000532	3.806988	0.0001
AR(1)	0.052980	0.026796	1.977185	0.0480
AR(5)	-0.051767	0.025326	-2.044036	0.0409
AR(10)	0.043848	0.023284	1.883224	0.0597
Variance Equation				
C	2.33E-05	4.79E-06	4.855218	0.0000
ARCH(1)	0.193760	0.016770	11.55393	0.0000
GARCH(1)	0.780635	0.019751	39.52430	0.0000
R-squared	0.009958	Mean dependent var		0.000887
Adjusted R-squared	0.005947	S.D. dependent var		0.028288
S.E. of regression	0.028204	Akaike info criterion		-4.758010
Sum squared resid	1.178060	Schwarz criterion		-4.733053
Log likelihood	3546.960	F-statistic		2.482635
Durbin-Watson stat	2.033701	Prob(F-statistic)		0.021527
Inverted AR Roots	.72	.61+.44i	.61-.44i	.23-.68i
	.23+.68i	-.23+.71i	-.23-.71i	-.57+.42i
	-.57-.42i	-.74		

Quadro 5.5: A justamento do modelo (5.32) aos retornos do Ibovespa

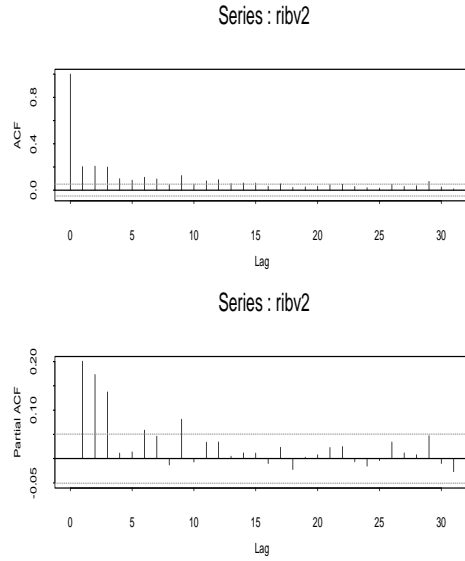


Figura 5.5: F.a.c. e f.a.c.p. dos quadrados dos retornos do Ibovespa

A aplicação do teste ML, com  $p = 12$ , para confirmar a ausência de heteroscedasticidade condicional nos resíduos do modelo (5.32) forneceu  $F = 0,9863$  ( $P = 0,459$ ). Apresentamos, na Figura 5.6, a estimativa do desvio padrão condicional ( $\sqrt{h_t}$ ) dos retornos.

Vemos, pelo gráfico, que a influência da queda da bolsa Nasdaq não aparece destacada, utilizando o modelo ajustado.

Consideremos, agora, o caso  $\varepsilon_t \sim t_\nu$ . No Quadro 5.6 temos o ajuste do modelo usando o software **S+FinMetrics**, resultando

$$\begin{aligned} X_t &= 0,0019 + 0,0625X_{t-1} - 0,0531X_{t-5} + 0,0485X_{t-10} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= 0,00002 + 0,1687a_{t-1}^2 + 0,8099h_{t-1}. \end{aligned} \tag{5.33}$$

Os valores das estatísticas de Ljung-Box (e p-valores), tanto para os resíduos como para os quadrados dos resíduos, indicam que o modelo é adequado para

descrever o comportamento da série. Comparando os valores dos critérios AIC e BIC (divididos por 1498, no caso do S+FinMetrics), concluímos que o modelo com inovações  $t$  está melhor ajustado que o modelo com inovações gaussianas. A estimativa da volatilidade está na Figura 5.7.

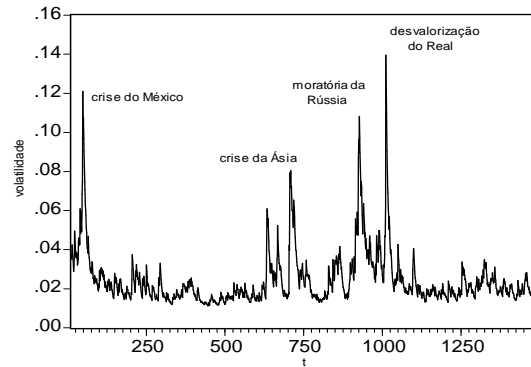


Figura 5.6: Estimativa do desvio padrão condicional para os retornos do Ibovespa usando o modelo (5.32)

## 5.5 Extensões do Modelo GARCH

Há uma literatura muito grande sobre extensões dos modelos ARCH-GARCH. Nesta seção vamos nos concentrar apenas em alguns. No que segue faremos a exposição para extensões do modelo GARCH(1,1).

### 5.5.1 Modelos EGARCH

Vimos que os modelos ARCH e GARCH tratam simetricamente os retornos, pois a volatilidade é uma função quadrática dos mesmos. Mas também é sabido que a volatilidade reage de forma assimétrica aos retornos, tendendo a ser maior para retornos negativos.



---

Mean Equation:  $ibv \sim \text{ARMA}(10, 0)$   
 Conditional Variance Equation:  $\sim \text{garch}(1, 1)$   
 Conditional Distribution:  $t$  with estimated parameter and standard error 1.53819

---

-----  
 Estimated Coefficients:

	Value	Std.Error	t value	Pr(> t )
C	0.0019444	0.0004924	3.95	4.11e-005
AR(1)	0.0625537	0.0272386	2.30	1.09e-002
AR(2)	0.0000000	NA	NA	NA
AR(3)	0.0000000	NA	NA	NA
AR(4)	0.0000000	NA	NA	NA
AR(5)	-0.0530871	0.0257115	-2.06	1.96e-002
AR(6)	0.0000000	NA	NA	NA
AR(7)	0.0000000	NA	NA	NA
AR(8)	0.0000000	NA	NA	NA
AR(9)	0.0000000	NA	NA	NA
AR(10 )	0.0485335	0.0244152	1.99	2.35e-002
A	0.0000187	0.0000056	3.34	4.26e-004
ARCH(1)	0.1686792	0.0243302	6.93	3.06e-012
GARCH(1 )	0.8098845	0.0255631	31.68	0.00e+000

---

AIC(15) = -7134.33   e   BIC(15) = -7054.65

Ljung-Box test for standardized residuals:

Statistic	P-value	Chi^2-d.f.
12.1	0.435	12

Ljung-Box test for squared standardized residuals:

Statistic	P-value	Chi^2-d.f.
16	0.19	12

Lagrange multiplier test:

Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8
2.67	1.03	-0.16	0.585	-0.328	-0.84	-1.77	-0.0923

Lag 9	Lag 10	Lag 11	Lag 12	C
-0.148	-0.46	1.15	-0.128	0.292

TR^2	P-value	F-stat	P-value
14.6	0.261	1.34	0.303

---

Quadro 5.6: Ajustamento do modelo (5.33) aos retornos do Ibovespa

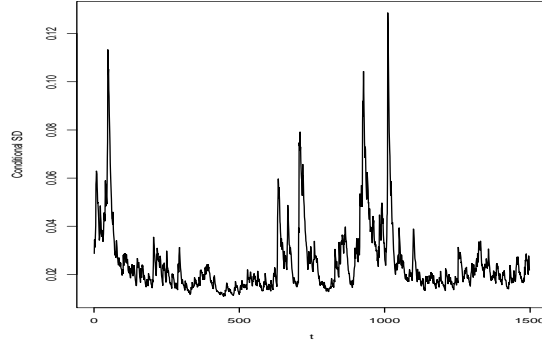


Figura 5.7: Estimativa do desvio padrão condicional para os retornos do Ibovespa usando o modelo (5.33)

Tendo em vista o exposto, Nelson (1991) introduziu os modelos EGARCH (“exponential GARCH”).

**Definição 5.3.** Um modelo  $EGARCH(1,1)$  é dado por

$$X_t = \sqrt{h_t} \varepsilon_t, \quad (5.34)$$

$$\ln(h_t) = \alpha_0 + \alpha_1 g(\varepsilon_{t-1}) + \beta_1 \ln(h_{t-1}), \quad (5.35)$$

em que  $\varepsilon_t$  são v.a. i.i.d. com média zero e variância um e  $g(\cdot)$  é a curva de impacto de informação, dada por

$$g(\varepsilon_t) = \theta \varepsilon_t + \gamma \{|\varepsilon_t| - E(|\varepsilon_t|)\}. \quad (5.36)$$

Aqui,  $\theta$  e  $\gamma$  são parâmetros reais, e  $|\varepsilon_t| - E(|\varepsilon_t|)$  é uma seqüência de v.a. i.i.d. com média zero.

Note que  $E\{g(\varepsilon_t)\} = 0$  e podemos escrever

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{se } \varepsilon_t \geq 0 \\ (\theta - \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{se } \varepsilon_t < 0. \end{cases} \quad (5.37)$$

Para que retornos negativos tenham maior impacto na volatilidade esperamos  $\gamma < 0$ . Para uma normal padrão, não é difícil ver que  $E(|\varepsilon_t|) = \sqrt{2/\pi}$ .

Esta assimetria permite que a volatilidade responda mais rapidamente a retornos negativos do que a positivos, fato este conhecido como “efeito alavancagem”.

No caso geral, a equação (5.35) para um EGARCH  $(r, s)$  fica

$$\ln(h_t) = \alpha_0 + \frac{1 + b_1 B + \dots + b_s B^s}{1 - a_1 B - \dots - a_r B^r} g(\varepsilon_{t-1}),$$

onde os operadores em  $B$  têm raízes fora do círculo unitário. Note que em (5.35) e no modelo geral usamos o logaritmo de  $h_t$ , o que implica que os coeficientes do modelo não necessitam ser não-negativos.

O modelo mais simples é o EGARCH(1,0), que pode ser escrito

$$\ln(h_t) = \alpha_0 + \frac{g(\varepsilon_{t-1})}{1 - \alpha_1 B} g(\varepsilon_{t-1}).$$

Se usarmos o EViews para estimar modelos, este software considera o modelo EGARCH na forma:

$$\ln(h_t) = w + \beta \ln(h_{t-1}) + \alpha \left| \frac{X_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \frac{X_{t-1}}{\sqrt{h_{t-1}}}. \quad (5.38)$$

O S+FinMetrics usa uma especificação similar. Quando  $\gamma \neq 0$  o efeito assimétrico deve ser incorporado ao modelo GARCH. Observe também que podemos escrever  $\varepsilon_t$  no lugar de  $X_t/\sqrt{h_t}$  nas fórmulas acima.

**Exemplo 5.7.** A série de retornos diários da Petrobrás já foi analisada no exemplo 5.5, no qual ajustamos um modelo AR(1)-ARCH(3). Um modelo AR(1)-GARCH(1,1) com erros gaussianos também pode ser ajustado, resultando

$$\begin{aligned} X_t &= 0,13173X_{t-1} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\ h_t &= 0,00003 + 0,1384a_{t-1}^2 + 0,8334h_{t-1}, \end{aligned} \quad (5.39)$$

com  $\widehat{\text{Var}}(a_t) = 0,0011$ .

Estes dois modelos tratam simetricamente os retornos. Vamos agora ajustar um modelo EGARCH(1,1), verificando se um efeito assimétrico deve ser

incorporado ao modelo anterior. O Quadro 5.7 apresenta o ajustamento do modelo AR(1)-EGARCH(1,1), usando o EViews

$$\begin{aligned} X_t &= 0,1576X_{t-1} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\ \ln h_t &= -0,5231 + 0,9496 \ln h_{t-1} + 0,2196|\varepsilon_{t-1}| - 0,1420\varepsilon_{t-1}, \end{aligned} \quad (5.40)$$

A análise dos resultados revela que todos os parâmetros do modelo são significativos ( $P = 0,0000$ ); isto significa que  $\gamma \neq 0$  e que o efeito assimétrico deve ser incorporado ao modelo.

O teste de Ljung-Box aplicado aos resíduos padronizados resulta  $Q(10) = 9,789$  ( $P = 0,368$ ) e  $Q(20) = 15,973$  ( $P = 0,659$ ), e aos quadrados dos resíduos padronizados resulta  $Q(10) = 7,905$  ( $P = 0,544$ ) e  $Q(20) = 12,323$  ( $P = 0,871$ ). Esses resultados indicam que os resíduos não são correlacionados e não apresentam heteroscedasticidade condicional. Essa última afirmação pode também ser comprovada pela aplicação do teste ML aos quadrados dos resíduos, obtendo-se  $F = 0,6891$  ( $P = 0,7634$ ).

	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	0.157626	0.025427	6.199105	0.0000
Variance Equation				
C	-0.523085	0.051894	-10.07983	0.0000
RES /SQR[GARCH](1)	0.219585	0.019876	11.04753	0.0000
RES/SQR[GARCH](1)	-0.142020	0.012840	-11.06066	0.0000
EGARCH(1)	0.949623	0.006794	139.7649	0.0000
R-squared	0.009289	Mean dependent var		0.001134
Adjusted R-squared	0.006633	S.D. dependent var		0.034517
S.E. of regression	0.034402	Akaike info criterion		-4.293177
Sum squared resid	1.765815	Schwarz criterion		-4.275438
Log likelihood	3218.443	Durbin-Watson stat		2.082962
Inverted AR Roots	.16			

Quadro 5.7: Ajustamento do modelo (5.40) aos retornos da Petrobrás

A estimativa do desvio padrão condicional encontra-se na Figura 5.8. Vemos que esta não revela a influência da queda da Nasdaq, em abril de 2000.

Para comparar os três modelos, (5.22), (5.37) e (5.38), ajustados aos retornos da Petrobrás, apresentamos na Tabela 5.1, os valores de alguns critérios de ajustamentos, a saber, AIC, BIC e log-verossimilhança. Por esses resultados, podemos dizer que o modelo AR(1)-EGARCH(1,1) ajusta melhor a série, pois apresenta o max (log-verossimilhança), min(AIC) e min(BIC).

Tabela 5.1: Valores mínimos de AIC, BIC e máximos da log-verossimilhança para os modelo ajustados aos retornos da Petrobrás.

Modelo	AR(1)+ ARCH(3)	AR(1)+ GARCH(1,1)	AR(1)+ EGARCH(1,1)
log-verossim.	3152,52	3181,79	3218,44
AIC	-4,205	-4,245	-4,293
BIC	-4,187	-4,231	-4,275

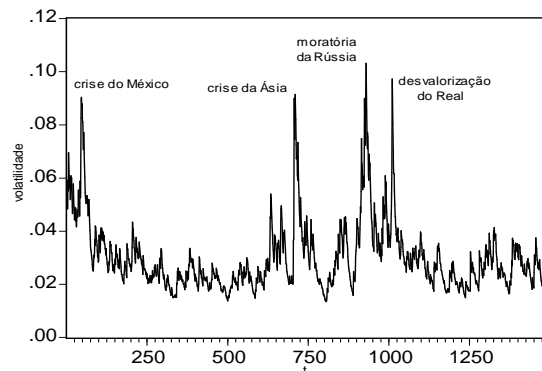


Figura 5.8: Estimativa do desvio padrão condicional para os retornos da Petrobrás usando o modelo (5.40)

### 5.5.2 Modelos TARCH

O modelo TARCH (“threshold ARCH”) é um caso particular do modelo ARCH não-linear, e a volatilidade agora segue a forma funcional

$$h_t^\gamma = \alpha_0 + \alpha_1 g^{(\gamma)}(\varepsilon_{t-1}) + \beta_1 h_{t-1}^\gamma,$$

em que

$$g^{(\gamma)}(\varepsilon_t) = \theta I_{\{\varepsilon_t > 0\}} |\varepsilon_t|^\gamma + (1 - \theta) I_{\{\varepsilon_t \leq 0\}} |\varepsilon_t|^\gamma.$$

Para  $\gamma = 1$  temos o modelo de Zakoian (1994) e para  $\gamma = 2$  o modelo GJR (de Glosten, Jagannathan and Runkle, 1993).

O EViews e o S+FinMetrics usam a formulação

$$h_t = w + \alpha X_{t-1}^2 + \gamma X_{t-1}^2 d_{t-1} + \beta h_{t-1},$$

com

$$d_t = \begin{cases} 1, & \text{se } X_t < 0 \text{ ("bad news")} \\ 0, & \text{se } X_t \geq 0 \text{ ("good news")}. \end{cases}$$

Se  $\gamma \neq 0$ , há um impacto de informação assimétrica. Esperamos que  $\gamma > 0$  para que “bad news” tenha um impacto maior.

**Exemplo 5.8.** Considere a série de retornos diários do Banespa, no mesmo período das séries anteriores. Não é difícil verificar que um modelo MA(1)-ARCH(2) é adequado para descrever esta série. Vamos verificar se existe impacto de informação assimétrica, ajustando um modelo MA(1)-TARCH(1,1) com erros gaussianos aos retornos. Os resultados estão no Quadro 5.8. O modelo ajustado é

$$\begin{aligned} X_t &= a_t - 0,0686a_{t-1}, \\ a_t &= \sqrt{h_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\ h_t &= 0,00042 + 0,1660a_{t-1}^2 + 0,0993a_{t-1}^2 d_{t-1} + 0,5870h_{t-1}, \end{aligned} \tag{5.41}$$

com  $d_t$  dado acima.

Analisando os resultados no Quadro 5.8, notamos que todos os parâmetros são significativos, incluindo o parâmetro  $\gamma$ , indicando presença de informação assimétrica. Os testes Ljung-Box e ML indicam que as suposições do modelo estão verificadas. Na Figura 5.9 está representada o desvio padrão condicional,

$\sqrt{h_t}$ . Uma comparação dos critérios AIC, BIC e log-verossimilhança mostra que o modelo assimétrico ajusta-se melhor que o simétrico.

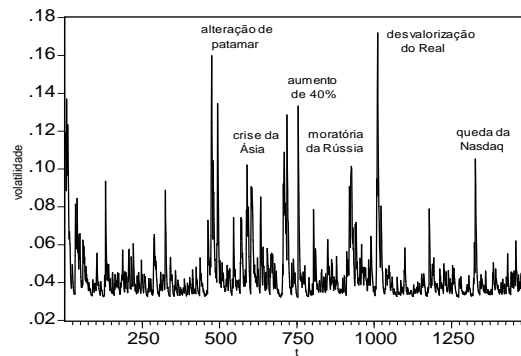


Figura 5.9: Estimativa do desvio padrão condicional para os retornos do Banespa usando o modelo (5.41)

	Coefficient	Std. Error	z-Statistic	Prob.
MA(1)	0.068916	0.029476	2.338048	0.0194
Variance Equation				
C	0.000425	3.16E-05	13.47387	0.0000
ARCH(1)	0.166031	0.017633	9.415653	0.0000
(RESID<0)*ARCH(1)	0.099313	0.032932	3.015696	0.0026
GARCH(1)	0.586986	0.027408	21.41671	0.0000
R-squared	0.012086	Mean dependent var		0.001403
Adjusted R-squared	0.009439	S.D. dependent var		0.047593
S.E. of regression	0.047367	Akaike info criterion		-3.480284
Sum squared resid	3.349798	Schwarz criterion		-3.462554
Log likelihood	2611.732	Durbin-Watson stat		1.873088
Inverted MA Roots	-.07			

Quadro 5.8: A justamento do modelo (5.41) aos retornos do Banespa

## 5.6 Modelos de Volatilidade Estocástica

Os modelos da família ARCH supõem que a variância condicional depende de retornos passados. O modelo de volatilidade estocástica (MVE), primeiro proposto por Taylor (1980, 1986) não faz essa suposição. Este modelo tem como premissa o fato que a volatilidade presente depende de seus valores passados, mas é independente dos retornos passados.

Mudando um pouco a notação, denotemos agora a variância condicional por  $\sigma_t^2$ , ou seja,  $\sigma_t^2 = E(X_t^2 | \mathcal{F}_{t-1})$ .

**Definição 5.4.** Dizemos que a série  $X_t$  segue um modelo de volatilidade estocástica se

$$X_t = \sigma_t \varepsilon_t, \quad (5.42)$$

$$\sigma_t = e^{h_t/2}, \quad (5.43)$$

em que  $\varepsilon_t$  é uma seqüência estacionária, com média zero e variância um, e  $h_t$  é uma seqüência que pode se estacionária ou não, com uma densidade de probabilidade  $f(h)$ .

A formulação mais simples do modelo supõe que o logaritmo da volatilidade,  $h_t$ , seja dado por

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad (5.44)$$

na qual  $\eta_t$  é uma seqüência estacionária gaussiana, de média zero e variância  $\sigma_\eta^2$ , independente de  $\varepsilon_t$ . Segue-se que devemos ter  $|\alpha_1| < 1$ .

Outras formulações do MVE foram apresentadas na literatura, dentre as quais destacamos as seguintes.

(1) Forma canônica de Kim et al. (1998).

Aqui, o MVE é escrito na forma

$$X_t = \beta e^{h_t/2} \varepsilon_t, \quad (5.45)$$

$$h_{t+1} = \mu + \alpha_1(h_t - \mu) + \sigma_\eta \eta_t, \quad (5.46)$$



com

$$h_t \sim \mathcal{N}\left(\mu, \frac{\sigma_\eta^2}{1 - \alpha_1^2}\right),$$

sendo  $\varepsilon_t, \eta_t \sim \mathcal{N}(0, 1)$ , e se  $\beta = 1$ , então  $\mu = 0$ .

(2) Formulação de Jaquier et al. (1994), na qual

$$X_t = \sqrt{h_t} \varepsilon_t, \quad (5.47)$$

$$\ln(h_t) = \alpha_0 + \alpha_1 \ln(h_{t-1}) + \sigma_\eta \eta_t. \quad (5.48)$$

Sabemos que se  $\varepsilon_t \sim \mathcal{N}(0, 1)$ , então  $\ln(\varepsilon_t^2)$  tem uma distribuição chamada *log-qui-quadrado*, de tal sorte que

$$\begin{aligned} E(\ln(\varepsilon_t^2)) &\simeq -1,27 \\ \text{Var}(\ln(\varepsilon_t^2)) &= \pi^2/2. \end{aligned}$$

De (5.42) e (5.44) obtemos

$$\ln(X_t^2) = \ln(\sigma_t^2) + \ln(\varepsilon_t^2), \quad (5.49)$$

$$h_t = \ln(\sigma_t^2) = \alpha_0 + \alpha_1 h_{t-1} + \eta_t. \quad (5.50)$$

Chamando  $\xi_t = \ln(\varepsilon_t^2) - E(\ln(\varepsilon_t^2)) \simeq \ln(\varepsilon_t^2) + 1,27$ , temos que  $E(\xi_t) = 0$ ,  $\text{Var}(\xi_t) = \pi^2/2$  e

$$\ln(X_t^2) = -1,27 + h_t + \xi_t, \quad \xi_t \sim \text{i.i.d. } (0, \pi^2/2), \quad (5.51)$$

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\eta^2). \quad (5.52)$$

Aqui, supomos  $\xi_t$  e  $\eta_t$  independentes.

#### Propriedades

Vamos calcular agora alguns parâmetros associados ao MVE, considerando-

se a forma (5.44):

(i)  $E(X_t) = E(\sigma_t \varepsilon_t) = E(\sigma_t)E(\varepsilon_t) = 0$ , dado que  $\sigma_t$  e  $\varepsilon_t$  so independentes.

(ii)  $\text{Var}(X_t) = E(X_t^2) = E(\sigma_t^2 \varepsilon_t^2) = E(\sigma_t^2)E(\varepsilon_t^2) = E(\sigma_t^2)$ .

Dado que supusemos  $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$ , e  $h_t$  estacionário, com  $\mu_h = E(h_t) = \alpha_0/(1 - \alpha_1)$ ,  $\sigma_h^2 = \text{Var}(h_t) = \sigma_\eta^2/(1 - \alpha_1^2)$ , então obtemos

$$h_t \sim \mathcal{N}\left(\frac{\alpha_0}{1 - \alpha_1}, \frac{\sigma_\eta^2}{1 - \alpha_1^2}\right). \quad (5.53)$$

Como  $h_t$  é normal,  $\sigma_t^2$  é log-normal, logo temos

$$E(X_t^2) = E(\sigma_t^2) = e^{\mu_h + \sigma_h^2/2}.$$

Não é difícil mostrar que

$$E(X_t^4) = 3e^{2\mu_h + 2\sigma_h^2},$$

da qual obtemos a curtose

$$K = \frac{3e^{2\mu_h + 2\sigma_h^2}}{e^{2\mu_h + \sigma_h^2}} = 3e^{\sigma_h^2} > 3, \quad (5.54)$$

como deveríamos esperar, ou seja, caudas longas sob o MVE.

(iii) A função de auto-covariância da série  $X_t$  é dada por

$$\gamma_X(u) = E(X_t X_{t+u}) = E(\sigma_t \sigma_{t+u} \varepsilon_t \varepsilon_{t+u}) = 0,$$

pois  $\varepsilon_t$  e  $\eta_t$  são independentes. Logo  $X_t$  é serialmente não-correlacionada, mas não independente, pois existe correlação em  $\ln(X_t^2)$ . Denotando-se  $Y_t = \ln(X_t^2)$ , então a auto-covariância de  $Y_t$  é dada por

$$\gamma_Y(u) = E[(Y_t - E(Y_t))(Y_{t+u} - E(Y_{t+u}))].$$

Como o primeiro termo entre parênteses igual a  $h_t - E(h_t) + \xi_t$  e  $h_t$  é independente de  $\xi_t$ , obtemos que

$$\begin{aligned}\gamma_Y(u) &= E[(h_t - E(h_t) + \xi_t)(h_{t+u} - E(h_{t+u}) + \xi_{t+u})] \\ &= E[(h_t - E(h_t))(h_{t+u} - E(h_{t+u})) + E(\xi_t \xi_{t+u})],\end{aligned}$$

e chamando as auto-covariâncias do segundo membro de  $\gamma_h(\cdot)$  e  $\gamma_\xi(\cdot)$ , respectivamente, teremos

$$\gamma_Y(u) = \gamma_h(u) + \gamma_\xi(u),$$

para todo  $u$ .

Como estamos supondo (5.44), ou seja, um AR(1), temos que

$$\gamma_h(u) = \alpha_1^u \frac{\sigma_\eta^2}{1 - \alpha_1^2}, \quad u > 0,$$

enquanto que  $\gamma_\xi(u) = 0$ , para  $u > 0$ . Logo,  $\gamma_Y(u) = \gamma_h(u)$ , para todo  $u \neq 0$ , e podemos escrever a função de auto-correlação de  $Y_t$  como

$$\rho_Y(u) = \frac{\gamma_Y(u)}{\gamma_Y(0)} = \frac{\alpha_1^u \sigma_\eta^2 / (1 - \alpha_1^2)}{\gamma_h(0) + \gamma_\xi(0)}, \quad u > 0,$$

do que obtemos

$$\rho_Y(u) = \frac{\alpha_1^u}{1 + \pi^2 / 2\sigma_h^2}, \quad u > 0, \quad (5.55)$$

que tende a zero exponencialmente a partir do lag 2, o que indica que  $Y_t = \ln(X_t^2)$  pode ser modelada por um modelo ARIMA(1, 1).

Na prática, obtemos valores de  $\alpha_1$  próximos de um, o que implica o aparecimento de altas correlações para volatilidades e conseqüentes grupos de volatilidades na série.

Um MVE geral será obtido admitindo-se um modelo AR(p) para  $h_t$ :

$$X_t = \sigma_t \varepsilon_t, \quad (5.56)$$

$$(1 - \alpha_1 B - \dots - \alpha_p B^p)h_t = \alpha_0 + \eta_t, \quad (5.57)$$

com as suposições anteriores sobre as inovações, e agora supondo-se que as raízes do polinômio  $1 - \alpha_1 B - \dots - \alpha_p B^p$  estejam fora do círculo unitário.

MVE foram estendidos para incluir o fato que a volatilidade tem memória longa, no sentido que a função de auto-correlação de  $\ln(X_t^2)$  decai lentamente, embora, como vimos, os  $X_t$  não tenham correlação serial.

#### Estimação

Os MVE são difíceis de estimar. Podemos usar a abordagem de Durbin e Koopman (1997a, 1997b, 2000), que consiste em usar o procedimento de quase-verossimilhança, por meio do Filtro de Kalman. Aqui, o modelo (5.42)-(5.43) é reescrito na forma

$$X_t = \sigma \varepsilon_t e^{h_t/2}, \quad (5.58)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (5.59)$$

em que  $\sigma = \exp\{\alpha_0/2\}$ . Uma forma equivalente é dada por

$$\ln(X_t^2) = \kappa + h_t + u_t, \quad (5.60)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (5.61)$$

em que  $u_t = \ln(\varepsilon_t^2) - E(\ln(\varepsilon_t^2))$  e  $\kappa = \ln(\sigma^2) + E(\ln(\varepsilon_t^2))$ .

As equações (5.60)-(5.61) estão na formulação denominada de espaço de estados: (5.60) é a *equação de observação* e (5.61) é a *equação de estado*.

#### Observações:

- (i) Quando  $\phi$  for próximo de 1, o ajustamento de um MVE é similar ao de um GARCH(1,1), com  $\alpha_1 + \beta_1$  próximo de 1.
- (ii) Quando  $\phi = 1$ ,  $h_t$  é um passeio aleatório e o ajustamento de um MVE é similar ao de um modelo IGARCH(1,1).
- (iii) Quando algumas observações forem iguais a zero, o que pode ocorrer na prática, não podemos fazer a transformação logaritmica especificada em (5.60). Uma solução sugerida por Fuller e analisada por Breidt e Carriquiry (1996) é fazer a seguinte transformação baseada numa expansão de Taylor:

$$\ln(X_t^2) = \ln(X_t^2 + cS_X^2) - cS_X^2/(X_t^2 + cS_X^2), \quad t = 1, \dots, T,$$

em que  $S_X^2$  é a variância amostral da série  $X_t$  e  $c$  é um número pequeno.

O programa STAMP (Koopman et al., 1995) pode ser utilizado na obtenção de estimadores de quase-verossimilhança (QMV) dos parâmetros do modelo, escrito na forma de espaço de estados. Este programa incorpora a transformação acima, com um valor “default”  $c = 0,02$ . Uma das vantagens da utilização do procedimento de QMV é que ele pode ser aplicado sem a especificação de uma particular distribuição para  $\varepsilon_t$ .

Shephard e Pitt (1997) propuseram o uso de amostragem ponderada (“importance sampling”) para estimar a função de verossimilhança.

Como o MVE é um modelo hierárquico, Jaquier et al. (1994) propuseram uma análise bayesiana para o mesmo. Veja também Shephard e Pitt (1997) e Kim et al. (1998). Uma resenha do problema de estimação do MVE é feita por Motta (2001).

**Exemplo 5.9.** Vamos re-analisar a série de retornos diários do Ibovespa usando o MVE e programa STAMP. O Quadro 5.9 apresenta os resultados da estimação do modelo, fornecendo

$$\begin{aligned} \ln(X_t^2) &= -8,6220 + h_t + u_t, \\ h_t &= 0,9858h_{t-1} + \eta_t, \\ \hat{\text{Var}}(u_t) &= 2,7418, \\ \hat{\text{Var}}(\eta_t) &= 0,0236. \end{aligned} \tag{5.62}$$

A f.a.c. residual, juntamente com o teste de Ljung-Box estão mostrados no Quadro 5.10, sugerindo que os resíduos do modelo são não-correlacionados. Uma análise residual mais completa é fornecida pela Figura 5.10. A estimativa da volatilidade ( $e^{h_t}$ ) e do desvio padrão condicional é apresentada na Figura 5.11, revelando as influências das crises do México e da Ásia, da moratória na Rússia e da desvalorização do Real, que já haviam sido detectadas quando usamos o modelo AR(10)-GARCH(1,1). Vale a pena observar que  $\hat{\phi} = 0,9858$ , o que indica uma equivalência de ajustamento entre os dois modelos.

---

```

Method of estimation is Maximum likelihood
The present sample is: 1 to 1498

SVretornos = Level + AR(1) + Irregular
Log-Likelihood is -817.302 (-2 LogL = 1634.6).
Prediction error variance is 2.96951

Estimated variances of disturbances.

Component      SVretornos (q-ratio)
Irr              2.7418 ( 1.0000)
Arl              0.023645 ( 0.0086)

Estimated autoregressive coefficient.
The AR(1) rho coefficient is 0.985828.

Estimated coefficients of final state vector.

Variable      Coefficient      R.m.s.e.      t-value
Lvl           -8.6220      0.27296      -31.587 [ 0.0000]
Arl            0.12543      0.51580

Goodness-of-fit results for Residual SVretornos
Prediction error variance (p.e.v)      2.969511
Prediction error mean deviation (m.d)  2.427108
Ratio p.e.v. / m.d in squares          0.952954
Coefficient of determination            R2      0.140840
... based on differences                RD2     0.447572
Information criterion of Akaike         AIC      1.092403
... of Schwartz (Bayes)                 BIC      1.103041

```

---

Quadro 5.9: A justamento de um MVE aos retornos do Ibovespa

## 5.7 Problemas

1. Ajuste modelos ARCH com erros gaussianos para as séries;
  - (a) log-retornos diários das ações do Banespa.
  - (b) log-retornos diários do índice Ibovespa.
  - (c) log-retornos mensais do S& P500, de janeiro de 1962 a dezembro de 1999 (arquivo m-sp62.99.dat).

---

Serial correlation statistics for Residual SVretornos.  
 Durbin-Watson test is 1.95137.  
 Asymptotic deviation for correlation is 0.0258457.

Lag	dF	SerCorr	BoxLjung	ProbChi2(dF)
1	0	0.0229		
2	0	0.0068		
3	0	0.0278		
4	1	0.0207	2.6609	[ 0.1028]
5	2	-0.0126	2.8982	[ 0.2348]
6	3	-0.0463	6.1222	[ 0.1058]
7	4	0.0184	6.6310	[ 0.1567]
8	5	-0.0510	10.5570	[ 0.0609]
9	6	-0.0401	12.9848	[ 0.0433]
10	7	-0.0108	13.1616	[ 0.0683]
11	8	-0.0204	13.7889	[ 0.0874]
12	9	0.0123	14.0161	[ 0.1218]
13	10	0.0051	14.0553	[ 0.1705]
14	11	0.0393	16.3964	[ 0.1270]
15	12	0.0112	16.5867	[ 0.1658]
16	13	-0.0211	17.2635	[ 0.1875]
17	14	-0.0168	17.6919	[ 0.2212]
18	15	0.0185	18.2104	[ 0.2518]
19	16	-0.0345	20.0160	[ 0.2195]
20	17	0.0121	20.2385	[ 0.2622]
21	18	-0.0101	20.3945	[ 0.3111]
22	19	0.0027	20.4053	[ 0.3706]
23	20	-0.0152	20.7587	[ 0.4115]
24	21	-0.0084	20.8656	[ 0.4672]
25	22	0.0066	20.9318	[ 0.5250]
26	23	-0.0258	21.9462	[ 0.5235]
27	24	-0.0212	22.6352	[ 0.5414]
28	25	0.0097	22.7780	[ 0.5905]
29	26	0.0400	25.2185	[ 0.5066]
30	27	-0.0443	28.2142	[ 0.4000]

---

Quadro 5.10 - F.a.c. dos resíduos do modelo (5.60) e estatística de Ljung-Box

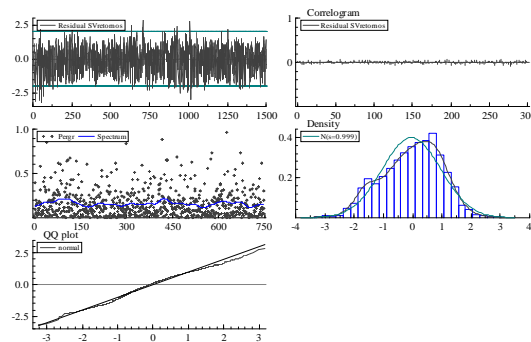


Figura 5.10: Análise residual do ajustamento do modelo (5.60) à série de retornos do Ibovespa

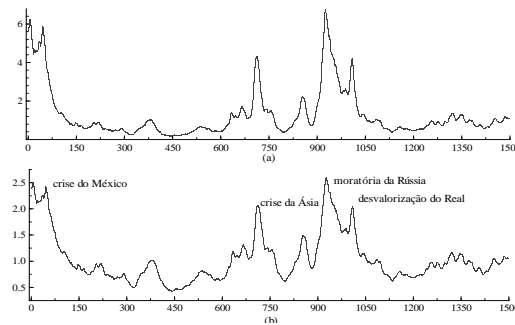


Figura 5.11: MVE para os retornos do Ibovespa (a) estimativa da volatilidade  
(b) estimativa do desvio padrão condicional

2. Ajuste modelos GARCH com erros gaussianos para as séries:
  - (a) log-retornos diários das ações da Petrobrás.
  - (b) log-retornos diários do índice Ibovespa.
3. Ajuste um modelo assimétrico (EGARCH ou TARARCH) com erros gaussianos à série de log-retornos mensais das ações da IBM de janeiro de 1926 a dezembro de 1997(arquivo m-ibm26.97.dat).
4. Ajuste um modelo de volatilidade estocástica às séries:
  - (a) log-retornos diários das ações do Banespa.
  - (b) log-retornos diários das ações da Petrobrás.
  - (c) log-retornos diários das ações da CEMIG, de 3/1/95 a 27/12/2000.
5. Obtenha as previsões de origem  $T$  e horizonte  $h$ ,  $h \geq 1$ , para um modelo GARCH(1,2).
6. Suponha que  $r_1, \dots, r_N$  sejam observações de uma série de log-retornos, seguindo um modelo AR(1)-GARCH(1,1),



$$\begin{aligned} r_t &= \mu + \phi_1 r_{t-1} + a_t, \\ a_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\ \sigma_t^2 &= \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

Obtenha a função de log-verossimilhança condicional dos dados.

7. Obtenha a expressão (5.29).
8. Refaça o problema 1, usando a distribuição  $t$  de Student para os erros.
9. Refaça o problema 3, usando a distribuição  $t$  de Student para os erros.
10. Refaça o problema 2, usando a distribuição de erro generalizada para os erros.
11. Encontre a log-verossimilhança para um modelo ARCH(1), supondo  $\varepsilon_t \sim t_\nu$ .
12. Mesmo problema anterior, para um modelo GARCH(1,1).
13. **Modelo PGARCH.** O modelo PGARCH (1,1) (de “power” ’GARCH’), proposto por Ding et al. (1993) é dado por

$$h_t^d = \alpha_0 + \alpha_1 (|X_{t-1}| + \gamma X_{t-1})^d + \beta_1 h_t^d,$$

onde  $d$  é um inteiro positivo e  $\gamma$  denota o coeficiente de alavancagem. Se  $d = 2$  e  $\gamma = 0$  temos o GARCH usual; se  $d = 2$  temos o GARCH com alavancagem. O modelo com  $d = 1$  é robusto a observações atípicas. Usando o **S+FinMetrics**:

- (a) Ajuste um modelo PGARCH (1,1), aos dados da Petrobrás.
- (b) Ajuste um modelo PGARCH (1,1), com  $d = 1$ , aos dados do Banespa.
14. Ajuste um modelo EGARCH(1,0) aos dados de retornos diários da CEMIG.
15. Determine  $E(|\varepsilon_t|)$  para uma distribuição  $t$  de Student padronizada, isto é,  $\varepsilon = X/\sqrt{\nu/(\nu-2)}$ , onde  $X \sim t_\nu$ .

## Apêndice 5. Algumas Distribuições Especiais

Além da distribuição normal e da distribuição  $t$ -Student, podemos usar outras distribuições para o erro  $\varepsilon_t$  de um modelo GARCH genérico. Nelson (1991) propôs a distribuição de erro generalizada (“generalized error distribution”, GED). Dizemos que a v.a.  $X$  segue uma GED, com média zero e variância um, se sua densidade é dada por

$$f(x) = \frac{\nu \exp[-(\frac{1}{2})|x/\lambda|^\nu]}{\lambda 2^{(\nu+1)/\nu} \Gamma(1/\nu)},$$

onde

$$\lambda = \left[ \frac{2^{-2/\nu} \Gamma(1/\nu)}{\Gamma(3/\nu)} \right]^{1/2}.$$

Aqui,  $\nu > 0$  governa o comportamento das caudas da distribuição. Se  $\nu = 2$  obtemos a distribuição normal padrão, e se  $\nu = 1$  obtemos a distribuição exponencial dupla,

$$f(x) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|}.$$

Quando  $\nu < 2$  a distribuição tem caudas mais pesadas do que a normal e quando  $\nu > 2$  tem caudas mais leves do que a normal.

Fernandez e Stell (1998) propuseram o uso da distribuição  $t$  assimétrica, cuja distribuição tem densidade

$$\begin{aligned} f(x|\iota, \nu) &= \frac{2}{\iota + 1/\iota} \left[ g(\iota(sx + m)|\nu) I_{(-\infty, 0)}(x + m/s) \right] \\ &+ \frac{2}{\iota + 1/\iota} \left[ g((sx + m)/\iota|\nu) I_{(0, +\infty)}(x + m/s) \right], \end{aligned}$$

onde  $g(\cdot|\nu)$  denota uma  $t$ -Student com  $\nu$  graus de liberdade,

$$m = \frac{\Gamma((\nu + 1)/2) \sqrt{\nu - 2}}{\sqrt{\pi} \Gamma(\nu/2)} (\iota - 1/\iota),$$

$$s = \sqrt{(\iota^2 + 1/\iota^2 - 1) - m^2},$$

sendo  $\iota$  o parâmetro de assimetria, com  $\iota = 1$  para a  $t$  tradicional.

Laurent e Peters (2002) desenvolveram um pacote na linguagem Ox para estimar modelos da família ARCH, com erros normais,  $t$ , GED e  $t$  assimétricos.



# Valor em Risco

## 6.1 Introdução

Neste capítulo estaremos interessados em calcular uma medida de um tipo particular de risco, o chamado risco de mercado. Tal medida é o **VaR** (valor em risco). O cálculo do **VaR** envolve o cálculo da volatilidade de um ativo financeiro ou de uma carteira de instrumentos financeiros. No Capítulo 1 discutimos vários aspectos relacionados com a volatilidade.

Trataremos, agora, somente de volatilidade estatística. Usaremos os modelos estudados no Capítulo 5, bem como discutiremos outros três enfoques para o cálculo do **VaR**: o já mencionado RiskMetrics, o baseado em quantis empíricos e o baseado na teoria dos valores extremos.

A abordagem baseada em modelos tipo ARCH é denominada às vezes de abordagem econométrica. A dificuldade com esta é que se torna difícil calcular o **VaR** de uma carteira com muitos ativos, pois seria necessário ajustar um modelo heteroscedástico condicional multivariado, que é uma tarefa complicada ou muitas vezes impossível. Por isso, o enfoque baseado no RiskMetrics, que supõe que se tenha um modelo de passeio casual com inovações normais, e modela a volatilidade e correlações por meio de modelos EWMA, é mais difundido e mais simples de implementar na prática.

As empresas em geral estão expostas a três classes de risco: operacional, estratégico e financeiro. Os riscos financeiros estão ligados às variações de variáveis financeiras (como juros e taxas de câmbio) que podem implicar em perdas financeiras.

Os riscos financeiros ainda podem ser de vários tipos, como operacional, de crédito, de liquidez, legal e de mercado. Estaremos interessados somente nos riscos financeiros de mercado, que estão relacionados às perdas potenciais associadas ao comportamento do mercado.

Portanto, o VaR será uma medida de risco financeiro de mercado e que dá uma medida do grau de incerteza sobre retornos líquidos futuros.

No caso de risco de crédito há algumas metodologias, como o *CreditRisk+*, desenvolvido pelo *Crédit Suisse Firts Bank* e o *CreditMetrics*, proposto pelo banco J.P. Morgan. Os interessados podem consultar os documentos preparados por estas instituições. Uma referência interessante e que compara as duas metodologias é Gordy (2000).

## 6.2 Valor em Risco

Informalmente, o VaR é uma medida da variação potencial máxima do valor de um ativo (ou carteira de ativos), sobre um período pré-fixado, com dada probabilidade. Ou seja, quanto se pode perder, com probabilidade  $p$ , sobre um horizonte  $h$  fixado.

Do ponto de vista de uma empresa, o VaR é uma medida de perda associada a um evento extremo, sob condições normais de mercado.

**Exemplo 6.1.** Suponha que exista uma chance de 95% de que a taxa de câmbio Real/USD não caia mais do que 1% em um dia. Suponha, ainda, que uma empresa tenha 100 milhões de reais aplicados num fundo cambial. Calculemos a perda potencial sobre este valor aplicado.

Uma série temporal do desvio padrão (volatilidade)  $\sigma_t$  dos retornos  $r_t$  da taxa de câmbio Real/USD pode dar uma indicação da variação da taxa. Admita o modelo (2.75) para os retornos  $r_t$ , ou seja, estamos admitindo normalidade. Suponha que uma estimativa do desvio padrão hoje seja  $\sigma_t = 0,46\%$  (veremos na seção 6.3 como obter essa estimativa). Então o VaR é calculado como

$$\text{VaR} = (1,65)(\sigma_t) = (1,65)(0,46) = 0,759\%.$$

Portanto, não se espera que a taxa de câmbio caia mais do que 0,759%, com 95% de probabilidade. O valor 1,65 é o (0,05)-quantil da  $\mathcal{N}(0,1)$ . Em

reais, o VaR é o valor de mercado da posição multiplicado pelo valor obtido acima, ou seja,

$$\text{Risco} = 100 \text{ milhões} \times 0,759\% = 759.000,00 \text{ reais.}$$

A conclusão é que em 95% das vezes, não se perderá mais do que R\$ 759.000,00 em um dia.

Passemos a definir formalmente o VaR. Para tanto definamos os dois tipos de posições financeiras em uso.

Uma posição financeira *comprada* (ou “long”) significa possuir determinado ativo (ou carteira de ativos). Uma posição financeira *vendida* (ou “short”) envolve vender um ativo que não se possui. Esta operação é realizada alugando-se o ativo. Em data futura, o vendedor é obrigado a comprar exatamente o mesmo número de cotas ou ações alugadas (e não o valor em moeda), para pagar o débito. Como o pagamento é em cotas ou ações, o vendedor ganha com a queda do preço do ativo.

Suponha que no instante  $t$  estejamos interessados em calcular o risco de uma posição financeira para o horizonte  $h > 0$ . Seja

$$\Delta P(h) = P(t+h) - P(t)$$

a variação do valor do ativo entre os dois instantes. A quantidade  $\Delta P(h)$  representa o lucro ou a perda (L & P) da posição sobre o horizonte  $h$ . Do ponto de vista prático, é o L & P obtido marcando-se a posição a mercado hoje e deixando-a sem mudanças até uma nova marcação  $h$  dias depois, digamos. Chamemos de  $F_h(\cdot)$  a função de distribuição acumulada (f.d.a.) de  $\Delta P(h)$ .

**Definição 6.1.** Definimos o VaR de uma posição comprada sobre o horizonte  $h$ , com probabilidade  $p$ ,  $0 < p < 1$ , por meio de

$$p = P(\Delta P(h) \leq \text{VaR}) = F_h(\text{VaR}). \quad (6.1)$$

Observemos que o VaR depende de  $p$  e de  $h$ , ou seja, deveríamos escrever  $\text{VaR}_{p,h}$ . Além disso, o valor em risco aumenta com  $p$  diminuindo ou com  $h$  aumentando.

Algumas observações são necessárias aqui:

- (i) O VaR em (6.1) é dado em unidades monetárias (u.m.), por exemplo, reais. Lembremos que os retornos simples,  $R_t$ , são dados em porcentagem e que os log-retornos  $r_t$  são aproximadamente iguais a  $R_t$ , logo podemos supor que os  $r_t$  medem, aproximadamente, variações percentuais. Assim sendo, usaremos log-retornos no que segue.
- (ii) A definição 6.1 mostra que o VaR é o  $p$ -quantil da distribuição  $F_h(\cdot)$ . Na prática, teremos que estimar este quantil, usando por exemplo a distribuição empírica dos retornos.
- (iii) O VaR calculado em (6.1) tem valor negativo, pois quem tem uma posição comprada sofre uma perda se  $\Delta P(h) < 0$ .
- (iv) A quantia em u.m. no cálculo do VaR é obtida como no exemplo 6.1, ou seja, multiplicando o valor da posição financeira pelo VaR do log-retorno. A posição financeira em u.m. é usualmente o valor do ativo marcado pelo mercado (“mark-to-market”).

No caso de uma posição vendida, há perda se  $\Delta P(h) > 0$ , ou seja, o preço do ativo aumenta. Neste caso o VaR é definido por

$$p = P(\Delta P(h) \geq \text{VaR}) = 1 - F_h(\text{VaR}), \quad (6.2)$$

que tipicamente é positivo para  $p$  pequeno. O sinal positivo aqui indica perda.

As definições (6.1) e (6.2) implicam que o VaR é calculado usando a cauda esquerda da distribuição  $F_h(\cdot)$ , para uma posição comprada e usando a cauda direita, para uma posição vendida. Também, a definição (6.1) aplica-se a uma posição vendida se usarmos a distribuição de  $-\Delta P(h)$ . Portanto, basta considerar o cálculo do VaR para uma dessas posições.

### 6.3 VaR Usando a Distribuição Normal

Nesta seção discutiremos a abordagem do RiskMetrics (veja Longestae and More, 1995). Aqui, a suposição é que a distribuição condicional dos retornos, dada a informação passada, é normal com média zero e variância  $\sigma_t^2$ , ou seja,



$$r_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma_t^2).$$

Além disso, para estimar a volatilidade  $\sigma_t^2$  é usado o modelo EWMA (“exponentially weighted moving average”)

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2, \quad 0 < \lambda < 1. \quad (6.3)$$

O valor do parâmetro de suavização  $\lambda$  pode ser escolhido minimizando-se a soma dos quadrados dos resíduos de ajuste ou de previsão.

Uma maneira equivalente de formular o modelo é usar a abordagem de modelos GARCH, notando-se que (6.3) pode ser pensado como um modelo IGARCH(1,1).

Vimos no Capítulo 1 que o log-retorno de  $k$  períodos,  $r_t[k]$ , do instante  $t + 1$  ao instante  $t + k$ , é dado por

$$r_t[k] = r_{t+1} + r_{t+2} + \dots + r_{t+k},$$

de modo que podemos escrever que

$$r_t[k] | \mathcal{F}_t \sim \mathcal{N}(0, \sigma_t^2[k]),$$

onde  $\sigma_t^2[k]$ , a volatilidade deste retorno, pode ser calculada usando resultados da modelagem GARCH. De fato, como podemos escrever (6.3) como

$$\begin{aligned} \log P_t - \log P_{t-1} &= r_t = a_t, \\ a_t &= \sigma_t \varepsilon_t \sim \text{IGARCH}(1, 1), \\ \varepsilon_t &\sim \mathcal{N}(0, 1), \end{aligned}$$

segue-se que

$$\begin{aligned} \sigma_t^2[k] &= \text{Var}(r_t[k] | \mathcal{F}_t) = \text{Var}\left(\sum_{i=1}^k r_{t+i} | \mathcal{F}_t\right) = \\ &= \sum_{i=1}^k \text{Var}(a_{t+i} | \mathcal{F}_t), \end{aligned}$$

dado que os  $a_t$  são não-correlacionados. Como  $a_{t-1} = r_{t-1} = \sigma_{t-1}\varepsilon_{t-1}$ , obtemos de (6.3)

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1 - \lambda)\sigma_{t-1}^2\varepsilon_{t-1}^2,$$

ou ainda,

$$\sigma_t^2 = \sigma_{t-1}^2 + (1 - \lambda)\sigma_{t-1}^2(\varepsilon_{t-1}^2 - 1).$$

Desta relação obtemos

$$\sigma_{t+i}^2 = \sigma_{t+i-1}^2 + (1 - \lambda)\sigma_{t+i-1}^2(\varepsilon_{t+i-1}^2 - 1), \quad i = 2, \dots, k.$$

Como  $E(\varepsilon_{t+i-1}^2 - 1) = 0$ ,  $i \geq 2$ , obtemos

$$E(\sigma_{t+i}^2 | \mathcal{F}_t) = E(\sigma_{t+i-1}^2 | \mathcal{F}_t), \quad i = 2, \dots, k. \quad (6.4)$$

Para previsão a um passo da volatilidade, de (6.3) obtemos

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1 - \lambda)r_t^2,$$

ou seja,  $\hat{\sigma}_t^2(1) = \lambda\sigma_t^2 + (1 - \lambda)r_t^2$ , e (6.4) mostra que

$$\text{Var}(r_{t+i} | \mathcal{F}_t) = \sigma_{t+1}^2, \quad i = 2, \dots, k,$$

ou seja,

$$\sigma_t^2[k] = k\hat{\sigma}_t^2(1). \quad (6.5)$$

Isto é, podemos escrever que

$$r_t[k] | \mathcal{F}_t \sim \mathcal{N}(0, k\hat{\sigma}_t^2(1)).$$

Portanto, sob o modelo adotado (6.3), a variância condicional dos log-retornos de  $k$  períodos é proporcional ao horizonte  $k$  e o desvio padrão condicional de  $r_t[k]$  é dado por  $\sqrt{k}\sigma_{t+1}$ , expressão esta usualmente chamada “regra da raiz quadrada do tempo”. Note que  $\sigma_{t+1} = \sqrt{\hat{\sigma}_t^2(1)}$ .

Por exemplo, se fixarmos em (6.1) a probabilidade  $p = 5\%$ , então o Risk-Metrics usa  $-1,65\sigma_{t+1}$  como VaR, que é o 0,05-quantil da normal com média

zero e variância  $\sigma_t^2$ . Normalmente o sinal negativo, que significa perda, é ignorado e

$$\text{VaR} = (\text{Valor da posição}) \times (1,65) \times (\sigma_{t+1}).$$

Esta expressão corresponde ao VaR de um período (um dia, por exemplo). O VaR de  $k$  períodos é dado por

$$\text{VaR}[k] = (\text{Valor da posição}) \times (1,65) \times \sqrt{k} \times \sigma_{t+1},$$

ou seja,

$$\text{VaR}[k] = \sqrt{k} \text{VaR}. \quad (6.6)$$

**Exemplo 6.1.** (continuação) Vimos que

$$\text{VaR} = (100 \text{ milhões}) \times (1,65) \times (0,46\%) = 759.000,00,$$

ao passo que o VaR de 30 dias é dado por

$$\text{VaR}[30] = (100 \text{ milhões}) \times (1,65) \times \sqrt{30} \times (0,46\%) = 4.157.214,00.$$

**Exemplo 6.2.** Suponha que uma empresa tenha uma posição comprada de 10 milhões de reais em ações da Petrobrás e queremos calcular o VaR de 1 e 15 dias. Consideremos os log-retornos diários da Petrobrás, com  $T = 1498$  observações, de 3/1/95 a 27/12/2000, mostrados na Figura 6.1. Supondo que o modelo do RiskMetrics seja válido, obtemos

$$\sigma_t^2 = 0,81\sigma_{t-1}^2 + 0,19r_{t-1}^2.$$

Dos dados obtemos  $r_{1498} = 0,011124$  e do modelo ajustado obtemos  $\hat{\sigma}_{1498}^2 = 0,000235$ . Logo a estimativa da volatilidade 1 passo a frente é dada por  $\hat{\sigma}_{1498}^2(1) = 0,000214$ . Se  $p = 0,05$ , o quantil da distribuição é  $-1,65\sqrt{0,000214} = -0,02413$ . O sinal negativo indica perda e o VaR de um dia será

$$\text{VaR} = 10.000.000,00 \times 0,02413 = 241.300,00,$$

ao passo que o VaR de 15 dias será

$$\text{VaR}[15] = \sqrt{15} \times 241.300,00 = 934.550,00.$$

O VaR de uma posição vendida no mesmo valor seria idêntico, a diferença é que o quantil a ser usado seria positivo, que nesse caso também indica perda.

Os cálculos podem ser feitos por meio de algum programa que estime  $\lambda$ , como por exemplo o EViews ou o S+FinMetrics. Usando este último, obtemos o gráfico de  $\sigma_t$  da Figura 6.2. Compare com a Figura 5.2.

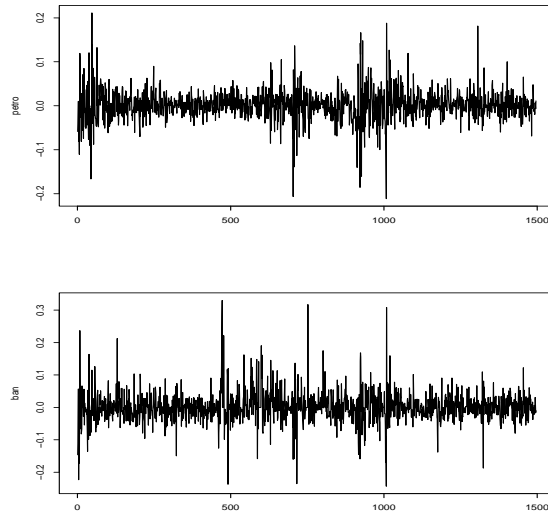


Figura 6.1: Retornos diários da Petrobrás e Banespa

**Observações.** (i) Se a média da série não for zero, o modelo fica

$$r_t = \mu + a_t, \quad a_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 \sim \text{IGARCH}(1, 1), \quad \varepsilon_t \sim \mathcal{N}(0, 1).$$

Neste caso o  $p$ -quantil é dado por  $\mu + z_p\sigma_{t+1}$ , onde  $z_p$  é o  $p$ -quantil da normal padrão; se  $p = 0,05$ , então  $z_p = -1,65$ .

Para  $k$  períodos, o  $p$ -quantil é  $k\mu + z_p\sqrt{k}\sigma_{t+1}$ ; se  $p = 0,05$ , este ficará  $k\mu - 1,65\sqrt{k}\sigma_{t+1} = \sqrt{k}(\sqrt{k}\mu - 1,65\sigma_{t+1})$ , que não é igual a  $\sqrt{k}\text{VaR}$ .

(ii) Suponha que se tenha agora uma carteira com  $m$  posições financeiras e sejam  $r_{1t}, \dots, r_{mt}$  os respectivos retornos. Sejam

$$\rho_{ij} = \text{Corr}(r_{it}, r_{jt}) = \frac{\gamma_{ij,t}}{\sigma_{ii,t}\sigma_{jj,t}},$$

para  $i < j = 1, \dots, m$ , as correlações entre os retornos. Então as covariâncias  $\gamma_{ij,t}$  são estimadas usando, no lugar de (6.3), o modelo

$$\gamma_{ij,t} = \lambda\gamma_{ij,t-1} + (1 - \lambda)r_{i,t-1}r_{j,t-1}. \quad (6.7)$$

Desta maneira é fácil ver que o VaR da carteira é dado por

$$\text{VaR} = \sqrt{\sum_{i=1}^m \text{VaR}_i^2 + 2 \sum_{i<j} \rho_{ij} \text{VaR}_i \text{VaR}_j}, \quad (6.8)$$

onde  $\text{VaR}_i$  é o valor em risco para o retorno  $r_{it}$ . O S+FinMetrics calcula as correlações por meio da função EWMA.

**Exemplo 6.3.** Considere uma carteira com dois ativos, sendo uma posição de 10 milhões de reais em ações da Petrobrás e outra de 5 milhões de reais em ações do Banespa. Os retornos diários do Banespa estão mostrados na Figura 6.1. Vamos calcular o VaR de um dia da carteira, usando (6.8). No exemplo anterior, vimos que  $\text{VaR}_1 = -0,02413$ . Para calcular o VaR para a segunda posição, suponha que  $\lambda = 0,85$ . Obtemos  $\hat{\sigma}_{2,1498}(1) = 0,000315$  e o  $\text{VaR}_2 = -1,65\sqrt{0,000315} = -0,02929$ . Em reais, este VaR é de  $5.000.000,00 \times 0,2929 = 146.450,00$ . O desvio padrão condicional dos retornos do Banespa estão na Figura 6.2. Compare com a Figura 5.7.

A seguir, usaremos (6.7) para estimar a covariância prevista no instante  $t = 1498$ . Usando  $\lambda = 0,90$  para a carteira, temos que

$$\hat{\gamma}_{12,1498}(1) = 0,90\gamma_{12,1498} + 0,10r_{1,1498}r_{2,1498},$$

sendo que  $r_{1,1498} = 0,011124$ ,  $r_{2,1498} = -0,00982$  e  $\gamma_{12,1498} = 0,0000706$ . Obtemos, então,  $\hat{\gamma}_{12,1498}(1) = 0,000052616$ . Segue-se que a correlação  $\rho_{12}$  entre os dois ativos é estimada por

$$\hat{\rho}_{12,1498}(1) = (0,000052616)/(0,0146 \times 0,0177) = 0,20356.$$

Logo, o VaR de um dia da carteira é

$$\text{VaR} = \sqrt{(0,02413)^2 + (0,02929)^2 + 2(0,20356)(0,02413)(0,02929)} = 0,04157.$$

Na Figura 6.3 temos ilustradas as covariâncias entre os retornos, ao longo do tempo. Neste exemplo usamos o programa S+FinMetrics para fazer os cálculos e figuras. Vemos que as covariâncias são em geral positivas.

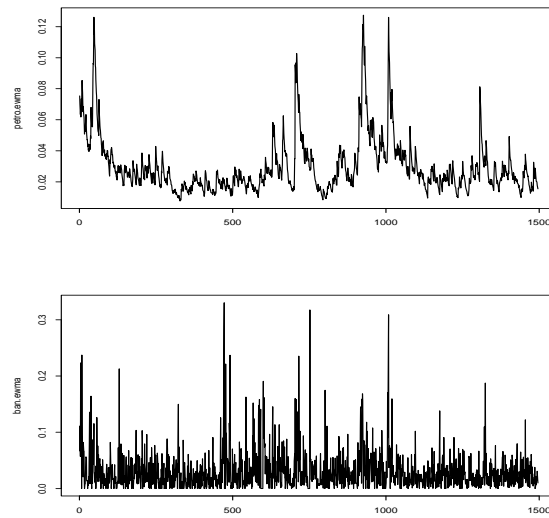


Figura 6.2: Desvios padrões condicionais dos retornos da Petrobrás e Banespa

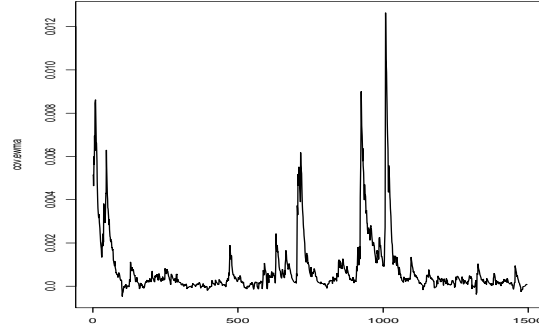


Figura 6.3: Covariâncias entre retornos da Petrobrás e Banespa

## 6.4 VaR Usando Modelos ARMA e GARCH

Lembremos que uma série de retornos é, em geral, não-correlacionada, mas dependente. Se este for o caso, a volatilidade é modelada por um dos modelos heteroscedásticos considerados no Capítulo 5. Mas vimos também que algumas séries de retornos ainda exibem a presença de auto-correlação, havendo a necessidade de eliminá-la por meio do ajuste inicial de um modelo linear, por exemplo da família ARMA.

A estratégia é, portanto, modelar a média da série de retornos  $r_t$  por meio de um modelo ARMA e depois modelar os resíduos  $a_t$  deste modelo por um membro da família ARCH. Por exemplo, se escolhermos um modelo GARCH(r,s) para usar, teremos o modelo ARMA(p,q)-GARCH(r,s):

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{j=1}^q \theta_j a_{t-j}, \quad (6.9)$$

$$a_t = \sigma_t \varepsilon_t, \quad (6.10)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i a_{t-i}^2 + \sum_{j=1}^s \beta_j \sigma_{t-j}^2. \quad (6.11)$$

Como estudamos no Capítulo 5 podemos escolher para  $\varepsilon_t$  uma distribuição normal,  $t$  ou distribuição de erro generalizada.

Se escolhermos  $\varepsilon_t \sim \mathcal{N}(0, 1)$ , resulta que

$$r_{t+1}|\mathcal{F}_t \sim \mathcal{N}(\hat{r}_t(1), \hat{\sigma}_t^2(1)),$$

onde  $\hat{r}_t(1)$  e  $\hat{\sigma}_t^2(1)$  são as previsões a um passo da média e variância usando (6.9) e (6.11), respectivamente.

Supondo-se, por exemplo,  $p = 0,05$ ,

$$\text{VaR} = \hat{r}_t(1) - 1,65\hat{\sigma}_t(1). \quad (6.12)$$

Escolhendo-se  $\varepsilon_t \sim t_\nu$ , o  $p$ -quantil é dado por  $\hat{r}_t(1) - t_\nu^*(p)\hat{\sigma}_t(1)$ , onde  $t_\nu^*(p)$  é o  $p$ -quantil da distribuição  $t_\nu$  padronizada.

Chamando de  $Q(p)$  o  $p$ -quantil de  $t_\nu$ , então temos

$$\begin{aligned} p = P(t_\nu \leq Q(p)) &= P\left(\frac{t_\nu}{\sqrt{\nu/(\nu-2)}} \leq \frac{Q(p)}{\sqrt{\nu/(\nu-2)}}\right) = \\ &= P(t_\nu^* \leq Q^*(p)), \quad \nu > 2, \end{aligned}$$

ou seja,  $Q^*(p)$  é  $p$ -quantil da distribuição  $t_\nu$  padronizada. Logo,

$$\text{VaR} = \hat{r}_t(1) - \frac{t_\nu(p)\hat{\sigma}_t(1)}{\sqrt{\nu/(\nu-2)}}. \quad (6.13)$$

**Exemplo 6.4.** Consideremos a mesma situação do exemplo 6.2 e calculemos o VaR, com  $p = 0,05$ . Um modelo adequado para a série é um AR(1) – GARCH(1,1) gaussiano, que estimado usando o S+FinMetrics resulta ser

$$r_t = 0,00223 + 0,12394r_{t-1} + a_t, \quad a_t = \sigma_t\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1),$$

$$\sigma_t^2 = 0,00003 + 0,1432a_{t-1}^2 + 0,8318\sigma_{t-1}^2.$$

Compare este modelo ajustado com o modelo (5.37).

Dos dados sabemos que  $r_{1498} = 0,011124$  e  $r_{1497} = -0,006689$ , de modo que a previsão da série um passo à frente é dada por

$$\hat{r}_{1498}(1) = 0,00223 + 0,12394r_{1498} = 0,003609.$$



A previsão da volatilidade um passo à frente é

$$\hat{\sigma}_{1498}^2(1) = 0,00003 + 0,1432a_{1498}^2 + 0,8318\sigma_{1498}^2.$$

Agora,  $a_{1498} = r_{1498} - 0,00223 - 0,12394r_{1497} = 0,009723$  e  $\sigma_{1498}^2 = 0,000459$  (obtida do modelo GARCH ajustado), de modo que

$$\hat{\sigma}_{1498}^2(1) = 0,00003 - (0,1432)(0,009723)^2 + (0,8318)(0,000459) = 0,000402.$$

O 0,05-quantil será, então,

$$\hat{r}_{1498}(1) - 1,65\hat{\sigma}_{1498}(1) = 0,003609 - 1,65(0,02004) = -0,0295.$$

Finalmente, o valor em risco de um dia é dado por

$$\text{VaR} = 10.000.000,00 \times 0,0295 = 295.0000,00.$$

Para se obter o VaR de  $k$  períodos, temos que obter  $r_t[k]$  como antes. Estando na origem  $T$ ,

$$r_T[k] = r_{T+1} + \dots + r_{T+k},$$

e usando (6.9) e (6.11) podemos obter a média e variância condicionais de  $r_T[k]$  dada a informação  $\mathcal{F}_T$  até o instante  $T$ . Então, a seguinte proposição pode ser provada (veja Tsay, 2002 e Problema 4).

**Proposição 6.1.** (a) A previsão da média do retorno  $r_t$  no período  $k$  é dada por

$$\hat{r}_T[k] = \hat{r}_T(1) + \dots + \hat{r}_T(k), \quad (6.14)$$

onde  $\hat{r}_T(h)$  é a previsão de origem  $T$  e horizonte  $h$  usando (6.9).

(b) O erro de previsão é dado por

$$e_T[k] = a_{T+k} + (1 + \psi_1)a_{T+k-1} + \dots + \left(\sum_{i=0}^{k-1} \psi_i\right)a_{T+1}, \quad (6.15)$$

onde os  $\psi_i$  são os pesos da representação do processo como uma média móvel infinita.

(c) A previsão da volatilidade do retorno no período  $k$  é dada por

$$\text{Var}(e_T[k]|\mathcal{F}_T) = \hat{\sigma}_T^2(k) + (1 + \psi_1)^2 \hat{\sigma}_T^2(k-1) + \dots + \left(\sum_{i=0}^{k-1} \psi_i\right)^2 \hat{\sigma}_T^2(1), \quad (6.16)$$

onde  $\hat{\sigma}_T^2(h)$  é a previsão  $h$  passos a frente da volatilidade usando (6.11).

**Exemplo 6.4.** (continuação). Calculemos o VaR de 5 dias para o exemplo anterior. Para um modelo AR(1)

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t,$$

sabemos que  $\psi_j = \phi_1^j$ ,  $j \geq 1$ . A previsão de origem  $T$  e horizonte  $h$  é dada por

$$\hat{r}_T(h) = \phi_0 + \phi_1 \hat{r}_T(h-1),$$

sendo que para  $h = 1$ ,  $\hat{r}_T(0) = r_T$ . Trabalhando com os valores do exemplo 6.4 é fácil ver que obtemos  $\hat{r}_{1498}[5] = 0,01394$ . Lembremos que as previsões  $\hat{r}_t(\ell)$  convergem para  $E(r_t) = 0,00255$ , quando  $\ell \rightarrow \infty$ .

Por outro lado, para o modelo GARCH(1,1) temos que

$$\hat{\sigma}_T^2(h) = \alpha_0 + (\alpha_1 + \beta_1) \hat{\sigma}_T^2(h-1), \quad h \geq 2,$$

sendo que para  $h = 1$  a previsão é  $\alpha_0 + \alpha_1 a_T^2 + \beta_1 \sigma_T^2$ . Usando os valores obtidos no exemplo 6.4 podemos calcular as previsões da volatilidade para  $h = 1, 2, \dots, 5$ . Finalmente, usando (6.16) obtemos a previsão da volatilidade do retorno de 5 dias, ou seja,

$$\sigma_T^2[5] = \hat{\sigma}_T^2(5) + (1 + \psi_1)^2 \hat{\sigma}_T^2(4) + \dots + (1 + \psi_1 + \dots + \psi_4)^2 \hat{\sigma}_T^2(1),$$

cujos valores resultam  $\sigma_T^2[5] = 0,0028415$ . O valor em risco de 5 dias é, então, com  $p = 0,05$ ,

$$\text{VaR}[5] = 0,01394 - 1,65\sqrt{0,0028415} = -0,074014.$$

Em u.m. obtemos o valor R\$740.140,00.

## 6.5 VaR Usando Quantis Empíricos

Uma maneira de estimarmos de forma não-paramétrica o VaR é por meio dos quantis empíricos dos dados observados.

Chamemos de  $r_1, \dots, r_T$  os retornos observados e considere as estatísticas de ordem  $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(T)}$ . Se indicarmos por  $Q(p)$  o  $p$ -quantil da distribuição (desconhecida) dos retornos, um estimador consistente é dado pelo  $p$ -quantil empírico, definido por

$$q_p = \begin{cases} r_{(i)}, & \text{se } p = p_i = (i - 0,5)/T, \ i = 1, \dots, T \\ (1 - f_i)r_{(i)} + f_i r_{(i+1)}, & \text{se } p_i < p < p_{i+1} \\ r_{(1)}, & \text{se } 0 < p < p_1 \\ r_{(T)}, & \text{se } p_T < p < 1, \end{cases} \quad (6.17)$$

onde  $f_i = (p - p_i)/(p_{i+1} - p_i)$ .

Ou seja, ordenados os dados,  $q_p$  é uma das estatísticas de ordem, se  $p$  for da forma  $p_i = (i - 0,5)/T$  e está na reta ligando os pontos  $(p_i, r_{(i)})$  e  $(p_{i+1}, r_{(i+1)})$ , se  $p$  estiver entre  $p_i$  e  $p_{i+1}$ . Tomamos  $p_i$  da forma escolhida e não como  $i/T$  para que, por exemplo, a mediana calculada segundo esta definição coincida com a definição usual.

Uma suposição aqui adotada é que a distribuição dos retornos continue válida para o período de previsão, o que pode não ser razoável.

Supondo-se os  $r_t$  i.i.d., com densidade  $f$  e  $f(Q(p)) \neq 0$ , pode-se demonstrar que  $r_{(j)}$  é aproximadamente normal, com média  $Q(p)$  e variância  $p(1 - p)/T[f(Q(p))]^2$ , se  $j = Tp$ . Ou seja, pela definição dada acima, os quantis empíricos são aproximadamente normais, para  $p = p_i$ ,  $0 < p < p_1$  e  $p_T < p < 1$ , e misturas de normais, para  $p_i < p < p_{i+1}$ .

**Exemplo 6.5.** Considere os mesmos dados do exemplo 6.2 e calculemos o VaR de um dia usando os quantis empíricos. Por (6.17) temos  $p_{74} = 0,0490654$ ,  $p_{75} = 0,0500667$  e  $f_{74} = 0,93$  de onde

$$q(0, 05) = (0, 07)r_{(74)} + (0, 93)r_{(75)} = -0, 4816.$$

Segue-se que o valor em risco de um dia da posição é  $\text{VaR} = 481.600,00$  reais.

## 6.6 VaR Usando a Teoria de Valores Extremos

Obtidas as estatísticas de ordem  $r_{(1)} \leq \dots \leq r_{(T)}$ , vamos nos fixar em  $r_{(1)} = \min\{r_1, \dots, r_T\}$  e  $r_{(T)} = \max\{r_1, \dots, r_T\}$ . O mínimo é relevante para o cálculo do VaR para posições financeiras compradas e o máximo é relevante para posições vendidas. Basta considerar um dos casos, devido ao fato que  $r_{(1)} = -\max\{s_1, \dots, s_T\}$ , onde  $s_t = -r_t$ ,  $t = 1, \dots, T$ .

A teoria de valores extremos (TVE) clássica estuda o comportamento de máximos, mínimos e outras estatísticas de ordem, para seqüências de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Extensões para o caso de séries estacionárias com dependência fraca e séries não-estacionárias foram consideradas na literatura. Veja Coles (2001) para detalhes. Mesmo que a série seja dependente, considerando-se os máximos de blocos, como veremos a seguir, a suposição de que estes máximos sejam independentes parece ser razoável na prática.

Um resumo da TVE, relevante aos nossos propósitos, é dada no Apêndice 6. Vamos nos concentrar no máximo  $r_{(T)}$ . A TVE procura obter a distribuição limite (aproximada) para o máximo normalizado

$$r_T^* = \frac{r_{(T)} - b_T}{a_T}, \quad (6.18)$$

para seqüências de constantes  $\{a_T > 0\}$  e  $\{b_T\}$ , que são escolhidas de modo a estabilizar a posição e escala do máximo, quando  $T \rightarrow \infty$ . Conforme mostrado no Apêndice 6, supondo-se os retornos independentes com distribuição  $F$ , se existirem seqüências como acima tais que a distribuição de (6.18) converge para a distribuição não-degenerada  $G(z)$ , então  $G$  pertence a uma de três famílias, que podem ser conjuntamente colocadas na forma

$$G(z) = \exp\left\{-\left[1 + \xi \left(\frac{z - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \quad (6.19)$$

definida sobre  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , para  $-\infty < \mu < \infty$ ,  $-\infty < \xi < \infty$ ,  $\sigma > 0$ .

A família (6.19) é chamada *distribuição generalizada de valores extremos* (GVE), sendo  $\mu$  o *parâmetro de posição*,  $\sigma$  o *parâmetro de escala* e  $\xi$  o *parâmetro de forma*. Como visto no apêndice, esta família é determinada pelo parâmetro  $\xi$ , de modo que se  $\xi = 0$  obtemos a família tipo I de Gumbel, se  $\xi > 0$  obtemos a família tipo II de Fréchet e se  $\xi < 0$  a família tipo III de Weibull.

Para aplicar a TVE a séries de retornos, procedemos como segue:

- (a) dividimos a série observada de retornos  $r_1, \dots, r_T$  em  $m$  blocos de tamanho  $n$ ;
- (b) obtemos o máximo de cada bloco,  $r_{n,i}, i = 1, \dots, m$ , aos quais a TVE pode ser aplicada, ou seja, ajustamos uma distribuição GVE a esses máximos;
- (c) estimamos os quantis desta distribuição, a partir do qual podemos obter o VaR de uma posição vendida. Note-se que

$$r_{n,i} = \max_{1 \leq j \leq n} \{r_{(i-1)n+j}\}, \quad i = 1, \dots, m. \quad (6.20)$$

**Observações.** (i) O comportamento das caudas de  $F$  é que determina a distribuição limite  $G$ .

(ii) A família tradicionalmente usada em gestão de risco é a família de Fréchet ( $\xi > 0$ ).

(iii) A escolha de  $a_T$  e  $b_T$  depende de  $F$ .

A coleção de máximos (6.20) pode ser usada para estimar os parâmetros do modelo GVE. Usualmente os blocos são escolhidos de modo a corresponder a um ano de observações, se tivermos por exemplo dados mensais. No caso de retornos diários, os valores usados são  $n = 21$  (um mês),  $n = 63$  (um trimestre) e  $n = 252$  (um ano). Como vimos no apêndice, podemos obter os EMV, supondo

$$z_{n,i} = \frac{r_{n,i} - b_n}{a_n}. \quad (6.21)$$

Estimadores dos quantis da distribuição dos máximos de grupos são obtidos invertendo-se a equação (6.19). Se  $0 < p^* < 1$ , o  $(1 - p^*)$ -quantil é dado por

$$z_{p^*} = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p^*)\}^{-\xi}], & \text{se } \xi \neq 0 \\ \mu - \sigma \log[-\log(1 - p^*)], & \text{se } \xi = 0, \end{cases} \quad (6.22)$$

com  $G(z_{p^*}) = 1 - p^*$ . Este quantil é, às vezes, chamado de *nível de retorno*, associado ao *período de retorno*  $1/p^*$ . A interpretação é que o valor  $z_{p^*}$  é excedido pelo máximo do período (anual, por exemplo), em qualquer período, com probabilidade  $p^*$ . Observe que deveríamos escrever  $\mu_n, \sigma_n$  e  $\xi_n$  em (6.22), pois para cada escolha de  $n$  temos estimativas diferentes para os parâmetros. O  $(1 - p^*)$ -quantil estimado é obtido substituindo-se os EMV de  $\mu, \sigma$  e  $\xi$  em (6.22).

Para obter o VaR da série de retornos original  $r_t$  temos que relacionar quantis desta série com os quantis da série dos máximos. Temos

$$p^* = P(r_{n,i} \geq z_{p^*}) = 1 - P(r_{n,i} \leq z_{p^*}) = 1 - [P(r_t \leq z_{p^*})]^n,$$

do que seque

$$1 - p^* = [1 - P(r_t \geq z_{p^*})]^n. \quad (6.23)$$

Para a série original de retornos,  $r_t$ , fixado  $p$ , o  $(1 - p)$ -ésimo quantil de  $r_t$  é  $z_{p^*}$  se a probabilidade  $p^*$  for escolhida de (6.23) com  $p = P(r_t \geq z_{p^*})$ , logo devemos ter

$$1 - p^* = (1 - p)^n.$$

De (6.22), o VaR de uma posição vendida será dado por

$$\text{VaR} = \begin{cases} \mu_n - \frac{\sigma_n}{\xi_n} \{1 - [-n \log(1 - p)]^{-\xi_n}\}, & \text{se } \xi_n \neq 0 \\ \mu_n - \sigma_n \log[-n \log(1 - p)], & \text{se } \xi_n = 0 \end{cases} \quad (6.24)$$

Para posições compradas vale um raciocínio análogo, trabalhando com  $r_{(1)}$  e a definição 6.1. Veja o Apêndice 6 para detalhes.

Resumindo, o procedimento para calcular o VaR no caso de uma posição financeira vendida, usando a TVE, é:

- (a) Selecione  $n$  e obtenha os máximos dos blocos,  $\{r_{n,i}\}$ .
- (b) Obtenha os EMV de  $\mu, \sigma, \xi$ , para o valor fixado de  $n$ .
- (c) Se o modelo ajustado for adequado, use (6.24) para calcular o VaR.

**Exemplo 6.6.** Considere a situação em que um fundo mantém uma posição vendida de 10 milhões de reais em ações do Banespa. Consideremos os  $T = 1470$  últimos log-retornos, deprezando-se os primeiros 28 dados, para podermos obter  $m = 70$  blocos de  $n = 21$  dias, ou seja, estamos usando máximos mensais. Aos 70 máximos de blocos assim obtidos ajustamos uma distribuição GVE. Utilizamos aqui o software desenvolvido por Coles (2001), que pode ser obtido do site do autor. Os estimadores obtidos são  $\hat{\mu} = 0,0671(0,00489)$ ,  $\hat{\sigma} = 0,0357(0,00394)$  e  $\hat{\xi} = 0,2242(0,10543)$ , onde colocamos entre parênteses os respectivos desvios padrões. O valor da log-verossimilhança resultante é  $-113,9054$ , que poderia ser usado para comparar diversos ajustes, por exemplo, para  $n = 63$ . A função `gev` do S+FinMetrics/EVIS também pode ser utilizada e aplicada neste caso produz os mesmos resultados. Substituindo estes valores em (6.24), com  $p = 0,05$  obtemos

$$\text{VaR} = 0,0671 - \frac{0,0357}{0,2242} \{1 - [-21 \log(0,95)]^{-0,2242}\} = 0,0645.$$

Segue-se que o VaR de um dia da posição é de  $(10\text{milhões}) \times 0,0645 = 645.000,00$ .

Uma abordagem diferente, usando excessos sobre (ou abaixo) de um limiar especificado, pode ser utilizada. Aqui, a distribuição relevante é a distribuição generalizada de Pareto e resultados sobre processos pontuais. Veja Coles (2001) e Tsay (2002) para detalhes. O VaR de  $k$  períodos é dado no Problema 5.

## 6.7 Tópicos Adicionais

Há muitas vantagens e limitações no uso do VaR (veja Alexander, 2001). Dentre as vantagens citamos duas: pode ser usado para comparar os riscos de mercado de todos os tipos de atividades de uma empresa e pode ser estendido a outros tipos de risco, como risco de crédito e operacional. Entre as desvantagens : o VaR somente considera riscos de curto prazo em circunstâncias normais de mercado, os custos para implementar um sistema de cálculo do VaR podem ser altos e as medidas de VaR podem ser imprecisas. De fato, o método mais utilizado faz a suposição não realista de distribuição normal para os retornos. Outra desvantagem é que o VaR não fornece informação sobre o tamanho das violações, quando elas ocorrem. Seria de interesse examinar as magnitudes dos excessos. Para uma análise empírica, veja Berkowitz and O' Brien (2002). Para uma proposta de análise dos excessos, veja Manteiga (2002).

Artzner et al. (1997) introduziram certos requisitos que medidas de risco deveriam cumprir. Uma medida de risco é dita *coerente* se a cada perda  $X$  associa uma medida de risco  $\rho(X)$  tal que:

- (i)  $X \geq Y \rightarrow \rho(X) \geq \rho(Y)$  : monotonicidade;
- (ii)  $\rho(tX) = t\rho(X)$  : homogeneidade;
- (iii)  $\rho(X + nr) = \rho(X) - n$ , onde  $r$  é a taxa livre de risco;
- (iv)  $\rho(X + Y) \leq \rho(X) + \rho(Y)$  : sub-aditividade.

As propriedades (i)-(iv) implicam que a função de risco é convexa. A propriedade (iv) nos diz que o risco total não é maior do que a soma dos riscos de posições individuais.

O VaR não é uma medida coerente de risco, pois não satisfaz necessariamente à propriedade (iv). Artzner et al. (1999) introduziram o conceito de VaR condicional, CVaR, que é uma medida coerente de risco: é a perda esperada, dado que esta exceda o VaR. Formalmente, se  $X$  representa a perda,

$$\text{CVaR} = E(X|X > \text{VaR}). \quad (6.25)$$



O CVaR é também chamado “perda média de excessos” ou ainda “tail VaR”. Segue-se que  $\text{CVaR} \geq \text{VaR}$ . O cálculo do CVaR está intimamente ligado à distribuição de valores extremos. De fato, temos que estimar a distribuição dos excessos sobre um limiar, que conduz à distribuição generalizada de Pareto. Veja Coles (2001) para detalhes.

Um problema importante é aquele de otimizar uma carteira de instrumentos financeiros com o objetivo de reduzir o risco. Trabalhos recentes têm considerado metodologias que visam otimizar o CVaR, em vez de minimizar o VaR. Veja Rockafellar and Uryasev (2000, 2002) e Andersson et al. (2001).

Engle and Manganelli (1999) propuseram uma outra abordagem: em vez de modelar toda a distribuição, modela-se diretamente o quantil por meio de uma especificação autoregressiva condicional, denominada *valor em risco autoregressivo condicional* – CAViaR. Um caso particular é

$$\text{VaR}_t = \beta_0 + \beta_1 \text{VaR}_{t-1} + \ell(\beta_2, X_{t-1}, \text{VaR}_{t-1}),$$

onde o papel de  $\ell(\cdot, \cdot, \cdot)$  é ligar o nível atual do valor em risco com nível anterior do retorno. Os parâmetros do modelo são estimados usando regressão quantílica (Koenker e Basset, 1978).

## 6.8 Problemas

1. Suponha que um investidor tenha uma posição vendida de um milhão de reais em ações da Cemig. Considere a série de preços diários dessas ações (arquivo d-cemig95.00.dat), de 2/1/95 a 27/12/2000, com  $T = 1499$  observações e obtenha os respectivos log-retornos. Use  $p = 0,05$  e calcule o VaR dessa posição para horizontes de 1 e 15 dias usando:
  - (a) o método do RiskMetrics;
  - (b) um modelo gaussiano ARMA-GARCH;
  - (c) um modelo ARMA-GARCH com erros seguindo uma distribuição  $t$  de Student.
2. Considere os log-retornos diários da Petrobrás (arquivo d-petro95.00.dat) e suponha uma posição vendida de cinco milhões de reais. Calcule o VaR de 1 e 30 dias usando:

- (a) um modelo gaussiano ARMA-GARCH;
  - (b) a teoria de valores extremos.
3. Suponha um fundo de investimentos que possua uma posição comprada com 10 milhões de reais em ações da IBM. Considere a série de log-retornos diários da IBM de 3/7/62 a 31/12/99 (arquivo d-ibm62.99.dat). Calcule o VaR de 1 e 5 dias, usando:
- (a) quantis empíricos;
  - (b) RiskMetrics;
  - (c) teoria de valores extremos.
4. Prove a Proposição 6.1.
5. Usando a TVE o VaR de  $k$  períodos para posições vendidas é dado por

$$\text{VaR}[k] = k^\xi \text{VaR},$$

onde  $\xi$  é o parâmetro de forma da distribuição GVE. Obtenha o Var de 20 dias para o exemplo 6.6. Estamos supondo, aqui, que o valor de  $\xi$  é positivo, ou seja, a distribuição ajustada é de Fréchet. No caso de posição comprada, usando-se mínimos, o fator multiplicativo torna-se  $k^{-\xi}$ . Neste caso, a distribuição de Fréchet conduz a  $\xi < 0$ .

6. Obtenha o VaR para uma posição comprada, usando (A.5) abaixo.
7. Obtenha o VaR de um dia de uma posição de compra de 5 milhões de reais aplicada em ações da Petrobrás. Use os retornos do arquivo d-petro95.00.dat.
8. Para o exemplo 6.4, considere o modelo AR(1) – GARCH(1,1), mas agora usando a distribuição  $t_\nu$ . Obtenha o VaR de um dia e de 30 dias.

## Apêndice 6. Teoria de Valores Extremos

### Teorema Limite Para Máximos

Considere o máximo  $r_{(T)} = \max\{r_1, \dots, r_T\}$  de uma série observada de retornos, supostos independentes com distribuição  $F(\cdot)$ . Então,

$$P(r_{(T)} \leq z) = P(r_1 \leq z, \dots, r_T \leq z) = \prod_{i=1}^T P(r_i \leq z) = [F(z)]^T. \quad (A.1)$$

Como  $F$  é desconhecida, (A.1) também o é. Seja  $z_+$  o menor valor de  $z$  tal que  $F(z) = 1$ . Então  $F^T(z) \rightarrow 0$ ,  $T \rightarrow \infty$ , logo a distribuição de  $r_{(T)}$  é degenerada em  $z_+$ . Esta dificuldade é contornada considerando-se seqüências  $\{a_T > 0\}$  e  $\{b_T\}$  tais que

$$r_T^* = \frac{r_{(T)} - b_T}{a_T} \quad (A.2)$$

convirja para uma distribuição não-degenerada, para  $T \rightarrow \infty$ . O seguinte resultado é, então, válido.

**Teorema.** Se existirem seqüências de constantes  $\{a_T > 0\}$  e  $\{b_T\}$  tais que a distribuição de  $r_T^*$  convirja para uma distribuição  $G(z)$ , para  $T \rightarrow \infty$ , com  $G$  não-degenerada, então  $G$  pertence a uma das seguintes famílias:

(i) Tipo I de Gumbel:

$$G(z) = \exp\{-\exp[-(z - b)/a]\}, \quad -\infty < z < +\infty;$$

(ii) Tipo II de Fréchet:

$$G(z) = \begin{cases} 0, & \text{se } z \leq b \\ \exp\{-(z - b)/a\}^{-\alpha}, & \text{se } z > b; \end{cases}$$

(iii) Tipo III de Weibull:

$$G(z) = \begin{cases} \exp\{-[-((z - b)/a)^\alpha]\}, & \text{se } z < b \\ 1, & \text{se } z \geq b, \end{cases}$$

para  $a > 0, \alpha > 0$  e  $b$  real.

Estas três classes são chamadas *distribuições de valores extremos*;  $a$  é o parâmetro de escala,  $b$  é o parâmetro de posição e  $\alpha$  é o parâmetro de forma.

É conveniente escrever as três famílias numa única forma

$$G(z) = \exp\{-[1 + \xi(\frac{z - \mu}{\sigma})]^{-1/\xi}\}, \quad (A.3)$$

definida em  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , com  $-\infty < \mu < \infty$ ,  $\sigma > 0$ ,  $-\infty < \xi < \infty$ , e chamada *família generalizada de valores extremos* (GVE). Aqui,  $\mu$  é o parâmetro de posição,  $\sigma$  é o parâmetro de escala e  $\xi$  é o parâmetro de forma. Temos que  $\xi > 0$  para a família de tipo II (Fréchet),  $\xi < 0$  para a família de tipo III (Weibull) e o caso  $\xi = 0$  é obtido como o limite de (A.3) para  $\xi \rightarrow 0$ , conduzindo à família de tipo I de Gumbel, com  $b = \mu$  e  $a = \sigma$ .

Com esta generalização, em vez de se ter que escolher uma família inicialmente, para depois estimar os parâmetros, a inferência se faz diretamente sobre o parâmetro  $\xi$ . Na Figura 6.4 temos as densidades das três famílias, com  $\xi = 0$  (Gumbel),  $\xi = 0,5$  (Fréchet) e  $\xi = -0,5$  (Weibull).

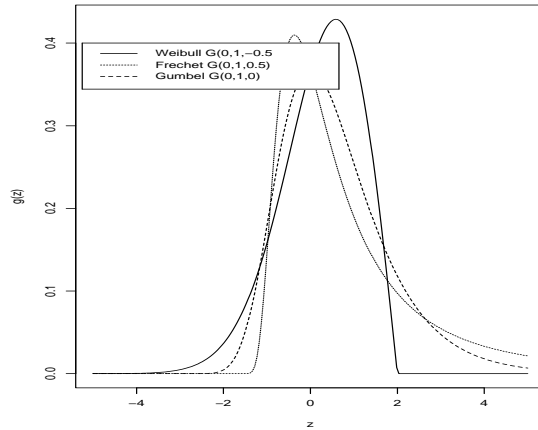


Figura 6.4: Densidades das famílias Fréchet, Weibull e Gumbel

O teorema acima pode ser usado para aproximar a distribuição de máximos de seqüências longas. Na prática, os dados são divididos em  $m$  blocos de tamanho  $n$ , digamos, gerando um conjunto de máximos  $r_{n,1}, \dots, r_{n,m}$ , aos quais a distribuição GVE pode ser ajustada. Em dados de séries temporais oceanográficas ou meteorológicas, por exemplo, os blocos são tomados como contendo um ano de observações. Veja o texto para o caso de retornos financeiros.

Em situações práticas há interesse em se determinar quantis da distribuição GVE. Invertendo (A.3) obtemos

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - [-\log(1-p)]^{-\xi}\}, & \text{se } \xi \neq 0 \\ \mu - \sigma \log[-\log(1-p)], & \text{se } \xi = 0, \end{cases} \quad (\text{A.4})$$

onde  $G(z_p) = 1 - p$ , ou seja,  $z_p$  é o  $(1-p)$ -quantil da distribuição  $G$  e é denominado de *nível de retorno*, associado ao *período de retorno*  $1/p$ . Note que seria melhor escrever  $z_{1-p}$ , usando a notação usual para quantis.

Ou seja,  $z_p$  é o valor tal que, com probabilidade  $p$ , é excedido pelos máximos no período. Podemos fazer gráficos de  $z_p$  contra  $y_p = -\log(1-p)$  (ou contra  $\log y_p$ ), que é linear se  $\xi = 0$ .

A escolha das seqüências  $\{a_T\}$  e  $\{b_T\}$  é que determina a forma da distribuição limite (A.1). Veja Coles (2001) para detalhes e exemplos de escolhas convenientes.

Um modelo para mínimos pode ser obtido facilmente. Suponha que  $r_{(1)} = \min\{r_1, \dots, r_T\}$  e tomemos  $s_t = -r_t, t = 1, \dots, T$ . Então, se  $s_{(T)}$  for o máximo da seqüência  $s_1, \dots, s_T$ , temos que  $r_{(1)} = -s_{(T)}$ . Logo, para  $T$  grande,

$$\begin{aligned} P(r_{(1)} \leq z) &= P(-s_{(T)} \leq z) = P(s_{(T)} \geq -z) = 1 - P(s_T \leq -z) \\ &\approx 1 - \exp\{-[1 + \xi(-z - \mu)/\sigma]^{-1/\xi}\} \\ &= 1 - \exp\{-[1 - \xi(z - \tilde{\mu})/\sigma]^{-1/\xi}\}, \end{aligned} \quad (\text{A.5})$$

sobre  $\{z : 1 - \xi(z - \tilde{\mu})/\sigma > 0\}$ , com  $\tilde{\mu} = -\mu$ . Esta é a distribuição GVE para mínimos. Daqui podemos obter o  $p$ -quantil  $z_p$  tal que  $G(z_p) = p$ .

## Inferência para GVE

Um primeiro ponto a ressaltar é que a escolha do tamanho dos blocos é um fator crítico na análise; se houver poucas observações num bloco, os estimadores serão viesados e se houver muitas observações, as variâncias dos estimadores serão grandes. Como em muitas outras situações, deve haver uma escolha de compromisso entre resolução (acurácia, viés) e estabilidade (precisão, variância). Mas como já salientamos, na prática muitas vezes há uma escolha conveniente, como por exemplo um ano de observações.

O parâmetro de forma  $\xi$  pode ser estimado por métodos não paramétricos. Não trataremos desse caso aqui e o leitor interessado pode consultar Tsay (2002) para detalhes. Vamos nos limitar aqui aos estimadores de máxima verossimilhança (EMV). Há o problema inicial de se verificar se as condições de regularidade necessárias para o uso de propriedades assintóticas dos EMV estão satisfeitas no caso do modelo GVE. Para nosso uso, basta saber que se  $\xi > -0,5$  os EMV são regulares e nos demais casos há problemas. Smith (1985) discute estes casos, mas parece que o caso  $\xi \leq -0,5$  raramente ocorre nas aplicações (corresponde à distribuições com caudas à direita muito curtas), de modo que não teremos problemas em usar o método de MV aqui. Os dados disponíveis são os máximos de blocos,  $z_1, \dots, z_m$ , que supomos serem independentes com distribuição GVE. A log-verossimilhança para o caso  $\xi \neq 0$  é dada por

$$\begin{aligned} \ell(\mu, \sigma, \xi | \mathbf{z}) = & -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^m \log[1 + (z_i - \mu)/\sigma] \\ & - \sum_{i=1}^m [1 + \xi(z_i - \mu)/\sigma]^{-1/\xi}, \end{aligned} \quad (A.6)$$

dado que  $1 + \xi(z_i - \mu)/\sigma > 0$ ,  $i = 1, \dots, m$ .

No caso  $\xi = 0$ , obtemos

$$\log(\mu, \sigma | \mathbf{z}) = -m \log \sigma - \sum_{i=1}^m (z_i - \mu)/\sigma - \sum_{i=1}^m \exp\{-(z_i - \mu)/\sigma\}. \quad (A.7)$$

A maximização numérica de (A.6) ou (A.7) conduz aos EMV  $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ . Devido às propriedades dos EMV, estes estimadores terão uma distribuição assintótica normal.

O  $(1 - p)$ -quantil pode ser estimado substituindo os EMV dos parâmetros em (A.5) e obtemos

$$\hat{z}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - y_p^{-\hat{\xi}}), & \text{se } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{se } \hat{\xi} = 0, \end{cases}$$

com  $y_p = -\log(1 - p)$ .

Pode-se estimar a variância de  $\hat{z}_p$  usando o método delta. Veja Coles (2001) para mais informações sobre este método e sobre o uso das ferramentas usuais de diagnóstico do modelo ajustado.

## Comentários Finais

Neste apêndice fornecemos um breve resumo sobre a TVE relevante ao cálculo do VaR, concentrando-nos no caso univariado. Há extensões para observações com certo grau de dependência, séries não estacionárias e séries multivariadas.

Em algumas situações há a possibilidade de se considerar outras estatísticas de ordem, além do máximo (ou mínimo). Por exemplo, podemos ter informação sobre as  $r$  maiores estatísticas de ordem em cada instante de tempo. Além disso, podemos considerar extremos que excedem um limiar (“threshold”) fixado, situação para a qual a distribuição generalizada de Pareto é aplicável. Finalmente, há a possibilidade de se caracterizar valores extremos que excedem um limiar usando a teoria de processos pontuais. Veja Coles (2001), Tsay (2002), Mendes (2004), Embrechts et al. (1997) e Reiss and Thomas (2001) para informação sobre esses tópicos.





# Modelos Lineares Multivariados

## 7.1 Introdução

Neste capítulo estaremos interessados em estabelecer modelos para uma série temporal vetorial  $\mathbf{X}_t$ , com  $n$  componentes  $X_{1t}, X_{2t}, \dots, X_{nt}$ , observadas em  $t = 0, \pm 1, \pm 2, \dots$ . Além da análise de cada componente individual  $X_{it}$ , como já tratamos em capítulos anteriores, onde a auto-correlação contida em cada série é importante, estaremos estudando as relações dinâmicas entre as séries componentes. Usaremos a notação  $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{nt})'$ ,  $t \in Z$  e  $X_{it}$  ou  $X_{i,t}$ , indistintamente, para a  $i$ -ésima componente,  $i = 1, \dots, n$ .

**Exemplo 7.1.** Podemos pensar o vetor  $\mathbf{X}_t$  como constituído pelos retornos de  $n$  ativos financeiros de um fundo de investimentos no instante  $t$  e o objetivo é analisar o desempenho do fundo ao longo do tempo. Numa outra situação, um investidor pode ter uma carteira com ações da Petrobrás, Banco do Brasil, Banespa e Banco Itaú e neste caso  $n = 4$ .

O vetor de médias de  $\mathbf{X}_t$  será denotado por

$$\boldsymbol{\mu}_t = E(\mathbf{X}_t) = (\mu_{1t}, \mu_{2t}, \dots, \mu_{nt})' \quad (7.1)$$

e depende, em geral, de  $t$ .

A matriz de covariâncias de  $\mathbf{X}_t$  é definida por

$$\mathbf{\Gamma}(t + \tau, t) = E\{(\mathbf{X}_{t+\tau} - \boldsymbol{\mu}_{t+\tau})(\mathbf{X}_t - \boldsymbol{\mu}_t)'\}, \quad (7.2)$$

que é uma matriz  $n \times n$  e que, em geral, também depende de  $t$ .

As quantidades (7.1) e (7.2) descrevem as propriedades de segunda ordem das séries  $X_{1t}, \dots, X_{nt}$ . Se estas tiverem uma distribuição normal multivariada, as propriedades das séries serão completamente especificadas pelas médias e covariâncias. Observe que (7.2) fornece as auto-covariâncias das séries individuais bem como as covariâncias entre séries diferentes.

Se denotarmos por  $\gamma_{ij}(t + \tau, t)$ ,  $i, j = 1, \dots, n$  as componentes da matriz  $\mathbf{\Gamma}(t + \tau, t)$ , então

$$\begin{aligned} \gamma_{ij}(t + \tau, t) &= \text{Cov}\{X_{i,t+\tau}, X_{j,t}\} \\ &= E\{(X_{i,t+\tau} - \mu_{i,t+\tau})(X_{j,t} - \mu_{j,t})\}, \end{aligned} \quad (7.3)$$

$i, j = 1, \dots, n$  é a covariância entre as séries  $X_{i,t+\tau}$  e  $X_{j,t}$ .

**Exemplo 7.2.** No exemplo 7.1, com  $\mathbf{X}_t = (X_{1t}, \dots, X_{4t})'$ ,  $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{4t})'$  é o vetor de médias e a matriz (7.2) ficará

$$\mathbf{\Gamma}(t + \tau, t) = \begin{bmatrix} \gamma_{11}(t + \tau, t) & \gamma_{12}(t + \tau, t) & \cdots & \gamma_{14}(t + \tau, t) \\ \gamma_{21}(t + \tau, t) & \gamma_{22}(t + \tau, t) & \cdots & \gamma_{24}(t + \tau, t) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{41}(t + \tau, t) & \gamma_{42}(t + \tau, t) & \cdots & \gamma_{44}(t + \tau, t) \end{bmatrix}$$

Na diagonal principal temos as auto-covariâncias das séries individuais, calculadas nos instantes  $t + \tau$  e  $t$ , enquanto fora da diagonal principal temos as *covariâncias cruzadas* entre as séries  $X_{i,t+\tau}$  e  $X_{j,t}$ ,  $i \neq j$ .

Um caso de interesse é quando tanto o vetor de médias quanto a matriz de covariâncias não depende de  $t$ . Obteremos séries (fracamente) estacionárias.

## 7.2 Séries Estacionárias

Vamos nos restringir neste capítulo ao caso de estacionariedade fraca ou de segunda ordem. Dizemos que a série  $n$ -variada  $\mathbf{X}_t$  é *estacionária* se a média

$\mu_t$  e a matriz de covariâncias  $\Gamma(t + \tau, t)$ ,  $t, \tau \in Z$ , não dependerem do tempo  $t$ . Nesta situação teremos

$$\mu = E(\mathbf{X}_t) = (\mu_1, \dots, \mu_n)', \quad (7.4)$$

e

$$\Gamma(\tau) = E\{(\mathbf{X}_{t+\tau} - \mu)(\mathbf{X}_t - \mu)'\} = [\gamma_{ij}(\tau)]_{i,j=1}^n, \quad (7.5)$$

$\tau \in Z$ . Neste caso,  $\gamma_{ii}(\tau)$  será a função de auto-covariância da série estacionária  $X_{it}$  e  $\gamma_{ij}(\tau)$  será a função de covariância cruzada de  $X_{it}$  e  $X_{jt}$ . Note-mos que, em geral,  $\gamma_{ij}(\tau) \neq \gamma_{ji}(\tau)$ .

No caso particular de  $\tau = 0$  em (7.5) obtemos

$$\Gamma(0) = E\{(\mathbf{X}_t - \mu)(\mathbf{X}_t - \mu)'\}, \quad (7.6)$$

que é a *matriz de covariâncias contemporâneas*. Em particular,  $\gamma_{ii}(0) = \text{Var}(X_{it})$ ,  $\gamma_{ij}(0) = \text{Cov}\{X_{it}, X_{jt}\}$ .

O *coeficiente de correlação contemporâneo* entre  $X_{it}$  e  $X_{jt}$  é então dado por

$$\rho_{ij}(0) = \frac{\gamma_{ij}(0)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}. \quad (7.7)$$

Obviamente,  $\rho_{ij}(0) = \rho_{ji}(0)$ ,  $\rho_{ii}(0) = 1$  e  $-1 \leq \rho_{ij}(0) \leq 1$ , para todo  $i, j = 1, \dots, n$ , do que decorre que  $\rho(0) = [\rho_{ij}(0)]_{i,j=1}^n$  é uma matriz simétrica, com elementos na diagonal principal todos iguais a um.

A *matriz de correlações* de lag  $\tau$  é definida por

$$\rho(\tau) = \mathbf{D}^{-1}\Gamma(\tau)\mathbf{D}^{-1}, \quad (7.8)$$

sendo  $\mathbf{D} = \text{diag}\{\sqrt{\gamma_{11}(0)}, \dots, \sqrt{\gamma_{nn}(0)}\}$ . Ou seja, denotando  $\rho(\tau) = [\rho_{ij}(\tau)]_{i,j=1}^n$ , temos

$$\rho_{ij}(\tau) = \frac{\gamma_{ij}(\tau)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}, \quad (7.9)$$

que é o coeficiente de correlação entre  $X_{i,t+\tau}$  e  $X_{j,t}$ .

Quando  $\tau > 0$ , este coeficiente mede a dependência linear de  $X_{it}$  sobre  $X_{jt}$ , que ocorreu antes do instante  $t + \tau$ . Então, se  $\rho_{ij}(\tau) \neq 0$ ,  $\tau > 0$ , dizemos que  $X_{jt}$  é *antecedente* a  $X_{it}$  ou que  $X_{jt}$  *lidera*  $X_{it}$  no lag  $\tau$ . De modo análogo,  $\rho_{ji}(\tau)$  mede a dependência linear de  $X_{jt}$  sobre  $X_{it}$ ,  $\tau > 0$ .

O fato que  $\rho_{ij}(\tau) \neq \rho_{ji}(\tau)$ , para todo  $i, j$ , vem, intuitivamente, do fato que estes dois coeficientes de correlação medem relações lineares diferentes entre  $X_{it}$  e  $X_{jt}$ . As matrizes  $\mathbf{\Gamma}(\tau)$  e  $\boldsymbol{\rho}(\tau)$  não são, em geral, simétricas. O que vale é a seguinte proposição.

**Proposição 7.1.** As seguintes propriedades são válidas:

- (i)  $\mathbf{\Gamma}(\tau) = \mathbf{\Gamma}'(-\tau)$ .
- (ii)  $|\gamma_{ij}(\tau)| \leq [\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}$ ,  $i, j = 1, \dots, n$ .
- (iii)  $\gamma_{ii}(\tau)$  é uma função de auto-covariância, para todo  $i$ .
- (iv)  $\sum_{j,k=1}^m \mathbf{a}'_j \mathbf{\Gamma}(j-k) \mathbf{a}_k \geq 0$ , para quaisquer  $m$  e  $\mathbf{a}_1, \dots, \mathbf{a}_m$  vetores de  $\mathbb{R}^n$ .

As demonstrações de (i)-(iii) são imediatas e (iv) decorre do fato que  $E(\sum_{j=1}^m \mathbf{a}'_j (\mathbf{X}_j - \boldsymbol{\mu}))^2 \geq 0$ . Observe também que de (i) obtemos que  $\gamma_{ij}(\tau) = \gamma_{ji}(-\tau)$ . A matriz  $\boldsymbol{\rho}(\tau)$  tem propriedades análogas, sendo que  $\rho_{ii}(0) = 1$ . Note que  $\rho_{ij}(0)$  não necessita ser igual a 1 e também é possível que  $|\gamma_{ij}(\tau)| > |\gamma_{ij}(0)|$ , se  $i \neq j$ ; o que vale é a propriedade (ii) acima.

**Exemplo 7.3.** Suponha que os processos  $X_{1t}$  e  $X_{2t}$  sejam dados por

$$\begin{aligned} X_{1t} &= a_t, \\ X_{2t} &= 0,3X_{2,t-1} + a_t, \end{aligned}$$

sendo  $a_t \sim \text{RB}(0, 1)$ . Se  $\mathbf{X}_t = (X_{1t}, X_{2t})'$ , então é fácil ver que  $\boldsymbol{\mu} = E(\mathbf{X}_t) = \mathbf{0}$ . Por outro lado,

$$\begin{aligned} \mathbf{\Gamma}(0) &= E(\mathbf{X}_t \mathbf{X}_t') = \\ E \begin{bmatrix} a_t^2 & 0,3X_{2,t-1}a_t + a_t^2 \\ 0,3X_{2,t-1}a_t + a_t^2 & (0,3X_{2,t-1} + a_t)^2 \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & 1,099 \end{bmatrix}, \end{aligned}$$

dado que  $E(a_t^2) = 1$ ,  $E(0, 3X_{2,t-1}a_t + a_t^2) = E(a_t^2) = 1$  ( $X_{2,t-1}$  só depende de  $a_{t-1}, a_{t-2}, \dots$ ) e  $E(0, 09X_{2,t-1}^2 + a_t^2 + 0, 6a_tX_{2,t-1}) = 0, 09\text{Var}(X_{2,t}) + 1 = (0, 09)(1/0, 91) + 1 = 1, 099$ .

Agora,

$$\Gamma(1) = E(\mathbf{X}_{t+1}\mathbf{X}_t') =$$

$$E \begin{bmatrix} a_{t+1}a_t & 0, 3X_{2,t-1}a_{t+1} + a_{t+1}a_t \\ 0, 3X_{2,t}a_t + a_t a_{t+1} & (0, 3X_{2,t} + a_{t+1})(0, 3X_{2,t-1} + a_t) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0, 3 & 0, 33 \end{bmatrix},$$

pois  $E(a_t a_{t+1}) = 0$ ,  $E(X_{2,t-1} a_{t+1}) = 0$ ,  $E(X_{2,t} a_t) = E(a_t^2) = 1$ ,  $E(X_{2,t} X_{2,t-1}) = 0, 3E(X_{2,t-1}^2) = (0, 3)(0, 91)$ , de modo que o valor de  $\gamma_{2,2}(1)$  é dado por  $E(0, 09X_{2,t} X_{2,t-1} + 0, 3X_{2,t} a_t + 0, 3X_{2,t-1} a_{t+1} + a_t a_{t+1}) = (0, 09)(0, 3)/0, 91 + 0, 3 = 0, 3/0, 91 = 0, 33$ .

Segue-se que

$$\Gamma(-1) = \Gamma'(1) = \begin{bmatrix} 0 & 0, 3 \\ 0 & 0, 33 \end{bmatrix}.$$

De modo análogo podemos calcular  $\Gamma(2)$ ,  $\Gamma(3)$ , etc.

Dizemos que a série  $\{\mathbf{a}_t, t \in Z\}$  é um *ruído branco multivariado* ( $n \times 1$ ), com média  $\mathbf{0}$  e matriz de covariâncias  $\Sigma$  se  $\mathbf{a}_t$  é estacionário com média  $\mathbf{0}$  e sua matriz de covariâncias é dada por

$$\Gamma(\tau) = \begin{cases} \Sigma, & \text{se } \tau = 0, \\ \mathbf{0}, & \text{se } \tau \neq 0. \end{cases} \quad (7.10)$$

Usaremos a notação  $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \Sigma)$ . Se além disso os vetores  $\mathbf{a}_t$  forem independentes e identicamente distribuídos, escreveremos  $\mathbf{a}_t \sim \text{IID}(\mathbf{0}, \Sigma)$ .

Um processo  $\mathbf{X}_t$  diz-se *linear* se

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \Psi_j \mathbf{a}_{t-j}, \quad (7.11)$$

onde  $\mathbf{a}_t$  é ruído branco multivariado e  $\Psi_j$  é uma sequência de matrizes cujas componentes são absolutamente somáveis. Segue-se que  $E(\mathbf{X}_t) = \mathbf{0}$  e a matriz de covariâncias de  $\mathbf{X}_t$  é dada por

$$\Gamma(\tau) = \sum_{j=0}^{\infty} \Psi_{j+\tau} \Sigma \Psi_j', \quad \tau \in Z. \quad (7.12)$$

### 7.3 Estimação de Médias e Covariâncias

Supondo que temos observações  $\{\mathbf{X}_t, t = 1, \dots, T\}$  do processo estacionário  $\{\mathbf{X}_t, t \in Z\}$ , a média  $\boldsymbol{\mu}$  pode ser estimada pelo vetor de médias amostrais

$$\bar{\mathbf{X}} = \frac{\sum_{t=1}^T \mathbf{X}_t}{T}. \quad (7.13)$$

Segue-se que a média  $\mu_j$  de  $X_{jt}$  é estimada por  $\sum_{t=1}^T X_{jt}/T$ .

Pode-se demonstrar, sob diversas condições sobre o processo  $\mathbf{X}_t$ , que:

- (i)  $E(\bar{\mathbf{X}} - \boldsymbol{\mu})'(\bar{\mathbf{X}} - \boldsymbol{\mu}) \rightarrow 0$ , se  $\gamma_{ii}(\tau) \rightarrow 0$ ,  $i = 1, \dots, n$ ;
- (ii)  $TE(\bar{\mathbf{X}} - \boldsymbol{\mu})'(\bar{\mathbf{X}} - \boldsymbol{\mu}) \rightarrow \sum_{i=1}^n \sum_{\tau} \gamma_{ii}(\tau)$ , se  $\sum_{\tau} |\gamma_{ii}(\tau)| < \infty$ ,  $i = 1, \dots, n$ ;
- (iii) o vetor  $\bar{\mathbf{X}}$  tem distribuição assintótica normal multivariada.

Veja Brockwell e Davis (1991) para detalhes. Para estimar  $\Gamma(\tau)$  usamos

$$\hat{\Gamma}(\tau) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-\tau} (\mathbf{X}_{t+\tau} - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})', & 0 \leq \tau \leq T-1 \\ \frac{1}{T} \sum_{t=-\tau+1}^T (\mathbf{X}_{t+\tau} - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})', & -T+1 \leq \tau \leq 0. \end{cases} \quad (7.14)$$

A matriz de correlações pode ser estimada por

$$\hat{\boldsymbol{\rho}}(\tau) = \hat{\mathbf{D}}^{-1} \hat{\Gamma} \hat{\mathbf{D}}^{-1}, \quad (7.15)$$

onde  $\hat{\mathbf{D}}$  é a matriz diagonal  $n \times n$  dos desvios padrões amostrais das séries individuais.

Veja Fuller (1996) para propriedades dos estimadores  $\hat{\Gamma}(\tau)$  e  $\hat{\boldsymbol{\rho}}(\tau)$ .

**Exemplo 7.4.** Suponha que  $X_{1t}$  represente os retornos diários da Petrobrás e  $X_{2t}$  os retornos diários do Ibovespa, de 3/1/95 a 27/12/2000,  $T = 1498$  e  $\mathbf{X}_t = (X_{1t}, X_{2t})'$ . O programa SCA fornece as matrizes de correlações

amostrais da Tabela 7.1. Uma maneira conveniente de representar estas matrizes é usar os símbolos  $+$ ,  $-$  e  $\cdot$ , quando o valor de uma correlação cruzada for, respectivamente, maior ou igual a  $2/\sqrt{T}$ , menor ou igual a  $-2/\sqrt{T}$  ou estiver entre  $-2/\sqrt{T}$  e  $2/\sqrt{T}$ . Estas matrizes pictóricas também estão apresentadas na Tabela 7.1.

Vemos, por exemplo, que

$$\hat{\rho}(0) = \begin{bmatrix} 1,00 & 0,83 \\ 0,83 & 1,00 \end{bmatrix},$$

enquanto

$$\hat{\rho}(1) = \begin{bmatrix} 0,05 & 0,02 \\ 0,10 & 0,11 \end{bmatrix}.$$

Como  $1/\sqrt{1498} = 0,02584$ , o elemento  $\rho_{12}(1)$  pode ser considerado estatisticamente nulo, de modo que a representação pictórica dessa matriz de correlações amostrais é

$$\begin{bmatrix} + & \cdot \\ + & + \end{bmatrix}.$$

Note que a correlação contemporânea entre as duas séries é 0,83.

Tabela 7.1: Matrizes de correlações amostrais para retornos diários do Ibovespa e Petrobrás, com notação simplificada.

lag 1	lag 2	lag 3	lag 4
$\begin{bmatrix} 0,05 & 0,02 \\ 0,10 & 0,11 \end{bmatrix}$	$\begin{bmatrix} 0,00 & -0,01 \\ 0,00 & -0,02 \end{bmatrix}$	$\begin{bmatrix} -0,06 & -0,08 \\ -0,04 & -0,06 \end{bmatrix}$	$\begin{bmatrix} -0,06 & -0,05 \\ -0,02 & -0,04 \end{bmatrix}$
$\begin{bmatrix} + & \cdot \\ + & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} - & - \\ \cdot & - \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & \cdot \end{bmatrix}$

**Exemplo 7.5.** Consideremos, agora, a série bivariada, consistindo dos retornos mensais do Ibovespa e da taxa de juros dos títulos C-Bond da dívida

brasileira, ambas de julho de 1994 a agosto de 2001,  $T = 86$ . Na Tabela 7.2 temos as matrizes de correlações amostrais até o lag 4. Vemos que  $\boldsymbol{\rho}(\tau)$ ,  $\tau = 1, 2, \dots, 4$ , podem ser consideradas nulas, o que sugere que estamos na presença de um ruído branco bivariado. É fácil verificar que a correlação contemporânea entre as duas séries é negativa  $(-0,77)$ .

Tabela 7.2: Matrizes de correlações amostrais para os retornos mensais do Ibovespa e C-Bond, com notação simplificada.

lag 1		lag 2		lag 3		lag 4	
0,11	-0,05	-0,20	0,08	-0,10	0,07	0,09	-0,12
0,10	-0,15	0,08	-0,04	0,04	-0,06	-0,04	0,07
⋮ ⋮		⋮ ⋮		⋮ ⋮		⋮ ⋮	

## 7.4 Modelos Auto-regressivos Vetoriais

Nesta seção estudaremos uma importante classe de modelos lineares multivariados, a dos modelos auto-regressivos vetoriais de ordem  $p$ , que denotaremos por  $\text{VAR}(p)$  (de “vector autoregression”). Não confundir a notação  $\text{VAR}$  com  $\text{VaR}$ , de valor em risco, utilizada no Capítulo 6.

Dizemos que o processo  $\mathbf{X}_t$ , de ordem  $n \times 1$ , segue um modelo  $\text{VAR}(p)$  se

$$\mathbf{X}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p} + \mathbf{a}_t, \quad (7.16)$$

onde  $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Phi}_0 = (\phi_{10}, \dots, \phi_{n0})'$  é um vetor  $n \times 1$  de constantes e  $\boldsymbol{\Phi}_k$  são matrizes  $n \times n$  constantes, com elementos  $\phi_{ij}^{(k)}$ ,  $i, j = 1, \dots, n$ ,  $k = 1, \dots, p$ . Se  $\mathbf{I}_n$  é a matriz identidade de ordem  $n$ , o modelo (7.16) pode ser escrito na forma

$$\boldsymbol{\Phi}(B) \mathbf{X}_t = \boldsymbol{\Phi}_0 + \mathbf{a}_t, \quad (7.17)$$

onde  $\boldsymbol{\Phi}(B) = \mathbf{I}_n - \boldsymbol{\Phi}_1 B - \dots - \boldsymbol{\Phi}_p B^p$  é o operador auto-regressivo vetorial de ordem  $p$ , ou ainda, um polinômio matricial  $n \times n$  em  $B$ . O elemento genérico de  $\boldsymbol{\Phi}(B)$  é  $[\delta_{ij} - \phi_{ij}^{(1)} B - \dots - \phi_{ij}^{(p)} B^p]$ , para  $i, j = 1, \dots, n$  e  $\delta_{ij} = 1$ , se  $i = j$  e igual a zero, caso contrário.



Vamos, agora, considerar, por simplicidade, o modelo VAR(1), ou seja,

$$\mathbf{X}_t = \Phi_0 + \Phi \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (7.18)$$

Um caso especial é quando  $n = 2$  e (7.18) reduz-se a

$$\begin{aligned} X_{1t} &= \phi_{10} + \phi_{11}X_{1,t-1} + \phi_{12}X_{2,t-1} + a_{1t}, \\ X_{2t} &= \phi_{20} + \phi_{21}X_{1,t-1} + \phi_{22}X_{2,t-1} + a_{2t}, \end{aligned} \quad (7.19)$$

onde desprezamos o índice 1 em  $\Phi_1$  e em  $\phi_{ij}^{(1)}$ . Denotemos os elementos de  $\Sigma$  por  $\sigma_{ij}$ ,  $i, j = 1, 2$ .

Observe que em (7.19) não fica explicitada a dependência *contemporânea* entre  $X_{1t}$  e  $X_{2t}$ . Dizemos que (7.19) e (7.18) são *modelos em forma reduzida*. É possível obter o modelo na forma *estrutural*, em que esta relação fica explicitada. Veja o Apêndice 7.D para detalhes. O modelo em forma reduzida é preferido por facilidades de estimação e previsão.

Retomemos (7.19). Se  $\phi_{12} = 0$ , a série  $X_{1t}$  não dependerá de  $X_{2,t-1}$  e, de modo análogo, se  $\phi_{21} = 0$ ,  $X_{2,t}$  não dependerá de  $X_{1,t-1}$ . Por outro lado, se  $\phi_{12} = 0$  e  $\phi_{21} \neq 0$ , existe uma relação linear unidirecional de  $X_{1t}$  para  $X_{2t}$ . Se  $\phi_{12} = \phi_{21} = 0$  dizemos que não existe relação linear entre as séries, ou que elas são *não-acopladas*. Finalmente, se  $\phi_{12} \neq 0$ ,  $\phi_{21} \neq 0$ , dizemos que existe uma relação de *feedback* entre as duas séries. Note também que se  $\sigma_{12} = 0$  em  $\Sigma$ , não existe relação linear contemporânea entre  $X_{1t}$  e  $X_{2t}$ .

O processo  $\mathbf{X}_t$  em (7.18) será estacionário se a média for constante e  $E(\mathbf{X}_{t+\tau} \mathbf{X}_t')$  independente de  $t$ . Neste caso, se  $\mu = E(\mathbf{X}_t)$ , teremos

$$\mu = (\mathbf{I}_n - \Phi)^{-1} \Phi_0.$$

Segue-se que o modelo poderá ser escrito na forma

$$\mathbf{X}_t - \mu = \Phi(\mathbf{X}_{t-1} - \mu) + \mathbf{a}_t,$$

ou ainda, se  $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \mu$ ,

$$\tilde{\mathbf{X}}_t = \Phi \tilde{\mathbf{X}}_{t-1} + \mathbf{a}_t. \quad (7.20)$$

Assim como no caso de um AR(1) univariado, obtemos de (7.20) que

$$\tilde{\mathbf{X}}_t = \mathbf{a}_t + \Phi \mathbf{a}_{t-1} + \Phi^2 \mathbf{a}_{t-2} + \dots, \quad (7.21)$$

ou seja, temos a representação MA( $\infty$ ) do modelo. Também, é fácil ver que  $\text{Cov}(\mathbf{a}_t, \mathbf{X}_{t-1}) = \mathbf{0}$  e  $\text{Cov}(\mathbf{a}_t, \mathbf{X}_t) = \Sigma$ .

Iremos indicar por  $|\mathbf{A}|$  o determinante da matriz quadrada  $\mathbf{A}$ .

**Proposição 7.2.** O processo  $\mathbf{X}_t$  seguindo um modelo VAR(1) será estacionário se todas as soluções de

$$|\mathbf{I}_n - \Phi z| = 0 \quad (7.22)$$

estiverem fora do círculo unitário.

Como as soluções de (7.22) são inversas dos autovalores de  $\Phi$ , uma condição equivalente é que todos os auto-valores de  $\Phi$  sejam menores do que um, em módulo. Ou ainda,  $|\mathbf{I}_n - \Phi z| \neq 0$ ,  $|z| \leq 1$ . A demonstração da Proposição 7.2 está dada no Apêndice 7.B.

**Exemplo 7.6.** No caso de um VAR(1) bivariado, temos que (7.22) fica

$$\begin{bmatrix} 1 - \phi_{11}z & -\phi_{12}z \\ -\phi_{21}z & 1 - \phi_{22}z \end{bmatrix} = (1 - \phi_{11}z)(1 - \phi_{22}z) - \phi_{12}\phi_{21}z^2 = 0,$$

ou seja, obtemos a equação

$$1 - \text{tr}(\Phi)z - |\Phi|z^2 = 0,$$

onde  $\text{tr}(\Phi) = \phi_{11} + \phi_{22}$  indica o traço de  $\Phi$ . Logo as duas séries são (conjuntamente) estacionárias se as soluções desta equação de segundo grau estiverem fora do círculo unitário. Por exemplo, se

$$\Phi = \begin{bmatrix} 0,5 & 0,3 \\ -0,6 & -0,1 \end{bmatrix},$$

então  $\text{tr}(\Phi) = 0,4$ ,  $|\Phi| = 0,13$  e as raízes da equação terão módulos maiores do que um.

**Exemplo 7.7.** Consideremos o modelo VAR(1) ( $n = 2$ )

$$\begin{aligned} X_{1,t} &= 0,4 + 0,5X_{1,t-1} + 0,3X_{2,t-1} + a_{1,t}, \\ X_{2,t} &= -1,7 - 0,6X_{1,t-1} - 0,1X_{2,t-1} + a_{2,t}, \end{aligned}$$

e vamos simulá-lo usando a função `simulate.VAR` do `S+FinMetrics`. Aqui,

$$\Phi_1 = \begin{bmatrix} 0,5 & 0,3 \\ -0,6 & -0,1 \end{bmatrix}, \quad \Phi_0 = \begin{bmatrix} 0,4 \\ -1,7 \end{bmatrix}, \quad \mu = \begin{bmatrix} 2,0 \\ -1,0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix}.$$

Temos na Figura 7.1 as duas séries simuladas. É fácil ver que este modelo é estacionário.

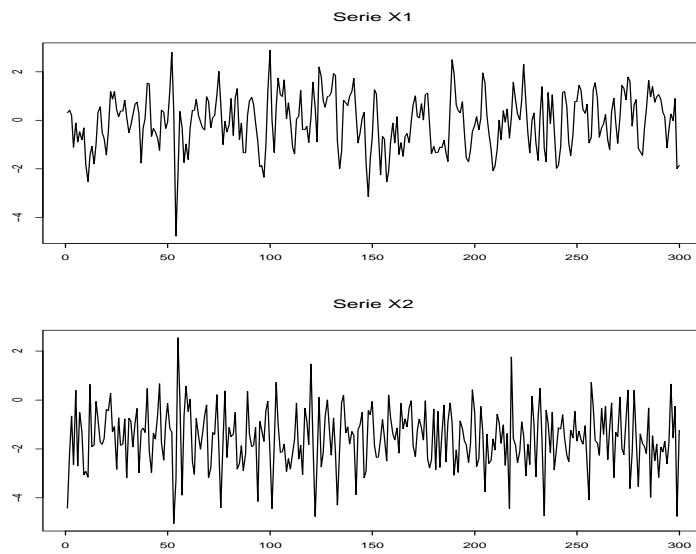


Figura 7.1: Modelo VAR(1) estacionário simulado

Calculemos a matriz de covariâncias de  $\mathbf{X}_t$ , admitindo-se o modelo (7.20). Usando (7.21) temos que

$$\mathbf{\Gamma}(0) = \mathbf{\Sigma} + \mathbf{\Phi}\mathbf{\Sigma}\mathbf{\Phi}' + \mathbf{\Phi}^2\mathbf{\Sigma}(\mathbf{\Phi}^2)' + \dots = \sum_{j=0}^{\infty} \mathbf{\Phi}^j\mathbf{\Sigma}(\mathbf{\Phi}^j)', \quad \mathbf{\Phi}_0^0 = \mathbf{I}_n.$$

Uma fórmula análoga vale para  $\mathbf{\Gamma}(\tau)$ , veja o Problema 4. No entanto essas fórmulas envolvem somas infinitas, pouco úteis para cálculos. Vejamos uma maneira mais atraente.

Se pós-multiplicarmos (7.20) por  $\tilde{\mathbf{X}}'_{t-\tau}$  e tomarmos a esperança, obteremos

$$E(\tilde{\mathbf{X}}_t\tilde{\mathbf{X}}'_{t-\tau}) = \mathbf{\Phi}E(\tilde{\mathbf{X}}_{t-1}\tilde{\mathbf{X}}'_{t-\tau}) + E(\mathbf{a}_t\tilde{\mathbf{X}}'_{t-\tau}).$$

Fazendo  $\tau = 0$  obtemos

$$\mathbf{\Gamma}(0) = \mathbf{\Phi}\mathbf{\Gamma}(-1) + \mathbf{\Sigma} = \mathbf{\Phi}\mathbf{\Gamma}(1)' + \mathbf{\Sigma}.$$

Ou seja, para calcular  $\mathbf{\Gamma}(0)$  necessitamos de  $\mathbf{\Phi}$ ,  $\mathbf{\Sigma}$  e  $\mathbf{\Gamma}(1)$ . Como o termo  $E(\mathbf{a}_t\tilde{\mathbf{X}}'_{t-\tau})$  é nulo para  $\tau > 0$ , obtemos

$$\mathbf{\Gamma}(\tau) = \mathbf{\Phi}\mathbf{\Gamma}(\tau - 1), \quad \tau > 0,$$

e por substituições sucessivas encontramos

$$\mathbf{\Gamma}(\tau) = \mathbf{\Phi}^\tau\mathbf{\Gamma}(0), \quad \tau > 0. \quad (7.23)$$

Segue-se que  $\mathbf{\Gamma}(1) = \mathbf{\Phi}\mathbf{\Gamma}(0)$ , donde

$$\mathbf{\Gamma}(0) = \mathbf{\Phi}\mathbf{\Gamma}(0)\mathbf{\Phi}' + \mathbf{\Sigma},$$

e tomando o operador  $\text{vec}$  em ambos os membros obtemos

$$\text{vec}(\mathbf{\Gamma}(0)) = \text{vec}(\mathbf{\Phi}\mathbf{\Gamma}(0)\mathbf{\Phi}') + \text{vec}(\mathbf{\Sigma}),$$

e como  $\text{vec}(\mathbf{\Phi}\mathbf{\Gamma}(0)\mathbf{\Phi}') = (\mathbf{\Phi} \otimes \mathbf{\Phi})\text{vec}(\mathbf{\Gamma}(0))$ , obtemos finalmente

$$\text{vec}(\mathbf{\Gamma}(0)) = (\mathbf{I}_{n^2} - \mathbf{\Phi} \otimes \mathbf{\Phi})^{-1}\text{vec}(\mathbf{\Sigma}).$$

Veja o Apêndice A para noções sobre o operador  $\text{vec}$  e o produto de Kronecker  $\otimes$  entre duas matrizes.

Nesta expressão a inversa existe pois  $|\mathbf{I}_{n^2} - \Phi \otimes \Phi| \neq 0$ , devido ao fato que os autovalores do produto  $\Phi \otimes \Phi$  são iguais aos produtos dos autovalores de  $\Phi$ , logo têm módulos menores do que 1.

Os resultados anteriores podem ser estendidos a processos VAR( $p$ ),  $p > 1$ , devido ao fato que tal processo pode sempre ser escrito na forma VAR(1). Veja o Apêndice 7.C. Segue-se que a seguinte proposição é válida.

**Proposição 7.3.** Para o modelo VAR( $p$ ) dado em (7.16) temos os seguintes resultados:

(i) O processo  $\mathbf{X}_t$  será estacionário se as soluções de

$$|\mathbf{I}_n - \Phi_1 B - \dots - \Phi_p B^p| = 0$$

estiverem fora do círculo unitário.

(ii) Se  $\mathbf{X}_t$  for estacionário,

$$\mu = E(\mathbf{X}_t) = (\mathbf{I}_n - \Phi_1 B - \dots - \Phi_p B^p)^{-1} \Phi_0.$$

(iii) Escrevendo (7.16) na forma

$$\tilde{\mathbf{X}}_t = \Phi_1 \tilde{\mathbf{X}}_{t-1} + \dots + \Phi_p \tilde{\mathbf{X}}_{t-p} + \mathbf{a}_t,$$

com  $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \mu$  e multiplicando esta equação por  $\tilde{\mathbf{X}}_{t-\tau}'$  obtemos

$$\Gamma(\tau) = \Phi_1 \Gamma(\tau - 1) + \dots + \Phi_p \Gamma(\tau - p), \quad \tau > 0,$$

que são as equações de Yule-Walker no caso de um modelo VAR( $p$ ).

A demonstração de (i) está no Apêndice 7.C e (ii) e (iii) são imediatas. Para uma condição equivalente a (i) veja o Problema 12. Observe que

$$\begin{aligned} \Gamma(0) &= \Phi_1 \Gamma(-1) + \dots + \Phi_p \Gamma(-p) + \Sigma \\ &= \Phi_1 \Gamma(1)' + \dots + \Phi_p \Gamma(p)' + \Sigma. \end{aligned}$$

Estas equações podem ser utilizadas para calcular  $\Gamma(\tau)$  recursivamente, para  $\tau \geq p$ . Para  $|\tau| < p$ , temos que usar a representação VAR(1) de um processo VAR( $p$ ). Veja o Apêndice 7.C.

## 7.5 Construção de Modelos VAR

A construção de modelos VAR segue o mesmo ciclo de identificação, estimação e diagnóstico usado para modelos univariados da classe ARMA.

### Identificação

Uma maneira de identificar a ordem  $p$  de um modelo VAR( $p$ ) consiste em ajustar sequencialmente modelos auto-regressivos vetoriais de ordens  $1, 2, \dots, k$  e testar a significância dos coeficientes (matrizes). Considere, pois, os modelos

$$\begin{aligned} \mathbf{X}_t &= \Phi_0^{(1)} + \Phi_1^{(1)} \mathbf{X}_{t-1} + \mathbf{a}_t^{(1)}, \\ \mathbf{X}_t &= \Phi_0^{(2)} + \Phi_1^{(2)} \mathbf{X}_{t-1} + \Phi_2^{(2)} \mathbf{X}_{t-2} + \mathbf{a}_t^{(2)}, \\ \dots &\quad \dots \\ \mathbf{X}_t &= \Phi_0^{(k)} + \Phi_1^{(k)} \mathbf{X}_{t-1} + \dots + \Phi_k^{(k)} \mathbf{X}_{t-k} + \mathbf{a}_t^{(k)}. \end{aligned} \tag{7.24}$$

Os parâmetros podem ser estimados por MQ ordinários, que fornecem estimadores consistentes e eficientes. Testamos, então,

$$\begin{aligned} H_0 &: \Phi_k^{(k)} = \mathbf{0}, \\ H_1 &: \Phi_k^{(k)} \neq \mathbf{0}, k = 1, 2, \dots \end{aligned} \tag{7.25}$$

O teste da razão de verossimilhanças é baseado nas estimativas das matrizes de covariâncias dos resíduos dos modelos ajustados. Para a  $k$ -ésima equação, considere

$$\hat{\mathbf{a}}_t^{(k)} = \mathbf{X}_t - \hat{\Phi}_0^{(k)} - \hat{\Phi}_1^{(k)} \mathbf{X}_{t-1} - \dots - \hat{\Phi}_k^{(k)} \mathbf{X}_{t-k}.$$

A matriz de covariância dos resíduos, que estima  $\Sigma$ , é dada então por

$$\hat{\Sigma}_k = \frac{1}{T-k} \sum_{t=k+1}^T \hat{\mathbf{a}}_t^{(k)} (\hat{\mathbf{a}}_t^{(k)})', \quad k \geq 0, \quad (7.26)$$

onde para  $k = 0$ ,  $\hat{\mathbf{a}}_t^{(0)} = \mathbf{X}_t - \bar{\mathbf{X}}$ . A estatística da razão de verossimilhanças para o teste (7.25) é dada por

$$\text{RV}(k) = (T-k) \ln \frac{|\hat{\Sigma}_{k-1}|}{|\hat{\Sigma}_k|}, \quad (7.27)$$

que tem distribuição qui-quadrado com  $n^2$  graus de liberdade,  $\chi^2(n^2)$ .

Outra maneira de identificar a ordem de um VAR é usar algum critério de informação, como:

$$\begin{aligned} \text{AIC}(k) &= \ln(|\hat{\Sigma}_k|) + 2kn^2/T \quad (\text{Akaike}), \\ \text{BIC}(k) &= \ln(|\hat{\Sigma}_k|) + kn^2 \ln(T)/T \quad (\text{Schwarz}), \\ \text{HQC}(k) &= \ln(|\hat{\Sigma}_k|) + kn^2 \ln(\ln(T))/T \quad (\text{Hannan-Quinn}). \end{aligned} \quad (7.28)$$

O programa EViews, por exemplo, fornece esses critérios mais os valores de FPE (“final prediction error”, de Akaike),  $\text{RV}(k)$  e o valor da log-verossimilhança. O programa SCA fornece os valores de  $\text{RV}(k)$  e AIC.

### Estimação

Identificado o valor de  $p$  e supondo  $\mathbf{a}_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$ , podemos estimar os coeficientes por máxima verossimilhança. Neste caso, os estimadores de MQ são equivalentes a estimadores de MV condicionais.

No caso de um VAR(1), os EMV condicionais são obtidos maximizando-se

$$\begin{aligned} \ell &= -\frac{n(T+1)}{2} \ln(2\pi) + \frac{(T-1)}{2} \ln |\Sigma^{-1}| \\ &\quad - \frac{1}{2} \sum_{t=2}^T (\mathbf{X}_t - \Phi \mathbf{X}_{t-1})' \Sigma^{-1} (\mathbf{X}_t - \Phi \mathbf{X}_{t-1}), \end{aligned} \quad (7.29)$$

obtendo-se

$$\hat{\Phi} = \left[ \sum_{t=2}^T \mathbf{X}_t \mathbf{X}_{t-1}' \right] \left[ \sum_{t=2}^T \mathbf{X}_{t-1} \mathbf{X}_{t-1}' \right]^{-1}, \quad (7.30)$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{a}}_t (\hat{\mathbf{a}}_t)', \quad (7.31)$$

$$\hat{\mathbf{a}}_t = \mathbf{X}_t - \hat{\Phi} \mathbf{X}_{t-1}. \quad (7.32)$$

No caso geral de um VAR( $p$ ), os EMV condicionais são obtidos por métodos de maximização numérica.

### Diagnóstico

Para testar se o modelo é adequado, usamos os resíduos para construir a versão multivariada da estatística de Box-Ljung-Pierce, dada por

$$Q(m) = T^2 \sum_{\tau=1}^m \frac{1}{T-\tau} \text{tr}(\hat{\Gamma}(\tau)' \hat{\Gamma}(0)^{-1} \hat{\Gamma}(\tau) \hat{\Gamma}(0)^{-1}), \quad (7.33)$$

que sob  $H_0$  : a série  $\mathbf{a}_t$  é ruído branco, tem distribuição  $\chi^2(n^2(m-p))$ . Para que o número de graus de liberdade seja positivo,  $m$  deve ser maior do que  $p$ .

### Previsão

Considere o VAR(1) dado em (7.18) e suponha que o parâmetro  $\Phi$  seja conhecido. A previsão de origem  $T$  e horizonte  $h$  é dada por

$$\hat{\mathbf{X}}_T(h) = \Phi \hat{\mathbf{X}}_T(h-1),$$

da qual segue

$$\hat{\mathbf{X}}_T(h) = \Phi^h \mathbf{X}_T, \quad h = 1, 2, \dots \quad (7.34)$$

Como

$$\mathbf{X}_{t+h} = \Phi \mathbf{X}_{t+h-1} + \mathbf{a}_{t+h},$$



temos que o erro de previsão  $h$  passos a frente é dado por

$$\mathbf{e}_T(h) = \mathbf{X}_{t+h} - \hat{\mathbf{X}}_T(h) = \sum_{j=0}^{h-1} \Phi^j \mathbf{a}_{T+h-j}, \quad (7.35)$$

de modo que o erro quadrático médio do previsor (7.34) fica

$$\Sigma(h) = \text{EQMP}(h) = \sum_{j=0}^{h-1} \Phi^j \Sigma(\Phi^j)'. \quad (7.36)$$

Considerando, agora, o modelo VAR( $p$ ), com parâmetros supostos conhecidos,  $\mathbf{a}_t$  uma sequência i.i.d. e  $\mathcal{F}_t = \{\mathbf{X}_s : s \leq t\}$ , obtemos

$$E(\mathbf{X}_{t+h}|\mathcal{F}_t) = \Phi_0 + \Phi_1 E(\mathbf{X}_{t+h-1}|\mathcal{F}_t) + \dots + \Phi_p E(\mathbf{X}_{t+h-p}|\mathcal{F}_t),$$

pois  $E(\mathbf{a}_{t+h}|\mathcal{F}_t) = 0$ , para todo  $h > 0$ .

Para  $h = 1$  obtemos

$$\hat{\mathbf{X}}_t(1) = \Phi_0 + \Phi_1 \mathbf{X}_t + \dots + \Phi_p \mathbf{X}_{t-p+1},$$

e para  $h = 2$  temos

$$\hat{\mathbf{X}}_t(2) = \Phi_0 + \Phi_1 \hat{\mathbf{X}}_t(1) + \Phi_2 \mathbf{X}_t + \dots + \Phi_p \mathbf{X}_{t-p+2},$$

de modo que as previsões podem ser obtidas recursivamente.

Neste caso, o erro de previsão de horizonte  $h$  é dado por

$$\mathbf{e}_T(h) = \sum_{j=0}^{h-1} \Psi_j \mathbf{a}_{T+h-j}, \quad (7.37)$$

onde as matrizes  $\Psi_j$  são obtidas recursivamente por

$$\Psi_j = \sum_{k=1}^{p-1} \Psi_{j-k} \Phi_k, \quad (7.38)$$

com  $\Psi_0 = \mathbf{I}_n$  e  $\Phi_j = 0, j > p$ . Segue-se que a matriz de EQM de previsão fica

$$\Sigma(h) = \sum_{j=0}^{h-1} \Psi_j \Sigma \Psi_j' \quad (7.39)$$

Quando os parâmetros do modelo VAR(p) são estimados, o melhor preditor de  $X_{T+h}$  é agora dado por

$$\tilde{\mathbf{X}}_T(h) = \hat{\Phi}_0 + \hat{\Phi}_1 \tilde{\mathbf{X}}_T(h-1) + \dots + \hat{\Phi}_p \tilde{\mathbf{X}}_T(h-p), \quad h > 1. \quad (7.40)$$

Neste caso, a matriz de EQM de previsão torna-se

$$\hat{\Sigma}(h) = \Sigma(h) + EQM(\mathbf{X}_{T+h} - \tilde{\mathbf{X}}_T(h)). \quad (7.41)$$

Na prática, o segundo termo em (7.41) é ignorado e  $\hat{\Sigma}(h)$  é calculada por

$$\hat{\Sigma}(h) = \sum_{j=0}^{h-1} \hat{\Psi}_j \hat{\Sigma} \hat{\Psi}_j', \quad (7.42)$$

com  $\hat{\Psi}_j = \sum_{k=1}^{p-1} \hat{\Psi}_{j-k} \hat{\Phi}_k$ . Lütkepohl (1991) dá uma aproximação para o segundo membro de (7.41).

**Exemplo 7.8.** Retornemos ao exemplo 7.4 e ajustemos um modelo VAR(p) à série  $\mathbf{X}_t$ , onde  $X_{1t}$  é a série de retornos diários do Ibovespa e  $X_{2t}$  é a série de retornos diários da Petrobrás, com  $T = 1498$  observações. Na Tabela 7.3 temos os resultados de ajustes de modelos auto-regressivos até ordem 8. Usando os valores de (7.27) ou dos AIC correspondentes, selecionamos a ordem  $p = 6$ . A tabela foi obtida usando-se o programa SCA.

Vamos usar, agora, o programa EViews para estimar o modelo VAR(6). A Tabela 7.4 é um resumo da saída do programa. O vetor constante, não significativo, tem componentes  $\phi_{10} = 0,00128(0,0007)$  e  $\phi_{20} = 0,00143(0,0009)$ , de modo que o modelo bivariado ajustado é

$$\begin{aligned} X_{1t} &= 0,1018X_{1,t-1} - 0,1113X_{1,t-6} - 0,0790X_{2,t-3} + a_{1t}, \\ X_{2t} &= -0,1338X_{1,t-6} + 0,1111X_{2,t-1} - 0,0973X_{2,t-3} + a_{2,t}. \end{aligned} \quad (7.43)$$

Tabela 7.3: Estatísticas resultantes de ajustes de modelos  $\text{VAR}(p)$ ,  $p = 1, \dots, 8$ , para os retornos diários do Ibovespa e Petrobrás.

Ordem	1	2	3	4	5	6	7	8
RV( $k$ )	65,80	5,14	10,14	10,24	13,72	11,00	3,52	4,86
AIC	-18,419	-18,417	-18,419	-18,420	-18,424	-18,427	-18,424	-18,422

Nota: Valores críticos da distribuição qui-quadrado com 4 g.l. são 9,5 (5%) e 13,3 (1%).

Na Tabela 7.5 temos as representações pictóricas dos coeficientes (matriciais). O **S+FinMetrics** fornece valores idênticos. Note que os retornos diários da Petrobrás são influenciados por valores passados dos retornos diários do Ibovespa, ou seja, dependem de preços do mercado financeiro. Por outro lado, os retornos do Ibovespa são influenciados por valores defasados dos retornos da Petrobrás, o que é razoável, dado que as ações da Petrobrás fazem parte do índice. Seque-se que há uma relação de “feedback” entre as duas séries. Do exemplo 7.4, obtemos que o coeficiente de correlação entre as duas séries é 0,83, o que mostra uma relação contemporânea forte.

Obtemos, também, que o determinante da matriz de covariâncias residuais é dado por  $2,78 \times 10^{-7}$ ,  $\log L = 7027,013$ ,  $\text{AIC} = -9,385$  e  $\text{BIC} = -9,292$ . O **EViews** também calcula valores da estatística  $Q$ . Por exemplo,  $Q(7) = 6,92$ , com 4 graus de liberdade e  $p$ -valor 0,14, mostrando que o ajuste é adequado. Na Figura 7.2 mostramos a f.a.c. amostral para os resíduos e quadrados dos resíduos do modelo ajustado. Vemos que há possibilidade de melhorar o modelo, introduzindo termos de médias móveis (veja a seção seguinte) e considerando um modelo heteroscedástico condicional multivariado para os resíduos, dada a dependência presente nos mesmos.

As previsões para horizontes  $h = 1, 2, \dots, 12$  para o modelo (7.43), usando o **S+FinMetrics**, estão na Tabela 7.6 e na Figura 7.3.

## 7.6 Modelos ARMA Vetoriais

O modelo  $\text{VAR}(p)$  estudado na seção anterior é um caso particular dos modelos  $\text{VARMA}(p,q)$  (de “vector autoregressive moving average”), se  $q = 0$ .

Se  $p = 0$  obtemos um modelo de médias móveis vetorial de ordem  $q$ , VMA( $q$ ), dado por

$$\mathbf{X}_t = \boldsymbol{\Theta}_0 + \mathbf{a}_t - \boldsymbol{\Theta}_1 \mathbf{a}_{t-1} - \dots - \boldsymbol{\Theta}_q \mathbf{a}_{t-q}, \quad (7.44)$$

sendo novamente  $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \boldsymbol{\Sigma})$  e  $\boldsymbol{\Theta}_i$  matrizes  $n \times n$  de constantes, e  $\boldsymbol{\Theta}_0$  o vetor  $n \times 1$  de médias do processo.

O modelo pode ser escrito compactamente na forma

$$\mathbf{X}_t = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}(B)\mathbf{a}_t, \quad (7.45)$$

na qual  $\boldsymbol{\Theta}(B) = \mathbf{I}_n - \boldsymbol{\Theta}_1 B - \dots - \boldsymbol{\Theta}_q B^q$  é um polinômio matricial  $n \times n$  de ordem  $q$ .

Tabela 7.4: Ajuste de um modelo VAR(6) aos retornos diários do Ibovespa e Petrobrás. Primeira linha: estimativas; segunda linha: desvios padrões.

$\Phi_1$	$\Phi_2$	$\Phi_3$
$\begin{bmatrix} 0,1018 & -0,0672 \\ -0,0096 & 0,1111 \end{bmatrix}$	$\begin{bmatrix} -0,0257 & -0,0093 \\ 0,0069 & -0,0423 \end{bmatrix}$	$\begin{bmatrix} 0,0197 & -0,0790 \\ 0,0517 & -0,0973 \end{bmatrix}$
$\begin{bmatrix} 0,047 & 0,039 \\ 0,057 & 0,047 \end{bmatrix}$	$\begin{bmatrix} 0,047 & 0,040 \\ 0,057 & 0,048 \end{bmatrix}$	$\begin{bmatrix} 0,047 & 0,040 \\ 0,057 & 0,048 \end{bmatrix}$
$\Phi_4$	$\Phi_5$	$\Phi_6$
$\begin{bmatrix} -0,0689 & 0,0196 \\ -0,0002 & -0,0186 \end{bmatrix}$	$\begin{bmatrix} -0,0363 & -0,0334 \\ 0,0300 & -0,0693 \end{bmatrix}$	$\begin{bmatrix} -0,1113 & 0,0243 \\ -0,1338 & 0,0412 \end{bmatrix}$
$\begin{bmatrix} 0,047 & 0,040 \\ 0,057 & 0,048 \end{bmatrix}$	$\begin{bmatrix} 0,047 & 0,040 \\ 0,057 & 0,048 \end{bmatrix}$	$\begin{bmatrix} 0,047 & 0,039 \\ 0,057 & 0,047 \end{bmatrix}$

Tabela 7.5: Representações pictóricas das matrizes da Tabela 7.4.

$\Phi_1$	$\Phi_2$	$\Phi_3$
$\begin{bmatrix} + & \cdot \\ \cdot & + \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & - \\ \cdot & - \end{bmatrix}$
$\Phi_4$	$\Phi_5$	$\Phi_6$
$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ - & \cdot \end{bmatrix}$

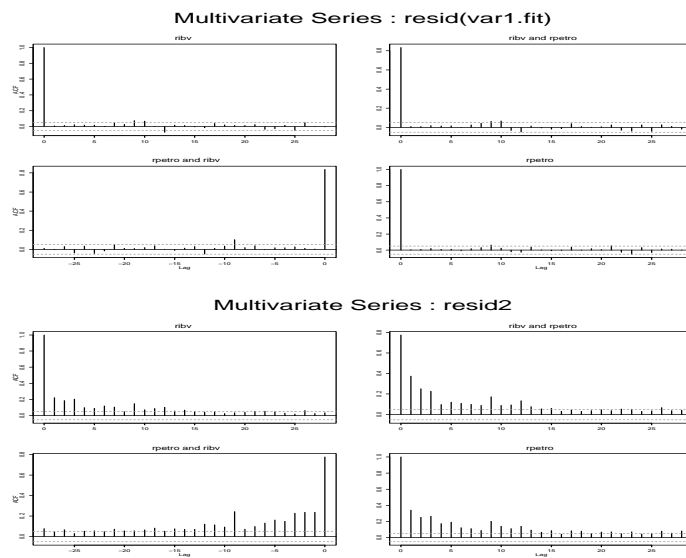


Figura 7.2: F.a.c. e f.c.c. amostrais dos resíduos e seus quadrados

Tabela 7.6: Valores previstos para o modelo (7.43),  $h = 1, 2, \dots, 12$ .

$h$	Ibovespa (e.p.)	Petrobrás (e.p.)
1	0.0031 (0.0282)	0.0003 (0.0340)
2	0.0068 (0.0282)	0.0086 (0.0342)
3	0.0001 (0.0282)	0.0030 (0.0342)
4	-0.0016 (0.0283)	0.0010 (0.0343)
5	-0.0022 (0.0283)	-0.0006 (0.0343)
6	-0.0022 (0.0284)	-0.0021 (0.0344)
7	0.0003 (0.0285)	0.0002 (0.0345)
8	0.0008 (0.0285)	0.0007 (0.0345)
9	0.0017 (0.0285)	0.0016 (0.0345)
10	0.0017 (0.0286)	0.0018 (0.0345)
11	0.0016 (0.0286)	0.0019 (0.0345)
12	0.0013 (0.0286)	0.0017 (0.0345)

**Proposição 7.4.** Para o modelo (7.44) temos:

- (i)  $\text{Cov}(\mathbf{X}_t, \mathbf{a}_t) = \mathbf{\Sigma}$ ;
- (ii)  $\mathbf{\Gamma}(0) = \mathbf{\Sigma} + \mathbf{\Theta}_1 \mathbf{\Sigma} \mathbf{\Theta}_1' + \dots + \mathbf{\Theta}_q \mathbf{\Sigma} \mathbf{\Theta}_q'$ ;
- (iii) a matriz de covariâncias de  $\mathbf{X}_T$  é dada por

$$\mathbf{\Gamma}(\tau) = \begin{cases} \sum_{j=\tau}^q \mathbf{\Theta}_j \mathbf{\Sigma} \mathbf{\Theta}_j', & \text{se } 1 \leq \tau \leq q, \\ \mathbf{0}, & \text{se } \tau > q, \end{cases} \quad (7.46)$$

sendo  $\theta_0 = -\mathbf{I}_n$ .

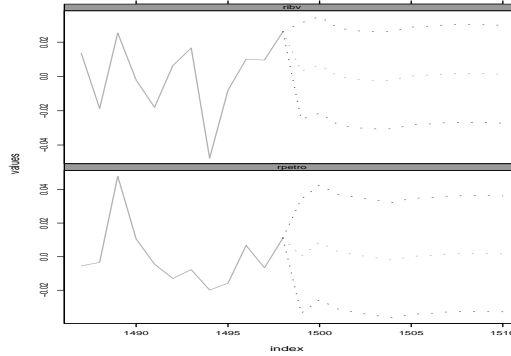


Figura 7.3: Previsões para o modelo (7.43),  $h = 1, \dots, 12$

De (7.46) segue-se que  $\rho(\tau) = \mathbf{0}$ ,  $\tau > q$ , logo a ordem  $q$  de um modelo VMA( $q$ ) pode ser identificada analisando-se as matrizes de correlações cruzadas amostrais.

**Exemplo 7.9.** O modelo de médias móveis vetorial mais simples é o VMA(1), dado por

$$\mathbf{X}_t = \Theta_0 + \mathbf{a}_t - \Theta \mathbf{a}_{t-1},$$

que tem todas as matrizes de correlações nulas a partir do lag 2. Como vimos acima,  $\Theta$  é a média do processo. No caso especial  $n = 2$  obtemos

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \theta_{10} \\ \theta_{20} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} - \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix} \begin{bmatrix} a_{1,t-1} \\ a_{2,t-1} \end{bmatrix}.$$

Assim como no caso de um VAR, relações entre as duas séries podem ser estabelecidas analisando-se os elementos  $\theta_{12}$  e  $\theta_{21}$  da matriz  $\Theta$ . Veja o Problema 1.

Estimativas de máxima verossimilhança condicionais podem ser obtidas, como no caso univariado, calculando-se valores dos choques  $\mathbf{a}_t$  recursivamente,

supondo-se  $\mathbf{a}_t = \mathbf{0}$ ,  $t \leq 0$ . A identificação é feita usando-se as matrizes de correlações amostrais ou os critérios de informação vistos anteriormente. O diagnóstico do modelo identificado é feito usando a estatística de Box-Pierce-Ljung e a previsão como no caso univariado.

O modelo VARMA geral é definido por

$$\begin{aligned} \mathbf{X}_t = & \Phi_0 + \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_p \mathbf{X}_{t-p} + \\ & + \mathbf{a}_t - \Theta_1 \mathbf{a}_{t-1} - \dots - \Theta_q \mathbf{a}_{t-q}, \end{aligned} \quad (7.47)$$

sendo novamente  $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \Sigma)$  e as matrizes  $\Phi_i$  e  $\Theta_i$  como antes.

Com os polinômios matriciais anteriormente definidos, podemos escrever o modelo na forma

$$\Phi(B)\mathbf{X}_t = \Phi_0 + \Theta(B)\mathbf{a}_t. \quad (7.48)$$

Este processo tem uma solução estacionária

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi_j \mathbf{a}_{t-j} \quad (7.49)$$

se  $|\Phi(z)| \neq 0$ , para todo  $z$  complexo tal que  $|z| \leq 1$ , sendo que as matrizes  $\Psi_j$  são determinadas univocamente por  $\Psi(z) = \Phi^{-1}(z)\Theta(z)$ .

Por outro lado, o processo (7.47) pode ser escrito na forma  $\text{AR}(\infty)$

$$\sum_{j=0}^{\infty} \Pi_j \mathbf{X}_{t-j} = \mathbf{a}_t, \quad (7.50)$$

se  $|\Theta(z)| \neq 0$ , para todo complexo  $z$  com  $|z| \leq 1$  e as matrizes  $\Pi_j$  são determinadas univocamente por  $\Pi(z) = \Theta^{-1}(z)\Phi(z)$ . Para detalhes veja Brockwell e Davis (1991). As matrizes  $\Psi_j$  e  $\Pi_j$  podem ser obtidas recursivamente. Veja os problemas 2 e 3.

Os modelos VARMA não são muito usados, devido a dificuldades computacionais e problemas de identificabilidade. A matriz de covariâncias (ou a matriz de densidade espectral) não determina univocamente  $\Phi$ ,  $\Theta$  e  $\Sigma$ , a menos que condições mais restritivas sejam impostas. Veja Dunsmuir e Hannan (1976). Não-identificabilidade implica que a superfície de verossimilhança



não tem um único máximo. Para exemplos de problemas de identificabilidade veja Tsay (2002).

Para a identificação da ordem  $(p, q)$  de uma modelo VARMA os critérios de informação descritos antes podem ser utilizados. A estimação pode ser feita via EMV condicionais ou exatos e a estatística de Box-Pierce-Ljung pode ser usada para fins de diagnóstico.

## 7.7 Causalidade de Granger

A elucidação de relações de causalidade entre variáveis é um dos principais problemas em pesquisa empírica.

Para sistemas temporais, Granger (1969) define causalidade em termos de *previsibilidade*: a variável  $X$  causa a variável  $Y$ , com respeito a um dado universo de informação (que inclui  $X$  e  $Y$ ), se o presente de  $Y$  pode ser previsto mais eficientemente usando valores passados de  $X$ , do que não usando esse passado, toda e qualquer outra informação disponível (incluindo valores passados de  $Y$ ) sendo usada em ambos os casos. A definição não requer que o sistema seja linear; se o for, as previsões serão lineares.

Seja  $\{A_t, t = 0, \pm 1, \pm 2, \dots\}$  o conjunto de informação relevante até (e incluindo) o instante  $t$ , contendo pelo menos  $X_t, Y_t$ . Defina  $\bar{A}_t = \{A_s : s < t\}$ ,  $\bar{\bar{A}}_t = \{A_s : s \leq t\}$ , e definições análogas para  $\bar{X}_t, \bar{Y}_t$ , etc. Seja  $P_t(Y|B)$  o preditor de EQM mínimo de  $Y_t$ , usando o conjunto de informação  $B$  e  $\sigma^2(Y|B)$  o correspondente EQM do preditor.

**Definição 7.1.** Dizemos que:

- (a)  $X_t \rightarrow Y_t$ :  $X_t$  causa  $Y_t$  no sentido de Granger se

$$\sigma^2(Y_t|\bar{A}_t) < \sigma^2(Y_t|\bar{A}_t - \bar{X}_t).$$

Ou seja,  $Y_t$  pode ser melhor prevista usando toda a informação disponível, incluindo o passado de  $Y_t$  e  $X_t$ .

Dizemos também que  $X_t$  é *exógena* ou *antecedente* a  $Y_t$ .

- (b)  $X_t \Rightarrow Y_t$ :  $X_t$  causa instantaneamente  $Y_t$  no sentido de Granger se:

$$\sigma^2(Y_t|A_t, \overline{\overline{X}}_t) < \sigma^2(Y_t|\overline{A}_t)$$

Ou seja, o valor presente de  $Y_t$  é melhor previsto se o valor presente de  $X_t$  for incluído.

- (c) Há *feedback*, e escrevemos  $X_t \leftrightarrow Y_t$ , se  $X_t$  causa  $Y_t$  e  $Y_t$  causa  $X_t$ .
- (d) Há *causalidade unidirecional* de  $X_t$  para  $Y_t$  se  $X_t \rightarrow Y_t$  e não há *feedback*.

É fácil ver que se  $X_t \Rightarrow Y_t$ , então  $Y_t \Rightarrow X_t$ . Portanto usualmente dizemos que há causalidade instantânea entre  $X_t$  e  $Y_t$ .

A definição estende-se para vetores aleatórios. Sejam  $\mathbf{X}_t = (X_{1t}, \dots, X_{rt})'$  e  $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{st})'$  dois vetores aleatórios. Defina  $P(Y_{it}|B_t), \sigma^2(Y_{it}|B_t)$ , etc como antes. O melhor preditor linear de  $\mathbf{Y}_t$  baseado em  $B_t$  é então  $P(\mathbf{Y}_t|B_t) = (P(Y_{1t}|B_t), \dots, P(Y_{st}|B_t))'$ .

**Definição 7.2.** Dizemos que:

- (a) O vetor  $\mathbf{X}_t$  *causa* o vetor  $\mathbf{Y}_t$  se

$$\sigma^2(Y_{it}|\overline{A}_t) < \sigma^2(Y_{it}|\overline{A}_t - \overline{\mathbf{X}}_t),$$

para pelo menos um valor de  $i = 1, \dots, s$ .

- (b) O vetor  $\mathbf{X}_t$  *não causa* o vetor  $\mathbf{Y}_t$  se

$$\sigma^2(Y_{it}|\overline{A}_t) = \sigma^2(Y_{it}|\overline{A}_t - \overline{\mathbf{X}}_t),$$

para todo  $i = 1, 2, \dots, s$ .

Há várias propostas para operacionalizar as definições anteriores. Pierce e Haugh (1977) propõem ajustar modelos ARIMA a transformações adequadas de ambas as séries e depois estabelecer padrões de causalidade entre os resíduos por meio de correlações cruzadas. Veja também Layton (1984). Hsiao (1979) sugere ajustar modelos auto-regressivos via AIC. No caso de mais de duas

séries, Boudjellaba et al. (1992) sugerem ajustar modelos VARMA às séries. Uma resenha desses procedimentos é feita por da Cunha (1997).

Neste capítulo trataremos do assunto por meio da representação VAR da série multivariada  $\mathbf{X}_t$ , de ordem  $n \times 1$ . A representação MA do processo é dada por (7.49), ou seja,

$$\mathbf{X}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(B)\mathbf{a}_t, \quad \boldsymbol{\Psi}_0 = \mathbf{I}_n. \quad (7.51)$$

Suponha que

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix},$$

onde  $\mathbf{Y}_t$  é um vetor  $r \times 1$  e  $\mathbf{Z}_t$  é um vetor  $s \times 1$ ,  $r + s = n$ . Então podemos escrever

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Psi}_{11}(B) & \boldsymbol{\Psi}_{12}(B) \\ \boldsymbol{\Psi}_{21}(B) & \boldsymbol{\Psi}_{22}(B) \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{bmatrix}, \quad (7.52)$$

particionando  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Psi}(B)$  e  $\mathbf{a}_t$  de acordo com a partição de  $\mathbf{X}_t$ . Se houver causalidade unidirecional de  $\mathbf{Y}_t$  para  $\mathbf{Z}_t$ , isto é, se  $\mathbf{Z}_t$  for melhor prevista pelo presente e passado de  $\mathbf{Y}_t$ , mas não o contrário, deveremos ter  $\boldsymbol{\Psi}_{12}(B) = 0$  e obteremos

$$\mathbf{Y}_t = \boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{11}(B)\mathbf{a}_{1t}, \quad (7.53)$$

$$\mathbf{Z}_t = \boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{21}(B)\mathbf{a}_{1t} + \boldsymbol{\Psi}_{22}(B)\mathbf{a}_{2t}. \quad (7.54)$$

Note que (7.54) pode ser escrita

$$\mathbf{Z}_t = \boldsymbol{\mu}_2 + \mathbf{V}(B)\mathbf{Y}_t + \boldsymbol{\psi}_{22}(B)\mathbf{a}_{2t}, \quad (7.55)$$

que é um modelo de regressão dinâmica. As condições  $\boldsymbol{\Psi}_{12}(B) = 0$  e  $\mathbf{V}(B) = 0$  implicam que  $\mathbf{Y}_t$  não causa  $\mathbf{Z}_t$  e vice-versa. Nesta situação, as duas séries serão não-acopladas ou relacionadas apenas instantaneamente.

Na realidade, é possível demonstrar o seguinte resultado, que é uma caracterização de não-causalidade de Granger. Note que, de (7.51),  $\boldsymbol{\Psi}(B) = \mathbf{I}_n + \boldsymbol{\Psi}_1 B + \boldsymbol{\Psi}_2 B^2 + \dots$  e

$$\Psi_i = \begin{bmatrix} \Psi_{11,i} & \Psi_{12,i} \\ \Psi_{21,i} & \Psi_{22,i} \end{bmatrix}, \quad i = 1, 2, \dots$$

**Proposição 7.5.** O previsor ótimo de  $\mathbf{Y}_t$  baseado em  $\bar{\bar{\mathbf{Z}}}_t$  é igual ao previsor ótimo de  $\mathbf{Y}_t$  baseado em  $\bar{\bar{\mathbf{Y}}}_t$  se e somente se  $\Psi_{12,i} = \mathbf{0}$ ,  $i = 1, 2, \dots$

Em outras palavras,  $\mathbf{Z}_t$  não causa  $\mathbf{Y}_t$  se e somente se  $\Psi_{12,i}(B) = \mathbf{0}$ , para todo  $i \geq 1$ . Para a demonstração veja Lütkepohl (1991). Esta proposição aplica-se não só a modelos VAR mas a qualquer processo que possa ser escrito na forma MA infinita, e dá uma maneira de verificar a não-existência de causalidade de Granger. Do ponto de vista prático, convém considerar o modelo VAR de ordem finita, ou seja,

$$\begin{aligned} \mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{11,1} & \boldsymbol{\Phi}_{12,1} \\ \boldsymbol{\Phi}_{21,1} & \boldsymbol{\Phi}_{22,1} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-1} \\ \mathbf{Z}_{t-1} \end{bmatrix} + \dots \\ &+ \begin{bmatrix} \boldsymbol{\Phi}_{11,p} & \boldsymbol{\Phi}_{12,p} \\ \boldsymbol{\Phi}_{21,p} & \boldsymbol{\Phi}_{22,p} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-p} \\ \mathbf{Z}_{t-p} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{bmatrix}, \end{aligned} \quad (7.56)$$

e a condição da Proposição 7.5 estará satisfeita se e somente se  $\boldsymbol{\Phi}_{12,i} = \mathbf{0}$ ,  $i = 1, 2, \dots, p$ . Ou seja, se  $\mathbf{X}_t$  seguir um modelo VAR( $p$ ), com matriz de covariâncias não-singular, então  $\mathbf{Z}_t$  não causa  $\mathbf{Y}_t$  se e somente se  $\boldsymbol{\Phi}_{12,i} = \mathbf{0}$ ,  $i = 1, 2, \dots, p$ .

Uma caracterização de não-existência de causalidade instantânea é dada pela proposição seguinte. A prova é dada em Lütkepohl (1991).

**Proposição 7.6.** Se  $\mathbf{X}_t$  for como em (7.56), com matriz de covariâncias não singular, então *não existe causalidade instantânea* entre  $\mathbf{Y}_t$  e  $\mathbf{Z}_t$  se e somente se  $E(\mathbf{a}_{1t}\mathbf{a}_{2t}') = 0$ .

A prova da proposição é baseada no fato que a matriz de covariâncias  $\boldsymbol{\Sigma}$  pode ser escrita como  $\boldsymbol{\Sigma} = \mathbf{T}\mathbf{T}'$ , onde  $\mathbf{T}$  é uma matriz triangular inferior com elementos positivos na diagonal principal. Segue-se que a representação MA de  $\mathbf{X}_t$  pode ser escrita como

$$\begin{aligned}
\mathbf{X}_t &= \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{T} \mathbf{T}^{-1} \mathbf{a}_{t-j} \\
&= \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Xi}_j \mathbf{b}_{t-j},
\end{aligned}$$

com  $\boldsymbol{\Xi}_j = \boldsymbol{\Psi}_j \mathbf{T}$ ,  $\mathbf{b}_t = \mathbf{T}^{-1} \mathbf{a}_t \sim \text{RB}(\mathbf{0}, \boldsymbol{\Sigma}_b)$ , e  $\boldsymbol{\Sigma}_b = \mathbf{T}^{-1} \boldsymbol{\Sigma} (\mathbf{T}^{-1})' = \mathbf{I}_n$ . Veja também o Apêndice 7.D.

A equação (7.56) pode ser particionada como:

$$\mathbf{Y}_t = \boldsymbol{\mu}_1 + \sum_{i=1}^p \boldsymbol{\Phi}_{11,i} \mathbf{Y}_{t-i} + \sum_{i=1}^p \boldsymbol{\Phi}_{12,i} \mathbf{Z}_{t-i} + \mathbf{a}_{1t}, \quad (7.57)$$

$$\mathbf{Z}_t = \boldsymbol{\mu}_2 + \sum_{i=1}^p \boldsymbol{\Phi}_{21,i} \mathbf{Y}_{t-i} + \sum_{i=1}^p \boldsymbol{\Phi}_{22,i} \mathbf{Z}_{t-i} + \mathbf{a}_{2t}. \quad (7.58)$$

Suponha, também, a matriz  $\boldsymbol{\Sigma}$  particionada como

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix},$$

sendo que  $\boldsymbol{\Sigma}_{ij} = E(\mathbf{a}_{it} \mathbf{a}_{jt}')$ ,  $i, j = 1, 2$ .

Então, como vimos acima:

- (i)  $\mathbf{Z}_t$  não causa  $\mathbf{Y}_t \leftrightarrow \boldsymbol{\Phi}_{12,i} = \mathbf{0}$ , para todo  $i$ ;
- (ii)  $\mathbf{Y}_t$  não causa  $\mathbf{Z}_t \leftrightarrow \boldsymbol{\Phi}_{21,i} = \mathbf{0}$ , para todo  $i$ .

Resultados equivalentes a (i) e (ii) são dados na proposição a seguir.

**Proposição 7.7** (i)  $\mathbf{Z}_t$  não causa  $\mathbf{Y}_t \leftrightarrow |\boldsymbol{\Sigma}_{11}| = |\boldsymbol{\Sigma}_1|$ , onde  $\boldsymbol{\Sigma}_1 = E(\mathbf{c}_{1t} \mathbf{c}_{1t}')$  é obtida da regressão restrita

$$\mathbf{Y}_t = \boldsymbol{\nu}_1 + \sum_{i=1}^p \mathbf{A}_i \mathbf{Y}_{t-i} + \mathbf{c}_{1t}. \quad (7.59)$$

(ii)  $\mathbf{Y}_t$  não causa  $\mathbf{Z}_t \leftrightarrow |\Sigma_{22}| = |\Sigma_2|$ , onde  $\Sigma_2 = E(\mathbf{c}_{2t}\mathbf{c}_{2t}')$  é obtida da regressão restrita

$$\mathbf{Z}_t = \nu_2 + \sum_{i=1}^p \mathbf{C}_i \mathbf{Z}_{t-i} + \mathbf{c}_{2t}. \quad (7.60)$$

As regressões (7.57)-(7.60) podem ser estimadas por MQO e a partir dos resíduos de MQ as matrizes de covariâncias envolvidas são estimadas por:

$$\hat{\Sigma}_i = (T-p)^{-1} \sum_{t=p+1}^T \hat{\mathbf{c}}_{it} \hat{\mathbf{c}}_{it}',$$

$$\hat{\Sigma}_{ii} = (T-p)^{-1} \sum_{t=p+1}^T \hat{\mathbf{a}}_{it} \hat{\mathbf{a}}_{it}', \quad i = 1, 2.$$

Os testes e respectivas estatísticas da razão de verossimilhanças são dados por:

(i)  $H_{01} : \Phi_{12,i} = \mathbf{0}$ , para todo  $i$  ( $\mathbf{Z}_t$  não causa  $\mathbf{Y}_t$ ),

$$RV_1 = (T-p)[\log |\hat{\Sigma}_1| - \log |\hat{\Sigma}_{11}|] \sim \chi^2(prs).$$

(ii)  $H_{02} : \Phi_{21,i} = \mathbf{0}$ , para todo  $i$  ( $\mathbf{Y}_t$  não causa  $\mathbf{Z}_t$ ),

$$RV_2 = (T-p)[\log |\hat{\Sigma}_2| - \log |\hat{\Sigma}_{22}|] \sim \chi^2(prs).$$

Testes e estatísticas semelhantes podem ser estabelecidas para causalidade instantânea. Veja o Apêndice 7.D.

**Exemplo 7.10.** Para o exemplo 7.8, vemos que  $X_{1t} \rightarrow X_{2t}$ , ou seja, retornos diários do Ibovespa causam, no sentido de Granger, retornos diários da

Petrobrás. Também,  $X_{2t} \rightarrow X_{1t}$ . Logo, há “feedback” entre as duas séries de retornos.

**Exemplo 7.11.** Um modelo VAR(1) para as séries de retornos diários do Banespa ( $X_{1t}$ ) e da Petrobrás ( $X_{2t}$ ) é dado por

$$\begin{aligned} X_{1t} &= 0,122X_{1,t-1} + a_{1t}, \\ X_{2t} &= 0,333X_{1,t-1} + 0,081X_{2,t-1} + a_{2t}, \end{aligned}$$

sendo  $\begin{bmatrix} 1 & 0,007 \\ 0,007 & 1 \end{bmatrix}$  a matriz de correlação estimada. Vemos que Banespa causa Petrobrás, mas não o contrário. Procedendo como no Apêndice 7.C e escrevendo o modelo na forma estrutural, obtemos

$$X_{2t} = 0,005X_{1t} + 0,332X_{1,t-1} + 0,081X_{2,t-1} + b_{2t}, \quad (7.61)$$

o que mostra que provavelmente não há causalidade instantânea entre as séries, dada a magnitude do coeficiente de  $X_{1t}$ . Mas um teste formal teria que ser feito, utilizando os resultados do Apêndice 7.C.

## 7.8 Problemas

1. Para o modelo VMA(1) do exemplo 7.9, verifique o que acontece se  $\theta_{12} = 0$  e se  $\theta_{21} = 0$ , separadamente; depois, se ambos são nulos simultaneamente; finalmente, se ambos são diferentes de zero.
2. Mostre que as matrizes  $\Psi_j$  e  $\Pi_j$  das representações (7.49) e (7.50) podem ser obtidas recursivamente por:

$$\begin{aligned} \Psi_0 &= \mathbf{I}_n = \Pi_0 \\ \Psi_j &= \sum_{i=1}^j \Phi_i \Psi_{j-i} + \Theta_j, \quad j \geq 1, \\ \Pi_j &= -\sum_{i=1}^j \theta_i \Pi_{j-i} - \Phi_j, \quad j \geq 1, \end{aligned}$$

onde  $\Theta_j = \mathbf{0}, j > q$  e  $\Phi_i = \mathbf{0}, i > p$ .

3. Prove que a matriz de covariâncias de um modelo VARMA é dada por  $\Gamma(\tau) = \sum_{k=0}^{\infty} \Psi_{\tau+k} \Sigma \Psi_k'$ ,  $\tau \in \mathbb{Z}$ .
4. Da representação (B.1) do Apêndice 7.B mostre que  $E(\mathbf{X}_t) = \boldsymbol{\mu}$  e  $\Gamma(\tau) = \sum_{j=0}^{\infty} \Phi^{\tau+j} \Sigma (\Phi^j)'$  e, em particular, obtenha  $\Gamma(0)$ .
5. Obtenha o modelo (7.61).
6. Ajuste um modelo VAR às séries de retornos diários do Ibovespa e da Cemig.
7. Use (C.2) do Apêndice 7.C para obter as matrizes de covariâncias de lags 0, 1 e 2 para o exemplo de modelo VAR(1) dado no Apêndice 7.D.
8. Estabeleça relações de causalidade para as séries dos exemplos 7.8 e 7.9.
9. Para o exemplo 7.8, encontre as equações estruturais. Veja o Apêndice 7.D.
10. Para o exemplo 7.9, encontre as equações estruturais para  $X_{1t}$  e  $X_{2t}$ . Veja o Apêndice 7.D.
11. Prove que a condição (i) da Proposição 7.3 é equivalente a dizer que os auto-valores da matriz

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

têm módulos menores do que um. Veja também o Apêndice 7.C.

12. Como ficaria o problema anterior para um modelo VAR(2)?
13. Prove (7.30) e (7.31).
14. Prove (7.38).
15. Verifique se há causalidade instantânea para os exemplos 7.10 e 7.11.



## Apêndice 7.A. Alguns Resultados sobre Matrizes

### A.1 Conceitos Básicos

A notação  $\mathbf{A} = [a_{ij}]$  indicará uma matriz de ordem  $m \times n$ . A transposta de  $\mathbf{A}$  será indicada por  $\mathbf{A}'$ , a soma de  $\mathbf{A}$  e  $\mathbf{B}$  por  $\mathbf{A} + \mathbf{B}$ , se ambas têm a mesma ordem. Se  $\mathbf{A}$  é de ordem  $m \times n$  e  $\mathbf{B}$  é de ordem  $n \times r$ , então o produto  $\mathbf{AB}$  é a matriz  $\mathbf{C} = [c_{ij}]$ , de ordem  $m \times r$ , cujos elementos  $c_{ij}$  são dados por  $\sum_{k=1}^n a_{ik}b_{kj}$ . Não é verdade que  $\mathbf{AB} = \mathbf{BA}$ , em geral, mas  $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C} = \mathbf{ABC}$ , desde que os produtos estejam definidos. A matriz identidade de ordem  $n$  será indicada por  $\mathbf{I}_n$  e  $\mathbf{0}$  indicará a matriz nula.

Dizemos que  $\mathbf{A}$  é *ortogonal* se  $m = n$  e  $\mathbf{A}'\mathbf{A} = \mathbf{I}_n$ . Uma matriz  $\mathbf{A}$  é *simétrica* se  $m = n$  e  $\mathbf{A}' = \mathbf{A}$ . Dizemos que  $\mathbf{A}$  é *não-negativa definida* se for simétrica e se

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i,j=1}^m a_{ij}x_ix_j \geq 0,$$

para todo vetor  $\mathbf{x}$  de ordem  $m \times 1$ . Denotaremos tal matriz por  $\mathbf{A} \geq \mathbf{0}$ . Dizemos que  $\mathbf{A}$  é *positiva definida* se  $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ , para todo vetor  $\mathbf{x} \neq \mathbf{0}$  e escrevemos  $\mathbf{A} > \mathbf{0}$ . A expressão  $\mathbf{x}'\mathbf{A}\mathbf{x}$  diz-se uma *forma quadrática* nas variáveis  $x_1, \dots, x_m$  e as nomenclaturas acima aplicam-se, também, às respectivas formas quadráticas.

### A.2 Determinante, Traço e Posto.

Características importantes de matrizes são o determinante, traço e posto. Se  $\mathbf{A}$  é quadrada de ordem  $m$ , o *traço* de  $\mathbf{A}$  é definido por  $\text{tr}(\mathbf{A}) = \sum_{i=1}^m a_{ii}$ .

É fácil ver que  $\text{tr}(\mathbf{A}') = \text{tr}(\mathbf{A})$ ,  $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$  e  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ , se  $\mathbf{A}$  for  $m \times n$  e  $\mathbf{B}$  for  $n \times m$ .

Se  $\mathbf{A}$  é quadrada de ordem  $m$ , real, o *determinante* de  $\mathbf{A}$ , denotado  $|\mathbf{A}|$ , é a única função real de elementos de  $\mathbf{A}$  tal que

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|,$$

para toda matriz  $\mathbf{B}$  de ordem  $m$  e  $|\mathbf{I}| = 1$ , se

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

para todo  $\gamma$ .

O posto da matriz  $\mathbf{A}$ , denotado  $\rho(\mathbf{A})$ , é o número de linhas (ou colunas) linearmente independentes de  $\mathbf{A}$ ; ou é a ordem da maior submatriz de  $\mathbf{A}$  com determinante não nulo.

Uma matriz quadrada  $\mathbf{A}$ , de ordem  $m$ , diz-se *não-singular* se  $\rho(\mathbf{A}) = m$ , isto é, se  $|\mathbf{A}| \neq 0$ . Neste caso, existe uma única matriz  $\mathbf{A}^{-1}$ , de ordem  $m$ , chamada a *inversa* de  $\mathbf{A}$ , tal que

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_m.$$

As seguintes propriedades são válidas:

- (1) Se  $\mathbf{A}$  for ortogonal,  $\mathbf{A}^{-1} = \mathbf{A}'$ ;
- (2)  $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$ ;
- (3)  $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$ ;
- (4)  $|\alpha\mathbf{A}| = \alpha^m|\mathbf{A}|$ ;
- (5)  $|\mathbf{A}^{-1}| = (|\mathbf{A}|)^{-1}$  se  $\mathbf{A}$  não-singular;
- (6)  $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ ;
- (7)  $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$ ;
- (8)  $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ , com ordens apropriadas;
- (9)  $\rho(\mathbf{AB}) = \rho(\mathbf{A})$ , se  $\mathbf{B}$  for não-singular;
- (10)  $\rho(\mathbf{A}) = \rho(\mathbf{A}'\mathbf{A})$ ;
- (11)  $\rho(\mathbf{AB}) \leq \min\{\rho(\mathbf{A}), \rho(\mathbf{B})\}$ .

### A.3 Matrizes Particionadas

Dizemos que a matriz  $\mathbf{A}$  é *particionada* (ou em blocos) se seus elementos também são matrizes:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1n} \\ \mathbf{A}_{21} & \cdots & \mathbf{A}_{2n} \\ \cdots & \cdots & \cdots \\ \mathbf{A}_{m1} & \cdots & \mathbf{A}_{mn} \end{bmatrix},$$

onde  $\mathbf{A}_{j1}, \dots, \mathbf{A}_{jn}$  são matrizes com o mesmo número de linhas,  $j = 1, \dots, m$  e  $\mathbf{A}_{1k}, \dots, \mathbf{A}_{mk}$  são matrizes com o mesmo número de colunas,  $k = 1, \dots, n$ .

Por exemplo, se  $\mathbf{A}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\mathbf{A}_{12} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ,  $\mathbf{A}_{21} = \begin{bmatrix} 1 & 1 \end{bmatrix}$ ,  $\mathbf{A}_{22} = [0]$ , então

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

é um exemplo. Matrizes em blocos podem ser somadas, multiplicadas, etc, se as dimensões dos elementos (matrizes) são apropriadas.

### A.4 Produto de Kronecker e Vetorização

Seja  $\mathbf{A}$  uma matriz  $m \times n$  e  $\mathbf{B}$  uma matriz de ordem  $r \times s$ . Então o *produto de Kronecker* (ou produto tensorial)  $\mathbf{A} \otimes \mathbf{B}$  é a matriz em blocos de ordem  $(mr) \times (ns)$  dada por

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{A}b_{11} & \cdots & \mathbf{A}b_{1s} \\ \mathbf{A}b_{21} & \cdots & \mathbf{A}b_{2s} \\ \cdots & \cdots & \cdots \\ \mathbf{A}b_{r1} & \cdots & \mathbf{A}b_{rs} \end{bmatrix}.$$

Por exemplo,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \otimes \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} a\alpha & b\alpha & a\beta & b\beta \\ c\alpha & d\alpha & c\beta & d\beta \\ a\gamma & b\gamma & a\delta & b\delta \\ c\gamma & d\gamma & c\delta & d\delta \end{bmatrix}.$$

As seguintes propriedades do produto de Kronecker são importantes:

- (1)  $(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C})$ ;
- (2)  $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$ ;
- (3)  $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$ ;
- (4) Se  $\mathbf{A}$  e  $\mathbf{B}$  são matrizes quadradas, então  $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = [\text{tr}(\mathbf{A})][\text{tr}(\mathbf{B})]$ ;
- (5)  $\rho(\mathbf{A} \otimes \mathbf{B}) = \rho(\mathbf{A})\rho(\mathbf{B})$ ;
- (6) Se  $\mathbf{A}$  e  $\mathbf{B}$  são inversíveis,  $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ ;
- (7) Se  $\mathbf{A}$  é quadrada de ordem  $m$  e  $\mathbf{B}$  é quadrada de ordem  $n$ , então  $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n |\mathbf{B}|^m$ ;
- (8) Se  $\mathbf{A}$  e  $\mathbf{B}$  são matrizes quadradas, com autovalores  $\lambda_A, \lambda_B$ , respectivamente, e correspondentes autovetores  $\mathbf{v}_A$  e  $\mathbf{v}_B$ , então  $\lambda_A \lambda_B$  é um autovalor de  $\mathbf{A} \otimes \mathbf{B}$  com autovetor  $\mathbf{v}_A \otimes \mathbf{v}_B$ .

Um procedimento muito útil em determinadas situações é o de *vetorização*. Seja  $\mathbf{A}$  uma matriz de ordem  $m \times n$ . Então  $\text{vec}(\mathbf{A})$  denotará o vetor de ordem  $(mn) \times 1$  tal que

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_j \\ \vdots \\ \mathbf{a}_n \end{bmatrix},$$

onde  $\mathbf{a}_j$  denota a  $j$ -ésima coluna de  $\mathbf{A}$ .

Por exemplo,

$$\text{vec} \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} a \\ c \\ b \\ d \end{bmatrix}.$$

Os seguintes resultados são importantes.

- (1)  $\text{vec}(\mathbf{A} + \mathbf{B}) = \text{vec}(\mathbf{A}) + \text{vec}(\mathbf{B})$ ;
- (2)  $\text{vec}(\mathbf{AB}) = (\mathbf{I} \otimes \mathbf{A}) \text{vec}(\mathbf{B}) = (\mathbf{B}' \otimes \mathbf{I}) \text{vec}(\mathbf{A})$ ;
- (3)  $\text{vec}(\mathbf{AXB}') = (\mathbf{A} \otimes \mathbf{B}) \text{vec}(\mathbf{X})$ .
- (4)  $\text{tr}(\mathbf{ABC}) = \text{vec}(\mathbf{A}')'(\mathbf{C}' \otimes \mathbf{I}) \text{vec}(\mathbf{B})$ .

### A.5 Decomposição de Matrizes

Seja  $\mathbf{A}$  uma matriz quadrada de ordem  $m$ . Segue-se que  $|\mathbf{A} - \lambda \mathbf{I}|$  é um polinômio de ordem  $m$  em  $\lambda$  e terá  $m$  raízes complexas  $\lambda_1, \dots, \lambda_m$ . Essas raízes são chamadas *raízes características* ou *autovalores* de  $\mathbf{A}$ .

Como  $\mathbf{A} - \lambda_j \mathbf{I}$  é singular,  $j = 1, \dots, m$ , existe um vetor  $\mathbf{a}_j$ , cujas coordenadas não são todas nulas, tal que  $(\mathbf{A} - \lambda_j \mathbf{I})\mathbf{a}_j = \mathbf{0}$ , ou seja,  $\mathbf{A}\mathbf{a}_j = \lambda_j \mathbf{a}_j$ ,  $j = 1, \dots, m$ . Os vetores  $\mathbf{a}_1, \dots, \mathbf{a}_m$  são chamados *vetores característicos* ou *autovetores* de  $\mathbf{A}$ .

Os seguintes resultados são válidos.

- (1)  $\rho(\mathbf{A})$  dá o número de autovalores de  $\mathbf{A}$  não nulos.
- (2)  $\text{tr}(\mathbf{A}) = \sum_{j=1}^m \lambda_j$ .
- (3)  $|\mathbf{A}| = \prod_{j=1}^m \lambda_j$ .
- (4) Se  $\mathbf{A}$  é uma matriz simétrica, real, todos os seus autovalores são reais, e para cada autovalor real existe um autovetor real.
- (5) Se  $\mathbf{A}$  é simétrica, real, os autovetores correspondentes a autovalores distintos são ortogonais.

- (6) Se  $\mathbf{A}$  é não negativa definida, então  $\lambda_j \geq 0$ ,  $j = 1, \dots, m$ .
- (7) Se  $\mathbf{A}$  é simétrica, de ordem  $m \times m$ , existe uma matriz ortogonal  $\mathbf{X}$ , tal que

$$\mathbf{X}'\mathbf{A}\mathbf{X} = \mathbf{\Lambda} = \text{diag}\{\lambda_1, \dots, \lambda_m\},$$

ou  $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}'$ , onde os  $\lambda_j$  são os autovalores de  $\mathbf{A}$  e as colunas de  $\mathbf{X}$  são os correspondentes autovetores.

O resultado (7) é chamado o *teorema espectral* para matrizes simétricas. Segue-se que a *decomposição espectral* de  $\mathbf{A}$  é dada por

$$\mathbf{A} = \sum_{j=1}^m \lambda_j \mathbf{x}_j \mathbf{x}_j',$$

onde  $\mathbf{x}_j$  é o autovetor correspondente a  $\lambda_j$ .

Se  $\mathbf{A}$  é uma matriz quadrada de ordem  $m$ , positiva definida, existe uma matriz triangular inferior  $\mathbf{T}$ , com elementos da diagonal principal positivos, tal que

$$\mathbf{T}^{-1}\mathbf{A}(\mathbf{T}')^{-1} = \mathbf{I}_m, \quad \text{ou} \quad \mathbf{A} = \mathbf{T}\mathbf{T}'.$$

Esta decomposição de  $\mathbf{A}$  é chamada *decomposição de Choleski*. Veja o Apêndice 7.D para exemplos.

## Apêndice 7.B. Demonstração da Proposição 7.2

De (7.18) obtemos, recursivamente,

$$\begin{aligned} \mathbf{X}_t &= (\mathbf{I}_n + \mathbf{\Phi} + \dots + \mathbf{\Phi}^j)\mathbf{\Phi}_0 + \\ &+ \mathbf{\Phi}^{j+1}\mathbf{X}_{t-j-1} + \sum_{i=0}^j \mathbf{\Phi}^i \mathbf{a}_{t-i}. \end{aligned}$$

Se todos os autovalores de  $\mathbf{\Phi}$  tiverem módulos menores do que um, a sequência  $\mathbf{\Phi}^i$ ,  $i \geq 0$ , é absolutamente somável e a soma  $\sum_{i=0}^{\infty} \mathbf{\Phi}^i \mathbf{a}_{t-i}$  existe em

média quadrática. Também  $(\mathbf{I}_n + \Phi + \dots + \Phi^j)\Phi_0 \rightarrow (\mathbf{I}_n - \Phi)^{-1}\Phi_0$  e  $\Phi^{j+1} \rightarrow 0$ , quando  $j \rightarrow \infty$ .

Portanto, a solução estacionária para  $\mathbf{X}_t$  é dada por

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \Phi^i \mathbf{a}_{t-i}, \quad t \in Z, \quad (B.1)$$

onde  $\boldsymbol{\mu} = (\mathbf{I}_n - \Phi)^{-1}\phi_0$ .

## Apêndice 7.C. Modelo VAR( $p$ ) na Forma VAR(1)

Suponha que  $\mathbf{X}_t$  seja dado por (7.16). Defina o seguinte processo VAR(1):

$$\mathbf{Y}_t = \mathbf{C} + \mathbf{F}\mathbf{Y}_{t-1} + \mathbf{b}_t, \quad (C.1)$$

onde

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{X}_t \\ \mathbf{X}_{t-1} \\ \vdots \\ \mathbf{X}_{t-p+1} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \Phi_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{b}_t = \begin{bmatrix} \mathbf{a}_t \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

são vetores de ordem  $np \times 1$  e

$$\mathbf{F} = \begin{bmatrix} \Phi_1 & \Phi_2 & \cdots & \Phi_{p-1} & \Phi_p \\ \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_n & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_n & \mathbf{0} \end{bmatrix}$$

é uma matriz  $np \times np$ .

Pela discussão sobre modelos VAR(1), o processo  $\mathbf{Y}_t$  é estacionário se

$$|\mathbf{I}_{np} - \mathbf{F}z| \neq 0, \quad |z| \leq 1.$$

É fácil ver que  $|\mathbf{I}_{np} - \mathbf{F}z| = |\mathbf{I}_n - \Phi_1 z - \dots - \Phi_p z^p|$ , logo o processo VAR( $p$ ) é estacionário se (i) da Proposição 7.3 for válida.

Segue-se de (C.1) que a matriz de covariâncias de  $\mathbf{Y}_t$  é dada por

$$\Gamma_Y(\tau) = \sum_{i=0}^{\infty} \mathbf{F}^{\tau+i} \Sigma_b (\mathbf{F}^i)',$$

onde  $\Sigma_b = E(\mathbf{b}_t \mathbf{b}_t')$ . Novamente, sendo esta uma soma infinita, é mais conveniente usar (iii) da Proposição 7.3. De (C.1) temos

$$\mathbf{Y}_t - \boldsymbol{\nu} = \mathbf{F}(\mathbf{Y}_{t-1} - \boldsymbol{\nu}) + \mathbf{b}_t,$$

onde  $\boldsymbol{\nu} = (\boldsymbol{\mu}', \dots, \boldsymbol{\mu}')' = E(\mathbf{Y}_t)$ . Segue-se do caso VAR(1) que

$$\Gamma_Y(0) = \mathbf{F} \Gamma_Y(0) \mathbf{F}' + \Sigma_b$$

e as matrizes  $\Gamma_Y(\tau)$ ,  $\tau = -p+1, \dots, p-1$  podem ser obtidas de

$$\text{vec}(\Gamma_Y(0)) = (\mathbf{I}_{(np)^2} - \mathbf{F} \otimes \mathbf{F})^{-1} \text{vec}(\Sigma_b). \quad (C.2)$$

Note que

$$\begin{aligned} \Gamma_Y(0) &= E \begin{bmatrix} \mathbf{X}_t - \boldsymbol{\nu} \\ \vdots \\ \mathbf{X}_{t-p+1} - \boldsymbol{\nu} \end{bmatrix} \begin{bmatrix} (\mathbf{X}_t - \boldsymbol{\nu})', \dots, (\mathbf{X}_{t-p+1} - \boldsymbol{\nu})' \end{bmatrix} \\ &= \begin{bmatrix} \Gamma(0) & \Gamma(1) & \dots & \Gamma(p-1) \\ \vdots & \vdots & & \vdots \\ \Gamma(-p+1) & \Gamma(-p+2) & \dots & \Gamma(0) \end{bmatrix}. \end{aligned}$$

## Apêndice 7.D. Modelos Estruturais

Como vimos, o modelo VAR( $p$ ) (7.16) está na forma reduzida, porque não explicita uma relação linear contemporânea entre as séries  $X_{it}, i = 1, \dots, n$ . Esta dependência explícita pode ser obtida por meio de uma transformação linear, especificamente, usando uma decomposição de Choleski da matriz  $\Sigma$ . Dado que esta é positiva definida, existe uma matriz triangular inferior  $\mathbf{L}$ , com elementos diagonais unitários e uma matriz diagonal  $\mathbf{\Lambda}$  tal que  $\Sigma = \mathbf{L} \mathbf{\Lambda} \mathbf{L}'$ . Segue-se que  $\mathbf{L}^{-1} \Sigma (\mathbf{L}')^{-1} = \mathbf{\Lambda}$ . Se indicarmos por  $\mathbf{b}_t = \mathbf{L}^{-1} \mathbf{a}_t$ , então temos



que  $E(\mathbf{b}_t) = \mathbf{0}$  e a matriz de covariâncias de  $\mathbf{b}_t$  é dada por  $\mathbf{\Lambda}$ , ou seja, os elementos  $b_{it}$  de  $\mathbf{b}_t$  são não-correlacionados.

Pré-multiplicando a equação (7.16) por  $\mathbf{L}^{-1}$  obtemos

$$\mathbf{L}^{-1}\mathbf{X}_t = \mathbf{L}^{-1}\mathbf{\Phi}_0 + \mathbf{L}^{-1}\mathbf{\Phi}_1\mathbf{X}_{t-1} + \dots + \mathbf{L}^{-1}\mathbf{\Phi}_p\mathbf{X}_{t-p} + \mathbf{b}_t. \quad (D.1)$$

É fácil ver que a última linha de  $\mathbf{L}^{-1}$  é da forma  $(a_{n1}, a_{n2}, \dots, a_{n,n-1}, 1)$  de modo que a última equação do modelo (D.1) é

$$X_{nt} + \sum_{i=1}^{n-1} a_{ni} X_{it} = \phi_{n,0}^* + \sum_{i=1}^n \mathbf{\Phi}_{ni}^{(1)*} X_{i,t-1} + \dots + \sum_{i=1}^n \mathbf{\Phi}_{ni}^{(p)*} X_{i,t-p} + b_{nt}. \quad (D.2)$$

Como  $b_{nt}$  é não-correlacionado com  $b_{it}$ ,  $1 \leq i < n$ , a equação (D.2) mostra a relação linear contemporânea de  $X_{nt}$  com  $X_{it}$ ,  $1 \leq i < n$ . Esta equação é chamada *equação estrutural* para  $X_{nt}$ . Para qualquer outra componente  $X_{jt}$  podemos rearranjar o modelo VAR( $p$ ) de modo que  $X_{jt}$  apareça como a última componente de  $\mathbf{X}_t$ .

Considere, por exemplo, o modelo VAR(1) dado por

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 0,5 & 0,1 \\ 0,4 & 0,5 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix},$$

com

$$\mathbf{\Sigma} = \begin{bmatrix} 2,25 & 0,3 \\ 0,3 & 0,64 \end{bmatrix}.$$

É fácil ver que a matriz  $\mathbf{L}$ , triangular inferior, necessária para diagonalizar  $\mathbf{\Sigma}$ , e a matriz  $\mathbf{\Lambda}$  são, respectivamente,

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0,13 & 1 \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} 2,25 & 0 \\ 0 & 0,60 \end{bmatrix}.$$

Pré-multiplicando o modelo por  $\mathbf{L}^{-1}$  obtemos

$$X_{2t} = 0,13X_{1,t} + 0,34X_{1,t-1} + 0,49X_{2,t-1} + b_{2t},$$

mostrando que  $X_{2t}$  depende do valor presente e do passado de  $X_{1,t}$ , além de depender também de seu próprio passado.

Rearranjando o modelo na forma

$$\begin{bmatrix} X_{2t} \\ X_{1t} \end{bmatrix} = \begin{bmatrix} 0,4 & 0,5 \\ 0,5 & 0,1 \end{bmatrix} \begin{bmatrix} X_{2,t-1} \\ X_{1,t-1} \end{bmatrix} + \begin{bmatrix} a_{2t} \\ a_{1t} \end{bmatrix},$$

com

$$\Sigma = \begin{bmatrix} 0,64 & 0,3 \\ 0,3 & 2,25 \end{bmatrix},$$

obtemos que  $\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0,47 & 1 \end{bmatrix}$  e  $\mathbf{\Lambda} = \begin{bmatrix} 0,64 & 0 \\ 2,11 & 1 \end{bmatrix}$ , de modo que

$$X_{1t} = 0,47X_{2t} + 0,5X_{2,t-1} + 0,1X_{1,t-1} + c_{2,t},$$

mostrando também a dependência contemporânea de  $X_{1t}$  sobre  $X_{2t}$ .

Se pré-multiplicarmos o sistema (7.57)-(7.58) por  $\begin{bmatrix} \mathbf{I}_r & -\Sigma_{12}\Sigma_{22}^{-1} \\ -\Sigma'_{12}\Sigma_{11}^{-1} & \mathbf{I}_s \end{bmatrix}$  obtemos

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu}_1 + \sum_{i=1}^p \Phi_{11,i}^* \mathbf{Y}_{t-i} + \sum_{i=0}^p \Phi_{12,i}^* \mathbf{Z}_{t-i} + \mathbf{a}_{1t}^*, \\ \mathbf{Z}_t &= \boldsymbol{\mu}_2 + \sum_{i=1}^p \Phi_{21,i}^* \mathbf{Y}_{t-i} + \sum_{i=0}^p \Phi_{22,i}^* \mathbf{Z}_{t-i} + \mathbf{a}_{2t}^*, \end{aligned}$$

onde  $\mathbf{a}_{1t}^* = \mathbf{a}_{1t} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{a}_{2t}$  e  $\mathbf{a}_{2t}^* = \mathbf{a}_{2t} - \Sigma'_{12}\Sigma_{11}^{-1}\mathbf{a}_{1t}$ .

Chamando  $\Sigma_i^* = E[\mathbf{a}_{it}^*(\mathbf{a}_{it}^*)']$ , temos a seguinte proposição.

**Proposição D.1.** Há causalidade instantânea entre  $\mathbf{Y}_t$  e  $\mathbf{Z}_t$  se e somente se  $\Phi_{12,0}^* \neq \mathbf{0}$  e  $\Phi_{21,0}^* \neq \mathbf{0}$ , ou ainda, se e somente se  $|\Sigma_{11}| > |\Sigma_1^*|$  e  $|\Sigma_{22}| > |\Sigma_2^*|$ .

O testes de não-causalidade instantânea pode ser conduzidos como nos casos vistos na seção 7.7 para não causalidade (hipóteses  $H_{01}$  e  $H_{02}$ ), estimando-se as matrizes  $\Sigma_i^*$  com os resíduos de MQ das regressões acima.

# Processos Co-Integrados

## 8.1 Introdução

Vimos que um processo  $X_t$  é integrado de ordem  $d$  se  $\Delta^d X_t$  for estacionário, e escrevemos  $X_t \sim I(d)$ . Em particular, um processo estacionário é  $I(0)$ . Estudamos, em particular, a classe dos processos  $ARIMA(p, d, q)$ . Para estes, após tomarmos  $d$  diferenças, o processo estacionário resultante é representado por um modelo  $ARMA(p, q)$ .

No Capítulo 7 tratamos de modelos VAR estacionários, ou seja, as séries envolvidas são  $I(0)$ . A teoria usual de MQO também aplica-se a séries  $I(0)$ . Se algumas ou todas as séries de um modelo de regressão são  $I(1)$ , os resultados estatísticos usuais em geral não são mais válidos. Este é o problema da *regressão espúria*, tratado por Granger e Newbold (1974). Estes autores verificaram, através de simulações, que dadas duas séries completamente não-correlacionadas, mas  $I(1)$ , a regressão de uma sobre a outra tenderá a produzir uma relação aparentemente significativa. Veja também Phillips (1986). Há, portanto, a necessidade de se desenvolver técnicas para analisar relações entre séries não-estacionárias.

Neste capítulo estaremos interessados em analisar modelos para descrever co-movimentos dinâmicos de duas ou mais séries temporais, como séries de preços de ativos ou taxas de câmbio. É comum que preços de ativos apresentem uma tendência estocástica comum no longo prazo, ou seja, que sejam *co-integrados*.

Preços e taxas (de câmbio, de juros) em geral são  $I(1)$  e é usual analisar

os logaritmos destas séries, para investigar co-integração. Estabelecida uma relação de equilíbrio de longo prazo entre log-preços, por exemplo, ajusta-se um modelo que corrige desvios de curto prazo da relação de equilíbrio. Este modelo é chamado *modelo de correção de erros* (MCE).

Se  $X_t$  e  $Y_t$  forem processos  $I(d)$ , então a combinação linear  $Z_t = Y_t - \alpha X_t$  será, em geral, também  $I(d)$ . Mas é possível que  $Z_t$  seja integrado de ordem menor, digamos  $I(d-b)$ ,  $b > 0$ . Se  $d = b = 1$ , então  $X_t$  e  $Y_t$  serão  $I(1)$  e  $Z_t$  será  $I(0)$ . Neste caso, dizemos que  $X_t$  e  $Y_t$  são co-integrados. Todavia, não é geralmente verdade que exista  $\alpha$  tal que  $Z_t \sim I(0)$  ou, em geral,  $Z_t \sim I(d-b)$ .

No caso de um vetor  $\mathbf{X}_t$ , de ordem  $n \times 1$ , dizemos que ele é integrado de ordem  $d$ ,  $I(d)$ , se  $d$  for a maior ordem de integração das séries individuais. Ou seja, se  $\mathbf{X}_t = (X_{1t}, \dots, X_{nt})'$ ,  $X_{it} \sim I(d_i)$ , então  $d = \max(d_1, \dots, d_n)$ .

**Definição 8.1.** As componentes do vetor  $\mathbf{X}_t$  são *co-integradas de ordem*  $(d, b)$ , e escrevemos,  $\mathbf{X}_t \sim \text{C.I.}(d, b)$ , se:

- (a) todas as componentes de  $\mathbf{X}_t$  são  $I(d)$ ;
- (b) existe um vetor  $\beta = (\beta_1, \dots, \beta_n)'$ , não-nulo, tal que

$$\beta' \mathbf{X}_t = \beta_1 X_{1t} + \dots + \beta_n X_{nt} \sim I(d-b), \quad d \geq b > 0. \quad (8.1)$$

O vetor  $\beta$ , de ordem  $n \times 1$ , é chamado *vetor co-integrado* (ou vetor de co-integração).

**Exemplo 8.1** (Engle e Granger, 1987). Considere  $n = 2$  e as séries  $X_{1t}$  e  $X_{2t}$ , dadas por

$$X_{1t} + \beta X_{2t} = u_t, \quad (8.2)$$

$$u_t = \phi_1 u_{t-1} + \varepsilon_{1t}, \quad (8.3)$$

$$X_{1t} + \alpha X_{2t} = v_t, \quad (8.4)$$

$$v_t = \phi_2 v_{t-1} + \varepsilon_{2t}, \quad (8.5)$$

onde supomos os  $\varepsilon_{it}$  independentes e normais, com média zero e com  $E(\varepsilon_{it}\varepsilon_{js}) = 0, i, j = 1, 2$ . Suponha  $\phi_i \neq 0, i = 1, 2$ . Então temos os seguintes casos a analisar:

(i)  $\phi_i < 1, i = 1, 2$ .

Neste caso,  $X_{1t}$  e  $X_{2t}$  serão  $I(0)$ , mas os parâmetros  $\alpha$  e  $\beta$  não são identificados.

(ii)  $\phi_1 = 1, \phi_2 < 1$ .

As séries são ambas  $I(1)$  e  $(1, \alpha)'$  é o vetor co-integrado. A equação (8.4) é identificada.

(iii)  $\phi_1 < 1, \phi_2 = 1$ .

Similar ao anterior, o vetor co-integrado é  $(1, \beta)'$  e a equação (8.2) é identificada.

**Observações:** (i) O vetor de co-integração  $\beta$  não é único, pois se  $\lambda \neq 0$ , então  $\lambda\beta$  é também um vetor de co-integração. Tipicamente, uma das variáveis é usada para normalizar  $\beta$ , fixando-se seu coeficiente igual a um; usualmente toma-se  $\beta = (1, -\beta_2, \dots, -\beta_n)'$ , de modo que

$$\beta' \mathbf{X}_t = X_{1t} - \beta_2 X_{2t} - \dots - \beta_n X_{nt}.$$

Por exemplo, se  $\beta' \mathbf{X}_t \sim I(0)$ , temos que

$$X_{1t} = \beta_2 X_{2t} + \dots + \beta_n X_{nt} + u_t,$$

com  $u_t \sim I(0)$ . Dizemos que  $u_t$  é o *resíduo de co-integração*. Em equilíbrio de longo prazo,  $u_t = 0$  e a relação de equilíbrio de longo prazo é

$$X_{1t} = \beta_2 X_{2t} + \dots + \beta_n X_{nt}.$$

(ii) Todas as variáveis devem ser integradas de *mesma ordem*. Se elas forem integradas de ordens diferentes, não podem ser co-integradas. Veja o Problema 2.

(iii) Se  $\mathbf{X}_t$  tiver  $n > 2$  componentes, podem existir vários vetores de co-integração. Se existirem exatamente  $r$  vetores de co-integração linearmente

independentes, com  $0 < r \leq n - 1$ , então eles podem ser reunidos numa matriz  $\mathbf{B}$ , de ordem  $n \times r$ , com posto  $r$ , chamado o *posto de co-integração*. Neste caso,

$$\mathbf{B}'\mathbf{X}_t = \begin{bmatrix} \beta_1'\mathbf{X}_t \\ \vdots \\ \beta_r'\mathbf{X}_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{bmatrix}$$

é estacionária, isto é,  $I(0)$ . Por exemplo, se  $n = 3$  e  $r = 2$ , com  $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})'$  e  $\beta_2 = (\beta_{21}, \beta_{22}, \beta_{23})'$ , então

$$\mathbf{B}'\mathbf{X}_t = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \beta_1'\mathbf{X}_t \\ \beta_2'\mathbf{X}_t \end{bmatrix},$$

de modo que obtemos  $\beta_1'\mathbf{X}_t \sim I(0)$  e  $\beta_2'\mathbf{X}_t \sim I(0)$ . Note que se  $\beta_3 = c_1\beta_1 + c_2\beta_2$ , então  $\beta_3$  é também um vetor co-integrado.

## 8.2 Tendências Comuns

Vimos no exemplo 2.7 que log-preços de ativos podem ser modelados por um passeio aleatório, ou seja, na notação do exemplo,

$$\Delta p_t = \mu + \varepsilon_t,$$

onde  $\varepsilon_t \sim \text{i.i.d.} \mathcal{N}(0, \sigma^2)$ . Logo, a melhor previsão de qualquer valor futuro é o valor de hoje mais um “drift”. Mas se existe uma relação de co-integração entre dois ou mais log-preços, um modelo multivariado pode dar informação sobre o equilíbrio de longo prazo entre as séries.

Preços co-integrados têm uma tendência estocástica comum, um fato apontado por Stock e Watson (1988). Ou seja, eles caminharão juntos no longo prazo porque uma combinação linear deles é reversível à média (estacionária).

**Exemplo 8.2.** Suponha que

$$X_{1t} = \mu_{1t} + \varepsilon_{1t}, \quad (8.6)$$

$$X_{2t} = \mu_{2t} + \varepsilon_{2t}, \quad (8.7)$$

onde  $\mu_{it}$  é um passeio aleatório representando a tendência estocástica da variável  $X_{it}$ ,  $i = 1, 2$  e  $\varepsilon_{it}$ ,  $i = 1, 2$  é estacionário. Suponha que  $X_{1t}$  e  $X_{2t}$  sejam  $I(1)$  e que existam constantes  $\beta_1$  e  $\beta_2$  tais que  $\beta_1 X_{1t} + \beta_2 X_{2t}$  seja  $I(0)$ , ou seja,

$$\beta_1 X_{1t} + \beta_2 X_{2t} = (\beta_1 \mu_{1t} + \beta_2 \mu_{2t}) + (\beta_1 \varepsilon_{1t} + \beta_2 \varepsilon_{2t})$$

seja estacionário. Então devemos ter o primeiro termo do segundo membro nulo, ou seja,  $\mu_{1t} = -(\beta_2/\beta_1)\mu_{2t}$ , o que mostra que as tendências são as mesmas, a menos de um escalar.

De modo geral, se o vetor  $\mathbf{X}_t$  for co-integrado, com  $r$  vetores de co-integração,  $0 < r < n$ , então existirão  $n - r$  tendências estocásticas comuns.

O fato de duas séries serem co-integradas não significa que elas apresentem alta correlação. O seguinte exemplo mostra esse fato (Alexander, 2001).

**Exemplo 8.3.** Suponha

$$X_{1t} = u_t + \varepsilon_{1t}, \quad (8.8)$$

$$X_{2t} = u_t + \varepsilon_{2t}, \quad (8.9)$$

$$u_t = u_{t-1} + \varepsilon_t, \quad (8.10)$$

onde os erros são i.i.d. e independentes entre si. Como  $u_t$  é passeio aleatório,  $X_{1t}$  e  $X_{2t}$  são  $I(1)$  e são co-integradas porque  $X_{1t} - X_{2t} = \varepsilon_{1t} - \varepsilon_{2t} \sim I(0)$ .

As séries também têm uma tendência estocástica comum, dada por  $u_t$ , que é passeio aleatório sem “drift”. Chamando  $\text{Var}(\varepsilon_t) = \sigma^2$ ,  $\text{Var}(\varepsilon_{1t}) = \sigma_1^2$ ,  $\text{Var}(\varepsilon_{2t}) = \sigma_2^2$  e notando que

$$\Delta X_{1t} = X_{1t} - X_{1,t-1} = \varepsilon_t + (\varepsilon_{1t} - \varepsilon_{1,t-1}),$$

temos que

$$\text{Var}(\Delta X_{1t}) = \sigma^2 + 2\sigma_1^2.$$

De modo análogo, obtemos  $\text{Var}(\Delta X_{2,t}) = \sigma^2 + 2\sigma_2^2$  e  $\text{Cov}(\Delta X_{1t}, \Delta X_{2t}) = \sigma^2$ , de modo que o coeficiente de correlação

$$\rho(\Delta X_{1t}, \Delta X_{2t}) = \frac{\sigma^2}{\sqrt{(\sigma^2 + 2\sigma_1^2)(\sigma^2 + 2\sigma_2^2)}}.$$

Esta correlação será pequena se  $\sigma^2 \ll \sigma_1^2$  ou  $\sigma^2 \ll \sigma_2^2$ .

**Exemplo 8.4.** Considere as séries

$$X_{1t} = \beta_2 X_{2t} + u_t, \quad (8.11)$$

$$X_{2t} = X_{2,t-1} + v_t, \quad (8.12)$$

onde  $u_t$  e  $v_t$  são ambas  $I(0)$ . Segue-se que  $X_{2t}$  é um passeio casual e representa a tendência estocástica comum, ao passo que (8.11) representa a relação de equilíbrio de longo prazo. O vetor de co-integração é  $\beta = (1, -\beta_2)'$ . Na Figura 8.1 temos as séries simuladas, com  $\beta_2 = 1$ ,  $u_t = 0,6u_{t-1} + a_t$ ,  $a_t$  e  $v_t$  independentes  $\mathcal{N}(0, 1)$ , independentes entre si. Veja os problemas 6 e 7 para outros exemplos de sistemas co-integrados. As equações (8.11) e (8.12) constituem a *representação triangular* de Phillips (1991).

### 8.3 Modelo de Correção de Erros

Nesta seção seguiremos de perto os desenvolvimentos em Lütkepohl (1991) e Hendry e Juselius (2000, 2001). Vimos, na seção 8.1, que se duas ou mais séries são co-integradas, existe uma relação de equilíbrio de longo prazo entre elas. Muitas variáveis econômicas apresentam relações de equilíbrio, como preços de um mesmo produto em diferentes mercados. Suponha, por exemplo, que  $P_{1t}$  e  $P_{2t}$  sejam tais preços em dois mercados distintos e que a relação (normalizada) de equilíbrio entre eles seja  $P_{1t} - \beta P_{2t} = 0$ . Suponha, ainda, que variações em  $P_{1t}$  dependam de desvios deste equilíbrio no instante  $t - 1$ , ou seja,



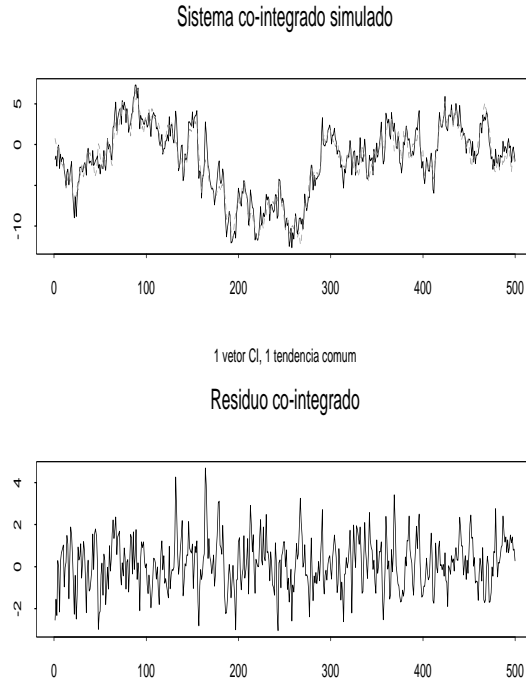


Figura 8.1: Sistema bivariado co-integrado:  $X_{1t}$  (linha cheia) e  $X_{2t}$  (linha tracejada)

$$\Delta P_{1t} = \alpha_1(P_{1,t-1} - \beta P_{2,t-1}) + a_{1t}, \quad (8.13)$$

e uma relação similar valha para  $P_{2t}$ :

$$\Delta P_{2t} = \alpha_2(P_{1,t-1} - \beta P_{2,t-1}) + a_{2t}. \quad (8.14)$$

Suponha que  $P_{1t}$  e  $P_{2t}$  sejam  $I(1)$ ; como  $\Delta P_{it} \sim I(0)$ , os segundos membros devem ser  $I(0)$ . Supondo os erros  $a_{it}$  ruídos brancos, e portanto estacionários, segue-se que  $\alpha_i(P_{1,t-1} - \beta P_{2,t-1}) \sim I(0)$ . Logo, se  $\alpha_1 \neq 0$  e  $\alpha_2 \neq 0$ , segue que  $P_{1t} - \beta P_{2t} \sim I(0)$  e representa uma relação de co-integração entre  $P_{1t}$  e  $P_{2t}$ .

O mesmo vale para um mecanismo de correção de erro mais geral. Suponha que  $X_{1t}$  e  $X_{2t}$  sejam duas séries  $I(1)$ ,  $u_t = X_{1t} - \beta X_{2t} = 0$  seja a relação de equilíbrio e

$$\Delta X_{1t} = \alpha_1(X_{1,t-1} - \beta X_{2,t-1}) + a_{11}(1)\Delta X_{1,t-1} + a_{12}(1)\Delta X_{2,t-1} + a_{1t}, \quad (8.15)$$

$$\Delta X_{2t} = \alpha_2(X_{1,t-1} - \beta X_{2,t-1}) + a_{21}(1)\Delta X_{1,t-1} + a_{22}(1)\Delta X_{2,t-1} + a_{2t}. \quad (8.16)$$

Este é um modelo VAR(1) nas primeiras diferenças com um termo de correção de erro adicionado. Os parâmetros  $\alpha_1$  e  $\alpha_2$  são relacionados à velocidade de ajustamento. Se ambos forem nulos não há relação de longo prazo e não temos um modelo como o acima.

Se  $\mathbf{X}_t = (X_{1t}, X_{2t})'$ , podemos escrever (8.15)-(8.16) como

$$\Delta \mathbf{X}_t = \alpha \beta' \mathbf{X}_{t-1} + \mathbf{A} \Delta \mathbf{X}_{t-1} + \mathbf{a}_t, \quad (8.17)$$

com

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 \\ -\beta \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11}(1) & a_{12}(1) \\ a_{21}(1) & a_{22}(1) \end{bmatrix}.$$

Vemos que (8.17) também pode ser escrita

$$\mathbf{X}_t - \mathbf{X}_{t-1} = \alpha \beta' \mathbf{X}_{t-1} + \mathbf{A}(\mathbf{X}_{t-1} - \mathbf{X}_{t-2}) + \mathbf{a}_t,$$

ou

$$\mathbf{X}_t = (\mathbf{I}_n + \mathbf{A} + \alpha \beta') \mathbf{X}_{t-1} - \mathbf{A} \mathbf{X}_{t-2} + \mathbf{a}_t, \quad (8.18)$$

logo variáveis que são co-integradas podem ser geradas por um processo VAR.

Considere, agora, um modelo VAR(1)  $n$ -dimensional,

$$\mathbf{X}_t = \Phi \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (8.19)$$

Pela Proposição 7.2, o processo  $\mathbf{X}_t$  será estacionário se todas as soluções de  $|\mathbf{I}_n - \Phi z| = 0$  estiverem fora do círculo unitário. Suponha que o processo seja não-estacionário, com uma ou mais raízes sobre o círculo unitário. Isto é equivalente a dizer que um ou mais auto-valores de  $\Phi$  são iguais a um, os demais estando dentro do círculo unitário.

Como  $|\mathbf{I}_n - \Phi| = 0$ , a matriz  $\mathbf{\Pi} = \mathbf{I}_n - \Phi$  é singular. Suponha que o seu posto seja  $\rho(\mathbf{\Pi}) = r < n$ , de modo que  $\mathbf{\Pi}$  pode ser decomposta como  $\mathbf{\Pi} = \alpha\beta'$ , onde  $\alpha$  e  $\beta$  têm ordem  $n \times r$  e posto  $r$ .

Suponha que as componentes de  $\mathbf{X}_t$  sejam todas  $I(1)$ . Então, de (8.19),

$$\mathbf{X}_t - \mathbf{X}_{t-1} = -(\mathbf{I}_n - \Phi)\mathbf{X}_{t-1} + \mathbf{a}_t,$$

ou

$$\Delta\mathbf{X}_t = -\mathbf{\Pi}\mathbf{X}_{t-1} + \mathbf{a}_t. \quad (8.20)$$

Portanto, como

$$\alpha\beta'\mathbf{X}_{t-1} = -\Delta\mathbf{X}_t + \mathbf{a}_t,$$

o segundo termo é  $I(0)$ , logo  $\alpha\beta'\mathbf{X}_{t-1}$  é  $I(0)$  e continua a ser  $I(0)$  se o multiplicarmos por  $(\alpha'\alpha)^{-1}\alpha'$ , resultando  $\beta'\mathbf{X}_{t-1} \sim I(0)$  e, finalmente,  $\beta'\mathbf{X}_t \sim I(0)$ . Segue-se que cada linha de  $\beta'\mathbf{X}_t$  representará uma relação de co-integração.

Conclui-se que, a partir de um VAR(1)  $n$ -dimensional, obtemos um modelo nas primeiras diferenças com variáveis co-integradas.

É fácil ver que, para um VAR(2)

$$\mathbf{X}_t = \Phi_1\mathbf{X}_{t-1} + \Phi_2\mathbf{X}_{t-2} + \mathbf{a}_t, \quad (8.21)$$

obtemos

$$\Delta\mathbf{X}_t = \mathbf{D}_1\Delta\mathbf{X}_{t-1} - \mathbf{\Pi}\mathbf{X}_{t-2} + \mathbf{a}_t, \quad (8.22)$$

com  $\mathbf{\Pi} = \mathbf{I}_n - \Phi_1 - \Phi_2$  e  $\mathbf{D}_1 = -(\mathbf{I}_n - \Phi_1)$ . Veja o Problema 4.

Outra maneira de escrever (8.21) é

$$\Delta\mathbf{X}_t = \mathbf{F}_1\Delta\mathbf{X}_{t-1} - \mathbf{\Pi}\mathbf{X}_{t-1} + \mathbf{a}_t, \quad (8.23)$$

onde  $\mathbf{F}_1 = -\Phi_2$ ,  $\mathbf{\Pi}$  como antes. Esta forma é a chamada *forma de correção de equilíbrio* ou *de correção de erros*. Em (8.23) a matriz de níveis defasados,  $\mathbf{\Pi}$ , está no instante  $t-1$ , mas pode ser escolhida estar em qualquer “lag”, sem mudar a verossimilhança, como em (8.22). A forma (8.23) é mais apropriada se quisermos discriminar entre efeitos de ajustamento de curto prazo a relações

de longo prazo e os efeitos de variações nas diferenças defasadas (Hendry e Juselius, 2001).

Por outro lado, tomando-se uma diferença em (8.19), obtemos

$$\Delta \mathbf{X}_t = \Phi \Delta \mathbf{X}_{t-1} + \Delta \mathbf{a}_t,$$

que tem uma parte de médias móveis não invertível, logo não obtemos uma representação VAR para as primeiras diferenças. O mesmo ocorre com (8.21).

Considere, agora, um modelo VAR( $p$ ) genérico,

$$\mathbf{X}_t = \Phi_0 + \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_p \mathbf{X}_{t-p} + \mathbf{a}_t. \quad (8.24)$$

Este diz-se *co-integrado de posto  $r$*  se  $\Pi = \mathbf{I}_n - \Phi_1 - \dots - \Phi_p$  tiver posto  $r$  e portanto puder ser escrita como  $\Pi = \alpha \beta'$ ,  $\alpha$  e  $\beta$  de ordem  $n \times r$  e posto  $r$ . Dizemos que  $\beta$  é a *matriz de co-integração* ou de *vetores co-integrados* e  $\alpha$  é a *matriz de cargas*. Se  $r = 0$ , então  $\Delta \mathbf{X}_t$  tem uma representação VAR( $p-1$ ) estacionária e se  $r = n$ , então o vetor  $\mathbf{X}_t$  tem uma representação VAR( $p$ ) estacionária.

Neste caso, a representação (8.20) ou (8.22) fica

$$\Delta \mathbf{X}_t = \Phi_0 + \mathbf{D}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{D}_{p-1} \Delta \mathbf{X}_{t-p+1} - \Pi \mathbf{X}_{t-p} + \mathbf{a}_t, \quad (8.25)$$

com  $\mathbf{D}_i = -(\mathbf{I}_n - \Phi_1 - \dots - \Phi_i)$ ,  $i = 1, 2, \dots, p-1$ .

O processo (8.24) terá uma representação ou *modelo de correção de erros* (MCE)

$$\Delta \mathbf{X}_t = \Phi_0 - \alpha \beta' \mathbf{X}_{t-1} + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t, \quad (8.26)$$

onde  $\mathbf{F}_i = -(\Phi_{i+1} + \dots + \Phi_p)$ ,  $i = 1, 2, \dots, p-1$ .

Como  $\Delta \mathbf{X}_t \sim \mathbf{I}(0)$  e  $\beta' \mathbf{X}_{t-1} \sim \mathbf{I}(0)$ , estes termos têm média constante; sejam  $E(\Delta \mathbf{X}_t) = \mathbf{c}$ , um vetor  $n \times 1$ , representando taxas de crescimento, e  $E(\beta' \mathbf{X}_{t-1}) = \mu$ , um vetor  $r \times 1$ , representando interceptos nas relações de co-integração. Temos, então, que

$$\Phi_0 = (\mathbf{I}_n - \mathbf{F}_1 - \dots - \mathbf{F}_{p-1})\mathbf{c} + \alpha \mu.$$

Segue-se que o termo contante é a soma de duas parcelas, uma relacionada com o crescimento dos dados e outra com os interceptos nas relações de co-integração. Podemos, então, escrever (8.26) como

$$\Delta \mathbf{X}_t - \mathbf{c} = \sum_{i=1}^{p-1} \mathbf{F}_i (\Delta \mathbf{X}_{t-i} - \mathbf{c}) - \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{X}_{t-1} - \boldsymbol{\mu}) + \mathbf{a}_t, \quad (8.27)$$

e vemos que há duas formas de correção de equilíbrio em (8.27): uma do crescimento dos dados em relação à sua média e, outra, dos vetores de co-integração em relação à sua média. Em análises de séries reais, temos que verificar se  $\mathbf{c}$  e  $\boldsymbol{\mu}$  são diferentes de zero ou não.

Podemos obter estimadores de máxima verossimilhança para os parâmetros  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ ,  $\mathbf{F}$  e  $\boldsymbol{\Sigma}$  do modelo VAR( $p$ ) co-integrado, onde  $\boldsymbol{\Sigma}$  é a matriz de covariâncias de  $\mathbf{a}_t$ . Veja Lütkepohl (1991) para detalhes.

**Exemplo 8.5.** (Hendry e Juselius, 2001). Sejam  $P_{1t}$  e  $P_{2t}$  os preços de gasolina em dois locais e  $P_{3t}$  o preço do petróleo. Uma relação de co-integração entre preços de gasolina existiria se, por exemplo, diferenciais de preços entre dois locais quaisquer fossem estacionários. Considere o modelo VAR(1)

$$\Delta \mathbf{P}_t = \boldsymbol{\Phi}_0 \mathbf{D}_t + \boldsymbol{\Phi}_1 \Delta \mathbf{P}_{t-1} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{P}_{t-1} + \mathbf{a}_t,$$

onde  $\mathbf{P}_t = (P_{1t}, P_{2t}, P_{3t})'$ ,  $\mathbf{a}_t = (a_{1t}, a_{2t}, a_{3t})' \sim \text{i.i.d } \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ ,  $\boldsymbol{\Phi}_1 = [\phi_{ij}]$ ,  $i, j = 1, 2, 3$ ,

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

Segue-se que podemos explicar variações de preços entre dois períodos consecutivos como resultados de:

- (a) um termo contendo constantes e variáveis “dummies”, como por exemplo alguma intervenção no mercado global;
- (b) um ajustamento a variações de preços no instante anterior, com impactos dados pelos  $\phi_{ij}$ ;

- (c) um ajustamento ao desequilíbrio anterior entre preços em diferentes locais ( $P_{1t} - P_{2t}$ ) e entre o preço no local 2 e o preço do petróleo ( $P_{2t} - P_{3t}$ ), com impactos  $\alpha_{i,1}$  e  $\alpha_{i1}$ ;
- (d) choques aleatórios  $a_{it}$ .

Neste exemplo, teremos duas relações de co-integração, dadas por  $u_{1t} = P_{1t} - P_{2t}$ ,  $u_{2t} = P_{2t} - P_{3t}$ , se  $P_{it} \sim I(1)$  e  $u_{it} \sim I(0)$ . Estas relações significam que os três preços seguem relações de equilíbrio de longo prazo. Veja Hendry e Juselius (2001) para uma análise de um exemplo de preços de gasolina nos Estados Unidos.

## 8.4 Testes para Co-integração

Para se concluir que duas ou mais séries são co-integradas poderíamos pensar que bastaria analisar os seus gráficos. Todavia isso não é suficiente. Co-integração pode ou não existir entre séries que parecem ter uma tendência comum de longo prazo. É necessário usar testes formais e nesta seção veremos dois procedimentos para testar a existência de co-integração entre duas ou mais séries.

Suponha o vetor  $\mathbf{X}_t$ , de ordem  $n \times 1$ , com todas as componentes  $I(1)$ . Podemos destacar duas situações:

- (a) há, no máximo, um vetor de co-integração; este caso foi tratado por Engle e Granger (1987);
- (b) há  $r$ ,  $0 \leq r < n$ , vetores de co-integração, caso considerado por Johansen (1988).

Além dessas referências, veja Zivot e Wang (2003), que também é uma referência para o uso do program `S+FinMetrics`.

### 8.4.1 Procedimento de Engle e Granger

Seja  $u_t = \beta' \mathbf{X}_t$  o resíduo de co-integração. O teste de Engle e Granger consiste em dois passos:

- (i) forme os resíduos de co-integração;

(ii) faça um teste de raízes unitárias para determinar se esses resíduos são  $I(0)$ . Temos as hipóteses:

$$\begin{aligned} H_0 &: u_t \sim I(1) : \text{não há co-integração,} \\ H_1 &: u_t \sim I(0) : \text{há co-integração.} \end{aligned} \quad (8.28)$$

Temos, ainda, dois casos a considerar:

[1] O vetor de co-integração é conhecido e fixado

Por exemplo, o vetor é especificado pela teoria econômica. Use um teste ADF ou PP para testar  $H_0$  contra  $H_1$ .

**Exemplo 8.6.** Consideremos as duas séries geradas no exemplo 8.4, onde  $\beta = (1, -1)'$ . Usando a função `unitroot` do `S+FinMetrics` obtemos o Quadro 8.1, teste ADF com 2 lags e constante na regressão.

```
Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root
Type of Test: t test
Test Statistic: -12.39
P-value: 2.562e-24

Coefficients:
      Value Std. Error  t value Pr(>|t|)
lag1  -0.5436   0.0439   -12.3905  0.0000
lag2   0.1136   0.0444    2.5560  0.0109
constant -0.0826  0.0463   -1.7835  0.0751

Regression Diagnostics:

      R-Squared 0.2552
Adjusted R-Squared 0.2522
Durbin-Watson Stat 1.9787

Residual standard error: 1.024 on 494 degrees of freedom
F-statistic: 84.82 on 2 and 495 degrees of freedom, the p-value is 0
```

Quadro 8.1: Teste ADF para o exemplo 8.6

Vemos que o valor observado da estatística ADF é  $-12,39$ , o que conduz

à rejeição da hipótese  $H_0$ , que  $X_{1t}$  e  $X_{2t}$  não sejam co-integradas, com vetor de co-integração  $\beta = (1, -1)'$ .

[2] O vetor de co-integração é estimado

Para o caso de duas séries  $\mathbf{X}_t = (X_{1t}, X_{2t})'$ , considere a regressão

$$X_{2t} = \alpha + \beta X_{1t} + u_t,$$

e use os resíduos de MQO  $\hat{u}_t$  para o teste de raiz unitária.

No caso geral de  $\mathbf{X}_t$ , de ordem  $n \times 1$ , considere  $\mathbf{X}_t = (X_{1t}, \mathbf{X}_{2t}')'$ , com  $\mathbf{X}_{2t} = (X_{2t}, \dots, X_{nt})'$  e a regressão

$$X_{1t} = \alpha + \beta_2' \mathbf{X}_{2t} + u_t, \quad (8.29)$$

para obter os resíduos de MQO  $\hat{u}_t$  e testar  $H_0$  contra  $H_1$ .

Formalmente, o teste ADF não é apropriado;  $\hat{u}_t$  é o resíduo de uma regressão na qual o vetor de co-integração é estimado e sob a hipótese de não-co-integração, tal vetor não será identificado: usando os valores críticos de  $\hat{\tau}_\mu$  rejeitaremos tal hipótese muito freqüentemente. Além disso, quando há mais de duas séries  $I(1)$  o procedimento de Engle e Granger pode ter um viés importante. Ou seja, estimativas diferentes do vetor de co-integração são obtidas dependendo da escolha da variável resposta.

Phillips e Ouliaris (1990) desenvolveram testes de raiz unitária apropriadas aos resíduos de (8.29); as estatísticas usadas não têm a distribuição assintótica usual ADF ou PP sob a hipótese  $H_0$  de não existência de co-integração. A distribuição assintótica é também uma função do movimento browniano padrão e depende dos termos determinísticos incluídos em (8.29) e do número de variáveis em  $\mathbf{X}_{2t}$ . Estas distribuições são conhecidas como distribuições de Phillips e Ouliaris (PO). Hansen (1992) mostrou que as distribuições de PO dependem também do comportamento das tendências presentes em  $X_{1t}$  e  $\mathbf{X}_{2t}$ . Os casos por ele considerados foram:

- (a)  $X_{1t}$  e  $\mathbf{X}_{2t}$  são ambos  $I(1)$ , sem “drift”. Use as estatísticas de PO, com constante e dimensão  $n - 1$ ;
- (b)  $\mathbf{X}_{2t}$  é  $I(1)$  com “drift” e  $X_{1t}$  pode ou não ser  $I(1)$  com “drift”. Use PO com constante e tendência, dimensão  $n - 2$ . Se  $n = 2$ , use as distribuições usuais ADF e PP, com constante e tendência;



(c)  $X_{1t}$  é  $I(1)$  com “drift” e  $\mathbf{X}_{2t}$  é  $I(1)$  sem “drift”. Neste caso, considere a regressão

$$X_{1t} = \alpha_0 + \alpha_1 t + \beta_2' \mathbf{X}_{2t} + u_t. \quad (8.30)$$

Use PO para os resíduos  $\hat{u}_t$  de (8.30), com constante, tendência e dimensão  $n - 1$ .

Para obter quantis e p-valores das distribuições de PO, use as funções `qpoint` e `ppoint` do `S+FinMetrics`.

**Exemplo 8.7.** Retomemos o exemplo anterior e suponha, agora, que temos que estimar  $\beta = (1, -\beta_2)'$ . Aqui, o modelo é

$$X_{2t} = \alpha + \beta_2 X_{1t} + u_t.$$

Os EMQO dos parâmetros são  $\hat{\alpha} = -0,2714$  e  $\hat{\beta}_2 = 0,9293$ . Observe que a estimativa de  $\beta_2$  está próxima do valor verdadeiro, um. Como a média dos  $\hat{u}_t$  é zero, os testes de raiz unitária são feitos para modelo sem constante ou tendência. Nos quadros 8.2 e 8.3 temos os resultados da aplicação dos testes ADF e PP.

Como  $X_{1t}$  e  $X_{2t}$  são  $I(1)$  sem “drift”, obtemos os quantis  $-3,0448$ ,  $-3,3361$  e  $-3,8967$  com os níveis 0,10, 0,05 e 0,01, respectivamente. Ambas as estatísticas rejeitam a existência de raiz unitária, com nível 0,01. Logo  $u_t$  é  $I(0)$  e as séries são co-integradas.

#### 8.4.2 Procedimento de Johansen

O procedimento de Johansen é uma generalização multivariada do teste de DF. Considere o modelo (8.26) re-escrito na forma

$$\Delta \mathbf{X}_t = \Phi_0 \mathbf{D}_t + \alpha \beta' \mathbf{X}_{t-1} + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t, \quad (8.31)$$

onde agora  $\Pi = \Phi_1 + \dots + \Phi_p - \mathbf{I}_n$  e  $\mathbf{D}_t$  contém termos determinísticos (constantes, tendências etc).

Test for Unit Root: Augmented DF Test

Null Hypothesis: there is a unit root

Type of Test: t test

Test Statistic: -11.66

P-value: 2.348e-23

Coefficients:

	Value	Std. Error	t value	$Pr(>  t )$
lag1	-0.4942	0.0424	-11.6636	0.0000
lag2	0.0876	0.0445	1.9674	0.0497

Regression Diagnostics:

R-Squared 0.2346

Adjusted R-Squared 0.2330

Durbin-Watson Stat 1.9793

Residual standard error: 0.9681 on 494 degrees of freedom

F-statistic: 76 on 2 and 496 degrees of freedom, the p-value is 0

Quadro 8.2: Teste ADF para o exemplo 8.7

O procedimento de Johansen (1988, 1995) para testar a existência de co-integração é baseado nos seguintes passos:

- (i) verificar a ordem de integração das séries envolvidas; verificar a existência de tendências lineares;
- (ii) especificar e estimar um modelo  $\text{VAR}(p)$  para  $\mathbf{X}_t$ , que supomos  $I(1)$ ;
- (iii) construir testes da razão de verossimilhanças (RV) para se determinar o número de vetores de co-integração, que sabemos ser igual ao posto de  $\mathbf{\Pi}$ ;
- (iv) dados os vetores de co-integração (normalizados apropriadamente), estimar o MCE (via EMV).

```

Test for Unit Root: Phillips-Perron Test

Null Hypothesis: there is a unit root
Type of Test: t test
Test Statistic: -12.09
P-value: 1.98e-24
using bartlett window with bandwidth 5

Coefficients:
      Value Std. Error  t value Pr(>|t|)
lag1  -0.4539   0.0373  -12.1836   0.0000

Regression Diagnostics:

      R-Squared 0.2296
Adjusted R-Squared 0.2296
Durbin-Watson Stat 1.9053

Residual standard error: 0.97 on 495 degrees of freedom
F-statistic: 148.4 on 1 and 498 degrees of freedom, the p-value is 0

```

### Quadro 8.3: Teste PP para o exemplo 8.7

Segundo Johansen (1994, 1995), os termos determinísticos em (8.31) são restritos à forma

$$\Phi_0 \mathbf{D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t. \quad (8.32)$$

Para verificarmos o efeitos dos termos determinísticos no modelo VAR, consideremos um caso especial:

$$\Delta \mathbf{X}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{a}_t, \quad (8.33)$$

onde  $\boldsymbol{\mu}_0$  e  $\boldsymbol{\mu}_1$  são ambos vetores  $n \times 1$ . Vamos decompor estes dois vetores em relação à média das relações de co-integração e em relação à média das taxas de crescimento,

$$\begin{aligned} \boldsymbol{\mu}_0 &= \boldsymbol{\alpha} \boldsymbol{\rho}_0 + \mathbf{c}_0, \\ \boldsymbol{\mu}_1 &= \boldsymbol{\alpha} \boldsymbol{\rho}_1 + \mathbf{c}_1. \end{aligned} \quad (8.34)$$

Então podemos escrever

$$\begin{aligned}\Delta \mathbf{X}_t &= \alpha \rho_0 + \mathbf{c}_0 + \alpha \rho_1 t + \mathbf{c}_1 t + \alpha \beta' \mathbf{X}_{t-1} + \mathbf{a}_t \\ &= \alpha (\rho_0, \rho_1, \beta') \begin{pmatrix} 1 \\ t \\ X_{t-1} \end{pmatrix} + (\mathbf{c}_0 + \mathbf{c}_1 t) + \mathbf{a}_t,\end{aligned}$$

ou ainda,

$$\Delta \mathbf{X}_t = \alpha \begin{pmatrix} \rho_0' \\ \rho_1' \\ \beta \end{pmatrix} \mathbf{X}_{t-1}^* + (\mathbf{c}_0 + \mathbf{c}_1 t) + \mathbf{a}_t, \quad (8.35)$$

com  $\mathbf{X}_{t-1}^* = (1, t, \mathbf{X}_{t-1}')'$ .

Podemos sempre escolher  $\rho_0$  e  $\rho_1$  tais que o erro de equilíbrio  $(\beta^*)' \mathbf{X}_t^* = \mathbf{v}_t$  tenha média zero, logo

$$E(\Delta \mathbf{X}_t) = \mathbf{c}_0 + \mathbf{c}_1 t. \quad (8.36)$$

Note que se  $\mathbf{c}_0 \neq \mathbf{0}$  temos um crescimento constante nos dados e se  $\mathbf{c}_1 \neq \mathbf{0}$  temos uma tendência linear nas diferenças ou tendência quadrática nos níveis das variáveis.

Há cinco casos a considerar.

**Caso 1.** constante nula,  $\mu_t = \mathbf{0}$ ; neste caso,  $\rho_0 = \rho_1 = \mathbf{0}$  e o modelo não possui qualquer componente determinística, com  $\mathbf{X}_t \sim I(1)$  sem “drift” (não há crescimento dos dados) e as relações de co-integração têm média zero. A menos que  $\mathbf{X}_0 = \mathbf{0}$ , este caso tem pouco interesse nas aplicações práticas.

**Caso 2.** constante restrita,  $\mu_t = \mu_0 = \alpha \rho_0$ ; neste caso,  $\rho_1 = \mathbf{0}, \mathbf{c}_0 = \mathbf{0}$ , mas  $\rho_0 \neq \mathbf{0}$  e portanto não há tendência linear nos dados e as relações de co-integração têm média  $\rho_0$ .

**Caso 3.** constante irrestrita,  $\mu_t = \mu_0$ ; aqui,  $\rho_1 = \mathbf{0}$ , as séries de  $\mathbf{X}_t$  são  $I(1)$  sem “drift” e as relações de co-integração podem ter médias diferentes de zero.

**Caso 4.** tendência restrita,  $\mu_t = \mu_0 + \alpha \rho_1 t$ ; neste caso,  $\mathbf{c}_1 = \mathbf{0}$ , mas  $\mathbf{c}_0, \rho_0, \rho_1$

são irrestritos. As séries são  $I(1)$  com “drift” e as relações de co-integração têm uma tendência linear.

**Caso 5.** tendência irrestrita,  $\mu_t = \mu_0 + \mu_1 t$ ; não há nenhuma restrição sobre  $\mu_0$  e  $\mu_1$ , as séries são  $I(1)$  com tendência linear (logo tendência quadrática nos níveis) e as relações de co-integração têm tendência linear. Previsões podem ser ruins, logo deve-se ter certo cuidado em se adotar esta opção.

Veja Hendry e Juselius (2001) e Zivot e Wang (2003) para detalhes.

Sabemos que o posto de  $\Pi$  fornece também o número de autovalores não-nulos de  $\Pi$ ; suponha que os ordenemos  $\lambda_1 > \lambda_2 > \dots > \lambda_n$ . Se as séries são *não* co-integradas,  $\rho(\Pi) = 0$  e todas os auto-valores serão nulos, ou ainda  $\ell n(1 - \lambda_i) = 0$ , para todo  $i$ . Um teste da RV para testar o posto de  $\Pi$  é baseado na *estatística traço*

$$\lambda_{\text{traço}}(r_0) = -T \sum_{i=r_0+1}^n \ell n(1 - \hat{\lambda}_i), \quad (8.37)$$

onde  $\hat{\lambda}_i$  são os auto-valores estimados de  $\Pi$  e (8.37) testa

$$\begin{aligned} H_0 &: r \leq r_0, \\ h_1 &: r > r_0, \end{aligned} \quad (8.38)$$

sendo  $r$  o posto de  $\Pi$ . Se  $\rho(\Pi) = r_0$ , então  $\hat{\lambda}_{r_0+1}, \dots, \hat{\lambda}_n$  são aproximadamente nulas e a estatística (8.37) será pequena; caso contrário, será grande. Como dissemos acima, a distribuição assintótica de (8.37) é uma generalização multivariada da distribuição ADF e depende da dimensão  $n - r_0$  e da especificação dos termos determinísticos. Os valores críticos podem ser encontrados em Osterwald-Lenum (1992) para os casos (a)-(e) acima e  $n - r_0 = 1, \dots, 10$ .

Johansen também usa a *estatística do máximo auto-valor*

$$\lambda_{\text{max}}(r_0) = -T \ell n(1 - \hat{\lambda}_{r_0+1}), \quad (8.39)$$

para testar

$$\begin{aligned} H_0 &: r = r_0, \\ H_1 &: r = r_0 + 1. \end{aligned} \quad (8.40)$$

A distribuição assintótica de (8.39) também depende de  $n - r_0$  e da especificação de termos determinísticos. Valores críticos podem ser encontrados na referência acima citada.

Supondo-se que o posto de  $\Pi$  é  $r$ , Johansen (1988) prova que o estimador de máxima verossimilhança de  $\beta$  é dado por  $\hat{\beta}_{MV} = (\hat{v}_1, \dots, \hat{v}_r)$ , onde  $\hat{v}_i$  é o auto-vetor associado ao auto-valor  $\hat{\lambda}_i$  e os estimadores de máxima verossimilhança dos parâmetros restantes são obtidos por meio de uma regressão multivariada com  $\beta$  substituído pelo EMV. Johansen (1995) mostra a normalidade assintótica dos estimadores de  $\beta$ , com taxa de convergência  $T^{-1}$ . Veja também a seção 8.5.

**Exemplo 8.8.** Considere  $T = 250$  valores dos sistema dado no problema 6, sendo o vetor de co-integração  $\beta = (1; -0,5; -0,5)'$ ,  $u_t$  gerado por um modelo AR(1) com parâmetro 0,75 e erro  $\mathcal{N}(0, (0,5)^2)$ ,  $v_t, w_t$  ambos normais independentes, com média zero e desvio padrão 0,5. Veja a Figura 8.2. Usando o critério AIC (ou BIC) selecionamos a ordem  $p = 1$  para o VAR( $p$ ) a ser ajustado, de modo que  $p - 1 = 0$  no modelo de correção de erros.

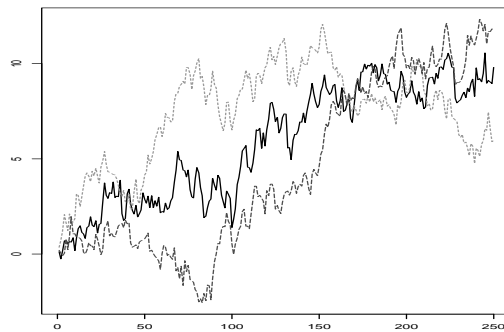


Figura 8.2: Sistema trivariado co-integrado:  $X_{1t}$  (linha cheia),  $X_{2t}$  (linha pontilhada) e  $X_{3t}$  (linha tracejada)

```

Call:
coint(Y = cbind(y1, y2, y3), lags = 0, trend = "rc")

Trend Specification:
H1*(r): Restricted constant

Trace tests significant at the 5% level are flagged by ' + '.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged by ' * '.
Max Eigenvalue tests significant at the 1% level are flagged by '***'.

Tests for Cointegration Rank:
      Eigenvalue Trace Stat  95% CV  99% CV Max Stat  95% CV
H(0)+ ** 0.1058    40.3476   34.9100  41.0700  27.8569  22.0000
H(1)    0.0360    12.4907   19.9600  24.6000   9.1353  15.6700
H(2)    0.0134     3.3554    9.2400  12.9700   3.3554   9.2400

      99% CV
H(0)+ ** 26.8100
H(1)    20.2000
H(2)    12.9700

      y1      y2      y3 Intercept*
1.268086 -0.6427168 -0.7006781 0.4312081
y1      y2      y3 Intercept*
-1 0.5068402 0.552548 -0.3400465

```

Quadro 8.4: Teste de Johansen para o exemplo 8.8

No Quadro 8.4 temos os valores das estatísticas  $\lambda_{\text{traço}} = 40,35$  e  $\lambda_{\text{max}} = 27,86$ . Notamos que ambas são significativas com o nível 0,05 para testar a hipótese  $H_0$  de que não há co-integração contra a alternativa que há mais que uma (uma, respectivamente) relações de integração. Por outro lado, a hipótese nula de uma relação de co-integração contra a alternativa de mais que uma (duas, respectivamente) é aceita, com o nível 0,05, sendo  $\lambda_{\text{traço}} = 12,49$  e  $\lambda_{\text{max}} = 9,14$ . Concluimos, então, que há somente um vetor de co-integração. O quadro apresenta, também, o vetor não-normalizado e o vetor normalizado, além do coeficiente de tendência, supondo o Caso 2 acima. O vetor de co-integração estimado é  $\hat{\beta} = (1; -0,507; -0,553)'$ , sendo que o vetor verdadeiro tem  $\beta_2 = \beta_3 = 0,5$ .

No Quadro 8.5 temos o resultado da estimação do MCE. Vemos que o termo de tendência não é significativo, o mesmo ocorrendo com os coeficientes da segunda e terceira equação, de modo que o modelo reduz-se a

```

Call:
VECM(test = coint.rc)

Cointegrating Vectors:
      coint.1
      1.0000

      y2 -0.4933
(std.err) 0.0855
(t.stat) -5.7694

      y3 -0.5320
(std.err) 0.1016
(t.stat) -5.2363

      Trend* -0.3698
(std.err) 1.7663
(t.stat) -0.2094

VECM Coefficients:
      y1      y2      y3
coint.1 -0.1886  0.0323  0.0065
(std.err) 0.0449  0.0411  0.0406
(t.stat) -4.2006  0.7867  0.1593

Intercept -0.0300  0.0343  0.0487
(std.err) 0.0389  0.0356  0.0351
(t.stat) -0.7713  0.9660  1.3865

Regression Diagnostics:
      y1      y2      y3
R-squared 0.0667  0.0025  0.0001
Adj. R-squared 0.0629 -0.0015 -0.0039
Resid. Scale 0.5563  0.5087  0.5027

Information Criteria:
      logL      AIC      BIC      HQ
-532.1524 1068.3049 1075.3398 1071.1365

      total residual
Degree of freedom: 249      247

```

Quadro 8.5: Estimação do MCE para o exemplo 8.8



$$\Delta X_{1t} = -0,1865(X_{1,t-1} - 0,507X_{2,t-1} - 0,553X_{3,t-1}) + a_{1t}. \quad (8.41)$$

Na Figura 8.3 temos ilustrados os resíduos de co-integração, que parecem ser estacionários.

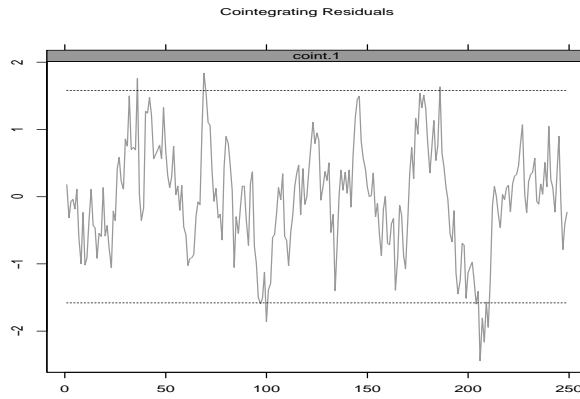


Figura 8.3: Resíduos de co-integração para o exemplo 8.8

**Exemplo 8.9.** Como um último exemplo, consideremos as séries do Ibovespa e dos preços das ações da Petrobrás, mostradas na Figura 8.4. O Quadro 8.6 traz os resultados da aplicação do Teste de Johansen, mostrando que as séries não são co-integradas. Foi ajustado um VAR(6) e usou-se o Caso 2 para o termo determinístico.

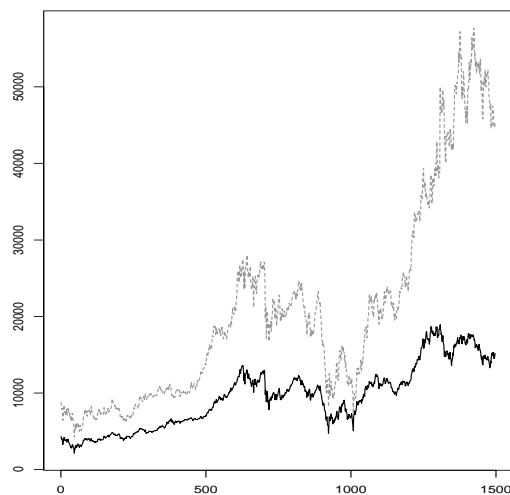


Figura 8.4: Índice Ibovespa(linha cheia) e Petrobrás (linha tracejada)

## 8.5 Comentários Finais

Encerramos este capítulo com algumas observações sobre processos co-integrados, bem como uma apresentação um pouco mais detalhada sobre estimadores de máxima verossimilhança (EMV) do MCE.

### 8.5.1 Enfoque Bayesiano

Existem alguns trabalhos sobre o tratamento bayesiano de processos co-integrados, como Koop (1992), Tsurumi e Wago (1994) e Phillips (1994). Veja Lopes e Lima (1995) para mais detalhes.

```

Call:
coint(Y = cbind(ibv, petro), lags = 5, trend = "rc")

Trend Specification:
H1*(r): Restricted constant

Trace tests significant at the 5% level are flagged by ' +'.
Trace tests significant at the 1% level are flagged by '++'.
Max Eigenvalue tests significant at the 5% level are flagged by ' *'.
Max Eigenvalue tests significant at the 1% level are flagged by '***'.

Tests for Cointegration Rank:
      Eigenvalue Trace Stat  99% CV  95% CV Max Stat  99% CV
H(0)    0.0059    11.9308   19.9600  24.6000   8.7635  15.6700
H(1)    0.0021     3.1673    9.2400  12.9700   3.1673   9.2400

      95% CV
H(0)    20.2000
H(1)    12.9700

Unnormalized Cointegrating Vectors:
      ibv  petro Intercept*
coint.1 -0.0006  0.0002  2.1125
coint.2 -0.0001  0.0000  2.0162

Unnormalized Adjustment Coefficients:
      coint.1  coint.2
ibv   -4.1648  11.0152
petro -44.6770  16.4132

Long Run Impact Matrix: alpha*beta'
      ibv  petro
ibv   0.0021 -0.0012
petro  0.0280 -0.0086

```

Quadro 8.6: Estimação do MCE para o exemplo 8.9

### 8.5.2 Termos Determinísticos no MCE

Os Casos (1)-(5) considerados na seção 8.4.2 são usualmente referidos como  $H_2(r)$ ,  $H_1^*(r)$ ,  $H_1(r)$ ,  $H^*(r)$  e  $H(r)$ , respectivamente. Esta nomenclatura é também adotada nos programas computacionais, veja o Quadro 8.4, por exemplo.

O MCE irrestrito é denotado  $H(r)$ , significando que  $\rho(\mathbf{\Pi}) \leq r$ . Obtemos, então, uma seqüência de modelos hierárquicos  $H(0) \subset \dots \subset H(r) \subset \dots \subset H(n)$ , onde  $H(0)$  indica o modelo VAR não co-integrado, com  $\mathbf{\Pi} = \mathbf{0}$  e  $H(n)$  indica o modelo VAR(p) irrestrito estacionário.

### 8.5.3 Procedimento de Máxima Verossimilhança

Para efeito de ilustração, consideremos o modelo VAR(2) escrito na forma de correção de erros, com um termo constante, uma tendência linear e um vetor de variáveis “dummies”,

$$\Delta \mathbf{X}_t = \mathbf{\Phi}_0 + \delta t + \mathbf{\Psi} \mathbf{d}_t + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \alpha \mathbf{\beta}' \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (8.42)$$

Vimos que, para que (8.42) seja internamente consistente,  $\mathbf{\beta}$  deve ter posto reduzido  $r < n$ . Vamos descrever brevemente o procedimento de máxima verossimilhança para estimar (8.42). Os detalhes estão em Johansen (1988, 1995) e Banerjee et al. (1993). A idéia é usar uma verossimilhança concentrada. Defina

$$\begin{aligned} \mathbf{Z}_{0t} &= \Delta \mathbf{X}_t, \\ \mathbf{Z}_{1t} &= \mathbf{X}_{t-1}, \\ \mathbf{Z}_{2t} &= (1, t, \mathbf{d}_t, \Delta \mathbf{X}_{t-1})'. \end{aligned}$$

No caso de um VAR( $p$ ), as diferenças  $\Delta \mathbf{X}_{t-j}$  são incluídas em  $\mathbf{Z}_{2t}$ . Então (8.42) fica

$$\mathbf{Z}_{0t} = \mathbf{\Theta} \mathbf{Z}_{2t} + \alpha \mathbf{\beta}' \mathbf{Z}_{1,t-1} + \mathbf{a}_t,$$

onde  $\mathbf{\Theta} = (\mathbf{\Phi}_0, \delta, \mathbf{\Psi}, \mathbf{F}_1)'$ . Concentrando fora da verossimilhança os efeitos de ajustamento de curto prazo,  $\mathbf{\Theta} \mathbf{Z}_{2t}$ , obtemos um modelo mais simples, por meio das seguintes regressões de MQO:

$$\begin{aligned} \mathbf{Z}_{0t} &= \hat{\mathbf{R}}_1' \mathbf{Z}_{2t} + \mathbf{u}_{1t}, \\ \mathbf{Z}_{1t} &= \hat{\mathbf{R}}_2' \mathbf{Z}_{2t} + \mathbf{u}_{2t}, \end{aligned}$$

em que  $\mathbf{u}_{1t} = \mathbf{Z}_{0t} - \mathbf{M}_{02}\mathbf{M}_{22}^{-1}\mathbf{Z}_{2t}$ ,  $\mathbf{u}_{2t} = \mathbf{Z}_{1t} - \mathbf{M}_{12}\mathbf{M}_{22}^{-1}\mathbf{Z}_{2t}$  são os resíduos de MQO e

$$\mathbf{M}_{ij} = \frac{1}{T} \sum_t \mathbf{Z}_{it} \mathbf{Z}_{jt}'$$

são matrizes de momentos, de modo que  $\hat{\mathbf{R}}_1' = \mathbf{M}_{02}\mathbf{M}_{22}^{-1}$  e  $\hat{\mathbf{R}}_2' = \mathbf{M}_{12}\mathbf{M}_{22}^{-1}$ .

O modelo concentrado pode, então, ser escrito

$$\mathbf{u}_{1t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{u}_{2t} + \mathbf{u}_t, \quad (8.43)$$

que inclui somente relações de equilíbrio de longo prazo.

Os EMV são obtidos em dois estágios:

- (i) para obter um estimador de  $\boldsymbol{\alpha}$ , supomos  $\boldsymbol{\beta}$  conhecido de modo que  $\boldsymbol{\alpha}$  pode ser estimado por MQO em (8.43);
- (ii) insira  $\boldsymbol{\alpha} = \boldsymbol{\alpha}(\hat{\boldsymbol{\beta}})$  na expressão da verossimilhança concentrada e obtenha o EMV de  $\boldsymbol{\beta}$  por otimização não-linear ou por uma regressão multivariada de posto reduzido (Johansen, 1988). Esta solução fornece os  $n$  auto-valores  $\hat{\lambda}_1 > \dots > \hat{\lambda}_n$ . Como vimos, o estimador de  $\boldsymbol{\beta}$  é dado pela matriz  $n \times r$  de auto-vetores, correspondentes aos  $r$  maiores auto-valores  $\hat{\lambda}_i$ . Obtido o EMV de  $\boldsymbol{\beta}$ , calcule  $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\hat{\boldsymbol{\beta}})$ .

## 8.6 Problemas

1. Mostre que, se uma relação de equilíbrio  $X_t + \alpha Y_t \sim I(0)$  existe, ela é única.
2. Sejam  $X(t) \sim I(d_1)$  e  $Y_t \sim I(d_2)$ ,  $d_2 > d_1$ . Mostre que qualquer combinação linear de  $X_t$  e  $Y_t$  é  $I(d_2)$ .
3. Sejam

$$\begin{aligned} X_{1t} &= \beta X_{2t} + \gamma \Delta X_{2,t} + \varepsilon_{1t}, \\ \Delta^2 X_{2t} &= \varepsilon_{2t}, \end{aligned}$$

onde  $\varepsilon_{it}$  são como no exemplo 8.1.

- (a) Mostre que ambas as séries são  $I(2)$ .
- (b) Mostre que  $X_{1t}, X_{2t}$  e  $\Delta X_{2t}$  são co-integradas. Qual é o vetor co-integrado?
- (c) Mostre que  $X_{1t}$  e  $X_{2t}$  são co-integração(2,1).
4. Prove (8.22).
5. Mostre que (8.19) pode ser escrito na forma

$$\Delta^2 \mathbf{X}_t = \mathbf{G} \Delta \mathbf{X}_{t-1} - \mathbf{\Pi} \mathbf{X}_{t-1} + \mathbf{a}_t,$$

onde  $\mathbf{G} = -\mathbf{I}_n - \mathbf{\Phi}_2$  e  $\mathbf{\Pi}$  como antes.

6. Simule o sistema co-integrado (trivariado):

$$\begin{aligned} X_{1t} &= \beta_2 X_{2t} + \beta_3 X_{3t} + u_t, \\ X_{2t} &= X_{2,t-1} + v_t, \\ X_{3t} &= X_{3,t-1} + w_t, \end{aligned}$$

onde  $u_t, v_t$  e  $w_t$  são todas  $I(0)$ . O vetor de co-integração é  $\boldsymbol{\beta} = (1, -\beta_2, -\beta_3)'$ , a primeira equação representa a relação de equilíbrio de L.P. e as duas outras constituem as tendências estocásticas comuns. Os  $u_t$  são os resíduos de co-integração.

7. Simule o sistema co-integrado (trivariado):

$$\begin{aligned} X_{1t} &= \alpha_1 X_{3t} + u_t, \\ X_{2t} &= \alpha_2 X_{3t} + v_t, \\ X_{3t} &= X_{3,t-1} + w_t, \end{aligned}$$

onde  $u_t, v_t$  e  $w_t$  são todas  $I(0)$ . Neste caso, as duas primeiras equações descrevem relações de equilíbrio de L.P. e a terceira descreve a tendência estocástica comum. Há dois vetores de co-integração,  $\boldsymbol{\beta}_1 = (1, 0, -\alpha_1)'$ ,  $\boldsymbol{\beta}_2 = (1, 0, -\alpha_2)'$ , e  $u_t, v_t$  são os resíduos de co-integração.

8. O modelo de demanda por moeda especifica que (em logaritmos, exceto para  $r_t$ )

$$m_t = \beta_0 + \beta_1 p_t + \beta_2 y_t + \beta_3 r_t + e_t,$$

onde :

$m_t$ : demanda por moeda a longo prazo;

$p_t$ : nível de preço;

$y_t$ : renda real (PIB);

$r_t$ : taxa de juros (de curto prazo);

$e_t$ : erro estacionário.

- (a) Supondo as quatro séries  $I(1)$ , mostre que as séries são co-integradas e obtenha o vetor de co-integração normalizado.
- (b) Suponha que exista a seguinte relação entre  $m_t$  e  $y_t$ :

$$m_t = \gamma_0 + \gamma_1(y_t + p_t) + e_{1t},$$

onde o erro  $e_{1t}$  é estacionário. Mostre que neste caso existem dois vetores de co-integração. Especifique a matriz  $\mathbf{B}$ , de posto 2, que contém estes dois vetores.

9. Use o teste de Engle e Granger para testar se as séries simuladas nos problemas 6 e 7 são co-integradas.
10. Refaça o problema anterior com o teste de Johansen.
11. Use os testes de Engle e Granger e Johansen para testar se as séries IBOVESPA (d-ibv95.00-dat), preços de ações do Banespa (d-ban95.00.dat) e preços de ações da Petrobras (d-petro95.00.dat) são co-integradas.
12. Mesmo problema para as séries mensais do Ibovespa (m-ibv94.01.dat) e C-Bond (m-cbond94.01.dat).





# Processos com Memória Longa

## 9.1 Introdução

O processo ARMA( $p, q$ ) é considerado um processo de “memória curta”, uma vez que a f.a.c.  $\rho_j$  decresce rapidamente para zero. Na realidade, pode-se demonstrar que, para tal processo,

$$|\rho_j| \leq Cr^{-j}, \quad j = 1, 2, \dots \quad (9.1)$$

onde  $C > 0$  e  $0 < r < 1$ . A expressão (9.1) garante que a função de auto-correlação decai para zero exponencialmente.

Um processo de memória longa é um processo estacionário em que a função de auto-correlação decresce hiperbolicamente para zero, isto é,

$$\rho_j \sim Cj^{-\alpha}, \quad j \rightarrow \infty, \quad (9.2)$$

onde  $C > 0$  e  $0 < \alpha < 1$ . Pode-se usar o *coeficiente de Hurst*  $H = 1 - \alpha/2$ , de modo que  $1/2 < H < 1$ . Quanto maior  $H$ , maior a memória longa do processo. Pode-se provar que o espectro  $f(\lambda)$  do processo, cuja função de auto-correlação é como em (9.2), tende a  $C_f \lambda^{\alpha-1}$ , para  $\lambda \rightarrow 0$ , onde  $C_f > 0$  constante. Ou seja, a função densidade espectral de um processo de memória longa diverge para  $+\infty$  na frequência zero.

Estudos empíricos, principalmente em Climatologia e Hidrologia (década de 50) revelaram a presença de memória longa (ML) em dados de séries temporais e espaciais. Estas séries apresentam persistência nas auto-correlações amostrais, isto é, dependência significativa entre observações separadas por um longo intervalo de tempo. Estas autocorrelações apresentam o comportamento dado por (9.2). Outra característica desse tipo de série é que sua função densidade espectral é não limitada na frequência zero, o que equivale a dizer que sua função de auto-correlação não é absolutamente somável.

Formalmente, temos a

**Definição 9.1.** *Suponha que  $X_t$  tenha auto-correlação  $\rho_j$ . Dizemos que  $X_t$  possui memória longa se*

$$\lim_{n \rightarrow \infty} \sum_{j=-n}^n |\rho_j| \quad (9.3)$$

*é não-finita.*

O fenômeno de ML foi notado por Hurst (1951, 1957), Mandelbrot e Wallis (1968) e McLeod e Hipel (1978), em conjunção com problemas na área de Hidrologia. Modelos de ML também são de interesse na análise de estudos climáticos, como no estudo da aparente tendência crescente em temperaturas globais devido ao efeito estufa. Veja Seater (1993), por exemplo.

Recentemente (década de 80), os economistas notaram que há evidências que processos de ML descrevem de modo satisfatório dados econômicos e financeiros, tais como taxas de juros e de inflação. Estudos recentes na modelagem da volatilidade de ativos financeiros mostram que tais processos são de grande utilidade. Uma excelente revisão sobre processos de ML em econometria é feita por Baillie (1996).

**Exemplo 9.1.** A Figura 9.1 mostra a conhecida série de índices de preços anuais de trigo de Beveridge (1925), de 1500 a 1869 ( $T = 370$ ), e suas auto-correlações amostrais, notando o seu lento decaimento. Os preços são médias sobre várias localidades na Europa (arquivo a-bev00.69.dat).

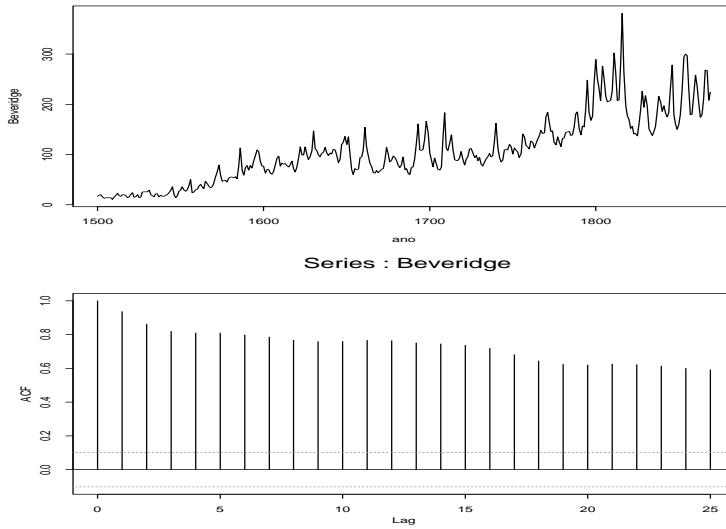


Figura 9.1: (a) Série de índices de preços de trigo de Beveridge (b) f.a.c. amostral.

**Exemplo 9.2.** Temos, na Figura 9.2, as auto-correlações amostrais das séries de valores absolutos dos retornos diários do Ibovespa, Dow Jones, Banespa e Petrobrás. Estes valores absolutos representam a volatilidade da série. O lento decaimento das auto-correlações mostra claramente a persistência da volatilidade. As figuras mostram, também, as auto-correlações de modelos auto-regressivos  $AR(p)$  ajustados às séries. Os valores de  $p$  para as séries do Ibovespa, Dow, Banespa e Petrobrás são, respectivamente, 12, 12, 6 e 17. A série de volatilidades menos persistente é a do Banespa. Vemos que as auto-correlações dos modelos auto-regressivos são boas estimativas para “lags” baixos. Notamos, ainda, o número excessivo de parâmetros do modelo auto-regressivo necessários para capturar a dependência nas séries.

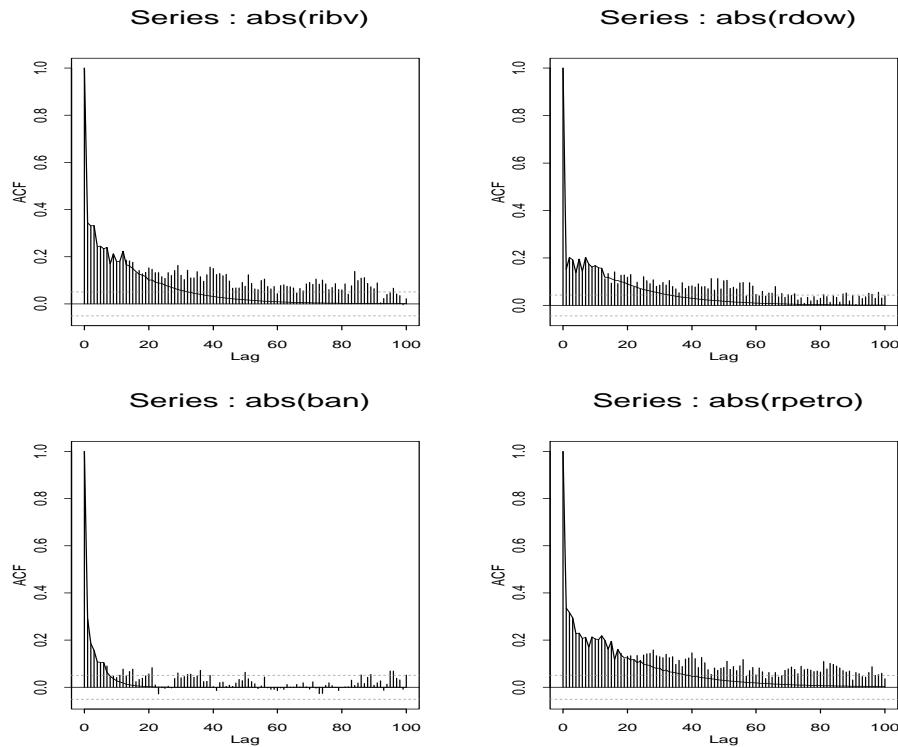


Figura 9.2: Auto-correlações amostrais da volatilidade (a) Ibovespa (b) Dow Jones (c) Banespa (d) Petrobrás, com f.a.c. dos modelos AR ajustadas.

Uma outra característica de séries com memória longa é que as auto-correlações da série original indicam não-estacionariedade, ao passo que a série diferenciada pode parecer ser “super-diferenciada”. Ou seja, processos de ML situam-se entre processos  $I(0)$  e  $I(1)$ .

Procurando respeitar as características de uma série de memória longa, citadas anteriormente, foram definidos dois modelos importantes, nos quais a função de densidade espectral é proporcional a  $\lambda^{-r}$ ,  $1 < r < 2$ , para  $\lambda$  próximo de zero e o decaimento da função de auto-correlação é do tipo (9.2). Primeiro foi introduzido o ruído gaussiano fracionário por Mandelbrot e Van Ness (1968). Mais tarde Granger e Joyeux (1980) e Hosking (1981) introduzi-

ram o modelo ARIMA fracionário (ou ARFIMA), que é uma generalização do modelo ARIMA.

Há trabalhos recentes incorporando ML a processos GARCH, como nos processos FIGARCH (“fractionally integrated generalized autoregressive conditional heteroskedasticity”), introduzidos por Baillie et al. (1996). Também, processos de ML associados a modelos de volatilidade estocástica foram considerados por Harvey (1998) e Breidt et al. (1993). Veja a seção 9.6.

## 9.2 Estimação e Testes para Memória Longa

Nesta seção apresentaremos dois procedimentos para testar se uma série temporal apresenta memória longa e estimar o parâmetro de longa dependência. Um é baseado na estatística R/S e outro no periodograma, que é um estimador do espectro de um processo estacionário.

O modelo proposto para a série  $X_t$  é o *processo integrado fracionário*

$$(1 - B)^d(X_t - \mu) = u_t, \quad (9.4)$$

onde  $u_t$  é um processo estacionário, com espectro  $f_u(\lambda)$ , e para qualquer número real  $d > -1$ , define-se o operador de diferença fracionária

$$\begin{aligned} (1 - B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\ &= 1 - dB + \frac{1}{2!}d(d-1)B^2 - \frac{1}{3!}d(d-1)(d-2)B^3 + \dots, \end{aligned} \quad (9.5)$$

ou seja,

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

A relação existente entre  $d$  e  $H$  é  $d = H - 1/2$ . Se  $0 < d < 1/2$ , então  $X_t$  é estacionário com memória longa. Se  $-1/2 < d < 0$ , dizemos que  $X_t$  é anti-persistente.

### 9.2.1 Estatística R/S

A estatística R/S foi introduzida por Hurst (1951) com o nome “rescaled range” (ou “range over standard deviation”), com o propósito de testar a existência de memória longa numa série temporal.

Dadas as observações  $X_1, \dots, X_T$ , a estatística R/S é dada por

$$Q_T = \frac{1}{S_T} \left[ \max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) \right], \quad (9.6)$$

onde  $\bar{X}$  é a média amostral e  $S_T^2$  é a variância amostral.

Pode-se demonstrar que se  $X_t$  são i.i.d. normais, então  $Q_T/\sqrt{T}$  converge fracamente para uma v.a. que está no domínio de atração de uma ponte browniana. Lo (1991) fornece os quantis desta variável limite. Ele nota que a estatística definida por (9.6) não é robusta à dependência de curta memória e propõe substituir  $Q_T$  por

$$\tilde{Q}_T = \frac{1}{\hat{\sigma}_T(q)} \left[ \max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) \right], \quad (9.7)$$

onde  $\hat{\sigma}_T(q)$  é a raiz quadrada do estimador da variância de longo prazo de Newey-West, com largura de faixa  $q$ , dado por

$$\hat{\sigma}_T^2(q) = S_T^2 \left( 1 + \frac{2}{T} \sum_{j=1}^q w_{qj} r_j \right),$$

sendo  $w_{qj} = 1 - j/(q+1)$ ,  $q < T$  e  $r_j$  são as auto-correlações amostrais usuais de  $X_t$ . Newey and West (1987) sugerem escolher  $q = [4(T/100)^{2/9}]$ .

Se o processo  $X_t$  não tiver ML, a estatística S/L converge para sua distribuição limite à taxa  $T^{1/2}$ , mas se há ML presente, a taxa de convergência é  $T^H$ .

Estes fatos sugerem construir gráficos (na escala log-log) de R/S contra o tamanho amostral. Para uma série com MC os pontos devem estar ao longo

de uma reta com inclinação  $1/2$ , ao passo que para uma série com ML, a reta deve ter inclinação  $H > 1/2$ , para grandes amostras.

Para a construção deste gráfico, considerar os valores de  $R/S$  contra  $k_i$ , para  $k_i = fk_{i-1}$ ,  $i = 2, \dots, s$ ,  $k_1$  grande inicialmente e  $f$  um fator conveniente. Por exemplo, divida a amostra em  $[T/k_i]$  blocos.

A função `rosTest` do `S+FinMetrics` calcula (9.7) com  $q = [4(T/100)^{1/4}]$ . Esta função pode ser usada para testar se há ML na série temporal. A função `d.ros` estima o valor de  $H$  segundo esse procedimento gráfico.

**Exemplo 9.3.** Considere os retornos diários do Ibovespa de 1995 a 2000 e a série de volatilidades, dada pelos valores absolutos dos retornos. Esta série é mostrada na Figura 9.3. O Quadro 9.1 mostra o resultado da aplicação da função `rosTest`. O valor da estatística  $\tilde{Q}_T$  é 2,4619, significativa com o nível 0,01, o que confirma que a série apresenta memória longa. A Figura 9.4 apresenta o log-log plot de  $R/S$ , com a reta ajustada. O valor estimado de  $H = 0,71$ , do que resulta  $d = 0,21$ . O gráfico foi feito com  $k_1 = 5$  e  $f = 2$ . A reta pontilhada no gráfico indica MC ( $H = 1/2$ ).

---

Test for Long Memory: Modified R/S Test  
Null Hypothesis: no long-term dependence

Test Statistics:

2.4619\*\*

\*: significant at 5% level

\*\*: significant at 1% level

Total Observ.: 1498

Bandwidth: 7

---

Quadro 9.1: Teste para ML para volatilidade do Ibovespa.

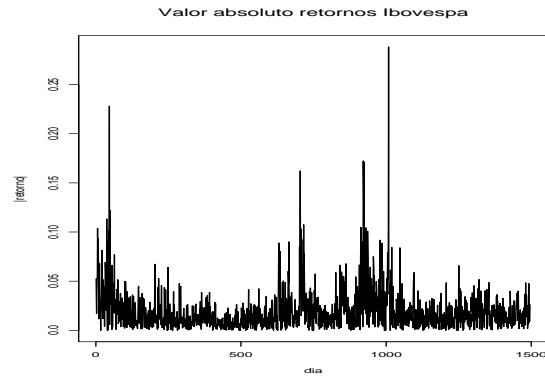


Figura 9.3: Valores absolutos dos retornos do Ibovespa.

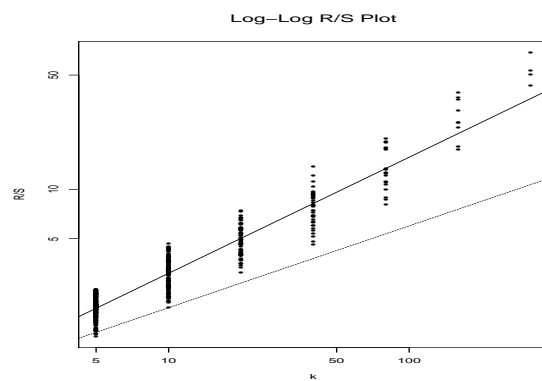


Figura 9.4: Plot R/S para valores absolutos dos retornos do Ibovespa.

### 9.2.2 Procedimento GPH

Este método para estimação do parâmetro de longa memória foi proposto por Geweke e Porter-Hudak (1983) e se baseia na equação que exhibe relação entre os espectros de  $X_t$  e de  $u_t$  em (9.4). Tal equação foi reescrita para que se assemelhasse a uma equação de regressão linear, onde o coeficiente de inclinação envolve o parâmetro  $d$ .

De (9.4) temos que



$$f_x(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_u(\lambda), \quad (9.8)$$

em que  $f_u(\lambda)$  é o espectro de  $u_t$ .

Multiplicando ambos os lados de (9.8) por  $f_u(0)$  e aplicando o logaritmo obtemos

$$\ln f_x(\lambda) = \ln f_u(0) - d \ln |1 - e^{-i\lambda}|^2 + \ln \left( \frac{f_u(\lambda)}{f_u(0)} \right). \quad (9.9)$$

Substituindo  $\lambda$  por  $\lambda_j = 2\pi j/T$  (frequência de Fourier) e adicionando  $\ln(I_x(\lambda_j))$ , a ambos os lados de (9.9), obtemos

$$\begin{aligned} \ln I_x(\lambda_j) &= \ln f_u(0) - d \ln \left( 4 \sin^2 \left( \frac{\lambda_j}{2} \right) \right) \\ &\quad + \ln \left( \frac{f_u(\lambda_j)}{f_u(0)} \right) + \ln \left( \frac{I_x(\lambda_j)}{f_x(\lambda_j)} \right), \end{aligned} \quad (9.10)$$

em que

$$I_x(\lambda_j) = (2\pi T)^{-1} \left| \sum_{t=1}^T X_t \exp(-i\lambda_j t) \right|^2$$

é o *periodograma* dos dados  $X_1, \dots, X_T$ .

O termo  $\ln \left( \frac{f_u(\lambda_j)}{f_u(0)} \right)$  pode ser desprezado quando se considerar apenas as frequências  $\lambda_j$  próximas de zero. Assim, podemos reescrever (9.10) como um modelo de regressão linear

$$Y_j = a - dX_j + \varepsilon_j, \quad j = 1, \dots, m, \quad (9.11)$$

em que

$$\begin{aligned} Y_j &= \ln I_x(\lambda_j), \\ X_j &= \ln \left( 4 \sin^2 \left( \frac{\lambda_j}{2} \right) \right), \\ \varepsilon_j &= \ln \left( \frac{I_x(\lambda_j)}{f_x(\lambda_j)} \right), \\ a &= \ln f_u(0) \text{ e } m = cT^\alpha, \quad 0 < \alpha < 1, \end{aligned}$$

$c$  uma constante. A relação linear (9.11) sugere a utilização de um estimador de mínimos quadrados para  $d$ , isto é,

$$\hat{d}_{MQ} = -\frac{\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^m (X_i - \bar{X})^2}. \quad (9.12)$$

Geweke e Porter-Hudak (1983) demonstram que

$$\hat{d}_{MQ} \xrightarrow{\mathcal{D}} \mathcal{N}\left(d, \frac{\pi^2}{6 \sum_{j=1}^m (X_j - \bar{X})^2}\right).$$

Sob  $H_0 : d = 0$ , isto é, o processo não tem memória longa, a estatística

$$t_{d=0} = \hat{d} \left( \frac{\pi^2}{6 \sum_{j=1}^m (X_j - \bar{X})^2} \right)^{-1/2} \quad (9.13)$$

tem distribuição normal padrão.

O estimador  $d$ , calculado por meio de (9.12) é chamado estimador GPH (de Geweke e Porter-Hudak).

O programa **S+FinMetrics** usa a função **gphTest** para estimar  $d$  e testar  $H_0$  dada acima, usando como “default”,  $m = T^\alpha$ , com  $\alpha = 0,5$ .

**Observação:** o parâmetro  $d$  pode também ser estimado por máxima verossimilhança, juntamente com os parâmetros de um processo ARFIMA ajustado à uma série temporal com memória longa. Veja a seção seguinte.

**Exemplo 9.4.** Consideremos novamente os valores absolutos dos retornos diários do Ibovespa, do exemplo 9.3. O Quadro 9.2 apresenta o resultado da aplicação da função **gphTest**. O valor da estatística é 4,3335, e a hipótese nula de que não há memória longa é rejeitada com o nível 1%. O valor estimado de  $d$  é  $\hat{d} = 0,5256$ , o que sugere ML e não-estacionariedade. Além disso, baseado no erro padrão assintótico fornecido, 0,12128, obtemos o intervalo de confiança  $[0,283; 0,768]$  para  $d$ , com coeficiente de confiança de 95%.

---

Test for Long Memory: GPH Test	
Null Hypothesis: $d = 0$	
Test Statistics:	
$d$	0.5256
stat	4.3335**
*: significant at 5% level	
**: significant at 1% level	
Total Observ.: 1498	
Number of Freq: 38	
[1]	0.1212815

---

Quadro 9.2: Teste para ML para volatilidade do Ibovespa, usando GPH

### 9.3 Modelos ARFIMA

Nesta seção estudaremos uma classe de modelos que são capazes de descrever, simultaneamente, as dinâmicas de memórias curta e longa de processos estacionários.

**Definição 9.2.** Dizemos que  $\{X_t\}$  é um processo auto-regressivo fracionário integrado de média móveis, ou ARFIMA( $p, d, q$ ) com  $d \in (-\frac{1}{2}, \frac{1}{2})$ , se  $\{X_t\}$  for estacionário e satisfizer a equação

$$\phi(B)(1 - B)^d X_t = \theta(B)a_t, \quad (9.14)$$

onde  $a_t \sim RB(0, \sigma_a^2)$  e  $\phi(B)$  e  $\theta(B)$  são polinômios em  $B$  de graus  $p$  e  $q$ , respectivamente.

A razão da escolha dessa família de processos, para fins de modelagem das séries com comportamento de memória longa, é que o efeito do parâmetro  $d$  em observações distantes decai hiperbolicamente conforme a distância aumenta, enquanto os efeitos dos parâmetros  $\phi$  e  $\theta$  decaem exponencialmente. Então,  $d$  deve ser escolhido com o objetivo de explicar a estrutura de correlação de ordens altas da série, enquanto os parâmetros  $\phi$  e  $\theta$  explicam a estrutura de correlação de ordens baixas.

## A) Estacionariedade e Invertibilidade

Hosking (1981) demonstra que o processo ARFIMA( $p, d, q$ ), dado por (9.14) é:

- (i) estacionário se  $d < \frac{1}{2}$  e todas as raízes de  $\phi(B) = 0$  estiverem fora do círculo unitário;
- (ii) invertível se  $d > -\frac{1}{2}$  e todas as raízes de  $\theta(B) = 0$  estiverem fora do círculo unitário.

## B) Funções de auto-correlação e densidade espectral

Hosking (1981) também mostra que se  $X_t$ , dado por (9.14), for estacionário e invertível e se  $f(\lambda)$  for a função densidade espectral de  $X_t$ , então

- (i)  $\lim_{\lambda \rightarrow 0} \lambda^{2d} f(\lambda)$  existe e é finito;
- (ii)  $\lim_{k \rightarrow \infty} k^{1-2d} \rho_k$  existe e é finito.

**Exemplo 9.5.** O caso mais simples é o *ruído branco fracionário*, ou seja, um ARFIMA( $0, d, 0$ ), dado por

$$(1 - B)^d X_t = a_t, \quad a_t \sim \text{RB}(0, \sigma_a^2). \quad (9.15)$$

Se  $a_t$  é gaussiano, teremos o *ruído gaussiano fracionário*.

Quando  $d < \frac{1}{2}$ ,  $X_t$  é um processo estacionário e tem representação na forma  $X_t = \psi(B)a_t$  com os pesos dados por

$$\psi_k = \frac{d(1+d) \cdots (k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!}.$$

Como  $\Gamma(d+k) = d(d+1) \cdots (d+k-1)/\Gamma(d)$ , podemos escrever

$$\psi_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)},$$

e temos

$$\psi_k \sim \frac{k^{d-1}}{(d-1)!} = c_1 k^{d-1}, k \rightarrow \infty$$

sendo  $c_1$  uma constante.

Quando  $d > -\frac{1}{2}$  o processo é invertível e tem representação na forma  $\pi(B)X_t = a_t$  com os pesos dados por

$$\pi_k = \frac{-d(1-d) \cdots (k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!},$$

e como  $\Gamma(k-d) = (k-d-1) \cdots (1-d)(-d)\Gamma(-d)$ , podemos também escrever

$$\pi_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}$$

e

$$\pi_k \sim \frac{k^{-d-1}}{(-d-1)!} = c_2 k^{-d-1}, k \rightarrow \infty,$$

$c_2$  constante. A seguir, assumiremos  $-\frac{1}{2} < d < \frac{1}{2}$ .

As funções de densidade espectral, auto-correlação, auto-correlação parcial e a variância são dadas, respectivamente, por

$$f(\lambda) = \begin{cases} \frac{\sigma_a^2}{2\pi} \left(2 \operatorname{sen} \left(\frac{\lambda}{2}\right)\right)^{-2d}, & 0 < \lambda \leq \pi, \\ \lambda^{-2d}, & \lambda \rightarrow 0, \end{cases} \quad (9.16)$$

$$\rho_h = \frac{(-d)!(h+d-1)!}{(d-1)!(h-d)!} = \prod_{0 < k \leq h} \frac{k-1+d}{k-d}, \quad h=1, 2, \dots \quad (9.17)$$

$$\phi_{hh} = \frac{d}{h-d}, \quad h=1, 2, \dots$$

$$\gamma_0 = \frac{(-2d)!}{(-d)!^2}.$$

Em particular, temos que

$$\rho_1 = \frac{d}{1-d}, \quad (9.18)$$

$$\rho_h \sim \frac{(-d)!h^{2d-1}}{(d-1)!} = c_3 h^{2d-1}, \quad h \rightarrow \infty,$$

sendo  $c_3$  constante e

$$f(\lambda) \sim \lambda^{-2d}. \quad (9.19)$$

A Figura 9.5(a) apresenta  $N = 100$  observações simuladas de um modelo ARFIMA(0,  $d$ , 0) com  $d = 0,45$  e a Figura 9.6 (a) apresenta o gráfico das auto-correlações.

**Exemplo 9.6.** Consideremos, agora, o processo ARFIMA(1,  $d$ , 0), dado por

$$(1 - B)^d(1 - \phi B)X_t = a_t,$$

que é um processo estacionário e invertível se  $|d| < \frac{1}{2}$  e  $|\phi| < 1$ .

Além disso, temos que

- (a) os pesos  $\psi_j$  e  $\pi_j$  das representações  $X_t = \psi(B)a_t$  e  $\pi(B)X_t = a_t$  são dados por

$$\psi_j = \frac{(j+d-1)!}{j!(d-1)!} F(1, -j; 1-d-j, \phi) \sim \frac{j^{d-1}}{(1-\phi)(d-1)!}$$

e

$$\pi_j = \frac{(j-d-2)!}{(j-1)!(-d-1)!} \{1 - \phi - (1+d)/j\} \sim \frac{(1-\phi)}{(-d-1)!} j^{-d-1},$$

respectivamente, em que  $F(a, b; c, z) = 1 + \frac{ab}{c}z + \frac{a(a+1)b(b+1)}{c(c+1)}z^2 + \dots$  é a *função hipergeométrica* e a aproximação vale para  $j \rightarrow \infty$ ;

- (b) a função densidade espectral é

$$f(\lambda) = \begin{cases} \frac{\sigma_a^2}{2\pi} \frac{(2\text{sen}(\frac{\lambda}{2}))^{-2d}}{1+\phi^2-2\phi\cos\lambda}, & 0 < \lambda \leq \pi, \\ \frac{\lambda^{-2d}}{(1-\phi)^2}, & \lambda \rightarrow 0; \end{cases}$$

- (c) a expressão para a f.a.c. é bastante complicada mas, em particular, temos que

$$\rho_1 = \frac{(1 + \phi^2)F(1, d; 1 - d; \phi) - 1}{\phi[2F(1, d; 1 - d; \phi) - 1]}$$

e

$$\rho_j = \frac{(-d)!(1 + \phi)j^{2d-1}}{(d-1)!(1 - \phi)^2 F(1, 1 + d; 1 - d; \phi)}, \quad j \rightarrow \infty.$$

Além disso,

$$\gamma_0 = \frac{(-2d)!F(1, 1 + d; 1 - d; \phi)}{(1 + \phi)[(-d)!]^2}.$$

A Figura 9.5(b) apresenta  $N = 100$  observações simuladas de um processo ARFIMA(1,  $d$ , 0) com  $\phi = 0,8$  e  $d = 0,45$  e a Figura 9.6 (b) apresenta o gráfico das f.a.c.

**Exemplo 9.7.** Considere, agora, um processo ARFIMA(0,  $d$ , 1), dado por

$$(1 - B)^d X_t = (1 - \theta B)a_t,$$

que pode ser visto como uma média móvel de primeira ordem de um ruído branco fracionário;  $X_t$  é estacionário e invertível se  $|\theta| < 1$  e  $|d| < \frac{1}{2}$ . Além disso, temos que:

- (a) os pesos  $\psi_j$  e  $\pi_j$  das representações auto-regressiva e de médias móveis infinitas são dadas por

$$\psi_j = \frac{(j - d - 1)!}{j!(-d - 1)!} F(1, -j; 1 + d - j, \theta) \sim \frac{j^{-d-1}}{(1 - \theta)(-d - 1)!}$$

e

$$\pi_j = \frac{(j + d - 2)!}{(j - 1)(d - 1)!} \left[ 1 - \theta - \frac{(1 + d)}{j} \right] \sim \frac{(1 - \theta)}{(d - 1)!} j^{d-1},$$

respectivamente, em que  $F(\cdot)$  é a função hipergeométrica dada no Exemplo 9.6 e a aproximação vale para  $j \rightarrow \infty$ ;

- (b)  $f(\lambda) = \frac{\sigma_a^2}{2\pi} [1 + \theta^2 - 2\theta \cos \lambda] \left[ 2 \operatorname{sen} \left( \frac{\lambda}{2} \right) \right]^{-2d} \sim [(1 - \theta)^2 \lambda^{-2d}]$  quando  $\lambda \rightarrow 0$ ;

- (c) a expressão para a f.a.c. é bastante complicada mas, em particular, temos que

$$\rho_1 = \frac{(1 + \theta^2)d(2 - d) - 2\theta(1 - d + d^2)}{(1 - d)(2 - d)\{1 + \theta^2 - 2\theta d/(1 - d)\}}$$

e

$$\rho_j = \frac{(-d)!}{(d-1)!} a j^{2d-1}, \quad j \rightarrow \infty,$$

em que  $a = \frac{(1-\theta^2)}{(1+\theta^2-2\theta d/(1-d))}$ .

A Figura 9.5 (c) apresenta  $N = 100$  observações de um processo ARFIMA  $(0, d, 1)$  com  $d = 0,45$  e  $\theta = 0,3$  e a Figura 9.6 (c) apresenta o respectivo gráfico das f.a.c.

**Exemplo 9.8.** Finalmente, a Figura 9.5 (d) apresenta  $N = 100$  observações simuladas de um processo ARFIMA  $(1, d, 1)$  com  $\phi = 0,8$ ,  $\theta = 0,3$  e a Figura 9.6 (d) apresenta o gráficos das auto-correlações.

Em todos os exemplos citados do processo ARFIMA  $(p, d, q)$  podemos notar o comportamento da função de auto-correlação, que tem decaimento hiperbólico.

Para mais detalhes, veja Hosking (1981) e Granger e Joyeux (1980).



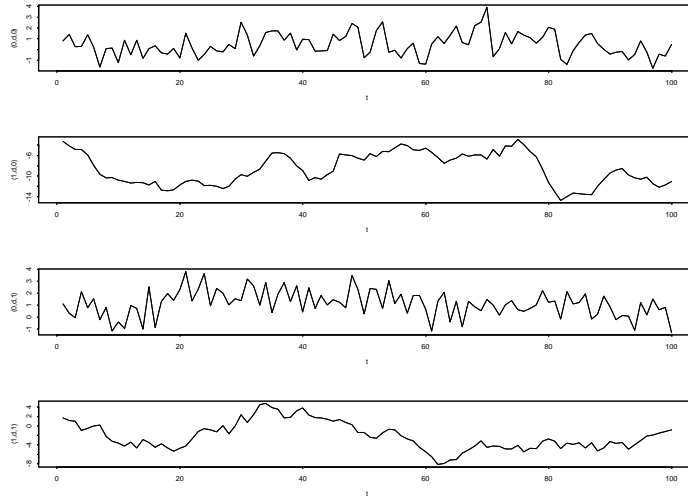


Figura 9.5: Séries ARFIMA( $p, d, q$ ) geradas de acordo com os Exemplos 9.5 a 9.8.

## 9.4 Estimação de modelos ARFIMA

Nesta seção vamos apresentar o método de máxima verossimilhança para obter estimadores dos parâmetros do modelo (9.14). Um outro método que pode ser utilizado é a estimação semi-paramétrica no domínio da frequência, por meio do procedimento GPH visto anteriormente. Para detalhes, veja Morettin e Toloi (2006).

A função de verossimilhança de  $\mathbf{X} = (X_1, \dots, X_T)$  proveniente de um processo ARFIMA( $p, d, q$ ) pode ser expressa na forma

$$L(\boldsymbol{\eta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-T/2} (r_0 \cdots r_{T-1})^{-1/2} \exp \left[ -\frac{1}{2\sigma_a^2} \sum_{j=1}^T (X_j - \hat{X}_j)^2 / r_{j-1} \right], \quad (9.20)$$

em que  $\boldsymbol{\eta} = (d, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ ,  $\hat{X}_j$ ,  $j = 1, \dots, T$ , são as previsões um passo à frente e  $r_{j-1} = (\sigma_a^2)^{-1} E(X_j - \hat{X}_j)^2$ .

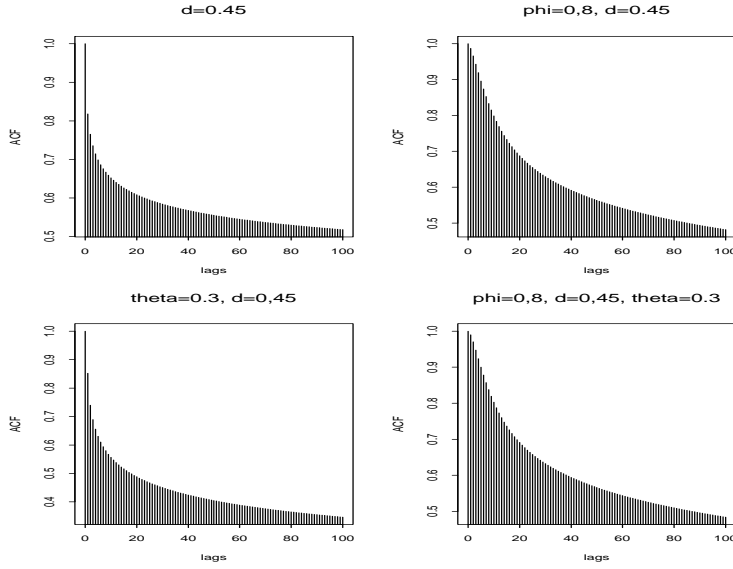


Figura 9.6: F.a.c. das séries ARFIMA( $p, d, q$ ) geradas de acordo com os Exemplos 9.5 a 9.8.

Os estimadores de máxima verossimilhança dos parâmetros são dados por

$$\hat{\sigma}_{MV}^2 = T^{-1} S(\hat{\boldsymbol{\eta}}_{MV}), \quad (9.21)$$

onde

$$S(\hat{\boldsymbol{\eta}}_{MV}) = \sum_{j=1}^T (X_j - \hat{X}_j)^2 / r_{j-1}$$

e  $\hat{\boldsymbol{\eta}}_{MV}$  é o valor de  $\boldsymbol{\eta}$  que minimiza

$$\ell(\boldsymbol{\eta}) = \ln(S(\boldsymbol{\eta})) + T^{-1} \sum_{j=1}^T \ln r_{j-1}.$$

Entretanto, o cálculo de  $\ell(\boldsymbol{\eta})$  é bastante lento. Um procedimento alternativo é considerar uma aproximação para  $\ell(\boldsymbol{\eta})$  dada por

$$\ell(\boldsymbol{\eta}) \simeq \ell_*(\boldsymbol{\eta}) = \ln \frac{1}{T} \sum_j \frac{I_T(w_j)}{2\pi f(w_j; \boldsymbol{\eta})}, \quad (9.22)$$

em que

$$I_T(w_j) = \frac{1}{T} \left| \sum_{t=1}^T X_t e^{-itw_j} \right|^2$$

é o periodograma dos dados,

$$f(w_j; \boldsymbol{\eta}) = \frac{\sigma_a^2}{2\pi} \frac{|1 - \theta_1 e^{-iw_j} - \dots - \theta_q e^{-qiw_j}|^2}{|1 - \phi_1 e^{-iw_j} - \dots - \phi_p e^{-piw_j}|^2} \cdot |1 - e^{-iw_j}|^{-2}$$

é a função densidade espectral do processo  $X_t$  e  $\sum_j$  é a soma sobre todas as frequências de Fourier,  $w_j = 2\pi j/T \in (-\pi, \pi]$ ,  $j = 0, 1, \dots, [T/2]$ .

Hannan (1973) e Fox e Taqqu (1986) mostram que:

- (i) o estimador  $\hat{\boldsymbol{\eta}}_{MV}$  que minimiza (9.22) é consistente;
- (ii) se  $d > 0$ ,

$$\hat{\boldsymbol{\eta}}_{MV} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\eta}, T^{-1} A^{-1}(\boldsymbol{\eta})), \quad (9.23)$$

em que  $A(\boldsymbol{\eta})$  é uma matriz de ordem  $(p + q + 1) \times (p + q + 1)$  com  $(j, k)$ -ésimo elemento dado por

$$A_{jk}(\boldsymbol{\eta}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial \ln f(\lambda; \boldsymbol{\eta})}{\partial \eta_j} \frac{\partial \ln f(\lambda; \boldsymbol{\eta})}{\partial \eta_k} d\lambda;$$

- (iii) a variância  $\sigma_a^2$  é estimada por

$$\hat{\sigma}_{MV}^2 = \frac{1}{T} \sum_j \frac{I_T(w_j)}{2\pi f(w_j; \hat{\boldsymbol{\eta}}_{MV})}.$$

O estimador de  $d$  obtido desta maneira é chamado *estimador de Whittle* ou *estimador de Fox-Taqqu*. O programa **Splus** utiliza a função `arima.fracdiff` para estimar modelos ARFIMA, incluindo modelos da forma  $(0, d, 0)$ . Contudo, é necessário especificar os valores de  $p$  e  $q$  da parte ARIMA. Como vimos, é difícil especificar estes valores e uma possibilidade é encontrar valores  $p \leq p_{\max}$  e  $q \leq q_{\max}$  que minimizam o AIC ou BIC.

O programa **S+FinMetrics** utiliza a função **FARIMA**, baseada numa extensão de modelos ARFIMA proposta por Beran (1995), que supõe  $d > -1/2$ . Veja Zivot e Wang (2003) para detalhes. A função `d.whittle` também pode ser usada para o caso ARFIMA  $(0, d, 0)$ .

**Exemplo 9.9.** Consideremos a série de valores absolutos dos retornos da Petrobrás, de 2/1/95 a 27/12/2000. Usando a função **FARIMA** do **S+FinMetrics**, vamos ajustar um modelo ARFIMA  $(p, d, q)$ . Para determinar as ordens  $p$  e  $q$ , consideremos  $p = 0, 1, 2$  e  $q = 0, 1, 2$  e escolhemos o modelo que minimiza o BIC. O modelo escolhido é um ARIMA  $(2, d, 1)$ , com  $\hat{d} = 0,346$ , indicando a presença de ML. O diagnóstico do modelo pode ser feito com o comando `plot(petro.fit)`, que produz diversos gráficos que ajudam a verificar se o modelo é adequado. O Quadro 9.3 mostra a saída do programa.

Segue-se que o modelo ajustado é

$$(1 + 1,087B + 0,171B^2)(1 - B)^{0,346}X_t = (1 + 0,935)a_t.$$

Os comandos apropriados são, indicando os retornos por `rpetro`:

```
>petro.fit=FARIMA(abs(rpetro), p.range=c(0,2), q.range=c(0,2),mmax=0),
>summary(petro.fit).
```

## 9.5 Previsão de modelos ARFIMA

Considere o processo ARFIMA  $(p, d, q)$  estacionário e invertível,

$$\phi(B)(1 - B)^d X_t = \theta_0 + \theta(B)a_t, \quad -0,5 < d < 0,5. \quad (9.24)$$

Podemos reescrever o processo na forma de choques aleatórios,

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad (9.25)$$

---

```
FARIMA(x = abs(rpetro), p.range = c(0, 2), q.range = c(0, 2), mmax = 0)
```

Coefficients:

	Value	Std. Error	t value	Pr(> t )
d	0.3461	0.0630	5.4907	0.0000
AR(1)	-1.0869	0.0489	-22.2462	0.0000
AR(2)	-0.1710	0.0475	-3.5984	0.0003
MA(1)	-0.9345	0.0210	-44.4532	0.0000

Information Criteria:

log-likelihood	BIC
3560.959	-7092.676

Residual scale estimate: 0.0224

	total	residual
Degree of freedom:	1496	1491

BIC of all models estimated:

	q=0	q=1	q=2
p=0	-7080.255	-7089.267	-7081.146
p=1	-7085.676	-7078.541	-7080.092
p=2	-7085.192	-7092.676	-7077.863

---

Quadro 9.3: Estimação do modelo ARFIMA(2,d,1).

e na forma invertida

$$\sum_{j=0}^{\infty} \pi_j Z_{t-j} = \theta_0 + a_t, \quad (9.26)$$

onde

$$\sum_{j=0}^{\infty} \psi_j B^j = \theta(B) \phi^{-1}(B) (1 - B)^{-d}$$

e

$$\sum_{j=0}^{\infty} \pi_j B^j = \phi(B) \theta^{-1}(B) (1 - B)^d.$$

Assim, podemos fazer previsões de valores futuros do processo  $X_t$ , utilizando as equações (9.25) ou (9.26). A variância do erro de previsão, também pode ser calculada de modo usual.

Uma outra forma é usar a da equação de diferenças

$$\varphi(B)X_t = \theta_0 + \theta(B)a_t, \quad (9.27)$$

em que  $\varphi(B) = \phi(B)(1 - B)^d = \phi(B)D(B)$  e  $D(B) = 1 - d_1B - d_2B^2 - \dots$  é um polinômio em  $B$ , com coeficientes dados por

$$d_j = \frac{-\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0 < k \leq j} \frac{k-1-d}{k}, \quad j = 0, 1, 2, \dots \quad (9.28)$$

e  $\Gamma(\cdot)$  é a função gama, dada por

$$\Gamma(x) = \begin{cases} \int_0^{\infty} t^{x-1} e^{-t} dt, & x > 0, \\ \infty, & x = 0, \\ x^{-1}\Gamma(1+x), & x < 0. \end{cases}$$

Utilizando (9.28) e as expressões (9.25) e (9.26), podemos fazer previsões para a série de memória longa  $X_t$ .

Note que  $D(B)$  é um polinômio de ordem infinita. Na prática, quando temos uma série com  $T$  observações, utilizamos somente os  $L$  primeiros termos desse polinômio,  $L < T$ .

Para mais detalhes, ver Brockwell e Davis (1991). O programa S+FinMetrics usa a função `predict` para obter previsões de um modelo ARFIMA.

**Exemplo 9.10.** Para o exemplo 9.9, onde `petro.fit` indica o nome do modelo ARFIMA ajustado, usamos os comandos:

```
> petro.pred=predict(petro.fit,n.predict=100, ar.approx=100)
> class(petro.pred)
> summary(petro.pred)
> plot(petro.pred,abs(rpetro),n.old=300)
```

Aqui, `n.predict` indica o número de passos a frente, `ar.approx` indica a ordem do modelo AR usado e `n.old` é o número de observações da série original usadas no gráfico. Os primeiros 32 valores previstos, constantes do objeto `petro.pred`, estão mostrados no Quadro 9.4.

---

Predicted Values:

[1]	0.0823	-0.0370	0.0742	-0.0287	0.0629	-0.0199	0.0542	-0.0121
[9]	0.0471	-0.0061	0.0415	-0.0012	0.0371	0.0029	0.0336	0.0062
[17]	0.0308	0.0087	0.0285	0.0107	0.0268	0.0126	0.0254	0.0139
[25]	0.0242	0.0151	0.0233	0.0160	0.0226	0.0167	0.0222	0.0174

---

Quadro 9.4: Valores previstos pelo modelo ARFIMA.

## 9.6 Processos de Volatilidade com ML

Vimos que os modelos da família GARCH modelam a dinâmica de retornos juntamente com sua volatilidade, considerada uma variável latente (não-observada).

Por outro lado, os modelos ARFIMA podem ser usados para modelar o comportamento de ML da volatilidade, supondo-se que esta seja uma variável observada. Uma idéia é considerar  $|X_t|$  ou  $X_t^2$ , se  $X_t$  indicar os log-retornos. Veja Taylor (1986), Ding et al. (1993) e Dacorogna et al. (1993), para estudos de  $|X_t|$  (ações),  $|X_t|^p$  (ações) e  $X_t^2$  (taxas de câmbio), respectivamente. Ou podemos usar a chamada volatilidade de Garman-Klass, baseada nos preços de abertura, máximo, mínimo e de fechamento de determinado ativo (ação). Veja o Apêndice 9.

Um modelo GARCH (r,s) pode ser escrito como um modelo ARMA (q,s), com  $q = \max(r, s)$ , na forma (veja (5.27))

$$\phi(B)X_t^2 = \alpha_0 + \theta(b)\nu_t, \quad (9.29)$$

onde  $\nu_t = X_t^2 - \sigma_t^2$ , são as inovações seguindo um processo diferença martingale e

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_q B^q,$$

$$\theta(B) = 1 - \beta_1 B - \dots - \beta_s B^s,$$

com  $\phi_i = \alpha_i + \beta_i$ ,  $i = 1, \dots, q$ .

Baillie et al. (1996) consideraram um processo de ML na variância condicional, denominado “Garch integrado fracionário”, ou FIGARCH, que implica em decaimento hiperbólico nas inovações quadráticas e pesos persistentes.

O modelo FIGARCH (q,d,s) é definido por

$$\phi(B)(1-B)^d X_t^2 = \alpha_0 + \theta(B)\nu_t, \quad (9.30)$$

onde  $\phi(B)$  e  $\theta(B)$  são como definidos acima e com raízes fora do círculo unitário. Se  $d = 0$  obtemos um modelo GARCH e se  $d = 1$  obtemos um modelo IGARCH.

No Capítulo 5 estudamos alguns modelos assimétricos, em particular, o modelo EGARCH. Notando que este pode ser escrito como um modelo ARMA em termos da  $\ln \sigma_t^2$  (e portanto garante que  $\sigma_t^2 > 0$ ), Bollerslev e Mikkelesen (1996) propuseram o modelo EGARCH integrado fracionário, FIEGARCH.

Um modelo FIEGARCH(q,d,s) é definido por

$$\phi(B)(1-B)^d \ln \sigma_t^2 = \alpha_0 + \sum_{j=1}^s (\beta_j |\varepsilon_{t-j}| + \gamma_j \varepsilon_{t-j}), \quad (9.31)$$

onde  $\varepsilon_t$  são os resíduos padronizados, ou seja,  $\varepsilon_t = X_t/\sigma_t$ . Em (9.31), os parâmetros  $\gamma_j$  dão o efeito de alavancagem (“leverage”). O processo é estacionário se  $0 < d < 1$ .

O programa S+FinMetrics usa a função `fgarch` para estimar modelos FIGARCH e FIEGARCH, mantendo todas as demais facilidades dos modelos GARCH.

**Exemplo 9.11.** Vamos considerar os retornos diários (preços de fechamento) do DJIA, do exemplo 1.2, mas agora multiplicados por 100. Ajustamos modelos FIGARCH(1,d,1) e FIEGARCH(1,d,1), com resultados nos quadros 9.5 e



9.6, respectivamente. Como os retornos são não-correlacionados, não há necessidade de ajustar previamente um modelo ARMA. Na Tabela 9.1 temos a comparação entre os modelos, mostrando que o modelo FIEGARCH tem os menores AIC e BIC. O valor estimado de  $d$  é  $\hat{d} = 0,585$ , para o modelo FIEGARCH, o que indica que o modelo é estacionário.

O diagnóstico e previsão desses modelos podem ser feitos de modo similar aos modelos ARFIMA e GARCH. Veja Zivot e Wang (2003) para detalhes.

Tabela 9.1: Comparação entre os modelos FIGARCH e FIEGARCH ajustados aos retornos do DJIA.

	FIGARCH (1,1)	FIEGARCH (1,1)
AIC	5853	5769
BIC	5881	5803
log-verossim.	-2921	-2879

## 9.7 Problemas

1. Mostre que um processo ARFIMA  $(0, d, 0)$  com  $-0,5 < d < 0,5$  tem representação nas formas  $X_t = \psi(B)a_t$  e  $\pi(B)X_t = a_t$ , com pesos dados por

$$\psi_k = \frac{(k+d-1)!}{k!(d-1)!} \text{ e } \pi_k = \frac{(k-d-1)!}{k!(-d-1)!},$$

respectivamente.

2. Considere um processo ARFIMA  $(0, d, 0)$ , com  $d = 0,4$ . Represente graficamente os pesos  $\psi_k$  e  $\pi_k$ ,  $k=1,2,3,\dots$ , encontrados usando o problema anterior.
3. Considere a série de valores absolutos dos retornos diários do Banespa (arquivo d-ban95.00.dat).

```

-----
fgarch(formula.mean = mdow ~ 1, formula.var = ~ figarch(1, 1))
Mean Equation: mdow ~ 1
Conditional Variance Equation: ~ figarch(1, 1)
-----
Estimated Coefficients:
-----
              Value Std.Error      t value   Pr(>|t|)
      C  0.07564  2.202e-002   3.435e+000  3.029e-004
      A  0.01926  3.520e-003   5.471e+000  2.515e-008
GARCH(1)  0.87490  1.148e-002   7.623e+001  0.000e+000
  ARCH(1) -0.05828  2.421e-002  -2.407e+000  8.082e-003
fraction  1.00000  2.906e-009   3.442e+008  0.000e+000
-----
AIC(5) = 5852.895
BIC(5) = 5880.877

Normality Test:
-----
Jarque-Bera P-value Shapiro-Wilk   P-value
      422.8         0         0.9804 7.759e-011

Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
      21.57 0.04261         12

Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
      11.73 0.4675         12

Lagrange multiplier test:
-----
Lag 1   Lag 2 Lag 3   Lag 4   Lag 5 Lag 6 Lag 7   Lag 8   Lag 9
1.825 -0.07819 -1.41 -0.7683 -0.4089 1.315 -0.29 -0.1051 -0.4721

Lag 10 Lag 11 Lag 12   C
0.1306 0.1221 0.07977 1.81

TR^2 P-value F-stat P-value
11.52 0.4847 1.054 0.5084
-----

```

Quadro 9.5: Modelo FIGARCH(1,1) ajustado aos retornos do DJIA.

```

-----
fgarch(formula.mean = mdow ~ 1, formula.var = ~ fiegarch(1, 1),
leverage = T)
Mean Equation: mdow ~ 1
Conditional Variance Equation: ~ fiegarch(1, 1)
-----
Estimated Coefficients:
-----
              Value Std.Error t value    Pr(>|t|)
      C  0.06021    0.02043   2.948 1.618e-003
      A -0.14163    0.02190  -6.466 6.321e-011
GARCH(1) 0.39300    0.11183   3.514 2.255e-004
  ARCH(1) 0.18322    0.02767   6.623 2.265e-011
    LEV(1) -0.15623    0.01841  -8.487 0.000e+000
fraction 0.58496    0.04236  13.808 0.000e+000
-----
AIC(6) = 5769.173
BIC(6) = 5802.752

Normality Test:
-----
      Jarque-Bera P-value Shapiro-Wilk    P-value
      115.2      0      0.9828 0.00001087

Ljung-Box test for standardized residuals:
-----
Statistic P-value Chi^2-d.f.
      26.43 0.009338      12

Ljung-Box test for squared standardized residuals:
-----
Statistic P-value Chi^2-d.f.
      17.61 0.1279      12

Lagrange multiplier test:
-----
      Lag 1 Lag 2 Lag 3 Lag 4 Lag 5 Lag 6 Lag 7 Lag 8 Lag 9
0.02189 -0.593 -1.571 -0.2955 0.0652 2.703 0.1065 1.719 -0.1345

      Lag 10 Lag 11 Lag 12      C
      1.025 1.287 0.08342 -1.588

      TR^2 P-value F-stat P-value
      18 0.1157 1.651 0.1745
-----

```

Quadro 9.6: Modelo FIEGARCH(1,1) ajustado aos retornos do DJIA.

- (a) Verifique se existe necessidade de uma transformação na série a fim de estabilizar sua variância.
  - (b) Verifique a necessidade de tomar diferenças com o objetivo de tornar a série estacionária.
  - (c) Após as transformações adequadas, identifique um modelo ARFIMA  $(0, d, 0)$  estimando  $d$  através do método de regressão.
4. Considere um processo ARFIMA  $(0, d, 0)$  estacionário e invertível. Mostre que a função densidade espectral é dada por (9.16).
- Sugestão: Utilize as propriedades de filtragem linear.
5. Deduza a expressão da função densidade espectral de um processo ARFIMA  $(0, d, 1)$  estacionário e invertível.
6. Mostre que a função densidade espectral de um processo ARFIMA  $(1, d, 0)$  estacionário e invertível é dada por

$$f(\lambda) = \frac{[2\sin(0,5\lambda)]^{-2d}}{1 + \phi^2 - 2\cos\lambda}$$

e que

$$f(\lambda) \approx \frac{\lambda^{-2d}}{(1 - \phi)^2}, \quad \lambda \rightarrow 0.$$

7. Simule  $N = 1000$  observações  $\{X_t\}$  de um processo ARFIMA  $(1, d, 0)$ , com  $\phi = 0,6$  e  $d = 0,45$ .
- (a) Faça um gráfico dos dados simulados e comente.
  - (b) Calcule as f.a.c. e f.a.c.p. amostrais e comente.
  - (c) Estime os parâmetros do modelo, testando a significância de cada um deles.
  - (d) Ajuste um modelo ARMA a  $Y_t = (1 - B)X_t$ .
  - (e) Compare o ajustamento dos modelos ARFIMA (item (c)) e ARMA (item (d)).

8. Considere a série  $r_t$  de retornos diários (fechamento) do DJIA (arquivo d-dow95.02.dat). Considere a série de retornos ao quadrado,  $v_t = r_t^2$ , como representando a série de volatilidades.
  - (a) Calcule as funções de auto-correlação e auto-correlação parcial amostrais da série  $v_t$  e comente.
  - (b) Identifique um modelo de memória longa para a série  $v_t$ , utilizando o método de regressão.
9. Utilize o modelo ajustado no Problema 8 para fazer previsões, a partir da última observação, considerando  $h = 1, 2, \dots, 12$ .
10. Estime um modelo de memória longa ajustado à série de valores absolutos dos retornos diários da Cemig (arquivo d-cemig95.00.dat). Verifique se o modelo é adequado e, caso necessário, faça as modificações apropriadas para torná-lo adequado. Utilize o modelo final para fazer previsões até 12 passos à frente. Faça um gráfico da série original e das previsões obtidas.
11. Mostre que o modelo FIGARCH dado em (9.30) pode ser escrito na forma

$$\sigma_t^2 = \alpha_0(1 - \theta(1))^{-1} + \lambda(B)X_t^2,$$

onde  $\lambda(B) = 1 - [1 - \theta(B)]^{-1}\phi(B)(1 - B)^d$ .

12. Ajuste modelos FIGARCH e FIEGARCH aos retornos do Ibovespa (arquivo d-ibv95.00.dat) e do Banespa (arquivo d-ban95.00.dat). Se necessário, tome retornos em porcentagem.

## Apêndice 9. Volatilidade de Garman-Klass

Como vimos no Capítulo 5, a volatilidade é uma variável não-observada, e uma maneira de estimá-la é ajustar algum modelo de volatilidade, como ARCH, MVE, etc. Se tivermos dados de alta frequência (intra-diários) é

possível estimar a volatilidade diária por meio da chamada *volatilidade realizada*. Por exemplo, se tivermos os preços de uma ação, negociada na Bolsa, a cada 5 minutos no dia  $t$ , e se a Bolsa opera 6 horas neste dia, teremos 72 observações,  $P_{t,1}, \dots, P_{t,72}$ . Considerando os log-retornos, teremos  $r_{t,1}, \dots, r_{t,72}$ , que podem ser usados para obter  $v_t = r_{t,0}^2 + \sum_{j=1}^{72} r_{t,j}^2$ , onde  $r_{t,0}$  é o retorno *overnight*, obtido da última cotação do dia  $t-1$  e primeira do dia  $t$ .

Como muitos ativos não têm suficiente liquidez, não é possível obter dados intra-diários regularmente, mas usualmente temos disponíveis os preços de abertura (A), fechamento (F), máximo (M) e mínimo (m). A volatilidade diária pode, então, ser estimada segundo fórmula proposta por Garman e Klass (1980), dada por

$$\hat{\sigma}_{t,GK}^2 = a \frac{(A_t - F_{t-1})}{f} + (1-a) \frac{\hat{\sigma}_{t,GK^*}^2}{1-f}, \quad (A.1),$$

onde:

$A_t$ : log do preço de abertura do dia  $t$ ;

$F_t$ : log do preço de fechamento do dia  $t$ ;

$a$ : constante, com valor 0,12, que minimiza a variância

$$\hat{\sigma}_{t,GK^*}^2 = 0,511(u_t - d_t)^2 - 0,019[F_t(u_t - d_t) - 2u_t d_t] - 0,383c_t^2,$$

onde:  $u_t = M_t - A_t$ ,  $d_t = m_t - A_t$ ,  $c_t = F_t - A_t$ .

Em (A.1),  $f$  é a fração do dia em que o mercado está fechado; o SPlus utiliza  $f = 0.192$ . O programa S+FinMetrics implementa o estimador (A.1) por meio da função `TA.garmanKlass`.

**Exemplo 9.11.** Consideremos a série de índices diários do DJIA, do exemplo 1.2, com  $T = 1992$ . Na Figura A.1 temos a série de log-volatilidades de GK e a respectiva f.a.c., mostrando o seu caráter de ML. Inicialmente, vamos estimar o valor de  $d$ , usando um modelo ARFIMA (0,d,0) para a log-volatilidade de GK. Usando a função `FARIMA` do S+FinMetrics obtemos  $\hat{d} = 0,4365$ . A seguir, usando o BIC como critério de escolha do modelo, ajustamos um modelo ARFIMA (p,d,q) e obtemos  $\hat{d} = 0,4868$ , com  $p = 0, q = 1$ , ou seja, um modelo ARFIMA (0,d,1), sendo  $\hat{\theta} = 0,1293$ . O quadro A.1 mostra este ajuste.

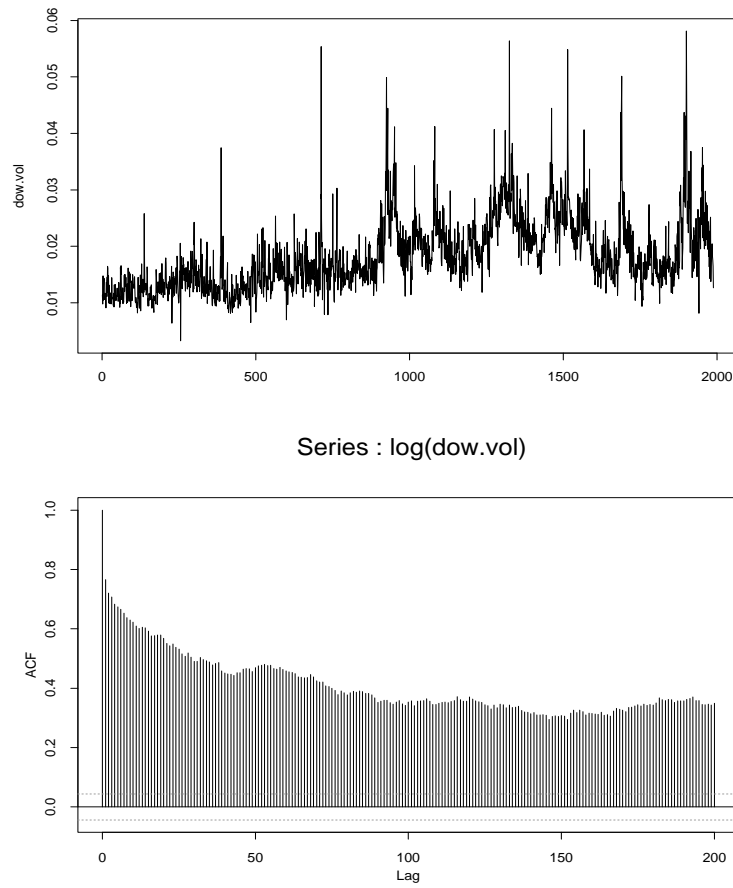


Figura A.1: Volatilidade de GK e f.a.c.

-----  
Call:

```
FARIMA(x = log(dow.vol), p.range = c(0, 2), q.range = c(0, 2), mmax = 0
)
```

Coefficients:

	Value	Std. Error	t value	Pr(> t )
d	0.4868	0.0257	18.9395	0.0000
MA(1)	0.1293	0.0325	3.9746	0.0001

Information Criteria:

log-likelihood	BIC
511.2452	-1007.2997

Residual scale estimate: 0.1869

	total residual
Degree of freedom:	1989      1986

BIC of all models estimated:

	q=0	q=1	q=2
p=0	-996.1286	-1007.2997	-1002.1456
p=1	-1003.1272	-1006.0611	-1002.4372
p=2	-985.8263	-1000.5414	-994.1989

-----

Quadro A.1: Ajuste do modelo ARFIMA (0,d,1).



---

# Referências

Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, **60**, 255–265.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Akaike, H. (1977). On entropy maximization principle. In *Applications of Statistics* (P.R. Krishnaiah, ed.), 27–41. Amsterdam: North-Holland.

Alexander, C. (2001). *Market Models*. Wiley.

Andersson, F., Mausser, H., Rosen, D. and Uryasev, S. (2001). Credit risk optimization with conditional value-at-risk criterion. *Mathematical Programming*, Series B, **89**, 273–291.

Aparicio, F.M. and Estrada, J. (2001). Empirical distributions of stock returns: European securities markets, 1990-1995. *European Journal of Finance*, **7**, 1–21.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1997). Thinking coherently. *Risk*, **10**, 68–71.

Artzner, P., Delbaen, F., Eber, J.M. and Heath, D. (1999). Coherent measures

of risk. *Mathematical Finance*, **9**, 203–228.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5–59.

Banerjee, A., Dolado, J.J., Galbraith, J.W. and Hendry, D.F. (1993). *Cointegration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Bera, A.K. and Jarque, C.M. (1981). An efficient large sample test for normality of observations and regression residuals. Working paper in Econometrics N0. 40, Australian National University, Canberra.

Beran, J. (1995) Maximum likelihood estimation of the differencing parameter for invertible short and long memory ARIMA models. *Journal of the Royal Statistical Society, Series B*, **57**, 659–672.

Berkowitz, J. and O'Brien, J. (2002). How accurate are the Value-at-Risk models at commercial banks? *The Journal of Finance*, Vol. LVII, 1093–11.

Beveridge, W.H. (1921). Weather and harvest cycles. *Economics Journal*, **31**, 429–452.

Black, F. (1976). Studies of stock market volatility changes. Proc. 1976 Meeting of the American Statistical Association, Business and Economic Statistics Section, 177–181.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 635–654.

Blattberg, R.C. and Gonedes, N.J. (1974). A comparison of stable and Student distributions as statistical models for stock prices. *The Journal of Business*, **47**, 244–280.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative process and rates of return. *Review of Economics and Statistics*,

69, 542–547.

Bollerslev, T. (1988). On the correlation structure for the generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, **9**, 121–132.

Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, **52**, 5–59.

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). Arch Models. In *Handbook of Econometrics*, Vol. IV (eds. R.F. Engle and D.L. McFadden), 2959–3038. New York: North Holland.

Bollerslev, T. and Mikkelsen, H.O. (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, **73**, 151–184.

Box, G.E.P. and Pierce, D.A. (1970). Distribution of autocorrelations in autoregressive moving average models. *Journal of The American Statistical Association*, **65**, 1509–1526.

Box, G.E.P., Jenkins, G.M. and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*. Third Edition. Englewood Cliffs: Prentice Hall.

Boudjellaba, H., Dufour, J.-M. and Roy, R. (1992). Testing causality between two vectors in multivariate autoregressive moving average models. *Journal of the American Statistical Association*, **87**, 1082–1090.

Breidt, F.J. and Carriquiry, A.L. (1996). Improved quasi-maximum likelihood estimation for stochastic volatility models. In *Modelling and Prediction: Honoring Seymour Geisser* ( J.C. Lee and A. Zellner, eds.), 228–247. New York: Springer.

Breidt, F.J., Crato, N. and de Lima, P.J.F. (1993). Modeling long-memory stochastic volatility. Working paper (Johns Hopkins University, Baltimore, MD).

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*. Second Edition. New York: Springer.

- Campbell, J.Y., Lo, A.W. and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
- Carmona, R.A. (2004). *Statistical Analysis of Financial Data in S-Plus*. New York: Springer.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extremes*. London: Springer.
- Cramér, H. and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes*. New York: Wiley.
- da Cunha, D. M. S. (1997). *Análise de Causalidade em Séries Temporais*. Dissertação de mestrado, IME-USP, São Paulo.
- Dacorogna, M.M., Muller, U.A., Nagler, R.J., Olsen, R.B. and Pictet, O.V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance*, **12**, 413–438.
- Dickey, D.A. and Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**, 427–431.
- Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**, 1052–1072.
- Dickey, D.A. and Pantula, S. (1987). Determining the order of differencing in autoregressive processes. *Journal of Business and Economic Statistics*, **15**, 455–461.
- Ding, Z., Granger, C.W.J. and Engle, R.F. (1993). A long memory property of stock returns and a new model. *Journal of Empirical Finance*, **1**, 83–106.
- Dolado, J.J., Jenkinson, T. and Sosvilla-Rivero, S. (1990). Cointegration and unit roots. *Journal of Economic Surveys*, **4**, 249–273.
- Dunsmuir, W. and Hannan, E.J. (1976). Vector linear time series models. *Advances in Applied Probability*, **8**, 339–364.

- Durbin, J. (1970). An alternative to the bounds test for testing serial correlation in least-squares regression. *Econometrica*, **38**, 422–429.
- Durbin, J. and Koopman, S.J. (1997a). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, **84**, 669–684.
- Durbin, J. and Koopman, S.J. (1997b). Time series analysis of Non-Gaussian observations based on state space models. Preprint, London School of Economics.
- Durbin, J. and Koopman, S.J. (2000). Time series analysis of Non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of The Royal Statistical Society, Series B*, **62**, 3–56.
- Einstein, A. (1905). On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat. *Ann. Physics*, **17**.
- Elliot, G., Rothenberg, T.J. and Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, **64**, 813–836.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. New York: Springer.
- Enders, W. (1995). *Applied Econometric Analysis*. New York: Wiley.
- Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987–1008.
- Engle, R.F. and Granger, C.W.J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, **55**, 251–276.
- Engle, R.F. and Yoo, B.S. (1987). Forecasting and testing in cointegrated systems. *Journal of Econometrics*, **35**, 143–159.
- Engle, R.F. and Manganelli, S. (1999). CAViaR: Conditional autoregressive value at risk by regression quantiles. Preprint, University of California, San Diego.
- Engle, R.F. and Patton, A.J. (2001). What good is a volatility model. *Quan-*

*titative Finance*, **1**, 237–245.

Fernandez, C. and Stell, M. (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.

Fox, R. and Taqqu, M. S. (1986). Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, **14**, 517–532.

Franses, P.H. (1998). *Time Series Models for Business and Economic Forecasting*. Cambridge: Cambridge University Press.

Garman, M.B. and Klass, M.J. (1980). On the estimation of security price volatility from historical data. *Journal of Business*, **53**, 67–78.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221–238.

Glosten, L.R., Jagannathan, R. and Runkle, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48**, 1779–1801.

Gordy, M.B. (2000). A comparative anatomy of credit risk models. *Journal of Banking and Finance*, **26**, 1335–1353.

Granger, C.W.J. (1969). Investigating causal relationships by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.

Granger, C. M. G. and Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, **1**, 15–29.

Granger, C.W.J. and Newbold, P.E. (1974). Spurious regression in econometrics. *Journal of Econometrics*, **2**, 111–120.

Granger, C.W.J. and Andersen, A.P. (1978). *An Introduction to Bilinear Time Series Models*. Gottingen: Vandenhoeck and Ruprecht.

Hamilton, J.D. (1990). Analysis of time series subject to change in regime. *Journal of Econometrics*, **45**, 39–70.

- Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hannan, E. J. (1973). The asymptotic theory of linear time series models. *Journal of Applied Probability*, **10**, 130–145.
- Hannan, E.J. (1980). The estimation of the order of an ARMA process. *The Annals of Statistics*, **8**, 1071–1081.
- Hannan, E.J. (1982). Testing for autocorrelation and Akaike's criterion. In *Essays in Statistical Science*, special volume 19A of *Journal of Applied Probability*. The Applied Probability Trust, Sheffield, 403–412.
- Hansen, B.E. (1992). Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends. *Journal of Econometrics*, **53**, 87–121.
- Hendry, D.F. and Juselius, K. (2000). Explaining cointegration analysis: Part I. *The Energy Journal*, **21**, 1–42.
- Hendry, D.F. and Juselius, K. (2001). Explaining cointegration analysis: Part II. *The Energy Journal*, **22**, 75–120.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, **68**, 165–176.
- Hsiao, C. (1979). Autoregressive modelling of Canadian money and income data. *Journal of the American Statistical Association*, **74**, 553–560.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **16**, 770–799.
- Hurst, H. E. (1957). A suggested statistical model of time series that occur in nature. *Nature*, **180**, 494.
- Hurvish, C.M. and Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**, 297–307.
- Jaquier, E., Polson, N.G. and Rossi, P.E. (1994). Bayesian analysis of stochastic volatility models (With Discussion). *Journal of Business and Economic Statistics*, **12**, 371–417.

- Jenkins, G.M. and Watts, D.G. (1968). *Spectral Analysis and its Applications*. San Francisco: Holden-Day.
- Joanes, D.N. and Gill, C.A. (1998). Comparing measures of sample skewness and kurtosis. *The Statistician*, **47**, 183–189.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, **12**, 231–254.
- Johansen, S. (1994). The role of the constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, **13**, 205–229.
- Johansen, S. (1995). *Likelihood Based Inference in Cointegrated Vector Error Correction Models*. Oxford: Oxford University Press.
- Kim, S., Shephard, N. and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **85**, 361–393.
- Kim, C.-J. and Nelson, C.R. (1999). *State Space Models With Regime Switching*. Cambridge: The MIT Press.
- Koenker, R.W. and Basset, G.W. (1978). Regression quantiles. *Econometrica*, **46**, 33–50.
- Koop, G. (1992). Aggregate shocks and macroeconomic fluctuations: A Bayesian approach. *Journal of Applied Econometrics*, **7**, 395–411.
- Koopman, S.J., Harvey, A.C., Doornick, J.A. and Shephard, N. (1995). STAMP 5.0: Structural Time Series Analyser, Modeller and Predictor. London: Timberlake.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, **54**, 159–178.
- Laurent, S. and Peters, J.P. (2002). GARCH 2.3: An Ox package for estimating and forecasting various ARCH models. *Journal of Economic Surveys*, **16**,



447–485.

Layton, A. P. (1984). A further note on the detection of Granger instantaneous causality. *Journal of Time Series Analysis*, **5**, 15–18.

Ljung, G. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297–303.

Lo, A.W. (1991). Long term memory in stochastic market prices. *Econometrica*, **59**, 1279–1313.

Longerstaeey, J. and More, L. (1995). Introduction to RiskMetrics<sup>TM</sup>, 4th Edition. New York: Morgan Guaranty Trust Company.

Lopes, H.F. and Lima, E.C.R. (1995). *Co-integração: Enfoques Clássico e Bayesiano*. Vitória: Sexta Escola de Séries Temporais e Econometria.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Heidelberg: Springer Verlag.

Mandelbrot, B.B. and Van Ness, J.W. (1968). Fractional Brownian motions, fractional Brownian noises and applications. *SIAM Review*, **10**, 422–437.

Mandelbrot, B.B. and Wallis, J. (1968). Noah, Joseph and operational hydrology. *Water Resources Research*, **4**, 909–918.

Manteiga, S.M. (2002). Comparação de Metodologias para Estimação de Volatilidades para Cálculo do VaR- Valor-no-Risco e Modelagem de Perdas Não Previstas pelo VaR em Momentos de Crise. Dissertação de Mestrado Profissionalizante, FEA-IME, Universidade de São Paulo.

McLeod, A. I. and Hipel, K. W. (1978). Preservation of the rescaled adjusted range, 1: A reassessment of the Hurst phenomenon. *Water Resources Research*, **14**, 491–508.

Mendes, B.V.M. (2004). *Introdução à Análise de Eventos Extremos*. Rio de Janeiro: E-Papers Serviços Editoriais.

Merton, R.C. (1973). Theory of option pricing. *Bell. J. Econ. Manag. Sci.*, **4**, 141–183.

- Mills, T.C. (1999). *The Econometric Modelling of Financial Time Series*. Second Edition. Cambridge University Press.
- Mittnik, S., Rachev, S.T. and Paoletta, M.S. (1998). Stable Paretian modeling in finance: Some empirical and theoretical aspects. In *A Practical Guide to Heavy Tails* (R.J. Adler, R.E. Feldman and M.S. Taqqu, editors), p. 79–110. Boston: Birkhäuser.
- Morettin, P.A. (1984). The Levinson algorithm and its applications in time series analysis. *International Statistical Review*, **52**, 83–92.
- Morettin, P.A. and Toloi, C.M.C. (2006). *Análise de Séries Temporais*. Segunda Edição. São Paulo: Editora E. Blücher-Associação Brasileira de Estatística.
- Motta, A.C.O. (2001). Modelos de Espaço de Estados Não-Gaussianos e o Modelo de Volatilidade Estocástica. Dissertação de mestrado, IMECC-UNICAMP.
- Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns. *Econometrica*, **59**, 347–370.
- Nelson, D.B. and Cao, C.Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, **10**, 229–235.
- Newey, W.K. and West, K.D. (1987). A simple positive semidefinite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708.
- Ng, S. and Perron, P. (1995). Unit roots tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, **90**, 268–281.
- Osterwald-Lenum, M. (1992). A note with quantiles of the asymptotic distribution of maximum likelihood cointegration rank statistics. *Oxford Bulletin of Economics and Statistics*, **54**, 461–472.
- Pagan, A.R. and Schwert, G.W. (1990). Alternative models for conditional stochastic volatility. *Journal of Econometrics*, **45**, 267–290.

- Peña, D., Tsay, R.S. and Tiao, G.C. (2001). *A Course in Time Series Analysis*. New York: Wiley.
- Perron, P. (1988). Trends and random walks in macroeconomic time series: Further evidence from a new approach. *Journal of Economic Dynamics and Control*, **12**, 297–332.
- Perron, P. and Ng, S. (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies*, **63**, 435–463.
- Phillips, P.C.B. (1986). Understanding spurious regression in econometrics. *Journal of Econometrics*, **33**, 311–340.
- Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica*, **55**, 277–301.
- Phillips, P.C.B. (1991). Optimal inference in cointegrated systems. *Econometrica*, **59**, 283–306.
- Phillips, P.C.B. (1994). Model determination and macroeconomic activity. Cowles Foundation Discussion Paper # 1083, Yale University.
- Phillips, P.C.B. and Perron, P. (1988). Testing for unit roots in time series regression. *Biometrika*, **75**, 335–346.
- Phillips, P.C.B. and Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, **58**, 73–93.
- Phillips, P.C.B. and Shimotsu, K. (2004). Local Whittle estimation in non-stationary and unit root cases. *The Annals of Statistics*, **32**, 656–692.
- Pierce, D.A. and Haugh, L.D. (1977). Causality in temporal systems: Characterizations and a survey. *Journal of Econometrics*, **5**, 265–293.
- Quenouille, M.H. (1949). Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, Series B*, **11**, 68–84.
- Rissanen, J. (1978). Modelling by shortest data description. *Automatica*, **14**, 465–471.

- Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Second Edition. Birkhäuser.
- Rockafellar, R.T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21–41.
- Rockafellar, R.T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, **26** 1443–1471.
- Sáfadi, T. and Morettin, P.A. (2000). A Bayesian analysis of threshold autoregressive moving average models. *Sankhya*, Series B, **62**, 353–371.
- Said, S.E. and Dickey, D.A. (1985). Hypothesis testing in ARIMA (p,1,q) models. *Journal of the American Statistical Association*, **80**, 369–374.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461–464.
- Schwert, W. (1989). Test for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics*, **7**, 147–159.
- Seater, J. J. (1993). World temperature – Trend uncertainties and their implications for economic policy. *Journal of Business and Economic Statistics*, **11**, 265–277.
- Shephard, N. and Pitt, M.K. (1997). Likelihood analysis of Non-Gaussian measurement time series. *Biometrika*, **84**, 653–667.
- Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike’s information criterion. *Biometrika*, **63**, 117–126.
- Smith, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika*, **72**, 67–90.
- Stock, J.H. and Watson, M.W. (1988). Testing for common trends. *Journal of The American Statistical Association*, **83**, 1097–1107.
- Subba Rao, T. (1981). On the theory of bilinear time series models. *Journal of the Royal Statistical Society*, Series B, **43**, 244–255.

- Subba Rao, T. and Gabr, M.M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Berlin: Springer.
- Taylor, S.J. (1980). Conjectured models for trend in financial prices tests as forecasts. *Journal of The Royal Statistical Society, Series B*, **42**, 338–362.
- Taylor, S.J. (1986). *Modeling Financial Time Series*. New York: Wiley.
- Tong, H. (1983). *Threshold Models in Non-Linear Time Series Analysis*. Lecture Notes in Statistics, **21**. Heidelberg: Springer
- Tong, H. (1990). *Non-Linear Time Series Models*. Oxford: Oxford University Press.
- Tsay, R.S. (2002). *Analysis of Financial Time Series*. New York: Wiley.
- Tsurumi, H. and Wago, H. (1994). A Bayesian analysis of unit root and cointegration with an application to a Yen–Dollar exchange rate model. Preprint.
- Venables, W.N. and Ripley, B.D. (2001). *Modern Applied Statistics with S-Plus*. Second Edition. New York: Springer.
- Wiener, N. (1923). Differential space. *Journal of Mathematical Physics*, **2**, 131–174.
- Zakoian, J.M. (1994). Threshold heteroskedasticity models. *Journal of Economic Dynamics and Control*, **18**, 931–955.
- Zivot, E. and Wang, J. (2003). *Modelling Financial Time Series With SPLUS*. New York: Springer.



---

## Séries Usadas no Texto

As seguintes séries podem ser obtidas de <http://www.ime.usp.br/~pam/ef.html>.

BAN-D: Preços diários das ações do Banespa, de 2/1/95 a 27/12/2000 ( $T=1499$ )  
d-ban95.00.dat

CEMIG-D: Preços diários das ações da CEMIG, de 3/1/95 a 27/12/2000 ( $T=1499$ ).  
d-cemig95.00.dat

DOW-D: Índices diários do Dow Jones Industrial Average, de 3/1/95 a 26/11/02 ( $T=1992$ )  
d-dow95.02.dat

IBM-D: Log-retornos diários da IBM de 3/7/1962 a 31/12/1999 ( $T = 9442$ )  
d-ibm62.99.dat

IBV-D: Índices diários do Ibovespa, de 2/1/95 a 27/12/2000 ( $T=1499$ )  
d-ibv95.00.dat

PETRO-D: Preços diários das ações da Petrobrás PN, de 2/1/95 a 27/12/2000 ( $T=1499$ )  
d-petro95.00.dat

US/REAL-D: Taxas de câmbio diárias USD/Real, de 30/06/1994 a 01/07/1998 ( $T = 997$ )

d-usre94.98.dat

CBOND-M: Dados mensais dos juros do C-Bond brasileiro, de julho de 1994 a agosto de 2001 ( $T=86$ ).

m-cbond94.01.dat

IBM-M: Log-retornos mensais da IBM de janeiro de 1926 a dezembro de 1997 ( $T=864$ )

m-ibm26.97.dat

IBV-M: Índices mensais do Ibovespa, de julho de 1994 a agosto de 2001 ( $T=86$ )

m-ibv94.01.dat

ICV-M: Índices mensais de custo de vida de São Paulo, de janeiro de 1970 a junho de 1979 ( $T=114$ )

m-icv70.79.dat

IPI-M: Índices mensais de produção industrial do Brasil-setor de alimentação, de janeiro de 1985 a julho de 2000 ( $T=187$ ).

m-ipi85.00.dat

PFI-M: Índices mensais de produção física industrial geral do Brasil, de janeiro de 1985 a julho de 2000 ( $T=187$ )

m-pfi85.00.dat

SP-M: Log-retornos mensais do S&P500, de janeiro de 1962 a dezembro de 1999 ( $T=456$ )

m-sp62.99.dat

IBV-ID: Índices intra-diários do Ibovespa, a cada 15 minutos, de 6/4/1998 a 13/8/2003 ( $T = 37961$ )

id-ibv98.03.dat

TLM-ID: Preços intra-diários da Telemar PN, a cada 15 minutos, de 2/1/2002 a 31/3/2005 ( $T = 21429$ )

id-tel02.05.dat

BEV-A: Índices de preços anuais de trigo na Europa, de 1500 a 1869 ( $T = 370$ )

a-bev00.69.dat



---

# Índice

- agregação
  - cross-section, 12
  - temporal, 12
- AIC, 90
  - correção, 92
  - de modelo AR, 92
  - de modelo ARMA, 91
- aleatório
  - passeio, 43
- AR(p)
  - f.a.c., 77
  - f.a.c.p., 80
  - vetorial, 224
- ARCH, 1
  - estimação, 154
  - identificação, 153
  - modelos, 150
  - momentos, 150
  - previsão, 156
  - verificação, 155
- ARFIMA
  - estacionariedade, 300
  - estimação, 305
  - f.a.c., 300
  - modelos, 299
- ARIMA
  - identificação, 76
  - modelos, 75
- ARMA, 1
  - modelos, 75
- ARMA(p,q), 52
  - auto-covariâncias, 54
  - estacionariedade, 53
  - f.a.c., 78
  - f.a.c.p., 80
  - invertibilidade, 53
- assimetria, 16
- auto-covariância
  - continuidade, 39
  - propriedades, 37
- auto-regressivos
  - AR(1), 44
  - AR(p), 44
  - processos, 44
  - vetoriais, 224

- BIC, 92
  - de modelo ARMA, 92
- Browniano
  - geométrico, 66
  - movimento, 62
  - padrão, 62
- causalidade
  - de Granger, 241
  - instantânea, 244
  - unidirecional, 242
- Choleski
  - decomposição, 256
- correlações
  - matriz de, 219
- covariâncias
  - estimação, 222
  - matriz de, 217
- curtose, 16
- curva
  - de impacto de informação, 170
- CVaR, 208
- DF
  - teste, 82
- diagnóstico
  - de modelos ARIMA, 100
- diferenças
  - equação de, 47
- distribuição
  - de erro generalizada, 186
  - de valores extremos, 205, 212
  - t assimétrica, 186
- distribuições
  - finito-dimensionais, 33
- Durbin-Levinson
  - algoritmo, 80
- ergodicidade, 15
- erro
  - de previsão, 106
  - de previsão, variância, 106
- estacionário
  - estrito, 35
  - fraco, 36
  - processo, 31
- estacionários
  - lineares, 41
- estacionariedade
  - condição, 47
- estatística
  - R/S, 294
- estimação, 94
  - de modelos AR, 94
  - de modelos MA, 96
  - de modelos VAR, 231
- estimadores
  - de MV exatos, 97
  - variância dos, 98
- EViews, 23, 100, 155, 231, 234
- EWMA, 22
- FIEGARCH
  - modelos, 312
- FIGARCH
  - modelos, 312
- Fréchet, 212
- função
  - de auto-correlação, 39
  - de auto-correlação parcial, 78
  - de auto-covariância, 34
- funcional

- teorema limite central, 68
- GARCH
  - estimação, 164
  - extensões, 168
  - identificação, 163
  - modelos, 160
  - previsão, 164
- GARCH(1,1), 163
- GPH
  - estimador, 298
  - procedimento, 296
- Gumbel, 212
- GVE, 212
  - inferência, 214
- heteroscedasticidade, 18
- Hurst
  - coeficiente, 289
- identificação
  - de modelos ARIMA, 76
  - procedimento de, 82
- log-retorno, 8
  - da carteira, 11
- longa
  - memória, 289
- média
  - de um processo, 34
- MA(q), 48
  - auto-covariância, 49
  - f.a.c., 78
  - f.a.c.p., 80
  - invertibilidade, 51
- martingales, 69
- convergência, 71
- diferença de, 71
- matrizes, 249
  - autovalores, 253
  - autovetores, 253
  - decomposição, 253
  - determinante de, 249
  - particionadas, 251
  - posto de, 249
  - produto de Kronecker, 251
  - traço de, 249
  - vetorização, 251
- memória
  - curta, 289
  - longa, 289
- memória longa, 289
  - testes, 293
- Minitab, 99
- modelos
  - ARCH, 150
  - ARFIMA, 299
  - ARIMA, 59
  - ARMA vetorias, 235
  - auto-regressivos vetoriais, 224
  - bilineares, 149
  - de volatilidade estocástica, 176
  - EGARCH, 168
  - estruturais, 256
  - FIEGARCH, 312
  - FIGARCH, 312
  - GARCH, 160
  - IGARCH, 164
  - lineares por parte, 149
  - mistos, 97
  - multivariados, 217

- não-lineares, 148
- polinomiais, 148
- sazonais, 110
- sazonais multiplicativos, 112
- switching, 149
- TARCH, 173
- VARMA, 240
- momentos, 34
  - média, 34
- MVE, 176
  - estimação, 180
  - forma canônica, 176
  - formulação de Jaquier, 177
  - propriedades, 177
- previsão
  - com modelos ARIMA, 104
  - erro, 106
  - intervalo de confiança, 108
- processo
  - auto-similar, 64
  - com memória longa, 289
  - complexo, 40
  - contínuo em mq, 39
  - especificação de um, 33
  - estocástico, 31
  - gaussiano, 37
  - integrado, 59
  - linear, 41
  - linear geral, 54
  - linear multivariado, 221
  - não-estacionário, 56
  - puramente aleatório, 43
  - realização, 32
  - trajetória, 32
  - trend-stationary, 57
- raiz unitária, 123
  - teste ADF, 128
  - teste DF, 126
  - teste PP, 134
- resíduos, 101
- retornos
  - agregação, 10
  - anualização, 9
  - brutos simples, 8
  - compostos, 8
  - distribuição, 13
  - fatos estilizados, 18
  - líquidos simples, 8
  - multiperíodos, 8
- risco, 189
  - medidas coerentes, 208
  - valor em, 189
- RiskMetrics, 22
- ruído branco
  - discreto, 42
- S+FinMetrics, 23, 155
- SARIMA
  - modelos, 112
- sazonais
  - modelos, 110
- sazonal
  - diferença, 111
- SCA, 222, 234
- seqüência
  - i.i.d.*, 42
  - aleatória, 41
- SPlus, 23
- STAMP, 23, 181
- teste

- ADF, 128
- de auto-correlação residual, 101
- de Box-Pierce-Ljung, 102
- DF, 124
- DF-extensões, 129
- Phillips-Perron, 134
- unitárias
  - raízes, 82, 123
- valor
  - em risco, 189
- valores extremos
  - teoria, 211
- VAR
  - construção, 230
  - diagnóstico, 232
  - equações de Yule-Walker, 229
  - estacionário, 229
  - estimação, 231
  - identificação, 230
  - modelos, 224
  - previsão, 232
- VaR, 191
  - condicional, 208
  - definição, 191
  - RiskMetrics, 192
  - usando modelos GARCH, 199
  - usando quantis empíricos, 203
  - usando TVE, 204
- VARMA
  - modelos, 240
- volatilidade, 1, 20
  - estatística, 20
  - estocástica, 176
  - histórica, 20
  - implícita, 20
  - modelagem, 147
  - realizada, 22
- Weibull, 212
- Wold
  - teorema de, 56
- Yule-Walker
  - equações, 48, 78
  - estimadores, 48