

# Econometria Financeira

**Um Curso em Séries Temporais Financeiras**

Pedro A. Morettin

Departamento de Estatística

Instituto de Matemática e Estatística

Universidade de São Paulo

Terceira Edição

São Paulo, março de 2017

---

# Conteúdo

<b>Prefácio da Terceira Edição</b>	<b>xix</b>
<b>Prefácio da Segunda Edição</b>	<b>xxi</b>
<b>Prefácio</b>	<b>xxiii</b>
<b>1 Preliminares</b>	<b>1</b>
1.1 Introdução . . . . .	1
1.2 Tipos de Dados . . . . .	2
1.3 Retornos . . . . .	7
1.4 Agregação de Retornos . . . . .	10
1.5 Distribuição de Retornos . . . . .	12
1.6 Assimetria e Curtose . . . . .	16
1.7 Fatos Estilizados Sobre os Retornos . . . . .	18
1.8 Volatilidade . . . . .	20
1.9 Aspectos Computacionais . . . . .	22
1.10 Problemas . . . . .	23
1.11 Apêndice 1.A. Distribuições Estáveis . . . . .	25
1.12 Apêndice 1.B. Teste de Normalidade . . . . .	29
<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 Autocovariância . . . . .	36
2.4 Processos Estocásticos Complexos . . . . .	39
2.5 Processos Lineares Estacionários . . . . .	40
2.5.1 Processos Autorregressivos . . . . .	42
2.5.2 Processos de Médias Móveis . . . . .	46
2.5.3 Processos Autorregressivos e de Médias Móveis . . . . .	49

2.6	Processos Não Estacionários . . . . .	52
2.7	Movimento Browniano . . . . .	58
2.8	Martingales . . . . .	64
2.9	Problemas . . . . .	66
<b>3</b>	<b>Modelos ARIMA</b>	<b>69</b>
3.1	Introdução . . . . .	69
3.2	Identificação . . . . .	70
3.2.1	Procedimento de Identificação . . . . .	74
3.2.2	Formas Alternativas de Identificação . . . . .	80
3.3	Estimação . . . . .	84
3.3.1	Modelos Autorregressivos . . . . .	85
3.3.2	Modelos de Médias Móveis . . . . .	86
3.3.3	Modelos Mistos . . . . .	87
3.3.4	Variâncias dos Estimadores . . . . .	88
3.4	Diagnóstico . . . . .	90
3.4.1	Teste de Autocorrelação Residual . . . . .	90
3.4.2	Teste de Box-Pierce-Ljung . . . . .	91
3.5	Previsão com Modelos ARIMA . . . . .	92
3.6	Modelos Sazonais . . . . .	97
3.7	Problemas . . . . .	100
<b>4</b>	<b>Raízes Unitárias</b>	<b>105</b>
4.1	Introdução . . . . .	105
4.2	O Teste de Dickey-Fuller . . . . .	106
4.3	Extensões do Teste DF . . . . .	110
4.4	Comentários Finais . . . . .	115
4.5	Problemas . . . . .	119
4.6	Apêndice 4: Provas dos Teoremas 4.1 e 4.2 . . . . .	120
<b>5</b>	<b>Modelos para a Volatilidade</b>	<b>123</b>
5.1	Introdução . . . . .	123
5.2	Modelos ARCH . . . . .	124
5.3	Modelos GARCH . . . . .	134
5.4	Extensões do Modelo GARCH . . . . .	139
5.4.1	Modelos EGARCH . . . . .	139
5.4.2	Modelos TGARCH . . . . .	142
5.5	Modelos de Volatilidade Estocástica . . . . .	143
5.6	Tópicos Adicionais . . . . .	153
5.7	Problemas . . . . .	153
5.8	Apêndice 5. Algumas Distribuições Especiais . . . . .	155

<b>6 Processos com Memória Longa</b>	<b>157</b>
6.1 Introdução . . . . .	157
6.2 Estimação e Testes para Memória Longa . . . . .	160
6.2.1 Estatística R/S . . . . .	161
6.2.2 Procedimento GPH . . . . .	163
6.3 Modelos ARFIMA . . . . .	165
6.4 Estimação de modelos ARFIMA . . . . .	169
6.5 Previsão de modelos ARFIMA . . . . .	172
6.6 Processos de Volatilidade com ML . . . . .	174
6.7 Problemas . . . . .	176
6.8 Apêndice 6. Volatilidade de Garman-Klass . . . . .	179
<b>7 Valor em Risco</b>	<b>181</b>
7.1 Introdução . . . . .	181
7.2 Valor em Risco . . . . .	182
7.3 VaR Usando a Distribuição Normal . . . . .	184
7.4 VaR Usando Modelos ARMA e GARCH . . . . .	189
7.5 VaR Usando Quantis Empíricos . . . . .	193
7.6 VaR Usando a Teoria de Valores Extremos . . . . .	193
7.7 VaR Usando o Método POT . . . . .	196
7.8 Tópicos Adicionais . . . . .	200
7.9 Problemas . . . . .	202
7.10 Apêndice 7. Teoria de Valores Extremos . . . . .	203
<b>8 Análise de Dados de Alta Frequência</b>	<b>209</b>
8.1 Introdução . . . . .	209
8.2 Volatilidade Realizada . . . . .	212
8.3 Modelo de Duração Condicional . . . . .	226
8.4 Modelagem da Volatilidade . . . . .	232
8.5 Modelos R-GARCH . . . . .	233
8.6 Comentários Adicionais . . . . .	242
8.7 Problemas . . . . .	243
8.8 Apêndice 8: Notas Complementares . . . . .	244
<b>9 Modelos Lineares Multivariados</b>	<b>247</b>
9.1 Introdução . . . . .	247
9.2 Séries Estacionárias . . . . .	248
9.3 Estimação de Médias e Covariâncias . . . . .	251
9.4 Modelos Autorregressivos Vetoriais . . . . .	253
9.5 Construção de Modelos VAR . . . . .	258
9.6 Modelos ARMA Vetoriais . . . . .	264
9.7 Causalidade de Granger . . . . .	266

9.8 Problemas . . . . .	271
9.9 Apêndice 9.A. Alguns Resultados sobre Matrizes . . . . .	273
9.10 Apêndice 9.B. Demonstração da Proposição 9.2 . . . . .	277
9.11 Apêndice 9.C. Modelo VAR( $p$ ) na Forma VAR(1) . . . . .	278
9.12 Apêndice 9.D. Modelos Estruturais . . . . .	279
<b>10 Processos Cointegrados</b>	<b>283</b>
10.1 Introdução . . . . .	283
10.2 Tendências Comuns . . . . .	286
10.3 Modelo de Correção de Erros . . . . .	288
10.4 Testes para Cointegração . . . . .	292
10.4.1 Procedimento de Engle e Granger . . . . .	293
10.4.2 Procedimento de Johansen . . . . .	295
10.5 Comentários Finais . . . . .	302
10.5.1 Enfoque Bayesiano . . . . .	302
10.5.2 Termos Determinísticos no MCE . . . . .	302
10.5.3 Procedimento de Máxima Verossimilhança . . . . .	303
10.5.4 Cointegração Dinâmica . . . . .	305
10.6 Problemas . . . . .	305
<b>11 Análise de Dependência e Cópulas</b>	<b>309</b>
11.1 Introdução . . . . .	309
11.2 Medidas de Dependência . . . . .	310
11.3 Cópulas . . . . .	314
11.4 Famílias Paramétricas de Cópulas . . . . .	319
11.5 Ajuste de Cópulas Paramétricas . . . . .	325
11.5.1 Estimadores de Máxima Verossimilhança . . . . .	326
11.5.2 Cópulas Empíricas . . . . .	327
11.5.3 Estimadores de Pseudo-MV . . . . .	328
11.6 Cópulas para Séries Temporais . . . . .	330
11.7 Valor em Risco e Cópulas . . . . .	332
11.8 Comentários Adicionais . . . . .	335
11.9 Problemas . . . . .	337
<b>12 Modelos GARCH Multivariados</b>	<b>341</b>
12.1 Introdução . . . . .	341
12.2 Generalizações do Modelo GARCH Univariado . . . . .	343
12.2.1 Modelos VEC . . . . .	344
12.2.2 Modelos BEKK . . . . .	348
12.3 Combinações Lineares de Modelos GARCH . . . . .	354
12.3.1 Modelos Fatoriais . . . . .	354
12.3.2 Modelos Ortogonais . . . . .	355

12.3.3	Modelo Fatorial via Componentes Principais . . . . .	356
12.3.4	Modelo GICA-GARCH . . . . .	360
12.4	Combinações Não Lineares de Modelos GARCH . . . . .	362
12.4.1	Modelos com Correlações Condicionais . . . . .	362
12.4.2	Modelo Dinâmico Geral . . . . .	369
12.5	Tópicos Adicionais . . . . .	370
12.6	Problemas . . . . .	372
<b>Referências</b>		<b>375</b>
<b>Séries Usadas no Texto</b>		<b>395</b>
<b>Índice Remissivo</b>		<b>398</b>



---

# Prefácio da Terceira Edição

Nesta edição, acrescentamos o Capítulo 12, que trata dos modelos GARCH multivariados. A dificuldade desses modelos reside em vários aspectos, mas principalmente no computacional, já que os programas existentes não tratam, obviamente, de todos os modelos sugeridos na literatura. Também corrigimos erros existentes na segunda edição e acrescentamos material novo em três capítulos. No Capítulo 1, o apêndice sobre distribuições estáveis foi aumentado, tendo em vista uma nova seção sobre os modelos R-GARCH no Capítulo 8. No Capítulo 5, introduzimos a estimação bayesiana dos modelos de volatilidade estocástica. Outra novidade, é que estamos disponibilizando, em página na Internet (veja o Capítulo 1), soluções da maioria dos exemplos do livro por meio de pacotes do R. Agradeço a Kim Samejima pela inestimável colaboração para que isso fosse possível. Estamos, também, providenciando as soluções dos problemas sugeridos, que estarão à disposição oportunamente, na página do livro, <http://www.ime.usp.br/~pam/ef.html>.

São Paulo, março de 2017.

Pedro A. Morettin



---

# Prefácio da Segunda Edição

Esta segunda edição difere da anterior em vários sentidos. Primeiramente, a ordem dos capítulos foi alterada, para dar uma continuidade melhor aos assuntos tratados no livro. Em segundo lugar, um novo capítulo foi introduzido, tratando da análise da dependência entre duas ou mais séries temporais, incluindo o tópico de cópulas, que tem sido bastante usado em finanças recentemente. Finalmente, introduzimos dados atualizados de várias séries, itens novos em algumas sessões e corrigimos erros existentes na primeira edição. Agradecemos a vários colegas que enviaram sugestões e correções, especialmente a Sumaia A. Latif, que produziu várias figuras do Capítulo 11.

São Paulo, março de 2011.

Pedro A. Morettin



---

# 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.

A área de Econometria Financeira teve um desenvolvimento explosivo nos últimos anos, podendo-se dizer que resultou da fusão de conhecimentos provenientes das áreas de finanças empíricas e econometria de séries temporais. Importantes para esse crescimento foram os avanços em coleta e armazenamento de grandes bancos de dados e velocidade de processamento dos equipamentos computacionais.

Os três temas principais em finanças atualmente são o aprecamento de ativos, alocação de carteiras e gestão de riscos, sendo que áreas recentes de pesquisa incluem volatilidade e correlação de retornos de ativos financeiros, escolha ótima de carteiras e microestrutura de mercados. Neste trabalho, abordaremos apenas alguns tópicos da primeira área, com menção ao cálculo do VaR (valor em risco) de um ativo ou de uma carteira de ativos.

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 nessas 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 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, Clélia M. C. Toloi, 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. O conteúdo do Capítulo 10 foi baseado fortemente na dissertação de mestrado de Alberto F. Berti e nas teses de doutorado de Adriana Bruscato e Juan Carlos Ruilova. Agradeço a eles a permissão para usar dados e exemplos. 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 17º SINAPE, em 2006, baseado em parte do conteúdo deste livro.

Os dados usados e outros aspectos computacionais relacionados ao livro podem ser acessados em minha página <http://www.ime.usp.br/~pam>. Comentários e sugestões podem ser enviados ao meu endereço eletrônico, pam@ime.usp.br.

São Paulo, fevereiro de 2008.

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 a 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. Nesses 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. Desse modo, modelos lineares, como aqueles pertencentes à família dos modelos ARMA (autorregressivos 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, diversas séries apresentam alguma forma de autocorrelação, de modo que modelos ARMA podem ser inicialmente aplicados para remover essa 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 et al. (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 3, fazemos uma revisão sobre a construção de modelos ARIMA. No Capítulo 4, tratamos do importante problema de raízes unitárias e de como testar sua presença em dada série. No Capítulo 5, introduzimos os modelos heteroscedásticos condicionais e, no Capítulo 6, introduzimos os processos com memória longa e modelos apropriados para representá-los, como os modelos ARFIMA.

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 frequentemente usado. No Capítulo 7, introduzimos várias formas de se calcular o valor em risco. No Capítulo 8, introduzimos noções básicas sobre análise de dados de alta frequência e, no Capítulo 9, estendemos o estudo dos modelos estudados nos capítulos 2 e 3 para englobar modelos lineares multivariados, analisando os modelos VAR (modelos autorregressivos vetoriais) e VARMA (modelos autorregressivos e de médias móveis vetoriais). O conceito de cointegração e o modelo de correção de erros são discutidos no Capítulo 10. No Capítulo 11, apresentamos algumas medidas de dependência entre variáveis aleatórias e séries temporais, com ênfase no estudo de cópulas. Finalmente, no Capítulo 12, estudamos os modelos GARCH multivariados.

## 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, normalmente 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. Nesses casos, os intervalos entre observações são variáveis aleatórias (as chamadas “durações”) e podemos ter também várias observações (negócios) coincidindo num mesmo instante de tempo. Esse 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/ef.html>.

**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 4 de julho de 1994 a 29 de setembro de 2010, num total de  $T = 4.019$  observações. O arquivo d-ibv94.10.dat contém

---

esses dados, que aparecem no arquivo na forma apresentada no Quadro 1.1.

Quadro 1.1: Dados do índice Ibovespa

Date	IBOVESPA
04/07/1994	3580,80
05/07/1994	3564,30
06/07/1994	3753,50
07/07/1994	3904,90
08/07/1994	4051,90
...	...

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  $T = 1.992$  observações. Estes dados encontram-se no arquivo d-dow95.02.dat. No Quadro 1.2, temos parte do arquivo de dados, onde aparecem os índices de abertura, máximo, mínimo, fechamento e o volume negociado no dia.

Quadro 1.2: Dados do índice Dow Jones

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
...	...	...	...	...	...

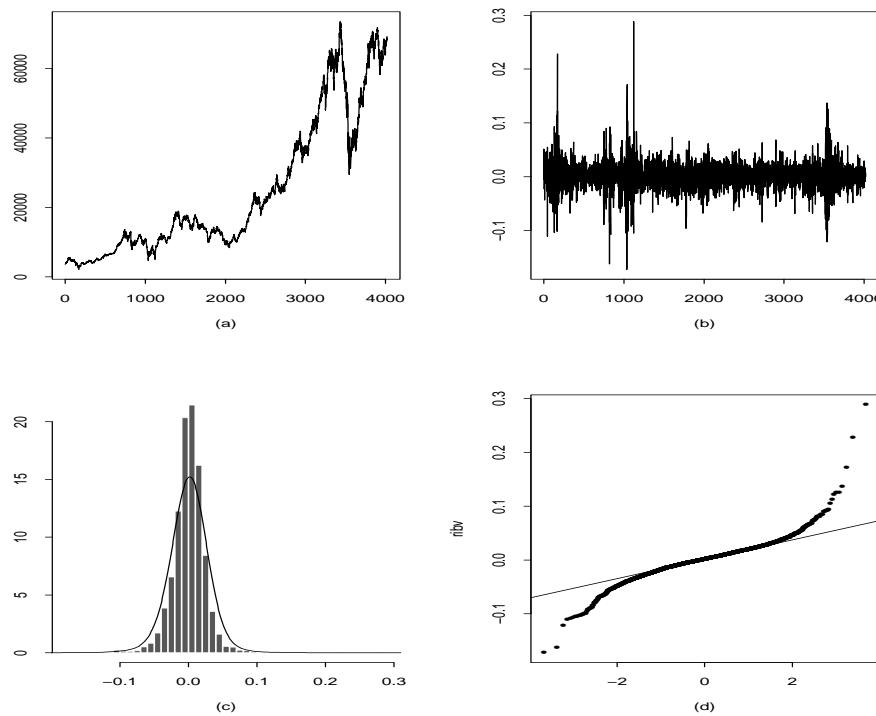


Figura 1.1: (a) Gráfico da série Ibovespa (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$ .

Dados de alta frequência podem ser registrados de diferentes formas, dependendo do tipo de ativo. No Quadro 1.3 temos um *layout* típico de dados de ações negociadas na Bolsa de Valores de São Paulo.

Quadro 1.3: Dados de alta frequência						
Data	Código	Empresa	Tipo	Pr.	Negócio	Hora
2003-02-03;	BBDC4;	BRADESCO;	PN*N1;	9.9900;	1101;	10;
2003-02-03;	BBDC4;	BRADESCO;	PN*N1;	10.0000;	1101;	20;
2003-02-03;	BBDC4;	BRADESCO;	PN*N1;	10.0000;	1101;	30;
2003-02-03;	BBDC4;	BRADESCO;	PN*N1;	10.0500;	1102;	40;
2003-02-03;	BBDC4;	BRADESCO;	PN*N1;	10.0500;	1102;	50;
...	...	...	...	...	...	...

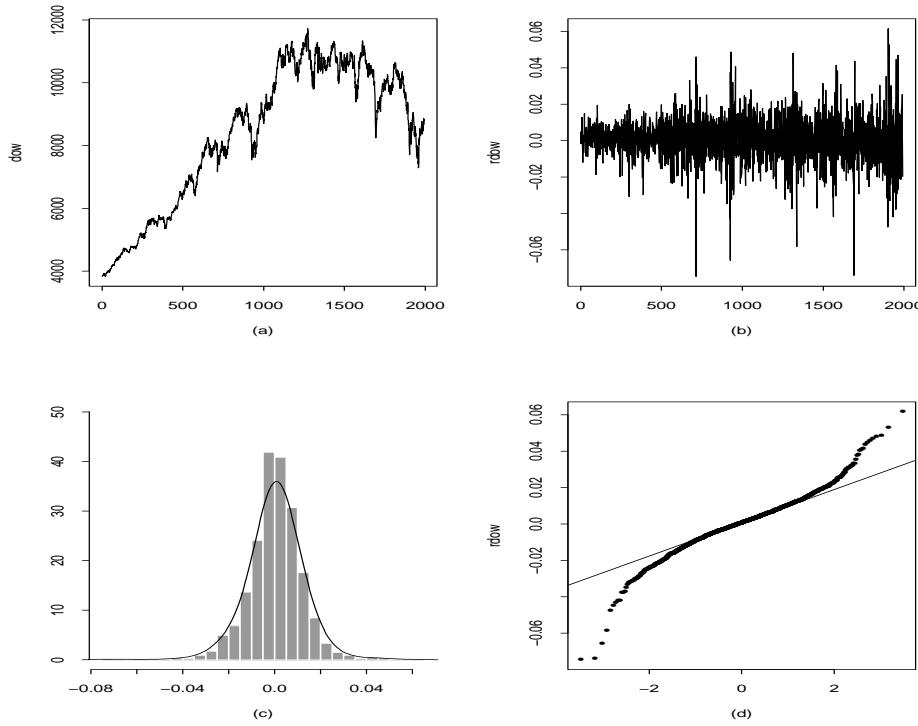


Figura 1.2: (a) Gráfico da série DJIA (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$ .

Este arquivo de dados traz os preços das ações Bradesco PN (na quinta coluna), no dia 3 de fevereiro de 2003. As outras colunas trazem informação sobre o código do ativo, a hora 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 vários 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. Nesses arquivos, os dados foram interpolados, de modo a se ter observações a cada 15 minutos.

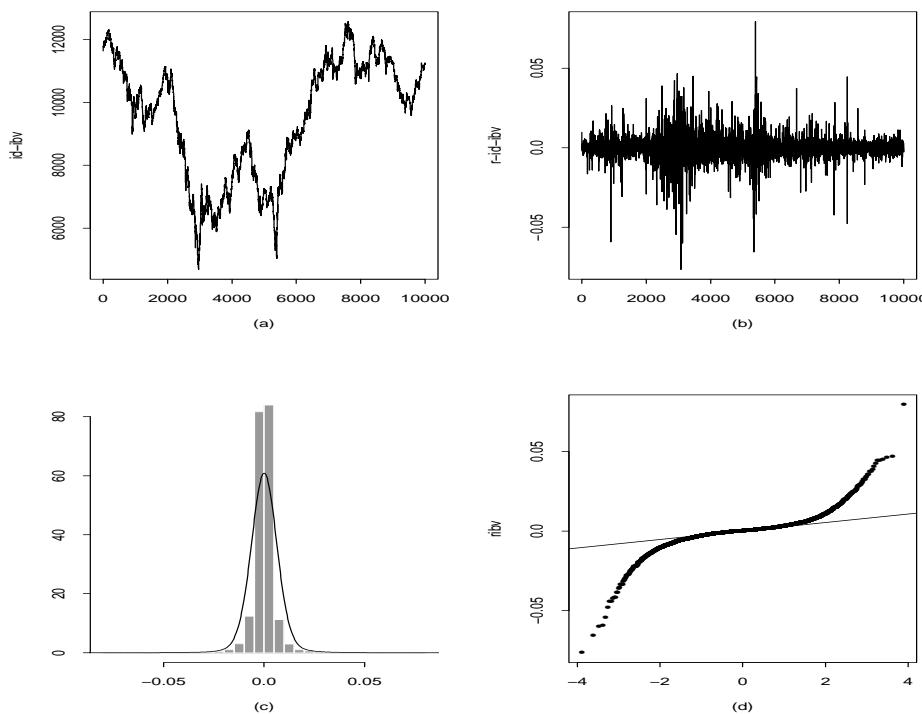


Figura 1.3: (a) Gráfico da série Ibovespa intradiária (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$ .

**Exemplo 1.3.** Na Figura 1.3 (a), temos o gráfico de parte dos dados do Ibovespa, observados a cada 15 minutos, de 6 de abril de 1998 a 13 de agosto de 2003, num total de 1.309 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.

Em todos os exemplos, apresentamos os gráficos das séries financeiras, que em geral são séries não estacionárias, como veremos no Capítulo 2. Na próxima seção, iremos considerar aquelas séries que efetivamente serão nosso objeto principal de estudo, os *retornos*.

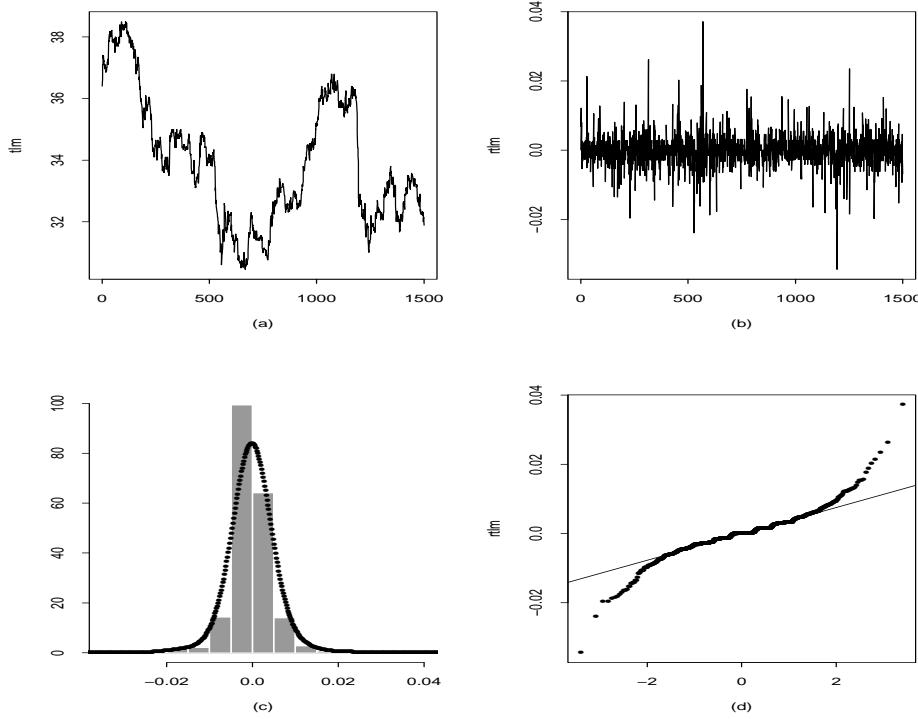


Figura 1.4: (a) Gráfico da série Telemar intradiária (b) série dos retornos (c) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$ .

### 1.3 Retornos

Um dos objetivos em finanças é a avaliação de riscos de uma carteira de ativos (instrumentos) financeiros. O risco é frequentemente 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 definida 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

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

Essa 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 e 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

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

de modo que

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

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

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

ou seja,

$$1 + R_t[k] = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}).$$

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. Veja o Problema 10.

Por sua vez, o log-retorno de período  $k$  fica, usando (1.4),

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

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

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

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  $P_1 = 3.580,80, \dots, P_5 = 4.051,90$ . Então,

$$\begin{aligned} R_2 &= \frac{P_2 - P_1}{P_1} = -0,004608, \\ r_2 &= \log \frac{P_2}{P_1} = -0,004619, \\ R_5[3] &= \frac{P_5 - P_2}{P_2} = 0,136801, \\ r_5[3] &= r_5 + r_4 + r_3 = \log \frac{P_5}{P_2} = 0,128218. \end{aligned}$$

**Exemplo 1.1.** (continuação) Na Figura 1.1 (b), temos o gráfico dos retornos diários do Ibovespa. Note que esses 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 os retornos são valores atípicos (*outliers*), quando comparados com os demais dias.

**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* (transversal), 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 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  $i = 1, \dots, N$ ,

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

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

para retornos simples e log-retornos, respectivamente.

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

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

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

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

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

**Exemplo 1.5.** Na Figura 1.5 (a), temos os índices mensais do Ibovespa, enquanto na Figura 1.5 (b) temos os respectivos retornos, no período de junho de 1994 a agosto de 2001, com  $T = 86$  dados (arquivo m-ibv94.01.dat). Esses 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.

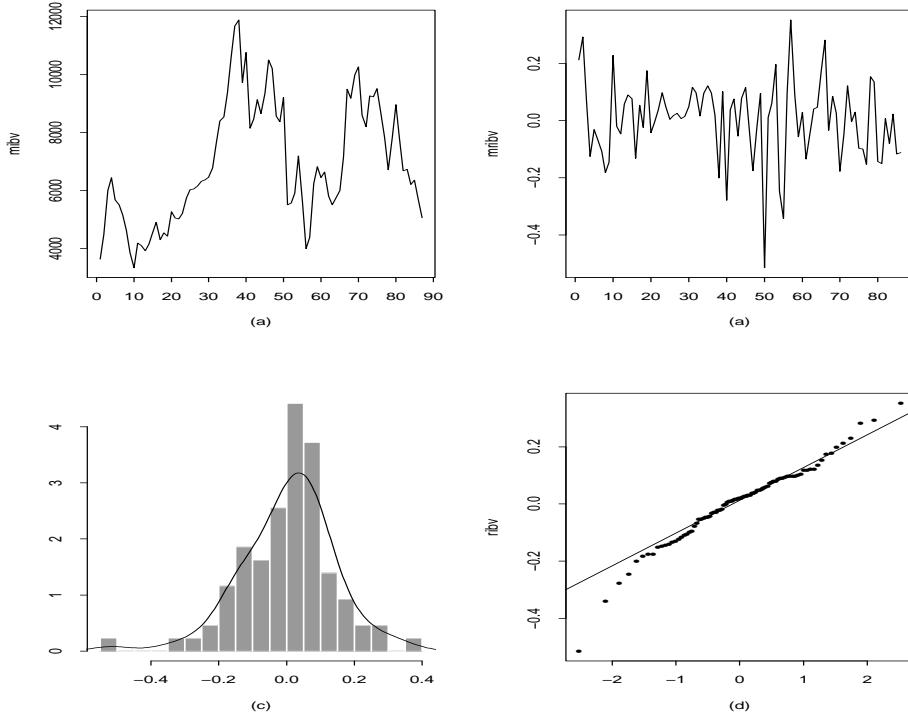


Figura 1.5: (a) Gráfico da série dos retornos mensais do Ibovespa (b) histograma dos retornos com densidade ajustada (d) gráfico  $Q \times Q$ .

## 1.5 Distribuição de Retornos

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

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

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 esses 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. Nessa 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 processso é ergódico se pudermos estimar características de interesse (média, autocovariância etc) a partir de uma única trajetória do processo. Assim, 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 covariáveis  $\mathbf{Y}$  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  $\boldsymbol{\theta} = (\mu_t, \sigma_t^2, t = 1, \dots, n)$  e a função de verossimilhança ficará

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

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

Como vimos, podemos considerar  $N$  ativos ao longo do tempo,  $r_{1t}, \dots, r_{Nt}$ , 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 reside no estudo das distribuições condicionais  $F_t(\mathbf{r}_t | \mathbf{r}_1, \dots, \mathbf{r}_{t-1}, \mathbf{Y}, \boldsymbol{\theta})$ .

Há vários dispositivos gráficos que podemos utilizar para avaliar a forma da distribuição dos retornos: o histograma, estimativas da função densidade e gráficos quantis-quantis (os chamados “Q×Q plots”).

### Histogramas

Este é um gráfico bastante conhecido e consiste em construir retângulos contíguos, a partir da divisão do espaço amostral em intervalos, geralmente com o mesmo comprimento. A partir dos dados  $x_1, \dots, x_n$  o histograma pode ser definido por

$$H(x) = \sum_{i=1}^n I\{x - \tilde{x}_i; h\},$$

onde  $\tilde{x}_i$  é o centro do intervalo onde a observação  $x_i$  cai e  $I\{z; h\}$  é o indicador do intervalo  $[-h, h]$ . Algum tipo de escalamento é feito para que a área do histograma seja um.

Há críticas ao uso do histograma, sendo que as principais são que o comportamento do histograma depende da escolha de  $h$  e da posição inicial da grade, e além disso informação é perdida, pois substituímos  $x_i$  pelo ponto médio do intervalo ao qual ele pertence.

Para evitarmos estas dificuldades, estimadores mais suaves da densidade  $f(x)$  podem ser usados. Um dos mais utilizados substitui retângulos por uma função *núcleo* (*kernel*) mais suave, obtendo-se

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n k(x - x_i; h),$$

onde  $k$  é o núcleo, em geral também uma função densidade, cuja variância é controlada por  $h$ , chamada *largura de faixa (bandwidth)*. O comportamento de  $\hat{f}$  vai depender de  $h$ , de modo que uma crítica ao histograma permanece. Mas é mais fácil comparar estimadores deste tipo do que histogramas. Uma escolha usual para  $k$  é  $k(z; h) = \phi(z; h)$ , a densidade normal com média zero e desvio padrão  $h$ .

O programa SPlus usa a função `hist` para obter o histograma de um conjunto de dados e as funções `ksmooth` e `density` para obter  $\hat{f}$  e pode-se escolher entre os núcleos retangular, triangular, cosseno e gaussiano, usando a opção `window`.

#### $Q \times Q$ Plots

Suponha que a v.a.  $X$  tenha distribuição contínua, com f.d.a.  $F$ . Então, para  $0 \leq p \leq 1$ , o  $p$ -quantil de  $F$  é o valor  $Q_p$  satisfazendo  $F(Q_p) = p$ , ou seja,

$$F(Q_p) = P(X \leq Q_p) = p.$$

Se existir a inversa de  $F$ , então  $Q_p = F^{-1}(p)$ . No caso de  $X$  ser discreta, a definição tem que ser modificada: o  $p$ -quantil é o valor  $Q_p$  satisfazendo

$$\begin{aligned} P(X \leq Q_p) &\geq p, \\ P(X \geq Q_p) &\geq 1 - p. \end{aligned}$$

Dado um conjunto de observações, podemos calcular os *quantis empíricos*. Uma maneira é considerar a *função de distribuição empírica*  $\hat{F}_n$  como estimador de  $F$ , ou seja, dadas as observações  $X_1, \dots, X_n$  de  $X$ ,

$$\hat{F}_n(x) = \frac{1}{n} \#\{i : 1 \leq i \leq n, X_i \leq x\}.$$

Então, o quantil  $Q_p$  é estimado pelo  $p$ -quantil de  $\hat{F}_n$ . Ou seja, o  $p$ -quantil estimado,  $q_p$ , seria definido por  $\hat{F}_n(q_p) = p$ . Contudo, usaremos um enfoque um pouco diferente.

Chamemos de  $r_1, \dots, r_T$  os retornos observados e considere as estatísticas de ordem  $r_{(1)} \leq r_{(2)} \leq \dots \leq r_{(T)}$ . Um estimador consistente de  $Q_p$  é dado pelo  $p$ -quantil empírico, definido por

$$q_p = \begin{cases} r_{(i)}, & \text{se } p = p_i = (i - 0,5)/T, i = 1, \dots, T \\ (1 - f_i)r_{(i)} + f_i r_{(i+1)}, & \text{se } p_i < p < p_{i+1} \\ r_{(1)}, & \text{se } 0 < p < p_1 \\ r_{(T)}, & \text{se } p_T < p < 1, \end{cases} \quad (1.21)$$

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.

Há dois tipos de gráficos  $Q \times Q$ : teóricos e empíricos. O primeiro tipo é usado para verificar se um conjunto de dados vem de determinada distribuição. O segundo tipo é usado para verificar se dois conjuntos de dados têm uma mesma distribuição. Para verificar se um conjunto de dados provém de uma distribuição especificada, consideramos o gráfico em que, no eixo horizontal, colocamos os quantis teóricos da distribuição hipotetizada para os dados, e no eixo vertical, os quantis empíricos dos dados, ambos calculados nos pontos  $p_i$ , acima. Se as observações realmente são provenientes da distribuição em questão, os pontos deverão estar distribuídos ao longo de uma reta.

**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. Esses 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  $Q \times Q$  com respeito aos quantis da distribuição normal padrão. 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  $r_t$  são 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

$$E(R_t) = e^{\mu+\sigma^2/2} - 1, \quad (1.22)$$

$$\text{Var}(R_t) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \quad (1.23)$$

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

$$A(X) = E\left(\frac{(X - \mu)^3}{\sigma^3}\right), \quad (1.24)$$

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

$$K(X) = E\left(\frac{(X - \mu)^4}{\sigma^4}\right). \quad (1.25)$$

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

$$\hat{A}(X) = \frac{m_3}{m_2^{3/2}} = \frac{1}{T} \sum_{t=1}^T \left( \frac{X_t - \bar{X}}{\hat{\sigma}} \right)^3, \quad (1.26)$$

$$\hat{K}(X) = \frac{m_4}{m_2^2} = \frac{1}{T} \sum_{t=1}^T \left( \frac{X_t - \bar{X}}{\hat{\sigma}} \right)^4, \quad (1.27)$$

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

$$\hat{A} \sim \mathcal{N}(0, 6/T), \quad \hat{K} \sim \mathcal{N}(3, 24/T). \quad (1.28)$$

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 intradiários. 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 não são, em geral, autocorrelacionados;
  2. os quadrados dos retornos são autocorrelacionados, apresentando uma correlação de defasagem 1 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 aqui.

**Exemplo 1.1.** (continuação) Na Figura 1.1 (b), temos a série de retornos do Ibovespa, na qual notamos os fatos estilizados apontados 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; queda da bolsa Nasdaq, em abril de 2000; início do governo Lula, em 2002; crise da *subprime* americana, em 2007 e a crise econômica internacional, em 2008.

Na Tabela 1.1, apresentamos algumas estatísticas das séries Ibovespa e DJIA. Notamos que as curtosas são altas, enquanto 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	0,000737	0,000410
Mediana	0,001364	0,000606
Desvio padrão	0,023983	0,011705
Assimetria	0,396310	-0,302922
Curtoza	11,385750	4,0184030
Mínimo	-0,172258	-0,074541
Máximo	0,288248	0,061554

**Exemplo 1.6.** Na Figura 1.6 (a), temos a série de preços diários das ações da Petrobras PN, no período de 18 de agosto de 1998 a 29 de setembro de 2010(arquivo d-petro98.10.dat), com  $T = 2.999$  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 Petrobras; 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, os retornos, o histograma e o 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 o desvio padrão 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 europeias. Essa 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 aqui, casos usuais são  $p = 2$  e  $p = 1$ .

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 7 para detalhes.

Outra possibilidade é utilizar os preços de abertura, mínimo, máximo e de fechamento (veja o Exemplo 1.2, para o caso do índice Dow Jones) para o cálculo de uma estimativa da volatilidade diária. Voltaremos a este assunto no Capítulo 6.

Para dados intradiários, pode-se estimar a volatilidade diária por meio da *volatilidade realizada*, que é a soma dos quadrados dos retornos obtidos em intervalos regulares durante o dia, por exemplo, a cada 15 minutos. Veja o Capítulo 8.

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.

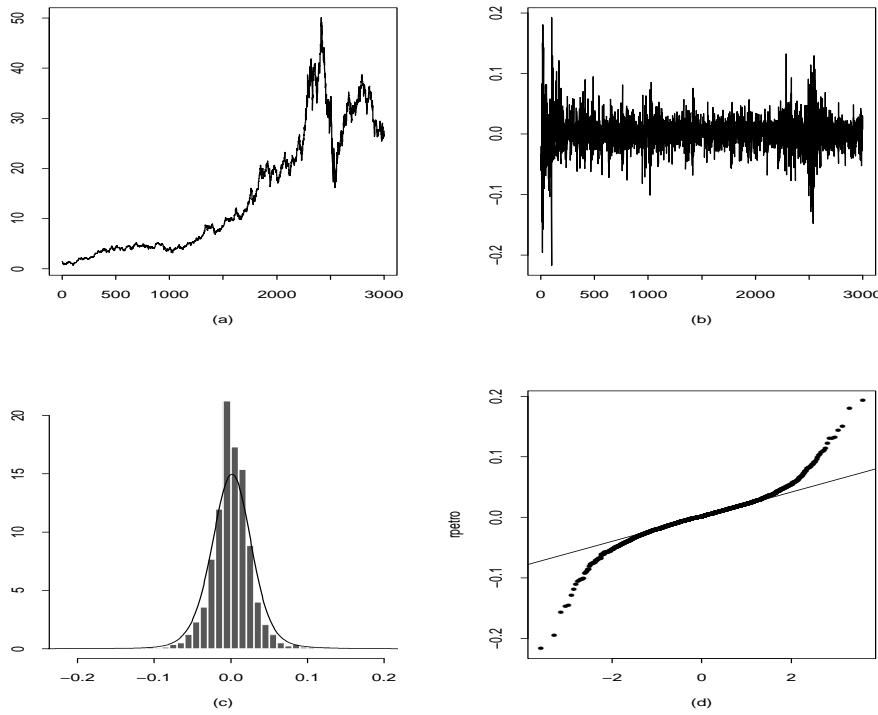


Figura 1.6: (a) Gráfico da série Petrobras (b) Retornos diários da Petrobras (c) Histograma com densidade ajustada (d) Gráfico  $Q \times Q$ .

Mas em toda atividade financeira (gestão de risco, apreçamento 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

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

$$h_t = E((r_t - \mu_t)^2 | \mathcal{F}_{t-1}) = E_{t-1}((r_t - \mu_t)^2), \quad (1.30)$$

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

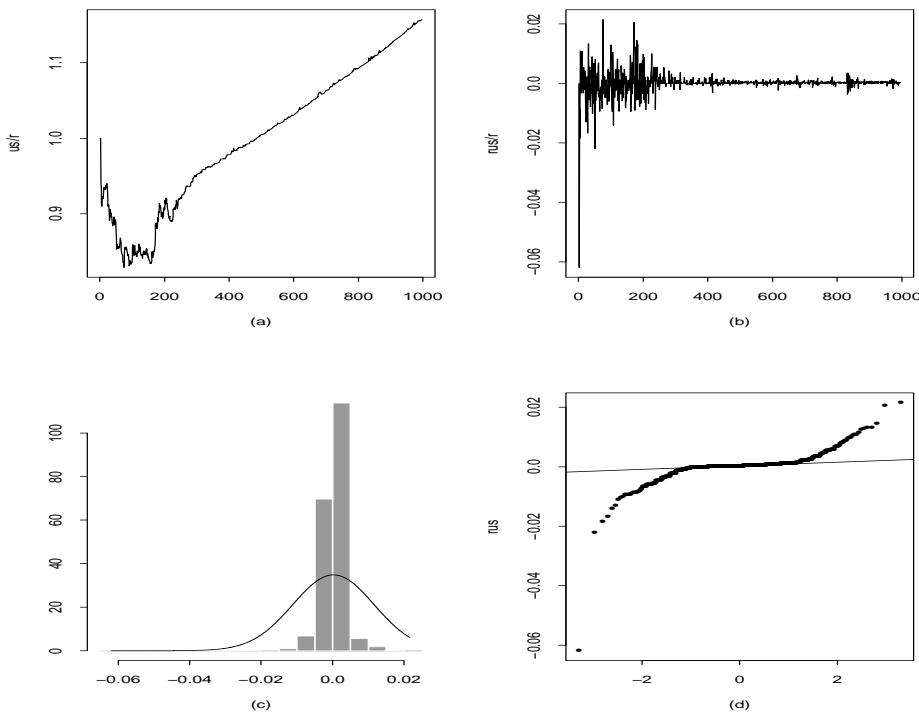


Figura 1.7: (a) Gráfico da série de taxas de câmbio USD/Real (b) Série dos retornos (c) Histograma com densidade ajustada (d) Gráfico  $Q \times Q$ .

Um modelo típico para a volatilidade é da forma

$$r_t = \mu_t + \sqrt{h_t} \varepsilon_t, \quad (1.31)$$

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 (1.29), (1.30) 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 eles citamos o EViews, o módulo S+FinMetrics do SPlus, o repositório de pacotes computacionais R, o MatLab e o STAMP.

Utilizaremos, preferencialmente, o S+FinMetrics, o EViews e o R, neste livro. Recomendamos aos leitores a consulta aos manuais desses pacotes ou a textos que

os utilizam de maneira sistemática, como Zivot e Wang (2006) 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.

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. As soluções da maioria dos exemplos do livro, usando pacotes do R, podem ser encontradas no sítio [rpubs.com/EconFin](http://rpubs.com/EconFin).

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)
> par(mfrow=c(1,1))
```

## 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 diárioss da Vale, de 31 de agosto de 1998 a 29 de setembro de 2010 (arquivo d-vale.98.10.dat), com  $T = 2.990$  observações:
- (a) Calcule as estatísticas : média, variância, coeficiente de assimetria e curtose, quartis, máximo e mínimo. Use algum programa, como o EViews ou o S+FinMetrics.
  - (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 da Vale, no começo de setembro de 1998, qual seria o valor do investimento no final de setembro de 2010? [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 diárioss da Petrobras (veja o arquivo d-petro98.10.dat).
7. Use a estatística (1.34), do Apêndice 1.B, para testar se os log-retornos diárioss da Vale têm uma distribuição normal.
8. Mesmo problema, para os log-retornos diárioss do Ibovespa (veja o arquivo d-ibv94.10.dat).
9. Mesmo problema, para os log-retornos diárioss 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  $e^x = 1 + x + x^2/2 + \dots$ ].
-

11. Mostre que, se  $X$  tem distribuição exponencial com parâmetro  $\beta$ , então o  $p$ -quantil é dado por  $Q_p = \frac{1}{\beta} \log\left(\frac{1}{1-p}\right)$ .
12. Na tabela abaixo estão os preços mensais de uma ação, durante dez meses.

mês	1	2	3	4	5	6	7	8	9	10
preço	14,2	15,0	12,0	19,0	12,5	13,0	17,0	17,5	14,8	16,0

Utilize um gráfico  $Q \times Q$  para avaliar se a distribuição dos preços é compatível com uma distribuição uniforme no intervalo  $[10, 20]$ .

13. Nos gráficos  $Q \times Q$  teóricos da Figura 1.8, feitos para verificar a suposição de normalidade, dê sua conclusão para cada caso e justifique-a. Comente sobre a distribuição em cada caso (simetria, caudas etc.), se você achar que não é normal.

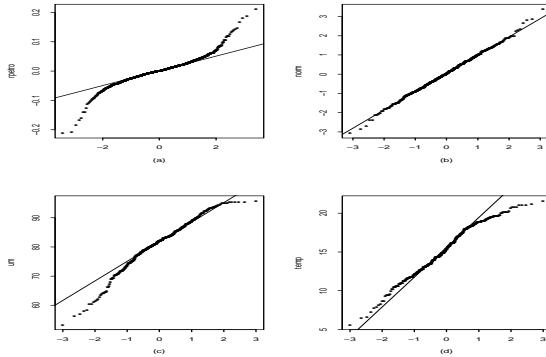


Figura 1.8: Gráficos  $Q \times Q$  para normalidade.

## 1.11 Apêndice 1.A. Distribuições Estáveis

Sabemos que se  $X_1, X_2, \dots$  são v.a.'s 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.'s 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

$$a_n X + b_n \xrightarrow{d} X_1 + X_2 + \dots + X_n,$$

onde  $\stackrel{d}{=}$  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 1.9, temos representadas a normal padrão e a Cauchy com densidade

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

com  $\sigma = 1$  e  $\mu = 0$ . Veja (1.32) abaixo para as definições de  $\sigma$  e  $\mu$ .

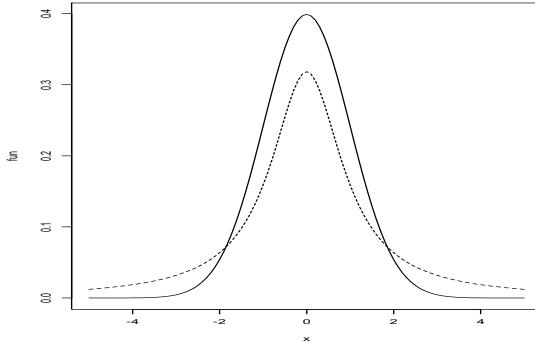


Figura 1.9: Distribuições estáveis: normal (linha cheia) e Cauchy (linha tracejada).

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 desse tipo.

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

$$\log \varphi(t) = it\mu - \sigma|t|^\alpha [1 - i\beta \operatorname{sgn}(t) \operatorname{tg}(\pi\alpha/2)], \quad (1.32)$$

para  $0 < \alpha < 1$ .

Nesta expressão,  $\mu$  é um parâmetro de localização, real,  $\sigma > 0$  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

$$\log \varphi(t) = i\mu t - \sigma|t| [1 - i\beta \operatorname{sgn}(t) \frac{2}{\pi} \log |t|]. \quad (1.33)$$

Indicaremos uma v.a com distribuição estável por  $S_\alpha(\sigma, \beta, \mu)$ .

Nas Figuras 1.10 e 1.11 temos diversos gráficos de densidades estáveis, para vários valores dos parâmetros.

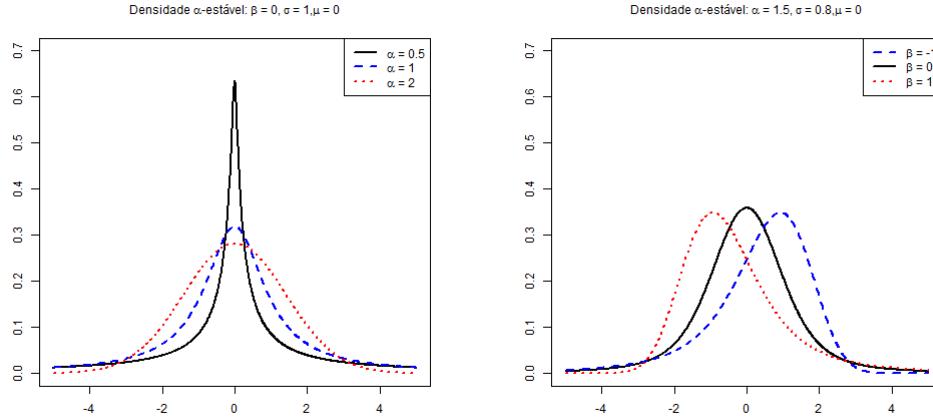


Figura 1.10: Parâmetro de estabilidade e parâmetro de assimetria.

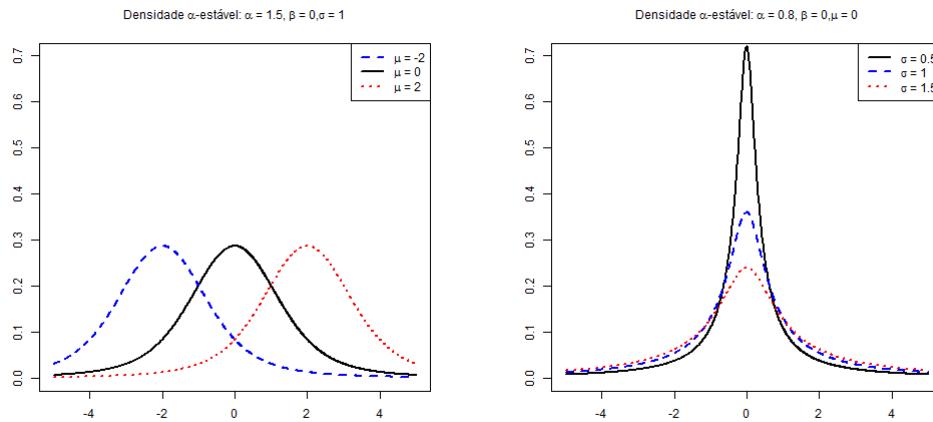


Figura 1.11: Parâmetro de posição e parâmetro de escala.

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  $\mu$ , 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 a direita (a esquerda).

Alguns outros fatos sobre distribuições estáveis:

- (i) Se  $X$  for estável, então  $X$  terá uma densidade limitada e contínua;
- (ii) Se  $X$  for estável, *simétrica*, então sua função característica será dada por  $\varphi(t) = e^{-c|t|^\alpha}$ .
- (iii) Se  $X$  for estável, então  $X$  será *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  $X_{n1} + \dots + X_{nn}$ .

A classe das distribuições estáveis inclui a possibilidade de contemplar caudas pesadas e assimetria. No campo das finanças, tais propriedades são particularmente importantes, pois, como vimos, os retornos financeiros apresentam esses fatos estilizados. Também, distribuições estáveis podem acomodar a característica de leptocurticidade, que não está presente, por exemplo, na distribuição  $t$  de Student. Outra característica importante é que a agregação de ativos de uma carteira com distribuições estáveis também terá distribuição estável.

A presença de heteroscedasticidade condicional pode ser uma causa da ocorrência de caudas pesadas nos retornos. Veja, por exemplo, De Vries (1991). Tradicionalmente, modelos GARCH com inovações  $t$  de Student são considerados para acomodar o excesso de curtose. No Capítulo 8 trataremos de uma generalização desses modelos, denominados modelos R-GARCH (*randomized GARCH*).

A ausência de densidades em forma fechada (exceto em três casos, normal, Cauchy e Lévy) impede a utilização de estimadores de máxima verossimilhança para a estimativa de parâmetros de uma distribuição estável. Do mesmo modo, a inexistência de momentos de ordem maior do que dois impede a utilização do método dos momentos. Métodos baseados em quantis e na função característica empírica são alternativas para esse propósito. Veja McCulloch (1986) e Koutrovelis (1981).

Um fato importante é que variáveis aleatórias com distribuição estável podem ser facilmente simuladas. Desse modo, a abordagem usada por Gouriéroux e Monfort (1996) pode ser uma solução para a estimativa de parâmetros de distribuições estáveis e de outros modelos baseados nessas distribuições.

Seja  $W$  uma v.a com distribuição exponencial de média 1 e seja  $U$  uma v.a uniformemente distribuída sobre  $[-\pi/2, \pi/2]$ . Além disso, sejam  $\zeta = \arctan(\beta \tan(\frac{\pi\alpha}{2})/\alpha)$  e  $\eta = [1 + \beta^2 \tan^2(\frac{\pi\alpha}{2})]^{1/(2\alpha)}$ . Então

$$X = \begin{cases} \eta \frac{\sin \alpha(\zeta+U)}{(\cos U)^{1/\alpha}} \left[ \frac{\cos(U-\alpha(U+\zeta))}{W} \right]^{\frac{1-\alpha}{\alpha}} & \text{se } \alpha \neq 1, \\ \frac{2}{\pi} [(\pi/2 + \beta U) \tan U] - \beta \ln \frac{\pi/2 W \cos U}{\frac{\pi}{2} + \beta U} & \text{se } \alpha = 1, \end{cases} \quad (1.34)$$

tem distribuição  $S_\alpha(1, \beta, 0)$ . Podemos obter uma v.a com distribuição  $S_\alpha(\sigma, \beta, \mu)$

por meio da fórmula

$$Z = \begin{cases} \sigma X + \mu, & \text{se } \alpha \neq 1, \\ \sigma X + \frac{2}{\pi} \beta \sigma \log \sigma + \mu, & \text{se } \alpha = 1. \end{cases} \quad (1.35)$$

Veja Weron e Weron (1995) para detalhes.

## 1.12 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 (1.24) deve ser igual a zero. Podemos usar, então, o resultado (1.28) 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 (1.25), 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, 1987), que combina esses dois testes, usando a estatística

$$S = \left( \frac{T}{6} \right) \hat{A}^2 + \left( \frac{T}{24} \right) (\hat{K} - 3)^2, \quad (1.36)$$

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 (1.34) 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 (1.34) 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.*

Nessas 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.'s) envolvidas sejam reais, mas elas podem ser complexas.

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

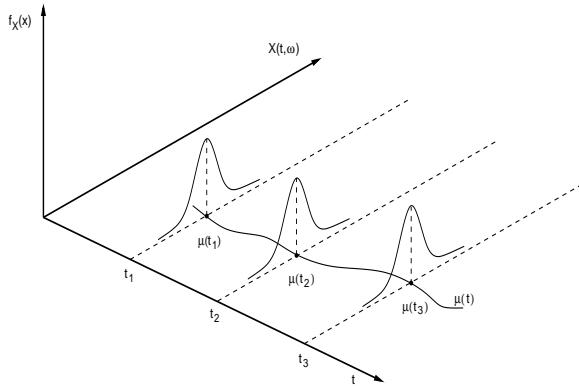


Figura 2.1: Processo estocástico como uma família de variáveis aleatórias.

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 ideia 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 (ou tempo) discreto*. Se  $\mathcal{T}$  for um intervalo de  $\mathbb{R}$  teremos um processo com *parâmetro (ou tempo) 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.

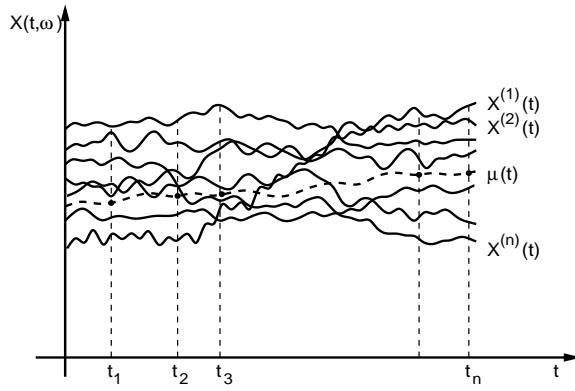


Figura 2.2: Processo estocástico como uma família de trajetórias

## 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 forem conhecidas as *distribuições finito-dimensionais* (2.1), para todo  $n \geq 1$ . Isto significa que, para  $n = 1$ , nós conhecemos as distribuições unidimensionais da v.a.  $X(t_1), t_1 \in \mathcal{T}$ , para  $n = 2$ , nós conhecemos as distribuições bidimensionais 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 essas 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.'s  $X(t_1), \dots, X(t_n)$ , 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 autocovariância* 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 autocovariância  $\gamma(t_1, t_2)$  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 autocovariância (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 curtoza.

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 unidimensionais 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 bidimensionais dependem de diferenças de tempos. De fato, para  $t_1, t_2 \in \mathcal{T}$ , temos por exemplo que  $\gamma(t_1, t_2) = \gamma(t_1 + t, t_2 + t)$  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 autocovariância* 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 por meio 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 2.3 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.'s  $X(t_1), X(t_2), \dots, X(t_n)$  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,  $t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ , o processo será escrito  $\{X_t, t \in \mathbb{Z}\}$ , 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 autocovariância do processo estacionário com tempo discreto será denotada por  $\gamma_\tau$ , ao passo que a do processo com tempo contínuo será denotada por  $\gamma(\tau)$ .

## 2.3 Propriedades da Função de Autocovariância

Seja  $\{X_t, t \in \mathbb{Z}\}$  um processo estacionário real com tempo 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  $\mathbb{Z}$ .

**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 prova deste fato, ver Cramér e Leadbetter (1967, p. 80).

Tipicamente, a f.a.c.v. de um processo estacionário tende a zero, para  $|\tau| \rightarrow \infty$ . A Figura 2.3 mostra esse 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.

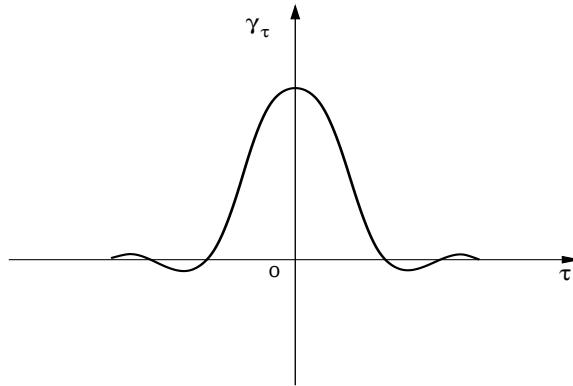


Figura 2.3: Função de autocovariância.

A *função de autocorrelação* (f.a.c.) do processo é definida por

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

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 continuidade de  $\gamma(\tau)$  para todo  $\tau$ .

**Prova.** Usando a desigualdade de Schwarz para duas v.a.'s 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 autocovariância  $\gamma_j$ ,

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

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.'s 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

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

e a variância por

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

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, nesta seção, alguns exemplos de processos estocásticos estacionários que são utilizados com frequência. Daremos ênfase aos processos autorregressivos (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 é adequado.

### Exemplo 2.1. Sequência Aleatória

Consideremos  $\{X_n, n = 1, 2, \dots\}$  uma sequência de v.a.'s definidas no mesmo espaço amostral  $\Omega$ . Aqui,  $\mathcal{T} = \{1, 2, \dots\}$  e temos um processo com parâmetro discreto, ou uma sequência aleatória. Para todo  $n \geq 1$ , podemos escrever

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

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 sequência  $\{X_n, n \geq 1\}$  de v.a.'s *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.'s  $X_1, X_2, \dots$  tiverem todas a mesma distribuição, teremos, então, uma sequência de v.a.'s 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.

### Exemplo 2.2. Ruído Branco

Uma forma mais fraca de uma sequência i.i.d. é dada pela definição a seguir.

**Definição 2.6.** Dizemos que  $\{\varepsilon_t, t \in \mathbb{Z}\}$  é um ruído branco com tempo discreto se as v.a.'s  $\varepsilon_t$  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.'s  $\varepsilon_t$  são independentes, elas também serão não correlacionadas. Uma sequência de v.a.'s i.i.d., como definida acima, é chamada um *processo puramente aleatório* ou *ruído branco forte*. O que definimos como ruído branco, seria chamado *ruído branco fraco*. Contudo, usaremos simplesmente a nomenclatura da Definição 2.6.

Ilustramos na Figura 2.4 a função de autocorrelação de um ruído branco. De agora em diante, vamos reservar a notação  $\{\varepsilon_t, t \in \mathbb{Z}\}$  para um ruído branco com tempo 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

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

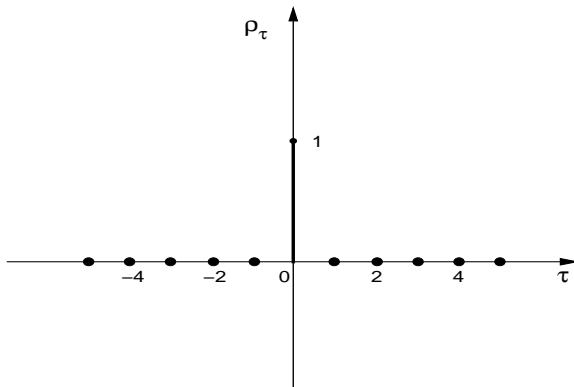


Figura 2.4: F.a.c. de um ruído branco.

### 2.5.1 Processos Autorregressivos

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

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \phi_2(X_{t-2} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t, \quad (2.23)$$

onde  $\mu, \phi_1, \dots, \phi_p$  são parâmetros reais e  $\varepsilon_t \sim RB(0, \sigma^2)$ . 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  $B$  por meio de  $B^s X_t = X_{t-s}, s \geq 1$ . Então (2.23) pode ser escrita

$$\phi(B) \tilde{X}_t = \varepsilon_t, \quad (2.24)$$

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

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

$$X_t = \phi X_{t-1} + \varepsilon_t. \quad (2.25)$$

Aqui,  $\phi(B) = 1 - \phi B$ . Por meio 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

$$X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}, \quad (2.26)$$

onde a convergência é em média quadrática. Logo, a condição  $|\phi| < 1$  é suficiente para  $X_t$  ser estacionário. Multiplicando ambos os membros de (2.25) por  $X_{t-\tau}$ , e tomado a esperança, obtemos

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

Mas, de (2.26), obtemos

$$\gamma_0 = \sigma_X^2 = \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma^2}{1 - \phi^2}, \quad (2.27)$$

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

$$\gamma_\tau = \frac{\sigma^2}{1 - \phi^2} \phi^{|\tau|}, \quad \tau \in \mathbb{Z}. \quad (2.28)$$

A f.a.c. de  $X_t$  é obtida de (2.28), ou seja,

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0} = \phi^{|\tau|}, \quad \tau \in \mathbb{Z}. \quad (2.29)$$

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

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

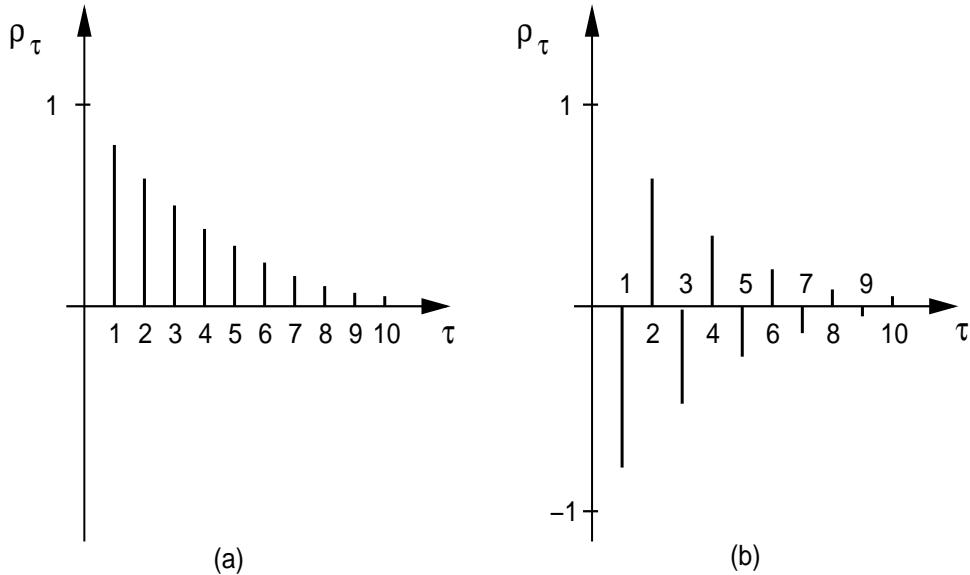


Figura 2.5: F.a.c. de um processo AR(1): (a)  $\phi = 0, 0.8$  (b)  $\phi = -0.8$ .

Procuremos solução para (2.23) na forma (2.26), isto é,

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}. \quad (2.30)$$

De (2.24), 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 (2.30) 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.

Pode-se demonstrar (ver Box et al., 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 (2.23) por  $X_{t-\tau}$  e tomindo valores esperados, obtemos

$$\sigma_X^2 = \frac{\sigma^2}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p}, \quad \text{para } \tau = 0, \quad (2.31)$$

$$\gamma_\tau = \phi_1 \gamma_{\tau-1} + \phi_2 \gamma_{\tau-2} + \dots + \phi_p \gamma_{\tau-p}, \quad \text{para } \tau > 0. \quad (2.32)$$

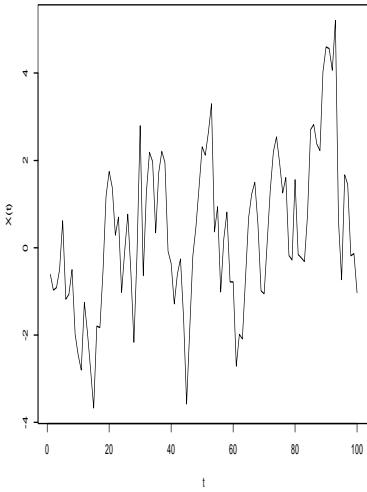


Figura 2.6: Processo AR(1) simulado,  $\phi = 0,8$ .

A mesma equação de diferenças é satisfeita por  $\rho_\tau$ , bastando dividir todos os termos de (2.32) por  $\gamma_0$ .

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

$$\gamma_\tau = A_1 G_1^\tau + A_2 G_2^\tau + \dots + A_p G_p^\tau, \quad (2.33)$$

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 (2.32), obtemos

$$\Gamma_p \phi_p = \gamma_p, \quad (2.34)$$

onde  $\Gamma_p = [\gamma_{ij}]$ , 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 (2.34) pode ser utilizada para obter estimadores dos parâmetros  $\phi_j$ 's, substituindo-se as f.a.c.v.'s por suas estimativas. Esses estimadores são chamados *estimadores de Yule-Walker*.

Uma análise de (2.33) nos permite concluir que a f.a.c.v. de um processo autoregressivo de ordem  $p$  é uma mistura de exponenciais (correspondentes às raízes

$G_i$  reais) e/ou senoides (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  $\{X_t, t \in \mathbb{Z}\}$  é um processo de médias móveis de ordem  $q$ , denotado por MA(q), se satisfizer à equação de diferenças

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (2.35)$$

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

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

$$\sigma_X^2 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2). \quad (2.36)$$

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,

$$\gamma_\tau = \begin{cases} \sigma^2(-\theta_\tau + \theta_1 \theta_{\tau+1} + \dots + \theta_q \theta_{q-\tau}), & \text{se } \tau = 1, \dots, q \\ 0, & \text{se } \tau > q \\ \gamma_{-\tau}, & \text{se } \tau < 0. \end{cases} \quad (2.37)$$

De (2.36) e (2.37) obtemos a f.a.c. do processo MA(q):

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

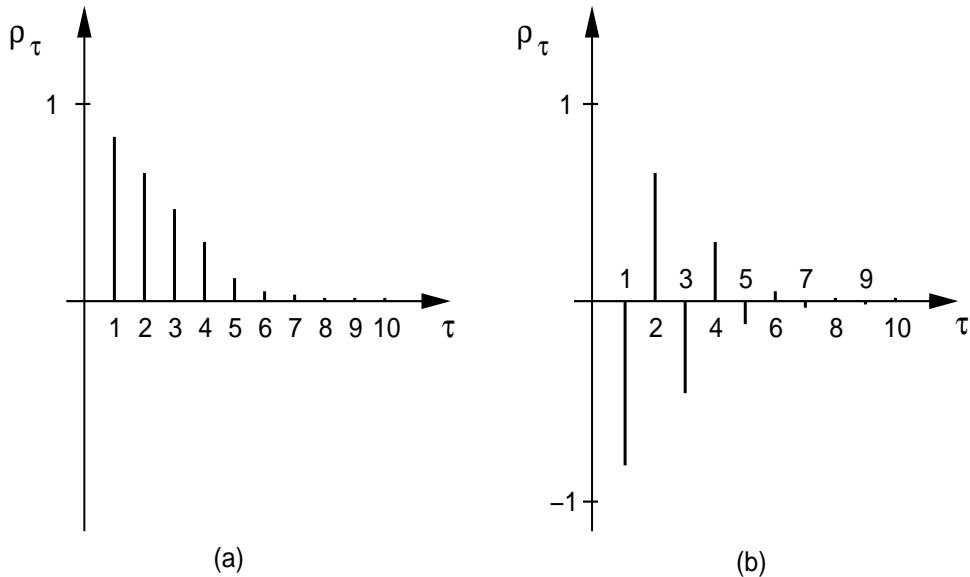


Figura 2.7: F.a.c. de dois processos AR(2) (a)  $\phi_1 = 0,5$ ,  $\phi_2 = 0,3$  (b)  
 $\phi_1 = 1,0$ ,  $\phi_2 = -0,89$ .

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),

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1}, \quad (2.39)$$

obtemos

$$\text{Var}(X_t) = \sigma_X^2 = \sigma^2(1 + \theta^2),$$

$$\rho_\tau = \begin{cases} \frac{-\theta}{1+\theta^2}, & \text{se } \tau = \pm 1 \\ 0, & \text{se } |\tau| > 1. \end{cases} \quad (2.40)$$

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 (2.35) pode ser escrito

$$X_t = \theta(B)\varepsilon_t. \quad (2.41)$$

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

$$X_t = -\theta X_{t-1} - \theta^2 X_{t-2} - \dots + \varepsilon_t, \quad (2.42)$$

se  $|\theta| < 1$ , para que a série do lado direito de (2.42) converja. Nesta equação, temos  $X_t$  escrito como um processo autoregressivo de ordem infinita. Dizemos que  $|\theta| < 1$  é uma *condição de invertibilidade* para o processo MA(1).

De modo geral, o processo (2.35) poderá ser escrito na forma

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + \varepsilon_t, \quad (2.43)$$

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

A relação (2.43) pode ser escrita

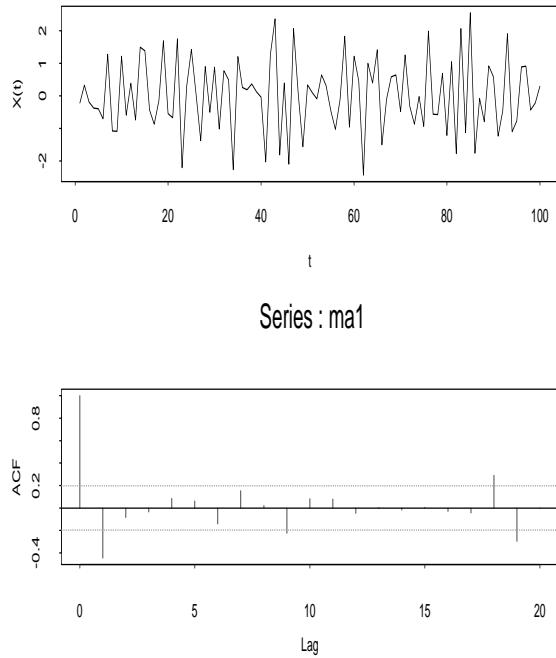
$$\pi(B)X_t = \varepsilon_t, \quad (2.44)$$

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

$$X_t = \varepsilon_t - 0,8\varepsilon_{t-1}, \quad \varepsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1). \quad (2.45)$$

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 autocorrelações 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 Autorregressivos e de Médias Móveis

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

$$X_t - \mu = \phi_1(X_{t-1} - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}, \quad (2.46)$$

onde  $\varepsilon_t \sim RB(0, \sigma^2)$ . Segue-se que a média do processo é  $\mu$ . Usando os operadores autorregressivo e de médias móveis, definidos anteriormente, podemos escrever (2.46) na forma

$$\phi(B)\tilde{X}_t = \theta(B)\varepsilon_t, \quad (2.47)$$

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

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

$$X_t = \phi X_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}. \quad (2.48)$$

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 (2.43), com pesos  $\pi_j = \theta^{j-1}(\theta - \phi)$ ,  $j \geq 1$ .

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

$$\begin{aligned}\gamma_\tau &= \phi_1\gamma_{\tau-1} + \phi_2\gamma_{\tau-2} + \dots + \phi_p\gamma_{\tau-p} + \gamma_{X\varepsilon}(\tau) \\ &\quad - \theta_1\gamma_{X\varepsilon}(\tau-1) - \dots - \theta_q\gamma_{X\varepsilon}(\tau-q),\end{aligned}\tag{2.49}$$

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

$$\gamma_\tau = \phi_1\gamma_{\tau-1} + \phi_2\gamma_{\tau-2} + \dots + \phi_p\gamma_{\tau-p}, \quad \tau > q.\tag{2.50}$$

A conclusão é que as autocovariâncias (e, portanto, as autocorrelações, 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 autorregressivos.

Para o caso do modelo (2.48), 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

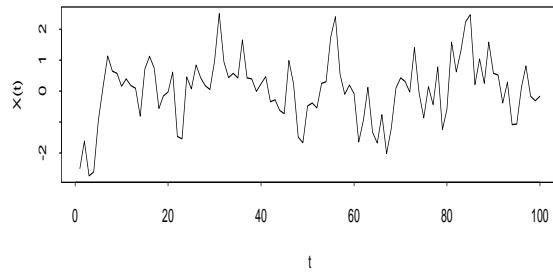
$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (2.51)$$

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

$$\sigma_X^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2. \quad (2.52)$$

Também, de (2.51), vemos que  $E\{X_t\} = 0$  e para  $\tau > 0$ ,

$$\gamma_\tau = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\tau}, \quad (2.53)$$



Series : arma

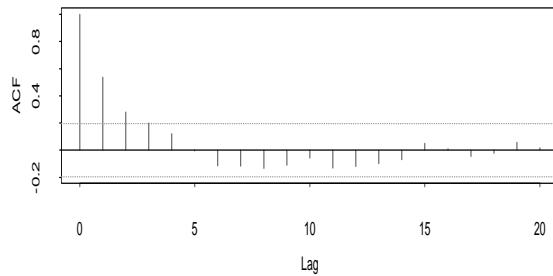


Figura 2.9: Processo ARMA(1,1) simulado,  $\phi = 0,8, \theta = 0,3$  e f.a.c.

admitindo-se que a série do segundo membro de (2.53) 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, essa é a condição de estacionariedade para o PLG.

De (2.52) e (2.53) 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 (2.51), como no teorema de Wold abaixo.

É imediato verificar que um processo MA(q) é caso particular de (2.51), com  $\psi_j = 0, j > q$ . Também, o processo AR(1) é obtido de (2.51), colocando-se  $\psi_j = \phi^j$ . Não é difícil verificar que um processo AR(p) genérico é caso particular do PLG.

Pelo que vimos aqui, um processo AR(p), p finito, pode ser escrito como um processo de médias móveis de ordem infinita e um processo MA(q), q finito, pode ser escrito como um processo autorregressivo 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*

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad \psi_0 = 1, \tag{2.54}$$

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

Iniciamos esta seção com um processo estocástico importante, o *passeio aleatório*. Passeios aleatórios têm grande importância em econometria e finanças. Uma hipótese célebre é que os preços de ativos financeiros seguem um passeio aleatório.

**Exemplo 2.4.** *Passeio Aleatório*

Considere uma sequência aleatória  $\{\varepsilon_t, t \geq 1\}$ , de v.a.'s i.i.d.  $(\mu_\varepsilon, \sigma_\varepsilon^2)$ . Defina a sequência

$$X_t = \varepsilon_1 + \dots + \varepsilon_t. \quad (2.55)$$

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 autocovariância de  $X_t$  depende de  $t_1$  e  $t_2$ . O processo (2.55) 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  $\varepsilon_t = X_t - X_{t-1}$ , esse processo tem *incrementos ortogonais ou não correlacionados*.

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 (autorregressivos integrados e de médias móveis) etc. No caso de não estacionariedade em variância, certas transformações (como a logaritmica) 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 (2.54) uma tendência determinística, 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 raíz unitária, da forma

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

com  $\psi(1) \neq 0$ . Esse 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 2.5.** Considere um caso especial de (2.56),

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

sendo  $\varepsilon_t$  um  $\text{RB}(0, \sigma^2)$ . 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) Extraindo-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  $\theta_0 = d! \beta_d$ . Esse modelo é, novamente, estacionário, mas não invertível.

**Exemplo 2.6.** 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

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

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  $i = 1, \dots, p$ . Nesse 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

$$(1 - B)\varepsilon_t = \frac{\theta(B)}{(1 - G_2B) \cdots (1 - G_pB)} a_t = \psi^*(B)a_t, \quad (2.63)$$

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 autorregressivo  $\phi(B)$  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 a seguir:

**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:

**Definição 2.8.** Se  $\Delta^d X_t \sim ARMA(p,q)$  dizemos que  $X_t$  segue um modelo  $ARIMA(p,d,q)$ : autorregressivo 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,

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

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 em que  $\phi = 1$ , teremos um passeio aleatório, estudado no Exemplo 2.6,

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

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 sequência suave, mas não estacionária.

Vimos aqui 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 2.7.** 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 2.8.** 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  $\mu + P_{t-1}$  e variância  $\sigma^2$ . Esse 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,

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

Esse modelo supõe que a variância seja constante. Uma suposição mais adequada é admitir que a variância (condicional) 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 (adotando a notação da Seção 1.8)

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

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 7.

## 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  $W = \{W(t), 0 \leq t \leq 1\}$  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.'s  $W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_k) - W(t_{k-1})$  são independentes;
- (c) para quaisquer  $s, t$  e  $\tau$  no intervalo  $[0, 1]$ , as v.a.'s  $W(t) - W(s)$  e  $W(t + \tau) - W(s + \tau)$  têm a mesma distribuição;
- (d) para todo  $0 \leq t \leq 1$ ,  $W(t) \sim \mathcal{N}(0, t)$ ;
- (e) as trajetórias de  $W(t)$  são contínuas, com probabilidade um.

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  $\text{Var}(W(t)) = t$ ,  $0 \leq t \leq 1$ . 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

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

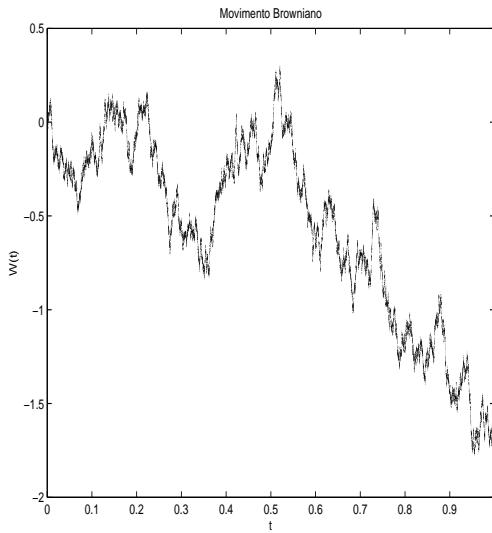


Figura 2.10: Movimento Browniano simulado.

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 elas 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, em decorrência dos 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 esse processo é *autossimilar*, con-

forme a definição a seguir.

**Definição 2.10.** Um processo estocástico  $X = \{X(t), t \geq 0\}$  é  $f$ -autossimilar, para 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  $t_1, \dots, t_n \geq 0, n \geq 1$ .

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

**Fato 3.** O MBP é  $0,5$ -autossimilar, 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  $\tau$  e para quaisquer  $t_1, \dots, t_n$ , no intervalo  $[0, 1]$ ,  $n \geq 1$ .

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 2.9.** *Movimento Browniano Geral.* 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,

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

Então,  $X(t)$  será gaussiano, com  $E(X(t)) = \mu t$  e  $\gamma_X(t, s) = \sigma^2 \min(t, s)$ , para  $s, t \geq 0$ . O parâmetro  $\mu$  é chamado o *drift*, enquanto  $\sigma$  é chamado “volatilidade”.

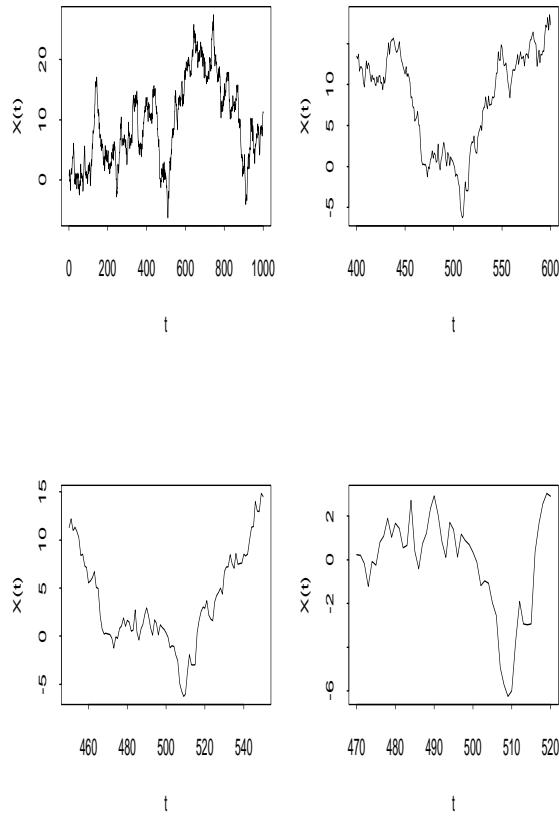


Figura 2.11: Um processo autossimilar.

**Exemplo 2.10.** *Movimento Browniano Geométrico.* Como vimos no Exemplo 2.8, 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

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

Segue-se que  $\log\{X(t)\}$  é um Movimento Browniano Geral. 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 sequência de v.a.'s 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 concluimos que a sequência 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 (2.85).

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, esse 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  $0 \leq r_1 \leq \dots \leq r_k \leq 1$ ,

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

(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  $\lambda \rightarrow \infty$ .

Nessas 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 essa 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 *submartingale* e se tivermos  $E(X_t | \mathcal{F}_s) \leq X_s$ , temos um *supermartingale*.

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

Se  $\mathcal{T} = \mathbb{Z}$  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 2.11.** (a) Se  $X_1, X_2, \dots$  são v.a.’s 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 submartingale e  $Z_n = \sum_{i=1}^n (X_i - \mu_i)$  é um martingale.

(c) Se  $X_{n+1} = X_n + a_n$ , onde  $a_n \sim RB(0, \sigma^2)$ , 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 dada 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 sequência 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 submartingale. Se  $\varphi$  for ainda crescente e  $X_n$  um submartingale, então  $\varphi(X_n)$  será também um submartingale.
5. O resultado a seguir é um teorema de convergência para submartingales.

**Teorema 2.2.** (Doob) Suponha que  $\{X_n, \mathcal{F}_n, n \geq 1\}$  seja um submartingale, 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 sequência de variáveis que passamos a definir.

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

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

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,  $\bar{Y}_N = \sum_{n=1}^N Y_n / N$  e temos as seguinte 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,  $N^{1/2}\bar{Y}_N \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$ .

## 2.9 Problemas

1. Use a equação (2.8) para provar que, se  $X(t)$  for estritamente estacionário, então  $\mu(t)$  e  $\sigma^2(t)$  são constantes.
2. 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.'s 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?
3. Considere as observações:

$t$	1	2	3	4	5	6	7
$Z_t$	15	19	13	17	22	18	20

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

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

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

- (a) Obtenha a média do processo  $Z_t$ ;
- (b) Calcule  $\gamma_\tau$ ,  $\tau = 0, \pm 1, \pm 2, \dots$
- (c) Calcule  $\rho_\tau$ ,  $\tau = 0, \pm 1, \dots$  e faça o seu gráfico.
5. Suponha  $\{a_t, t = 1, 2, \dots\}$  uma sequência de v.a.'s 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? (Neste e nos problemas seguintes, estacionário significa fracamente 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 autocovariância  $\gamma_Z$ . Um novo processo é definido por  $Y_t = Z_t - Z_{t-1}$ . Obtenha a média e a função de autocovariância 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-ibv94.10.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) autocorrelações 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 2.10. 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 martingale com respeito à  $\sigma$ -álgebra natural  $\mathcal{F}_t = \sigma\{W(s), s \leq t\}$ .

16. Mesmo problema para o processo  $W(t)^2 - t$ .

17. **Processos de Lévy.** Dizemos que um processo estocástico  $\{X_t\}$  é *càdlàg* (em francês, “continue à droite avec limite à gauche”) se suas trajetórias são (em quase toda a parte) contínuas a direita e têm limites a esquerda.

Um processo de Lévy  $\{X_t\}$  é um processo *càdlàg* com incrementos independentes e estacionários, começando na origem, isto é,  $P(X_0 = 0) = 1$ .

Mostre que o processo de Poisson e o Movimento Browniano são processos de Lévy.

18. Mostre que, para um modelo AR(2) estacionário, a representação em médias móveis infinita é tal que  $\psi_0 = 1, \psi_1 = \phi_1$  e  $\psi_i = \phi_1\psi_{i-1} + \phi_2\psi_{i-2}, i \geq 2$ .

19. Considere os seguintes modelos (todos com  $\mu = E(X_t) = 0$ ):

- (a)  $X_t = 0, 6X_{t-1} = \varepsilon_t$ ;
- (b)  $X_t = \varepsilon_t + 0, 8\varepsilon_{t-1}$ ;
- (c)  $X_t = 0, 3X_{t-1} - 0, 6X_{t-2} + \varepsilon_t$ ;
- (d)  $X_t = 0, 4X_{t-1} = \varepsilon_t - 0, 3\varepsilon_{t-1} + 0, 8\varepsilon_{t-2}$ ;
- (e)  $X_t = 0, 3\varepsilon_{t-1} + 0, 6\varepsilon_{t-2} + \varepsilon_t$ .

Obtenha os primeiros três pesos  $\psi_j$  e  $\pi_j$  para cada modelo.

20. Considere o modelo  $X_t + bX_{t-3} = \varepsilon_t$ . Obtenha a condição de estacionariedade para  $X_t$ . Encontre a representação de  $X_t$  na forma de médias móveis infinita. Obtenha a f.a.c. de  $X_t$ .

---

# Modelos ARIMA

## 3.1 Introdução

Neste capítulo, trataremos de modelar séries temporais por meio de processos autorregressivos 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  $\varepsilon_t \sim RB(0, \sigma_\varepsilon^2)$ . Com os operadores autorregressivos 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

$$X_t = \theta_0 + \phi_1X_{t-1} + \dots + \phi_pX_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q},$$

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  $\{X_t, t \in Z\}$  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) estimativa;

(iii) diagnóstico.

Estudaremos, a seguir, com algum detalhe, cada um desses estágios. Para mais informações o leitor pode consultar Morettin e Toloi (2006) e Box et al. (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 autocorrelações e autocorrelações parciais estimadas (veja abaixo o conceito de autocorrelação parcial), que esperamos representem adequadamente as respectivas quantidades teóricas, que são desconhecidas.

Lembremos (Seção 2.3) que a f.a.c.  $\rho_j$  é estimada por

$$r_j = \frac{c_j}{c_0}, \quad j = 0, 1, \dots, T-1, \quad (3.4)$$

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

$$c_j = \frac{1}{T} \sum_{t=1}^{T-j} [(X_t - \bar{X})(X_{t+j} - \bar{X})], \quad j = 0, 1, \dots, T-1, \quad (3.5)$$

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

Como veremos adiante, será necessário uma verificação 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

$$\text{Var}(r_j) \simeq \frac{1}{T} \sum_{v=-\infty}^{\infty} [\rho_v^2 + \rho_{v+j}\rho_{v-j} - 4\rho_j\rho_v\rho_{v-j} + 2\rho_v^2\rho_j^2]. \quad (3.6)$$

Para um processo em que as autocorrelações são nulas para  $v > q$ , todos os termos do lado direito de (3.6) anulam-se para  $j > q$ , exceto o primeiro, obtendo-se

$$\text{Var}(r_j) \simeq \frac{1}{T} \left[ 1 + 2 \sum_{v=1}^q \rho_v^2 \right], \quad j > q. \quad (3.7)$$

Como desconhecemos as autocorrelações  $\rho_v$ , substituímo-las por  $r_v$ , obtendo uma estimativa para (3.7),

$$\widehat{\sigma^2}(r_j) \simeq \frac{1}{T} \left[ 1 + 2 \sum_{v=1}^q r_v^2 \right], \quad j > q. \quad (3.8)$$

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 (3.7) (Jenkins e Watts, 1968, p. 187). Assim, pode-se construir um intervalo de confiança aproximado para as autocorrelações,

$$r_j \pm t_\gamma \cdot \widehat{\sigma}(r_j), \quad (3.9)$$

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, aproximadamente, a  $\gamma = 0,95$ , de modo que podemos considerar  $\rho_j$  significativamente diferente de zero se

$$|r_j| > 2\widehat{\sigma}(r_j), \quad j > q. \quad (3.10)$$

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

- (i) um processo AR( $p$ ) tem f.a.c. que decai de acordo com exponenciais e/ou senoides 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 senoides amortecidas após o “lag”  $q - p$ .

Essas 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 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 et al. (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

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix} \quad (3.11)$$

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

$$\begin{aligned} \phi_{11} &= \rho_1, \\ \phi_{22} &= \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \\ \phi_{33} &= \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}} \end{aligned}$$

e, em geral,

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

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

A quantidade  $\phi_{kk}$ , encarada como função de  $k$ , é chamada *função de autocorrelação 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

---


$$\begin{aligned} &\text{Corr}(X_t - \phi_{11}X_{t-1}, X_{t-2} - \phi_{11}X_{t-1}) \\ &= \text{Corr}(X_t - \rho_1X_{t-1}, X_{t-2} - \rho_1X_{t-1}) \\ &= \frac{\text{Cov}(X_t - \rho_1X_{t-1}, X_{t-2} - \rho_1X_{t-1})}{(\text{Var}(X_t - \rho_1X_{t-1})\text{Var}(X_{t-2} - \rho_1X_{t-1}))^{1/2}} \end{aligned}$$


---

$$\begin{aligned}
&= \frac{\gamma_2 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0}{\gamma_0 - 2\rho_1\gamma_1 + \rho_1^2\gamma_0} \\
&= \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_{22}.
\end{aligned}$$

Será necessário estimar a f.a.c.p. de um processo AR, MA ou ARMA. Uma maneira consiste em estimar, sucessivamente, modelos autorregressivos 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 (3.11) 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 (3.11) utilizando o algoritmo de Durbin-Levinson. Veja Morettin (1984) para detalhes.

Pode-se demonstrar (veja Box et al., 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 senoides 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( $p$ ),

$$\text{Var}(\hat{\phi}_{jj}) \simeq \frac{1}{T}, \quad j > p, \tag{3.12}$$

de modo que

$$\hat{\sigma}(\hat{\phi}_{jj}) \simeq \frac{1}{\sqrt{T}}, \quad j > p. \quad (3.13)$$

Além disso, para  $T$  grande e sob a hipótese que o processo seja AR( $p$ ),  $\hat{\phi}_{jj}$  terá distribuição aproximadamente normal, com média zero e variância (3.12), de modo que consideraremos  $\hat{\phi}_{jj}$  significativamente diferente de zero se

$$|\hat{\phi}_{jj}| > \frac{2}{\sqrt{T}}, \quad j > p. \quad (3.14)$$

**Exemplo 3.2.** Na Tabela 3.1, temos as estimativas das autocorrelações e de seus respectivos desvios padrões, bem como as autocorrelações 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  $\hat{\phi}_{jj}$  estão indicados em linha pontilhada.

Note que os valores estimados podem diferir bastante dos valores verdadeiros. Por exemplo, no caso do modelo AR(1),  $\phi = 0,8$  e  $\rho_1 = 0,8$ ,  $\rho_2 = 0,64$  etc. Para o modelo MA(1),  $\theta = 0,8$  e  $\rho_1 = -\theta/(1+\theta^2) = -0,488$ , sendo as demais nulas. Para o 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.

### 3.2.1 Procedimento de Identificação

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

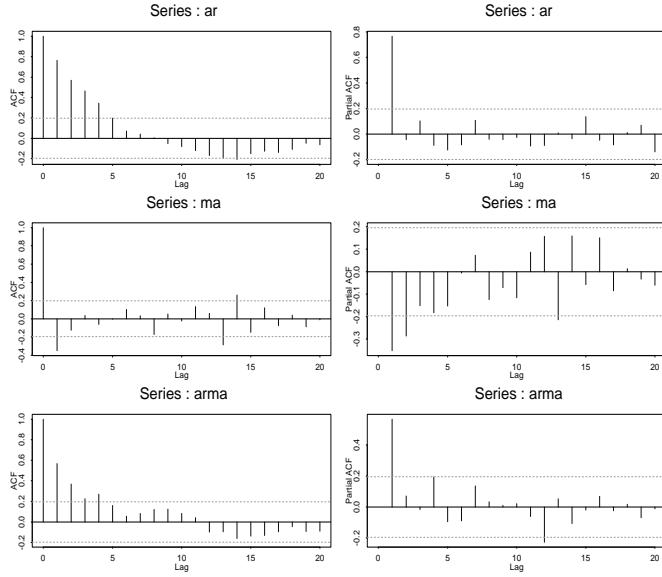


Figura 3.1: Autocorrelações e autocorrelações parciais amostrais para as séries simuladas AR(1), MA(1) e ARMA(1,1).

- verificar se existe necessidade de uma transformação na série original, com o objetivo de estabilizar sua variância;
- 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 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. Nesse estágio, a utilização de um teste para verificar a existência de raízes unitárias no polinômio autorregressivo, pode ser de grande utilidade. O teste de Dickey e Fuller (1979) será estudado no Capítulo 4;
- identificar o processo ARMA( $p, q$ ) resultante por meio da análise das autocorrelações e autocorrelações parciais estimadas, cujos comportamentos devem imitar os comportamentos das respectivas quantidades teóricas. Esses comportamentos foram abordados no Capítulo 2, e um resumo das propriedades destes modelos encontra-se na Tabela 3.2, extraída de Box et al. (1994).

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

$$\rho_j = A_1 G_1^j + A_2 G_2^j + \cdots + A_p G_p^j, \quad j > q - p, \quad (3.15)$$

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 (3.15) que,

se nenhuma raiz estiver muito próxima do círculo unitário, as autocorrelações  $\rho_j$  decairão para zero, para valores moderados de  $j$ .

Tabela 3.1: Autocorrelações amostrais e respectivos desvios padrões e autocorrelações parciais amostrais para as séries simuladas AR(1), MA(1) e ARMA(1,1)

lag	AR(1)			MA(1)			ARMA(1,1)			
	$j$	$r_j$	$\hat{\sigma}(r_j)$	$\hat{\phi}_{jj}$	$r_j$	$\hat{\sigma}(r_j)$	$\hat{\phi}_{jj}$	$r_j$	$\hat{\sigma}(r_j)$	$\hat{\phi}_{jj}$
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		

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, nesse estágio do procedimento, é evitar excesso de diferenças. Na prática,  $d = 0, 1$  ou  $2$ , e é suficiente inspecionar as primeiras 15 ou 20 autocorrelações 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 a 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

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

o que mostra que só  $\phi_{11} \neq 0$ ; os valores de  $\{r_j\}$  sugerem um decaimento e, portanto,  $\{\hat{\phi}_{jj}\}$  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

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

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 expon.	somente $\rho_1 \neq 0$
comportamento de $\phi_{kk}$	somente $\phi_{11} \neq 0$	decaim. exp. dom.
região de admissibilidade	$-1 < \phi < 1$	$-1 < \theta < 1$
Ordem	(2, $d$ , 0)	(0, $d$ , 2)
comportamento de $\rho_k$	mistura de expon. ou senoides amortecidas	só $\rho_1 \neq 0$ e $\rho_2 \neq 0$
comportamento de $\phi_{kk}$	só $\phi_{11} \neq 0$ e $\phi_{22} \neq 0$	dom. por mist. exp. ou senoides amortecidas
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 expon. após o lag 1	
região de admissibilidade	$-1 < \phi < 1, \quad -1 < \theta < 1$	

**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 logarítmica

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

A Figura 3.2 apresenta a série original e a série transformada. A Tabela 3.4 e a Figura 3.3 apresentam os valores e as correspondentes representações gráficas das funções de autocorrelação e autocorrelação parcial da série  $Y_t$  e de suas diferenças de ordens 1 e 2.

Analizando 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. Esse 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  $d = 1$ .

Analizando 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

$$\begin{aligned}\bar{W} &= 0,02116, \\ \text{Var}(\bar{W}) &= \frac{c_0(1 + r_1)}{n(1 - r_1)} = \frac{0,00014(1 + 0,49)}{113(1 - 0,49)} = 3,62 \times 10^{-6},\end{aligned}$$

ou seja,  $\hat{\sigma}(\bar{W}) = 0,0019$ . 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.

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

$$(1 - B)(1 - \phi_1 B)Y_t = \theta_0 + \varepsilon_t, \quad (3.16)$$

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

**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

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

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)	
$\frac{c_0(1 + r_1)(1 - 2r_1^2 + r_2)}{n(1 - r_1)(1 - r_2)}$	$\frac{c_0(1 + 2r_1 + 2r_2)}{n}$	

Tabela 3.4: Autocorrelações e autocorrelações 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

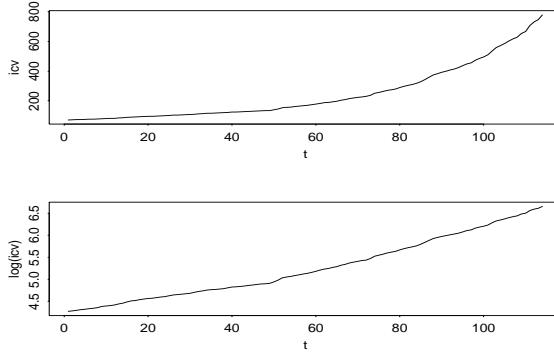


Figura 3.2: Gráfico das séries ICV e  $\ln(\text{ICV})$ .

O gráfico da série, as f.a.c. e f.a.c.p. amostrais estão na Figura 3.5, mostrando que essa 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 ideia é escolher as ordens  $k$  e  $l$  que minimizem a quantidade

$$P(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k + l) \frac{C(T)}{T}, \quad (3.17)$$

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.

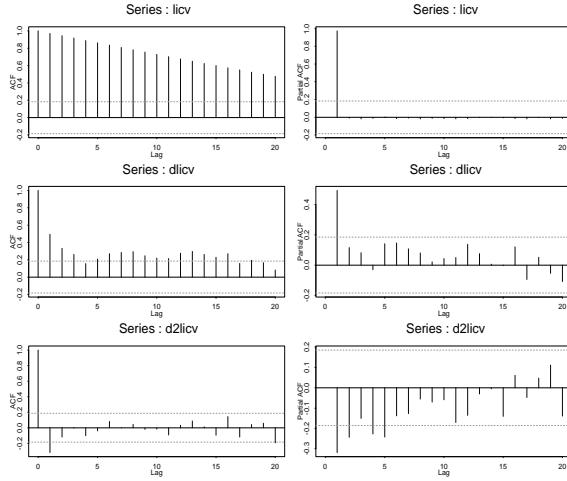


Figura 3.3: (a) Autocorrelações e (b) autocorrelações parciais das séries  $\ln(\text{ICV})$ ,  $(1 - B)\ln(\text{ICV})$  e  $(1 - B)^2\ln(\text{ICV})$ .

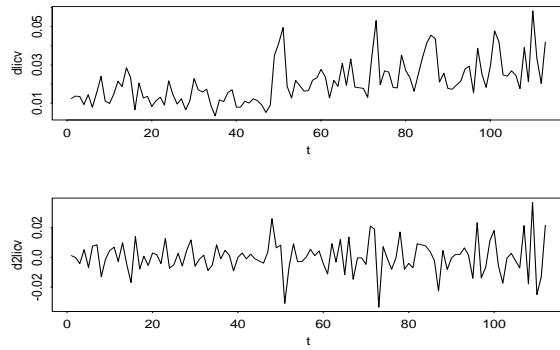


Figura 3.4: Gráfico das séries  $(1 - B)\ln(\text{ICV})$  e  $(1 - B)^2\ln(\text{ICV})$ .

A quantidade  $(k + l)C(T)/T$ , denominada termo penalizador, aumenta quando o número de parâmetros aumenta, enquanto a variância residual  $\hat{\sigma}_{k,l}^2$  diminui. Assim, minimizar (3.17) 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.

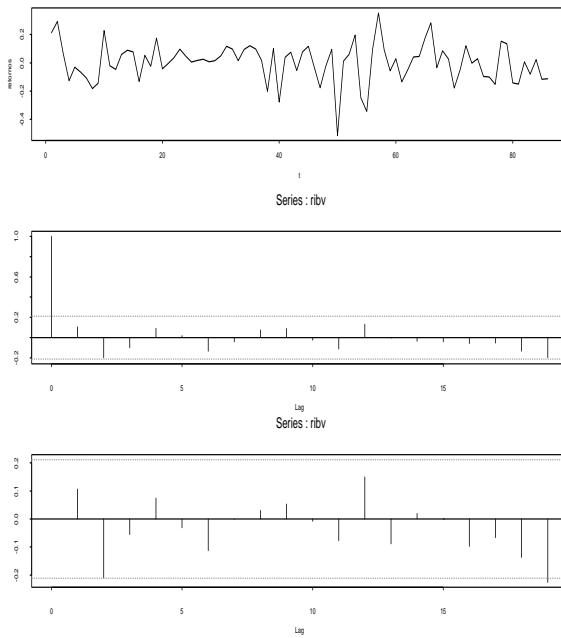


Figura 3.5: Retornos Ibovespa: (a) gráfico da série; (b) f.a.c. amostral; (c) f.a.c.p. amostral.

#### A. Critério de Informação AIC

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

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

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).

O que se faz, então, é estipular valores limites superiores  $K$  e  $L$  para  $k$  e  $l$ , e calcular (3.18) 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

$$\text{AIC}(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{T}, \quad k \leq K. \quad (3.19)$$

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

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

e, utilizando simulações, mostram que essa 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), Rissanem (1978) e Schwarz (1978) sugerem minimizar o Critério de Informação Bayesiano, dado por

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

na qual FV é a função de verossimilhança.

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

$$\text{BIC}(k, l) = \ln \hat{\sigma}_{k,l}^2 + (k+l) \frac{\ln T}{T}. \quad (3.20)$$

Hannan (1980) mostra que, sob determinadas condições, as estimativas  $k$  e  $l$  que minimizam a expressão (3.20) são fortemente consistentes.

**Exemplo 3.6.** Na Figura 3.6, temos o gráfico da série de retornos  $r_t$  diários da Petrobras, de 18/08/1998 a 29/09/2010, com  $T = 2.998$  observações, suas f.a.c. e f.a.c.p. amostrais; essas indicam que a série é autocorrelacionada e que um modelo autorregressivo é apropriado. Provavelmente, um modelo AR(3) é suficiente, mas vamos fixar uma ordem máxima  $K = 15$ . Na Tabela 3.5, temos alguns valores de AIC, BIC e do logaritmo da verossimilhança para cada modelo estimado. Como os critérios escolhem ordens diferentes, vamos nos fixar no AIC, e ajustar um AR(15). Estimando este modelo, somente os coeficientes  $\phi_1, \phi_2, \phi_3, \phi_{10}$  e  $\phi_{15}$  são significativos (veja a Seção 3.3), ou seja, o modelo final é

$$r_t = \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \phi_{10} r_{t-10} + \phi_{15} r_{t-15} + \varepsilon_t.$$

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

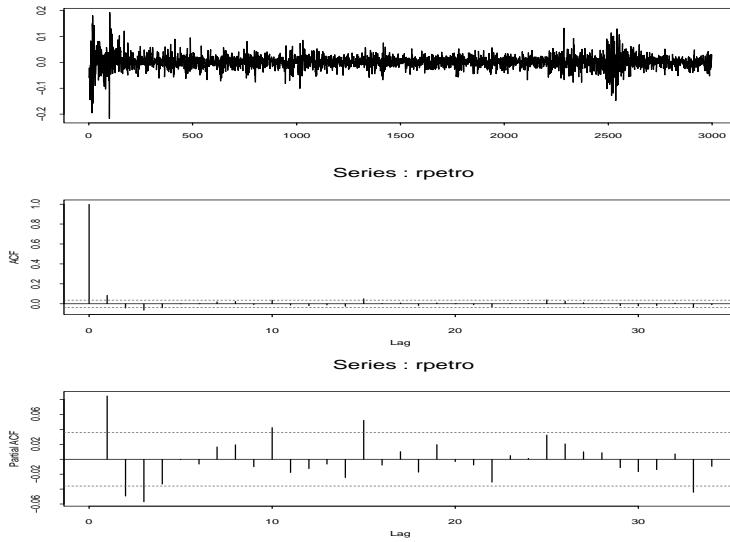


Figura 3.6: Retornos Petrobras: (a) gráfico da série (b) f.a.c. amostral  
(c) f.a.c.p. amostral.

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

Modelo	AIC	BIC	log-verossimilhança
AR(1)	-4,40647	-4,40447	6.604,10
AR(2)	-4,40923	-4,40522	6.607,03
AR(3)	-4,41287	-4,40685	6.611,27
AR(4)	-4,41321	-4,40557	6.611,14
AR(7)	-4,42430	-4,41025*	6.623,55
AR(8)	-4,42376	-4,40770	6.621,52
AR(9)	-4,42374	-4,40566	6.620,28
AR(10)	-4,42750	-4,40741	6.624,69*
AR(15)	-4,43088*	-4,40071	6.623,66

### 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 Autorregressivos

Consideremos o modelo AR(p)

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t, \quad (3.21)$$

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

$$\begin{aligned} \ell(\boldsymbol{\xi}|\mathbf{X}, x_0) &\approx (-T/2) \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T (X_t - \phi_0 - \phi_1 X_{t-1})^2 \\ &= (-T/2) \log \sigma^2 - \frac{1}{2\sigma^2} S(\boldsymbol{\phi}|\mathbf{X}, x_0). \end{aligned} \quad (3.22)$$

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 (3.21), 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

$$\hat{\sigma}^2 = \frac{\sum_{t=p+1}^T \hat{\varepsilon}_t^2}{T - 2p - 1}. \quad (3.23)$$

O valor do denominador em (3.23) 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)

$$X_t = \theta_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (3.24)$$

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 (2006). 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 et al. (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 ideia da precisão dos estimadores encontrados, devemos construir intervalos de confiança para os parâmetros. Seja  $\boldsymbol{\eta} = (\phi, \theta)$ , 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

$$\hat{\boldsymbol{\eta}} \xrightarrow{\mathcal{D}} \mathcal{N}_k(\boldsymbol{\eta}, \mathbf{V}^{-1}),$$

$$\mathbf{V} = 2\sigma^2 \begin{bmatrix} \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1^2} & \dots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_1 \partial \eta_k} \\ \vdots & & \vdots \\ \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k \partial \eta_1} & \dots & \frac{\partial^2 S(\boldsymbol{\eta})}{\partial \eta_k^2} \end{bmatrix}, \quad (3.25)$$

onde  $S(\boldsymbol{\eta})$  é a soma de quadrados que aparece em (3.22), 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 (3.25) 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)

$$\hat{\phi} = 0,8286,$$

$$\hat{\sigma}(\hat{\phi}) = 0,0812,$$

$$P = 0,0000.$$

(b) Modelo MA(1)

$$\hat{\theta} = 0,8241,$$

$$\hat{\sigma}(\hat{\theta}) = 0,0804,$$

$$P = 0,0000.$$

(c) Modelo ARMA(1,1)

$$\hat{\phi} = 0,9185, \quad \hat{\sigma}(\hat{\phi}) = 0,0714 \text{ e } P = 0,000,$$

$$\hat{\theta} = 0,3911, \quad \hat{\sigma}(\hat{\theta}) = 0,1626 \text{ e } P = 0,020.$$

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 (3.16), 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 Petrobras, do Exemplo 3.6, o programa EViews forneceu as seguintes estimativas:

$$\hat{\phi}_1 = 0,0782, \quad \hat{\sigma}(\hat{\phi}_1) = 0,018, \quad P = 0,0000,$$

$$\hat{\phi}_2 = -0,049, \quad \hat{\sigma}(\hat{\phi}_2) = 0,018, \quad P = 0,0073,$$

$$\hat{\phi}_3 = -0,057, \quad \hat{\sigma}(\hat{\phi}_3) = 0,018, \quad P = 0,0017,$$

$$\hat{\phi}_{10} = 0,038, \quad \hat{\sigma}(\hat{\phi}_{10}) = 0,018, \quad P = 0,0334,$$

$$\hat{\phi}_{15} = 0,054, \quad \hat{\sigma}(\hat{\phi}_{15}) = 0,018, \quad P = 0,0029.$$

### 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 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. Esse 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 esse modelo for verdadeiro, então os “erros verdadeiros”  $\varepsilon_t$  constituirão um ruído branco.

#### 3.4.1 Teste de Autocorrelação Residual

Estimados  $\hat{\phi}$  e  $\hat{\theta}$ , 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 autocorrelações 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 autocorrelações  $\hat{r}_k$  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 et al. (1994) provaram que, para um modelo AR(1), tem-se

---


$$\text{Var}(\hat{r}_k) \simeq \frac{1}{n}[1 - \phi^{2(k-1)}(1 - \phi^2)]$$

$$\text{Cov}(\hat{r}_i, \hat{r}_j) \simeq \frac{1}{n} \{ \delta_{ij} - \phi^{i+j-2} (1 - \phi^2) \},$$

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 autocorrelações estimadas dos resíduos 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  $k$ , esses limites subestimarão a significância de qualquer discrepância.

### 3.4.2 Teste de Box-Pierce-Ljung

Box e Pierce (1970) sugeriram um teste para as autocorrelações 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 desse teste foi proposta por Ljung e Box (1978), que é apresentada a seguir.

Se o modelo for apropriado, a estatística

$$Q(K) = n(n+2) \sum_{k=1}^K \frac{\hat{r}_k^2}{(n-k)} \quad (3.26)$$

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, verificar 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). A Figura 3.7 apresenta as f.a.c. e f.a.c.p dos resíduos com os correspondentes intervalos de confiança. Os valores das autocorrelações e autocorrelações parciais não indicam nenhuma quebra de comportamento de ruído branco nos resíduos.

Assim, o modelo ajustado à série  $\ln(ICV)$  é dada por

$$(1 - 0,5119B)(1 - B)\ln(ICV) = 0,01036 + \varepsilon_t, \quad (3.27)$$

com  $\hat{\sigma}^2 = 0,0000923$ .

**Exemplo 3.12.** Retomemos o exemplo 3.10, no qual obtivemos as estimativas dos parâmetros do modelo AR(15) ajustado à série de retornos da Petrobras. Uma estimativa da variância residual é  $\hat{\sigma}^2 = 0,00069$ . 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. No Capítulo 5, veremos que essa 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.

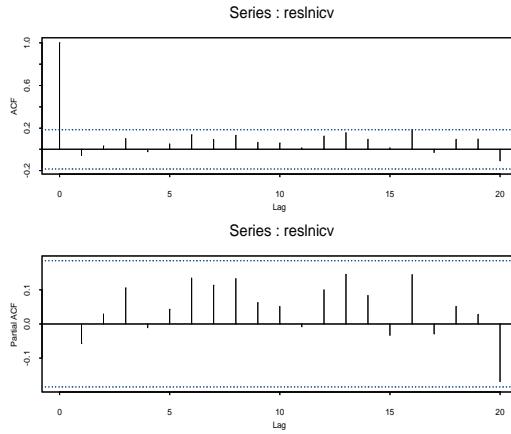


Figura 3.7: F.a.c. e f.a.c.p. dos resíduos do modelo ARIMA(1,1,0) ajustado à série  $\ln(ICV)$ .

### 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,

$$\begin{aligned} \hat{X}_T(h) &= E(\varphi_1 X_{T+h-1} + \dots + \varphi_{p+d} X_{T+h-p-d} \\ &\quad + \theta_0 + \varepsilon_{T+h} - \theta_1 \varepsilon_{T+h-1} - \dots - \theta_q \varepsilon_{T+h-q} \mid X_T, X_{T-1}, \dots). \end{aligned}$$

Para calcular as previsões, usamos os fatos:

$$(a) E(X_{T+j}|X_T, X_{T-1}, \dots) = \begin{cases} X_{T+j}, & \text{se } j \leq 0 \\ \hat{X}_T(j), & \text{se } j > 0 \end{cases}$$

$$(b) E(\varepsilon_{T+j}|X_T, X_{T-1}, \dots) = \begin{cases} \varepsilon_{T+j}, & \text{se } j \leq 0 \\ 0, & \text{se } j > 0 \end{cases}$$

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.

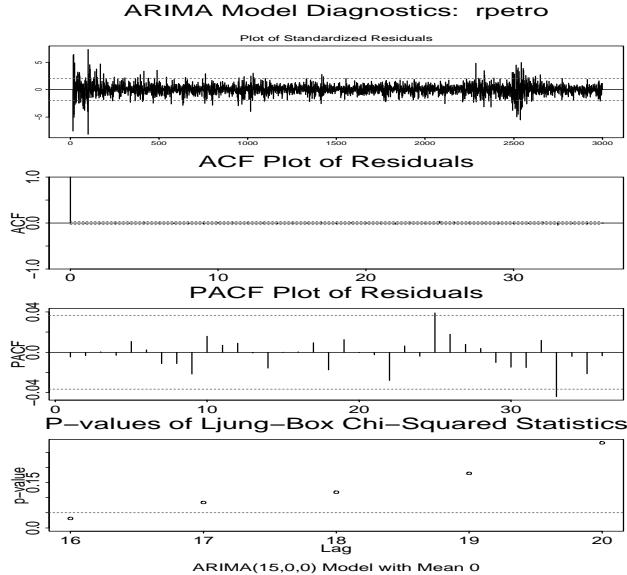


Figura 3.8: Diagnóstico do modelo do Exemplo 3.10.

**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 (infinita), pode-se provar que o *erro de previsão* é dado por

$$e_T(h) = X_{T+h} - \hat{X}_T(h) = \varepsilon_{T+h} + \psi_1 \varepsilon_{T+h-1} + \dots + \psi_{h-1} \varepsilon_{T+1}, \quad (3.28)$$

e a variância do erro de previsão é dada por

$$V(h) = \text{Var}(e_T(h)) = \sigma^2(1 + \psi_1^2 + \dots + \psi_{h-1}^2). \quad (3.29)$$

em que os pesos  $\psi_j$  vêm de  $\psi(B) = \varphi^{-1}(B)\theta(B)$ .

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. Nesse caso, (3.29) fica

$$\hat{V}(h) = \hat{\sigma}^2(1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2), \quad (3.30)$$

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,  $\text{Var}(e_T(1)) = \sigma^2$ . Essa 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 seis meses de 1979 e os primeiros seis meses de 1980).

Sabemos como calcular a variância do erro de previsão, que é dada por (3.29). 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  $E(\varepsilon_t a_s) = 0$ ,  $t \neq s$ , 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

$$\hat{X}_T(h) - u_\gamma \hat{\sigma} \left[ 1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2 \right]^{1/2} \leq X_{T+h} \leq \hat{X}_T(h) + u_\gamma \hat{\sigma} \left[ 1 + \sum_{j=1}^{h-1} \hat{\psi}_j^2 \right]^{1/2}. \quad (3.31)$$

**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 (3.31) 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 mesmo modo, podem ser obtidos intervalos de confiança para  $Y_{T+1}, Y_{T+2}$  etc. Observe que a variância aumenta com  $h$ ; logo, as amplitudes desses 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.

**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  $Y_t = \Delta X_t$ .

Para obtermos uma série livre de componente sazonal tomamos uma *diferença sazonal de ordem 12*,

$$Z_t = \Delta_{12} Y_t = (1 - B^{12}) Y_t, \quad (3.32)$$

ou seja,

$$Z_t = Y_t - Y_{t-12}.$$

Segue-se que podemos escrever

$$Z_t = (1 - B)(1 - B^{12}) X_t. \quad (3.33)$$

Note que essa 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 a f.a.c.p. da série  $Z_t$ .

Uma análise da f.a.c. e da f.a.c.p. de  $Z_t$  pode nos dar uma sugestão do modelo final para a série. Todavia, como já salientamos, essas funções têm, em geral, uma forma complicada, e pode ser difícil identificarmos um modelo a partir delas.

Um modelo sugerido por Box et al. (1994) e bastante utilizado em aplicações práticas é o chamado *airline model*, que tem a forma

$$(1 - B)(1 - B^s)X_t = (1 - \theta B)(1 - \Theta B^s)\varepsilon_t, \quad (3.34)$$

com  $|\theta| < 1$ ,  $|\Theta| < 1$ . Em (3.34),  $\theta$  é o parâmetro “regular” de médias móveis,  $\Theta$  é o parâmetro de médias móveis sazonal e  $s$  é o período sazonal.

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

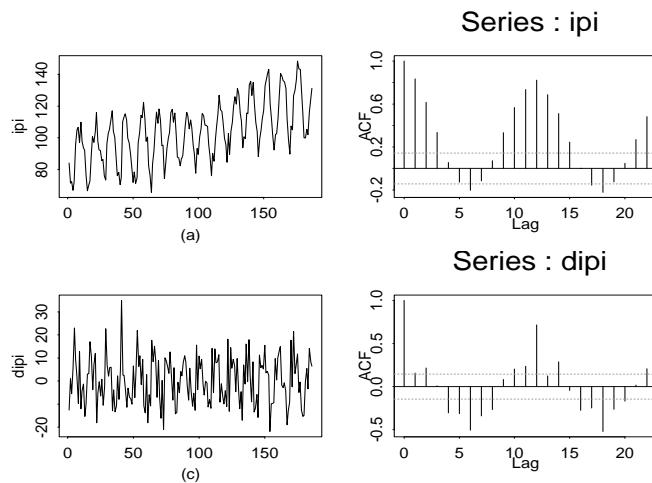
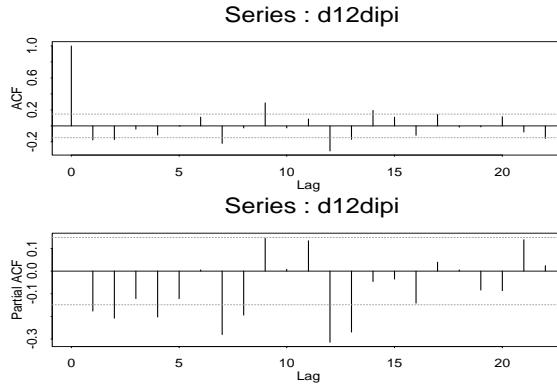


Figura 3.9: (a) Série IPI; (b) f.a.c. da série; (c) Primeira diferença da série IPI ;  
(d) f.a.c. da primeira diferença.

Figura 3.10: Série  $Z_t = \Delta\Delta_{12}X_t$ : (a) f.a.c. (b) f.a.c.p.

O modelo (3.34) é um caso particular do chamado modelo ARIMA *sazonal multiplicativo* (SARIMA) de ordem  $(p, d, q) \times (P, D, Q)_s$ , dado por

$$\phi(B)\Phi(B^s)\Delta^d\Delta_s^D X_t = \theta(B)\Theta(B^s)\varepsilon_t, \quad (3.35)$$

onde

$\phi(B) = 1 - \phi_1B - \dots - \phi_pB^p$  é o operador autorregressivo estacionário de ordem  $p$ ;

$\theta(B) = 1 - \theta_1B - \dots - \theta_qB^q$  é o operador de médias móveis invertível de ordem  $q$ ;

$\Phi(B^s) = 1 - \Phi_1B^s - \dots - \Phi_PB^{sP}$  é o operador autorregressivo sazonal de ordem  $P$ , estacionário;

$\Theta(B^s) = 1 - \Theta_1B^s - \dots - \Theta_QB^{sQ}$  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 (3.34) para a série IPI, obtemos

$$(1 + 0,2562B^4 + 0,1587B^5 + 0,2984B^7)(1 - B)(1 - B^{12})X_t \\ = (1 - 0,5409B)(1 - 0,6584B^{12})\varepsilon_t, \quad (3.36)$$

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 na Figura 3.11. 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 (3.36), 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
<hr/>			$EQMP_{180} = 58,17$

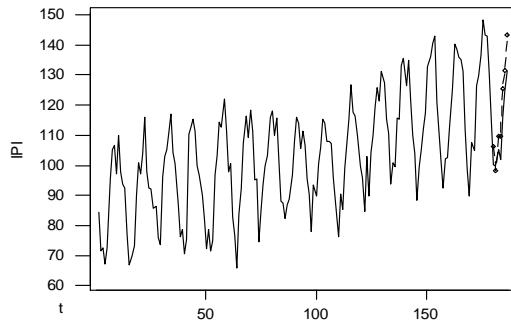


Figura 3.11: Série IPI com previsões utilizando o modelo (3.36), com origem em dezembro de 1999 e  $h = 1, 2, \dots, 7$ .

### 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,  $\text{Var}(r_1) = \frac{1}{T}(1-\phi^2)$ .

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),  $\text{Var}(\bar{W}) = \frac{c_0(1+r_1)}{n(1-r_1)}$ .
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	-0,73	-0,05	0,25	0,20	0,12

$T = 100$ ,  $\bar{X} = -0,08$ ,  $S_X^2 = 2,40$ . Identifique um modelo para  $X_t$ .

5. Considere o modelo  $(1 - \phi B)X_t = (1 - \theta B)\varepsilon_t$ . Mostre que:
  - (a)  $X_t = (1 + \phi B)(1 - \theta B)\varepsilon_t$ , se  $\phi$  pequeno;
  - (b)  $X_t = \{1 + (\phi - \theta)B\}\varepsilon_t$ , 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 por que 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 ARMA para a série de log-retornos diárias da Vale (arquivo d-vale98.10.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 autocorrelações:
- | $k$            | 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 autocorrelações
- | $j$                        | 1    | 2    | 3    | 4    | 5    | 6     | 7     | 8    | 9    | 10    |
|----------------------------|------|------|------|------|------|-------|-------|------|------|-------|
| $r_{\hat{\varepsilon}}(j)$ | 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 isso 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 SARIMA( $0, 1, 2$ )  $\times$  ( $0, 1, 1$ )<sub>12</sub>:

$$\Delta\Delta_{12}X_t = (1 - \Theta B^{12})(1 - \theta_1 B - \theta_2 B^2)\varepsilon_t.$$

- (a) Escreva o modelo na forma de um modelo ARIMA.
- (b) Qual a ordem do modelo ARIMA resultante?
- (c) Obtenha a f.a.c. do modelo.

18. Para o modelo SARIMA( $0, 1, 1$ )  $\times$  ( $0, 1, 1$ )<sub>12</sub>:

- (a) escreva-o explicitamente;
- (b) obtenha a região de invertibilidade;
- (c) obtenha as autocorrelações do processo.

19. Obtenha as estimativas dos parâmetros dos modelos identificados nos Problemas 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. Estime o modelo (3.34) para a série IPI e mostre que obtemos (3.36).
22. 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 autorregressivo apresenta uma raiz sobre o círculo unitário. Isso 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. Isso pode indicar que os dados foram superdiferençados. 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,

$$X_t = \theta_0 + \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (4.1)$$

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,

$$\hat{\phi}_{MV} \xrightarrow{a} \mathcal{N}(\phi, (1 - \phi^2)/T). \quad (4.2)$$

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

$$\sqrt{T}(\hat{\phi}_{MV} - \phi) \xrightarrow{a} \mathcal{N}(0, (1 - \phi^2)), \quad (4.4)$$

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,

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{RB}\mathcal{N}(0, \sigma^2). \quad (4.9)$$

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 (4.10). 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.

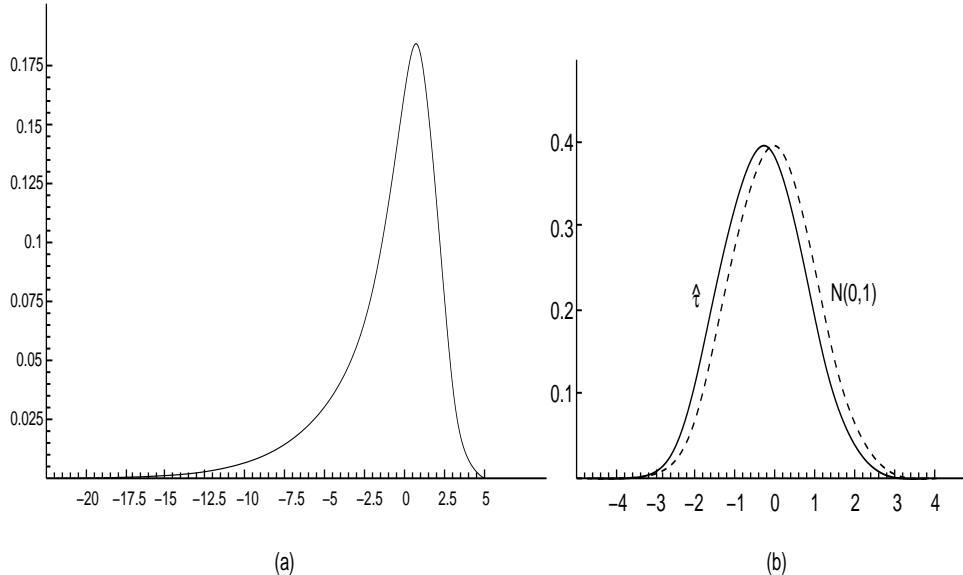


Figura 4.1: Distribuições, sob  $H_0 : \phi = 1$ , de (a)  $T(\hat{\phi}_{MQ} - 1)$  e (b)  $\hat{\tau}$ .

Suponha, agora, que a média não seja zero e temos o modelo (4.1). Nesse 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. Nesse 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  $H_0 : \phi = 1 | \theta_0 = 0$ . 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

$$\Phi_1 = \frac{\text{SQR}(r\text{estricta}) - \text{SQR}(irrestrita)/r}{\text{SQR}(irrestrita)/(T-k)}, \quad (4.23)$$

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

$$\Phi_1 = \frac{(\sum \Delta X_t^2 - \sum \hat{\varepsilon}_t^2)/2}{\sum \hat{\varepsilon}_t^2/(T-2)}. \quad (4.24)$$

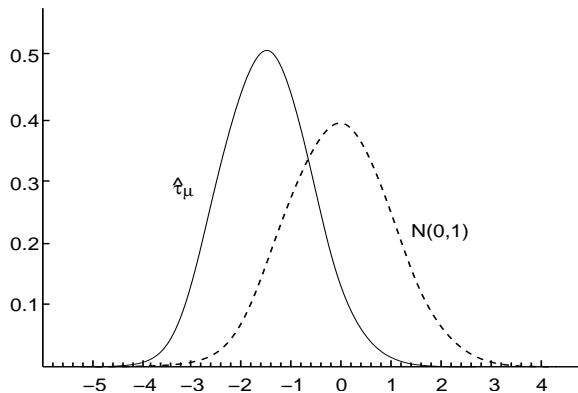


Figura 4.2: Distribuição de  $\hat{\tau}_\mu$  sob  $H_0 : \phi = 1$ .

Note que  $\sum \hat{\varepsilon}_t^2 = \sum (\Delta X_t - \hat{\theta}_0 - \hat{\phi}_{MQ} X_{t-1})^2$ . Valores críticos de  $\Phi_1$  para níveis de significância 0,05 e 0,01 são, respectivamente, 4,59 e 6,43.

### 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 autorregressivo  $\phi(B)$  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 autorregressivo 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  $\hat{\tau}_\mu = \hat{\phi}_1^*/\widehat{e.p.}(\hat{\phi}_1^*)$  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

$$p_{\max} = \left[ 12 \left( \frac{T}{100} \right)^{1/4} \right], \quad (4.28)$$

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  $X_0 = 0$ ,  $\phi = 0,7$ ,  $\varepsilon_t \sim \mathcal{N}(0, 1)$  e  $T = 200$ . O gráfico está na Figura 4.3, juntamente com a f.a.c. e a f.a.c.p.

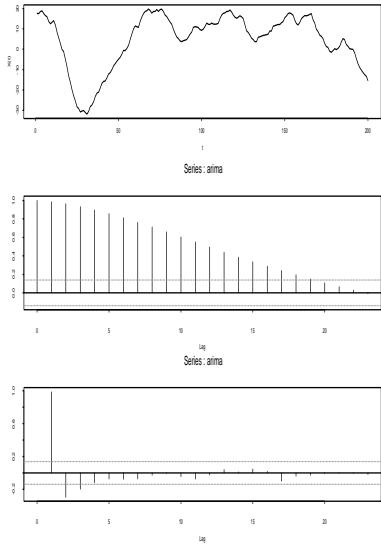


Figura 4.3: Série  $X_t$ , f.a.c. e f.a.c.p.

Se não soubéssemos como os dados  $X_t$  foram gerados, estas funções sugeriam 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

$$\hat{\phi} = 0,716, \hat{\sigma}^2 = 0,982, \widehat{\text{d.p.}}(\hat{\phi}) = 0,0313,$$

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, ajustariammos o modelo

$$(1 - 0,716B)(1 - B)X_t = \varepsilon_t, \quad \varepsilon_t \sim RB(0; 0, 982).$$

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, 0,0935, 0,0012, 0,0451 e 0,0451.

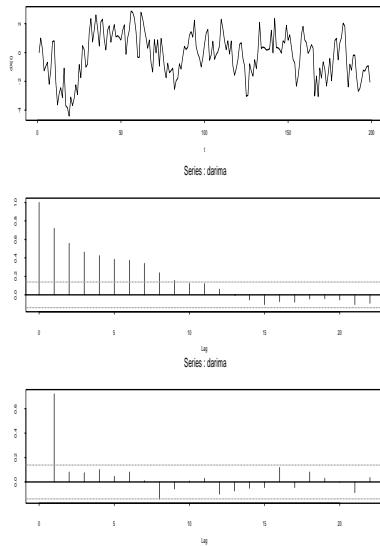


Figura 4.4: Série  $\Delta X_t$ , f.a.c. e f.a.c.p.

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

$$\hat{\tau}_\tau = \frac{\hat{\phi}_{1,MQ}^* - 1}{\widehat{\text{e.p.}}(\hat{\phi}_{1,MQ}^*)}, \quad (4.35)$$

cuja 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  $\sigma^2$ ,

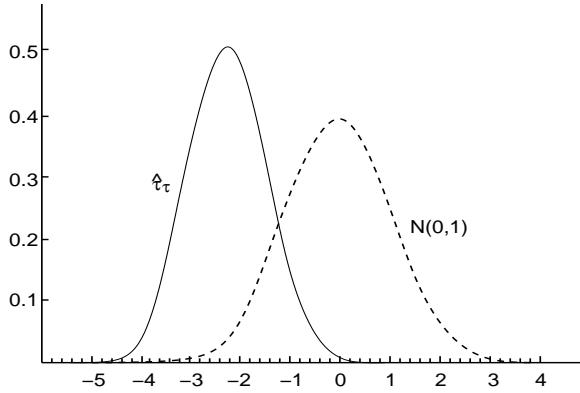
$$\hat{\tau}_\tau \xrightarrow{\mathcal{D}} \frac{\frac{1}{2}([W(1)]^2 - 1) - W(1) \int_0^1 W(r) dr + A}{[\int_0^1 [W(r)]^2 dr - (\int_0^1 W(r) dr)^2 + B]^{1/2}}, \quad (4.36)$$

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 autocorrelação 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

$$\begin{aligned} \sigma^2 &= \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(u_t^2), \\ \lambda^2 &= \lim_{T \rightarrow \infty} \sum_{t=1}^T E(T^{-1} S_T^2), \quad S_T = \sum_{t=1}^T u_t, \end{aligned}$$

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 da f.a.c.p. estão na Figuras 4.6 (b) e 4.6 (c). 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 que o valor da estatística de teste é  $\hat{\tau}_\mu = 1,25866$ , com valor- $p$  0,9985 e não rejeitamos a hipótese nula em (4.11). Como a constante do modelo parece ser não significativa, podemos considerar o caso (4.10), e a estatística do teste tem valor  $\hat{\tau} = 2,60781$ , 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  $T = 1.992$  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 4.1 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 4.2, do qual segue que não rejeitamos a hipótese nula de raiz unitária. No Quadro 4.3, temos o resultado da aplicação do teste PP, que também detecta raiz unitária na série.

## 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, isso tem a ver com um teste de estacionariedade, onde a hipótese nula específica que o processo é estacionário ao redor de uma tendência determinística (*trend-stationary*) e a hipótese alternativa especifica que o processo é I(1).

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 esse teste, veja Kwiatkowski 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).

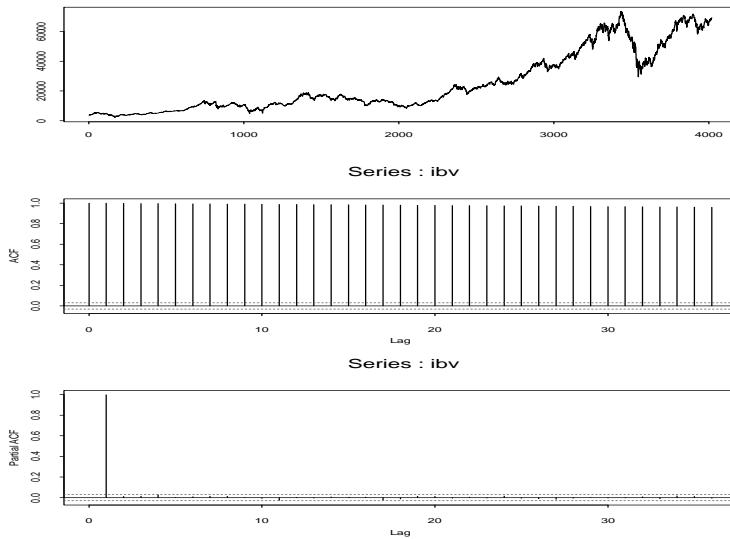


Figura 4.6: Série Ibovespa, f.a.c. e f.a.c.p.

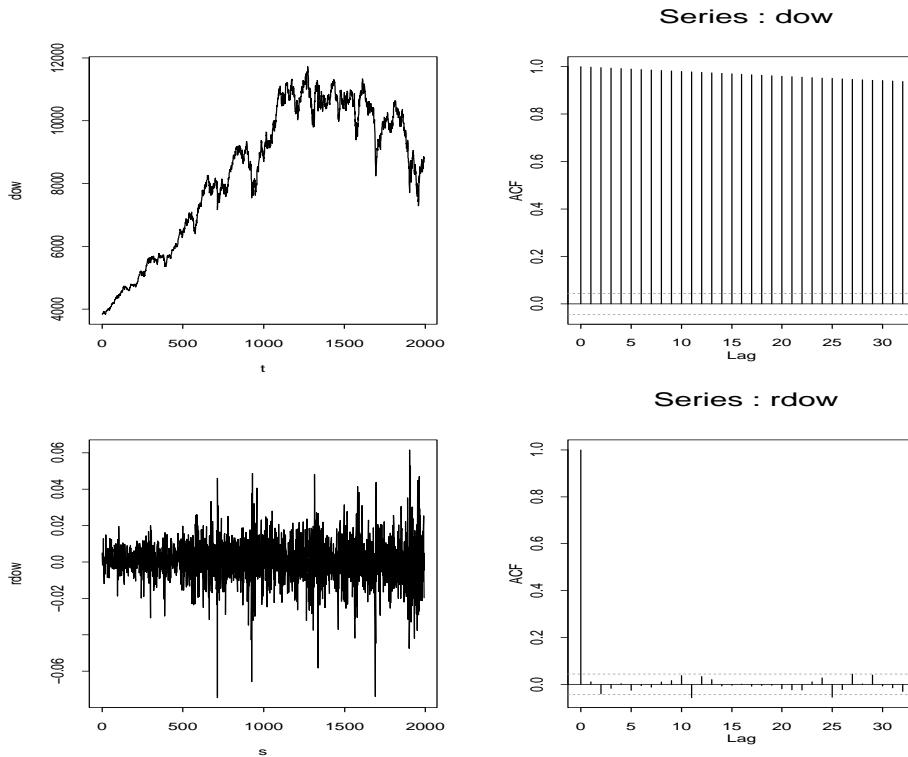


Figura 4.7: (a) Série DJIA; (b) f.a.c. da série; (c) Retornos do DJIA; (d) f.a.c. dos retornos.

[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 esses problemas.

Quadro 4.1: 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: -2.019
        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  $p = 1$ , 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 de Phillips-Perron 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					

[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 6 para noções sobre processos de memória longa.

[6] Sob o ponto de vista bayesiano, testes de raízes unitárias foram considerados inicialmente por Sims (1988) e Sims e Uhlig (1991). As críticas feitas por esses autores aos procedimentos frequentistas levaram Phillips (1991a,b) a responder, o que gerou um debate saudável, resumido por Bauwens et al. (1999). Outras referências são Lubrano (1995) e DeJong e Whiteman (1991).

## 4.5 Problemas

1. Considere a série de preços diários de ações da Petrobras (arquivo d-petro98.10.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 da Vale (arquivo d-vale98.10.dat);
  - (b) índices mensais da Bovespa (arquivo m-ibv94.01.dat);

- (c) dados mensais dos juros do C-Bond brasileiro, de julho de 1994 a agosto de 2001,  $T = 86$  (arquivo m-cbond94.01.dat).

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 de a 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 (4.34) com  $k = 1$  (ou  $p = 2$ ). Por meio de substituições sucessivas, mostre que se  $\phi = 1$  e  $\beta_1 \neq 0$ ,  $X_t$  conterá uma tendência quadrática.

*Observação.* É sempre possível que  $\beta_1 \neq 0$  e, nesse 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  $\hat{\tau}_\mu$ . 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.

## 4.6 Apêndice 4: Provas dos Teoremas 4.1 e 4.2

Nesse 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.'s 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

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T X_{t-1} \varepsilon_t \xrightarrow{\mathcal{D}} \frac{1}{2}(X - 1), \quad (A.4)$$

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

$$S_T(r) = [\sqrt{T}Y_T(r)]^2 \xrightarrow{\mathcal{D}} \sigma^2[W(r)]^2, \quad (A.5)$$

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

$$\frac{1}{T^2} \sum_{t=2}^T X_{t-1}^2 \xrightarrow{\mathcal{D}} \sigma^2 \int_0^1 [W(r)]^2 dr. \quad (A.6)$$

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).

# Modelos para a 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 heteroscedasticity*), introduzida por Engle (1982) e suas extensões. Esses modelos são não lineares no que se refere à variância, como veremos a seguir. 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.

Como dissemos acima, o objetivo será modelar a volatilidade de um retorno. A notação a ser usada é aquela estabelecida na Seção 1.7. Consideraremos uma série de retornos,

$$r_t = \ln(P_t) - \ln(P_{t-1}),$$

e sejam

$$\mu_t = E(r_t | \mathcal{F}_{t-1}), \quad h_t = \text{Var}(r_t | \mathcal{F}_{t-1}) \quad (5.1)$$

a média e variância condicional de  $r_t$ , onde  $\mathcal{F}_{t-1}$  denotará a informação até o instante  $t-1$ , que consideraremos ser  $\{r_{t-1}, r_{t-2}, \dots, r_1\}$ .

Em algumas situações, iremos supor que  $\mu_t = 0$ , de modo que neste caso  $h_t = E(r_t^2 | \mathcal{F}_{t-1})$ .

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

$$r_t = g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (5.2)$$

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

$$r_t = \varepsilon_t + \alpha \varepsilon_{t-1}^2,$$

é não linear na média, pois  $g(\cdot) = \alpha \varepsilon_{t-1}^2$  e  $h(\cdot) = 1$ , ao passo que o modelo ARCH(1)

$$r_t = \varepsilon_t \sqrt{\alpha r_{t-1}^2}$$

é não linear na variância, pois  $g(\cdot) = 0$  e  $h(\cdot) = \sqrt{\alpha r_{t-1}^2}$ , e  $r_{t-1}$  depende de  $\varepsilon_{t-1}$ .

## 5.2 Modelos ARCH

Os modelos ARCH, ou modelos autorregressivos com heteroscedasticidade condicional, foram introduzidos por Engle (1982), com o objetivo de estimar a variância da inflação. A ideia básica aqui é que o retorno  $r_t$  é não correlacionado serialmente, mas a volatilidade (variância condicional) depende de retornos passados por meio de uma função quadrática.

A Teoria dos Mercados Eficientes diz que os preços incorporam a informação disponível a todos os participantes do mercado, implicando que as variações de preços (retornos) não sejam previsíveis (Fama, 1970; Campbell et al., 1997). Há três formas de eficiência, dependendo da informação disponível aos participantes:

- (a) *eficiência fraca*: retornos não podem ser previstos a partir da história prévia dos preços;
- (b) *eficiência semiforte*: o conjunto de informações inclui a história passada dos preços mais toda a informação pública disponível;
- (c) *eficiência forte*: o conjunto de informações inclui toda a informação pública e privada.

Quando dizemos que os retornos não são previsíveis, queremos dizer que não podemos prever retornos “anormais”, isto é, retornos que não são previstos por um modelo suposto verdadeiro pelo mercado.

Eficiência fraca implica que os retornos sejam independentes ao longo do tempo, ou seja, um modelo adequado seria  $r_t = \mu + \varepsilon_t$ , com  $\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2)$ , o que é muito

restritivo. Os modelos ARCH permitem que os erros sejam não correlacionados, mas dependentes, ou seja,  $\text{Cov}(r_t^2, r_{t-1}^2)$  pode ser não nula.

**Definição 5.1.** Um modelo  $ARCH(m)$  é definido por

$$r_t = \sqrt{h_t} \varepsilon_t, \quad (5.3)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2, \quad (5.4)$$

onde  $\varepsilon_t$  i.i.d. com média zero,  $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, m-1, \alpha_m > 0$ .

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  $r_t$ .

Pela própria definição, valores grandes de  $r_t$  são seguidos por outros valores grandes da série.

Para investigar algumas propriedades dos modelos ARCH, consideremos o caso especial  $m = 1$ , ou seja, temos o modelo

$$r_t = \sqrt{h_t} \varepsilon_t, \quad (5.5)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2, \quad (5.6)$$

com  $\alpha_0 > 0, \alpha_1 > 0$ .

Calculemos a média e variância incondicionais da série.

$$(i) E(r_t) = E\{E(r_t | \mathcal{F}_{t-1})\} = 0;$$

$$\begin{aligned} (ii) \text{Var}(r_t) &= E(r_t^2) = E\{E(r_t^2 | \mathcal{F}_{t-1})\} \\ &= E(\alpha_0 + \alpha_1 r_{t-1}^2) = \alpha_0 + \alpha_1 E(r_{t-1}^2). \end{aligned}$$

Se o processo  $\{r_t\}$  for estacionário de segunda ordem, então, para todo  $t$ ,  $E(r_t^2) = E(r_{t-1}^2) = \text{Var}(r_t)$ , do que decorre

$$\text{Var}(r_t) = \frac{\alpha_0}{1 - \alpha_1}. \quad (5.7)$$

Como  $\text{Var}(r_t) > 0$ , deveremos ter  $0 < \alpha_1 < 1$ .

$$\begin{aligned} (iii) \text{Cov}(r_t, r_{t+k}) &= E(r_t r_{t+k}) \\ &= E[E(r_t r_{t+k} | \mathcal{F}_{t+k-1})] = E[r_t E(r_{t+k} | \mathcal{F}_{t+k-1})] \\ &= E[r_t E(\sqrt{h_{t+k}} \varepsilon_{t+k} | \mathcal{F}_{t+k-1})] = 0, \text{ para } k > 0, \end{aligned}$$

pois  $r_t$  está em  $\mathcal{F}_{t+k-1}$  e  $E(\varepsilon_{t+k} | \mathcal{F}_{t+k-1}) = 0$ .

Dessa forma,

$$\gamma_r(k) = 0, \quad k \geq 1,$$

indicando que  $r_t$  é uma sequência de variáveis não correlacionadas (ruído branco), com média zero e variância dada por (5.7).

Sabemos que os retornos apresentam geralmente caudas longas, de modo que a curtose é maior do que 3. Para calcular a curtose, supondo que  $r_t$  siga o modelo (5.3)-(5.4), é necessário calcular o momento de quarta ordem de  $r_t$ . Suponha que os  $\varepsilon_t$  sejam normais, para facilidade de cálculo. Então temos:

$$E(r_t^4 | \mathcal{F}_{t-1}) = E(h_t^2 \varepsilon_t^4 | \mathcal{F}_{t-1}) = 3(\alpha_0 + \alpha_1 r_{t-1}^2)^2, \quad (5.8)$$

pois  $E(\varepsilon_t^4) = 3$ , do que decorre

$$E(r_t^4) = 3E(\alpha_0 + \alpha_1 r_{t-1}^2)^2 = 3E(\alpha_0^2 + 2\alpha_0\alpha_1 r_{t-1}^2 + \alpha_1^2 r_{t-1}^4).$$

Admitindo-se que o processo seja estacionário de quarta ordem, o momento de quarta ordem pode ser escrito  $\mu_4 = E(r_t^4)$ , e teremos

$$\begin{aligned} \mu_4 &= 3(\alpha_0^2 + 2\alpha_0\alpha_1 \text{Var}(r_t) + \alpha_1^2 \mu_4) \\ &= 3(\alpha_0^2 + 2\alpha_0\alpha_1[\alpha_0/(1 - \alpha_1)] + \alpha_1^2 \mu_4) \\ &= 3\alpha_0^2(1 + 2\alpha_1/(1 - \alpha_1)) + 3\alpha_1^2 \mu_4. \end{aligned}$$

Daqui, obtemos, finalmente,

$$\mu_4 = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}. \quad (5.9)$$

Supondo-se que momentos de quarta ordem sejam finitos e positivos, de (5.9) 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. Isso é verdade para o modelo geral ARCH( $m$ ).

A curtose de  $r_t$  será, então, dada por

$$K = \frac{\mu_4}{[\text{Var}(r_t)]^2} = 3 \frac{\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \frac{(1 - \alpha_1)^2}{\alpha_0^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3. \quad (5.10)$$

Vemos, pois, que se admitirmos que  $r_t$  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, em razão do fato de termos retornos ao quadrado, alguns retornos grandes e isolados podem conduzir a superprevisões.

Pode-se mostrar (veja o Problema 16) que se  $\varepsilon_t \sim t_\nu$ , a curtoza é dada por

$$K = \lambda \frac{1 - \alpha_1^2}{1 - \lambda \alpha_1^2}, \quad (5.11)$$

se  $\alpha_1 < \sqrt{1/\lambda}$ , onde  $\lambda = 3(\nu - 2)/(\nu - 4)$  é a curtoza da distribuição  $t_\nu$ , que existe se  $\nu > 4$ . Nesse caso, as caudas são mais pesadas que no caso normal.

Utilizando (5.3) e (5.4), e calculando  $r_t^2 - h_t$ , temos que

$$r_t^2 - (\alpha_0 + \alpha_1 r_{t-1}^2) = h_t(\varepsilon_t^2 - 1),$$

ou seja,

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + v_t, \quad (5.12)$$

na qual

$$v_t = h_t(\varepsilon_t^2 - 1) = h_t(X - 1). \quad (5.13)$$

Supondo  $\varepsilon_t \sim N(0, 1)$ , vemos que  $X$  é uma v.a. com distribuição  $\chi^2(1)$ , o que mostra que temos um modelo AR(1) para  $r_t^2$ , mas com erros não gaussianos. Ainda, é fácil ver que  $\{v_t\}$  é uma sequência de v.a.'s de média zero, não correlacionadas, mas com variância não constante.

De (5.12), temos que a f.a.c. de  $r_t^2$  é dada por

$$\rho_{r^2}(k) = \alpha_1^k, \quad k > 0.$$

Para um modelo ARCH( $m$ ) teremos

$$r_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + v_t, \quad (5.14)$$

onde os  $v_t$  são como no caso  $m = 1$ . Ou seja, temos um modelo AR( $m$ ) para  $r_t^2$ , com inovações não gaussianas, se  $\varepsilon_t$  é gaussiano. Além disso, pode-se demonstrar que os retornos  $r_t$  também formam um ruído branco, com variância dada por

$$\text{Var}(r_t) = \frac{\alpha_0}{1 - \sum_{i=1}^m \alpha_i}.$$

### 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 esse for o caso, teremos

$$\phi(B)r_t = \theta_0 + \theta(B)a_t,$$

sendo que  $a_t \sim \text{ARCH}(m)$ . No que segue, quando nos referirmos a  $r_t$ , 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  $r_t^2$ .

- (i) Teste de Box-Pierce-Ljung para  $r_t^2$ . 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  $i = 1, \dots, m$ , na regressão

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 + u_t,$$

para  $t = m+1, \dots, T$ . A estatística do teste é  $S = TR^2$ , que tem distribuição assintótica  $\chi^2(m)$  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

$$F = \frac{(SQR_0 - SQR_1)/m}{SQR_1/(T - 2m - 1)} \sim F(m, T - 2m - 1), \quad (5.15)$$

na qual  $SQR_0 = \sum_{t=m+1}^T (r_t^2 - \bar{r})^2$  e  $SQR_1 = \sum_{t=m+1}^T \hat{u}_t^2$ , com  $\bar{r}$  a média amostral dos  $r_t^2$  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 (5.4) de modelarmos a volatilidade e dado que  $r_t^2$  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 autorregressivo. Segue-se que a função de autocorrelação parcial de  $r_t^2$  pode ser usada para encontrar a ordem  $r$  de um modelo ARCH( $s$ ).

### Estimação

Os estimadores dos parâmetros do modelo são obtidos pelo método de máxima verossimilhança condicional. A função de verossimilhança é dada por

$$L(r_1, \dots, r_T | \boldsymbol{\alpha}) = f(r_T | \mathcal{F}_{T-1})f(r_{T-1} | \mathcal{F}_{T-2}) \cdots f(r_{m+1} | \mathcal{F}_m)f(r_1, \dots, r_m | \boldsymbol{\alpha}),$$

e, supondo normalidade dos  $\varepsilon_t$ , podemos escrever

$$L(r_1, \dots, r_T | \boldsymbol{\alpha}) = \prod_{t=m+1}^T (\sqrt{2\pi h_t})^{-1} \exp\left\{-\frac{r_t^2}{2h_t}\right\} f(r_1, \dots, r_m | \boldsymbol{\alpha}).$$

Para  $T$  grande,  $f(r_1, \dots, r_m | \boldsymbol{\alpha})$  pode ser desprezado. Veja Engle (1982) para detalhes. Logo, temos que maximizar a função de verossimilhança condicional

$$L(r_{m+1}, \dots, r_T | \boldsymbol{\alpha}, r_1, \dots, r_m) = \prod_{t=m+1}^T (\sqrt{2\pi h_t})^{-1} \exp\left\{-\frac{r_t^2}{2h_t}\right\}, \quad (5.16)$$

onde a volatilidade  $h_t$  é obtida recursivamente.

No caso particular de um modelo ARCH(1), temos

$$L(r_2, \dots, r_T | \alpha_0, \alpha_1, r_1) = f(r_T | r_{T-1}) f(r_{T-1} | r_{T-2}) \cdots f(r_2 | r_1),$$

em que

$$(r_t | r_{t-1}) \sim \mathcal{N}(0, h_t),$$

e  $h_t = \alpha_0 + \alpha_1 r_{t-1}^2$ . Segue-se que

$$L(r_2, \dots, r_T | \alpha_0, \alpha_1, r_1) = (2\pi)^{-T/2} \prod_{t=2}^T (\alpha_0 + \alpha_1 r_{t-1}^2)^{-1/2} \exp\left\{-\frac{r_t^2}{2(\alpha_0 + \alpha_1 r_{t-1}^2)}\right\}.$$

A log-verossimilhança fica

$$\ell(r_2, \dots, r_T | \alpha_0, \alpha_1, r_1) \propto -\frac{1}{2} \sum_{t=2}^T \ln(\alpha_0 + \alpha_1 r_{t-1}^2) - \frac{1}{2} \sum_{t=2}^T \left( \frac{r_t^2}{\alpha_0 + \alpha_1 r_{t-1}^2} \right). \quad (5.17)$$

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, R, RATS e PcGIVE, dentre outros, podem ser usados para estimar modelos ARCH (e GARCH, em geral).

### Verificação

Para um modelo ARCH(m), com  $\varepsilon_t$  normal ou  $t$ -Student, os resíduos

$$\tilde{r}_t = \frac{r_t}{\sqrt{h_t}}$$

(padronizados) são v.a.'s 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  $\tilde{r}_t$ . Além disso, podemos calcular os coeficientes de assimetria e curtose estimados, e fazer um gráfico Q×Q para avaliar a suposição de normalidade (ou  $t_\nu$ ).

Para se verificar se ainda existe heteroscedasticidade condicional nos resíduos, pode-se aplicar o teste ML para a sequência  $\tilde{r}_t^2$ .

### Previsão

As previsões para a volatilidade utilizando o modelo ARCH(m) dado em (5.3)-(5.4) são obtidas recursivamente. Assim,

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 r_t^2 + \dots + \alpha_m r_{t-m+1}^2, \quad (5.18)$$

é a previsão de  $h_{t+1}$ , com origem fixada no instante  $t$ . As previsões  $\ell$  passos à frente, com origem em  $t$ , são dadas por

$$\hat{h}_t(\ell) = \alpha_0 + \sum_{i=1}^m \alpha_i \hat{h}_t(\ell - i), \quad (5.19)$$

em que  $\hat{h}_t(\ell - i) = r_{t+\ell-i}^2$ , se  $\ell - i \leq 0$ .

**Exemplo 5.1.** Vamos ajustar um modelo ARCH aos retornos diáários da Petrobras, com  $T = 2.998$  observações, série esta já estudada no exemplo 3.6. Lá, ajustamos um modelo AR(15), com termos de primeira, segunda, terceira, décima e décima quinta ordens significativos. Veja os Exemplos 3.10 e 3.12. A função de autocorrelação parcial dos quadrados dos resíduos  $\hat{a}_t$ , da Figura 5.1, sugere um modelo ARCH(9).

Um primeiro modelo proposto, então, é

$$\begin{aligned} r_t &= \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \phi_{10} r_{t-10} + \phi_{15} r_{t-15} + a_t, \\ a_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \dots + \alpha_9 a_{t-9}^2. \end{aligned}$$

(a) Admitindo  $\varepsilon_t \sim \mathcal{N}(0, 1)$ , e utilizando o programa EViews, obtemos que  $\phi_2$  não é significativamente diferente de zero. Embora o coeficiente  $\alpha_4$  seja não significativo, vamos mantê-lo no modelo. Obtemos o seguinte modelo ajustado:

$$\begin{aligned} r_t &= 0,090 r_{t-1} - 0,047 r_{t-3} + 0,058 r_{t-10} + 0,0323 r_{t-15} + a_t, \\ a_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= 0,0002 + 0,079 a_{t-1}^2 + 0,106 a_{t-2}^2 + \dots + 0,068 a_{t-9}^2. \end{aligned} \quad (5.20)$$

Os resultados estão no Quadro 5.1. A Figura 5.2 apresenta a f.a.c. dos resíduos padronizados e de seus quadrados, indicando que o modelo é adequado para modelar a dependência linear entre sucessivos retornos e que a heteroscedasticidade condicional foi adequadamente modelada. As estatísticas usuais (Box-Pierce-Ljung

para resíduos e quadrados dos resíduos, e estatística do teste de multiplicador de Lagrange) conduzem à mesma conclusão sobre a validade do modelo ajustado.

No gráfico da volatilidade da Figura 5.3, observamos que os maiores picos (em ordem cronológica) na volatilidade estimada correspondem a: (a) setembro de 1998, após a moratória na Rússia; (b) janeiro a abril de 1999, desvalorização do Real; (c) agosto de 2000, queda da bolsa Nasdaq, (d) novembro de 2007, crise da *subprime*, nos EUA; (e) outubro a dezembro de 2008, crise internacional.

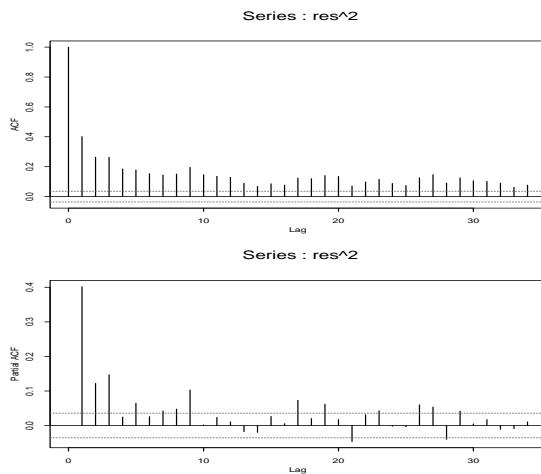


Figura 5.1: F.a.c. e f.a.c.p. dos quadrados dos resíduos  $\hat{a}_t$ , Exemplo 5.5.

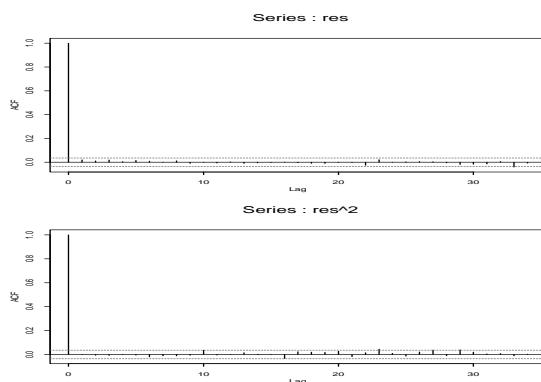


Figura 5.2: (a) F.a.c. dos resíduos padronizados; (b) f.a.c. dos quadrados dos resíduos padronizados.

Quadro 5.1: Ajuste de um modelo AR(15)-ARCH(9) aos retornos da Petrobras, erros normais				
Coefficient	Value	Std. Error	z-Statistic	P-value.
AR(1)	0.090085	0.019066	4.724946	0.0000
AR(3)	-0.047380	0.018540	-2.555553	0.0106
AR(10)	0.057937	0.015812	3.664181	0.0002
AR(15)	0.032284	0.016055	2.010847	0.0443
Variance		Equation		
C	0.000191	1.30E-05	14.73652	0.0000
RESID(-1) <sup>2</sup>	0.079356	0.015180	5.227592	0.0000
RESID(-2) <sup>2</sup>	0.106033	0.021303	4.977412	0.0000
RESID(-3) <sup>2</sup>	0.143646	0.020849	6.889890	0.0000
RESID(-4) <sup>2</sup>	0.019102	0.016731	1.141720	<b>0.2536</b>
RESID(-5) <sup>2</sup>	0.047102	0.017268	2.727660	0.0064
RESID(-6) <sup>2</sup>	0.073929	0.017776	4.158840	0.0000
RESID(-7) <sup>2</sup>	0.104454	0.016898	6.181589	0.0000
RESID(-8) <sup>2</sup>	0.036162	0.013544	2.669978	0.0076
RESID(-9) <sup>2</sup>	0.068166	0.020052	3.399508	0.0007
Adjusted R-squared	0.006271	S.D. dependent var	0.026506	
S.E. of regression	0.026423	Akaike info criterion	-4.736237	
Sum squared resid	2.072878	Schwarz criterion	-4.708074	
Log likelihood	7078.097	Hannan-Quinn criter.	-4.726104	
Durbin-Watson stat	2.009871			

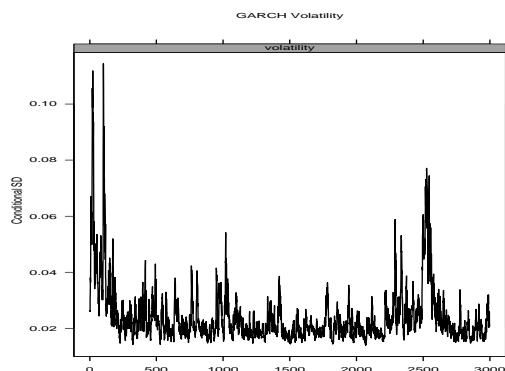


Figura 5.3: Estimativa da volatilidade dada pelo modelo (5.20) com erros gaussianos.

(b) Vamos supor, agora, que  $\varepsilon_t \sim t_\nu$ . Temos, no Quadro 5.2, obtido pelo EViews, os resultados do ajuste do mesmo modelo:

$$\begin{aligned}
 r_t &= 0,093r_{t-1} - 0,038r_{t-3} + 0,052r_{t-10} + 0,030r_{t-15} + a_t, \\
 a_t &= \sqrt{h_t}\varepsilon_t, \\
 h_t &= 0,0002 + 0,068a_{t-1}^2 + 0,113a_{t-2}^2 + \dots + 0,074a_{t-9}^2.
 \end{aligned} \tag{5.21}$$

O valor estimado de  $\nu$  foi, aproximadamente, 8 graus de liberdade. As estatísticas usadas antes mostram que esse modelo ajustado é adequado. Para comparar esse modelo ajustado com o modelo (5.20), devemos comparar os valores AIC e BIC. Vemos que esses valores são menores do que os fornecidos, quando consideramos os erros normais. A estimativa da volatilidade é bastante semelhante à Figura 5.3, que considera inovações gaussianas.

Quadro 5.2: Ajuste de um modelo AR(15)-ARCH(9) aos retornos da Petrobras, com erros $t$				
Coefficient	Value	Std. Error	z-Statistic	P-value.
AR(1)	0.093314	0.018684	4.994193	0.0000
AR(3)	-0.037607	0.018927	-1.986933	0.0469
AR(10)	0.052384	0.015813	3.312695	0.0009
AR(15)	0.029618	0.015692	1.887460	0.0591
Variance		Equation		
C	0.000195	1.88E-05	10.38501	0.0000
RESID(-1) <sup>2</sup>	0.068364	0.022184	3.081641	0.0021
RESID(-2) <sup>2</sup>	0.113265	0.027940	4.053827	0.0001
RESID(-3) <sup>2</sup>	0.145549	0.030012	4.849665	0.0000
RESID(-4) <sup>2</sup>	0.030873	0.022473	1.373785	0.1695
RESID(-5) <sup>2</sup>	0.054656	0.023227	2.353132	0.0186
RESID(-6) <sup>2</sup>	0.060621	0.023776	2.549614	0.0108
RESID(-7) <sup>2</sup>	0.099329	0.025418	3.907819	0.0001
RESID(-8) <sup>2</sup>	0.027099	0.019363	1.399475	0.1617
RESID(-9) <sup>2</sup>	0.073774	0.026047	2.832301	0.0046
R-squared	0.010207	Mean dependent var		0.001146
Adjusted R-squared	0.005538	S.D. dependent var		0.026506
S.E. of regression	0.026433	Akaike info criterion		-4.760212
Sum squared resid	2.073708	Schwarz criterion		-4.730037
Log likelihood	7114.856	Hannan-Quinn criter.		-4.749355
Durbin-Watson stat	2.014829			

### 5.3 Modelos GARCH

Uma generalização dos modelos ARCH foi sugerida por Bollerslev (1986, 1987, 1988), o chamado modelo GARCH (de *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  $GARCH(m, n)$  é definido por

$$r_t = \sqrt{h_t} \varepsilon_t \quad (5.22)$$

$$h_t = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + \sum_{j=1}^n \beta_j h_{t-j}, \quad (5.23)$$

em que  $\varepsilon_t$  são v.a.'s i.i.d., com média zero,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, m-1$ ,  $\beta_j \geq 0$ ,  $j = 1, \dots, n-1$ ,  $\alpha_m > 0$ ,  $\beta_n > 0$ ,  $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$ ,  $q = \max(m, n)$ .

Coeficientes 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

$$\nu_t = r_t^2 - h_t, \quad (5.24)$$

de modo que, substituindo em (5.23) obtemos

$$r_t^2 = \alpha_0 + \sum_{i=1}^q (\alpha_i + \beta_i) r_{t-i}^2 + \nu_t - \sum_{j=1}^n \beta_j \nu_{t-j}, \quad (5.25)$$

ou seja, temos um modelo ARMA( $q, n$ ) para  $r_t^2$ , 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, para todo  $t$ ,

$$E(\nu_t) = E(r_t^2 - h_t) = E(h_t \varepsilon_t^2 - h_t) = E(h_t)(E(\varepsilon_t^2) - 1) = 0,$$

$$E(\nu_t | \mathcal{F}_{t-1}) = E(r_t^2 | \mathcal{F}_{t-1}) - E(h_t | \mathcal{F}_{t-1}) = h_t - h_t = 0.$$

Segue-se, em particular, que

$$E(r_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^q (\alpha_i + \beta_i)}.$$

A longo prazo, a volatilidade convergirá para essa média.

Um modelo bastante usado na prática é o GARCH(1, 1), para o qual a volatilidade é expressa como

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}, \quad (5.26)$$

com  $0 < \alpha_1, \beta_1 < 1, \alpha_1 + \beta_1 < 1$ .

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 (5.26), obtemos facilmente

$$K = \frac{E(r_t^4)}{[E(r_t^2)]^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3, \quad (5.27)$$

dado que o denominador seja positivo, o que novamente mostra que se  $r_t$  segue um modelo GARCH, as caudas de  $r_t$  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 que 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

$$\sum_{t=1}^T (r_t^2 - h_t)^2.$$

Veja Mills (1999), Pagan e Schwert (1990) e Bollerslev et al. (1994).

Os estimadores dos parâmetros do modelo (5.22)-(5.23) são obtidos pelo método de máxima verossimilhança condicional. Supondo normalidade dos  $\varepsilon_t$ , temos que a log-verossimilhança, condicional às primeiras  $m$  observações, é dada por

$$\ell(r_{m+1}, \dots, r_T | \boldsymbol{\alpha}, \boldsymbol{\beta}, r_1, \dots, r_m) \propto -\frac{1}{2} \sum_{t=m+1}^T \ln(h_t) - \frac{1}{2} \sum_{t=m+1}^T \frac{r_t^2}{h_t}. \quad (5.28)$$

Bollerslev (1986) utiliza em (5.28),  $h_t = \hat{\sigma}^2, t = 1, \dots, n$ , com  $\hat{\sigma}^2 = \sum_{t=1}^T r_t^2 / T$ .

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 (5.26), são dadas por

---


$$\hat{h}_t(1) = \alpha_0 + \alpha_1 r_t^2 + \beta_1 h_t,$$


---

e para  $\ell > 1$ ,

$$\begin{aligned}\hat{h}_t(\ell) &= \alpha_0 + \alpha_1 \hat{r}_t^2(\ell - 1) + \beta_1 \hat{h}_t(\ell - 1), \\ &= \alpha_0 + \alpha_1 \hat{h}_t(\ell - 1) \hat{\varepsilon}_t^2(\ell - 1) + \beta_1 \hat{h}_t(\ell - 1),\end{aligned}$$

pois  $r_t = \sqrt{h_t} \varepsilon_t$ . Substituindo  $\hat{\varepsilon}_t^2(\ell - 1)$  por  $E(\varepsilon_{t+\ell-1}^2) = 1$ , temos que

$$\hat{h}_t(\ell) = \alpha_0 + (\alpha_1 + \beta_1) \hat{h}_t(\ell - 1), \quad \ell > 1. \quad (5.29)$$

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*). Nesse caso, teremos

$$\begin{aligned}r_t &= \sqrt{h_t} \varepsilon_t, \\ h_t &= \alpha_0 + \beta_1 h_{t-1} + (1 - \beta_1) r_{t-1}^2,\end{aligned}$$

com  $0 < \beta_1 < 1$ . Mas nesse caso, a variância incondicional de  $r_t$  não estará definida.

**Exemplo 5.2.** Vamos ajustar um modelo GARCH à série de retornos diárias do Ibovespa (veja o Exemplo 1.1). A Figura 5.4 reapresenta a f.a.c. e a f.a.c.p. da série, que indicam que um modelo apropriado é um autorregressivo, pois algumas autocorrelações parciais são significativas. Analisando a f.a.c. e a f.a.c.p. dos quadrados dos retornos, Figura 5.5, vemos uma forte dependência, sem um padrão bem definido.

(a) Iremos usar, então, um modelo AR(10)-GARCH(1,1), com erros gaussianos. Eliminando os coeficientes não significativos, o modelo ajustado foi

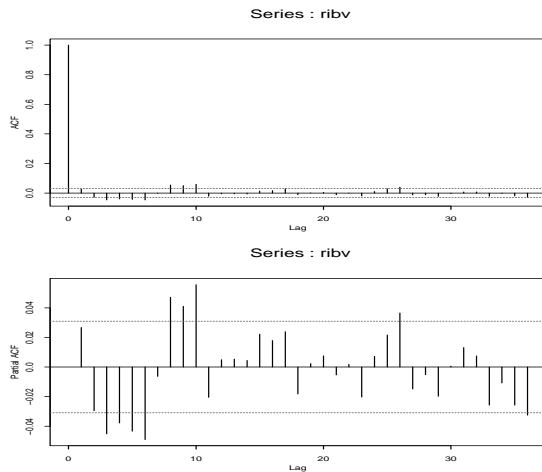


Figura 5.4: F.a.c. e f.a.c.p. dos retornos do Ibovespa.

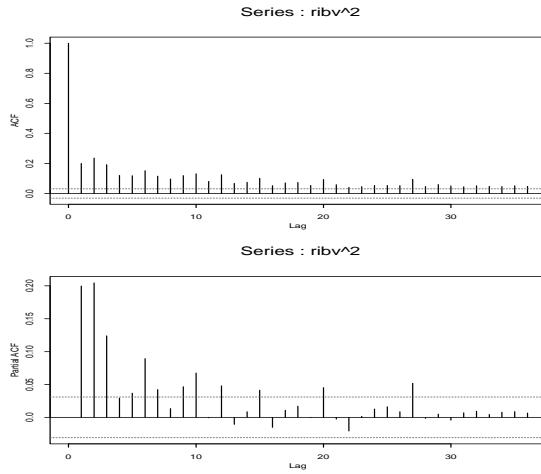


Figura 5.5: F.a.c. e f.a.c.p. dos quadrados dos retornos do Ibovespa.

$$\begin{aligned} r_t &= 0,051r_{t-10} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= 0,000014 + 0,117a_{t-1}^2 + 0,857h_{t-1}. \end{aligned} \tag{5.30}$$

O ajustamento do modelo é apresentado no Quadro 5.3. Os valores da estatística de Ljung-Box para os resíduos padronizados,  $\tilde{a}_t$ , são dados por  $Q(10) = 13,103$  ( $P=0,16$ ) e  $Q(20)=19,834$  ( $P = 0,41$ ), enquanto, para os quadrados dos resíduos padronizados, temos  $Q(10) = 13,34$  ( $P = 0,15$ ) e  $Q(20) = 22,451$  ( $P = 0,26$ ). Logo, podemos concluir que o modelo (5.30) é adequado para modelar a série.

A aplicação do teste ML, com  $p = 12$ , para confirmar a ausência de heteroscedasticidade condicional nos resíduos do modelo (5.30) forneceu  $F = 1,167$  ( $P = 0,30$ ). Apresentamos, na Figura 5.6, a estimativa do desvio padrão condicional ( $\sqrt{h_t}$ ) dos retornos.

(b) Consideremos, agora, o caso  $\varepsilon_t \sim t_\nu$ . No Quadro 5.4, temos o ajuste do modelo usando o software S+FinMetrics, resultando

$$\begin{aligned} r_t &= 0,055r_{t-10} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= 0,000010 + 0,097a_{t-1}^2 + 0,883h_{t-1}. \end{aligned} \tag{5.31}$$

Quadro 5.3: Ajuste de um modelo AR(10)-GARCH(1,1) aos retornos do Ibovespa, com erros gaussianos

Coefficient	Value	Std. Error	z-Statistic	P-value.
AR(10)	0.050503	0.015146	3.334394	0.0009
Variance		Equation		
C	1.40E-05	1.94E-06	7.227228	0.0000
RESID(-1) <sup>2</sup>	0.117340	0.006915	16.96882	0.0000
GARCH(-1)	0.856673	0.009128	93.84808	0.0000
R-squared	0.002675	Mean dependent var	0.000709	
Adjusted R-squared	0.001928	S.D. dependent var	0.023967	
S.E. of regression	0.023944	Akaike info criterion	-4.960518	
Sum squared resid	2.295532	Schwarz criterion	-4.954235	
Log likelihood	9944.879	Hannan-Quinn criter.	-4.958291	
Durbin-Watson stat	1.949702			

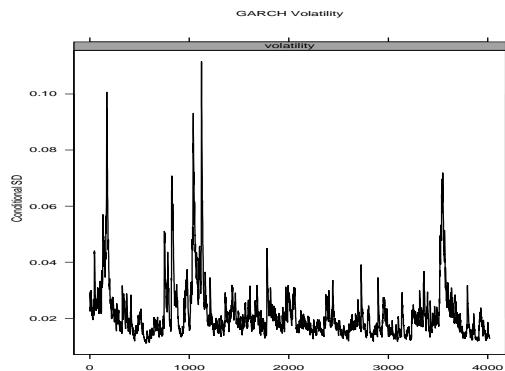


Figura 5.6: Estimativa do desvio padrão condicional para os retornos do Ibovespa, usando o modelo (5.30) com erros gaussianos.

Como estimativa de  $\nu$ , obtivemos 8 graus de liberdade. 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, concluímos que o modelo com inovações  $t$  está mais bem ajustado que o modelo com inovações gaussianas. A estimativa da volatilidade resulta ser similar à apresentada na Figura 5.6.

Quadro 5.4: Ajuste de um modelo AR(10)-GARCH (1,1) aos retornos do Ibovespa, erros $t$				
Coefficient	Value	Std. Error	z-Statistic	P-value.
AR(10)	0.054719	0.015158	3.609852	0.0003
Variance Equation				
C	1.04E-05	2.24E-06	4.653245	0.0000
RESID(-1) <sup>2</sup>	0.096692	0.010422	9.277384	0.0000
GARCH(-1)	0.883069	0.012169	72.56551	0.0000
R-squared	0.002740	Mean dependent	var	0.000709
Adjusted R-squared	0.001743	S.D. dependent var		0.023967
S.E. of regression	0.023946	Akaike info criterion		-4.990006
Sum squared resid	2.295383	Schwarz criterion		-4.982152
Log likelihood	10004.97	Hannan-Quinn criter.		-4.987223
Durbin-Watson stat	1.949938			

## 5.4 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).

### 5.4.1 Modelos EGARCH

Vimos que os modelos ARCH e GARCH tratam simetricamente os retornos, pois a volatilidade é uma função quadrática deles. Mas também é sabido que a volatilidade reage de forma assimétrica aos retornos, tendendo a ser maior para retornos negativos.

Tendo em vista o exposto, Nelson (1991) introduziu os modelos EGARCH (*exponential GARCH*).

**Definição 5.3.** Um modelo EGARCH(1,1) é dado por

$$r_t = \sqrt{h_t} \varepsilon_t, \quad (5.32)$$

$$\ln(h_t) = \alpha_0 + \alpha_1 g(\varepsilon_{t-1}) + \beta_1 \ln(h_{t-1}), \quad (5.33)$$

em que  $\varepsilon_t$  são v.a.'s i.i.d. com média zero e  $g(\cdot)$  é a curva de impacto de informação, dada por

$$g(\varepsilon_t) = \theta \varepsilon_t + \gamma \{ |\varepsilon_t| - E(|\varepsilon_t|) \}. \quad (5.34)$$

Aqui,  $\theta$  e  $\gamma$  são parâmetros reais, e  $|\varepsilon_t| - E(|\varepsilon_t|)$  é uma sequência de v.a.'s i.i.d. com média zero.

Note que  $E\{g(\varepsilon_t)\} = 0$  e podemos escrever

$$g(\varepsilon_t) = \begin{cases} (\theta + \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{se } \varepsilon_t \geq 0 \\ (\theta - \gamma)\varepsilon_t - \gamma E(|\varepsilon_t|), & \text{se } \varepsilon_t < 0. \end{cases} \quad (5.35)$$

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 (5.33) para um EGARCH  $(m, n)$  fica

$$\ln(h_t) = \alpha_0 + \frac{1 + b_1 B + \dots + b_{n-1} B^{n-1}}{1 - a_1 B - \dots - a_m B^m} g(\varepsilon_{t-1}),$$

onde os operadores em B têm raízes fora do círculo unitário. Note que em (5.33) 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

$$\ln(h_t) = \alpha_0 + \frac{g(\varepsilon_{t-1})}{1 - \alpha_1 B}.$$

Se usarmos o **EViews** para estimar modelos, esse software considera o modelo EGARCH na forma:

$$\ln(h_t) = w + \beta \ln(h_{t-1}) + \alpha \left| \frac{r_{t-1}}{\sqrt{h_{t-1}}} \right| + \gamma \frac{r_{t-1}}{\sqrt{h_{t-1}}}. \quad (5.36)$$

O **S+FinMetrics** usa uma especificação similar. Quando  $\gamma < 0$  o efeito assimétrico deve ser incorporado ao modelo GARCH. Observe também que podemos escrever  $\varepsilon_t$  no lugar de  $r_t/\sqrt{h_t}$  nas fórmulas acima.

**Exemplo 5.3.** A série de retornos diários da Petrobras já foi analisada no Exemplo 5.1, no qual ajustamos um modelo AR(15)-ARCH(9). Um modelo AR(15)-GARCH(1,1) com erros gaussianos também pode ser ajustado, resultando

$$\begin{aligned} r_t &= 0,086r_{t-1} - 0,042r_{t-3} + 0,048r_{t-10} + 0,042r_{t-15} + a_t, \\ a_t &= \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\ h_t &= 0,000015 + 0,082a_{t-1}^2 + 0,890h_{t-1}. \end{aligned} \quad (5.37)$$

Esses 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. Usando o **EViews**, obtemos

$$\begin{aligned}
 r_t &= 0,097r_{t-1} - 0,033r_{t-3} + 0,057r_{t-10} + 0,050r_{t-15} + a_t, \\
 a_t &= \sqrt{h_t}\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\
 \ln h_t &= -0,375 + 0,967 \ln h_{t-1} + 0,163|\varepsilon_{t-1}| - 0,079\varepsilon_{t-1},
 \end{aligned} \tag{5.38}$$

A análise dos resultados revela que  $\gamma < 0$  e que o efeito assimétrico deve ser incorporado ao modelo. O teste de Ljung-Box aplicado aos resíduos e aos quadrados dos resíduos 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.

A estimativa do desvio padrão condicional encontra-se na Figura 5.7.

Para comparar os três modelos, (5.20), (5.37) e (5.38), ajustados aos retornos da Petrobras, 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(15)-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 modelos ajustados aos retornos da Petrobras

Modelo	AR(15)+ ARCH(9)	AR(15)+ GARCH(1,1)	AR(15)+ EGARCH(1,1)
log-verossim.	7078,097	7082,285	7110,447
AIC	-4,7362	-4,7437	-4,7619
BIC	-4,7081	-4,7297	-4,7458

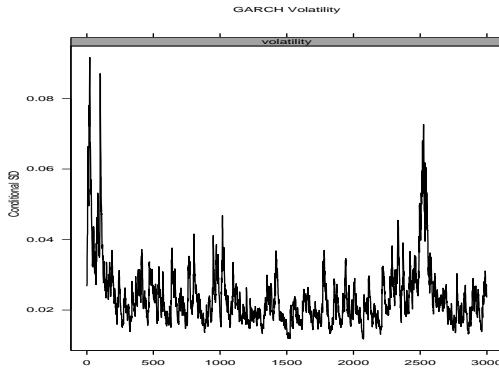


Figura 5.7: Estimativa do desvio padrão condicional para os retornos da Petrobras usando o modelo (5.38).

### 5.4.2 Modelos TGARCH

O modelo TGARCH (*threshold GARCH*) 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 e Runkle, 1993).

O EViews e o S+FinMetrics usam a formulação

$$h_t = w + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 d_{t-1} + \beta h_{t-1},$$

com

$$d_t = \begin{cases} 1, & \text{se } r_t < 0 \text{ (\textit{bad news})} \\ 0, & \text{se } r_t \geq 0 \text{ (\textit{good news})}. \end{cases}$$

Se  $\gamma \neq 0$ , há um impacto de informação assimétrica. Esperamos que  $\gamma > 0$  para que *bad news* tenha um impacto maior.

**Exemplo 5.4.** Considere a série de retornos diáários da Vale, no período de 31/08/1998 a 29/09/2010. Não é difícil verificar que um modelo AR(3)-GARCH(1,1) é adequado para descrevê-la. Verificaremos se existe impacto de informação assimétrica, ajustando um modelo AR(3)-TGARCH(1,1) com erros gaussianos aos retornos. O modelo ajustado é

$$\begin{aligned} r_t &= -0,051r_{t-3} + a_t, \\ a_t &= \sqrt{h_t} \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1) \\ h_t &= 0,00013 + 0,044a_{t-1}^2 + 0,063a_{t-1}^2 d_{t-1} + 0,910h_{t-1}, \end{aligned} \tag{5.39}$$

com  $d_t$  dado acima.

Analizando os resultados, 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. Uma comparação dos critérios AIC e BIC mostra que o modelo assimétrico ajusta-se melhor que o simétrico. Uma figura similar à Figura 5.7 é obtida para o desvio padrão condicional,  $\sqrt{h_t}$ .

## 5.5 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, 1982, 2008) não faz essa suposição. Esse modelo tem como premissa o fato de 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,  $\sigma_t^2 = E(r_t^2 | \mathcal{F}_{t-1})$ .

**Definição 5.4.** *Dizemos que a série  $r_t$  segue um modelo de volatilidade estocástica se*

$$r_t = \sigma_t \varepsilon_t, \quad (5.40)$$

$$\sigma_t = e^{h_t/2}, \quad (5.41)$$

em que  $\varepsilon_t$  é uma sequência estacionária, com média zero e  $h_t$  é uma sequência estacionária, com densidade de probabilidade  $f(h)$ .

A formulação mais simples do modelo, dada por Taylor (1982), supõe que o logaritmo da volatilidade,  $h_t = \log \sigma_t^2$ , seja dado por

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad (5.42)$$

na qual  $\eta_t$  é uma sequência 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

$$r_t = \beta e^{h_t/2} \varepsilon_t, \quad (5.43)$$

$$h_{t+1} = \mu + \alpha_1(h_t - \mu) + \sigma_\eta \eta_t, \quad (5.44)$$

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

$$r_t = \sqrt{h_t} \varepsilon_t, \quad (5.45)$$

$$\ln(h_t) = \alpha_0 + \alpha_1 \ln(h_{t-1}) + \sigma_\eta \eta_t. \quad (5.46)$$

Sabemos que se  $\varepsilon_t \sim \mathcal{N}(0, 1)$ , então  $\ln(\varepsilon_t^2)$  tem uma distribuição chamada *log-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 (5.40) e (5.42) obtemos

$$\ln(r_t^2) = \ln(\sigma_t^2) + \ln(\varepsilon_t^2), \quad (5.47)$$

$$h_t = \ln(\sigma_t^2) = \alpha_0 + \alpha_1 h_{t-1} + \eta_t. \quad (5.48)$$

Chamando  $\xi_t = \ln(\varepsilon_t^2) - E(\ln(\varepsilon_t^2)) \simeq \ln(\varepsilon_t^2) + 1,27$ , temos que  $E(\xi_t) = 0$ ,  $\text{Var}(\xi_t) = \pi^2/2$

$$\ln(r_t^2) = -1,27 + h_t + \xi_t, \quad \xi_t \sim \text{i.i.d. } (0, \pi^2/2), \quad (5.49)$$

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \eta_t, \quad \eta_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\eta^2). \quad (5.50)$$

Aqui, supomos  $\xi_t$  e  $\eta_t$  independentes.

### Propriedades

Vamos calcular agora alguns parâmetros associados ao MVE, considerando-se a forma (5.42):

(i)  $E(r_t) = E(\sigma_t \varepsilon_t) = E(\sigma_t)E(\varepsilon_t) = 0$ , dado que  $\sigma_t$  e  $\varepsilon_t$  são independentes.

(ii)  $\text{Var}(r_t) = E(r_t^2) = E(\sigma_t^2 \varepsilon_t^2) = E(\sigma_t^2)E(\varepsilon_t^2) = E(\sigma_t^2)$ .

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

$$h_t \sim \mathcal{N}\left(\frac{\alpha_0}{1-\alpha_1}, \frac{\sigma_\eta^2}{1-\alpha_1^2}\right). \quad (5.51)$$

Como  $h_t$  é normal,  $\sigma_t^2$  é log-normal, logo temos

$$E(r_t^2) = E(\sigma_t^2) = e^{\mu_h + \sigma_h^2/2}.$$

Não é difícil mostrar que

$$E(r_t^4) = 3e^{2\mu_h + 2\sigma_h^2},$$

da qual obtemos a curtose

$$K = \frac{3e^{2\mu_h + 2\sigma_h^2}}{e^{2\mu_h + \sigma_h^2}} = 3e^{\sigma_h^2} > 3, \quad (5.52)$$

como deveríamos esperar, ou seja, caudas longas sob o MVE.

(iii) A função de autocovariância da série  $r_t$  é dada por

$$\gamma_r(u) = E(r_t r_{t+u}) = E(\sigma_t \sigma_{t+u} \varepsilon_t \varepsilon_{t+u}) = 0,$$

pois  $\varepsilon_t$  e  $\eta_t$  são independentes. Logo  $r_t$  é serialmente não correlacionada, mas não independente, pois existe correlação em  $\ln(r_t^2)$ . Denotando-se  $Y_t = \ln(r_t^2)$ , então a autocovariância 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 é 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 autocovariâncias 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 (5.42), 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  $\gamma_\xi(u) = 0$ , para  $u > 0$ . Logo,  $\gamma_Y(u) = \gamma_h(u)$ , para todo  $u \neq 0$ , e podemos escrever a função de autocorrelação 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

$$\rho_Y(u) = \frac{\alpha_1^u}{1 + \pi^2/2\sigma_h^2}, \quad u > 0, \quad (5.53)$$

que tende a zero exponencialmente a partir do lag 2, o que indica que  $Y_t = \ln(r_t^2)$  pode ser modelada por um modelo ARMA(1, 1).

Na prática, obtemos valores de  $\alpha_1$  próximos de um, o que implica o aparecimento de altas correlações para volatilidades e consequentes grupos de volatilidades na série.

Um MVE geral será obtido admitindo-se um modelo AR(p) para  $h_t$ :

$$r_t = \sigma_t \varepsilon_t, \quad (5.54)$$

$$(1 - \alpha_1 B - \dots - \alpha_p B^p) h_t = \alpha_0 + \eta_t, \quad (5.55)$$

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 de que a volatilidade tem memória longa, no sentido que a função de autocorrelação de  $\ln(r_t^2)$  decai lentamente, embora, como vimos, os  $r_t$  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 (5.40)-(5.42) é reescrito na forma

$$r_t = \sigma \varepsilon_t e^{h_t/2}, \quad (5.56)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (5.57)$$

em que  $\sigma = \exp\{\alpha_0/2\}$ . Uma forma equivalente é dada por

$$\ln(r_t^2) = \kappa + h_t + u_t, \quad (5.58)$$

$$h_t = \alpha_1 h_{t-1} + \eta_t, \quad (5.59)$$

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 (5.58)-(5.59) estão na formulação denominada de espaço de estados: (5.58) é a *equação de observação* e (5.59) é a *equação de estado*.

### Observações:

- (i) Quando  $\alpha_1$  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  $\alpha_1 = 1$ ,  $h_t$  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 logarítmica especificada em (5.58). Uma solução sugerida por Fuller e analisada por Breidt e Carriquiry (1996) é fazer a seguinte transformação baseada numa expansão de Taylor:

$$\ln(r_t^2) = \ln(r_t^2 + cS_r^2) - cS_r^2/(r_t^2 + cS_r^2), \quad t = 1, \dots, T,$$

em que  $S_r^2$  é a variância amostral da série  $r_t$  e  $c$  é um número pequeno.

O programa STAMP (Koopman et al., 2000) pode ser utilizado na obtenção de estimadores de quase-verossimilhança (QMV) dos parâmetros do modelo, escrito na forma de espaço de estados. Esse 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 ele. Veja também Shephard e Pitt (1997) e Kim et al. (1998). Uma resenha do problema de estimação do MVE é feita por Motta (2001).

O pacote stochvol do R produz estimativas para os MVE, usando uma abordagem bayesiana via MCMC. Veja Kastner (2016). Utiliza-se a parametrização de Kim et al. (1998), descrita em (5.43)-(5.44).

Sob o ponto de vista bayesiano, podemos escrever:

$$r_t|h_t \sim \mathcal{N}(0, e^{h_t}), \quad (5.60)$$

$$h_t|\{h_{t-1}, \mu, \alpha_1, \sigma_\eta\} \sim \mathcal{N}(\mu + \alpha_1(h_{t-1} - \mu), \sigma_\eta^2), \quad (5.61)$$

$$h_t|\{h_0, \mu, \alpha_1, \sigma_\eta\} \sim \mathcal{N}(\mu, \sigma_\eta^2/(1 - \alpha_1^2)). \quad (5.62)$$

Chamemos  $\theta = (\mu, \alpha_1, \sigma_\eta)'$  o vetor de parâmetros. A distribuição a priori para  $\theta$  é considerada da forma

$$p(\theta) = p(\mu)p(\alpha_1)p(\sigma_\eta), \quad (5.63)$$

sendo que as prioris para  $\mu$ ,  $\alpha_1$  e  $\sigma_\eta$  são dadas por

$$p(\mu) \sim \mathcal{N}(b_\mu, B_\mu), \quad (5.64)$$

$$(\alpha_1 + 1)/2 \sim \text{Beta}(a_0, b_0), \quad (5.65)$$

$$\sigma_\eta^2 \sim B_\sigma \times \chi_1^2 = \text{Gamma}(1/2, 1/2B_\sigma). \quad (5.66)$$

Para prioris mais informativas, veja Kim et al. (1998), Jacquier et al. (2004) e Omori et al. (2007). Para mais detalhes sobre a escolha dessas prioris, veja Kastner (2016).

**Exemplo 5.5.** Vamos analisar a série de retornos diários do Ibovespa, mas no período 03/01/1995 a 27/12/2000, com  $T = 1.498$  observações, usando o MVE e o programa STAMP. O Quadro 5.5 apresenta os resultados da estimação do modelo, fornecendo

$$\begin{aligned}\ln(r_t^2) &= -8,6220 + h_t + u_t, \\ h_t &= 0,9858h_{t-1} + \eta_t, \\ \hat{\text{Var}}(u_t) &= 2,7418, \\ \hat{\text{Var}}(\eta_t) &= 0,0236.\end{aligned}\tag{5.67}$$

Quadro 5.5: Ajustamento de um MVE aos retornos do Ibovespa

```
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)
Ar1 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]
Ar1 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
```

A análise da f.a.c. residual, juntamente com o teste de Ljung-Box, sugerem que os resíduos do modelo são não correlacionados. Uma análise residual mais completa

é fornecida pela Figura 5.8. A estimativa da volatilidade ( $e^{ht}$ ) e do desvio padrão condicional é apresentada na Figura 5.9. Vale a pena observar que  $\hat{\alpha}_1 = 0,9858$ , o que indica uma equivalência de ajustamento entre os modelos GARCH (1,1) e VE.

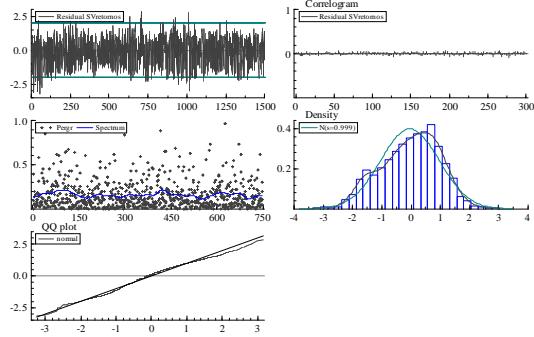


Figura 5.8: Análise residual do ajustamento do modelo (5.58) à série de retornos do Ibovespa.

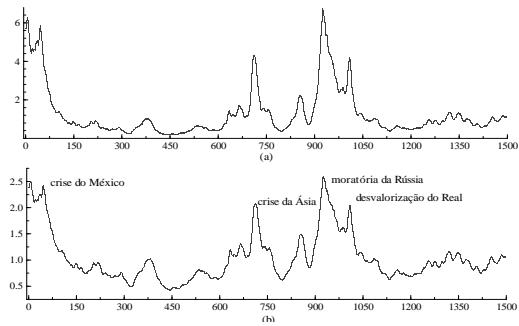


Figura 5.9: MVE para os retornos do Ibovespa; (a) estimativa da volatilidade; (b) estimativa do desvio padrão condicional.

Vamos, agora, ajustar um MVE para a série usando o programa `stochvol`. O comando `svsample` simula, via MCMC, a distribuição a posteriori baseada no modelo

proposto (5.60)-(5.62) e nas prioris (5.64)-(5.66). Chamando de `rt.ibov` os retornos do Ibovespa, temos:

```
> post.ibv <- svsample(rt.ibov)
> summary(post.ibov, showlatent=FALSE)
```

No Quadro 5.6 temos a descrição dos hiperparâmetros das distribuições a priori: para  $\mu$ , `priormu=c(0,100)`, ou seja, uma normal com média 0 e desvio padrão 100; para  $\phi$ , os valores de  $a_0$  e  $b_0$  (`priorphi`) são 5 e 1,5, resultando numa distribuição Beta com média 0,54 e desvio padrão 0,31; finalmente, para  $\sigma_\eta$ , tomamos  $B_\sigma = 1$ .

Quadro 5.6: Distribuições a priori utilizadas

```
svsample:
y: # Série de log-retornos
burnin=1000: #Número de observações iniciais a serem descartadas
draws=10000: # número de observações simuladas após o descarte
priormu=c(0,100): #hiperparâmetros da distribuição de mu
priorphi=c(5,1.5): # hiperparâmetros da distribuição de phi
priorsigma=1: # hiperparâmetro da distribuição de sigma.
```

No Quadro 5.7 temos o resultado da estimação, juntamente com os percentis 5%, 50% e 95% das distribuições a posteriori.

Quadro 5.7: Estimadores e percentis das distribuições a posteriori

Prior distributions:						
$\mu$	$\sim \text{Normal}(\text{mean} = 0, \text{sd} = 100)$					
$(\phi+1)/2$	$\sim \text{Beta}(a_0 = 5, b_0 = 1.5)$					
$\sigma^2$	$\sim 1 * \text{Chisq}(\text{df} = 1)$					
Posterior draws of parameters (thinning = 1):						
	mean	sd	5%	50%	95%	ESS
$\mu$	-7.670	0.2531	-8.058	-7.674	-7.277	3077
$\phi$	0.973	0.0084	0.959	0.974	0.986	419
$\sigma$	0.226	0.0269	0.184	0.225	0.273	234
$\exp(\mu/2)$	0.022	0.0028	0.018	0.022	0.026	3077
$\sigma^2$	0.052	0.0124	0.034	0.051	0.075	234

Observe que obtivemos  $\hat{\phi} = 0,973$ , comparado com 0,9858 com o STAMP. Também, compare as estimativas das variâncias de  $\eta_t$ , 0,052 contra 0,0236, respectivamente.

A Figura 5.10 mostra as estimativas da volatilidade, os traços dos parâmetros estimados via simulação MCMC e as respectivas densidades a posteriori. A Figura

5.11 mostra o gráfico dos resíduos e do gráfico  $Q \times Q$ , indicando um bom ajuste. Os comandos necessários são:

```
> plot(post.ibv),
> resid.ibov<-resid(post.ibv),
> plot(resid.ibv,cex=0.5).
```

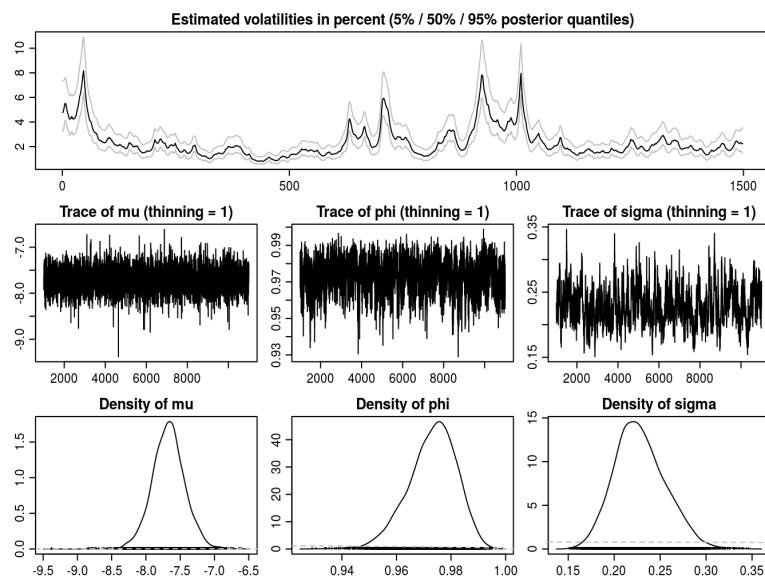


Figura 5.10: Estimativas da volatilidade, traços dos parâmetros e densidades a posteriori.

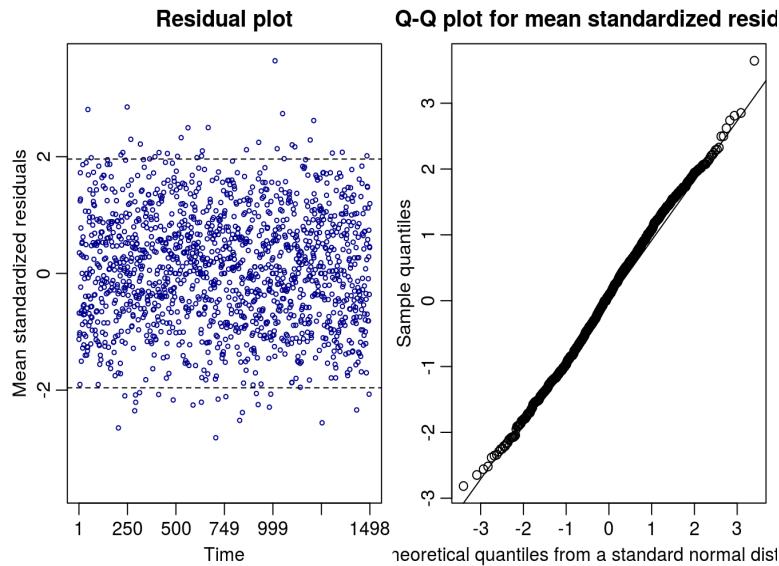


Figura 5.11: Resíduos e  $Q \times Q$  plot dos resíduos.

A Figura 5.12 compara os quantis da distribuição a posteriori simulada da volatilidade com a volatilidade observada (dada por  $r_t^2$ ). Vemos que a primeira capta os principais picos correspondentes a eventos observados no período. Os comandos são:

```
> plot.ts(rt.ibov2),
> volplot(post.ibv).
```

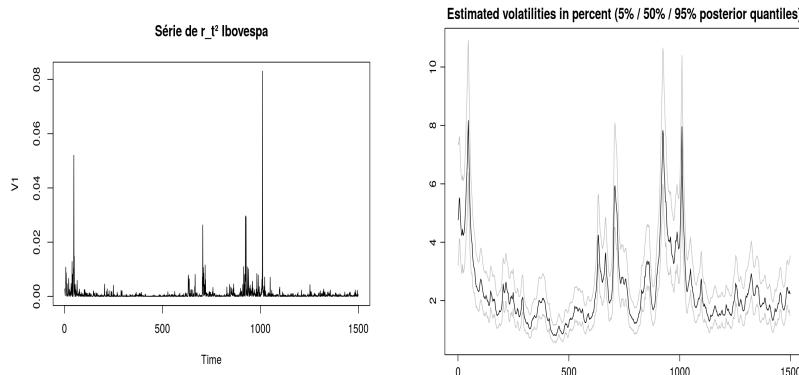


Figura 5.12: Série dos quadrados dos retornos e estimativa da volatilidade.

## 5.6 Tópicos Adicionais

[1] O MVE que aprendemos supõe que os erros das duas equações sejam normais. Trabalhos recentes utilizam outras distribuições na equação de volatilidades, como a família de mistura de normais na escala, com o objetivo de tratar as caudas pesadas presentes nos retornos financeiros. Casos especiais dessa família são a distribuição *t* de Student, a distribuição *Slash* e a distribuição *variance gamma*. Além disso, modelos com efeito alavanca (*leverage effect*) também podem ser combinados a esses modelos com erros seguindo uma mistura de normais na escala. Para detalhes veja Cathy et al. (2008) e Abanto-Valle et al. (2010).

[2] Além dos modelos estudados neste capítulo, há uma variedade muito grande de modelos da família ARCH que foram considerados na literatura. Veja o Problema 13 para um exemplo importante e Bollerslev (2008).

[3] Há trabalhos que usam distribuições estáveis, tanto na equação dos retornos, como na equação da volatilidade. Veja Mittnik et al. (1999), Nowicka-Zagrajek e Weron, (2001) e Sampaio e Morettin (2015).

## 5.7 Problemas

1. Ajuste modelos ARCH com erros gaussianos para as séries:
    - (a) log-retornos diários das ações da Vale, de 31/08/1998 a 29/9/2010.
    - (b) log-retornos diários do índice Ibovespa, de 04/07/1994 a 29/09/2010.
    - (c) log-retornos mensais do S& P500, de janeiro de 1962 a dezembro de 1999.
  2. Ajuste modelos GARCH com erros gaussianos para as séries:
    - (a) log-retornos diários das ações da Petrobras.
    - (b) log-retornos diários do índice Ibovespa.
  3. Ajuste um modelo assimétrico (EGARCH ou TGARCH) com erros gaussianos à série de log-retornos mensais das ações da IBM de janeiro de 1926 a dezembro de 1997.
  4. Ajuste um modelo de volatilidade estocástica às séries:
    - (a) log-retornos diários das ações da Vale, no período de 31/08/1998 a 29/9/2010.
    - (b) log-retornos diários das ações da Petrobras, no período de 18/08/1998 a 29/09/2010.
-

- (c) log-retornos diários das ações da CEMIG, de 03/01/95 a 27/12/2000.

Use os programas STAMP e `stochvol`.

5. Obtenha as previsões de origem  $T$  e horizonte  $h$ ,  $h \geq 1$ , para um modelo GARCH(1,2).
6. Suponha que  $r_1, \dots, r_T$  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 (5.27).
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

$$h_t^d = \alpha_0 + \alpha_1(|r_{t-1}| + \gamma r_{t-1})^d + \beta_1 h_{t-1}^d,$$

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 Petrobras do Problema 4 (a).
- (b) Ajuste um modelo PGARCH (1,1), com  $d = 1$ , aos dados da Vale do Problema 1 (a).
14. Ajuste um modelo EGARCH(1,0) aos dados de retornos diários da CEMIG, do Problema 4 (c).

15. Determine  $E(|\varepsilon|)$  para uma distribuição  $t$  de Student padronizada, isto é,  $\varepsilon = X/\sqrt{\nu/(\nu-2)}$ , onde  $X \sim t_\nu$ .
16. Prove (5.11).
17. Mostre que um modelo GARCH(1,1) é equivalente a um modelo ARCH( $\infty$ ), com pesos exponencialmente decrescentes.

## 5.8 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

$$\begin{aligned} f(x|\iota, \nu) &= \frac{2}{\iota + 1/\iota} [g(\iota(sx+m)|\nu) I_{(-\infty, 0)}(x+m/s)] \\ &\quad + \frac{2}{\iota + 1/\iota} [g((sx+m)/\iota|\nu) I_{(0, +\infty)}(x+m/s)], \end{aligned}$$

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.

---

# Processos com Memória Longa

## 6.1 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,

$$|\rho_j| \leq Cr^j, \quad j = 1, 2, \dots \quad (6.1)$$

onde  $C > 0$  e  $0 < r < 1$ . A expressão (6.1) garante que a função de autocorrelação decai para zero exponencialmente.

Um processo de memória longa é um processo estacionário em que a função de autocorrelação decresce hiperbolicamente para zero, isto é,

$$\rho_j \sim Cj^{-\alpha}, \quad j \rightarrow \infty, \quad (6.2)$$

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 autocorrelação é como em (6.2), 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 1950) revelaram a presença de memória longa (ML) em dados de séries temporais e espaciais. Essas séries apresentam persistência nas autocorrelações amostrais, isto é, dependência significativa entre observações separadas por um longo intervalo de tempo. Essas autocorrelações apresentam o comportamento dado por (6.2). 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 autocorrelação não é absolutamente somável.

Formalmente, temos a

**Definição 6.1.** Suponha que  $X_t$  tenha autocorrelação  $\rho_j$ . Dizemos que  $X_t$  possui memória longa se

$$\lim_{n \rightarrow \infty} \sum_{j=-n}^n |\rho_j| \quad (6.3)$$

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, devida ao efeito estufa. Veja Seater (1993), por exemplo.

Recentemente (década de 1980), 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 6.1.** A Figura 6.1 mostra a conhecida série de índices de preços anuais de trigo de Beveridge (1921), de 1500 a 1869 ( $T = 370$ ), e suas autocorrelações amostrais, notando o seu lento decaimento. Os preços são médias sobre várias localidades na Europa (arquivo a-bev00.69.dat).

**Exemplo 6.2.** Temos, na Figura 6.2, as autocorrelações amostrais das séries de valores absolutos dos retornos diários do Ibovespa, Dow Jones, Vale e Petrobras. Esses valores absolutos representam a volatilidade da série. O lento decaimento das autocorrelações mostra claramente a persistência da volatilidade. As figuras mostram, também, as autocorrelações de modelos autorregressivos AR( $p$ ) ajustados às séries. Os valores de  $p$  para as séries do Ibovespa, Dow, Vale e Petrobras são, respectivamente, 27, 12, 27 e 27. Vemos que as autocorrelações dos modelos autorregressivos são boas estimativas para *lags* baixos. Notamos, ainda, o número excessivo de parâmetros do modelo autorregressivo necessários para capturar a dependência nas séries.

---

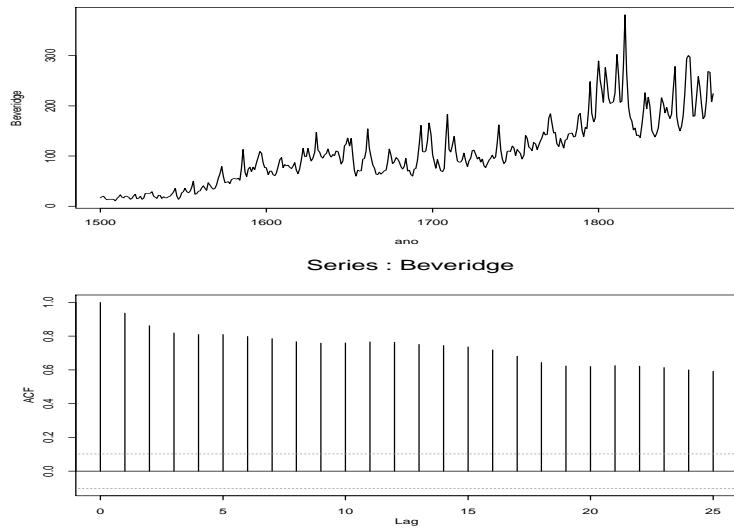


Figura 6.1: (a) Série de índices de preços de trigo de Beveridge; (b) f.a.c. amostral.

Uma outra característica de séries com memória longa é que as autocorrelações da série original indicam não estacionariedade, ao passo que a série diferenciada pode parecer ser “superdiferençada”. 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 autocorrelação é do tipo (6.2). Primeiro, foi introduzido o ruído gaussiano fracionário por Mandelbrot e Van Ness (1968). Mais tarde, Granger e Joyeux (1980) e Hosking (1981) introduziram 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 memória longa associados a modelos de volatilidade estocástica foram considerados por Harvey (1998) e Breidt et al. (1993). Veja a Seção 6.6.

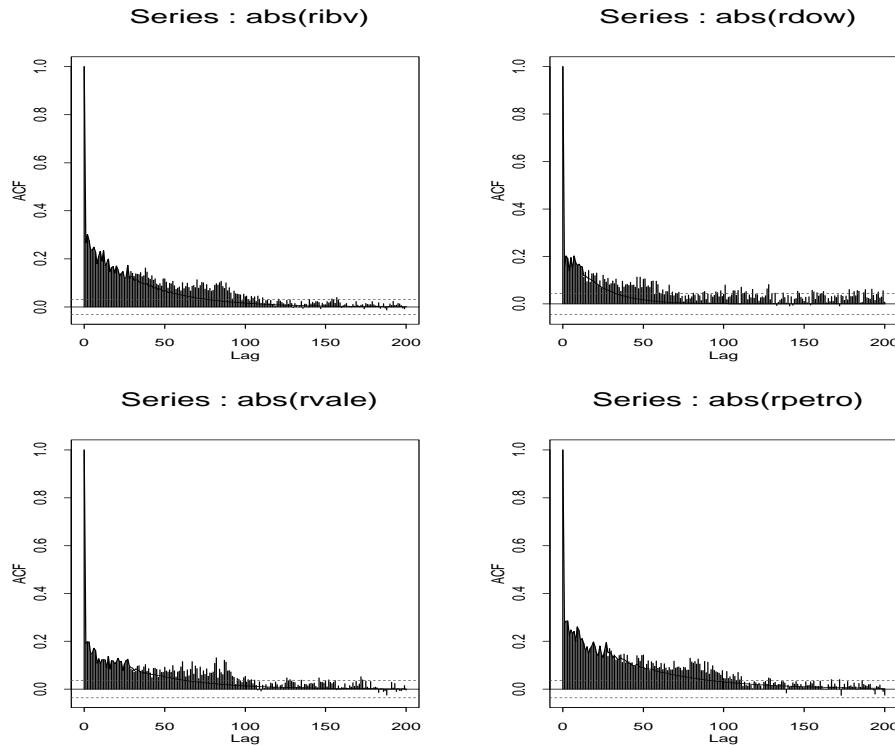


Figura 6.2: Autocorrelações amostrais da volatilidade: (a) Ibovespa; (b) Dow Jones; (c) Vale; (d) Petrobras, com f.a.c. dos modelos AR ajustadas.

## 6.2 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*

$$(1 - B)^d(X_t - \mu) = u_t, \quad (6.4)$$

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

$$\begin{aligned} (1 - B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\ &= 1 - dB + \frac{1}{2!} d(d-1)B^2 - \frac{1}{3!} d(d-1)(d-2)B^3 + \dots, \end{aligned} \quad (6.5)$$

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 antipersistente.

### 6.2.1 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

$$Q_T = \frac{1}{S_T} \left[ \max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) \right], \quad (6.6)$$

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 dessa variável limite. Ele nota que a estatística definida por (6.6) não é robusta à dependência de curta memória e propõe substituir  $Q_T$  por

$$\tilde{Q}_T = \frac{1}{\hat{\sigma}_T(q)} \left[ \max_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) - \min_{1 \leq k \leq T} \sum_{j=1}^k (X_j - \bar{X}) \right], \quad (6.7)$$

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 autocorrelações 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 R/S converge para sua distribuição limite à taxa  $T^{1/2}$ , mas se há ML presente, a taxa de convergência é  $T^H$ .

Esses 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 desse 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.

A função `rosTest` do `S+FinMetrics` calcula (6.7) com  $q = [4(T/100)^{1/4}]$ . Essa 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 6.3.** Considere os retornos diários do Ibovespa de 1994 a 2010 e a série de volatilidades, representada pelos valores absolutos dos retornos. Essa série é mostrada na Figura 6.3. O Quadro 6.1 mostra o resultado da aplicação da função `rosTest`. O valor da estatística  $\tilde{Q}_T$  é 2,741, significativa com o nível 0,01, o que confirma que a série apresenta memória longa. A Figura 6.4 apresenta o log-log plot de R/S, com a reta ajustada. O valor estimado de  $d$  é  $\hat{d} = 0,17$ . O gráfico foi feito com  $k_1 = 5$  e  $f = 2$ . A reta pontilhada no gráfico indica MC ( $H = 1/2$ ).

Quadro 6.1: Teste para ML para volatilidade do Ibovespa

Test for Long Memory: Modified R/S Test  
Null Hypothesis: no long-term dependence

Test Statistics:  
2.741\*\*  
\*: significant at 5% level  
\*\*: significant at 1% level

Total Observ.: 4018  
Bandwidth: 10

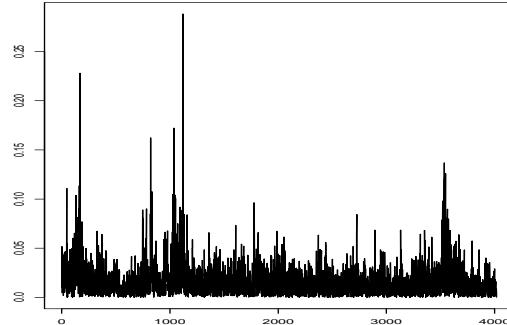


Figura 6.3: Valores absolutos dos retornos do Ibovespa.

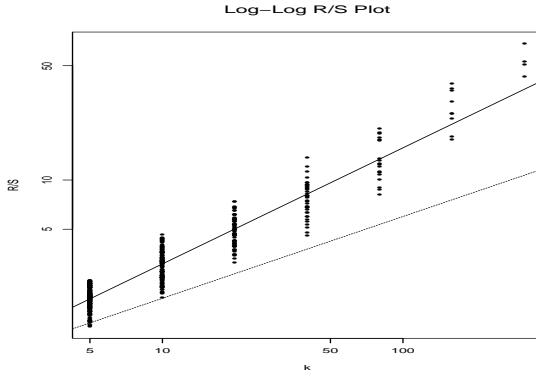


Figura 6.4: Plot R/S para valores absolutos dos retornos do Ibovespa.

### 6.2.2 Procedimento GPH

Esse método para estimativa do parâmetro de longa memória foi proposto por Geweke e Porter-Hudak (1983) e baseia-se na equação que exibe relação entre os espectros de  $X_t$  e de  $u_t$  em (6.4). Tal equação foi reescrita para que se assemelhasse a uma equação de regressão linear, onde o coeficiente de inclinação envolve o parâmetro  $d$ .

De (6.4) temos que

$$f_x(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_u(\lambda), \quad (6.8)$$

em que  $f_u(\lambda)$  é o espectro de  $u_t$ .

Multiplicando ambos os lados de (6.8) por  $f_u(0)$  e aplicando o logaritmo, obtemos

$$\ln f_x(\lambda) = \ln f_u(0) - d \ln |1 - e^{-i\lambda}|^2 + \ln \left( \frac{f_u(\lambda)}{f_u(0)} \right). \quad (6.9)$$

Substituindo  $\lambda$  por  $\lambda_j = 2\pi j/T$  (frequência de Fourier) e adicionando  $\ln(I_x(\lambda_j))$ , a ambos os lados de (6.9), obtemos

$$\begin{aligned} \ln I_x(\lambda_j) &= \ln f_u(0) - d \ln \left( 4 \operatorname{sen}^2 \left( \frac{\lambda_j}{2} \right) \right) \\ &\quad + \ln \left( \frac{f_u(\lambda_j)}{f_u(0)} \right) + \ln \left( \frac{I_x(\lambda_j)}{f_x(\lambda_j)} \right), \end{aligned} \quad (6.10)$$

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  $\ln(\frac{f_u(\lambda_j)}{f_u(0)})$  pode ser desprezado quando se considerar apenas as frequências  $\lambda_j$  próximas de zero. Assim, podemos reescrever (6.10) como um modelo de regressão linear

$$Y_j = a - dX_j + \varepsilon_j, \quad j = 1, \dots, m, \quad (6.11)$$

em que

$$\begin{aligned} Y_j &= \ln I_x(\lambda_j), \\ X_j &= \ln(4\sin^2(\frac{\lambda_j}{2})), \\ \varepsilon_j &= \ln\left(\frac{I_x(\lambda_j)}{f_x(\lambda_j)}\right), \\ a &= \ln f_u(0) \text{ e } m = cT^\alpha, \quad 0 < \alpha < 1, \end{aligned}$$

$c$  uma constante. A relação linear (6.11) sugere a utilização de um estimador de mínimos quadrados para  $d$ , isto é,

$$\hat{d}_{MQ} = -\frac{\sum_{i=1}^m (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^m (X_i - \bar{X})^2}. \quad (6.12)$$

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

$$t_{d=0} = \hat{d} \left( \frac{\pi^2}{6 \sum_{j=1}^m (X_j - \bar{X})^2} \right)^{-1/2} \quad (6.13)$$

tem distribuição normal padrão.

O estimador  $d$ , calculado por meio de (6.12) 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  $\alpha = 0, 5$ .

O parâmetro  $d$  pode também ser estimado por máxima verossimilhança, juntamente com os parâmetros de um processo ARFIMA ajustado a uma série temporal com memória longa. Veja a seção seguinte.

**Exemplo 6.4.** Consideremos novamente os valores absolutos dos retornos diários do Ibovespa, do Exemplo 6.3. O Quadro 6.2 apresenta o resultado da aplicação da função `gphTest`. O valor da estatística é 3,9466, e a hipótese nula de que não há

memória longa é rejeitada com o nível 1%. O valor estimado de  $d$  é  $\hat{d} = 0,3557$ . Além disso, baseado no erro padrão assintótico fornecido, 0,0901, obtemos o intervalo de confiança [0,1757; 0,5357] para  $d$ , com coeficiente de confiança de 95%.

Quadro 6.2: Teste para ML para volatilidade do Ibovespa, usando GPH

Test for Long Memory: GPH Test Null Hypothesis: $d = 0$  Test Statistics: $d$ 0.3557 stat 3.9466** *: significant at 5% level **: significant at 1% level  Total Observ.: 4018 Number of Freq: 63 [1] 0.0901
---

### 6.3 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 6.2.** Dizemos que  $\{X_t\}$  é um processo autorregressivo 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

$$\phi(B)(1 - B)^d X_t = \theta(B)a_t, \quad (6.14)$$

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 (6.14) 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 autocorrelação e densidade espectral**

Hosking (1981) também mostra que se  $X_t$ , dado por (6.14), 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 6.5.** O caso mais simples é o *ruído branco fracionário*, ou seja, um ARFIMA(0,  $d$ , 0), dado por

$$(1 - B)^d X_t = a_t, \quad a_t \sim \text{RB}(0, \sigma_a^2). \quad (6.15)$$

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

$$\psi_k \sim \frac{k^{d-1}}{(d-1)!} = c_1 k^{d-1}, \quad k \rightarrow \infty$$

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, autocorrelação, autocorrelação parcial e a variância são dadas, respectivamente, por

$$f(\lambda) = \begin{cases} \frac{\sigma_a^2}{2\pi} \left(2\sin\left(\frac{\lambda}{2}\right)\right)^{-2d}, & 0 < \lambda \leq \pi, \\ \lambda^{-2d}, & \lambda \rightarrow 0, \end{cases} \quad (6.16)$$

$$\rho_h = \frac{(-d)!(h+d-1)!}{(d-1)!(h-d)!} = \prod_{0 < k \leq h} \frac{k-1+d}{k-d}, \quad h=1, 2, \dots \quad (6.17)$$

$$\begin{aligned} \phi_{hh} &= \frac{d}{h-d}, \quad h=1, 2, \dots \\ \gamma_0 &= \frac{(-2d)!}{(-d)!^2}. \end{aligned}$$

Em particular, temos que

$$\rho_1 = \frac{d}{1-d}, \quad (6.18)$$

$$\rho_h \sim \frac{(-d)!h^{2d-1}}{(d-1)!} = c_3 h^{2d-1}, \quad h \rightarrow \infty,$$

sendo  $c_3$  constante e

$$f(\lambda) \sim \lambda^{-2d}. \quad (6.19)$$

A Figura 6.5(a) apresenta  $N = 100$  observações simuladas de um modelo ARFIMA(0,  $d$ , 0) com  $d = 0,45$  e a Figura 6.6 (a) apresenta o gráfico das autocorrelações.

**Exemplo 6.6.** 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

$$f(\lambda) = \begin{cases} \frac{\sigma_a^2}{2\pi} \frac{(2\sin(\frac{\lambda}{2}))^{-2d}}{1+\phi^2-2\phi\cos\lambda}, & 0 < \lambda \leq \pi, \\ \frac{\lambda^{-2d}}{(1-\phi)^2}, & \lambda \rightarrow 0; \end{cases}$$

(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 6.5(b) apresenta  $N = 100$  observações simuladas de um processo ARFIMA(1,  $d$ , 0) com  $\phi = 0,8$  e  $d = 0,45$  e a Figura 6.6 (b) apresenta o gráfico das f.a.c.

**Exemplo 6.7.** 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 autorregressiva 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 do Exemplo 6.6 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 6.5 (c) apresenta  $N = 100$  observações de um processo ARFIMA  $(0, d, 1)$ , com  $d = 0,45$  e  $\theta = 0,3$  e a Figura 6.6 (c) apresenta o respectivo gráfico das f.a.c.

**Exemplo 6.8.** Finalmente, a Figura 6.5 (d) apresenta  $N = 100$  observações simuladas de um processo ARFIMA  $(1, d, 1)$  com  $\phi = 0,8$ ,  $\theta = 0,3$  e a Figura 6.6 (d) apresenta o gráficos das autocorrelações.

Em todos os exemplos citados do processo ARFIMA  $(p, d, q)$ , podemos notar o comportamento da função de autocorrelação, que tem decaimento hiperbólico.

Para mais detalhes, veja Hosking (1981) e Granger e Joyeux (1980).

## 6.4 Estimação de modelos ARFIMA

Nesta seção, vamos apresentar o método de máxima verossimilhança para obter estimadores dos parâmetros do modelo (6.14). Um outro método que pode ser utilizado é a estimação semiparamétrica 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

$$L(\boldsymbol{\eta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-T/2} (r_0 \cdots r_{T-1})^{-1/2} \exp \left[ -\frac{1}{2\sigma_a^2} \sum_{j=1}^T (X_j - \hat{X}_j)^2 / r_{j-1} \right], \quad (6.20)$$

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  $r_{j-1} = (\sigma_a^2)^{-1} E(X_j - \hat{X}_j)^2$ .

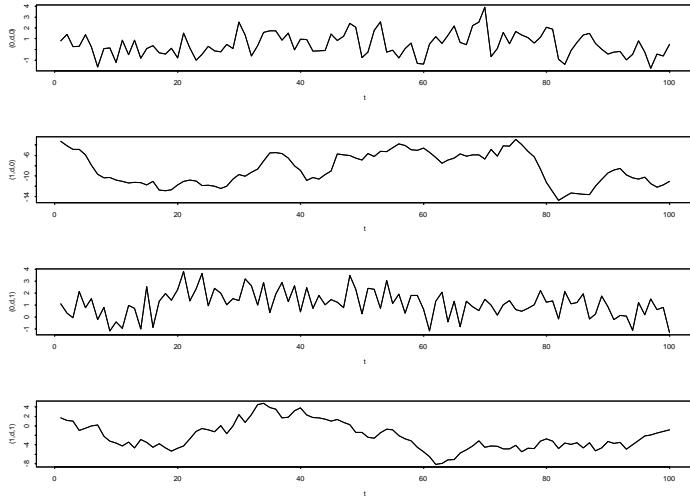


Figura 6.5: Séries ARFIMA( $p, d, q$ ) geradas de acordo com os Exemplos 6.5 a 6.8.

Os estimadores de máxima verossimilhança dos parâmetros são dados por

$$\hat{\sigma}_{MV}^2 = T^{-1} S(\hat{\boldsymbol{\eta}}_{MV}), \quad (6.21)$$

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

$$\ell(\boldsymbol{\eta}) \simeq \ell_*(\boldsymbol{\eta}) = \ln \frac{1}{T} \sum_j \frac{I_T(w_j)}{2\pi f(w_j; \boldsymbol{\eta})}, \quad (6.22)$$

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,

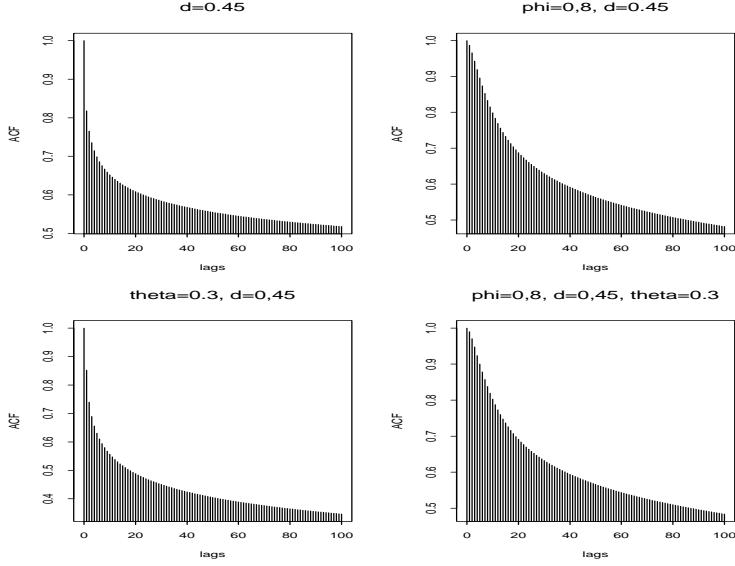


Figura 6.6: F.a.c. das séries ARFIMA( $p, d, q$ ) geradas de acordo com os Exemplos 6.5 a 6.8.

$$f(w_j; \boldsymbol{\eta}) = \frac{\sigma_a^2}{2\pi} \frac{|1 - \theta_1 e^{-iw_j} - \dots - \theta_q e^{-qiw_j}|^2}{|1 - \phi_1 e^{-iw_j} - \dots - \phi_p e^{-piw_j}|^2} \cdot |1 - e^{-iw_j}|^{-2}$$

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  $\hat{\eta}_{MV}$  que minimiza (6.22) é consistente;
- (ii) se  $d > 0$ ,

$$\hat{\boldsymbol{\eta}}_{MV} \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{\eta}, T^{-1} A^{-1}(\boldsymbol{\eta})), \quad (6.23)$$

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

$$\hat{\sigma}_{MV}^2 = \frac{1}{T} \sum_j \frac{I_T(w_j)}{2\pi f(w_j; \hat{\eta}_{MV})}.$$

O estimador de  $d$  obtido dessa 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 esses 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 (2006) para detalhes. A função d.whittle também pode ser usada para o caso ARFIMA  $(0,d,0)$ .

**Exemplo 6.9.** Consideremos a série de valores absolutos dos retornos diárias da Petrobras, de 18/08/1998 a 29/09/2010. 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 ARFIMA  $(2,d,1)$ , com  $\hat{d} = 0,157$ , 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 6.3 mostra um resumo da saída do programa.

Segue-se que o modelo ajustado é

$$(1 - 0,883B - 0,093B^2)(1 - B)^{0,157}X_t = (1 - 0,948)a_t.$$

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).
```

## 6.5 Previsão de modelos ARFIMA

Considere o processo ARFIMA  $(p, d, q)$  estacionário e invertível,

$$\phi(B)(1 - B)^d X_t = \theta_0 + \theta(B)a_t, \quad -0,5 < d < 0,5. \quad (6.24)$$

Podemos reescrever o processo na forma de choques aleatórios,

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j}, \quad (6.25)$$

e na forma invertida

$$\sum_{j=0}^{\infty} \pi_j Z_{t-j} = \theta_0 + a_t, \quad (6.26)$$

onde

$$\sum_{j=0}^{\infty} \psi_j B^j = \theta(B) \phi^{-1}(B) (1-B)^{-d}$$

e

$$\sum_{j=0}^{\infty} \pi_j B^j = \phi(B) \theta^{-1}(B) (1-B)^d.$$

Quadro 6.3: Estimação do modelo ARFIMA(2,d,1)

	Value	Std. Error	t value	P-value
d	0.1570	0.0329	4.7785	0.0000
AR(1)	0.8826	0.0380	23.2407	0.0000
AR(2)	0.0926	0.0333	2.7834	0.0054
MA(1)	0.9479	0.0063	149.9936	0.0000
	log-likelihood		BIC	
	7932.911		-15833.803	
BIC	of all	models	estimated:	
	q=0	q=1	q=2	
p=0	-15730.30	-15800.36	-15799.91	
p=1	-15793.87	-15829.81	-15806.90	
p=2	-15812.14	-15833.80	-15833.29	

Assim, podemos fazer previsões de valores futuros do processo  $X_t$ , utilizando as equações (6.25) ou (6.26). A variância do erro de previsão, também pode ser calculada de modo usual.

Uma outra forma é usar a da equação de diferenças

$$\varphi(B)X_t = \theta_0 + \theta(B)a_t, \quad (6.27)$$

em que  $\varphi(B) = \phi(B)(1-B)^d = \phi(B)D(B)$  e  $D(B) = 1 - d_1 B - d_2 B^2 - \dots$  é um polinômio em  $B$ , com coeficientes dados por

$$d_j = \frac{-\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0 < k \leq j} \frac{k-1-d}{k}, \quad j = 0, 1, 2, \dots \quad (6.28)$$

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 (6.28) e as expressões (6.25) e (6.26), 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 6.10.** Para o Exemplo 6.9, 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 36 valores previstos, constantes do objeto `petro.pred`, estão mostrados no Quadro 6.4.

## 6.6 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 essa seja uma variável observada. Uma ideia 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 6.

Quadro 6.4: Valores previstos para o modelo ARFIMA

-0.0019	-0.0021	-0.0035	-0.0042	-0.0046	-0.0049	-0.0051	-0.0051	-0.0051
-0.0050	-0.0050	-0.0049	-0.0047	-0.0045	-0.0044	-0.0042	-0.0039	-0.0037
-0.0035	-0.0033	-0.0030	-0.0028	-0.0025	-0.0023	-0.0020	-0.0018	-0.0015
-0.0013	-0.0010	-0.0008	-0.0006	-0.0003	-0.0001	0.0002	0.0005	0.0007

Um modelo GARCH ( $r,s$ ) pode ser escrito como um modelo ARMA ( $q,s$ ), com  $q = \max(r, s)$ , na forma (veja (5.25))

$$\phi(B)X_t^2 = \alpha_0 + \beta(B)\nu_t, \quad (6.29)$$

onde  $\nu_t = X_t^2 - \sigma_t^2$ , são as inovações seguindo um processo diferença martingale e

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_q B^q,$$

$$\beta(B) = 1 - \beta_1 B - \dots - \beta_s B^s,$$

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 decaimento hiperbólico nas inovações quadráticas e pesos persistentes.

O modelo FIGARCH (q,d,s) é definido por

$$\phi(B)(1 - B)^d X_t^2 = \alpha_0 + \beta(B)\nu_t, \quad (6.30)$$

onde  $\phi(B)$  e  $\beta(B)$  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. Note que

$$\beta(B)\sigma_t^2 = \alpha_0 + [\beta(B) - \phi(B)(1 - B)^d]X_t^2.$$

Para garantir que um processo FIGARCH seja estacionário e  $\sigma_t^2 > 0$ , restrições complicadas têm que ser impostas aos coeficientes do modelo. Veja Baillie et al. (1996) para detalhes.

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

$$\phi(B)(1 - B)^d \ln \sigma_t^2 = \alpha_0 + \sum_{j=1}^s (\beta_j |\varepsilon_{t-j}| + \gamma_j \varepsilon_{t-j}), \quad (6.31)$$

onde  $\phi(B)$  é definido como em (6.29) e  $\varepsilon_t$  são os resíduos padronizados, ou seja,  $\varepsilon_t = X_t / \sigma_t$ . Em (6.31), 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 6.11.** 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 6.5 e 6.6, respectivamente. Como os retornos são não correlacionados, não há necessidade de ajustar previamente um modelo ARMA. Na Tabela 6.1, 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. Note que o modelo FIGARCH ajustado tem  $d = 1$ , e portanto reduz-se a um modelo IGARCH.

O diagnóstico e previsão desses modelos podem ser feitos de modo similar aos modelos ARFIMA e GARCH. Veja Zivot e Wang (2006) para detalhes.

Tabela 6.1: 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

## 6.7 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 diárias da Vale (arquivo d-vale98.10.dat).
- (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  $d$  por meio de regressão.
4. Considere um processo ARFIMA  $(0, d, 0)$  estacionário e invertível. Mostre que a função densidade espectral é dada por (6.16).

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.

Quadro 6.5: Modelo FIGARCH(1,1) ajustado aos retornos do DJIA				
	Value	Std. Error	t value	P-value
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		
Ljung-Box	test for	standard. residuals:		
Statistic	P-value	Chi <sup>2</sup> -d.f.		
21.57	0.04261	12		
Ljung-Box test	for squared	stand. residuals:		
Statistic	P-value	Chi <sup>2</sup> -d.f.		
11.73	0.4675	12		
TR <sup>2</sup>	Lagrange	Multiplier	Test	
	P-value	F-stat	P-value	
11.52	0.4847	1.054	0.5084	

6. Mostre que a função densidade espectral de um processo ARFIMA( $1, d, 0$ ) estacionário de invertível é dada por

$$f(\lambda) = \frac{\sigma_a^2}{2\pi} \frac{[2\sin(0, 5\lambda)]^{-2d}}{1 + \phi^2 - 2\phi \cos \lambda}, \quad 0 < \lambda \leq \pi,$$

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  $d = 0, 45$ .
- Faça um gráfico dos dados simulados e comente.
  - Calcule as f.a.c. e f.a.c.p. amostrais e comente.
  - Estime os parâmetros do modelo, testando a significância de cada um deles.
  - Ajuste um modelo ARMA a  $Y_t = (1 - B)X_t$ .
  - Compare o ajustamento dos modelos ARFIMA (item (c)) e ARMA (item (d)).

Quadro 6.6: Modelo FIEGARCH(1,1) ajustado aos retornos do DJIA				
	Value	Std. Error	t value	P-value
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				
Ljung-Box	test for	stand. residuals:		
Statistic	P-value	Chi <sup>2</sup> -d.f.		
26.43	0.009338	12		
Ljung-Box	test for squared	stand. residuals:		
Statistic	P-value	Chi <sup>2</sup> -d.f.		
17.61	0.1279	12		
TR <sup>2</sup>	Lagrange	Multiplier	Test	
18	P-value	F-stat	P-value	
	0.1157	1.651	0.1745	

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.
- Calcule as funções de autocorrelação e autocorrelação parcial amostrais da série  $v_t$  e comente.
  - Identifique e estime um modelo de memória longa para a série  $v_t$ .
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 para a 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 (6.30) pode ser escrito na forma

$$\sigma_t^2 = \alpha_0 \beta(B)^{-1} + \lambda(B) X_t^2,$$

onde  $\lambda(B) = 1 - \beta(B)^{-1} \phi(B)(1 - B)^d$ .

12. Ajuste modelos FIGARCH e FIEGARCH aos retornos do Ibovespa (arquivo d-ibv94.10.dat) e da Vale (arquivo d-vale98.10.dat). Se necessário, tome retornos em porcentagem.

## 6.8 Apêndice 6. 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 (intradiários) é possível estimar a volatilidade diária por meio da chamada *volatilidade realizada*. Veja o Capítulo 8 para detalhes. Por exemplo, se tivermos os preços de uma ação, negociada na Bolsa, a cada cinco minutos no dia  $t$ , e se a Bolsa opera seis horas nesse dia, teremos 72 observações,  $P_{t,1}, \dots, P_{t,72}$ . Considerando os log-retornos, teremos  $r_{t,0}, r_{t,1}, \dots, r_{t,72}$ , 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 intradiários 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:  $u_t = M_t - A_t$ ,  $d_t = m_t - A_t$ ,  $c_t = F_t - A_t$ ,  $M_t$  é o logaritmo do preço máximo e  $m_t$  o logaritmo do preço mínimo.

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 6.12.** Consideremos a série de índices diários do DJIA, do Exemplo 1.2, com  $T = 1.992$ . 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 esse ajuste.

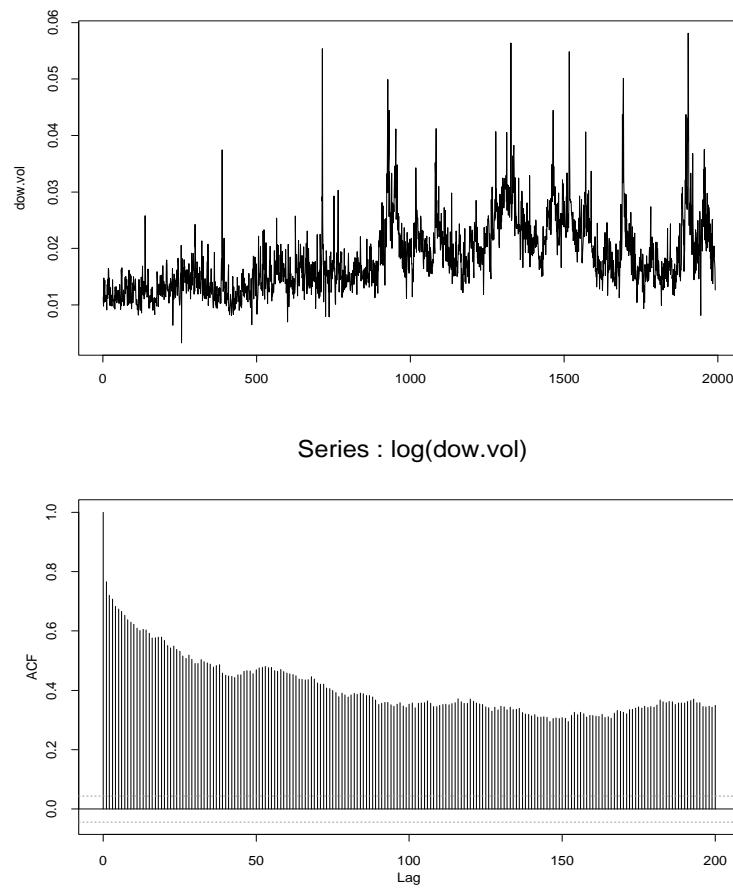


Figura A.1: Volatilidade de GK e f.a.c.

Quadro A.1: Ajuste do modelo ARFIMA (0,d,1)				
	Value	Std. Error	t value	P-value
d	0.4868	0.0257	18.9395	0.0000
MA(1)	0.1293	0.0325	3.9746	0.0001
log-likelihood		BIC		
511.2452		-1007.2997		
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	

# Valor em Risco

## 7.1 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. Nos Capítulos 1 e 5 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 ela 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 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+, de-

senvolvido pelo Crédit Suisse First Bank e o CreditMetrics, proposto pelo banco J.P. Morgan. Os interessados podem consultar os documentos preparados por essas instituições. Uma referência interessante e que compara as duas metodologias é Gordy (2000).

## 7.2 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 7.1.** Suponha que exista uma chance de 95% de que a taxa de câmbio Real/USD não caia em um dia. Suponha, ainda, que uma empresa tenha 100 milhões de reais aplicados num fundo cambial. Calculemos a perda potencial sobre esse 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 (0,95)-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. Essa operação é realizada alugando-se o ativo.  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 7.1.** Definimos o VaR de uma posição comprada sobre o horizonte  $h$ , com probabilidade  $p$ ,  $0 < p < 1$ , por meio de

$$p = P(\Delta P(h) \leq \text{VaR}) = F_h(\text{VaR}). \quad (7.1)$$

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 (7.1) 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 porcentuais. Assim sendo, usaremos log-retornos no que segue.
- (ii) A Definição 7.1 mostra que o VaR é o  $p$ -quantil da distribuição  $F_h(\cdot)$ . Na prática, teremos que estimar esse quantil, usando por exemplo a distribuição empírica dos retornos.
- (iii) O VaR calculado em (7.1) 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 7.1, 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

$$p = P(\Delta P(h) \geq \text{VaR}) = 1 - F_h(\text{VaR}), \quad (7.2)$$

que geralmente é positivo para  $p$  pequeno. O sinal positivo aqui indica perda.

As definições (7.1) e (7.2) 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 (7.1) 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.

### 7.3 VaR Usando a Distribuição Normal

Nesta seção, discutiremos a abordagem do RiskMetrics (veja Longerstaey e 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*)

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2, \quad 0 < \lambda < 1. \quad (7.3)$$

O valor do parâmetro de suavização  $\lambda$  pode ser escolhido minimizando-se o erro quadrático médio (EQM) de ajuste ou de previsão. A metodologia RiskMetrics sugere usar  $\lambda = 0,94$ . Manteiga (2002) encontra, para séries brasileiras, valores que variam entre 0,83 e 0,91. Para a série de retornos diárias da Petrobras, no período 1998-2010, obtivemos  $\lambda = 0,83$ . Contudo, por simplicidade, vamos usar o valor 0,90 em todas as nossas aplicações.

No caso de uma carteira com  $m$  ativos, poderemos proceder como segue. Seja  $\hat{\lambda}_i$  o valor ótimo (segundo o critério de menor EQM) para o  $i$ -ésimo ativo, e seja  $\tau_i$  o correspondente valor do EQM. Se  $\pi = \sum_{i=1}^m \tau_i$ , os pesos dos ativos são obtidos por meio de  $w_i = \theta_i^{-1} / \sum \theta_i^{-1}$ , onde  $\theta_i = \tau_i / \pi$ . O valor ótimo de  $\lambda$  é dado, então, por  $\tilde{\lambda} = \sum_{i=1}^m w_i \hat{\lambda}_i$ .

Uma maneira equivalente de formular o modelo é usar a abordagem de modelos GARCH, notando-se que (7.3) 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 desse retorno, pode ser calculada usando-se resultados da modelagem GARCH. De fato, como podemos escrever

$$\begin{aligned}\log P_t - \log P_{t-1} &= r_t = a_t, \\ a_t &= \sigma_t \varepsilon_t \sim \text{IGARCH}(1, 1), \\ \varepsilon_t &\sim \mathcal{N}(0, 1),\end{aligned}$$

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 (7.3)

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) \sigma_{t-1}^2 \varepsilon_{t-1}^2,$$

ou, ainda,

$$\sigma_t^2 = \sigma_{t-1}^2 + (1 - \lambda) \sigma_{t-1}^2 (\varepsilon_{t-1}^2 - 1).$$

Dessa 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  $E(\varepsilon_{t+i-1}^2 - 1 | \mathcal{F}_t) = 0$ ,  $i \geq 2$ , obtemos

$$E(\sigma_{t+i}^2 | \mathcal{F}_t) = E(\sigma_{t+i-1}^2 | \mathcal{F}_t), \quad i = 2, \dots, k. \quad (7.4)$$

Para previsão a um passo da volatilidade, de (7.3) 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 (7.4) mostra que

$$\text{Var}(r_{t+i} | \mathcal{F}_t) = \sigma_{t+1}^2, \quad i = 2, \dots, k,$$

ou seja,

$$\sigma_t^2[k] = k \hat{\sigma}_t^2(1). \quad (7.5)$$

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 (7.3), 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 essa usualmente chamada “regra da raiz quadrada do tempo”. Note que  $\sigma_{t+1} = \sqrt{\hat{\sigma}_t^2(1)}$ .

Por exemplo, se fixarmos em (7.1) a probabilidade  $p = 5\%$ , então o RiskMetrics usa  $-1,65\sigma_{t+1}$  como VaR, que é o 0,05-quantil da normal com média zero e variância  $\sigma_{t+1}^2$ . Normalmente o sinal negativo, que significa perda, é ignorado e

$$\text{VaR} = (\text{Valor da posição}) \times (1,65) \times (\sigma_{t+1}).$$

Essa 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,

$$\text{VaR}[k] = \sqrt{k} \text{VaR}. \quad (7.6)$$

**Exemplo 7.1.** (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 7.2.** Suponha que uma empresa tenha uma posição comprada de 10 milhões de reais em ações da Petrobras e queremos calcular o VaR de 1 e 15 dias. Consideremos os log-retornos diáridos da Petrobras, com  $T = 2.990$  observações, de 31/08/1998 a 29/09/2010, mostrados na Figura 7.1. Temos, então,

$$\sigma_t^2 = 0,90\sigma_{t-1}^2 + 0,10r_{t-1}^2.$$

Dos dados obtemos  $r_{2990} = 0,029522$  e, do modelo ajustado, obtemos  $\hat{\sigma}_{2990}^2 = 0,000545$ . Logo, a estimativa da volatilidade 1 passo a frente é dada por  $\sqrt{\hat{\sigma}_{2990}^2(1)} = \sqrt{0,000578} = 0,024404$ . Se  $p = 0,05$ , o quantil da distribuição é  $(-1,65)(0,024404) = -0,040267$ . O sinal negativo indica perda e o VaR de um dia será

$$\text{VaR} = 10.000.000,00 \times 0,040267 = 402.670,00,$$

ao passo que o VaR de 15 dias será

$$\text{VaR}[15] = \sqrt{15} \times 402.670,00 = 1.559.534,00.$$

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 7.2. Compare com a Figura 5.2.

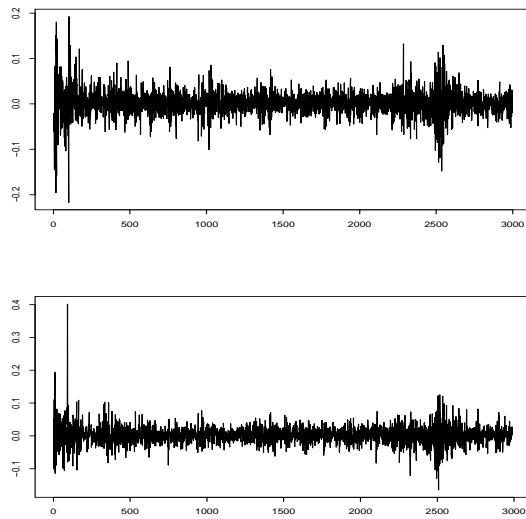


Figura 7.1: Retornos diáários da Petrobras e Vale.

**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 (7.3), o modelo

$$\gamma_{ij,t} = \lambda \gamma_{ij,t-1} + (1 - \lambda) r_{i,t-1} r_{j,t-1}. \quad (7.7)$$

Dessa maneira, é fácil ver que o VaR da carteira é dado por

$$\text{VaR} = \sqrt{\sum_{i=1}^m \text{VaR}_i^2 + 2 \sum_{i < j} \rho_{ij} \text{VaR}_i \text{VaR}_j}, \quad (7.8)$$

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 7.3.** Considere uma carteira com dois ativos, sendo uma posição de 10 milhões de reais em ações da Petrobras e outra de 5 milhões de reais em ações da Vale, ambos com  $T = 2.990$  observações. Os retornos diários da Vale estão mostrados na Figura 7.1. Vamos calcular o VaR de um dia da carteira, usando (7.8). No exemplo anterior, vimos que  $\text{VaR}_1 = -0,040267$ . Para calcular o VaR para a segunda posição, ainda supondo que  $\lambda = 0,90$ , obtemos  $\hat{\sigma}_{2990}^2(1) = 0,000215$  e o  $\text{VaR}_2 = -1,65\sqrt{0,000215} = -0,024206$ . Em reais, esse VaR é de  $5.000.000,00 \times 0,024206 = 121.030,00$ . O desvio padrão condicional dos retornos da Vale estão na Figura 7.2. Compare com a Figura 5.9.

A seguir, usaremos (7.7) para estimar a covariância prevista no instante  $t = 2.990$ . Usando  $\lambda = 0,90$  para a carteira, temos que

$$\hat{\gamma}_{12,2990}(1) = 0,90 \gamma_{12,2990} + 0,10 r_{1,2990} r_{2,2990},$$

sendo que  $r_{1,2990} = 0,02952$ ,  $r_{2,2990} = 0,004348$  e  $\gamma_{12,2990} = 0,0001041$ . Obtemos, então,  $\hat{\gamma}_{12,2990}(1) = 0,000107$ . Segue-se que a correlação  $\rho_{12}$  entre os dois ativos é estimada por

$$\hat{\rho}_{12,2990}(1) = (0,000107) / (0,024404 \times 0,014670) = 0,297545.$$

Logo, o VaR de um dia da carteira é

$$\begin{aligned} \text{VaR} &= \sqrt{(-0,040267)^2 + (-0,024206)^2 + 2(0,297545)(-0,040267)(-0,024206)} \\ &= 0,052796. \end{aligned}$$

O VaR da carteira em reais é 453.650,00, usando (7.8).

Na Figura 7.3, temos ilustradas as correlações entre os retornos, ao longo do tempo. Nesse exemplo usamos o programa S+FinMetrics para fazer os cálculos e as figuras.

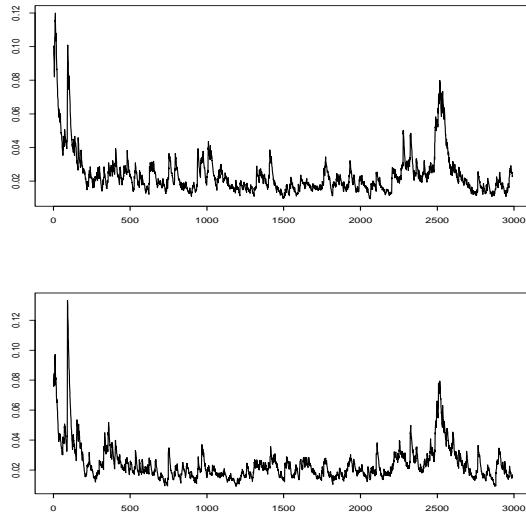


Figura 7.2: Desvios padrões condicionais dos retornos da Petrobras e Vale.

## 7.4 VaR Usando Modelos ARMA e GARCH

Lembremos que uma série de retornos é, em geral, não correlacionada, mas dependente. Se esse for o caso, a volatilidade será modelada por um dos modelos heteroscedásticos considerados no Capítulo 5. Mas vimos também que algumas séries de retornos ainda exibem a presença de autocorrelação, 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  $a_t$  desse modelo por um membro da família ARCH. Por exemplo, se escolhermos um modelo GARCH(m,n) para usar, teremos o modelo ARMA(p,q)-GARCH(m,n):

$$r_t = \phi_0 + \sum_{i=1}^p \phi_i r_{t-i} + a_t - \sum_{j=1}^q \theta_j a_{t-j}, \quad (7.9)$$

$$a_t = \sigma_t \varepsilon_t, \quad (7.10)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i a_{t-i}^2 + \sum_{j=1}^n \beta_j \sigma_{t-j}^2. \quad (7.11)$$

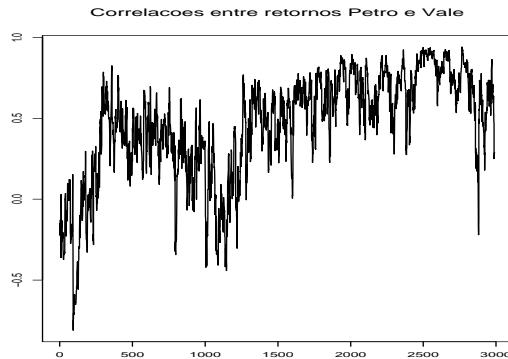


Figura 7.3: Correlações entre retornos da Petrobras e Vale.

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 (7.9) e (7.11), respectivamente.

Supondo-se, por exemplo,  $p = 0, 05$ ,

$$\text{VaR} = \hat{r}_t(1) - 1,65\hat{\sigma}_t(1). \quad (7.12)$$

Escolhendo-se  $\varepsilon_t \sim t_\nu$ , o VaR 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. De fato, chamando de  $t_\nu(p)$  o  $p$ -quantil de  $t_\nu$ , então temos

$$\begin{aligned} p &= P(t_\nu \leq t_\nu(p)) = P\left(\frac{t_\nu}{\sqrt{\nu/(\nu-2)}} \leq \frac{t_\nu(p)}{\sqrt{\nu/(\nu-2)}}\right) = \\ &= P(t_\nu^* \leq t_\nu^*(p)), \quad \nu > 2, \end{aligned}$$

ou seja,  $t_\nu^*(p) = \frac{t_\nu(p)}{\sqrt{\nu/(\nu-2)}}$  é o  $p$ -quantil da distribuição  $t_\nu$  padronizada. Logo,

$$\text{VaR} = \hat{r}_t(1) - \frac{t_\nu(p)\hat{\sigma}_t(1)}{\sqrt{\nu/(\nu-2)}}. \quad (7.13)$$

**Exemplo 7.4.** Consideremos a mesma situação do Exemplo 7.2 e calculemos o VaR, com  $p = 0,05$ . Um modelo adequado para a série é um AR(15) – GARCH(1,1) gaussiano, dado por (5.37). No entanto, vamos usar um modelo mais simples, AR(1) – GARCH(1,1), que também passa por todos os testes de diagnóstico:

$$\begin{aligned} r_t &= 0,00146 + 0,08020r_{t-1} + a_t, \\ a_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1), \\ \sigma_t^2 &= 0,000016 + 0,09772a_{t-1}^2 + 0,87624\sigma_{t-1}^2. \end{aligned}$$

Dos dados, sabemos que  $r_{2989} = 0,007519$  e  $r_{2990} = 0,029522$ , de modo que a previsão da série um passo à frente é dada por

$$\hat{r}_{2990}(1) = 0,00146 + 0,08020r_{2990} = 0,003828$$

A previsão da volatilidade um passo à frente é

$$\hat{\sigma}_{2990}^2(1) = 0,000016 + 0,09772a_{2990}^2 + 0,87624\sigma_{2990}^2.$$

Agora,  $a_{2990} = r_{2990} - 0,00146 - 0,0802r_{2989} = 0,027459$  e  $\sigma_{2990}^2 = 0,000533$  (obtida do modelo GARCH ajustado), de modo que

$$\hat{\sigma}_{2990}^2(1) = 0,000016 + (0,09772)(0,027459)^2 + (0,87624)(0,000533) = 0,000557.$$

O 0,05-quantil será, então,

$$\hat{r}_{2990}(1) - 1,65\hat{\sigma}_{2990}(1) = 0,003828 - 1,65(0,023601) = -0,035114.$$

Finalmente, o valor em risco de um dia é dado por

$$\text{VaR} = 10.000.000,00 \times 0,035114 = 351.140,00.$$

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 (7.9) e (7.11) 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, 2005 e Problema 4).

**Proposição 7.1.** (a) A previsão da média do retorno  $r_t$  no período  $k$  é dada por

$$\hat{r}_T[k] = \hat{r}_T(1) + \dots + \hat{r}_T(k), \quad (7.14)$$

onde  $\hat{r}_T(h)$  é a previsão de origem  $T$  e horizonte  $h$  usando (7.9).

(b) O erro de previsão é dado por

$$e_T[k] = a_{T+k} + (1 + \psi_1)a_{T+k-1} + \dots + (\sum_{i=0}^{k-1} \psi_i)a_{T+1}, \quad (7.15)$$

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

$$\text{Var}(e_T[k]|\mathcal{F}_T) = \hat{\sigma}_T^2(k) + (1 + \psi_1)^2 \hat{\sigma}_T^2(k-1) + \dots + (\sum_{i=0}^{k-1} \psi_i)^2 \hat{\sigma}_T^2(1), \quad (7.16)$$

onde  $\hat{\sigma}_T^2(h)$  é a previsão  $h$  passos à frente da volatilidade usando (7.11).

**Exemplo 7.4.** (continuação). Calculemos o VaR de cinco 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 7.4 é fácil ver que obtemos  $\hat{r}_{2990}[5] = 0,0015874$ . Lembremos que as previsões  $\hat{r}_T(\ell)$  convergem para  $E(r_t) = 0,00159$ , 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 7.4, podemos calcular as previsões da volatilidade para  $h = 1, 2, \dots, 5$ . Finalmente, usando (7.16) obtemos a previsão da volatilidade do retorno de cinco dias, ou seja,

$$\sigma_T^2[5] = \hat{\sigma}_T^2(5) + (1 + \psi_1)^2 \hat{\sigma}_T^2(4) + \dots + (1 + \psi_1 + \dots + \psi_4)^2 \hat{\sigma}_T^2(1),$$

cujo valor resulta  $\sigma_T^2[5] = 0,0034122$ . O valor em risco de cinco dias é, então, com  $p = 0,05$ ,

$$\text{VaR}[5] = 0,0015874 - 1,65 \sqrt{0,0034122} = -0,0947958.$$

Em u.m., obtemos o valor R\$ 947.958,00.

## 7.5 VaR Usando Quantis Empíricos

Uma maneira de estimarmos de forma não paramétrica o VaR é por meio dos quantis empíricos dos dados observados, conforme definidos em (1.21). Rememorando, chamando de  $r_1, \dots, r_T$  os retornos observados, o estimador do  $p$ -quantil é dado por

$$q_p = \begin{cases} r_{(i)}, & \text{se } p = p_i = (i - 0,5)/T, i = 1, \dots, T \\ (1 - f_i)r_{(i)} + f_i r_{(i+1)}, & \text{se } p_i < p < p_{i+1} \\ r_{(1)}, & \text{se } 0 < p < p_1 \\ r_{(T)}, & \text{se } p_T < p < 1, \end{cases} \quad (7.17)$$

onde  $f_i = (p - p_i)/(p_{i+1} - p_i)$ .

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 dada por  $p(1 - p)/T[f(Q(p))]^2$ , 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 7.5.** Considere os mesmos dados do exemplo 7.2 e calculemos o VaR de um dia usando os quantis empíricos. Por (7.17) temos  $p_{150} = 0,049867$ ,  $p_{151} = 0,050200$  e  $f_{150} = 0,40$  de onde

$$q(0,05) = (0,60)r_{(150)} + (0,40)r_{(151)} = -0,040005.$$

Segue-se que o valor em risco de um dia da posição é VaR = 400.050,00 reais.

## 7.6 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 sequências 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 esses máximos sejam independentes parece ser razoável na prática.

Um resumo da TVE, relevante aos nossos propósitos, é dada no Apêndice 7. Vamos nos concentrar no máximo  $r_{(T)}$ . A TVE procura obter a distribuição limite (aproximada) para o máximo normalizado

$$r_T^* = \frac{r_{(T)} - b_T}{a_T}, \quad (7.18)$$

para sequências 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 sequências como acima tais que a distribuição de (7.18) converge para a distribuição não degenerada  $G(z)$ , então  $G$  pertence a uma de três famílias, que podem ser conjuntamente colocadas na forma

$$G(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (7.19)$$

definida sobre  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ , para  $-\infty < \mu < \infty$ ,  $-\infty < \xi < \infty$ ,  $\sigma > 0$ .

A família (7.19) 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, essa 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 dessa distribuição, a partir do qual podemos obter o VaR de uma posição vendida. Note-se que

$$r_{n,i} = \max_{1 \leq j \leq n} \{r_{(i-1)n+j}\}, \quad i = 1, \dots, m. \quad (7.20)$$

**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 (7.20) 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

$$z_{n,i} = \frac{r_{n,i} - b_n}{a_n}. \quad (7.21)$$

Estimadores dos quantis da distribuição dos máximos de grupos são obtidos invertendo-se a equação (7.19). Se  $0 < p^* < 1$ , o  $(1 - p^*)$ -quantil é dado por

$$z_{1-p^*} = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - \{-\log(1 - p^*)\}^{-\xi}], & \text{se } \xi \neq 0 \\ \mu - \sigma \log[-\log(1 - p^*)], & \text{se } \xi = 0, \end{cases} \quad (7.22)$$

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  $z_{1-p^*}$  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 (7.22), 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 (7.22).

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

$$p^* = P(r_{n,i} \geq z_{1-p^*}) = 1 - P(r_{n,i} \leq z_{1-p^*}) = 1 - [P(r_t \leq z_{1-p^*})]^n,$$

do que segue

$$1 - p^* = [1 - P(r_t \geq z_{1-p^*})]^n. \quad (7.23)$$

Para a série original de retornos,  $r_t$ , fixado  $p$ , o  $(1 - p)$ -ésimo quantil de  $r_t$  é  $z_{1-p^*}$  se a probabilidade  $p^*$  for escolhida de (7.23) com  $p = P(r_t \geq z_{1-p^*})$ , logo devemos ter

$$1 - p^* = (1 - p)^n.$$

De (7.22), o VaR de uma posição vendida será dado por

$$\text{VaR} = \begin{cases} \mu_n - \frac{\sigma_n}{\xi_n} \{1 - [-n \log(1 - p)]^{-\xi_n}\}, & \text{se } \xi_n \neq 0 \\ \mu_n - \sigma_n \log[-n \log(1 - p)], & \text{se } \xi_n = 0 \end{cases} \quad (7.24)$$

Para posições compradas vale um raciocínio análogo, trabalhando com  $r_{(1)}$  e a Definição 6.1. Veja o Apêndice 7 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 EM  e  $\mu, \sigma, \xi$ , para o valor fixado de  $n$ .
- (c) Se o modelo ajustado for adequado, use (7.24) para calcular o VaR.

**Exemplo 7.6.** Considere a situação em que um fundo mantém uma posição vendida de 10 milhões de reais em ações da Vale. Consideremos os  $T = 2982$  últimos log-retornos, desprezando-se os primeiros oito dados, para podermos obter  $m = 142$  blocos de  $n = 21$  dias, ou seja, estamos usando máximos mensais. Aos 142 máximos de blocos assim obtidos ajustamos uma distribuição GVE. Utilizamos aqui a função `gev.fit`, constante do software desenvolvido por Coles (2001), que pode ser acessado do site do autor. Os estimadores obtidos são  $\hat{\mu} = 0,0368(0,00139)$ ,  $\hat{\sigma} = 0,0145(0,00118)$  e  $\hat{\xi} = 0,3166(0,0759)$ , onde colocamos entre parênteses os respectivos desvios padrões. O valor da log-verossimilhança resultante é  $-352.0177$ , 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 esses valores em (7.24), com  $p = 0,05$  obtemos  $\text{VaR} = 0,035770$ . Segue-se que o VaR de um dia da posição é de  $(10\text{ milhões}) \times 0,035770 = 357.700,00$ .

O VaR de  $k$  períodos é dado no Problema 5. Uma abordagem diferente, usando excessos sobre (ou abaixo) de um limiar especificado, é apresentada na seção a seguir.

## 7.7 VaR Usando o Método POT

Como sabemos, séries de retornos têm caudas pesadas, e as distribuições estáveis (veja o Capítulo 1) poderiam ser usadas para efeito de modelagem. O problema, como vimos, é que essas distribuições não são caracterizadas por fórmulas simples. Além disso, como as caudas são pesadas, os momentos dessas distribuições, em geral, não existem.

Uma classe de distribuições que trata do caso de decaimento polinomial das densidades é a classe das distribuições generalizadas de Pareto (DGP). A ideia é combinar um estimador não paramétrico para o centro da distribuição e um estimador paramétrico para as caudas. Um problema que surge aqui é identificar onde a cauda (ou caudas) começa( $m$ ). Para dados positivos, normalmente a cauda à direita é que interessa. Para dados que podem ser negativos, zero ou positivos (como os retornos), as duas caudas são de interesse.

Na classe DGP o decaimento é controlado por um parâmetro  $\xi$ , chamado parâmetro de forma ou *índice de cauda*. Por exemplo, no caso da distribuição de Cauchy,  $\xi = 1$ . O comportamento das caudas é comparado com  $(1 + \xi x)^{-1/\xi}$ , quando  $x \rightarrow -\infty$  ou

---

$x \rightarrow \infty$ . O valor de  $\xi$  pode ser diferente para as duas caudas e pode ser estimado por máxima verossimilhança, ou usando o estimador de Hill, mas usualmente obtemos estimadores instáveis. Por essa razão, usamos o método POT (*peaks-over-threshold*), baseado na frequência de pontos na amostra que excedem níveis (*thresholds*) fixados.

Veja o Apêndice 7 para uma exposição sobre as DGP.

Essencialmente, o método POT consiste nas duas etapas a seguir, tendo-se uma amostra  $x_1, x_2, \dots, x_n$  de  $F$ .

- (a) escolha um valor  $u_{sup}$ , no caso de uma cauda, denominado limiar ótimo, a partir do qual estimamos a cauda de  $F$ ; no caso de duas caudas, escolha dois limiares  $u_{inf}$  e  $u_{sup}$ ;
- (b) construa um estimador para a cauda de  $F$ , à direita de  $u_{sup}$  (ou à esquerda de  $u_{inf}$  e à direita de  $u_{sup}$ ), usando (A.8).

Como os parâmetros em (A.8) não são conhecidos, eles terão que ser estimados. Vimos que EMV podem ser utilizados. A função `gpd.tail` do S+FinMetrics pode ser usada para ajustar uma DGP a uma série de excessos acima de um limiar. Também, a função `gpd.fit` do software desenvolvido por Coles (2001) pode ser utilizada.

Mas é necessário escolher o limiar  $u$  antes da estimação do modelo. Suponha que temos os retornos observados  $r_1, \dots, r_T$ . Denotemos por  $r_{(1)}, \dots, r_{(k)}$  os retornos que excedem o limiar  $u$  e por  $z_j = r_{(j)} - u$ ,  $j = 1, \dots, k$ , os excessos sobre o limiar. Então, o objetivo é modelar esses excessos por meio de uma DGP. A escolha de  $u$  é equivalente à escolha do tamanho dos blocos, no caso do uso das distribuições GVE. **Há, basicamente, três métodos para a escolha do limiar: o primeiro, é baseado na média da DGP; o segundo, consiste em usar vários limiares e buscar a estabilidade dos parâmetros do modelo; o terceiro, consiste na análise de gráficos  $Q \times Q$ .**

No primeiro método, não é difícil mostrar que  $P(X - u | X > u)$  é uma função linear de  $u$ . Essa média pode ser estimada por  $\bar{z} = \sum_{j=1}^k z_j$ , onde  $z_j$  são os excessos definidos acima. Logo, fazendo-se um gráfico de  $\bar{z}$  em função de  $u < r_{(T)}$ , podemos identificar o valor de  $u$  tal que, acima dele, esse gráfico seja aproximadamente linear.

No segundo método, de modo similar, pode-se mostrar que, se  $u_0$  for um limiar válido para excessos seguindo uma DGP, estimativas de  $\xi$  e  $\sigma^*$  devem ser constantes acima de  $u_0$ , onde  $\sigma^* = \sigma_u - \xi u$  e  $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$ , se  $u > u_0$ . Logo, podemos selecionar  $u_0$  como o menor valor de  $u$  para o qual estimativas desses parâmetros sejam aproximadamente constantes, num gráfico dessas estimativas em função de  $u$ .

No terceiro método, podemos construir um gráfico quantil-quantil teórico para os dados, admitindo alguma distribuição, e verificando em que pontos se iniciam curvaturas indicativas de caudas pesadas.

Obtido o limiar  $u$ , os parâmetros da DGP podem ser estimados por MV. A log-verossimilhança pode ser escrita, se  $\xi \neq 0$ ,

$$\ell(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^k \log(1 + \xi z_i / \sigma),$$

no caso de  $1 + \xi z_i / \sigma > 0$ . Caso contrário,  $\ell = -\infty$ . Expressão similar para o caso  $\xi = 0$ . A maximização é feita usando-se métodos numéricos.

O diagnóstico do modelo ajustado é feito pelos meios usuais, por exemplo, gráficos  $Q \times Q$ .

Para o cálculo do VaR, suponha que temos uma posição vendida, de modo que estamos interessados na cauda superior. Chamando de  $z_{1-p}$  o  $(1-p)$ -quantil, temos que

$$P(r_t > z_{1-p} | Z > u) = \left[ 1 + \xi \left( \frac{z_{1-p} - \mu}{\sigma} \right) \right]^{-1/\xi},$$

e chamando  $p_u = P(r_t > u)$ , obtemos

$$P(r_t > z_{1-p}) = p_u \left[ 1 + \xi \left( \frac{z_{1-p} - \mu}{\sigma} \right) \right]^{-1/\xi},$$

logo, efetuando os cálculos, obtemos que o VaR é dado por

$$z_{1-p} = u + \frac{\sigma}{\xi} \left[ \left( \frac{p_u}{p} \right)^\xi - 1 \right]. \quad (7.25)$$

Substituindo-se os parâmetros por suas estimativas, obtemos

$$\hat{z}_{1-p} = u + \frac{\hat{\sigma}}{\hat{\xi}} \left[ \left( \frac{\hat{p}_u}{p} \right)^\hat{\xi} - 1 \right], \quad (7.26)$$

onde a estimativa de  $p_u$  é dada pela proporção de pontos acima de  $u$ , ou seja,  $\hat{p}_u = k/T$ . O VaR para uma posição comprada é obtido de forma similar. Veja o Problema 10.

**Exemplo 7.7.** Considere os dados do Exemplo 7.2. Vamos usar as funções apropiadas do software de Coles (2001). Os comandos são descritos a seguir.

```
> mrl.plot(rpetro) : # produz um gráfico das médias de excessos;
> gpd.fitrangle(rpetro, 0.008, 0.03, nint=20): # produz gráficos de parâmetros
   em função de u;
> gpd.fit(rpetro,0.02): # estima os parâmetros de escala e forma
> gpd.petro=gpd.fit(rpetro,0.02)
> gpd.diag(gpd.petro): # produz gráficos de diagnóstico
```

---

Analizando o gráfico quanti-quantil para normalidade e as Figuras 7.4 e 7.5, optamos pelo valor de  $u = 0,02$ . A função `gpd.fit` produz as estimativas, mostradas no Quadro 7.1, que é a saída editada da função `gpd.fit`. Vemos que  $\hat{\sigma} = 0,0137$ ,  $\hat{\xi} = 0,214$  e  $\hat{p}_u = 0,178$ . O número de excessos é 534. O logaritmo da verossimilhança é  $(-1.641,662)$  e os desvios padrões dos estimadores são mostrados na última linha do quadro.

Quadro 7.1: Estimativas dos parâmetros para o Exemplo 7.7

threshold: 0.02
nexc: 534
nnlh: -1641.662
mle: 0.01373293 0.21368425
rate: 0.1781187
se: 0.0009381164 0.0536592631

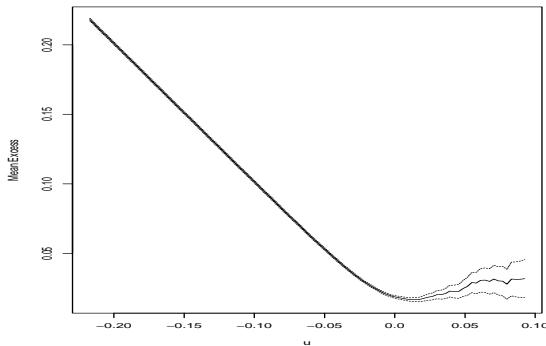


Figura 7.4: Médias de excessos em função de  $u$ .

O cálculo do valor em risco é feito usando (7.26), resultando  $\text{VaR} = 0,04$ , valor similar àqueles encontrados pelos outros métodos. Na Figura 7.6 temos alguns gráficos de diagnóstico, produzidos pela função `gpd.diag`, mostrando que o ajuste foi satisfatório.

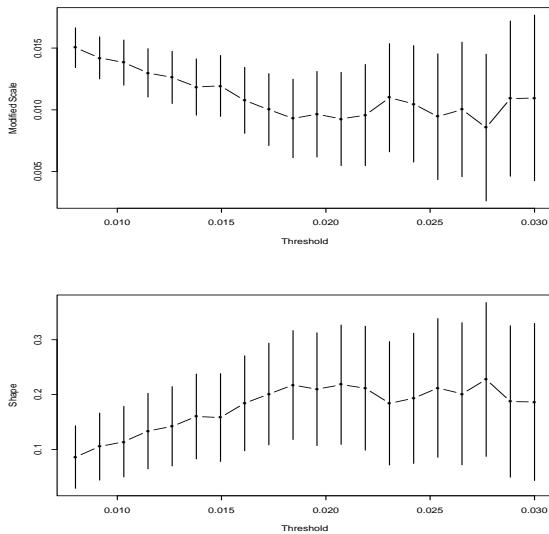


Figura 7.5: EMV e intervalos de confiança de 95% para parâmetros.

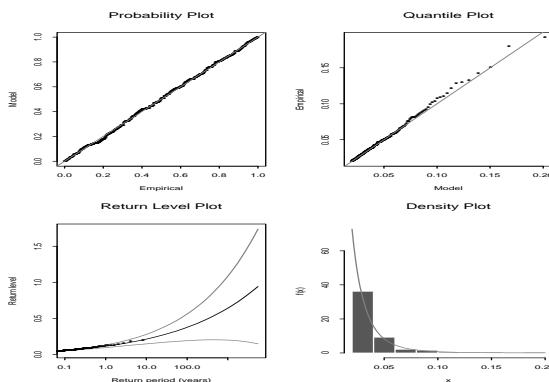


Figura 7.6: Diagnóstico para a DGP ajustada aos dados dos retornos da Petrobras.

## 7.8 Tópicos Adicionais

[1] Há muitas vantagens e limitações no uso do VaR (veja Alexander, 2001). Dentre as primeiras, 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 e O' Brien (2002). Para uma proposta de análise dos excessos, veja Manteiga (2002).

[2] Artzner et al. (1997) introduziram certos requisitos que medidas de risco devem 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)  $\rho(X + Y) \leq \rho(X) + \rho(Y)$ : subaditividade.

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 excede o VaR. Formalmente, se  $X$  representa a perda para uma posição vendida,

$$\text{CVaR} = E(X|X > \text{VaR}). \quad (7.27)$$

O CVaR é também chamado “perda média de excessos” ou ainda *expected short-fall* (ES). 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 a Seção 7.7 e o Apêndice 7 para detalhes.

Para uma posição vendida,

$$\text{CVaR}_p = \frac{1}{p} \int_p^1 \text{VaR}_p dp,$$

e, para uma posição comprada,

$$\text{CVaR}_p = \frac{1}{1-p} \int_0^p \text{VaR}_p dp.$$

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 e Uryasev (2000, 2002) e Andersson et al. (2001).

[3] Engle e Manganelli (2004) 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 – CAVaR*. 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 Bassett, 1978).

[4] Uma outra maneira de calcular medidas de risco é usando os modelos CARE (*conditional autoregressive expectile*), introduzidos por Taylor (2008). Veja também Kuan et al. (2009).

## 7.9 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 um e quinze 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 Petrobras (arquivo d-petro98.10.dat) e suponha uma posição vendida de cinco milhões de reais. Calcule o VaR de um e trinta dias usando:
  - (a) um modelo gaussiano ARMA-GARCH;
  - (b) a teoria de valores extremos (GVE).
  - (c) o método POT.
3. Suponha um fundo de investimentos que possua uma posição comprada com 10 milhoes 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 (GVE).
  - (c) O método POT.
4. Prove a Proposição 7.1.
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. Veja Danielsson e De Vries (1997). Obtenha o Var de 20 dias para o Exemplo 7.6. 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  $k^{-\xi}$ . Nesse 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 Petrobras, usando a teoria dos valores extremos. Use os retornos do arquivo d-petro98.10.dat.
8. Para o Exemplo 7.4, considere o modelo AR(15) – GARCH(1,1), mas agora usando a distribuição  $t_\nu$ . Obtenha o VaR de um dia e de cinco dias.
9. Prove (7.8).
10. Obtenha o VaR usando o método DGP para posições compradas.
11. Para o exemplo 7.6, use a função `gev.diag`, do software de Coles (2001), para verificar se o modelo GVE ajustado é satisfatório.
12. Prove que o CVaR é uma medida coerente de risco.

## 7.10 Apêndice 7. 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,

$$P(r_{(T)} \leq z) = P(r_1 \leq z, \dots, r_T \leq z) = \prod_{i=1}^T P(r_i \leq z) = [F(z)]^T. \quad (\text{A.1})$$

Como  $F$  é desconhecida, (A.1) também o é. Seja  $z_+$  o menor valor de  $z$  tal que  $F(z) = 1$ . Então  $F^T(z) \rightarrow 0$ ,  $T \rightarrow \infty$ , para todo  $z < z_+$ , logo a distribuição de  $r_{(T)}$  é degenerada em  $z_+$ . Esta dificuldade é contornada considerando-se sequências  $\{a_T > 0\}$  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 sequências de constantes  $\{a_T > 0\}$  e  $\{b_T\}$  tais que a distribuição de  $r_T^*$  converja 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.

Essas 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 essa 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 7.7 temos as densidades das três famílias, com  $\xi = 0$  (Gumbel),  $\xi = 0,5$  (Fréchet) e  $\xi = -0,5$  (Weibull).

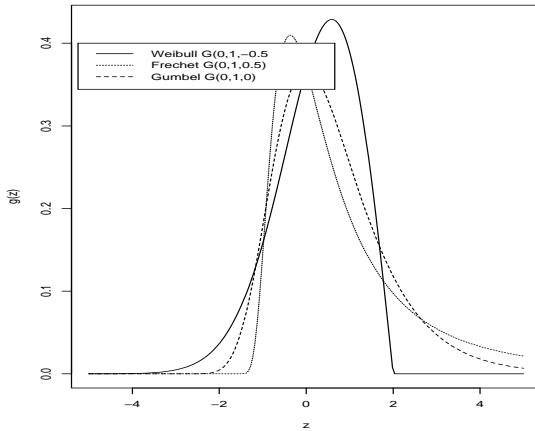


Figura 7.7: Densidades das famílias Fréchet, Weibull e Gumbel.

O teorema acima pode ser usado para aproximar a distribuição de máximos de sequências 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

$$z_{1-p} = \begin{cases} \mu - \frac{\sigma}{\xi} \{1 - [-\log(1-p)]^{-\xi}\}, & \text{se } \xi \neq 0 \\ \mu - \sigma \log[-\log(1-p)], & \text{se } \xi = 0, \end{cases} \quad (A.4)$$

onde  $G(z_p) = 1-p$ , ou seja,  $z_{1-p}$  é o  $(1-p)$ -quantil da distribuição  $G$  e é denominado de *nível de retorno*, associado ao *período de retorno*  $1/p$ .

Ou seja,  $z_{1-p}$  é o valor tal que, com probabilidade  $p$ , é excedido pelos máximos no período. Podemos fazer gráficos de  $z_{1-p}$  contra  $y_p = -\log(1-p)$  (ou contra  $\log y_p$ ), que é linear se  $\xi = 0$ .

A escolha das sequências  $\{a_T\}$  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 sequência  $s_1, \dots, s_T$ , temos que  $r_{(1)} = -s_{(T)}$ . Logo, para  $T$  grande,

$$\begin{aligned}
P(r_{(1)} \leq z) &= P(-s_{(T)} \leq z) = P(s_{(T)} \geq -z) = 1 - P(s_T \leq -z) \\
&= 1 - \exp\{-[1 + \xi(-z - \mu)/\sigma]^{-1/\xi}\} \\
&= 1 - \exp\{-[1 - \xi(z - \tilde{\mu})/\sigma]^{-1/\xi}\}, \tag{A.5}
\end{aligned}$$

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 (2005) 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 esses casos, mas parece que o caso  $\xi \leq -0,5$  raramente ocorre nas aplicações (corresponde a distribuições com caudas a 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

$$\begin{aligned}
\ell(\mu, \sigma, \xi | \mathbf{z}) &= -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^m \log[1 + (z_i - \mu)/\sigma] \\
&\quad - \sum_{i=1}^m [1 + \xi(z_i - \mu)/\sigma]^{-1/\xi}, \tag{A.6}
\end{aligned}$$

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  $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ . Em razão das propriedades dos EMV, esses estimadores terão uma distribuição assintótica normal.

O  $(1-p)$ -quantil pode ser estimado substituindo os EMV dos parâmetros em (A.4) e obtemos

$$\hat{z}_{1-p} = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - y_p^{-\hat{\xi}}), & \text{se } \hat{\xi} \neq 0 \\ \hat{\mu} - \hat{\sigma} \log y_p, & \text{se } \hat{\xi} = 0, \end{cases}$$

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.

### Distribuições de Pareto Generalizadas

Suponha que  $X_1, X_2, \dots$  sejam variáveis independentes, com a mesma f.d.  $F$ , desconhecida. Um evento é considerado extremo se  $X_i > u$ , onde  $u$  é um limiar alto. Segue-se que estaremos interessados em

$$P(X > u + z | X > u) = \frac{1 - F(u + z)}{1 - F(u)}, \quad z > 0.$$

Suponha que  $F$  seja tal que (A.3) valha. Então, para  $u$  suficientemente grande, a distribuição de  $X - u$ , condicional a  $X > u$ , é aproximadamente dada por

$$H(z) = 1 - \left(1 + \frac{\xi z}{\tilde{\sigma}}\right)^{-1/\xi}, \quad (A.8)$$

definida sobre o conjunto  $\{z : z > 0, 1 + \xi z / \tilde{\sigma} > 0\}$ , e  $\tilde{\sigma} = \sigma + \xi(u - \mu)$ . Os parâmetros  $\mu, \sigma$  e  $\xi$  são os mesmos que aparecem na distribuição GVE dada em (A.3). A classe de distribuições dada por (A.8) é denominada *classe de distribuições generalizada de Pareto* (DGP). Pode-se mostrar que  $\xi$  é invariante sob escolhas dos blocos, quando se considera a classe GVE. Veja Coles (2001) para detalhes e uma prova desse resultado. Notamos o seguinte:

- (a) quando  $\xi < 0$ ,  $H(z) = 0$ ,  $z < 0$  e  $H(z) = 1$ ,  $z > u - \tilde{\sigma}/\xi$  e  $H(z)$  é dada por (A.8), se  $0 \leq z \leq u - \tilde{\sigma}/\xi$ , ou seja, a distribuição de excessos é limitada superiormente;

(b) quando  $\xi > 0$ ,  $H(z) = 0$ ,  $z < 0$ , e é dada por (A.8) para  $z \geq 0$ , isto é, a distribuição não é limitada superiormente.

(c) quando  $\xi = 0$ , a distribuição é também ilimitada, e devemos tomar o limite em (A.8) para  $\xi \rightarrow 0$ , obtendo-se que

$$H(z) = 1 - e^{-z/\tilde{\sigma}}, \quad z > 0,$$

que é uma distribuição exponencial com parâmetro  $1/\tilde{\sigma}$ . A partir de agora vamos denotar  $\tilde{\sigma}$  simplesmente por  $\sigma$ , que deverá ser estimado juntamente com  $\mu$  e  $\xi$ .

### 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. Veja Coles (2001), Tsay (2005), Mendes (2004), Embrechts et al. (1997), Reiss e Thomas (2001) para informação sobre esses tópicos.

---

# Análise de Dados de Alta Frequência

## 8.1 Introdução

Dados financeiros de alta frequência são aqueles obtidos em intervalos muito pequenos de tempo, usualmente em escala intradiária e irregularmente espaçados no tempo. Algumas vezes, dados diários são também considerados de alta frequência. Para fixar nossa nomenclatura, consideraremos *dados de alta frequência* (DAF) como aqueles observados no decorrer de um dia de transações numa bolsa de valores, num mercado de taxas de câmbio etc.

No Capítulo 1, citamos duas séries de retornos intradiários: do Ibovespa e da Telemar PN, ambas observadas a cada 15 minutos (veja o Exemplo 1.3). Vimos, também, naquele capítulo, a forma típica como esses dados são apresentados. Os dados brutos dessas duas séries, também chamados dados *tick-by-tick*, ocorreram em instantes aleatórios de tempo e depois foram “limpos” e amostrados a cada 15 minutos. Para detalhes sobre o tratamento prévio de dados intradiários, veja Dacorogna et al. (2001) e Zivot (2005). Esses dados foram obtidos da Bolsa de Valores de São Paulo (Bovespa).

Outras bases de dados do mercado acionário são a TAQ (*Trades and Quotes*) da NYSE (*New York Stock Exchange*), Amex e Nasdaq. Como exemplo de base de dados de opções, temos a *Berkeley Options Data Base* e, como exemplo de base de dados de taxas de câmbio, temos aquela organizada por Olsen Associates. Normalmente, essas bases de dados não são acessíveis gratuitamente.

No mercado brasileiro temos poucos dados de ativos negociados em alta frequência. Tipicamente, os tempos de negociação são medidos em segundos e podemos ter várias negociações no mesmo instante de tempo. As variações de preços de um ativo são dadas em múltiplos de um *tick*, que pode ser, por exemplo, um centavo de real ou um centavo de dólar.

Dados de alta frequência são importantes no estudo da microestrutura dos mer-

cados financeiros. Nesses mercados, existem grandes movimentos dos preços dos ativos em intervalos de tempo menores do que um dia. No entanto, DAF também apresentam desafios, como a filtragem dos dados, pois estes podem ser registrados com erros e têm que ser corrigidos antes que qualquer análise possa ser feita.

Para as séries do Ibovespa e Telemar o número de observações de cada uma era bastante grande: 37.961 para a primeira e 21.429 para a segunda, respectivamente. Para séries de taxas de câmbio esses valores são muito maiores, pois os mercados de taxas de câmbio funcionam praticamente durante 24 horas, dadas as diferenças geográficas entre eles (veja o Apêndice 8). Assim, tamanhos de séries acima de 100.000 observações são comuns.

DAF são em geral não sincronizados. Ativos diferentes têm frequências de negociações diferentes. Além disso, para um determinado ativo, a intensidade das negociações varia ao longo do dia, sendo maior na abertura e fechamento dos pregões e menor no horário de almoço. Isso faz com que os tempos entre as negociações (ou durações, veja a Seção 8.3) apresentem um ciclo diário, na forma de “U” invertido.

Para retornos diários de ações, negociações não sincronizadas podem introduzir (Tsay, 2005):

- (a) correlação não nula no lag 1 para retornos de uma ação;
- (b) correlações e correlações cruzadas não nulas no lag 1 para uma carteira de ações;
- (c) em algumas situações, correlação negativa para retornos de uma ação.

Com referência ao item (c), considere o log-retorno  $r_t$  de uma ação e suponha que  $r_t \sim \text{i.i.d.}(\mu, \sigma^2)$ . Seja  $\pi$  a probabilidade que a ação não seja negociada em qualquer instante  $t$ . Chamemos  $r_t^*$  o retorno observado no instante  $t$ . Se não houver negociação no instante  $t$ ,  $r_t^* = 0$  e se existir,  $r_t^* = r_t + r_{t-1} + \dots + r_{t-k_t}$ , onde  $k_t$  é o maior inteiro não negativo tal que não ocorram negociações nos períodos  $t-k_t, t-k_t+1, \dots, t-1$ . Então (veja Tsay, 2005, para uma prova), temos que  $\text{Cov}(r_t^*, r_{t-j}^*) = -\pi^j \mu^2$ ,  $j \geq 1$ . Ou seja, quando  $\mu \neq 0$ , negociações não sincronizadas provocam correlação negativa na série de retornos.

Essa correlação negativa pode ser causada pelo chamado *bid-ask spread*. Chamemos de  $p_{b,t}$  o *bid price* (preço de venda para o público) e  $p_{a,t}$  o *ask price* (preço de compra para o público), ambos no instante  $t$ . A diferença  $s_t = p_{a,t} - p_{b,t}$  é chamada *bid-ask spread* no instante  $t$ , e é, em geral, um valor pequeno (alguns *ticks*).

O preço de um ativo num mercado “sem fricção” usualmente é definido por  $p_t^* = (p_{a,t} + p_{b,t})/2$ . Roll (1984) sugere um modelo para o verdadeiro preço de mercado de um ativo, em função de  $p_t^*$ . Em particular, para esse modelo, a correlação de lag 1 para os retornos é negativa e função do spread  $s_t$ . Se não tivermos interesse nos efeitos relacionados a  $s_t$ , podemos trabalhar com preços médios logarítmicos, ou seja,  $x_t = [\log(p_{a,t} \cdot p_{b,t})]/2$ , de modo que os retornos de período  $k$  serão dados por  $r_t[k] = x_t - x_{t-k}$ .

---

**Outra característica marcante de DAF é a ocorrência de múltiplas transações no mesmo instante, mesmo com preços diferentes. Mas a grande maioria das transações não apresenta mudanças de preços.**

**Exemplo 8.1.** Consideremos os dados de negociações das ações da Telemar no período de 8 a 10 de setembro de 2004, perfazendo 6.734 negociações. Foram retiradas as negociações que ocorreram fora do horário de funcionamento da bolsa, entre 10 horas e 17 horas, resultando 6.588 transações (146 negociações ocorreram no chamado período *after market*, horário em que a bolsa está fechada). Vários autores consideram que essas observações têm um padrão diferente das negociações que ocorrem durante o período normal de funcionamento da bolsa. Veja Engle (2000) e Zhang et al. (2001). Nesse caso, aproximadamente 2,17% das negociações referem-se às operações feitas por via eletrônica. Veja também Stoll e Whaley (1990).

Há inúmeras negociações que ocorreram ao mesmo tempo, a saber, 2.026, representando aproximadamente 30% do total. Dessas, apenas 35% apresentaram alterações de preços. Ou seja, transações simultâneas parecem não trazer grande informação sobre a microestrutura do mercado relevante para um tratamento estatístico. Essas negociações simultâneas são, em geral, excluídas na modelagem estatística de tempos entre negociações.

Na Tabela 8.1, selecionamos 4.096 valores selecionados entre os 4.562 que não apresentaram transações simultâneas, classificando os movimentos de preços da ação. As alterações de preços foram classificadas como aumento (+), estabilidade (0) e diminuição (-). A tabela mostra os movimentos de preços entre duas negociações consecutivas, ou seja, da  $(t - 1)$ -ésima para a  $t$ -ésima.

Tabela 8.1: Movimentação de preços em negociações consecutivas para a Telemar

neg. $t$ neg. $t - 1$	+	0	-	Total
+	168(4,1%)	390(9,5%)	263(6,4%)	821(20,0%)
0	397(9,7%)	1.547(37,8%)	444(10,8%)	2.388(58,3%)
-	255(6,3%)	451(11,0%)	181(4,4%)	887(21,7%)
Total	820(20,1%)	2.388(58,3%)	888(21,6%)	4.096(100%)

Notamos que:

- (a) aumentos ou diminuições consecutivas de preços são pequenas, representando apenas 8,5% das negociações;
- (b) em 37,8% das negociações, o preço permanece estável, havendo grande tendência de não haver variações de preço;
- (c) a porcentagem de negócios em que os preços mudaram de sentido, isto é, de queda para subida e vice-versa, é semelhante, da ordem de 6,3%, totalizando aproximadamente 13% das transações e indicando um processo de reversão de preços;

(d) a distribuição marginal de preços é semelhante em duas negociações consecutivas.

Além dos preços de uma ação, os tempos entre negociações ou *durações* são importantes para o estudo da microestrutura dos mercados acionários. Chamemos de  $\tau_0, \tau_1, \dots, \tau_T$  os instantes de tempo de negociações de uma ação, sendo que  $\tau_0$  é o instante inicial e  $T$  é o número de negociações no período avaliado. Então,  $x_t = \tau_t - \tau_{t-1}$ ,  $t = 1, \dots, T$  são as durações. Na Seção 8.3, trataremos de um modelo importante para as durações, o chamado modelo ACD (*autoregressive condicional duration*).

Na Figura 8.1, ilustramos a série dos 4.096 preços da Telemar considerados acima, a série de retornos, o histograma dos retornos e a respectiva f.a.c. Notamos uma correlação negativa no lag 1 da f.a.c., como discutido anteriormente. Na Figura 8.2, consideraremos as durações, com o respectivo histograma, f.a.c. e box-plot. O histograma das durações mostra um decaimento exponencial a partir da origem e a f.a.c. mostra que a série de durações tem memória longa.

Outra variável de interesse é o número de negociações em dado período, por exemplo, por dia, ou em cada intervalo de cinco minutos. A Figura 8.3 contém os gráficos do número de negociações em intervalos de 5 minutos para os dados da Telemar, no período de 4 de agosto a 10 de setembro de 2004 e da f.a.c. estimada para os lags de 0 a 260. A figura também mostra as durações das negociações, em intervalos de cinco minutos e respectiva f.a.c. Ambas as funções mostram uma periodicidade evidente de 84, correspondente ao número de intervalos de cinco minutos em cada dia de negociação.

A Figura 8.4 traz os gráficos do número médio de negociações e do tempo médio entre as negociações em intervalos de cinco minutos, no mesmo período. Ao todo são 84 médias. O primeiro gráfico apresenta um maior número de negócios nos períodos das 10h15 às 12h00 e das 14h40 às 16h55. O segundo gráfico apresenta obviamente um padrão inverso.

A Tabela 8.2 apresenta algumas estatísticas das negociações da Telemar nesse período, excluídas as simultâneas. O fato de a mediana das durações ser bem menor do que a média e o valor do coeficiente de assimetria ser relativamente alto (4,51 segundos) indica assimetria positiva dos dados, com concentração nos menores valores. Há, ainda, um grande número de valores extremos na cauda direita da distribuição das durações.

## 8.2 Volatilidade Realizada

No Capítulo 5, tratamos do problema da modelagem da volatilidade de um ativo, considerando-a como uma variável latente (não observada). Utilizamos modelos da família ARCH e modelos de volatilidade estocástica. Outras possibilidades para estimar volatilidade são usar a volatilidade implícita em modelos de opções ou volatilidade histórica.

---

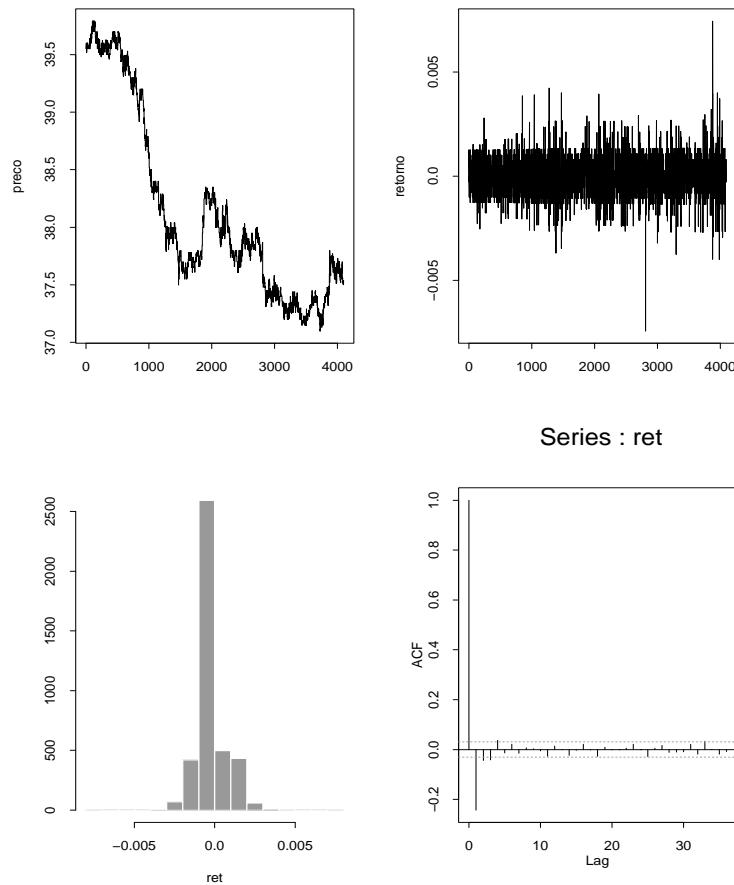


Figura 8.1: Séries de preços e retornos da Telemar, histograma e f.a.c. dos retornos.

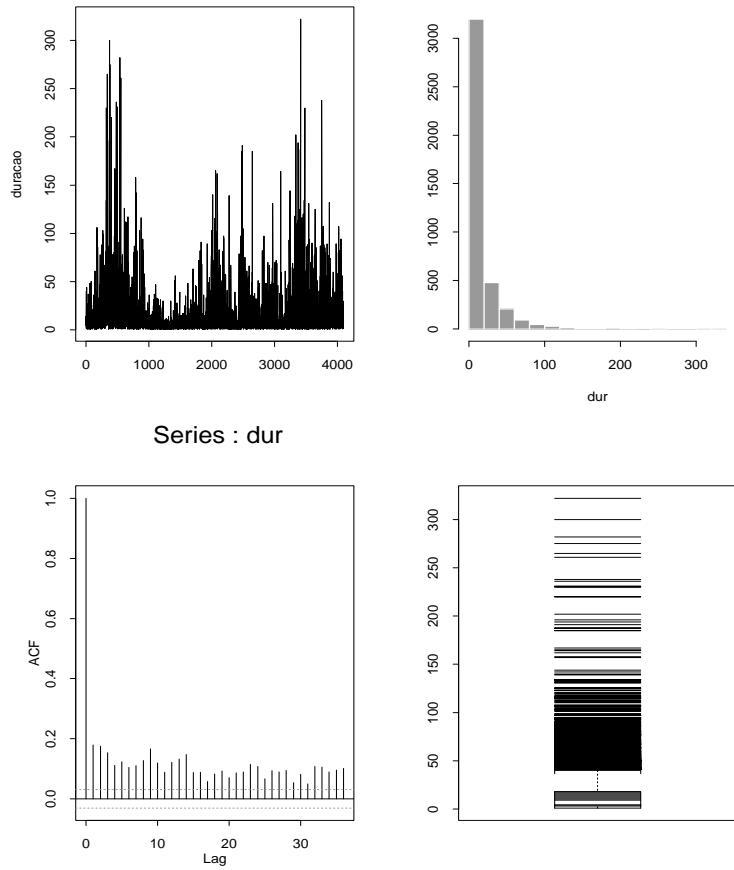


Figura 8.2: Série de durações da Telemar, histograma, f.a.c. e boxplot.

---

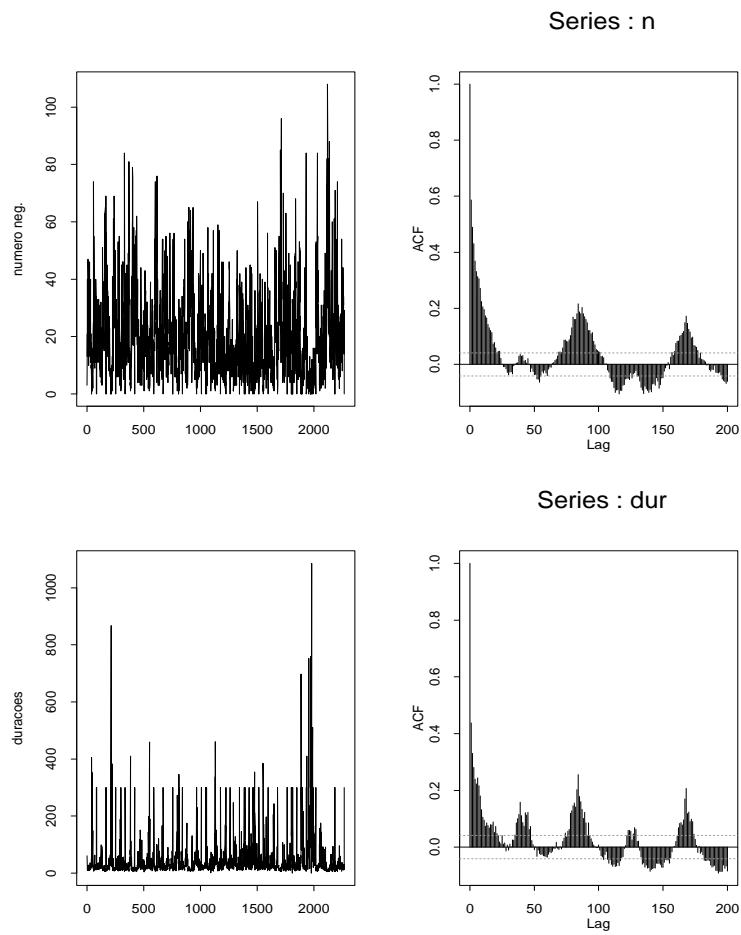


Figura 8.3: Número de negociações e durações da ação da Telemar em intervalos de 5 minutos e respectivas f.a.c.'s.

---

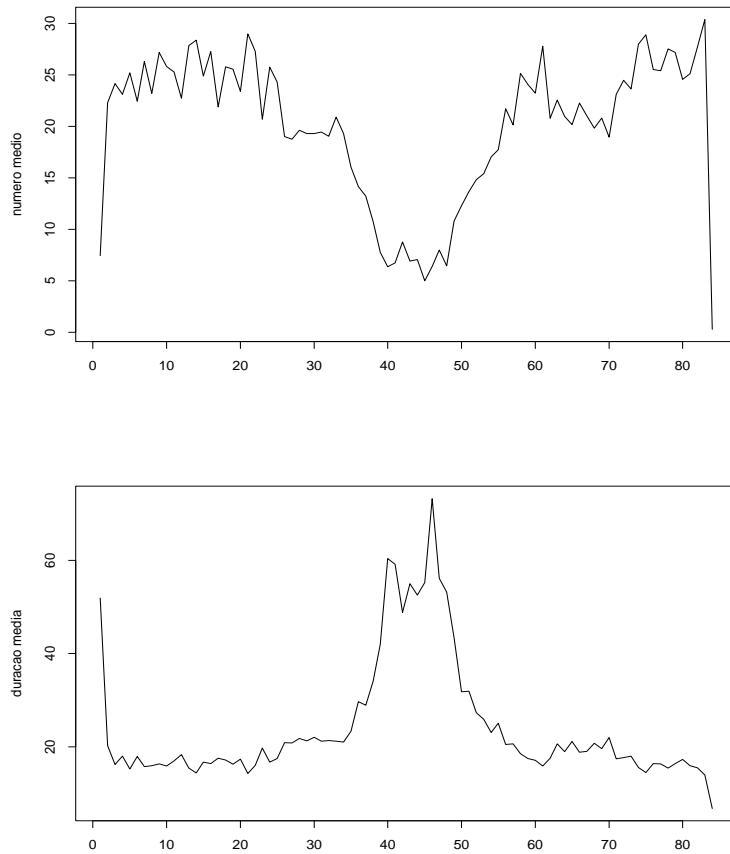


Figura 8.4: Número médio de negociações e durações médias da ação da Telemar em intervalos de cinco minutos.

Volatilidade e correlação são de importância fundamental em várias áreas de finanças, como apreçamento de ativos e gestão de riscos. Vimos que volatilidade dos ativos de uma carteira e as correlações entre os ativos são necessárias para o cálculo do VaR, valor em risco.

---

Tabela 8.2: Negociações da Telemar no período 8-10 de setembro de 2004

Estatística	Negociações (por dia)	Durações (em segundos)
Média	2.225	16,25
Desvio padrão	575	26,43
Mínimo	1.804	1
Máximo	2.895	322
Mediana		7
Assimetria		4,51

Todavia, a utilização de modelos para volatilidade é, em geral, restrita a dados diários ou de menor frequência e podem resultar em previsões não satisfatórias. Lembremos que os modelos GARCH, por exemplo, tendem a superestimar volatilidades futuras, incorporando resultados de eventos extremos que tenham ocorrido no passado. A escolha da distribuição dos erros nesses modelos também é um problema importante, a fim de modelar adequadamente os fatos estilizados presentes nos retornos, como agrupamentos de volatilidades e caudas pesadas. Os programas disponíveis aos usuários apresentam poucas possibilidades (normal, t, GED, veja o Apêndice ao Capítulo 5). Considerando-se uma série multivariada de retornos, a situação torna-se ainda mais complicada, pois modelos heteroscedásticos condicionais são complexos e difíceis de estimar, sendo seu uso limitado a uma carteira com poucos ativos.

Recentemente, a atenção está voltada para o uso de DAF intradiários para modelar e prever volatilidade e correlação por meio da chamada *volatilidade realizada*. A ideia é considerar a soma dos quadrados dos retornos obtidos em intervalos de alguns minutos durante um dia para estimar a volatilidade desse dia. Obteremos, então, uma série de volatildades diárias observadas, que pode ser modelada com modelos como ARIMA, ARFIMA etc.

Na impossibilidade de se obter dados intradiários, podemos considerar os preços de abertura, fechamento, máximo e mínimo de uma ação, observados em cada dia, para estimar a volatilidade desse dia, usando-se, por exemplo, a volatilidade de Garman-Klass, como mostrado no Apêndice 6.

Consideremos, primeiramente, o caso de um único ativo. Denote por  $P_t$  seu preço no instante  $t$  (um dia de negócios, por exemplo) e por  $p_t = \log P_t$ , sendo o logaritmo na base  $e$ . Então, com vimos, o log-retorno (que chamaremos simplesmente retorno) no dia  $t$  é dado por  $r_t = p_t - p_{t-1}$ .

Suponha, agora, que tenhamos preços intradiários desse ativo, observados em instantes irregularmente espaçados, mas depois amostrados a cada  $\Delta t$  unidades de tempo, por exemplo a cada 30 segundos ou a cada 15 minutos. Suponha, também, que fixemos a duração do pregão de uma bolsa (a Bovespa, por exemplo) em sete horas, das 10 horas às 17 horas (ou das 11 horas às 18 horas, no caso de horário de

verão). Se amostramos os preços a cada  $\Delta t = 15$  minutos, por exemplo, teremos  $M = 28$  preços ou  $M$  intervalos de 15 minutos por dia de negociações.

Denotemos por  $p_{d,m}$  o log-preço do ativo no período  $m$  ( $m = 1, 2, \dots, M$ ) do dia  $d$  ( $d = 1, 2, \dots, D$ ), ou seja, há  $D$  dias, de modo que o número total de observações é  $T = M \times D$ . O correspondente retorno intradiário no período  $m$  e dia  $d$  será denotado por

$$r_{d,m} = p_{d,m} - p_{d,m-1}, \quad m = 2, \dots, M, \quad d = 1, \dots, D, \quad (8.1)$$

com  $t = M(d-1) + m$ ,  $t = 1, \dots, T$ . Usaremos as notações  $r_t$  e  $r_{d,m}$  para denotar os retornos intradiários.

Com essa notação, o retorno diário  $r_d$  é dado por  $p_{d,M} - p_{d-1,M}$ , por exemplo, se  $M = 28$ ,  $r_t = p_{d,28} - p_{d-1,28}$ .

Além dos retornos definidos acima, há o chamado retorno *overnight*, que incorpora as informações relativas ao intervalo entre o fechamento do pregão do dia anterior e a abertura do pregão do dia atual, sendo definido por

$$r_{d,1} = p_{d,1} - p_{d-1,M}, \quad (8.2)$$

sendo  $p_{d,1}$  o log-preço do ativo na primeira cotação do dia  $d$  e  $p_{d-1,M}$  o log-preço do ativo na última cotação do dia  $d-1$ . Incluindo-se esse retorno no caso de  $\Delta t = 15$  minutos, teremos  $M = 29$  observações em cada dia.

**Exemplo 8.2.** Vamos considerar, agora, uma outra base de dados já mencionada no Exemplo 1.3, ou seja, dados intradiários do Ibovespa, obtidos de 1.381 dias úteis, em duas frequências de amostragem:

- (a)  $\Delta t = 15$  minutos, de 02/01/98 a 31/10/02, com 1.189 dias úteis, sendo que em cada dia há de 29 a 33 observações, totalizando 35.227 dados;
- (b)  $\Delta t = 30$  segundos, de 01/11/02 a 13/08/03, com 192 dias úteis, sendo que em cada dia há 865 a 872 observações, totalizando 166.236 dados.

Para a utilização desses dados originais, procedeu-se a uma limpeza deles. Os detalhes estão em Berti (2005). Como salientamos antes, a duração de cada pregão foi fixada em sete horas, que corresponde à duração normal do pregão de vivavoz, período em que é maior o volume de negociações. Um resumo dos ajustes é o seguinte:

- (a) foram excluídos os dados anteriores a 06/04/98, por apresentarem períodos irregulares e dados discrepantes;
  - (b) foram excluídos os dados de negociação eletrônica;
  - (c) foi excluída a primeira hora de negociação nos casos em que o pregão vivavoz possuía duração de oito horas;
-

- (d) foram excluídos dias com expediente reduzido (feriados, jogos da copa mundial de futebol etc.);
- (e) foram feitas correções em dados com sinais claros de erros de digitação, por meio de interpolações lineares.

Sobre o intervalo de amostragem,  $\Delta t$ , há vários estudos, dentre os quais destacamos Andersen e Bollerslev (1998), Giot e Laurent (2004) e Oomen (2001), que analisaram intervalos variando de cinco a vinte e cinco minutos.

Sob o ponto de vista teórico, para se obter consistência das medidas de volatilidade realizada é necessário que  $\Delta t \rightarrow 0$ , ou seja, um número grande de retornos para reduzir o erro estocástico de mensuração. Por outro lado, a microestrutura do mercado introduz um viés que cresce, quando a frequência amostral cresce. Um compromisso faz-se, então, necessário.

Bandi e Russel (2003) propõem um método para escolher  $\Delta t$  que minimiza o erro quadrático médio do erro de mensuração. Bandi e Russell (2006) estudam a influência de efeitos da microestrutura do mercado (*spread bid-ask*, efeitos de calendário etc.) no viés de medidas de volatilidade.

Andersen et al. (2001), analisando vários instrumentos financeiros, sugerem um intervalo de 30 minutos para taxas de câmbio mais líquidas, o que leva a considerar apenas 48 observações por dia.

Consideramos, então, em nosso exemplo, que um intervalo de amostragem de 15 minutos é razoável e como a maior parte dos dados está amostrada nessa frequência, no período de 01/11/02 a 13/08/03 (em que a frequência é a cada 30 segundos) também foram consideradas amostras a cada 15 minutos, resultando em cotações do Ibovespa de 06/04/98 a 13/08/03, ou seja, 1.309 dias, com 29 observações intradiárias cada um, totalizando 37.961 pontos.

Na Figura 8.5, temos o gráfico da série de índices do Ibovespa a cada 15 minutos, dos log-retornos, histograma dos retornos e f.a.c. Na Figura 8.6, temos os gráficos dos retornos diários e horários, com seus histogramas e f.a.c.'s, respectivamente.

Pelas figuras, vemos que há presença de correlação serial nas séries, mais evidentes para frequências mais altas. Já os quadrados dos retornos exibem autocorrelações maiores. Os valores da estatística de Ljung-Box mostram esses dois fatos, o que nos leva à rejeição da hipótese de ruído branco para retornos. Note a sazonalidade presente nas autocorrelações dos quadrados dos retornos intradiários a cada 15 minutos e horários, na Figura 8.7.

Na Tabela 8.3, apresentamos algumas estatísticas das séries de retornos, para um número menor de dias, 838, correspondendo a 24.302 observações da série do Ibovespa. Verifica-se que série de retornos diários possui um alto valor da curtose, que aumenta com a frequência de amostragem. A assimetria é positiva para retornos de um dia e negativa para frequências maiores. Os valores da estatística de Jarque-Bera mostra a não normalidade dos retornos em todas as frequências.

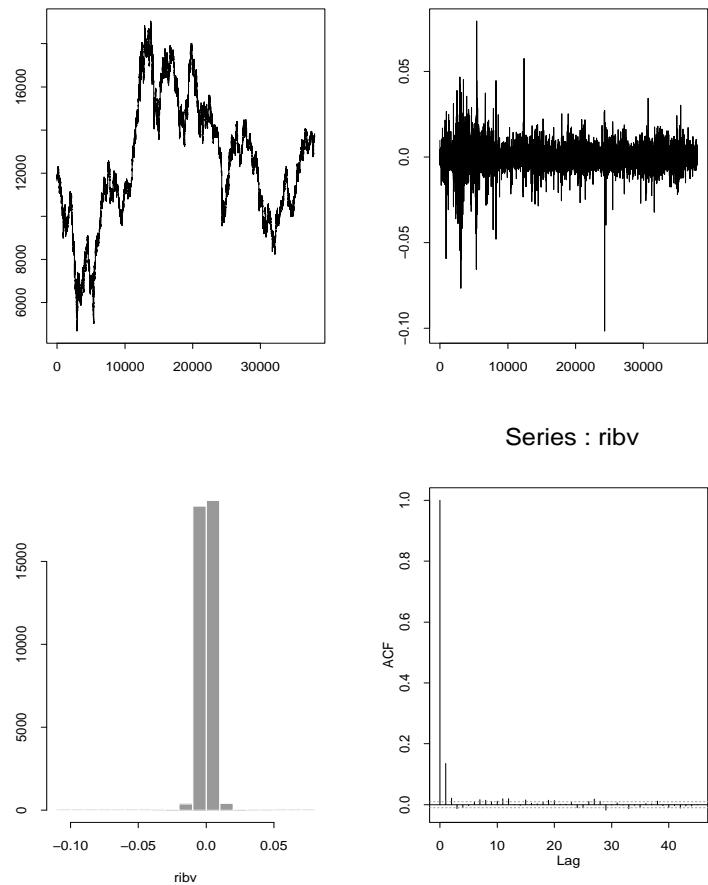


Figura 8.5: Série do Ibovespa a cada 15 min, retornos, histograma e f.a.c.

---

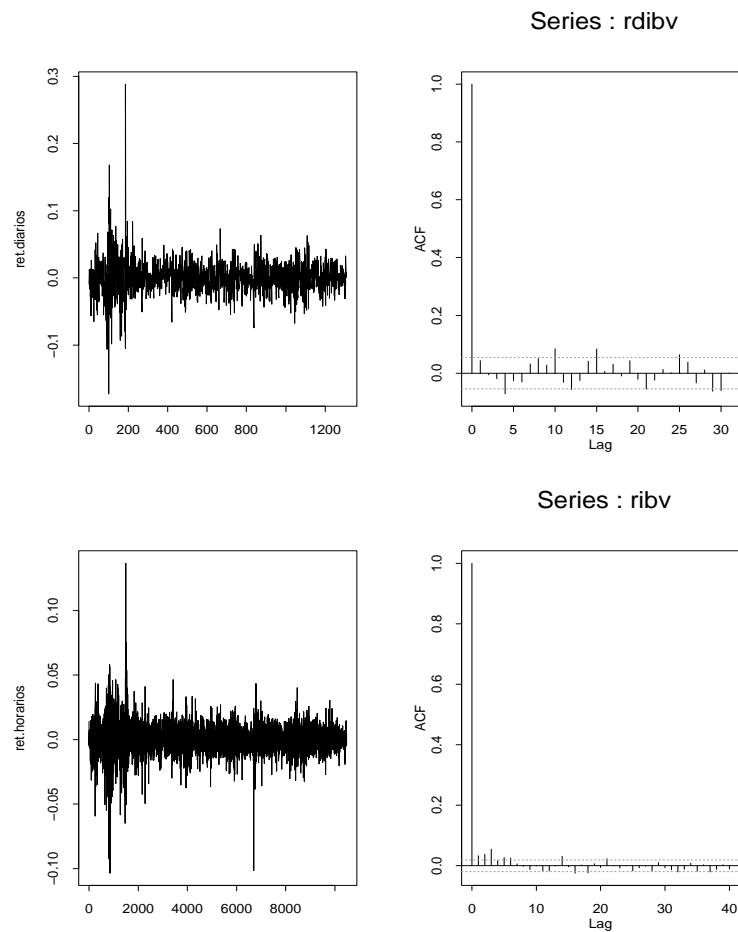


Figura 8.6: Retornos diários e horários do Ibovespa, e f.a.c.'s.

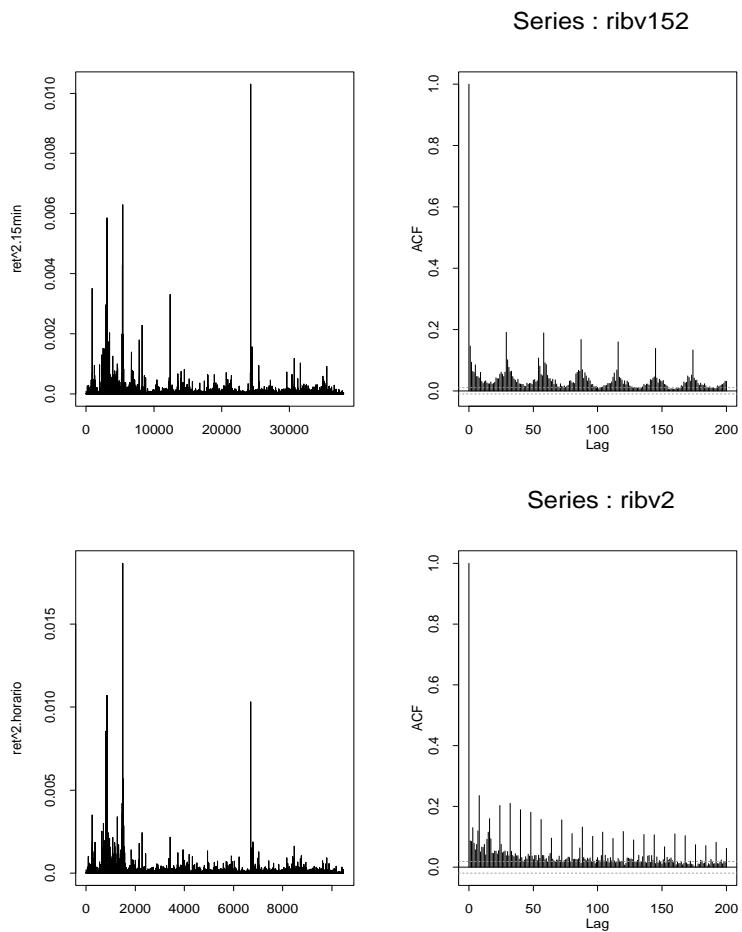


Figura 8.7: Quadrados dos retornos intradiários de 15 minutos e horários do Ibovespa e respectivas f.a.c.'s.

---

Tabela 8.3: Estatísticas para as séries de retornos do Ibovespa

Estatística	15 minutos	1 hora	1 dia	LVOLR	log(GK)
Mínimo	-10,15	-10,69	-17,27	-1,50	-1,89
Primeiro quartil	-0,16	-0,40	-1,47	0,49	0,33
Média	0,00	0,0	-0,01	1,11	0,99
Mediana	0,0	0,01	-0,04	0,99	0,93
Terceiro quartil	0,16	0,41	1,41	1,58	1,56
Máximo	7,93	13,30	28,82	4,93	5,67
Desvio padrão	0,43	0,97	2,84	0,91	1,03
Assimetria	-0,73	-0,24	1,16	0,80	0,56
Curtoza	42,90	17,64	16,33	1,21	1,30
Ljung-Box*	666,6	48,55	30,90	2.626	1.569
Ljung-Box(quadrado*)	1.389,0	972,1	106,0		
Jarque-Bera*	1.864,750	75.964	9385		

\* todas as estatísticas têm p-valor < 0,001

Vejamos como construir uma medida da volatilidade realizada. Primeiramente, definimos a *variância realizada* ( $VR_d$ ) para o ativo no dia  $d$  por meio de

$$VR_d = r_{d,1}^2 + \sum_{m=2}^M r_{d,m}^2, \quad d = 1, \dots, D, \quad (8.3)$$

com  $r_{d,m}$  e  $r_{d,1}$  definidos em (10.1) e (10.2), respectivamente.

A *volatilidade realizada* ( $VOLR_d$ ) para o ativo no dia  $d$  é, então, definida por

$$VOLR_d = \sqrt{VR_d}, \quad (8.4)$$

enquanto a *log-volatilidade realizada* no dia  $d$  é definida por

$$LVOLR_d = \log(VOLR_d). \quad (8.5)$$

No caso específico do Ibovespa intradiário, alguns dias foram excluídos no processo de filtragem dos dados e, nesse caso, temos que adaptar a equação (8.3), que toma a forma

$$VR_d = \begin{cases} \sum_{m=2}^{28} r_{d,m}^2, & \text{se o dia } d-1 \text{ foi excluído,} \\ r_{d,1}^2 + \sum_{m=2}^{28} r_{d,m}^2, & \text{caso contrário,} \end{cases} \quad (8.6)$$

para  $d = 1, \dots, 1.308$ .

Suponha, agora, que temos  $n$  ativos, com retornos  $r_{1,d;m}, \dots, r_{n,d;m}$  e considere o vetor  $\mathbf{r}_{d,m} = (r_{1,d;m}, \dots, r_{n,d;m})'$ ,  $m = 1, \dots, M$ ,  $d = 1, \dots, D$ . Então, para

cada ativo  $i$  teremos a VR dada por (8.3), que chamaremos  $VR_{i,d}$  e as respectivas volatilidade realizada ( $VOLR_{i,d}$ ) e log-volatilidade realizada ( $LVOLR_{i,d}$ ).

Para podermos calcular, por exemplo, o valor em risco de uma carteira com esses  $n$  ativos, será necessário calcular as correlações entre os retornos dos ativos. Definamos a matriz de covariâncias entre os retornos dos ativos para o dia  $d$  como

$$COVR_d = \sum_{m=1}^M \mathbf{r}_{d,m} \mathbf{r}'_{d,m}, \quad d = 1, \dots, D. \quad (8.7)$$

Se  $COVR_d = (c_{i,j,d}) = ([COVR_d]_{i,j})$ , então  $c_{i,j,d} = \sum_{m=1}^M r_{i,d;m} r_{j,d;m}$ ,  $i, j = 1, \dots, n$ ,  $d = 1, \dots, D$ . Esta matriz será positiva definida se  $n < M$ . A correlação entre os retornos dos ativos  $i$  e  $j$  é dada por

$$\rho_{i,j,d} = \frac{c_{i,j,d}}{\sqrt{VR_{i,d} \times VR_{j,d}}}. \quad (8.8)$$

**Exemplo 8.2** (continuação) Na Figura 8.8, temos o gráfico da volatilidade realizada,  $VOLR_t$ , que estima a volatilidade diária, obtida a partir dos retornos de 15 minutos do Ibovespa, bem como a sua f.a.c. Na mesma figura temos, também, o gráfico da  $LVOLR_d$  e sua f.a.c. Na Figura 8.9 temos a volatilidade de Garman-Klass, obtida a partir de preços de abertura, fechamento, mínimo e máximo de cada dia (conforme Apêndice 6), sua f.a.c., o logaritmo da volatilidade de GK e sua f.a.c., para efeito de comparação. Na Tabela 8.3, encontram-se as estatísticas do  $LVOLR_d$  e log(GK). Nas duas figuras, verificamos que as f.a.c.'s das volatilidades estimadas (VOLR, LVOLR, GK, log(GK)) decaem lentamente para zero, indicando que todas essas séries de volatilidades apresentam memória longa.

Analizando a f.a.c. dos quadrados dos retornos da Figura 8.7, notamos uma sazonalidade de 29 minutos para os dados de 15 minutos e de sete horas, para os dados horários. Esses padrões sazonais são importantes no procedimento de modelagem da volatilidade dos retornos intradiários. Aqui, é comum supor que estes sejam dados por  $r_t = \sigma_t s_t \varepsilon_t$ , ou seja, além da modelagem usual da volatilidade, introduzimos uma componente sazonal  $s_t$ .

Questões importantes são: (i) a retirada do padrão sazonal melhora a previsão dos modelos de volatilidade? (ii) qual é o melhor método para remover a componente sazonal? Para uma discussão dessas questões e dos diversos procedimentos de ajustamento sazonal veja Martens et al. (2002). Vamos nos limitar aqui em utilizar o seguinte procedimento, sugerido por Taylor e Xu (1997). Retomemos os retornos intradiários  $r_{d,m}$  e como estimativa da componente sazonal no período  $m$  considere

$$s_m = \left( \frac{1}{D} \sum_{d=1}^D r_{d,m}^2 \right)^{1/2}, \quad m = 1, \dots, M. \quad (8.9)$$

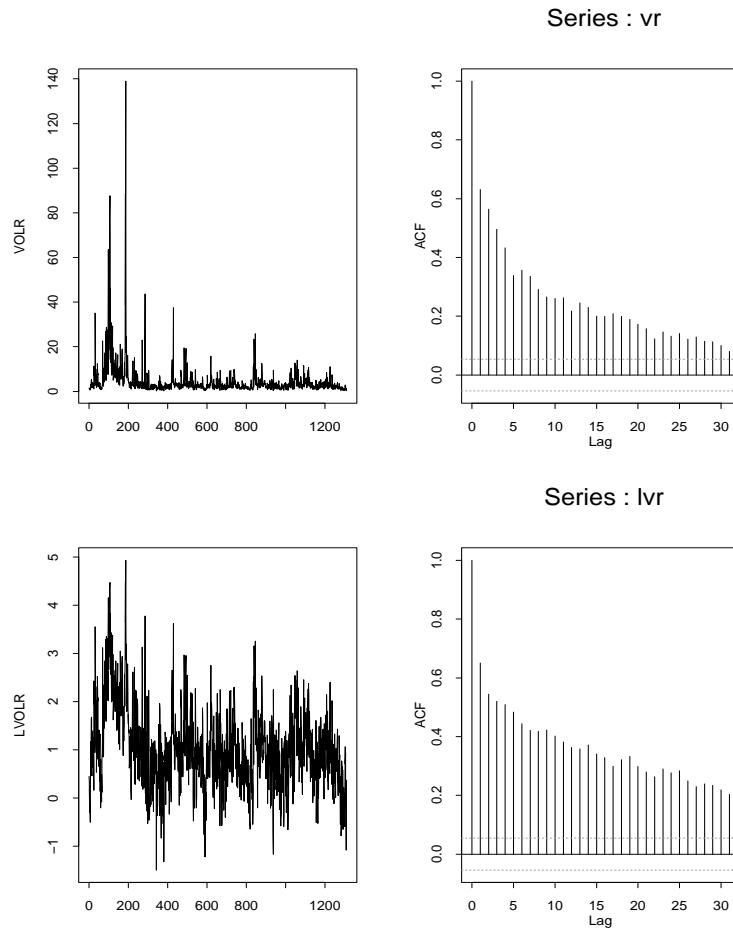


Figura 8.8: Volatilidade realizada e f.a.c.,  $\log(\text{volatilidade realizada})$  e f.a.c.  
para os retornos de 15 minutos do Ibovespa.

---

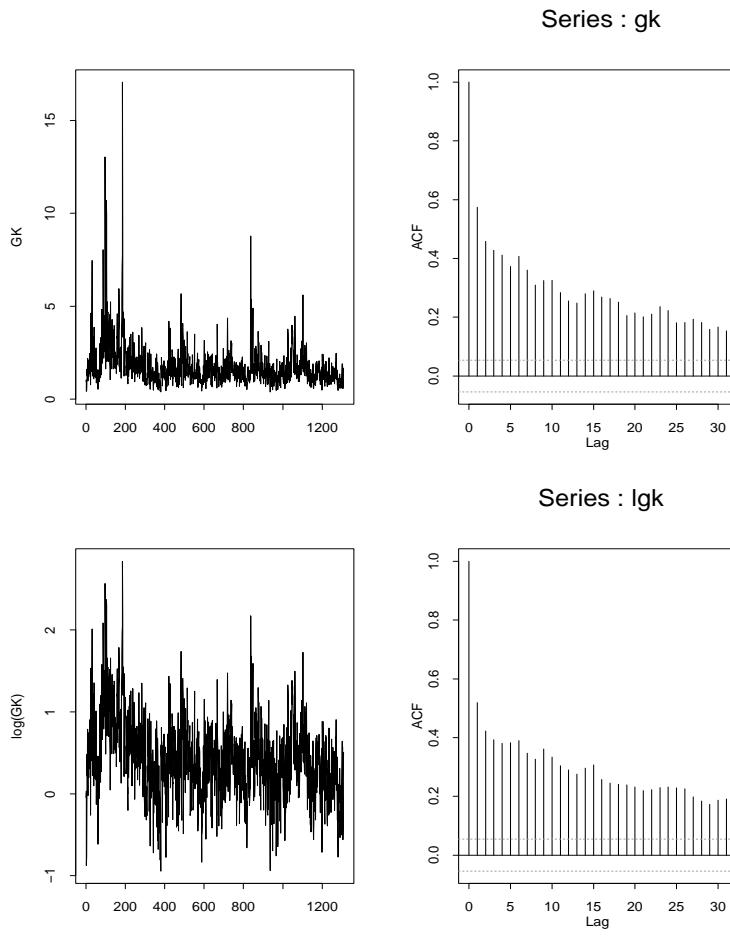


Figura 8.9: Volatilidade de GK e f.a.c,  $\log(\text{volatilidade de GK})$  e f.a.c para os retornos de 15 minutos do Ibovespa.

Segue-se que os retornos livres de componente sazonal são dados por  $r_t^* = r_t / s_t$ . Na Figura 8.10, temos a f.a.c dos quadrados dos retornos de 15 minutos e 1 hora dos dados do Ibovespa, após a retirada da componente sazonal, onde não se notam mais os padrões sazonais, os quais são mostrados na Figura 8.11.

### 8.3 Modelo de Duração Condicional

Usando conceitos similares aos dos modelos ARCH e GARCH desenvolvidos para a volatilidade, Engle e Russell (1998) e Bollerslev (1986) introduziram o modelo autorregressivo de duração condicional (ACD - *autoregressive conditional duration*) para descrever os tempos entre negociações de um particular ativo financeiro. Designemos por  $x_i$  a  $i$ -ésima duração e chamaremos  $\eta_i = E(x_i | \mathcal{F}_{i-1})$ , onde  $\mathcal{F}_{i-1}$  denota

a informação disponível até a  $(i - 1)$ -ésima negociação.

O modelo autorregressivo de duração condicional de ordens  $r$  e  $s$ , denotado por ACD( $r, s$ ), pode ser escrito na forma

$$x_i = \eta_i \varepsilon_i, \quad (8.10)$$

$$\eta_i = \omega + \sum_{j=1}^r \delta_j x_{i-j} + \sum_{j=1}^s \gamma_j \eta_{i-j}, \quad (8.11)$$

onde  $\varepsilon_i$  é uma sequência de v.a.'s i.i.d. não negativas, com média um. Podemos considerar para  $\varepsilon_i$  distribuições como a exponencial, Weibull e Gama generalizada. Como  $x_i$  é positivo, devemos ter  $\omega \geq 0$  e  $\delta_j > 0$ ,  $\gamma_j > 0$ .

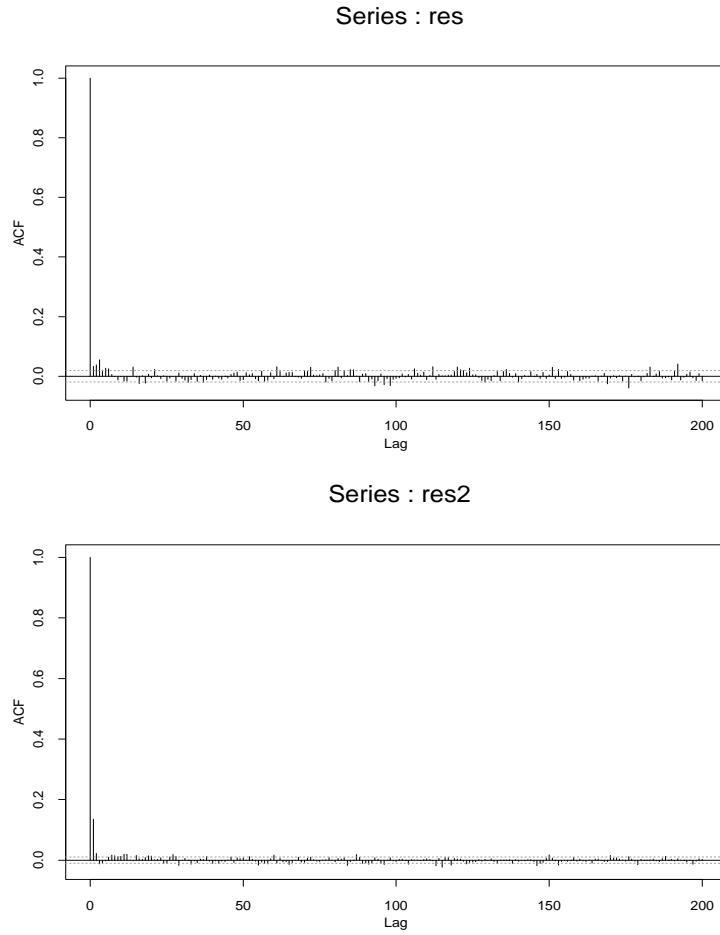


Figura 8.10: F.a.c.'s dos quadrados dos retornos horários e a cada 15 minutos do Ibovespa, sem a componente sazonal.

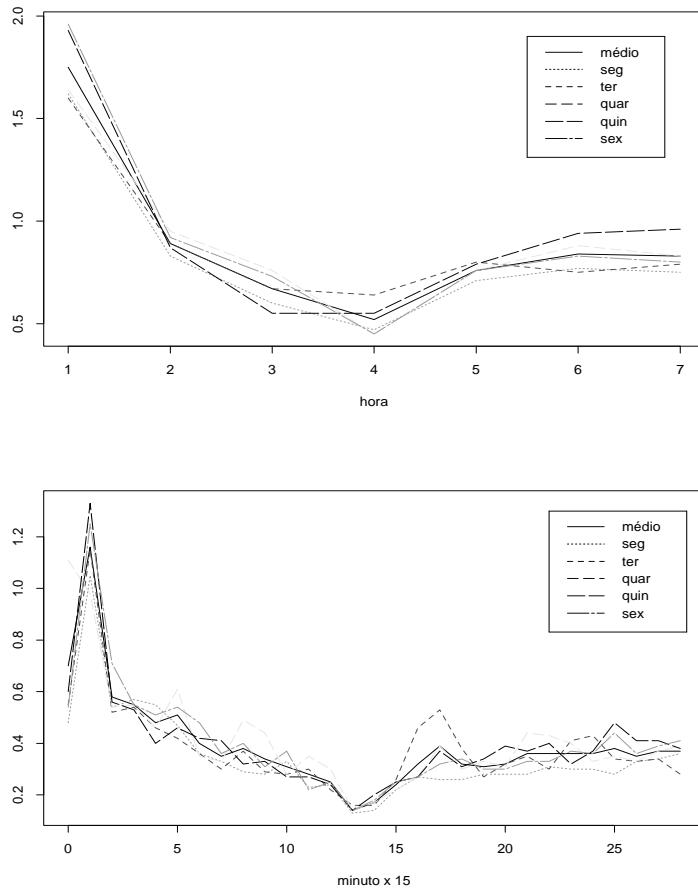


Figura 8.11: Padrões sazonais para os dados intradiários.

Como no caso de modelos GARCH, o processo  $\nu_i = x_i - \eta_i$  é uma sequência de diferenças martingales, com  $E(\nu_i | \mathcal{F}_{i-1}) = 0$ , não correlacionadas e com variância não constante. O modelo pode, então, ser escrito na forma

$$x_i = \omega + \sum_{j=1}^q (\delta_j + \gamma_j) x_{i-j} + \sum_{j=1}^s \gamma_j \nu_{i-j} + \nu_i, \quad (8.12)$$

onde  $q = \max(r, s)$ , ou seja, na forma de um modelo ARMA( $q, s$ ) com inovações não gaussianas e heteroscedásticas. Segue-se que

$$E(x_i) = \frac{\omega}{1 - \sum_{j=1}^q (\delta_j + \gamma_j)}, \quad (8.13)$$

com  $\sum_{j=1}^q (\delta_j + \gamma_j) < 1$ .

O modelo mais simples é o ACD(1, 1),

$$x_i = \eta_i \varepsilon_i, \quad (8.14)$$

$$\eta_i = \omega + \delta x_{i-1} + \gamma \eta_{i-1}, \quad (8.15)$$

supondo por exemplo  $\varepsilon_i \sim \text{i.i.d. } E(1)$  (exponencial de média um). Usando os mesmos argumentos usados para o modelo GARCH(1, 1) e supondo  $x_i$  estacionária, podemos obter (veja o Problema 3):

$$E(x_i) = \mu_x = \frac{\omega}{1 - \delta - \gamma}, \quad (8.16)$$

$$\text{Var}(x_i) = \mu_x \left( \frac{1 - \gamma^2 - 2\delta\gamma}{1 - 2\delta^2 - 2\delta\gamma - \gamma^2} \right). \quad (8.17)$$

De (8.17) devemos ter  $\gamma^2 + 2\delta\gamma + 2\delta^2 < 1$ , para que a variância incondicional seja constante no tempo.

Usando a forma ARMA(1, 1), teremos

$$x_i = \omega + (\delta + \gamma)x_{i-1} + \nu_i - \gamma\nu_{i-1},$$

da qual obtemos que a f.a.c de um modelo ACD(1, 1) é dada por

$$\rho_1 = \frac{\delta(1 - \gamma^2 - \delta\gamma)}{1 - \gamma^2 - 2\delta\gamma}, \quad (8.18)$$

$$\rho_j = (\delta + \gamma)\rho_{j-1}, \quad j \geq 2. \quad (8.19)$$

Veja o Problema 4.

Há várias alternativas sugeridas na literatura para o modelo ACD, dentre as quais destacamos: o modelo de duração estocástico (Bauwens e Vereda, 2000), a versão logarítmica (Bauwens e Giot, 2000), a versão não linear (Zhang et al., 2001), o modelo ACD assimétrico (Bauwens e Giot, 2003) e o modelo ACD aumentado (Fernandes e Gramming, 2005).

### Identificação

Assim como para um modelo GARCH, a especificação da ordem de um modelo ACD é uma tarefa complicada e, portanto, sugere-se que ajustem-se modelos de ordens não muito altas e utilize-se algum critério (como AIC, BIC ou log-verossimilhança) para a escolha do modelo.

### Estimação

Supondo que os erros  $\varepsilon_i$  sigam alguma distribuição, como uma das mencionadas aqui, podemos usar o método de máxima verossimilhança para obter os estimadores dos parâmetros do modelo ACD. Chamando de  $\boldsymbol{\theta}$  o vetor de parâmetros e  $\mathbf{x} = (x_1, \dots, x_T)'$  o vetor de durações observadas, obtemos a função de verossimilhança

$$L(\boldsymbol{\theta}; \mathbf{x}) = \left[ \prod_{i=2}^T f(x_i | \mathcal{F}_{i-1}, \boldsymbol{\theta}) \right] f(x_1 | \boldsymbol{\theta}). \quad (8.20)$$

A densidade marginal  $f(x_1 | \boldsymbol{\theta})$  tem importância decrescente, quando  $T$  cresce e pode ser omitida (Tsay, 2005), quando da maximização de (8.20) para obter os EMV.

**Exemplo 8.3.** Suponha o modelo ACD(1,1) com erros seguindo uma distribuição exponencial com média unitária. Então,

$$f(x_i | \mathcal{F}_{i-1}, \boldsymbol{\theta}) = \exp \left\{ -\frac{x_i}{\eta_i} \right\} \frac{1}{\eta_i},$$

e a log-verossimilhança fica

$$\begin{aligned} \ell(\boldsymbol{\theta}; \mathbf{x}) &= \sum_{i=2}^T [-\ln(\eta_i) - x_i/\eta_i] \\ &= \sum_{i=2}^T [-\ln(\omega + \delta x_{i-1} + \gamma \eta_{i-1}) - x_i / (\omega + \delta x_{i-1} + \gamma \eta_{i-1})], \end{aligned} \quad (8.21)$$

com  $\boldsymbol{\theta} = (\omega, \delta, \gamma)'$ .

### Diagnóstico

A verificação da adequação do modelo é feita nos moldes usuais, analisando a f.a.c. dos resíduos e dos quadrados dos resíduos do modelo ajustado e calculando as respectivas estatísticas de Box-Pierce-Ljung.

Como a série de durações apresenta uma periodicidade, como vimos na Seção 8.1, antes de ajustar o modelo, essa sazonalidade tem que ser removida, usando-se a mesma metodologia que foi vista para a volatilidade realizada.

**Exemplo 8.4.** Retomemos as durações em intervalos de cinco minutos da Telemar, no período de 4 de agosto a 10 de setembro de 2004. Removemos a sazonalidade usando um procedimento similar ao da Seção 10.2 e consideramos as 4.096 durações no período de 8 a 10 de setembro de 2004. A essa série livre de componente sazonal ajustamos um modelo ACD(1,1), com erros exponenciais de média 1. O programa

RATS foi utilizado para o procedimento de estimação. Veja Tsay (2005, p. 246) para um exemplo de tal programa. O modelo obtido é

$$x_i = \eta_i \varepsilon_i, \quad (8.22)$$

$$\eta_i = 0,01 + 0,07x_{i-1} + 0,918\eta_{i-1}, \quad (8.23)$$

onde os respectivos desvios padrões das estimativas são 0,002, 0,004 e 0,005. Na Figura 8.12, temos a f.a.c. dos resíduos e o plot  $Q \times Q$  (quantis empíricos dos resíduos *versus* quantis da  $\text{Exp}(1)$ ). Vemos que o modelo ajustado não consegue captar algumas durações extremas, que estão destacadas no plot  $Q \times Q$ . Uma possibilidade é considerar modelos ACD com coeficientes variando no tempo, a fim de levar em conta a não estacionariedade das durações e a sazonalidade presente. Veja Bruscato et al. (2010). Para fazer previsões com o modelo ajustado, é necessário recompor a série, por meio do padrão sazonal previamente estimado, ou seja,  $\hat{y}_i = \hat{x}_i s_5$ .

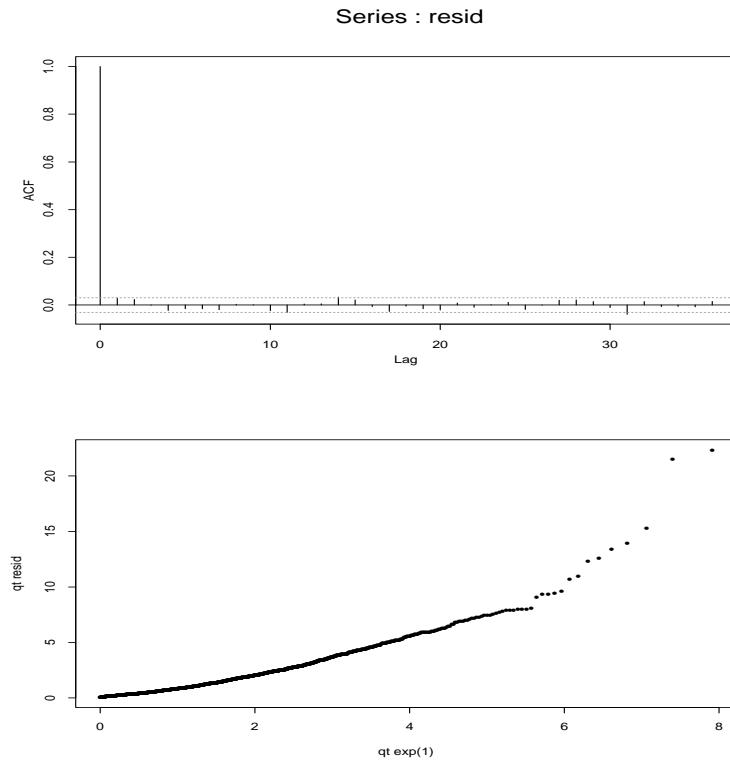


Figura 8.12: F.a.c dos resíduos do modelo (10.22)-(10.23) e plot  $Q \times Q$ .

## 8.4 Modelagem da Volatilidade

No Capítulo 5, vimos como modelar a volatilidade, supondo-a latente (não observada), usando modelos da família ARCH. Como a volatilidade realizada é efetivamente observada, podemos tentar modelá-la por meio de modelos da família ARIMA, por exemplo. Contudo, vimos que a volatilidade realizada apresenta características de processos de memória longa; logo, modelos da classe ARFIMA podem ser úteis.

Outra possibilidade é considerar modelos ARCH, GARCH e extensões para dados intradiários e usá-los para fornecer informação adicional para os dados diários. Veja Hol e Koopman (2002), por exemplo. A ideia é ajustar modelos GARCH, digamos, para frequências de uma hora ou 15 minutos e realizar previsões da volatilidade nessas frequências  $k$  passos à frente até cobrir o período de um dia ( $k = 7$  para dados horários e  $k = 29$  para dados a cada 15 minutos), agregando-se as previsões para obter a volatilidade diária. Observemos que, antes de ajustar um modelo ARMA-GARCH aos dados intradiários originais, é necessário retirar a sazonalidade, conforme visto na Seção 8.2.

**Exemplo 8.5.** Consideremos os dados intradiários do Ibovespa, com 1.309 dias de observações. Vamos considerar 838 dias para estimar modelos e 471 dias para previsões. No caso de dados horários, teremos 24.302 dados e no caso de dados a cada 15 minutos, teremos 5.886 dados para estimação. Para obter previsões da volatilidade diária, faremos previsões  $k$  passos à frente,  $k = 7$  ou  $k = 29$ , respectivamente.

A avaliação da acurácia preditora dos modelos de volatilidade é geralmente baseada em funções de perda, comparando os valores previstos do quadrado da volatilidade pelo modelo usado,  $\hat{h}_t^2$ , com os quadrados dos retornos,  $r_t^2$ . No entanto, Andersen e Bollerslev (1998) sugerem usar a volatilidade realizada, no lugar dos quadrados dos retornos. Se  $T$  indica o número total de dados e  $T_0$  o número de observações usadas para estimar o modelo, algumas medidas usadas são:

(a) Erro médio:

$$EM = \frac{1}{T - T_0} \sum_{t=T_0+1}^T (\hat{h}_t^2 - VOLR_t^2).$$

(b) Erro quadrático médio:

$$EQM = \frac{1}{T - T_0} \sum_{t=T_0+1}^T (\hat{h}_t^2 - VOLR_t^2)^2.$$

(c) Erro absoluto médio:

---

$$EAM = \frac{1}{T - T_0} \sum_{t=T_0+1}^T |\hat{h}_t^2 - VOLR_t^2|.$$

Berti (2005) ajustou várias combinações de modelos ARMA-GARCH (por GARCH, aqui entendemos GARCH, EGARCH, PGARCH etc.), sendo que os melhores modelos segundo medidas de perda, incluindo aquela acima, foram:

- (i) Para log-retornos a cada 15 minutos: modelo AR(2)-GARCH(1,2), com distribuição  $t$  para os erros.
- (ii) Para log-retornos horários: modelo AR(5)-GARCH(1,2), com distribuição  $t$  para os erros.
- (iii) Para log-retornos diários: modelo AR(4)-PGARCH(1,1), com distribuição  $t$  para os erros e com a variância realizada  $VR_{t-1}$  como covariável. Veja o Capítulo 5, Problema 13, para a definição do modelo *Power GARCH*.
- (iv) Para a volatilidade realizada: modelo ARFIMA(0,d,0), com  $\hat{d} = 0,438$ .

Veja Berti (2005) para detalhes e os Problemas 8, 9, 10 e 11. Duas observações sobre o ajuste desses modelos: (a) os modelos GARCH intradiários em geral superestimam as previsões da volatilidade; (b) os modelos ARFIMA, em geral, subestimam a volatilidade em períodos de alta volatilidade. Seria interessante também avaliar a eficácia dos modelos de volatilidade estocástica aplicados a dados intradiários.

## 8.5 Modelos R-GARCH

Vimos que modelos GARCH com inovações seguindo uma distribuição  $t$  de Student fornece uma alternativa que proporciona modelos com caudas mais pesadas, comparadas com modelos com inovações gaussianas. Contudo perdemos a propriedade de estabilidade, que é uma propriedade desejável, pois a classe das distribuições estáveis tem um domínio de atração que permite aproximações muito boas.

Vamos estudar nessa seção os modelos R-GARCH (*randomized GARCH*), introduzidos por Nowicka-Zagajek e Weron(2001).

O modelo R-GARCH(r,p,q), ( $r,p,q \in \mathbb{N}$ ) é definido pelas equações

$$r_t = \sqrt{h_t} \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \tag{8.24}$$

$$h_t = \sum_{i=1}^r \theta_i \eta_{t-i} + \sum_{j=1}^p \phi_j h_{t-j} + \sum_{k=1}^q \psi_k r_{t-k}^2, \tag{8.25}$$

sendo

$$r \geq 1, \theta_r > 0, \theta_i \geq 0, i = 1, \dots, r-1, \quad (8.26)$$

$$p \geq 0, \phi_p > 0, \phi_j \geq 0, j = 1, \dots, p-1, \quad (8.27)$$

$$q \geq 0, \psi_q > 0, \psi_k \geq 0, k = 1, \dots, q-1. \quad (8.28)$$

As inovações  $\epsilon_t$  são consideradas v.a's i.i.d. normais ( $\epsilon_t \sim \mathcal{N}(0, 1)$ ), as inovações  $\eta_t$  são v.a's i.i.d. positivas e  $\{\epsilon_t\}$  e  $\{\eta_t\}$  são independentes.

Os modelos R-GARCH têm as mesmas boas propriedades dos modelos GARCH. A equação (8.25) pode ser escrita na forma simbólica

$$h_t = \Theta(B)\eta_t + \Phi(B)h_t + \Psi(B)r_t^2$$

com

$$\Theta(B) = \theta_1 B + \theta_2 B^2 + \dots + \theta_r B^r,$$

$$\Phi(B) = \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p,$$

$$\Psi(B) = \psi_1 B + \psi_2 B^2 + \dots + \psi_q B^q.$$

A distribuição de  $r_t$ , conditional a  $\mathcal{F}_{t-1}$ , é gaussiana,

$$r_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, h_t).$$

Isso significa que a média condicional  $E(r_t | \mathcal{F}_{t-1}) = 0$  e a variância condicional  $\text{Var}(r_t | \mathcal{F}_{t-1}) = E(r_t^2 | \mathcal{F}_{t-1}) = h_t$ .

Supondo que  $\eta_t$  seja estável, os métodos usuais que supõem que momentos de segunda ordem existam não podem ser aplicados, mas pode-se obter resultados interessantes para distribuições incondicionais de modelos R-GARCH.

Consideremos o modelo R-GARCH(r,p,0), supondo que  $\eta_t$  tenha uma distribuição estável, assimétrica à direita, da forma

$$\eta_t \sim S_{\alpha/2} \left( 2 \left( \cos \frac{\pi \alpha}{4} \right)^{2/\alpha}, 1, 0 \right), \quad (8.29)$$

com  $0 < \alpha < 2$ . Se os polinômios  $\Theta(B)$  e  $1 - \Phi(B)$  não têm raízes comuns, então a variância condicional  $h_t$  para modelos R-GARCH(r,p,0), com inovações dadas por (8.29) pode ser representada na forma

$$h_t = \sum_{j=1}^{\infty} \delta_j \eta_{t-j}, \quad (8.30)$$

com  $\delta_j$ 's positivos, satisfazendo  $\delta_j < Q^j$  eventualmente<sup>1</sup>,  $Q > 1$ , se, e somente se, o polinômio  $1 - \Phi(B)$  não tenha raízes no disco fechado  $\{z : |z| \leq 1\}$ . A sequência

---

<sup>1</sup>“ $a_j < b_j$  eventualmente” significa que existe  $j_0$  tal que  $a_j < b_j$  para todo  $j > j_0$ .

$\{h_t\}$  é, então, estacionária, estável com índice de estabilidade  $\alpha/2$  e

$$h_t \sim S_{\alpha/2} \left( 2 \left( \sum_{j=1}^{\infty} \delta_j^{\alpha/2} \right)^{2/\alpha} \left( \cos \frac{\pi\alpha}{4} \right)^{2/\alpha}, 1, 0 \right). \quad (8.31)$$

Os  $\delta_j$ 's são os coeficientes na expansão de  $\frac{\Theta(B)}{1-\Phi(B)}$ ,  $|B| < 1$ .

É possível também determinar a distribuição incondicional de  $r_t$  no modelo R-GARCH(r,p,0). Se  $\eta_t$  for distribuída como em (8.29), o modelo R-GARCH(r,p,0) é estacionário, simétrico,  $\alpha$ -estável,  $0 < \alpha < 2$  e

$$r_t \sim S_\alpha \left( \left( \sum_{j=1}^{\infty} \delta_j^{\alpha/2} \right)^{1/\alpha}, 0, 0 \right). \quad (8.32)$$

Na Figura 8.13 temos um modelo R-GARCH (1,1,0) simulado, com parâmetros  $\alpha = 1,7$ ,  $\theta_1 = 10^{-7}$ ,  $\phi_1 = 0,8$ . A volatilidade correspondente está na Figura 8.14. Vemos que a série da volatilidade segue de perto os valores atípicos presentes na série de retornos.

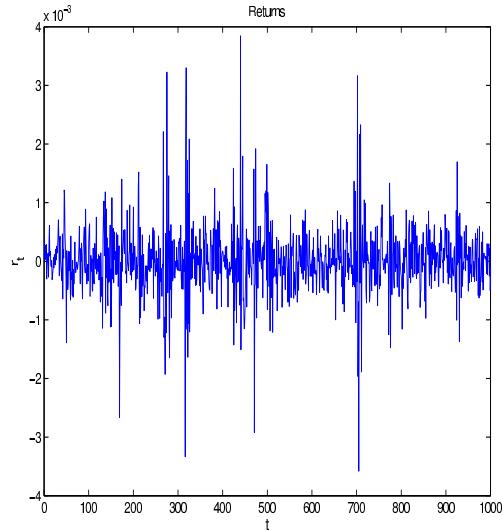


Figura 8.13: Modelo R-GARCH(1,1,0) simulado.

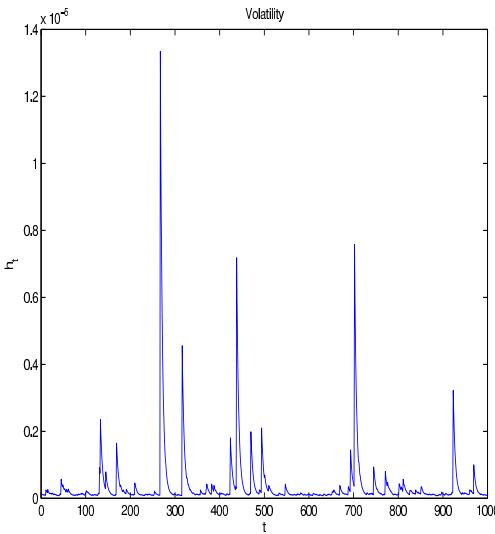


Figura 8.14: Volatilidade do modelo simulado R-GARCH(1,1,0).

Dada a impossibilidade de estimar os parâmetros do modelo pelos métodos tradicionais, iremos usar o procedimento chamado *inferência indireta*. Por simplicidade, iremos considerar o modelo R-GARCH(1,1,0), ou seja,

$$r_t = \sqrt{h_t} \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (8.33)$$

$$h_t = \theta_1 \eta_{t-1} + \phi_1 h_{t-1}, \quad (8.34)$$

onde as inovações  $\epsilon_t$  são i.i.d., normais padrão, as inovações  $\eta_t$  são distribuídas como em (8.29) e  $\{\epsilon_t\}$ ,  $\{\eta_t\}$  independentes.

A inferência indireta (veja Gouriéroux et al., 1993) é indicada para os casos em que a estimação do modelo de interesse é muito difícil de ser feita diretamente, em particular quando a estimação por máxima verossimilhança é muito difícil ou impossível, mas a simulação dos dados é simples. Nossa referência básica aqui é Lobardi e Calzolari (2008).

A ideia é a seguinte: suponha que tenhamos  $T$  observações  $\mathbf{y}$  e um modelo de interesse (MI), cuja função de verossimilhança  $\mathcal{L}^*(\mathbf{y}; \theta)$  seja difícil de maximizar. O modelo pode depender, também, de uma matriz de variáveis explicativas  $X$ . O estimador de máxima verossimilhança de  $\theta \in \Theta$  é dado por

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \sum_{t=1}^T \ln \mathcal{L}^*(\theta; y_t), \quad (8.35)$$

que não pode ser obtido. Como alternativa ao MI, consideremos um modelo auxiliar (MA), dependendo de um vetor de parâmetros  $\zeta \in \Upsilon$ , que seja fácil de manipular.

Suponha que decidimos usá-lo, no lugar do MI. Como o modelo não é especificado corretamente, o estimador

$$\hat{\zeta} = \arg \max_{\zeta \in \Upsilon} \sum_{t=1}^T \ln \tilde{\mathcal{L}}(\zeta; y_t) \quad (8.36)$$

não é, necessariamente, consistente: a ideia é explorar simulações realizadas sob o MI para corrigir o viés.

Primeiramente, considere  $\hat{\zeta}$  como o EMV de  $\zeta$  sob o MA. A seguir, simulamos um conjunto de  $S$  vetores de tamanho  $T$  do MI, com base no vetor de parâmetros arbitrário  $\hat{\theta}^{(0)}$ , que podemos pensar como um valor inicial. Denotemos esses vetores por  $\mathbf{y}^s(\hat{\theta}^{(0)})$ ,  $s = 1, \dots, S$ . Os valores simulados são então usados no MA, resultando

$$\hat{\zeta}_S(\hat{\theta}^{(0)}) = \arg \max_{\zeta \in \Upsilon} \sum_{s=1}^S \sum_{t=1}^T \ln \tilde{\mathcal{L}}[\zeta; y_t^s(\hat{\theta}^{(0)})]. \quad (8.37)$$

A ideia é atualizar, numericamente, o valor inicial  $\hat{\theta}^{(0)}$  a fim de minimizar a distância

$$[\hat{\zeta} - \hat{\zeta}_S]' \Omega [\hat{\zeta} - \hat{\zeta}_S],$$

na qual  $\Omega$  é uma matriz simétrica não negativa. Para dada estimativa  $\hat{\theta}^{(p)}$ , o procedimento resulta  $\hat{\theta}^{(p+1)}$ ; isso é repetido até que a série de estimativas  $\hat{\theta}^{(p)}$  converja, obtendo-se  $\hat{\theta} = \lim_{p \rightarrow \infty} \hat{\theta}^{(p)}$ .

A identificação de um MA apropriado é a questão básica. Deve-se notar que a dimensão do vetor de parâmetros do MA deve ser maior que ou igual a dimensão de  $\theta$ , para que a solução seja única. Quando os vetores de parâmetros têm a mesma dimensão e  $T$  é suficientemente grande, o estimador não depende da matriz  $\Omega$ .

Especificado um MA apropriado, a inferência indireta pode ser implementada com base na simulação da v.a  $\alpha$ -stable, dada no Capítulo 1, fórmula (1.34). Consideramos como MA um modelo que seja próximo ao modelo R-GARCH (1,1,0). Um candidato seria um modelo GARCH(1,1) com inovações  $t$  de Student:

$$\begin{aligned} r_t &= \sqrt{h_t} \epsilon_t, \quad t \in \mathbb{Z}, \\ h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned}$$

com  $\alpha_0 > 0$ ,  $(\alpha_1 + \beta_1) < 1$  e  $\epsilon_t \sim t(\nu)$ , sendo  $t(\nu)$  uma v.a  $t$  de Student com  $\nu$  graus de liberdade.

Denotando os parâmetros do MA por  $\zeta = (\alpha_0, \alpha_1, \beta_1)$  e os parâmetros do MI por  $\theta = (\alpha, \theta_1, \phi_1)$ , vemos que  $\dim \zeta = \dim \theta$  e, portanto, podemos escolher  $\Omega$  como a matriz identidade.

**Exemplo 8.6.** Vamos retomar os dados intradiários do Ibovespa, do Exemplo 1.3, amostrados a cada 15 minutos. Um total de  $T = 37.960$  retornos de abril a junho de 1998 são mostrados na Figura 8.15.

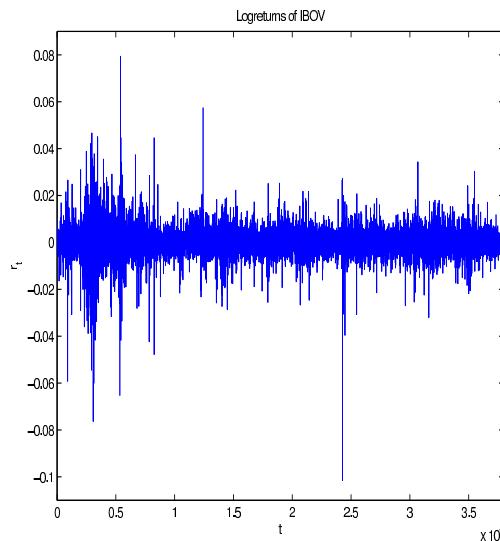


Figura 8.15: Retornos intradiários do Ibovespa.

O pacote MatLab foi usado para todos os cálculos dessa seção. Os códigos utilizados podem ser obtidos do autor. Veja Sampaio e Morettin (2015).

Baseados em 1.000 réplicas independentes e  $S = 4$  ajustamos um modelo R-GARCH(1,1,0) a esses retornos, com  $\nu = 3$  no MA. A Tabela 8.4 mostra as médias e desvios padrões Monte Carlo das estimativas dos parâmetros  $\hat{\alpha}$ ,  $\hat{\theta}_1$  e  $\hat{\phi}_1$ .

Observamos o pequeno valor de  $\hat{\theta}_1$ , e a estimativa de  $\alpha$ ,  $\hat{\alpha} = 1,34$ . Para efeito de comparação, ajustamos um modelo GARCH(0,1) com inovações  $t$  de Student model (note que o modelo R-GARCH(1,1,0) é um tipo de modelo GARCH(0,1) com o termo constante agora aleatório) e compararmos os erros quadráticos médios obtidos para os dois modelos. Veja a Tabela 8.5. Observamos que o EQM do modelo R-GARCH (1,1,0) é muito menor do que o EQM do modelo GARCH(0,1). A volatilidade estimada está na Figura 8.16.

Todavia, a Figura 8.17 mostra que a suposição de normalidade não é adequada, como era de se esperar.

Tabela 8.4: Médias e desvios padrões das estimativas via Monte Carlo para o Exemplo 8.5

	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\phi}_1$
Média	1,3376	$3,0078 \times 10^{-7}$	0,7112
Desvio padrão	0,0115	$4,9029 \times 10^{-9}$	0,0041

Tabela 8.5: EQM's do ajuste de modelos GARCH(0,1) e R-GARCH(1,1,0) aos retornos do Ibovespa

EQM para GARCH(0,1)	EQM para R-GARCH(1,1,0)
$2,4807 \times 10^{-5}$	$7,9848 \times 10^{-7}$

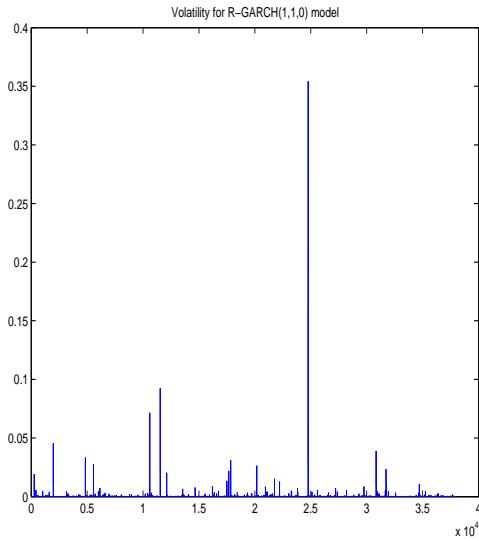


Figura 8.16: Volatilidade estimada para o modelo R-GARCH(1,1,0).

A seguir, exploramos o uso de inovações  $t$  de Student na equação dos retornos de um modelo R-GARCH.

Resultados teóricos similares àqueles para o R-GARCH com erros gaussianos são difíceis de obter, dado que a função característica para a distribuição  $t$  não está disponível. Definimos um modelo R-GARCH com inovações  $t$ , de modo análogo, substituindo a suposição para os  $\epsilon_t$ 's.

Como antes, vamos nos concentrar no caso R-GARCH (1,1,0) com  $\eta_t$ 's distribuídas como em (8.29). Especificamente, o modelo fica

$$\begin{aligned} r_t &= \sqrt{h_t} \epsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad \epsilon_t \sim \text{i.i.d. } t(\nu), \\ h_t &= \theta_1 \eta_{t-1} + \phi_1 h_{t-1}. \end{aligned}$$

O MA será o mesmo que antes:

$$\begin{aligned} r_t &= \sqrt{h_t} \epsilon_t, \quad t \in \mathbb{Z}, \\ h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned}$$

com  $\alpha_0 > 0$ ,  $(\alpha_1 + \beta_1) < 1$  e  $\epsilon_t \sim t(\nu)$ .

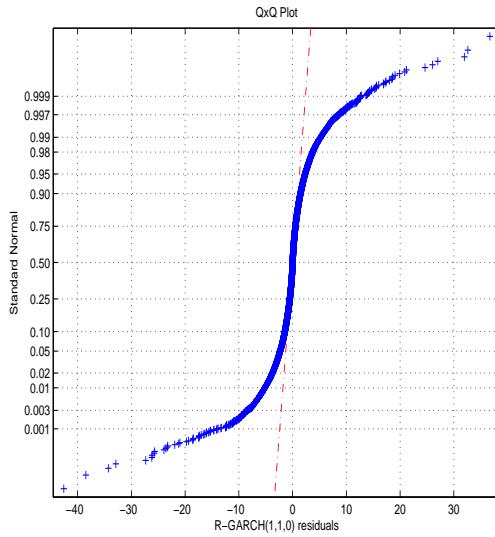


Figura 8.17: QxQ plot dos resíduos do modelo R-GARCH(1,1,0).

Continuamos a denotar o vetor de parâmetros para o MA por  $\zeta$  e o vetor para o MI por  $\theta$  e escolhemos  $\Omega$  como a matriz identidade.

A aplicação desse modelo aos dados dos retornos intradiários do Ibovespa, com 1.000 réplicas,  $S = 4$  e  $\nu = 3$ , resulta nos valores da Tabela 8.6.

Tabela 8.6: Médias e desvios padrões Monte Carlo para as estimativas usando a distribuição  $t(3)$

	$\hat{\alpha}$	$\hat{\theta}_1$	$\hat{\phi}_1$
<b>Média</b>	1,4743	$1,4330 \times 10^{-7}$	0,6800
<b>Desvio padrão</b>	0,0062	$3,4267 \times 10^{-9}$	0,0040

Na Tabela 8.7 temos a comparação entre os três modelos ajustados. Podemos ver que o EQM do modelo com distribuição  $t$  é menor que os demais. A volatilidade estimada com esse modelo está na Figura 8.18.

Tabela 8.7: EQM's para os ajustes dos modelos GARCH (0,1), R-GARCH (1,1,0) e R-GARCH(1,1,0) com distribuição  $t(3)$

<b>EQM para GARCH(0,1)</b>	$2,4807 \times 10^{-5}$
<b>EQM para R-GARCH(1,1,0)</b>	$7,9848 \times 10^{-6}$
<b>EQM para R-GARCH t</b>	$1,7408 \times 10^{-7}$

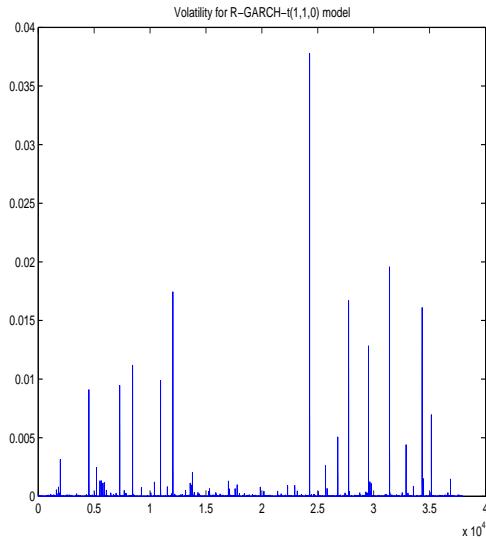


Figura 8.18: Volatilidade estimada para o modelo R-GARCH (1,1,0) com distribuição  $t$ .

O gráfico  $Q \times Q$  da Figura 8.19 mostra que o modelo com distribuição  $t$  é mais adequado.

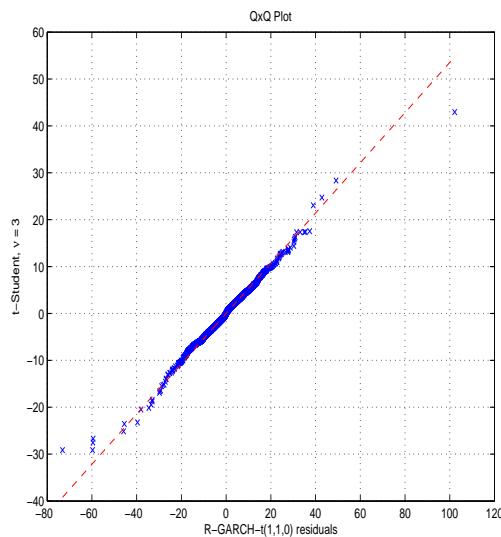


Figura 8.19: Gráfico  $Q \times Q$  para os resíduos do modelo R-GARCH com distribuição  $t$ .

É possível considerar modelos R-GARCH com inovações  $\alpha$ -estáveis, tanto na equação dos retornos, como na equação da volatilidade. Veja Sampaio e Morettin

(2016). Para outros modelos usando distribuições estáveis, veja Kim et al. (2010), Kim et al. (2011), Menn e Rachev (2009).

## 8.6 Comentários Adicionais

[1] Andersen et al. (2000, 2001, 2003) (designados brevemente por ABDL) e Barndorff-Nielsen e Shephard (2002, 2004) (designados brevemente por BS) desenvolveram os aspectos teóricos relativos à variância e volatilidade realizada. Seja  $p(t)$  o log-preço de um ativo,  $r(d, d - 1)$  o respectivo retorno diário e  $r(d) = p(d) - p(0)$  o retorno acumulado até  $d$ . A *variação quadrática (diária)* ( $VQ_d$ ) é definida por

$$VQ_d = p \lim \sum_{j=0}^{n-1} [p(s_{j+1}) - p(s_j)]^2,$$

onde  $0 = s_0 < s_1 < \dots < s_n = d$  e o limite é para  $\max_{1 \leq j \leq n} |s_j - s_{j-1}| \rightarrow 0$ ,  $n \rightarrow \infty$ . Então, BS provam que

$$VR_d \xrightarrow{P} VQ_d - VQ_{d-1}, \quad n \rightarrow \infty.$$

Ou seja, a variância realizada diária converge, em probabilidade, para incrementos diários da  $VQ_d$ . ABDL provam que

$$\text{Var}(r(d, d - 1) | \mathcal{F}_{d-1}) = E[VQ_d | \mathcal{F}_{d-1}],$$

ou seja, a variância condicional dos retornos diários é igual à esperança condicional do processo de variação quadrática diária.

Para a classe de processos de Ito contínuos, caracterizados pela equação diferencial estocástica

$$dp(t) = \mu(t)dt + \sigma(t)dW(t),$$

onde  $\sigma(t)$  é a volatilidade e  $W(t)$  é o MBP, a volatilidade integrada (latente) para o dia  $t$  é definida por

$$\sigma_t^{(d)} = \left( \int_{t-1d}^t \sigma^2(\omega) d\omega \right)^{1/2}.$$

Então, outra maneira de escrever é  $VR_d \xrightarrow{P} \sigma_t^{(d)}$ , para  $M \rightarrow \infty$ , sendo  $\Delta = 1d/M$ .

Ainda, para essa classe de modelos, resultados mais fortes podem ser obtidos, em particular,  $r(d, d - 1)/VOLR_d$  tem uma distribuição aproximadamente normal. Também é possível provar que a VOLR e LVOLR têm distribuições assintoticamente normais. Para detalhes, veja os autores acima mencionados.

[2] Podemos considerar outras classes de modelos para DAF. Müller et al. (1997) introduziram os modelos HARCH (*heterogeneous ARCH*). Um modelo HARCH(n) para retornos é dado por

$$\begin{aligned} r_t &= \sigma_t \varepsilon_t, \\ \sigma_t^2 &= c_0 + \sum_{j=1}^n c_j \left( \sum_{i=1}^j r_{t-i} \right)^2, \end{aligned}$$

com  $c_0 > 0, c_n > 0, c_j \geq 0$ , para  $j = 1, \dots, n - 1$ , e os  $\varepsilon_t$  são v.a.'s i.i.d. com média zero e variância um. Esse modelo leva em conta a heterogeneidade de informação do mercado financeiro, permitindo agregar retornos. Por exemplo, no modelo HARCH(2) a volatilidade toma a forma

$$\sigma_t^2 = c_0 + c_1 r_{t-1}^2 + c_2 (r_{t-1} + r_{t-2})^2.$$

Ruilova (2007) generaliza esse modelo para a classe GHARCH(m,p), que permite agregações de diferentes tamanhos, sendo  $m$  o número de componentes do mercado.

[3] Inspirado pela Hipótese do Mercado Heterogêneo, que deu origem aos modelos HARCH, e pela propagação assimétrica da volatilidade, Corsi (2009) propõe um modelo aditivo para componentes de volatilidade realizada diferentes, cada uma das quais gerada por um tipo diferente de participante do mercado. Tal modelo é chamado de Autorregressivo Heterogêneo (HAR, de *heterogeneous autoregressive*). Um modelo com três componentes de volatilidade realizada, diária, semanal e mensal, é dado por

$$VR_{t+1d}^{(d)} = c + \beta^{(d)} VR_t^{(d)} + \beta^{(w)} VR_t^{(w)} + \beta^{(m)} VR_t^{(m)} + \omega_{t+1d},$$

sendo que  $d$  representa dia,  $w$  semana e  $m$  mês. Ou seja, a volatilidade realizada num dia depende da volatilidade realizada do dia anterior, da volatilidade realizada da semana anterior e da volatilidade realizada do mês anterior. Teremos um modelo HAR(3) para a volatilidade realizada.

## 8.7 Problemas

1. Considere os dados intradiários da Telemar do Exemplo 8.1 (arquivo id-tibm.dat). Obtenha a volatilidade realizada diária, faça seu gráfico e obtenha sua f.a.c. Obtenha a f.a.c. dos quadrados dos retornos e verifique se existe sazonalidade. Caso haja, remova o padrão sazonal.
2. Mesmo problema para os dados intradiários da IBM (arquivo id-tibm.dat).

3. Prove as relações (8.16) e (8.17).
4. Prove as relações (8.18) e (8.19).
5. Obtenha (8.21) usando a distribuição de Weibull padrão para os erros  $\varepsilon_t$ :

$$f(\varepsilon, \alpha) = \alpha \cdot \varepsilon^{\alpha-1} \cdot \exp\{-\varepsilon^\alpha\}, \quad \alpha > 0.$$

6. Ajuste um modelo ACD para as durações da IBM (arquivo id-tlibm.dat), com erros exponenciais de média 1.
7. Mesmo problema, com distribuição Weibull padrão.
8. Ajuste os modelos ARMA-GARCH sugeridos na Seção 8.4 aos dados de retornos diários e intradiários do Ibovespa.
9. Ajuste o modelo ARFIMA sugerido na Seção 8.4 aos dados de volatilidade realizada do Ibovespa.
10. Ajuste modelos ARMA-GARCH para os dados intradiários da Telemar. Escolha o melhor modelo segundo as medidas dadas na Seção 8.4.
11. Ajuste modelos ARFIMA( $p, d, q$ ) para a volatilidade realizada da Telemar obtida no Problema 1. Escolha o modelo que fornece o melhor EQM de previsão.
12. Considere os dados de taxa de câmbio Euro/Dólar de 1 de janeiro de 1999 a 31 de dezembro de 2002, a cada cinco minutos, num total de  $T = 288.860$  observações (arquivo id-eudo99.02.dat). Obtenha o gráfico, histograma e f.a.c. dos retornos e dos quadrados dos retornos da série.
13. Para os dados do problema anterior, obtenha a volatilidade realizada, o logaritmo da volatilidade realizada, seus histogramas e f.a.c.'s.
14. Ajuste um modelo ARFIMA para a variância realizada dos dados do Euro/Dólar.

## 8.8 Apêndice 8: Notas Complementares

### A.8.1. Tipos de Mercados

Podemos dividir os mercados acionários em duas categorias:

- (a) *price driven markets*, nos quais há a figura dos especialistas ou *market-makers* (que trabalham para as próprias bolsas) e que são os responsáveis por comprar (ao *bid price*) e vender (ao *ask price*) lotes de ações, evitando movimentos abusivos nos preços das ações, e dando liquidez ao mercado. Brevemente, liquidez é definida como
-

a habilidade de comprar e vender rapidamente um grande volume de ações com um mínimo de impacto nos preços. A bolsa Nasdaq adota esse mecanismo.

(b) *order driven markets*, nos quais não há market-makers no processo de negociações, mas as ordens são feitas num *order book*, ou seja, cada ordem é colocada automaticamente no livro de ofertas de cada ação, sendo possível visualizar as melhores compras e vendas por lote, preço e contraparte. A Bovespa adota esse mecanismo, sendo que as corretoras contratadas por empresas para dar liquidez a seus papéis é que colocam as ordens. A NYSE adota um sistema híbrido. A Bovespa também poderia ser pensada dessa forma, supondo que as corretoras fazem o papel dos *market-makers*.

Para detalhes sobre esse assunto e sobre liquidez de mercados veja Bauwens e Giot (2001).

### A.8.2. Mercado de Taxa de Câmbio(FX)

Esse é um tipo especial de mercado, funcionando continuamente, 24 horas por dia, do tipo *price driven market*. Os participantes são *market-makers (dealers)*, representando grandes bancos.

Podemos considerar três zonas, correspondendo a três mercados geográficos: Ásia, Europa e Estados Unidos. Esses mercados operam da seguinte forma: após a meia-noite GMT os mercados de Tokyo, Hong Kong e Singapura estão ativos, com um decréscimo de operações ao redor das 4 horas GMT, que corresponde ao horário de almoço. Por volta das 8 horas GMT, as negociações começam na Europa (Londres, Frankfurt e Paris) e param na Ásia. A última zona começa por volta das 15 horas GMT com a abertura da bolsa de Nova Iorque. Este padrão intradiário é bem conhecido.

Outra característica dos mercados FX é que as negociações ocorrem somente entre os *dealers* e a informação sobre preços e volumes permanece confidencial. Veja Bauwens e Giot (2001) para detalhes.

### A.8.3. Algumas Distribuições Especiais

#### Distribuição exponencial

A v.a.  $X$  tem distribuição exponencial de parâmetro  $\beta > 0$  se sua densidade é dada por

$$f(x, \beta) = \frac{1}{\beta} \exp\{-x/\beta\}.I\{x \geq 0\}.$$

Se  $\beta = 1$ , obtemos a distribuição exponencial padrão,  $E(X) = 1$  e escrevemos  $X \sim E(1)$ .

**Distribuição Gama**

Dizemos que a v.a.  $X$  tem distribuição gama com parâmetros  $\alpha > 0$  e  $\beta > 0$ , e escrevemos  $X \sim \text{Gama}(\alpha, \beta)$ , se a densidade é dada por

$$f(x, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\{-x/\beta\}.I\{x \geq 0\}.$$

Se  $\beta = 1$ , obtemos a distribuição gama padrão com parâmetro  $\alpha$ .

**Distribuição de Weibull**

A v.a.  $X$  tem distribuição de Weibull com parâmetros  $\alpha > 0, \beta > 0$  se a densidade é dada por

$$f(x, \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\{-(x/\beta)^\alpha\}.I\{x \geq 0\}.$$

Se  $\alpha = 1$  obtemos a distribuição exponencial e se  $Y = \frac{X}{\beta \Gamma(1+1/\alpha)}$  obtemos a distribuição de Weibull padrão, com  $E(Y) = 1$  e densidade

$$f(y, \alpha) = 1 - \exp\{-[\Gamma(1 + 1/\alpha)y]^\alpha\}.I\{y \geq 0\}.$$

**Distribuição Gama Generalizada**

Dizemos que  $X$  tem distribuição gama generalizada com parâmetro de escala  $\beta$  e parâmetros de forma  $\alpha$  e  $\gamma$ , todos positivos, se a densidade é da forma

$$f(x, \gamma, \beta, \alpha) = \frac{\gamma}{\beta^{\gamma\alpha} \Gamma(\alpha)} x^{\gamma\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right].I\{x \geq 0\}.$$

Se  $Y = \lambda X / \beta$ , com  $\lambda = \Gamma(\alpha)/\Gamma(\alpha + 1/\gamma)$ , então  $E(Y) = 1$  e  $Y$  tem distribuição gama generalizada padrão. Se  $\alpha = 1$ , obtemos a distribuição de Weibull, logo exponencial e Weibull são casos particulares da gama generalizada.

# Modelos Lineares Multivariados

## 9.1 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 autocorrelação contida em cada série é importante, estaremos estudando as relações dinâmicas entre as séries componentes. Usaremos a notação  $\mathbf{X}_t = (X_{1t}, X_{2t}, \dots, X_{nt})'$ ,  $t \in \mathbb{Z}$  e  $X_{it}$  ou  $X_{i,t}$ , indistintamente, para a  $i$ -ésima componente,  $i = 1, \dots, n$ .

**Exemplo 9.1.** 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 Petrobras, Banco do Brasil, Telemar e Banco Itaú, e neste caso  $n = 4$ .

O vetor de médias de  $\mathbf{X}_t$  será denotado por

$$\boldsymbol{\mu}_t = E(\mathbf{X}_t) = (\mu_{1t}, \mu_{2t}, \dots, \mu_{nt})' \quad (9.1)$$

e depende, em geral, de  $t$ .

A matriz de covariâncias de  $\mathbf{X}_t$  é definida por

$$\boldsymbol{\Gamma}(t + \tau, t) = E\{(\mathbf{X}_{t+\tau} - \boldsymbol{\mu}_{t+\tau})(\mathbf{X}_t - \boldsymbol{\mu}_t)'\}, \quad (9.2)$$

que é uma matriz  $n \times n$  e que, em geral, também depende de  $t$ .

As quantidades (9.1) e (9.2) descrevem as propriedades de segunda ordem das séries  $X_{1t}, \dots, X_{nt}$ . Se essas tiverem uma distribuição normal multivariada, as propriedades das séries serão completamente especificadas pelas médias e covariâncias. Observe que (9.2) fornece as autocovariâncias 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  $\Gamma(t + \tau, t)$ , então

$$\begin{aligned}\gamma_{ij}(t + \tau, t) &= \text{Cov}\{X_{i,t+\tau}, X_{j,t}\} \\ &= E\{(X_{i,t+\tau} - \mu_{i,t+\tau})(X_{j,t} - \mu_{j,t})\},\end{aligned}\quad (9.3)$$

$i, j = 1, \dots, n$ , é a covariância entre as séries  $X_{i,t+\tau}$  e  $X_{j,t}$ .

**Exemplo 9.2.** No Exemplo 9.1, 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 (9.2) ficará

$$\Gamma(t + \tau, t) = \begin{bmatrix} \gamma_{11}(t + \tau, t) & \gamma_{12}(t + \tau, t) & \cdots & \gamma_{14}(t + \tau, t) \\ \gamma_{21}(t + \tau, t) & \gamma_{22}(t + \tau, t) & \cdots & \gamma_{24}(t + \tau, t) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{41}(t + \tau, t) & \gamma_{42}(t + \tau, t) & \cdots & \gamma_{44}(t + \tau, t) \end{bmatrix}$$

Na diagonal principal, temos as autocovariâncias 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.

## 9.2 Séries Estacionárias

Vamos nos restringir nesse capítulo ao caso de estacionariedade fraca ou de segunda ordem. Dizemos que a série  $n$ -variada  $\mathbf{X}_t$  é *estacionária* se a média  $\boldsymbol{\mu}_t$  e a matriz de covariâncias  $\Gamma(t + \tau, t)$ ,  $t, \tau \in \mathbb{Z}$ , não dependerem do tempo  $t$ . Nessa situação, teremos

$$\boldsymbol{\mu} = E(\mathbf{X}_t) = (\mu_1, \dots, \mu_n)', \quad (9.4)$$

e

$$\Gamma(\tau) = E\{(\mathbf{X}_{t+\tau} - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\} = [\gamma_{ij}(\tau)]_{i,j=1}^n, \quad (9.5)$$

$\tau \in \mathbb{Z}$ . Nesse caso,  $\gamma_{ii}(\tau)$  será a função de autocovariância da série estacionária  $X_{it}$  e  $\gamma_{ij}(\tau)$  será a função de covariância cruzada de  $X_{it}$  e  $X_{jt}$ . Notemos que, em geral,  $\gamma_{ij}(\tau) \neq \gamma_{ji}(\tau)$ .

No caso particular de  $\tau = 0$  em (9.5), obtemos

$$\Gamma(0) = E\{(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_t - \boldsymbol{\mu})'\}, \quad (9.6)$$

que é a *matriz de covariâncias contemporâneas*. Note que  $\gamma_{ii}(0) = \text{Var}(X_{it})$ ,  $\gamma_{ij}(0) = \text{Cov}\{X_{it}, X_{jt}\}$ .

O coeficiente de correlação contemporâneo entre  $X_{it}$  e  $X_{jt}$  é então dado por

$$\rho_{ij}(0) = \frac{\gamma_{ij}(0)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}. \quad (9.7)$$

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  $\boldsymbol{\rho}(0) = [\rho_{ij}(0)]_{i,j=1}^n$  é uma matriz simétrica, com elementos na diagonal principal todos iguais a um.

A matriz de correlações de lag  $\tau$  é definida por

$$\boldsymbol{\rho}(\tau) = \mathbf{D}^{-1}\boldsymbol{\Gamma}(\tau)\mathbf{D}^{-1}, \quad (9.8)$$

sendo  $\mathbf{D} = \text{diag}\{\sqrt{\gamma_{11}(0)}, \dots, \sqrt{\gamma_{nn}(0)}\}$ . Ou seja, denotando  $\boldsymbol{\rho}(\tau) = [\rho_{ij}(\tau)]_{i,j=1}^n$ , temos

$$\rho_{ij}(\tau) = \frac{\gamma_{ij}(\tau)}{[\gamma_{ii}(0)\gamma_{jj}(0)]^{1/2}}, \quad (9.9)$$

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  $\boldsymbol{\Gamma}(\tau)$  e  $\boldsymbol{\rho}(\tau)$  não são, em geral, simétricas. O que vale é a seguinte proposição.

**Proposição 9.1.** As seguintes propriedades são válidas:

- (i)  $\boldsymbol{\Gamma}(\tau) = \boldsymbol{\Gamma}'(-\tau)$ .
- (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 autocovariância, para todo  $i$ .
- (iv)  $\sum_{j,k=1}^m \mathbf{a}_j' \boldsymbol{\Gamma}(j-k) \mathbf{a}_k \geq 0$ , para quaisquer  $m$  e  $\mathbf{a}_1, \dots, \mathbf{a}_m$  vetores de  $\mathbb{R}^n$ .

As demonstrações de (i)-(iii) são imediatas. O item (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 9.3.** 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,

$$\boldsymbol{\Gamma}(0) = E(\mathbf{X}_t \mathbf{X}_t') =$$

$$E \begin{bmatrix} a_t^2 & 0,3X_{2,t-1}a_t + a_t^2 \\ 0,3X_{2,t-1}a_t + a_t^2 & (0,3X_{2,t-1} + a_t)^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1,099 \end{bmatrix},$$

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  $E(0,09X_{2,t-1}^2 + a_t^2 + 0,6a_tX_{2,t-1}) = 0,09\text{Var}(X_{2,t-1}) + 1 = 1,099$ .

Agora,

$$\boldsymbol{\Gamma}(1) = E(\mathbf{X}_{t+1} \mathbf{X}_t') =$$

$$E \begin{bmatrix} a_{t+1}a_t & 0,3X_{2,t-1}a_{t+1} + a_{t+1}a_t \\ 0,3X_{2,t}a_t + a_ta_{t+1} & (0,3X_{2,t} + a_{t+1})(0,3X_{2,t-1} + a_t) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0,3 & 0,33 \end{bmatrix},$$

pois

$$\begin{aligned} E(a_ta_{t+1}) &= 0, E(X_{2,t-1}a_{t+1}) = 0, E(X_{2,t}a_t) = E(a_t^2) = 1, \\ E(X_{2,t}X_{2,t-1}) &= 0, 3E(X_{2,t-1}^2) = (0,3)(0,91), \end{aligned}$$

de modo que o valor de  $\gamma_{2,2}(1)$  é dado por

$$E(0,09X_{2,t}X_{2,t-1} + 0,3X_{2,t}a_t + 0,3X_{2,t-1}a_{t+1} + a_ta_{t+1}) = (0,09)(0,3)/0,91 + 0,3 = 0,3/0,91 = 0,33.$$

Segue-se que

$$\boldsymbol{\Gamma}(-1) = \boldsymbol{\Gamma}'(1) = \begin{bmatrix} 0 & 0,3 \\ 0 & 0,33 \end{bmatrix}.$$

De modo análogo podemos calcular  $\boldsymbol{\Gamma}(2), \boldsymbol{\Gamma}(3)$  etc.

Dizemos que a série  $\{\mathbf{a}_t, t \in \mathbb{Z}\}$  é um *ruído branco multivariado* ( $n \times 1$ ), com média  $\mathbf{0}$  e matriz de covariâncias  $\boldsymbol{\Sigma}$ , se  $\mathbf{a}_t$  é estacionário com média  $\mathbf{0}$  e sua matriz de covariâncias é dada por

$$\boldsymbol{\Gamma}(\tau) = \begin{cases} \boldsymbol{\Sigma}, & \text{se } \tau = 0, \\ \mathbf{0}, & \text{se } \tau \neq 0. \end{cases} \quad (9.10)$$

Usaremos a notação  $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \boldsymbol{\Sigma})$ . Se além disso os vetores  $\mathbf{a}_t$  forem independentes e identicamente distribuídos, escreveremos  $\mathbf{a}_t \sim \text{IID}(\mathbf{0}, \boldsymbol{\Sigma})$ .

Um processo  $\mathbf{X}_t$  diz-se *linear* se

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{a}_{t-j}, \quad (9.11)$$

onde  $\mathbf{a}_t$  é ruído branco multivariado e  $\boldsymbol{\Psi}_j$  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

$$\boldsymbol{\Gamma}(\tau) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}_{j+\tau} \boldsymbol{\Sigma} \boldsymbol{\Psi}'_j, \quad \tau \in \mathbb{Z}. \quad (9.12)$$

### 9.3 Estimação de Médias e Covariâncias

Supondo que temos observações  $\{\mathbf{X}_t, t = 1, \dots, T\}$  do processo estacionário  $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ , a média  $\boldsymbol{\mu}$  pode ser estimada pelo vetor de médias amostrais

$$\bar{\mathbf{X}} = \frac{\sum_{t=1}^T \mathbf{X}_t}{T}. \quad (9.13)$$

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  $\boldsymbol{\Gamma}(\tau)$  usamos

$$\hat{\boldsymbol{\Gamma}}(\tau) = \begin{cases} \frac{1}{T} \sum_{t=1}^{T-\tau} (\mathbf{X}_{t+\tau} - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})', & 0 \leq \tau \leq T-1 \\ \frac{1}{T} \sum_{t=-\tau+1}^T (\mathbf{X}_{t+\tau} - \bar{\mathbf{X}})(\mathbf{X}_t - \bar{\mathbf{X}})', & -T+1 \leq \tau \leq 0. \end{cases} \quad (9.14)$$

A matriz de correlações pode ser estimada por

$$\hat{\rho}(\tau) = \hat{\mathbf{D}}^{-1} \hat{\boldsymbol{\Gamma}}(\tau) \hat{\mathbf{D}}^{-1}, \quad (9.15)$$

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  $\hat{\boldsymbol{\Gamma}}(\tau)$  e  $\hat{\rho}(\tau)$ .

**Exemplo 9.4.** Suponha que  $X_{1t}$  represente os retornos diários da Petrobras e  $X_{2t}$  os retornos diários do Ibovespa, de 19/08/1998 a 29/09/2010, com  $T = 2.998$

observações. Seja  $\mathbf{X}_t = (X_{1t}, X_{2t})'$ . O programa S+FinMetrics fornece as matrizes de correlações amostrais da Tabela 9.1. Uma maneira conveniente de representar essas matrizes é usar os símbolos +, - e ·, 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  $2/\sqrt{T}$ . Essas matrizes pictóricas também estão apresentadas na Tabela 9.1.

Vemos, por exemplo, que

$$\hat{\rho}(0) = \begin{bmatrix} 1,00 & 0,303 \\ 0,303 & 1,00 \end{bmatrix},$$

enquanto

$$\hat{\rho}(1) = \begin{bmatrix} 0,085 & 0,004 \\ 0,424 & 0,002 \end{bmatrix}.$$

Como  $2/\sqrt{2.998} = 0,037$ , os elementos  $\rho_{12}(1)$  e  $\rho_{22}(1)$  podem ser considerados estatisticamente nulos, de modo que a representação pictórica dessa matriz de correlações amostrais é

$$\begin{bmatrix} + & \cdot \\ + & \cdot \end{bmatrix}.$$

Note que a correlação contemporânea entre as duas séries é 0,303.

Tabela 9.1: Matrizes de correlações amostrais para retornos diários da Petrobras e Ibovespa, com notação simplificada.

lag 1	lag 2	lag 3	lag 4
$\begin{bmatrix} 0,085 & 0,004 \\ 0,424 & 0,002 \end{bmatrix}$	$\begin{bmatrix} -0,042 & -0,021 \\ 0,062 & 0,002 \end{bmatrix}$	$\begin{bmatrix} -0,065 & 0,006 \\ -0,026 & -0,041 \end{bmatrix}$	$\begin{bmatrix} -0,041 & -0,032 \\ -0,089 & -0,047 \end{bmatrix}$
$\begin{bmatrix} + & \cdot \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ + & \cdot \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ \cdot & - \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ - & - \end{bmatrix}$

**Exemplo 9.5.** 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 9.2 temos as matrizes de correlações amostrais até o lag 4. Vemos que  $\rho(\tau)$ ,  $\tau = 1, 2, \dots, 4$ , podem ser consideradas nulas, o que sugere que, possivelmente, estamos na presença de um ruído branco bivariado. É fácil verificar que a correlação contemporânea entre as duas séries é negativa (-0,77).

Tabela 9.2: Matrizes de correlações amostrais para os retornos mensais do Ibovespa e C-Bond, com notação simplificada.

lag 1	lag 2	lag 3	lag 4
$\begin{bmatrix} 0,11 & -0,05 \\ 0,10 & -0,15 \end{bmatrix}$	$\begin{bmatrix} -0,20 & 0,08 \\ 0,08 & -0,04 \end{bmatrix}$	$\begin{bmatrix} -0,10 & 0,07 \\ 0,04 & -0,06 \end{bmatrix}$	$\begin{bmatrix} 0,09 & -0,12 \\ -0,04 & 0,07 \end{bmatrix}$
$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$	$\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$

## 9.4 Modelos Autorregressivos Vetoriais

Nesta seção estudaremos uma importante classe de modelos lineares multivariados, a dos modelos autorregressivos vetoriais de ordem  $p$ , que denotaremos por VAR( $p$ ) (de *vector autoregression*). Não confundir a notação VAR com VaR, de valor em risco, utilizada no Capítulo 7.

Dizemos que o processo  $\mathbf{X}_t$ , de ordem  $n \times 1$ , segue um modelo VAR( $p$ ) se

$$\mathbf{X}_t = \Phi_0 + \Phi_1 \mathbf{X}_{t-1} + \dots + \Phi_p \mathbf{X}_{t-p} + \mathbf{a}_t, \quad (9.16)$$

onde  $\mathbf{a}_t \sim RB(\mathbf{0}, \Sigma)$ ,  $\Phi_0 = (\phi_{10}, \dots, \phi_{n0})'$  é um vetor  $n \times 1$  de constantes e  $\Phi_k$  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 (9.16) pode ser escrito na forma

$$\Phi(B)\mathbf{X}_t = \Phi_0 + \mathbf{a}_t, \quad (9.17)$$

onde  $\Phi(B) = \mathbf{I}_n - \Phi_1 B - \dots - \Phi_p B^p$  é o operador autorregressivo vetorial de ordem  $p$ , ou ainda, um polinômio matricial  $n \times n$  em  $B$ . O elemento genérico de  $\Phi(B)$  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,

$$\mathbf{X}_t = \Phi_0 + \Phi \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (9.18)$$

Um caso especial é quando  $n = 2$  e (9.18) reduz-se a

$$\begin{aligned} X_{1t} &= \phi_{10} + \phi_{11}X_{1,t-1} + \phi_{12}X_{2,t-1} + a_{1t}, \\ X_{2t} &= \phi_{20} + \phi_{21}X_{1,t-1} + \phi_{22}X_{2,t-1} + a_{2t}, \end{aligned} \quad (9.19)$$

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 (9.19), não fica explicitada a dependência *contemporânea* entre  $X_{1t}$  e  $X_{2t}$ . Dizemos que (9.19) e (9.18) são *modelos em forma reduzida*. É possível

obter o modelo na forma *estrutural*, em que essa relação fica explicitada. Veja o Apêndice 9.D para detalhes. O modelo em forma reduzida é preferido por facilidades de estimação e previsão.

Retomemos (9.19). 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 (9.18) será estacionário se a média for constante e  $E(\mathbf{X}_{t+\tau}\mathbf{X}'_t)$  independente de  $t$ . Neste caso, se  $\boldsymbol{\mu} = E(\mathbf{X}_t)$ , teremos

$$\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}_0.$$

Segue-se que o modelo poderá ser escrito na forma

$$\mathbf{X}_t - \boldsymbol{\mu} = \boldsymbol{\Phi}(\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \mathbf{a}_t,$$

ou ainda, se  $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{\mu}$ ,

$$\tilde{\mathbf{X}}_t = \boldsymbol{\Phi}\tilde{\mathbf{X}}_{t-1} + \mathbf{a}_t. \quad (9.20)$$

Assim como no caso de um AR(1) univariado, obtemos de (9.20) que

$$\tilde{\mathbf{X}}_t = \mathbf{a}_t + \boldsymbol{\Phi}\mathbf{a}_{t-1} + \boldsymbol{\Phi}^2\mathbf{a}_{t-2} + \dots, \quad (9.21)$$

ou seja, temos a representação MA( $\infty$ ) do modelo. Também, é fácil ver que temos  $\text{Cov}(\mathbf{a}_t, \mathbf{X}_{t-1}) = \mathbf{0}$  e  $\text{Cov}(\mathbf{a}_t, \mathbf{X}_t) = \boldsymbol{\Sigma}$ .

Iremos indicar por  $|\mathbf{A}|$  o determinante da matriz quadrada  $\mathbf{A}$ .

**Proposição 9.2.** O processo  $\mathbf{X}_t$  seguindo um modelo VAR(1) será estacionário se todas as soluções de

$$|\mathbf{I}_n - \boldsymbol{\Phi}z| = 0 \quad (9.22)$$

estiverem fora do círculo unitário.

Como as soluções de (9.22) são inversas dos autovalores de  $\boldsymbol{\Phi}$ , uma condição equivalente é que todos os autovalores de  $\boldsymbol{\Phi}$  sejam menores do que um, em módulo. Ou ainda,  $|\mathbf{I}_n - \boldsymbol{\Phi}z| \neq 0$ ,  $|z| \leq 1$ . A demonstração da Proposição 9.2 está dada no Apêndice 9.B.

**Exemplo 9.6.** No caso de um VAR(1) bivariado, temos que (9.22) fica

$$\begin{vmatrix} 1 - \phi_{11}z & -\phi_{12}z \\ -\phi_{21}z & 1 - \phi_{22}z \end{vmatrix} = (1 - \phi_{11}z)(1 - \phi_{22}z) - \phi_{12}\phi_{21}z^2 = 0,$$

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 dessa 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 9.7.** Consideremos o modelo VAR(1) ( $n = 2$ )

$$\begin{aligned} X_{1,t} &= 0,4 + 0,5X_{1,t-1} + 0,3X_{2,t-1} + a_{1,t}, \\ X_{2,t} &= -1,7 - 0,6X_{1,t-1} - 0,1X_{2,t-1} + a_{2,t}, \end{aligned}$$

com

$$\Sigma = \begin{bmatrix} 1 & 0,5 \\ 0,5 & 1 \end{bmatrix},$$

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 9.1, as duas séries simuladas. É fácil ver que este modelo é estacionário.

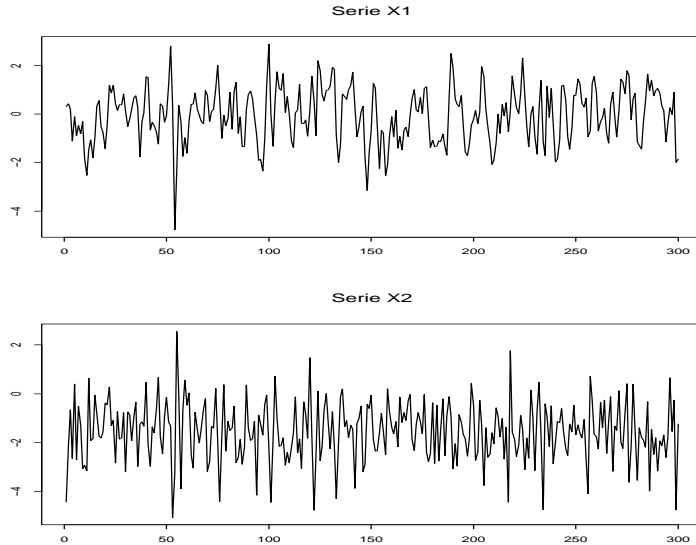


Figura 9.1: Modelo VAR(1) estacionário simulado.

Calculemos a matriz de covariâncias de  $\mathbf{X}_t$ , admitindo-se o modelo (9.20). Usando (9.21) temos que

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Sigma} + \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Phi}' + \boldsymbol{\Phi}^2\boldsymbol{\Sigma}(\boldsymbol{\Phi}^2)' + \dots = \sum_{j=0}^{\infty} \boldsymbol{\Phi}^j\boldsymbol{\Sigma}(\boldsymbol{\Phi}^j)', \quad \boldsymbol{\Phi}_0^0 = \mathbf{I}_n.$$

Uma fórmula análoga vale para  $\boldsymbol{\Gamma}(\tau)$ , veja o Problema 4. No entanto, essas fórmulas envolvem somas infinitas, pouco úteis para cálculos. Vejamos uma maneira mais atraente.

Se pós-multiplicarmos (9.20) por  $\tilde{\mathbf{X}}_{t-\tau}'$  e tomarmos a esperança, obteremos

$$E(\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_{t-\tau}') = \boldsymbol{\Phi} E(\tilde{\mathbf{X}}_{t-1} \tilde{\mathbf{X}}_{t-\tau}') + E(\mathbf{a}_t \tilde{\mathbf{X}}_{t-\tau}').$$

Fazendo  $\tau = 0$ , obtemos

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Phi}\boldsymbol{\Gamma}(-1) + \boldsymbol{\Sigma} = \boldsymbol{\Phi}\boldsymbol{\Gamma}(1)' + \boldsymbol{\Sigma}.$$

Ou seja, para calcular  $\boldsymbol{\Gamma}(0)$ , necessitamos de  $\boldsymbol{\Phi}$ ,  $\boldsymbol{\Sigma}$  e  $\boldsymbol{\Gamma}(1)$ . Como o termo  $E(\mathbf{a}_t \tilde{\mathbf{X}}_{t-\tau}')$  é nulo para  $\tau > 0$ , obtemos

$$\boldsymbol{\Gamma}(\tau) = \boldsymbol{\Phi}\boldsymbol{\Gamma}(\tau-1), \quad \tau > 0,$$

e, por substituições sucessivas, encontramos

$$\boldsymbol{\Gamma}(\tau) = \boldsymbol{\Phi}^\tau \boldsymbol{\Gamma}(0), \quad \tau > 0. \tag{9.23}$$

Segue-se que  $\boldsymbol{\Gamma}(1) = \boldsymbol{\Phi}\boldsymbol{\Gamma}(0)$ , donde

$$\boldsymbol{\Gamma}(0) = \boldsymbol{\Phi}\boldsymbol{\Gamma}(0)\boldsymbol{\Phi}' + \boldsymbol{\Sigma},$$

e, tomindo o operador vec em ambos os membros, obtemos

$$\text{vec}(\boldsymbol{\Gamma}(0)) = \text{vec}(\boldsymbol{\Phi}\boldsymbol{\Gamma}(0)\boldsymbol{\Phi}') + \text{vec}(\boldsymbol{\Sigma}),$$

e como  $\text{vec}(\boldsymbol{\Phi}\boldsymbol{\Gamma}(0)\boldsymbol{\Phi}') = (\boldsymbol{\Phi} \otimes \boldsymbol{\Phi})\text{vec}(\boldsymbol{\Gamma}(0))$ , obtemos, finalmente,

$$\text{vec}(\boldsymbol{\Gamma}(0)) = (\mathbf{I}_{n^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1}\text{vec}(\boldsymbol{\Sigma}).$$

Veja o Apêndice A para noções sobre o operador vec e o produto de Kronecker  $\otimes$  entre duas matrizes.

Nessa expressão a inversa existe pois  $|\mathbf{I}_{n^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi}| \neq 0$ , em razão do fato que os autovalores do produto  $\boldsymbol{\Phi} \otimes \boldsymbol{\Phi}$  são iguais aos produtos dos autovalores de  $\boldsymbol{\Phi}$ , logo têm módulos menores do que 1.

Os resultados anteriores podem ser estendidos a processos VAR( $p$ ),  $p > 1$ , porque tal processo pode sempre ser escrito na forma VAR(1). Veja o Apêndice 9.C. Segue-se que a seguinte proposição é válida.

**Proposição 9.3.** Para o modelo VAR( $p$ ) dado em (9.16) temos os seguintes resultados:

(i) O processo  $\mathbf{X}_t$  será estacionário se as soluções de

$$|\mathbf{I}_n - \boldsymbol{\Phi}_1 z - \dots - \boldsymbol{\Phi}_p z^p| = 0$$

estiverem fora do círculo unitário.

(ii) Se  $\mathbf{X}_t$  for estacionário,

$$\boldsymbol{\mu} = E(\mathbf{X}_t) = (\mathbf{I}_n - \boldsymbol{\Phi}_1 - \dots - \boldsymbol{\Phi}_p)^{-1}\boldsymbol{\Phi}_0.$$

(iii) Escrevendo (9.16) na forma

$$\tilde{\mathbf{X}}_t = \boldsymbol{\Phi}_1 \tilde{\mathbf{X}}_{t-1} + \dots + \boldsymbol{\Phi}_p \tilde{\mathbf{X}}_{t-p} + \mathbf{a}_t,$$

com  $\tilde{\mathbf{X}}_t = \mathbf{X}_t - \boldsymbol{\mu}$  e multiplicando esta equação por  $\tilde{\mathbf{X}}'_{t-\tau}$ , obtemos

$$\boldsymbol{\Gamma}(\tau) = \boldsymbol{\Phi}_1 \boldsymbol{\Gamma}(\tau-1) + \dots + \boldsymbol{\Phi}_p \boldsymbol{\Gamma}(\tau-p), \quad \tau > 0,$$

que são as equações de Yule-Walker, no caso de um modelo VAR( $p$ ).

A demonstração de (i) está no Apêndice 9.C e (ii) e (iii) são imediatas. Para uma condição equivalente a (i) veja o Problema 11. Observe que

$$\begin{aligned}\boldsymbol{\Gamma}(0) &= \boldsymbol{\Phi}_1\boldsymbol{\Gamma}(-1) + \dots + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}(-p) + \boldsymbol{\Sigma} \\ &= \boldsymbol{\Phi}_1\boldsymbol{\Gamma}(1)' + \dots + \boldsymbol{\Phi}_p\boldsymbol{\Gamma}(p)' + \boldsymbol{\Sigma}.\end{aligned}$$

Essas equações podem ser utilizadas para calcular  $\boldsymbol{\Gamma}(\tau)$  recursivamente, para  $\tau \geq p$ . Para  $|\tau| < p$ , temos que usar a representação VAR(1) de um processo VAR( $p$ ). Veja o Apêndice 9.C.

## 9.5 Construção de Modelos VAR

A construção de modelos VAR segue o mesmo ciclo de identificação, estimativa 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 autorregressivos vetoriais de ordens  $1, 2, \dots, k$  e testar a significância dos coeficientes (matrizes). Considere, pois, os modelos

$$\begin{aligned}\mathbf{X}_t &= \boldsymbol{\Phi}_0^{(1)} + \boldsymbol{\Phi}_1^{(1)}\mathbf{X}_{t-1} + \mathbf{a}_t^{(1)}, \\ \mathbf{X}_t &= \boldsymbol{\Phi}_0^{(2)} + \boldsymbol{\Phi}_1^{(2)}\mathbf{X}_{t-1} + \boldsymbol{\Phi}_2^{(2)}\mathbf{X}_{t-2} + \mathbf{a}_t^{(2)}, \\ &\dots \quad \dots \\ \mathbf{X}_t &= \boldsymbol{\Phi}_0^{(k)} + \boldsymbol{\Phi}_1^{(k)}\mathbf{X}_{t-1} + \dots + \boldsymbol{\Phi}_k^{(k)}\mathbf{X}_{t-k} + \mathbf{a}_t^{(k)}.\end{aligned}\tag{9.24}$$

Os parâmetros podem ser estimados por MQ ordinários, que fornecem estimadores consistentes e eficientes. Testamos, então,

$$\begin{aligned}H_0 &: \boldsymbol{\Phi}_k^{(k)} = \mathbf{0}, \\ H_1 &: \boldsymbol{\Phi}_k^{(k)} \neq \mathbf{0}, k = 1, 2, \dots\end{aligned}\tag{9.25}$$

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

$$\hat{\mathbf{a}}_t^{(k)} = \mathbf{X}_t - \hat{\boldsymbol{\Phi}}_0^{(k)} - \hat{\boldsymbol{\Phi}}_1^{(k)}\mathbf{X}_{t-1} - \dots - \hat{\boldsymbol{\Phi}}_k^{(k)}\mathbf{X}_{t-k}.$$

A matriz de covariância dos resíduos, que estima  $\boldsymbol{\Sigma}$ , é dada, então, por

$$\hat{\Sigma}_k = \frac{1}{T-k} \sum_{t=k+1}^T \hat{\mathbf{a}}_t^{(k)} (\hat{\mathbf{a}}_t^{(k)})', \quad k \geq 0, \quad (9.26)$$

onde para  $k=0$ ,  $\hat{\mathbf{a}}_t^{(0)} = \mathbf{X}_t - \bar{\mathbf{X}}$ . A estatística da razão de verossimilhanças para o teste (9.25) é dada por

$$RV(k) = (T-k) \ln \frac{|\hat{\Sigma}_{k-1}|}{|\hat{\Sigma}_k|}, \quad (9.27)$$

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:

$$\begin{aligned} AIC(k) &= \ln(|\hat{\Sigma}_k|) + 2kn^2/T \quad (\text{Akaike}), \\ BIC(k) &= \ln(|\hat{\Sigma}_k|) + kn^2 \ln(T)/T \quad (\text{Schwarz}), \\ HQC(k) &= \ln(|\hat{\Sigma}_k|) + kn^2 \ln(\ln(T))/T \quad (\text{Hannan-Quinn}). \end{aligned} \quad (9.28)$$

O programa **EViews**, por exemplo, fornece esses critérios mais os valores de FPE (*final prediction error*, de Akaike), RV(k) e o valor da log-verossimilhança. O programa **SCA** fornece os valores de 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. Nesse 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

$$\begin{aligned} \ell &= -\frac{n(T+1)}{2} \ln(2\pi) + \frac{(T-1)}{2} \ln |\Sigma^{-1}| \\ &\quad - \frac{1}{2} \sum_{t=2}^T (\mathbf{X}_t - \Phi \mathbf{X}_{t-1})' \Sigma^{-1} (\mathbf{X}_t - \Phi \mathbf{X}_{t-1}), \end{aligned} \quad (9.29)$$

obtendo-se

$$\hat{\Phi} = [\sum_{t=2}^T \mathbf{X}_t \mathbf{X}_{t-1}'] [\sum_{t=2}^T \mathbf{X}_{t-1} \mathbf{X}_{t-1}']^{-1}, \quad (9.30)$$

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{a}}_t (\hat{\mathbf{a}}_t)' , \quad (9.31)$$

$$\hat{\mathbf{a}}_t = \mathbf{X}_t - \hat{\Phi} \mathbf{X}_{t-1}. \quad (9.32)$$

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

$$Q(m) = T^2 \sum_{\tau=1}^m \frac{1}{T-\tau} \text{tr}(\hat{\Gamma}(\tau)' \hat{\Gamma}(0)^{-1} \hat{\Gamma}(\tau) \hat{\Gamma}(0)^{-1}), \quad (9.33)$$

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 (9.18) 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

$$\hat{\mathbf{X}}_T(h) = \Phi^h \mathbf{X}_T, \quad h = 1, 2, \dots \quad (9.34)$$

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

$$\mathbf{e}_T(h) = \mathbf{X}_{t+h} - \hat{\mathbf{X}}_T(h) = \sum_{j=0}^{h-1} \Phi^j \mathbf{a}_{t+h-j}, \quad (9.35)$$

de modo que o erro quadrático médio do previsor (9.34) fica

$$\Sigma(h) = \text{EQMP}(h) = \sum_{j=0}^{h-1} \Phi^j \Sigma (\Phi^j)' . \quad (9.36)$$

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

$$E(\mathbf{X}_{t+h}|\mathcal{F}_t) = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 E(\mathbf{X}_{t+h-1}|\mathcal{F}_t) + \dots + \boldsymbol{\Phi} E(\mathbf{X}_{t+h-p}|\mathcal{F}_t),$$

pois  $E(\mathbf{a}_{t+h}|\mathcal{F}_t) = 0$ , para todo  $h > 0$ .

Para  $h = 1$ , obtemos

$$\hat{\mathbf{X}}_t(1) = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{X}_t + \dots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p+1},$$

e, para  $h = 2$ , temos

$$\hat{\mathbf{X}}_t(2) = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \hat{\mathbf{X}}_t(1) + \boldsymbol{\Phi}_2 \mathbf{X}_t + \dots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p+2},$$

de modo que as previsões podem ser obtidas recursivamente.

Nesse caso, o erro de previsão de horizonte  $h$  é dado por

$$\mathbf{e}_T(h) = \sum_{j=0}^{h-1} \boldsymbol{\Psi}_j \mathbf{a}_{T+h-j}, \quad (9.37)$$

onde as matrizes  $\boldsymbol{\Psi}_j$  são obtidas recursivamente por

$$\boldsymbol{\Psi}_j = \sum_{k=1}^{p-1} \boldsymbol{\Psi}_{j-k} \boldsymbol{\Phi}_k, \quad (9.38)$$

com  $\boldsymbol{\Psi}_0 = \mathbf{I}_n$  e  $\boldsymbol{\Phi}_j = 0, j > p$ . Segue-se que a matriz de EQM de previsão fica

$$\boldsymbol{\Sigma}(h) = \sum_{j=0}^{h-1} \boldsymbol{\Psi}_j \boldsymbol{\Sigma} \boldsymbol{\Psi}'_j. \quad (9.39)$$

Quando os parâmetros do modelo VAR( $p$ ) são estimados, o melhor preditor de  $X_{T+h}$  é, agora, dado por

$$\tilde{\mathbf{X}}_T(h) = \hat{\boldsymbol{\Phi}}_0 + \hat{\boldsymbol{\Phi}}_1 \tilde{\mathbf{X}}_T(h-1) + \dots + \hat{\boldsymbol{\Phi}}_p \tilde{\mathbf{X}}_T(h-p), \quad h > 1. \quad (9.40)$$

Nesse caso, a matriz de EQM de previsão torna-se

$$\hat{\boldsymbol{\Sigma}}(h) = \boldsymbol{\Sigma}(h) + EQM(\mathbf{X}_{T+h} - \tilde{\mathbf{X}}_T(h)). \quad (9.41)$$

Na prática, o segundo termo em (9.41) é ignorado, e  $\hat{\boldsymbol{\Sigma}}(h)$  é calculada por

$$\hat{\boldsymbol{\Sigma}}(h) = \sum_{j=0}^{h-1} \hat{\boldsymbol{\Psi}}_j \hat{\boldsymbol{\Sigma}} \hat{\boldsymbol{\Psi}}'_j, \quad (9.42)$$

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 (9.41).

**Exemplo 9.8.** Ajustemos um modelo VAR( $p$ ) à série  $\mathbf{X}_t$ , onde  $X_{1t}$  é a série de retornos diários da Petrobras e  $X_{2t}$  é a série de retornos diários do Ibovespa, de 19/08/1998 a 29/09/2010, com  $T = 2.998$  observações. Na Tabela 9.3 temos os valores de BIC resultantes de ajustes de modelos autorregressivos vetoriais até ordem 8, usando a função VAR do S+FinMetrics:

```
> var.fit=VAR(my.df,max.ar=8,criterion="BIC"),
```

onde my.df é um `data.frame` contendo as duas séries de retornos. De acordo com os valores de BIC, selecionamos a ordem  $p = 3$ . Usando o mesmo programa para estimar os coeficientes do modelo identificado, obtemos a Tabela 9.4, de modo que o modelo bivariado ajustado é

$$\begin{aligned} X_{1t} &= 0,0010 + 0,0833X_{1,t-1} + 0,0534X_{1,t-2} - 0,0785X_{1,t-3} + a_{1t}, \\ X_{2t} &= 0,4012X_{1,t-1} - 0,2001X_{2,t-1} + 0,1066X_{1,t-2} \\ &\quad - 0,0751X_{2,t-2} + 0,0353X_{1,t-3} - 0,0474X_{2,t-3} + a_{2t}. \end{aligned} \tag{9.43}$$

O vetor de constantes estimado é dado por  $\hat{\Phi}_0 = (0,0010, 0,0004)'$ , sendo que somente o primeiro elemento do vetor é significativo.

Tabela 9.3: Valores de BIC resultantes de ajustes de modelos VAR( $p$ ),  $p = 1, \dots, 8$ , para os retornos diários da Petrobras e do Ibovespa.

Ordem	1	2	3	4	5	6	7	8
BIC	-28280,1	-28309,5	-28310,24	-28300,6	-28274,4	-28250,6	-28230,5	-28200,7

Na Tabela 9.5 temos as representações pictóricas dos coeficientes (matriciais). Note que os retornos diários da Petrobras não são influenciados por valores passados dos retornos diários do Ibovespa. Por outro lado, os retornos do Ibovespa são influenciados por valores defasados dos retornos da Petrobras, o que é razoável, dado que as ações da Petrobras fazem parte do índice. Segue-se que há uma relação de causalidade de  $X_{1t}$  para  $X_{2t}$ . Os valores da log-verossimilhança, AIC, BIC e HQ dados pelo programa são 14.209, 46, -28.390, 93, -28.306, 86 e -28.360, 68, respectivamente.

Tabela 9.4: Ajuste de um modelo VAR(3) aos retornos diários da Petrobras e do Ibovespa. Primeira linha: estimativas; segunda linha: desvios padrões

$\hat{\Phi}_1$	$\hat{\Phi}_2$	$\hat{\Phi}_3$
$\begin{bmatrix} 0,0833 & 0,0054 \\ 0,4012 & -0,2001 \end{bmatrix}$	$\begin{bmatrix} -0,0534 & 0,0332 \\ 0,1066 & -0,0751 \end{bmatrix}$	$\begin{bmatrix} -0,0785 & 0,0379 \\ 0,0353 & -0,0474 \end{bmatrix}$
$\begin{bmatrix} 0,0191 & 0,0253 \\ 0,0145 & 0,0191 \end{bmatrix}$	$\begin{bmatrix} 0,0214 & 0,0255 \\ 0,0162 & 0,0193 \end{bmatrix}$	$\begin{bmatrix} 0,0215 & 0,0228 \\ 0,0163 & 0,0172 \end{bmatrix}$

Tabela 9.5: Representações pictóricas das matrizes da Tabela 9.4

$\hat{\Phi}_1$	$\hat{\Phi}_2$	$\hat{\Phi}_3$
$\begin{bmatrix} + & \cdot \\ + & - \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ + & - \end{bmatrix}$	$\begin{bmatrix} - & \cdot \\ + & - \end{bmatrix}$

Na Figura 9.2, 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.

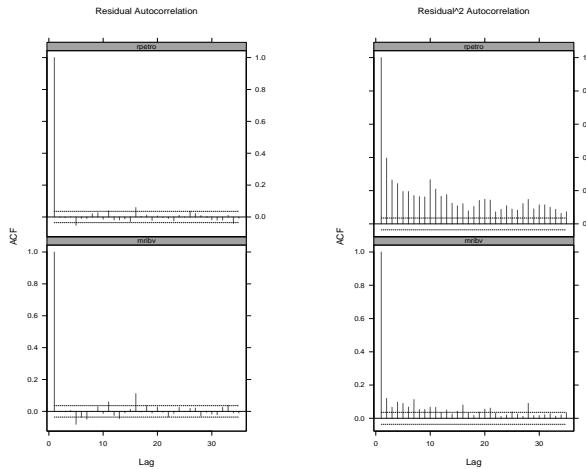


Figura 9.2: F.a.c. amostrais dos resíduos e seus quadrados.

As previsões para horizontes  $h = 1, 2, \dots, 12$  para o modelo (9.43), usando o S+FinMetrics, estão na Tabela 9.6.

Tabela 9.6: Valores previstos para o modelo (9.43),  $h = 1, 2, \dots, 10$ .

$h$	Petrobras (e.p.)	Ibovespa (e.p.)
1	0,0030 (0,0266)	0,0125 (0,0201)
2	-0,0006 (0,0267)	0,0023 (0,0226)
3	-0,0011 (0,0267)	0,0002 (0,0226)
4	0,0013 (0,0268)	-0,0008 (0,0226)
5	0,0013 (0,0268)	0,0008 (0,0227)
6	0,0011 (0,0268)	0,0009 (0,0227)
7	0,0009 (0,0268)	0,0008 (0,0227)
8	0,0010 (0,0268)	0,0007 (0,0227)
9	0,0010 (0,0268)	0,0007 (0,0227)
10	0,0010 (0,0268)	0,0007 (0,0227)

## 9.6 Modelos ARMA Vetoriais

O modelo VAR(p) estudado na seção anterior é um caso particular dos modelos 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

$$\mathbf{X}_t = \boldsymbol{\Theta}_0 + \mathbf{a}_t - \boldsymbol{\Theta}_1 \mathbf{a}_{t-1} - \dots - \boldsymbol{\Theta}_q \mathbf{a}_{t-q}, \quad (9.44)$$

sendo novamente  $\mathbf{a}_t \sim RB(\mathbf{0}, \boldsymbol{\Sigma})$  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

$$\mathbf{X}_t = \boldsymbol{\Theta}_0 + \boldsymbol{\Theta}(B)\mathbf{a}_t, \quad (9.45)$$

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  $q$ .

**Proposição 7.4.** Para o modelo (9.44) temos:

- (i)  $\text{Cov}(\mathbf{X}_t, \mathbf{a}_t) = \boldsymbol{\Sigma}$ ;
- (ii)  $\boldsymbol{\Gamma}(0) = \boldsymbol{\Sigma} + \boldsymbol{\Theta}_1 \boldsymbol{\Sigma} \boldsymbol{\Theta}_1' + \dots + \boldsymbol{\Theta}_q \boldsymbol{\Sigma} \boldsymbol{\Theta}_q'$ ;

(iii) a matriz de covariâncias de  $\mathbf{X}_t$  é dada por

$$\boldsymbol{\Gamma}(\tau) = \begin{cases} \sum_{j=\tau}^q \boldsymbol{\Theta}_j \boldsymbol{\Sigma} \boldsymbol{\Theta}'_{j-\tau}, & \text{se } 1 \leq \tau \leq q, \\ \mathbf{0}, & \text{se } \tau > q, \end{cases} \quad (9.46)$$

sendo  $\boldsymbol{\theta}_0 = -\mathbf{I}_n$ .

De (9.46) segue-se que  $\boldsymbol{\rho}(\tau) = \mathbf{0}$ ,  $\tau > q$ , logo a ordem  $q$  de um modelo VMA( $q$ ) pode ser identificada analisando-se as matrizes de correlações cruzadas amostrais.

**Exemplo 9.9.** O modelo de médias móveis vetorial mais simples é o VMA(1), dado por

$$\mathbf{X}_t = \boldsymbol{\Theta}_0 + \mathbf{a}_t - \boldsymbol{\Theta}\mathbf{a}_{t-1},$$

que tem todas as matrizes de correlações nulas a partir do lag 2, inclusive. Como vimos acima,  $\boldsymbol{\Theta}_0$  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  $\boldsymbol{\Theta}$ . 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 similarmente ao caso univariado.

O modelo VARMA geral é definido por

$$\begin{aligned} \mathbf{X}_t = & \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p} + \\ & + \mathbf{a}_t - \boldsymbol{\Theta}_1 \mathbf{a}_{t-1} - \dots - \boldsymbol{\Theta}_q \mathbf{a}_{t-q}, \end{aligned} \quad (9.47)$$

sendo novamente  $\mathbf{a}_t \sim \text{RB}(\mathbf{0}, \boldsymbol{\Sigma})$  e as matrizes  $\boldsymbol{\Phi}_i$  e  $\boldsymbol{\Theta}_i$  como antes.

Com os polinômios matriciais anteriormente definidos, podemos escrever o modelo na forma

$$\boldsymbol{\Phi}(B)\mathbf{X}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Theta}(B)\mathbf{a}_t. \quad (9.48)$$

---

Esse processo tem uma solução estacionária

$$\mathbf{X}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \mathbf{a}_{t-j} \quad (9.49)$$

se  $|\boldsymbol{\Phi}(z)| \neq 0$ , para todo  $z$  complexo tal que  $|z| \leq 1$ , sendo que as matrizes  $\boldsymbol{\Psi}_j$  são determinadas univocamente por  $\boldsymbol{\Psi}(z) = \boldsymbol{\Phi}^{-1}(z)\boldsymbol{\Theta}(z)$ .

Por outro lado, o processo (9.47) pode ser escrito na forma AR( $\infty$ )

$$\sum_{j=0}^{\infty} \boldsymbol{\Pi}_j \mathbf{X}_{t-j} = \mathbf{a}_t, \quad (9.50)$$

se  $|\boldsymbol{\Theta}(z)| \neq 0$ , para todo complexo  $z$  com  $|z| \leq 1$  e as matrizes  $\boldsymbol{\Pi}_j$  são determinadas univocamente por  $\boldsymbol{\Pi}(z) = \boldsymbol{\Theta}^{-1}(z)\boldsymbol{\Phi}(z)$ . Para detalhes veja Brockwell e Davis (1991). As matrizes  $\boldsymbol{\Psi}_j$  e  $\boldsymbol{\Pi}_j$  podem ser obtidas recursivamente. Veja os Problemas 2 e 3.

Os modelos VARMA não são muito usados, em virtude das dificuldades computacionais e dos problemas de identificabilidade. A matriz de covariâncias (ou a matriz de densidade espectral) não determina univocamente  $\boldsymbol{\Phi}$ ,  $\boldsymbol{\Theta}$  e  $\boldsymbol{\Sigma}$ , 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 (2005).

Para a identificação da ordem  $(p, q)$  de uma modelo VARMA os critérios de informação descritos antes podem ser utilizados. A estimativa pode ser feita via EMV condicionais ou exatos e a estatística de Box-Pierce-Ljung pode ser usada para fins de diagnóstico.

## 9.7 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 9.1.** 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 mais bem 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:

$$\sigma^2(Y_t|\bar{A}_t, \bar{\bar{X}}_t) < \sigma^2(Y_t|\bar{A}_t)$$

Ou seja, o valor presente de  $Y_t$  é mais bem 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 9.2.** Dizemos que:

- (a) O vetor  $\mathbf{X}_t$  causa o vetor  $\mathbf{Y}_t$ , se

$$\sigma^2(Y_{it}|\bar{A}_t) < \sigma^2(Y_{it}|\bar{A}_t - \bar{\mathbf{X}}_t),$$

para pelo menos um valor de  $i = 1, \dots, s$ .

- (b) O vetor  $\mathbf{X}_t$  não causa o vetor  $\mathbf{Y}_t$ , se

$$\sigma^2(Y_{it}|\bar{A}_t) = \sigma^2(Y_{it}|\bar{A}_t - \bar{\mathbf{X}}_t),$$

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 autorregressivos 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 (9.49), ou seja,

$$\mathbf{X}_t = \boldsymbol{\mu} + \boldsymbol{\Psi}(B)\mathbf{a}_t, \quad \boldsymbol{\Psi}_0 = \mathbf{I}_n. \quad (9.51)$$

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

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Psi}_{11}(B) & \boldsymbol{\Psi}_{12}(B) \\ \boldsymbol{\Psi}_{21}(B) & \boldsymbol{\Psi}_{22}(B) \end{bmatrix} \begin{bmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{bmatrix}, \quad (9.52)$$

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 mais bem prevista pelo presente e passado de  $\mathbf{Y}_t$ , mas não o contrário, deveremos ter  $\boldsymbol{\Psi}_{12}(B) = 0$  e obteremos

$$\mathbf{Y}_t = \boldsymbol{\mu}_1 + \boldsymbol{\Psi}_{11}(B)\mathbf{a}_{1t}, \quad (9.53)$$

$$\mathbf{Z}_t = \boldsymbol{\mu}_2 + \boldsymbol{\Psi}_{21}(B)\mathbf{a}_{1t} + \boldsymbol{\Psi}_{22}(B)\mathbf{a}_{2t}. \quad (9.54)$$

Note que (9.54) pode ser escrita

$$\mathbf{Z}_t = \boldsymbol{\mu}_2 + \mathbf{V}(B)\mathbf{Y}_t + \boldsymbol{\Psi}_{22}(B)\mathbf{a}_{2t}, \quad (9.55)$$

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. Nessa 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 (9.51),  $\boldsymbol{\Psi}(B) = \mathbf{I}_n + \boldsymbol{\Psi}_1 b + \boldsymbol{\Psi}_2 B^2 + \dots$  e

$$\boldsymbol{\Psi}_i = \begin{bmatrix} \boldsymbol{\Psi}_{11,i} & \boldsymbol{\Psi}_{12,i} \\ \boldsymbol{\Psi}_{21,i} & \boldsymbol{\Psi}_{22,i} \end{bmatrix}, \quad i = 1, 2, \dots$$

**Proposição 9.5.** O previsor ótimo de  $\mathbf{Y}_t$  baseado em  $\bar{\mathbf{Z}}_t$  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). Essa 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,

$$\begin{aligned}\mathbf{X}_t &= \begin{bmatrix} \mathbf{Y}_t \\ \mathbf{Z}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + \begin{bmatrix} \Phi_{11,1} & \Phi_{12,1} \\ \Phi_{21,1} & \Phi_{22,1} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-1} \\ \mathbf{Z}_{t-1} \end{bmatrix} + \dots \quad (9.56) \\ &\quad + \begin{bmatrix} \Phi_{11,p} & \Phi_{12,p} \\ \Phi_{21,p} & \Phi_{22,p} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{t-p} \\ \mathbf{Z}_{t-p} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_{1t} \\ \mathbf{a}_{2t} \end{bmatrix},\end{aligned}$$

e a condição da Proposição 9.5 estará satisfeita se, e somente se,  $\Phi_{12,i} = \mathbf{0}$ ,  $i = 1, 2, \dots, p$ . 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  $\Phi_{12,i} = \mathbf{0}$ ,  $i = 1, 2, \dots, p$ .

Uma caracterização de não existência de causalidade instantânea é dada pela proposição seguinte. A prova é dada em Lütkepohl (1991).

**Proposição 9.6.** Se  $\mathbf{X}_t$  for como em (9.56), 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  $\Sigma$  pode ser escrita como  $\Sigma = \mathbf{T}\mathbf{T}'$ , 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

$$\begin{aligned}\mathbf{X}_t &= \boldsymbol{\mu} + \sum_{j=0}^{\infty} \Psi_j \mathbf{T}\mathbf{T}^{-1} \mathbf{a}_{t-j} \\ &= \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Xi}_j \mathbf{b}_{t-j},\end{aligned}$$

com  $\boldsymbol{\Xi}_j = \Psi_j \mathbf{T}$ ,  $\mathbf{b}_t = \mathbf{T}^{-1} \mathbf{a}_t \sim RB(\mathbf{0}, \Sigma_b)$ , e  $\Sigma_b = \mathbf{T}^{-1} \Sigma (\mathbf{T}^{-1})' = \mathbf{I}_n$ . Veja também o Apêndice 7.D.

A equação (9.56) pode ser escrita como:

$$\mathbf{Y}_t = \boldsymbol{\mu}_1 + \sum_{i=1}^p \boldsymbol{\Phi}_{11,i} \mathbf{Y}_{t-i} + \sum_{i=1}^p \boldsymbol{\Phi}_{12,i} \mathbf{Z}_{t-i} + \mathbf{a}_{1t}, \quad (9.57)$$

$$\mathbf{Z}_t = \boldsymbol{\mu}_2 + \sum_{i=1}^p \boldsymbol{\Phi}_{21,i} \mathbf{Y}_{t-i} + \sum_{i=1}^p \boldsymbol{\Phi}_{22,i} \mathbf{Z}_{t-i} + \mathbf{a}_{2t}. \quad (9.58)$$

Suponha, também, a matriz  $\Sigma$  particionada como

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

sendo que  $\Sigma_{ij} = E(\mathbf{a}_{it}\mathbf{a}'_{jt})$ ,  $i, j = 1, 2$ .

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 9.7** (i)  $\mathbf{Z}_t$  não causa  $\mathbf{Y}_t \leftrightarrow |\Sigma_{11}| = |\Sigma_1|$ , onde  $\Sigma_1 = E(\mathbf{c}_{1t}\mathbf{c}'_{1t})$  é obtida da regressão restrita

$$\mathbf{Y}_t = \boldsymbol{\nu}_1 + \sum_{i=1}^p \mathbf{A}_i \mathbf{Y}_{t-i} + \mathbf{c}_{1t}. \quad (9.59)$$

(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

$$\mathbf{Z}_t = \boldsymbol{\nu}_2 + \sum_{i=1}^p \mathbf{C}_i \mathbf{Z}_{t-i} + \mathbf{c}_{2t}. \quad (9.60)$$

As regressões (9.57)-(9.60) 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 estabelecidos para causalidade instantânea. Veja o Apêndice 9.D.

**Exemplo 9.10.** Para o Exemplo 9.8, vemos que  $X_{1t} \rightarrow X_{2t}$ , ou seja, retornos diários da Petrobras causam, no sentido de Granger, retornos diários do Ibovespa. Vimos, também, que  $X_{2t}$  não causa  $X_{1t}$ .

**Exemplo 9.11.** Um modelo VAR(1) para as séries de retornos diários da Vale ( $X_{1t}$ ) e da Petrobras ( $X_{2t}$ ) é dado por

$$\begin{aligned} X_{1t} &= 0,0014 + a_{1t}, \\ X_{2t} &= 0,1467X_{1,t-1} + a_{2t}. \end{aligned}$$

Vemos que Vale causa Petrobras, mas não o contrário.

## 9.8 Problemas

1. Para o modelo VMA(1) do Exemplo 9.9, 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 (9.49) e (9.50) 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  $\Gamma(\tau) = \sum_{k=0}^{\infty} \Psi_{\tau+k} \Sigma \Psi_k'$ ,  $\tau \in Z$ .
4. Da representação (B.1) do Apêndice 9.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 (9.61).
6. Ajuste um modelo VAR às séries de retornos diários do Ibovespa e da Cemig, ambas de 02/01/95 a 27/12/2000.
7. Use (C.2) do Apêndice 9.C para obter as matrizes de covariâncias de lags 0, 1 e 2 para o exemplo de modelo VAR(1) dado no Apêndice 9.D.
8. Estabeleça relações de causalidade para as séries dos Exemplos 9.8 e 9.9.
9. Para o Exemplo 9.8, encontre as equações estruturais. Veja o Apêndice 9.D.
10. Para o Exemplo 9.9, encontre as equações estruturais para  $X_{1t}$  e  $X_{2t}$ . Veja o Apêndice 9.D.
11. Prove que a condição (i) da Proposição 9.3 é equivalente a dizer que os autovalores 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 9.C.

12. Como ficaria o problema anterior para um modelo VAR(2)?
  13. Prove (9.30) e (9.31).
  14. Prove (9.38).
  15. Verifique se há causalidade instantânea para os Exemplos 9.10 e 9.11.
  16. Prove a Proposição 9.1.
  17. Ajuste um modelo VAR aos retornos diários dos índices Ibovespa e Merval (Argentina) do arquivo d-indices.95.04.dat.
  18. Mesmo problema, para os retornos do Ibovespa e IPC (México).
-

## 9.9 Apêndice 9.A. Alguns Resultados sobre Matrizes

### A.1 Conceitos Básicos

A notação  $\mathbf{A} = [a_{ij}]$  indicará uma matriz de orde  $m \times n$ . 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

$$\mathbf{x}'\mathbf{Ax} = \sum_{i,j=1}^m a_{ij}x_i x_j \geq 0,$$

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  $\mathbf{x}'\mathbf{Ax} > 0$ , para todo vetor  $\mathbf{x} \neq \mathbf{0}$  e escrevemos  $\mathbf{A} > \mathbf{0}$ . A expressão  $\mathbf{x}'\mathbf{Ax}$  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  $|\mathbf{\Gamma}| = \gamma$ , 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

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_m.$$

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  $\mathbf{A}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ,  $\mathbf{A}_{12} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ,  $\mathbf{A}_{21} = [1 \ 1]$ ,  $\mathbf{A}_{22} = [0]$ , 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

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \cdots & \cdots & \cdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}.$$

Por exemplo,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \otimes \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} a\alpha & a\beta & b\alpha & b\beta \\ a\gamma & a\delta & b\gamma & b\delta \\ c\alpha & c\beta & d\alpha & d\beta \\ c\gamma & c\delta & d\gamma & d\delta \end{bmatrix}.$$

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)  $\text{vec}(\mathbf{AXB}') = (\mathbf{B} \otimes \mathbf{A}) \text{vec}(\mathbf{X})$ .
- (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 Cholesky*. Veja o Apêndice 9.D para exemplos.

## 9.10 Apêndice 9.B. Demonstração da Proposição 9.2

De (9.18) obtemos, recursivamente,

$$\mathbf{X}_t = (\mathbf{I}_n + \Phi + \dots + \Phi^j) \Phi_0 +$$

$$+ \Phi^{j+1} \mathbf{X}_{t-j-1} + \sum_{i=0}^j \Phi^i \mathbf{a}_{t-i}.$$

Se todos os autovalores de  $\Phi$  tiverem módulos menores do que um, a sequência  $\Phi^i$ ,  $i \geq 0$ , é absolutamente somável e a soma  $\sum_{i=0}^{\infty} \Phi^i \mathbf{a}_{t-i}$  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  $\boldsymbol{\mu} = (\mathbf{I}_n - \Phi)^{-1} \Phi_0$ .

### 9.11 Apêndice 9.C. Modelo VAR( $p$ ) na Forma VAR(1)

Suponha que  $\mathbf{X}_t$  seja dado por (9.16). 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 9.3 for válida.

Segue-se de (C.1) que a matriz de covariâncias de  $\mathbf{Y}_t$  é dada por

$$\boldsymbol{\Gamma}_Y(\tau) = \sum_{i=0}^{\infty} \mathbf{F}^{\tau+i} \boldsymbol{\Sigma}_b (\mathbf{F}^i)',$$

onde  $\boldsymbol{\Sigma}_b = E(\mathbf{b}_t \mathbf{b}_t')$ . Novamente, sendo esta uma soma infinita, é mais conveniente usar (iii) da Proposição 9.3. 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

$$\boldsymbol{\Gamma}_Y(0) = \mathbf{F}\boldsymbol{\Gamma}_Y(0)\mathbf{F}' + \boldsymbol{\Sigma}_b$$

e as matrizes  $\boldsymbol{\Gamma}_Y(\tau)$ ,  $\tau = -p+1, \dots, p-1$  podem ser obtidas de

$$\text{vec}(\boldsymbol{\Gamma}_Y(0)) = (\mathbf{I}_{(np)^2} - \mathbf{F} \otimes \mathbf{F})^{-1} \text{vec}(\boldsymbol{\Sigma}_b). \quad (C.2)$$

Note que

$$\begin{aligned} \boldsymbol{\Gamma}_Y(0) &= E \left[ \begin{array}{c} \mathbf{X}_t - \boldsymbol{\nu} \\ \vdots \\ \mathbf{X}_{t-p+1} - \boldsymbol{\nu} \end{array} \right] \left[ (\mathbf{X}_t - \boldsymbol{\nu})', \dots, (\mathbf{X}_{t-p+1} - \boldsymbol{\nu})' \right] \\ &= \left[ \begin{array}{cccc} \boldsymbol{\Gamma}(0) & \boldsymbol{\Gamma}(1) & \cdots & \boldsymbol{\Gamma}(p-1) \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Gamma}(-p+1) & \boldsymbol{\Gamma}(-p+2) & \cdots & \boldsymbol{\Gamma}(0) \end{array} \right]. \end{aligned}$$

## 9.12 Apêndice 9.D. Modelos Estruturais

Como vimos, o modelo VAR( $p$ ) (9.16) está na forma reduzida, porque não explica 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 Cholesky da matriz  $\boldsymbol{\Sigma}$ . Dado que esta é positiva definida, existe uma matriz triangular inferior  $\mathbf{L}$ , com elementos diagonais unitários e uma matriz diagonal  $\boldsymbol{\Lambda}$  tal que  $\boldsymbol{\Sigma} = \mathbf{L}\boldsymbol{\Lambda}\mathbf{L}'$ . Segue-se que  $\mathbf{L}^{-1}\boldsymbol{\Sigma}(\mathbf{L}')^{-1} = \boldsymbol{\Lambda}$ . 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  $\boldsymbol{\Lambda}$ , ou seja, os elementos  $b_{it}$  de  $\mathbf{b}_t$  são não-correlacionados.

Pré-multiplicando a equação (9.16) por  $\mathbf{L}^{-1}$  obtemos

$$\mathbf{L}^{-1}\mathbf{X}_t = \mathbf{L}^{-1}\boldsymbol{\Phi}_0 + \mathbf{L}^{-1}\boldsymbol{\Phi}_1\mathbf{X}_{t-1} + \dots + \mathbf{L}^{-1}\boldsymbol{\Phi}_p\mathbf{X}_{t-p} + \mathbf{b}_t. \quad (D.1)$$

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

$$X_{nt} + \sum_{i=1}^{n-1} a_{ni} X_{it} = \phi_{n,0}^* + \sum_{i=1}^n \boldsymbol{\Phi}_{ni}^{(1)*} X_{i,t-1} + \dots + \sum_{i=1}^n \boldsymbol{\Phi}_{ni}^{(p)*} X_{i,t-p} + b_{nt}. \quad (D.2)$$

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

$$\boldsymbol{\Sigma} = \begin{bmatrix} 2,25 & 0,3 \\ 0,3 & 0,64 \end{bmatrix}.$$

É fácil ver que a matriz  $\mathbf{L}$ , triangular inferior, necessária para diagonalizar  $\boldsymbol{\Sigma}$ , e a matriz  $\boldsymbol{\Lambda}$  são, respectivamente,

$$\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0,13 & 1 \end{bmatrix}, \quad \boldsymbol{\Lambda} = \begin{bmatrix} 2,25 & 0 \\ 0 & 0,60 \end{bmatrix}.$$

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

$$\boldsymbol{\Sigma} = \begin{bmatrix} 0,64 & 0,3 \\ 0,3 & 2,25 \end{bmatrix},$$

obtemos que  $\mathbf{L} = \begin{bmatrix} 1 & 0 \\ 0,47 & 1 \end{bmatrix}$  e  $\boldsymbol{\Lambda} = \begin{bmatrix} 0,64 & 0 \\ 0 & 2,11 \end{bmatrix}$ , 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 (9.57)-(9.58) por  $\begin{bmatrix} \mathbf{I}_r & -\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ -\boldsymbol{\Sigma}'_{12}\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{I}_s \end{bmatrix}$  obtemos

$$\begin{aligned}\mathbf{Y}_t &= \boldsymbol{\mu}_1 + \sum_{i=1}^p \boldsymbol{\Phi}_{11,i}^* \mathbf{Y}_{t-i} + \sum_{i=0}^p \boldsymbol{\Phi}_{12,i}^* \mathbf{Z}_{t-i} + \mathbf{a}_{1t}^*, \\ \mathbf{Z}_t &= \boldsymbol{\mu}_2 + \sum_{i=1}^p \boldsymbol{\Phi}_{21,i}^* \mathbf{Y}_{t-i} + \sum_{i=0}^p \boldsymbol{\Phi}_{22,i}^* \mathbf{Z}_{t-i} + \mathbf{a}_{2t}^*,\end{aligned}$$

onde  $\mathbf{a}_{1t}^* = \mathbf{a}_{1t} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \mathbf{a}_{2t}$  e  $\mathbf{a}_{2t}^* = \mathbf{a}_{2t} - \boldsymbol{\Sigma}_{12}^* \boldsymbol{\Sigma}_{11}^{-1} \mathbf{a}_{1t}$ .

Chamando  $\boldsymbol{\Sigma}_i^* = E[\mathbf{a}_{it}^* (\mathbf{a}_{it}^*)']$ , temos a seguinte proposição.

**Proposição D.1.** Há causalidade instantânea entre  $\mathbf{Y}_t$  e  $\mathbf{Z}_t$  se, e somente se,  $\boldsymbol{\Phi}_{12,0}^* \neq \mathbf{0}$  e  $\boldsymbol{\Phi}_{21,0}^* \neq \mathbf{0}$  ou, ainda, se, e somente se,  $|\boldsymbol{\Sigma}_{11}| > |\boldsymbol{\Sigma}_1^*|$  e  $|\boldsymbol{\Sigma}_{22}| > |\boldsymbol{\Sigma}_2^*|$ .

O teste de não causalidade instantânea pode ser conduzido como nos casos vistos na seção 9.7 para não causalidade (hipóteses  $H_{01}$  e  $H_{02}$ ), estimando-se as matrizes  $\boldsymbol{\Sigma}_i^*$  com os resíduos de MQ das regressões apresentadas aqui..



# Processos Cointegrados

## 10.1 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 9, 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. Esse é o problema da *regressão espúria*, tratado por Granger e Newbold (1974). Esses autores verificaram, por meio 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 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 *cointegrados*.

Preços e taxas (de câmbio, de juros), em geral, são  $I(1)$  e é usual analisar os logaritmos dessas séries, para investigar cointegração. 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. Esse 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  $I(0)$ . Nesse caso, dizemos que  $X_t$  e  $Y_t$  são cointegrados. 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  $d = \max(d_1, \dots, d_n)$ . Todavia, na definição a seguir, exigimos que todas as séries tenham a mesma ordem de integração.

**Definição 10.1.** As componentes do vetor  $\mathbf{X}_t$  são *cointegradas de ordem*  $(d, b)$ , e escrevemos,  $\mathbf{X}_t \sim C.I.(d, b)$ , se:

- (a) todas as componentes de  $\mathbf{X}_t$  são  $I(d)$ ;
- (b) existe um vetor  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)'$ , não nulo, tal que

$$\boldsymbol{\beta}' \mathbf{X}_t = \beta_1 X_{1t} + \dots + \beta_n X_{nt} \sim I(d - b), \quad d \geq b > 0. \quad (10.1)$$

O vetor  $\boldsymbol{\beta}$ , de ordem  $n \times 1$ , é chamado *vetor cointegrado* (ou vetor de cointegração).

**Observações:** (i) O vetor de cointegração  $\boldsymbol{\beta}$  não é único, pois se  $\lambda \neq 0$ , então  $\lambda\boldsymbol{\beta}$  é também um vetor de cointegração. Segue-se que alguma normalização é necessária para identificar  $\boldsymbol{\beta}$  univocamente. Uma normalização típica é  $\boldsymbol{\beta} = (1, -\beta_2, \dots, -\beta_n)'$ , de modo que

$$\boldsymbol{\beta}' \mathbf{X}_t = X_{1t} - \beta_2 X_{2t} - \dots - \beta_n X_{nt}.$$

Por exemplo, se  $\boldsymbol{\beta}' \mathbf{X}_t \sim I(0)$ , 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 cointegração*. 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 cointegradas. Veja o Problema 2.

(iii) Se  $\mathbf{X}_t$  tiver  $n > 2$  componentes, podem existir vários vetores de cointegração. Se existirem exatamente  $r$  vetores de cointegração 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 cointegração*. Nesse caso,

$$\mathbf{B}' \mathbf{X}_t = \begin{bmatrix} \boldsymbol{\beta}_1' \mathbf{X}_t \\ \vdots \\ \boldsymbol{\beta}_r' \mathbf{X}_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{rt} \end{bmatrix}$$

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 cointegrado.

**Exemplo 10.1.** Vejamos alguns exemplos de cointegração em economia.

- (a) Se  $C_t$  e  $Y_t$  indicarem, respectivamente, o consumo e a renda privada de um conjunto de indivíduos, o modelo da renda permanente implica que consumo e renda são cointegrados, ou seja,  $C_t = \alpha + \beta Y_t + e_t$ .
- (b) Modelos de demanda por moeda implicam relação de cointegração entre moeda, renda, preços e taxas de juros. Veja o Problema 8.
- (c) Paridade do poder de compra (PPC) implica cointegração entre a taxa de juros nominal de preços domésticos e internacionais. Seja  $P_t$  o índice de preços do país local,  $P_t^*$  o índice de preços do país estrangeiro e  $S_t$  a taxa de câmbio em moeda do país por unidade da moeda estrangeira. A teoria da PPC afirma que

$$P_t = S_t \cdot P_t^*,$$

ou, tomando o logaritmo,

$$p_t = s_t + p_t^*.$$

Em decorrência de uma variedade de fatores que limitam a arbitragem de preços internacionais, costuma-se trabalhar com uma versão mais fraca da PPC, ou seja,

$$s_t = p_t - p_t^* + u_t,$$

sendo  $u_t$  um processo estacionário. Se todas as séries forem  $I(1)$ , elas serão cointegradas.

**Exemplo 10.2** (Engle e Granger, 1987). Considere  $n = 2$  e as séries  $X_{1t}$  e  $X_{2t}$ , dadas por

$$X_{1t} + \beta X_{2t} = u_t, \tag{10.2}$$

$$u_t = \phi_1 u_{t-1} + \varepsilon_{1t}, \tag{10.3}$$

$$X_{1t} + \alpha X_{2t} = v_t, \tag{10.4}$$

$$v_t = \phi_2 v_{t-1} + \varepsilon_{2t}, \quad (10.5)$$

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)  $|\phi_i| < 1, i = 1, 2$ .

Nesse 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 cointegrado. A equação (10.4) é identificada.

(iii)  $|\phi_1| < 1, \phi_2 = 1$ .

Similar ao anterior, o vetor cointegrado é  $(1, \beta)'$  e a equação (10.2) é identificada.

## 10.2 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 cointegração entre dois ou mais log-preços, um modelo multivariado pode dar informação sobre o equilíbrio de longo prazo entre as séries.

Preços cointegrados 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 10.3.** Suponha que

$$X_{1t} = \mu_{1t} + \varepsilon_{1t}, \quad (10.6)$$

$$X_{2t} = \mu_{2t} + \varepsilon_{2t}, \quad (10.7)$$

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 cointegrado, com  $r$  vetores de cointegração,  $0 < r < n$ , então existirão  $n - r$  tendências estocásticas comuns.

O fato de duas séries serem cointegradas não significa que elas apresentem alta correlação. O seguinte exemplo mostra esse fato (Alexander, 2001).

**Exemplo 10.4.** Suponha

$$X_{1t} = u_t + \varepsilon_{1t}, \quad (10.8)$$

$$X_{2t} = u_t + \varepsilon_{2t}, \quad (10.9)$$

$$u_t = u_{t-1} + \varepsilon_t, \quad (10.10)$$

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 cointegradas 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

$$\rho(\Delta X_{1t}, \Delta X_{2t}) = \frac{\sigma^2}{\sqrt{(\sigma^2 + 2\sigma_1^2)(\sigma^2 + 2\sigma_2^2)}}.$$

Essa correlação será pequena se  $\sigma^2 \ll \sigma_1^2$  ou  $\sigma^2 \ll \sigma_2^2$ .

**Exemplo 10.5.** Considere as séries

$$X_{1t} = \beta_2 X_{2t} + u_t, \quad (10.11)$$

$$X_{2t} = X_{2,t-1} + v_t, \quad (10.12)$$

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 (10.11) representa a relação de equilíbrio de longo prazo. O vetor de cointegração é  $\beta = (1, -\beta_2)'$ . Na Figura 10.1 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 cointegrados. As equações (10.11) e (10.12) constituem a *representação triangular* de Phillips (1991a).

### 10.3 Modelo de Correção de Erros

Nesta seção seguiremos de perto os desenvolvimentos em Lütkepohl (1991) e em Hendry e Juselius (2000, 2001). Vimos, na Seção 10.1, que se duas ou mais séries são cointegradas, 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,

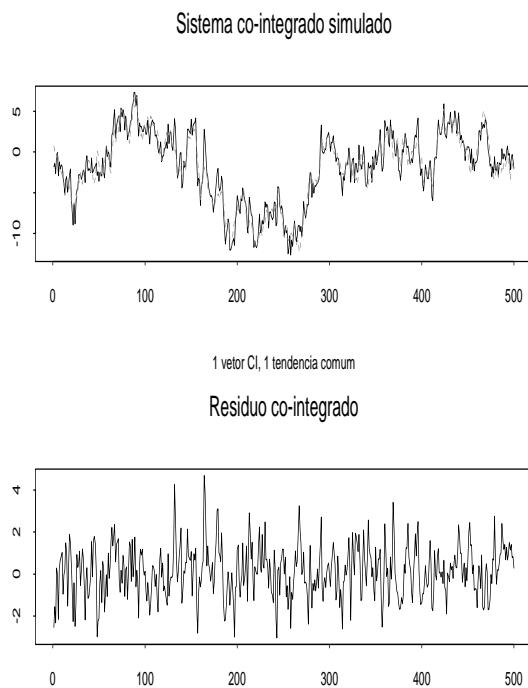


Figura 10.1: Sistema bivariado cointegrado:  $X_{1t}$  (linha cheia) e  $X_{2t}$  (linha tracejada).

$$\Delta P_{1t} = \alpha_1(P_{1,t-1} - \beta P_{2,t-1}) + a_{1t}, \quad (10.13)$$

e uma relação similar valha para  $P_{2t}$ :

$$\Delta P_{2t} = \alpha_2(P_{1,t-1} - \beta P_{2,t-1}) + a_{2t}. \quad (10.14)$$

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 cointegração 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

$$\Delta X_{1t} = \alpha_1(X_{1,t-1} - \beta X_{2,t-1}) + a_{11}(1)\Delta X_{1,t-1} + a_{12}(1)\Delta X_{2,t-1} + a_{1t}, \quad (10.15)$$

$$\Delta X_{2t} = \alpha_2(X_{1,t-1} - \beta X_{2,t-1}) + a_{21}(1)\Delta X_{1,t-1} + a_{22}(1)\Delta X_{2,t-1} + a_{2t}. \quad (10.16)$$

Esse 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 (10.15)-(10.16) como

$$\Delta \mathbf{X}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{A} \Delta \mathbf{X}_{t-1} + \mathbf{a}_t, \quad (10.17)$$

com

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 1 \\ -\beta \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} a_{11}(1) & a_{12}(1) \\ a_{21}(1) & a_{22}(1) \end{bmatrix}.$$

Vemos que (10.17) também pode ser escrita

$$\mathbf{X}_t - \mathbf{X}_{t-1} = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{A}(\mathbf{X}_{t-1} - \mathbf{X}_{t-2}) + \mathbf{a}_t,$$

ou

$$\mathbf{X}_t = (\mathbf{I}_n + \mathbf{A} + \boldsymbol{\alpha} \boldsymbol{\beta}') \mathbf{X}_{t-1} - \mathbf{A} \mathbf{X}_{t-2} + \mathbf{a}_t, \quad (10.18)$$

logo séries que são cointegradas podem ser geradas por um processo VAR.

Considere, agora, um modelo VAR(1)  $n$ -dimensional,

$$\mathbf{X}_t = \boldsymbol{\Phi} \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (10.19)$$

Pela Proposição 9.2, o processo  $\mathbf{X}_t$  será estacionário se todas as soluções de  $|\mathbf{I}_n - \boldsymbol{\Phi}| = 0$  estiverem fora do círculo unitário. Suponha que o processo seja não estacionário, com uma ou mais raízes sobre o círculo unitário. Isso é equivalente a dizer que um ou mais autovalores de  $\boldsymbol{\Phi}$  são iguais a um, os demais estando dentro do círculo unitário.

Como  $|\mathbf{I}_n - \boldsymbol{\Phi}| = 0$ , a matriz  $\boldsymbol{\Pi} = \mathbf{I}_n - \boldsymbol{\Phi}$  é singular. Suponha que o seu posto seja  $\rho(\boldsymbol{\Pi}) = r < n$ , de modo que  $\boldsymbol{\Pi}$  pode ser decomposta como  $\boldsymbol{\Pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$ , onde  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$  têm ordem  $n \times r$  e posto  $r$ .

Suponha que as componentes de  $\mathbf{X}_t$  sejam todas I(1). Então, de (10.19),

$$\mathbf{X}_t - \mathbf{X}_{t-1} = -(\mathbf{I}_n - \boldsymbol{\Phi})\mathbf{X}_{t-1} + \mathbf{a}_t,$$

ou

$$\Delta \mathbf{X}_t = -\boldsymbol{\Pi} \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (10.20)$$

Portanto, como

$$\boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} = -\Delta \mathbf{X}_t + \mathbf{a}_t,$$

o segundo termo é I(0), logo  $\boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1}$  é I(0) e continua a ser I(0) se o multiplicarmos por  $(\boldsymbol{\alpha}' \boldsymbol{\alpha})^{-1} \boldsymbol{\alpha}'$ , resultando  $\boldsymbol{\beta}' \mathbf{X}_{t-1} \sim I(0)$  e, finalmente,  $\boldsymbol{\beta}' \mathbf{X}_t \sim I(0)$ . Segue-se que cada linha de  $\boldsymbol{\beta}' \mathbf{X}_t$  representará uma relação de cointegração.

Conclui-se que, a partir de um VAR(1)  $n$ -dimensional, obtemos um modelo nas primeiras diferenças com variáveis cointegradas.

É fácil ver que, para um VAR(2)

$$\mathbf{X}_t = \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{X}_{t-2} + \mathbf{a}_t, \quad (10.21)$$

obtemos

$$\Delta \mathbf{X}_t = \mathbf{D}_1 \Delta \mathbf{X}_{t-1} - \boldsymbol{\Pi} \mathbf{X}_{t-2} + \mathbf{a}_t, \quad (10.22)$$

com  $\boldsymbol{\Pi} = \mathbf{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2$  e  $\mathbf{D}_1 = -(\mathbf{I}_n - \boldsymbol{\Phi}_1)$ . Veja o Problema 4.

Outra maneira de escrever (10.21) é

$$\Delta \mathbf{X}_t = \mathbf{F}_1 \Delta \mathbf{X}_{t-1} - \boldsymbol{\Pi} \mathbf{X}_{t-1} + \mathbf{a}_t, \quad (10.23)$$

onde  $\mathbf{F}_1 = -\boldsymbol{\Phi}_2$ ,  $\boldsymbol{\Pi}$  como antes. Essa forma é a chamada *forma de correção de equilíbrio ou de correção de erros*. Em (10.23) a matriz de níveis defasados,  $\boldsymbol{\Pi}$ , está no instante  $t-1$ , mas pode ser escolhida estar em qualquer “lag”, sem mudar a verossimilhança, como em (10.22). A forma (10.23) 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, tomado-se uma diferença em (10.19), obtemos

$$\Delta \mathbf{X}_t = \boldsymbol{\Phi} \Delta \mathbf{X}_{t-1} + \Delta \mathbf{a}_t,$$

que tem uma parte de médias móveis não invertível; logo, não obtemos uma representação VAR para as primeiras diferenças. O mesmo ocorre com (10.21).

Considere, agora, um modelo VAR( $p$ ) genérico,

$$\mathbf{X}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{X}_{t-1} + \dots + \boldsymbol{\Phi}_p \mathbf{X}_{t-p} + \mathbf{a}_t. \quad (10.24)$$

Este diz-se *cointegrado de posto r* se  $\boldsymbol{\Pi} = \mathbf{I}_n - \boldsymbol{\Phi}_1 - \dots - \boldsymbol{\Phi}_p$  tiver posto  $r$  e, portanto, puder ser escrita como  $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ ,  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$  de ordem  $n \times r$  e posto  $r$ . Dizemos que  $\boldsymbol{\beta}$  é a *matriz de cointegração* ou de *vetores cointegrados* e  $\boldsymbol{\alpha}$  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 (10.20) ou (10.22) fica

$$\Delta\mathbf{X}_t = \boldsymbol{\Phi}_0 + \mathbf{D}_1\Delta\mathbf{X}_{t-1} + \dots + \mathbf{D}_{p-1}\Delta\mathbf{X}_{t-p+1} - \boldsymbol{\Pi}\mathbf{X}_{t-p} + \mathbf{a}_t, \quad (10.25)$$

com  $\mathbf{D}_i = -(\mathbf{I}_n - \boldsymbol{\Phi}_1 - \dots - \boldsymbol{\Phi}_i)$ ,  $i = 1, 2, \dots, p - 1$ .

O processo (10.24) terá uma representação ou *modelo de correção de erros* (MCE)

$$\Delta\mathbf{X}_t = \boldsymbol{\Phi}_0 - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X}_{t-1} + \mathbf{F}_1\Delta\mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1}\Delta\mathbf{X}_{t-p+1} + \mathbf{a}_t, \quad (10.26)$$

onde  $\mathbf{F}_i = -(\boldsymbol{\Phi}_{i+1} + \dots + \boldsymbol{\Phi}_p)$ ,  $i = 1, 2, \dots, p - 1$ .

Como  $\Delta\mathbf{X}_t \sim I(0)$  e  $\boldsymbol{\beta}'\mathbf{X}_{t-1} \sim I(0)$ , esses termos têm média constante; sejam  $E(\Delta\mathbf{X}_t) = \mathbf{c}$ , um vetor  $n \times 1$ , representando taxas de crescimento, e  $E(\boldsymbol{\beta}'\mathbf{X}_{t-1}) = \boldsymbol{\mu}$ , um vetor  $r \times 1$ , representando interceptos nas relações de cointegração. Temos, então, que

$$\boldsymbol{\Phi}_0 = (\mathbf{I}_n - \mathbf{F}_1 - \dots - \mathbf{F}_{p-1})\mathbf{c} + \boldsymbol{\alpha}\boldsymbol{\mu}.$$

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 cointegração. Podemos, então, escrever (10.26) como

$$\Delta\mathbf{X}_t - \mathbf{c} = \sum_{i=1}^{p-1} \mathbf{F}_i(\Delta\mathbf{X}_{t-i} - \mathbf{c}) - \boldsymbol{\alpha}(\boldsymbol{\beta}'\mathbf{X}_{t-1} - \boldsymbol{\mu}) + \mathbf{a}_t, \quad (10.27)$$

e vemos que há duas formas de correção de equilíbrio em (10.27): uma, do crescimento dos dados em relação à sua média e, outra, dos vetores de cointegração 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( $p$ ) cointegrado, onde  $\boldsymbol{\Sigma}$  é a matriz de covariâncias de  $\mathbf{a}_t$ . Veja Lütkepohl (1991) para detalhes.

**Exemplo 10.6.** (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 cointegração entre preços de gasolina existiria se, por exemplo, diferenciais de preços entre dois locais quaisquer fossem estacionários. Considere o modelo VAR(1)

$$\Delta \mathbf{P}_t = \Phi_0 \mathbf{D}_t + \Phi_1 \Delta \mathbf{P}_{t-1} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{P}_{t-1} + \mathbf{a}_t,$$

onde  $\mathbf{P}_t = (P_{1t}, P_{2t}, P_{3t})'$ ,  $\mathbf{a}_t = (a_{1t}, a_{2t}, a_{3t})' \sim \text{i.i.d } \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ ,  $\Phi_1 = [\phi_{ij}]$ ,  $i, j = 1, 2, 3$ ,

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}.$$

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  $\alpha_{i,2}$ ;
- (d) choques aleatórios  $a_{it}$ .

Neste exemplo, teremos duas relações de cointegração, 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.

## 10.4 Testes para Cointegração

Para se concluir que duas ou mais séries são cointegradas, poderíamos pensar que bastaria analisar os seus gráficos. Todavia isso não é suficiente. Cointegração 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 cointegração 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 cointegração; esse caso foi tratado por Engle e Granger (1987);
- (b) há  $r$ ,  $0 \leq r < n$ , vetores de cointegração, caso considerado por Johansen (1988).

Além dessas referências, veja Zivot e Wang (2006), que também é uma referência para o uso do program S+FinMetrics.

### 10.4.1 Procedimento de Engle e Granger

Seja  $u_t = \beta' \mathbf{X}_t$  o resíduo de cointegração. O teste de Engle e Granger consiste em dois passos:

- (i) forme os resíduos de cointegração;
  - (ii) faça um teste de raízes unitárias para determinar se esses resíduos são  $I(0)$ .
- Temos as hipóteses:

$$\begin{aligned} H_0 &: u_t \sim I(1) : \text{não há cointegração}, \\ H_1 &: u_t \sim I(0) : \text{há cointegração}. \end{aligned} \quad (10.28)$$

Temos, ainda, dois casos a considerar:

- [1] O vetor de cointegração 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 10.7.** Consideremos as duas séries geradas no Exemplo 10.5, onde  $\beta = (1, -1)'$ . Usando a função `unitroot` do `S+FinMetrics`, obtemos o Quadro 10.1, teste ADF com dois lags e constante na regressão.

Quadro 10.1: Teste ADF para o Exemplo 10.7

```
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
```

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 cointegradas, com vetor de cointegração  $\beta = (1, -1)'$ .

[2] O vetor de cointegração 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

$$X_{1t} = \alpha + \beta'_2 \mathbf{X}_{2t} + u_t, \quad (10.29)$$

para obter os resíduos de MQO  $\hat{u}_t$  e testar  $H_0$  contra  $H_1$ . O teste ADF é feito sem termos determinísticos.

Formalmente, o teste ADF não é apropriado;  $\hat{u}_t$  é o resíduo de uma regressão na qual o vetor de cointegração é estimado e sob a hipótese de não cointegração, tal vetor não será identificado: usando os valores críticos de  $\hat{\tau}_\mu$ , rejeitaremos tal hipótese muito frequentemente. 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 cointegração são obtidas dependendo da escolha da variável resposta.

Phillips e Ouliaris (1990) desenvolveram testes de raiz unitária apropriados aos resíduos de (10.29); as estatísticas usadas não têm a distribuição assintótica usual ADF ou PP sob a hipótese  $H_0$  de não existência de cointegração. A distribuição assintótica é também uma função do movimento browniano padrão e depende dos termos determinísticos incluídos em (10.29) e do número de variáveis em  $\mathbf{X}_{2t}$ . Essas 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. A estatístico do teste ADF ou PP tem distribuição 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”. Nesse caso, considere a regressão

$$X_{1t} = \alpha_0 + \alpha_1 t + \beta'_2 \mathbf{X}_{2t} + u_t. \quad (10.30)$$

Use PO para os resíduos  $\hat{u}_t$  de (10.30), com constante, tendência e dimensão

$n - 1$ .

Para obter quantis e p-valores das distribuições de PO, use as funções `qcoint` e `pcoint` do `S+FinMetrics`.

**Exemplo 10.8.** 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 10.2 e 10.3, 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 cointegradas.

#### 10.4.2 Procedimento de Johansen

O procedimento de Yohansen é uma generalização multivariada do teste de DF. Considere o modelo (10.26) reescrito na forma

$$\Delta \mathbf{X}_t = \Phi_0 \mathbf{D}_t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t, \quad (10.31)$$

onde agora  $\boldsymbol{\Pi} = \Phi_1 + \dots + \Phi_p - \mathbf{I}_n$  e  $\mathbf{D}_t$  contém termos determinísticos (constantes, tendências etc.).

O procedimento de Johansen (1988, 1995) para testar a existência de cointegração 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 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 cointegração, que sabemos ser igual ao posto de  $\boldsymbol{\Pi}$ ;
- (iv) dados os vetores de cointegração (normalizados apropriadamente), estimar o MCE (via EMV).

Quadro 10.2: Teste ADF para o Exemplo 10.8

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 10.3: Teste PP para o Exemplo 10.8

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

---

Segundo Johansen (1994, 1995), os termos determinísticos em (10.31) são restritos à forma

$$\Phi_0 \mathbf{D}_t = \boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t. \quad (10.32)$$

Para verificarmos o efeitos dos termos determinísticos no modelo VAR, consideremos um caso especial:

$$\Delta \mathbf{X}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{a}_t, \quad (10.33)$$

onde  $\boldsymbol{\mu}_0$  e  $\boldsymbol{\mu}_1$  são ambos vetores  $n \times 1$ . Vamos decompor esses dois vetores em relação à média das relações de cointegração e em relação à média das taxas de crescimento,

$$\begin{aligned} \boldsymbol{\mu}_0 &= \boldsymbol{\alpha} \boldsymbol{\rho}_0 + \mathbf{c}_0, \\ \boldsymbol{\mu}_1 &= \boldsymbol{\alpha} \boldsymbol{\rho}_1 + \mathbf{c}_1. \end{aligned} \quad (10.34)$$

Então, podemos escrever

$$\begin{aligned} \Delta \mathbf{X}_t &= \boldsymbol{\alpha} \boldsymbol{\rho}_0 + \mathbf{c}_0 + \boldsymbol{\alpha} \boldsymbol{\rho}_1 t + \mathbf{c}_1 t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{a}_t \\ &= \boldsymbol{\alpha} (\boldsymbol{\rho}_0, \boldsymbol{\rho}_1, \boldsymbol{\beta}') \begin{pmatrix} 1 \\ t \\ \mathbf{X}_{t-1} \end{pmatrix} + (\mathbf{c}_0 + \mathbf{c}_1 t) + \mathbf{a}_t, \end{aligned}$$

ou, ainda,

$$\Delta \mathbf{X}_t = \boldsymbol{\alpha} \begin{pmatrix} \boldsymbol{\rho}_0' \\ \boldsymbol{\rho}_1' \\ \boldsymbol{\beta} \end{pmatrix} \mathbf{X}_{t-1}^* + (\mathbf{c}_0 + \mathbf{c}_1 t) + \mathbf{a}_t, \quad (10.35)$$

com  $\mathbf{X}_{t-1}^* = (1, t, \mathbf{X}_{t-1}^{'})'$ .

Podemos sempre escolher  $\boldsymbol{\rho}_0$  e  $\boldsymbol{\rho}_1$  tais que o erro de equilíbrio  $(\boldsymbol{\beta}^*)' \mathbf{X}_t^* = \mathbf{v}_t$  tenha média zero, logo

$$E(\Delta \mathbf{X}_t) = \mathbf{c}_0 + \mathbf{c}_1 t. \quad (10.36)$$

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,  $\boldsymbol{\mu}_t = \mathbf{0}$ ; neste caso,  $\boldsymbol{\rho}_0 = \boldsymbol{\rho}_1 = \mathbf{0}$  e o modelo não possui nenhuma componente determinística, com  $\mathbf{X}_t \sim I(1)$  sem *drift* (não há crescimento dos dados) e as relações de cointegração têm média zero. A menos que  $\mathbf{X}_0 = \mathbf{0}$ , este caso tem pouco interesse nas aplicações práticas.

Neste caso, o modelo (10.31) fica

$$\Delta \mathbf{X}_t = \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t.$$

Caso 2. constante restrita,  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 = \boldsymbol{\alpha} \boldsymbol{\rho}_0$ ; nesse caso,  $\boldsymbol{\rho}_1 = \mathbf{0}$ ,  $\mathbf{c}_0 = \mathbf{0}$ , mas  $\boldsymbol{\rho}_0 \neq \mathbf{0}$  e, portanto, não há tendência linear nos dados e as relações de cointegração têm média  $\boldsymbol{\rho}_0$ .

Podemos escrever o modelo na forma

$$\Delta \mathbf{X}_t = \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{X}_{t-1} + \rho_0) + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t.$$

Caso 3. constante irrestrita,  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0$ ; aqui,  $\boldsymbol{\rho}_1 = \mathbf{0}$ , as séries de  $\mathbf{X}_t$  são I(1) com drift e as relações de cointegração podem ter médias diferentes de zero.

O modelo será escrito como

$$\Delta \mathbf{X}_t = \mu_0 + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t,$$

Caso 4. tendência restrita,  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\alpha} \boldsymbol{\rho}_1 t$ ; neste caso,  $\mathbf{c}_1 = \mathbf{0}$ , mas  $\mathbf{c}_0, \boldsymbol{\rho}_0, \boldsymbol{\rho}_1$  são irrestritos. As séries são I(1) com drift e as relações de cointegração têm uma tendência linear.

Temos, para o modelo (10.31),

$$\Delta \mathbf{X}_t = \mu_0 + \boldsymbol{\alpha} (\boldsymbol{\beta}' \mathbf{X}_{t-1} + \rho_1 t) + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t.$$

Caso 5. constante e tendência irrestritas,  $\boldsymbol{\mu}_t = \boldsymbol{\mu}_0 + \boldsymbol{\mu}_1 t$ ; não há nenhuma restrição sobre  $\boldsymbol{\mu}_0$  e  $\boldsymbol{\mu}_1$ , as séries são I(1) com tendência linear (logo, tendência quadrática nos níveis) e as relações de cointegração têm tendência linear. Previsões podem ser ruins, logo deve-se ter certo cuidado em se adotar essa opção.

O modelo, nesse caso, é

$$\Delta \mathbf{X}_t = \mu_0 + \mu_1 t + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \dots + \mathbf{F}_{p-1} \Delta \mathbf{X}_{t-p+1} + \mathbf{a}_t.$$

Os casos mais relevantes para séries de taxas de juros, taxas de câmbio e preços de ativos são os Casos 2 e 3. Veja Hendry e Juselius (2001) e Zivot e Wang (2006) para detalhes.

Sabemos que o posto de  $\boldsymbol{\Pi}$  fornece também o número de autovalores não nulos de  $\boldsymbol{\Pi}$ ; suponha que os ordenemos  $\lambda_1 > \lambda_2 > \dots > \lambda_n$ . Se as séries são *não* cointegradas,  $\rho(\boldsymbol{\Pi}) = 0$  e todas os autovalores serão nulos ou, ainda,  $\ln(1 - \lambda_i) = 0$ , para todo  $i$ . Um teste da RV para testar o posto de  $\boldsymbol{\Pi}$  é baseado na *estatística traço*

$$\lambda_{\text{traço}}(r_0) = -T \sum_{i=r_0+1}^n \ln(1 - \hat{\lambda}_i), \quad (10.37)$$

onde  $\hat{\lambda}_i$  são os autovalores estimados de  $\boldsymbol{\Pi}$  e (10.37) testa

$$\begin{aligned} H_0 &: r \leq r_0, \\ H_1 &: r > r_0, \end{aligned} \quad (10.38)$$

sendo  $r$  o posto de  $\boldsymbol{\Pi}$ . Se  $\rho(\boldsymbol{\Pi}) = r_0$ , então  $\hat{\lambda}_{r_0+1}, \dots, \hat{\lambda}_n$  são aproximadamente nulas e a estatística (10.37) será pequena; caso contrário, será grande. Como dissemos aqui, a distribuição assintótica de (10.37) 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 (1)-(5) apresentados aqui e  $n - r_0 = 1, \dots, 10$ .

Johansen também usa a *estatística do máximo autovalor*

$$\lambda_{\max}(r_0) = -T \ell n(1 - \hat{\lambda}_{r_0+1}), \quad (10.39)$$

para testar

$$\begin{aligned} H_0 &: r = r_0, \\ H_1 &: r = r_0 + 1. \end{aligned} \quad (10.40)$$

A distribuição assintótica de (10.39) 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  $\boldsymbol{\Pi}$  é  $r$ , Johansen (1988) prova que o estimador de máxima verossimilhança de  $\boldsymbol{\beta}$  é dado por  $\hat{\boldsymbol{\beta}}_{\text{MV}} = (\hat{v}_1, \dots, \hat{v}_r)$ , onde  $\hat{v}_i$  é o autovetor associado ao autovalor  $\hat{\lambda}_i$  e os estimadores de máxima verossimilhança dos parâmetros restantes são obtidos por meio de uma regressão multivariada com  $\boldsymbol{\beta}$  substituído pelo EMV. Johansen (1995) mostra a normalidade assintótica dos estimadores de  $\boldsymbol{\beta}$ , com taxa de convergência  $T^{-1}$ . Veja também a Seção 10.5.

**Exemplo 10.9.** Considere  $T = 250$  valores dos sistema dado no Problema 6, sendo o vetor de cointegração  $\boldsymbol{\beta} = (1; -0,5; -0,5)'$ ,  $u_t$  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 10.2. 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. A função `coint` do `S+FinMetrics` será usada para fazer o teste.

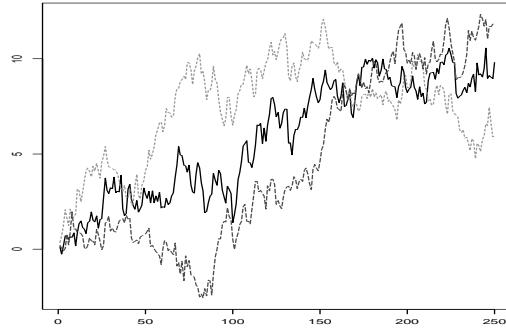


Figura 10.2: Sistema trivariado cointegrado:  $X_{1t}$  (linha cheia),  $X_{2t}$  (linha pontilhada) e  $X_{3t}$  (linha tracejada).

No Quadro 10.4, temos os valores das estatísticas  $\lambda_{\text{traço}} = 40,35$  e  $\lambda_{\max} = 27,86$ . Notamos que ambas são significativas com o nível 0,05 para testar a hipótese  $H_0$  de que não há cointegração contra a alternativa que há mais que uma (uma, respectivamente) relação de integração. Por outro lado, a hipótese nula de uma relação de cointegração contra a alternativa de mais que uma (duas, respectivamente) é aceita, com o nível 0,05, sendo  $\lambda_{\text{traço}} = 12,49$  e  $\lambda_{\max} = 9,14$ . Concluimos, então, que há somente um vetor de cointegração. O quadro apresenta, também, o vetor não-normalizado e o vetor normalizado, além do intercepto, supondo o Caso 2 apresentado. O vetor de cointegração estimado é  $\hat{\beta} = (1; -0,507; -0,553)'$ , sendo que o vetor verdadeiro tem  $\beta_2 = \beta_3 = 0,5$ .

No Quadro 10.5, temos o resultado da estimação do MCE. Estimadores de máxima verossimilhança são obtidos usando a função VECM do S+FinMetrics. O modelo completo, incluindo coeficientes não significativos, seria dado por

$$\begin{aligned}\Delta X_{1t} &= -0,030 - 0,189(X_{1,t-1} - 0,493X_{2,t-1} - 0,532X_{3,t-1} - 0,370) + a_{1t}, \\ \Delta X_{2t} &= 0,034 + 0,032(X_{1,t-1} - 0,493X_{2,t-1} - 0,532X_{3,t-1} - 0,370) + a_{2t}, \\ \Delta X_{3t} &= 0,049 + 0,007(X_{1,t-1} - 0,493X_{2,t-1} - 0,532X_{3,t-1} - 0,370) + a_{3t}.\end{aligned}$$

Todavia, vários coeficientes não são significativos e poderão ser eliminados do modelo, obtendo-se

$$\begin{aligned}\Delta X_{1t} &= -0,189(X_{1,t-1} - 0,493X_{2,t-1} - 0,532X_{3,t-1}) + a_{1t}, \\ \Delta X_{2t} &= a_{2t}, \\ \Delta X_{3t} &= a_{3t}.\end{aligned}$$

Vemos que somente  $X_{1t}$  responde ao erro de desequilíbrio defasado.

Na Figura 10.3, temos ilustrados os resíduos de cointegração, que parecem ser estacionários. Resultados similares são obtidos usando o Caso 3. Note que o Caso 2 contempla uma média para a relação de cointegração, que não é significativa no exemplo.

Quadro 10.4: Teste de Johansen para o Exemplo 10.9

```

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

```

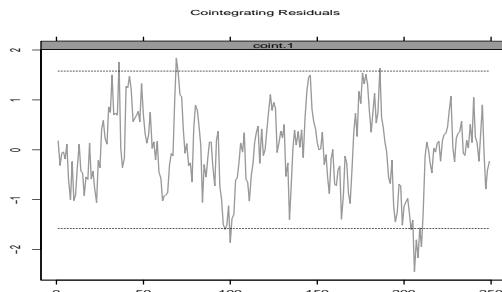


Figura 10.3: Resíduos de cointegração para o Exemplo 10.9.

**Exemplo 10.10.** Como um último exemplo, consideremos as séries do Ibovespa e dos preços das ações da Petrobras, no período de 1998 a 2010, com 2.999 observações diárias, mostradas na Figura 10.4. Usando-se o Caso 2 para o termo determinístico e a função `coint`, como no exemplo anterior, verifica-se que as séries não são cointegradas.

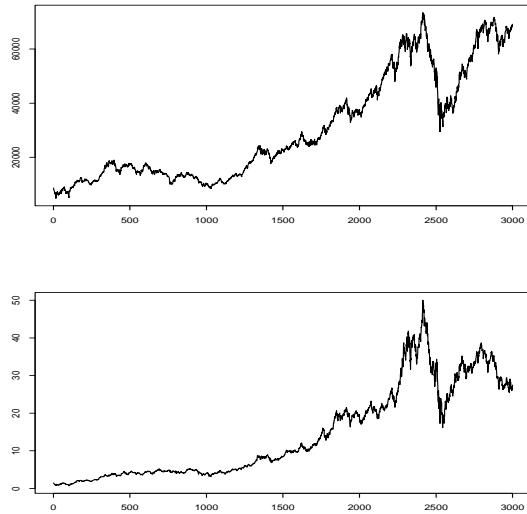


Figura 10.4: Índice Ibovespa e Petrobras.

## 10.5 Comentários Finais

Encerramos este capítulo com algumas observações sobre processos cointegrados, bem como uma apresentação um pouco mais detalhada sobre estimadores de máxima verossimilhança (EMV) do MCE.

### 10.5.1 Enfoque Bayesiano

Existem alguns trabalhos sobre o tratamento bayesiano de processos cointegrados, como Koop (1992), Tsurumi e Wago (1994) e Phillips (1994). Veja Lopes e Lima (1995) para mais detalhes.

### 10.5.2 Termos Determinísticos no MCE

Os Casos (1)-(5) considerados na Seção 10.4.2 são usualmente referidos como  $H_2(r)$ ,  $H_1^*(r)$ ,  $H_1(r)$ ,  $H^*(r)$  e  $H(r)$ , respectivamente. Essa nomenclatura é também adotada nos programas computacionais, veja o Quadro 10.4, por exemplo.

O MCE irrestrito é denotado  $H(r)$ , significando que  $\rho(\boldsymbol{\Pi}) \leq r$ . Obtemos, então, uma sequência de modelos hierárquicos  $H(0) \subset \cdots \subset H(r) \subset \cdots \subset H(n)$ , onde  $H(0)$  indica o modelo VAR não cointegrado, com  $\boldsymbol{\Pi} = \mathbf{0}$  e  $H(n)$  indica o modelo VAR(p) irrestrito estacionário.

### 10.5.3 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*,

$$\Delta \mathbf{X}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\delta} t + \boldsymbol{\Psi} \mathbf{d}_t + \mathbf{F}_1 \Delta \mathbf{X}_{t-1} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{X}_{t-1} + \mathbf{a}_t. \quad (10.41)$$

Vimos que, para que (10.41) seja internamente consistente,  $\boldsymbol{\beta}$  deve ter posto reduzido  $r < n$ . Vamos descrever brevemente o procedimento de máxima verossimilhança para estimar (10.41). Os detalhes estão em Johansen (1988, 1995) e Banerjee et al. (1993). A ideia 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 (10.42) fica

$$\mathbf{Z}_{0t} = \boldsymbol{\Theta} \mathbf{Z}_{2t} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{Z}_{1,t-1} + \mathbf{a}_t,$$

onde  $\boldsymbol{\Theta} = (\boldsymbol{\Phi}_0, \boldsymbol{\delta}, \boldsymbol{\Psi}, \mathbf{F}_1)'$ . Concentrando fora da verossimilhança os efeitos de ajustamento de curto prazo,  $\boldsymbol{\Theta} \mathbf{Z}_{2t}$ , 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

$$\mathbf{M}_{ij} = \frac{1}{T} \sum_t \mathbf{Z}_{1t} \mathbf{Z}_{jt}'$$

são matrizes de momentos, de modo que  $\hat{\mathbf{R}}_1' = \mathbf{M}_{02} \mathbf{M}_{22}^{-1}$  e  $\hat{\mathbf{R}}_2' = \mathbf{M}_{12} \mathbf{M}_{22}^{-1}$ .

Quadro 10.5: Estimação do MCE para o Exemplo 10.9			
Call:			
VECM(test = coint.rc)			
Cointegrating Vectors:			
coint.1 1.0000			
y2 -0.4933 (std.error) 0.0855 (t.stat) -5.7694			
y3 -0.5320 (std.error) 0.1016 (t.stat) -5.2363			
Trend* -0.3698 (std.error) 1.7663 (t.stat) -0.2094			
VECM Coefficients:			
y1           y2           y3 coint.1 -0.1886  0.0323  0.0065 (std.error) 0.0449  0.0411  0.0406 (t.stat) -4.2006  0.7867  0.1593			
Intercept -0.0300  0.0343  0.0487 (std.error) 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			

O modelo concentrado pode, então, ser escrito

$$\mathbf{u}_{1t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{u}_{2t} + \mathbf{u}_t, \quad (10.42)$$

que inclui somente relações de equilíbrio de longo prazo.

Os EMV são obtidos em dois estágios:

- (1) para obter um estimador de  $\boldsymbol{\alpha}$ , supomos  $\boldsymbol{\beta}$  conhecido de modo que  $\boldsymbol{\alpha}$  pode ser estimado por MQO em (10.42);
- (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). Essa solução fornece os  $n$  autovalores  $\hat{\lambda}_1 > \dots > \hat{\lambda}_n$ . Como vimos, o estimador de  $\boldsymbol{\beta}$  é dado pela matriz  $n \times r$  de autovetores, correspondentes aos  $r$  maiores autovalores  $\hat{\lambda}_i$ . Obtido o EMV de  $\boldsymbol{\beta}$ , calcule  $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}(\hat{\boldsymbol{\beta}})$ .

### 10.5.4 Cointegração Dinâmica

Observações empíricas mostram que relações de cointegração entre séries financeiras variam ao longo do tempo. Esse fato motivou autores como Hansen e Johansen (1999) e Quintos (1997) a proporem testes para verificarem a constância dos vetores de cointegração. Park e Hahn (1999) e Bierens e Martins (2010) consideram expansões dos vetores de cointegração, supostos variantes no tempo, em séries de Fourier e polinômios de Chebyshev, respectivamente. Finalmente, Fonseca (2017) considera expansões em ondaletas para o vetor cointegrado no modelo de correção de erros.

## 10.6 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 10.1.

- (a) Mostre que ambas as séries são  $I(2)$ .
- (b) Mostre que  $X_{1t}$ ,  $X_{2t}$  e  $\Delta X_{2t}$  são cointegradas. Qual é o vetor cointegrado?
- (c) Mostre que  $X_{1t}$  e  $X_{2t}$  são  $C.I.(2,1)$ .
4. Prove (10.22).
5. Mostre que (10.19) 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 cointegrado (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 cointegração é  $\beta = (1, -\beta_2, -\beta_3)'$ , 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 cointegração.

7. Simule o sistema cointegrado (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). Nesse caso, as duas primeiras equações descrevem relações de equilíbrio de L.P. e a terceira descreve a tendência estocástica comum. Há dois vetores de cointegração,  $\beta_1 = (1, 0, -\alpha_1)', \beta_2 = (1, 0, -\alpha_2)'$ , e  $u_t, v_t$  são os resíduos de cointegração.

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 cointegradas e obtenha o vetor de cointegração normalizado.  
 (b) Suponha que existe 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, nesse caso, existem dois vetores de cointegração. Especifique a matriz  $\mathbf{B}$ , de posto 2, que contém esses dois vetores.

9. Considere o processo linear  $\mathbf{Y}_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}$ , onde as matrizes  $\Psi_j$  decrescem para zero exponencialmente, de tal sorte que  $\Psi(z) = \sum_{j=0}^{\infty} \Psi_j z^j$  seja convergente. Dizemos que  $\mathbf{Y}_t$  é I(0) se  $\Psi(1) = \sum_{j=0}^{\infty} \Psi_j \neq 0$ . Um processo I(1) é aquele que se torna I(0) após uma diferença.

Seja  $\mathbf{X}_t = (X_{1t}, X_{2t}, X_{3t})'$ , com

$$\begin{aligned} X_{1t} &= \sum_{s=1}^t \varepsilon_{1s} + \varepsilon_{2t}, \\ X_{2t} &= \frac{1}{2} \sum_{s=1}^t \varepsilon_{1s} + \varepsilon_{3t}, \\ X_{3t} &= \varepsilon_{2t}. \end{aligned}$$

Aqui  $\varepsilon_{1t}, \varepsilon_{2t}, \varepsilon_{3t}$  são processos estacionários.

- (a) Prove que  $\mathbf{X}_t$  é I(1) (para isso, mostre que  $Y_t = \Delta \mathbf{X}_t$  é I(0), usando a definição acima).
- (b) Mostre que  $(1, -2, 0)'$  e  $(0, 0, 1)'$  são vetores de cointegração.
10. Use o teste de Engle e Granger para testar se as séries simuladas nos Problemas 6 e 7 são cointegradas.
11. Refaça o problema anterior com o teste de Johansen.
12. Verifique, usando o Exemplo 10.9 e o teste de Johansen, que as séries do Ibovespa e das ações da Petrobras não são cointegradas.
13. Use o teste de Johansen para testar se as séries:
- (a) diárias do Ibovespa, preços diárias de ações da Vale e preços diárias de ações da Petrobras são cointegradas, considerando o período de 31/8/1998 a 29/9/2010.
- (b) mensais do Ibovespa (m-ibv94.01.dat) e C-Bond são cointegradas, no período de junho de 1994 a agosto de 2001.
- (c) diárias dos índices Ibovespa e IPC são cointegradas, no período de 04/9/95 a 30/12/2004.
- (d) diárias da Petrobras3 e Petrobras4 são cointegradas, no período de 02/1/06 a 29/09/2010.



# Análise de Dependência e Cópulas

## 11.1 Introdução

Neste capítulo, iremos estudar como medir dependência entre variáveis aleatórias e séries temporais. No Capítulo 9, estudamos parte desse problema, por meio da função de covariância e do ajuste de modelos multivariados, em especial o modelo VAR. Usando esses modelos, introduzimos o conceito de Causalidade de Granger.

A função de covariância, assim como o caso particular do coeficiente de correlação, são funções de dependência globais. Além disso, o coeficiente de correlação é uma medida de dependência linear, apropriada se duas séries seguem uma distribuição normal bivariada (ou elíptica). Mas sabemos que séries financeiras, especialmente os retornos, não seguem uma distribuição normal, apresentam caudas pesadas e assimetria, logo é necessário encontrar outras medidas (não lineares) de dependência.

Veremos que cópulas são apropriadas para descrever dependência entre várias variáveis. A teoria de cópulas tem sido usada frequentemente para analisar séries financeiras e fatores de risco. Contudo, algum cuidado deve ser tomado aqui, pois praticamente toda a metodologia desenvolvida nesta área refere-se a dados provenientes de amostras de um conjunto de v.a.'s e nesse sentido teremos, em geral,  $n$ -plas independentes e identicamente distribuídas. E este não é o caso quando analisamos  $n$  variáveis financeiras.

Analisaremos, também, além das cópulas, algumas medidas de dependência local para v.a.'s e séries financeiras.

Suponha que temos duas v.a.'s  $X$  e  $Y$  e  $\delta(X, Y)$  seja uma medida de dependência entre elas. As seguintes propriedades são desejáveis para  $\delta$  (Embrechts et al., 2003):

- (i)  $\delta(X, Y) = \delta(Y, X);$

- (ii)  $-1 \leq \delta(X, Y) \leq 1$ ;
- (iii)  $\delta(X, Y) = 1$  se  $X$  e  $Y$  são comonotônicas e  $\delta(X, Y) = -1$  se  $X$  e  $Y$  são contramontotônicas;
- (iv) Se  $T$  for uma transformação monótona,

$$\delta(T(X), Y) = \begin{cases} \delta(X, Y), & \text{se } T \text{ crescente,} \\ -\delta(X, Y), & \text{se } T \text{ decrescente.} \end{cases}$$

- (v)  $\delta(X, Y) = 0$  se e somente se  $X$  e  $Y$  são independentes.

Dizemos que  $X$  e  $Y$  são comonotônicas se  $Y$  (ou  $X$ ) for uma função estritamente crescente de  $X$  (ou  $Y$ ) em quase toda parte e são contramontotônicas se a função for estritamente decrescente.

O coeficiente de correlação linear de Pearson,  $\rho_P$ , satisfaz (i)-(ii). Além disso,  $\rho_P$  requer que as variâncias de  $X$  e  $Y$  sejam finitas,  $\rho = 0$  não implica independência entre  $X$  e  $Y$ , a não ser que  $(X, Y)$  tenha uma distribuição normal bivariada. Também,  $\rho_P$  não é invariante sob transformações não lineares estritamente crescentes.

Outras medidas de dependência global entre duas variáveis  $X$  e  $Y$ , largamente usadas, são os coeficientes  $\tau_K$ , de Kendall, e o coeficiente  $\rho_S$ , de Spearman. Essas medidas sempre existem, são não paramétricas, baseadas em postos e não dependem das distribuições marginais de  $X$  e  $Y$ , mas somente da cópula  $C$  dessas duas variáveis. Diferente de  $\rho_P$ , essas medidas são invariantes por escala. Uma outra propriedade interessante é que, como dependem da cópula das duas variáveis,  $\tau_K$  e  $\rho_S$  podem ser usadas para ajustar uma cópula a dados, pois frequentemente os parâmetros de uma cópula são funções desses coeficientes (isto corresponde a encontrar estimadores pelo método dos momentos).

## 11.2 Medidas de Dependência

Nesta seção, estudaremos algumas medidas de dependência global entre duas variáveis. Começamos com o conhecido coeficiente de correlação.

### Coeficiente de Correlação Linear

Também chamado coeficiente correlação linear de Pearson, é definido por

$$\rho_P(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}, \quad (11.1)$$

onde  $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$  é a covariância entre  $X$  e  $Y$  e  $\sigma_X$ ,  $\sigma_Y$  indicam os desvios padrões de  $X$  e  $Y$ , respectivamente. Notamos que:

- (i)  $-1 \leq \rho_P \leq 1$ ;
- (ii)  $\rho_P = 0$  quando  $X$  e  $Y$  são independentes;
- (iii)  $\rho_P = \pm 1$  quando  $Y$  é uma função linear de  $X$ .

Um estimador de momentos de  $\rho_P$  é obtido estimando-se a covariância por

$$\widehat{\text{Cov}(X, Y)} = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y}) = \frac{1}{n} \sum_{t=1}^n x_t y_t - n\bar{x}\bar{y},$$

dada a amostra  $(x_1, y_1), \dots, (x_n, y_n)$  de  $(X, Y)$ , e onde  $\bar{x}$ ,  $\bar{y}$  denotam as médias amostrais de  $X$  e  $Y$ , respectivamente. O desvio padrão de  $X$  é estimado por  $\hat{\sigma}_X$ , onde

$$\hat{\sigma}_X^2 = \frac{1}{n} \sum_{t=1}^n (x_t - \bar{x})^2 = \frac{1}{n} \sum_{t=1}^n x_t^2 - \bar{x}^2$$

é o estimador da variância de  $X$ , expressão análoga para  $\hat{\sigma}_Y^2$ . Segue-se que um estimador para  $\rho_P$  é dado por

$$\hat{\rho}_P = \frac{\sum_{t=1}^n x_t y_t - n\bar{x}\bar{y}}{\left[ \left( \sum_{t=1}^n x_t^2 - n\bar{x}^2 \right) \left( \sum_{t=1}^n y_t^2 - n\bar{y}^2 \right) \right]^{1/2}}. \quad (11.2)$$

É um fato que  $-1 \leq \hat{\rho}_P \leq 1$  e, na prática, se o valor estimado estiver próximo de  $-1$  ou  $+1$ , as variáveis são fortemente associadas ou correlacionadas linearmente e se estiver próximo de zero, as variáveis são não correlacionadas. Se  $(X, Y)$  for normal bivariada, podemos falar em dependentes e independentes, respectivamente.

Convém reafirmar que  $\rho_P$  mede dependência linear entre  $X$  e  $Y$ , e não outro tipo de dependência. De fato, suponha que uma das variáveis possa ser expressa linearmente em termos da outra, por exemplo  $X = aY + b$ , e seja  $d = E(|X - aY - b|^2)$ . Então, pode-se provar (veja Problema 3) que o mínimo de  $d$  ocorre quando

$$a = \frac{\sigma_X}{\sigma_Y} \rho_P, \quad b = E(X) - aE(Y), \quad (11.3)$$

e o mínimo é dado por

$$\min d = \sigma_X^2 (1 - \rho_P^2). \quad (11.4)$$

Nota-se, portanto, que quanto maior o valor absoluto do coeficiente de correlação entre  $X$  e  $Y$ , melhor a acurácia com que uma das variáveis pode ser representada como uma combinação linear da outra. Obviamente, este mínimo anula-se se e somente se  $\rho_P = 1$  ou  $\rho_P = -1$ .

Segue-se imediatamente de (11.4) que

$$\rho_P(X, Y) = \frac{\sigma_X^2 - \min_{a,b} E(|X - aY - b|^2)}{\sigma_X^2}, \quad (11.5)$$

ou seja,  $\rho_P$  mede a redução relativa na variância de  $X$  por meio de uma regressão linear de  $X$  sobre  $Y$ .

Outro fato importante é que, se conhecermos somente as distribuições marginais de  $X$  e  $Y$ , mas não a relação de dependência entre as variáveis, então  $\rho_P$  pode variar num intervalo  $[\rho_1, \rho_2]$ , que pode ser muito menor do que  $[-1, 1]$ . Veja Carmona (2004) para detalhes.

### Coeficiente $\tau_K$ de Kendall

Suponha que a v.a.  $(X, Y)$  tenha distribuição  $F$ , contínua. As v.a.'s  $X$  e  $Y$  são concordantes se pequenos valores de uma estão associados a pequenos valores da outra, o mesmo ocorrendo com os valores grandes.

Formalmente, sejam  $(X_1, Y_1), (X_2, Y_2)$  dois pares independentes obtidos de  $F$ . Dizemos que estes pares são *concordantes* se  $X_1 > X_2$  sempre que  $Y_1 > Y_2$  e  $X_1 < X_2$  sempre que  $Y_1 < Y_2$  (isto é, se  $(X_1 - X_2)(Y_1 - Y_2) > 0$ ). Os pares são *discordantes* se ocorre  $X_1 > X_2$  sempre que  $Y_1 < Y_2$  e  $X_1 < X_2$  sempre que  $Y_1 > Y_2$  (ou seja, se  $(X_1 - X_2)(Y_1 - Y_2) < 0$ ).

Geometricamente, dois pontos distintos  $(X_1, Y_1), (X_2, Y_2)$  no plano são concordantes, se o segmento de reta que os une tem inclinação positiva, e discordantes, se o segmento tem inclinação negativa.

O *coeficiente  $\tau_K$  de Kendall* é definido por

$$\tau_K = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}. \quad (11.6)$$

Uma estimativa de  $\tau_K$ , baseada numa amostra de tamanho  $n$  de  $(X, Y)$ , é dada por

$$\hat{\tau}_K = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \text{sinal}((x_i - x_j)(y_i - y_j)), \quad (11.7)$$

ou seja, o quociente entre o número de pares concordantes menos o número de pares discordantes e o número total de pares. Nesta fórmula,  $\text{sinal}(x) = 1$ , se  $x > 0$ , 0 se  $x = 0$ , e  $-1$ , se  $x < 0$ .

Seja  $R_i$  o posto de  $X_i$  dentre  $X_1, \dots, X_n$  e  $S_i$  o posto de  $Y_i$  dentre  $Y_1, \dots, Y_n$ . Então,  $(X_i - X_j)(Y_i - Y_j) > 0$  se e somente se  $(R_i - R_j)(S_i - S_j) > 0$ , ou seja  $\hat{\tau}_K$  é uma função somente dos postos das observações.

Em aplicações financeiras, em que  $X$  e  $Y$  representam, por exemplo, retornos financeiros, o interesse está nos casos em que  $X$  e  $Y$  tomam valores muito grandes

ou muito pequenos ao mesmo tempo. Nestes casos,  $\tau_K$  fornece uma medida de concordância entre  $X$  e  $Y$ .

Outra medida de concordância é o coeficiente de Spearman.

#### Coeficiente $\rho_S$ de Spearman

Suponha que  $(X, Y)$  tenha distribuição  $F$  e que as distribuições marginais de  $X$  e  $Y$  sejam  $F_1$  e  $F_2$ , respectivamente. O *coeficiente  $\rho_S$  de Spearman* é também baseado em concordância e discordância. Suponha que  $(X_1, Y_1), (X_2, Y_2)$  e  $(X_3, Y_3)$  sejam três cópias independentes de  $(X, Y)$ . Então (Kruskal, 1958),

$$\rho_S = 3(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]). \quad (11.8)$$

Ou seja,  $\rho_S$  é definido com sendo proporcional à probabilidade de concordância menos a probabilidade de discordância dos dois vetores  $(X_1, Y_1)$  e  $(X_2, Y_3)$ . Note também que a f.d. de  $(X_1, Y_1)$  é  $F$ , enquanto a f.d. de  $(X_2, Y_3)$  é  $F_1(x)F_2(y)$ , pois  $X_2$  e  $Y_3$  são independentes.

Pode-se provar que

$$\rho_S = \rho_P(F_1(X), F_2(Y)), \quad (11.9)$$

ou seja, o coeficiente  $\rho_S$  é uma medida de correlação de postos. Veja o Problema 5 e a seção seguinte. Uma estimativa de  $\rho_S$ , baseada numa amostra de tamanho  $n$  de  $(X, Y)$ , é dada por

$$\hat{\rho}_S = \frac{12}{n(n^2 - 1)} \sum_{i=1}^n \left( R_i - \frac{n+1}{2} \right) \left( S_i - \frac{n+1}{2} \right), \quad (11.10)$$

onde  $R_i$  e  $S_i$  são os postos de  $X_i$  e  $y_i$  previamente definidos. Devido a (11.9), podemos também estimar  $\rho_S$  por

$$\hat{\rho}_S = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}}.$$

Pode-se definir uma *função de concordância*, que generaliza a definição de  $\tau_K$ . Veja Nelsen (2003) e o Problema 5.

#### Medidas de dependência de caudas

Estas medidas captam a dependência conjunta das caudas de distribuições bivariadas. Considere a v.a.  $(X, Y)$  e seja  $F_i^{-1}(q)$  o  $q$ -quantil de  $X_i$ ,  $i = 1, 2$ .

A medida de dependência de cauda superior,  $\lambda_u$ , é dada pela probabilidade de  $Y$  estar acima de um quantil alto, dado que  $X$  está acima de um quantil alto:

$$\lambda_u(X, Y) = \lim_{q \rightarrow 1, q < 1} P(Y > F_2^{-1}(q) | X > F_1^{-1}(q)). \quad (11.11)$$

De modo análogo, definimos a medida de cauda inferior,  $\lambda_\ell$ ,

$$\lambda_\ell(X, Y) = \lim_{q \rightarrow 0, q > 0} P(Y \leq F_2^{-1}(q) | X \leq F_1^{-1}(q)). \quad (11.12)$$

De modo geral,  $0 < \lambda_u \leq 1$ ,  $0 < \lambda_\ell \leq 1$ , sendo que  $\lambda_u = 0$  implica independência (assintótica) na cauda superior e  $\lambda_\ell = 0$  implica independência (assintótica) na cauda inferior.

### Dependência por quadrantes

Lehmann (1966) introduziu o conceito de dependência por quadrantes. Dizemos que as v.a.'s contínuas  $X$  e  $Y$  são *dependentes positivamente por quadrante* (*positively quadrant dependent*, PQD) se, para todo par  $(x, y)$  de  $\mathbb{R}^2$ , tivermos

$$P(X \leq x, Y \leq y) \geq P(X \leq x)P(Y \leq y). \quad (11.13)$$

Ou seja, a probabilidade de  $X$  e  $Y$  serem simultaneamente pequenas é pelo menos tão grande quanto no caso em que as v.a.'s são independentes.

A dependência negativa por quadrante (NQD) é definida de modo análogo, invertendo o sentido da desigualdade em (11.13).

## 11.3 Cópulas

A teoria de cópulas teve um desenvolvimento explosivo nos últimos anos, notadamente nas aplicações em estatística, finanças, gestão de risco e modelagem de dependências em carteiras de seguros.

De forma simples, uma cópula é uma função ligando distribuições conjuntas e suas marginais. O uso de cópulas acessa a distribuição multivariada por meio da modelagem de cada distribuição marginal e da estimação da cópula.

Por simplicidade, o desenvolvimento será feito para o caso de duas variáveis.

**Definição 11.1.** Uma *cópula* é uma função  $C : I^2 \rightarrow I$ , com  $I = [0, 1]$ , satisfazendo, para  $0 < u_1 \leq u_2 < 1$ ,  $0 < v_1 \leq v_2 < 1$ ,  $(u_1, v_1)$  e  $(u_2, v_2)$  em  $I^2$ , as condições:

- (i)  $C(u, 0) = C(0, v) = 0$ ,  $C(u, 1) = u$ ,  $C(1, v) = v$ ;
- (ii)  $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ .

As relações (i) significam uniformidade das margens, enquanto (ii) significa que, se  $(X, Y)$  tem f.d.  $C$ , então  $P(u_1 \leq X \leq u_2, v_1 \leq Y \leq v_2) \geq 0$ .

Um resultado fundamental nesta teoria é devido a Sklar (1959).

**Teorema 11.1** (a) Seja  $C$  uma cópula e  $F_1, F_2$  f.d.'s univariadas. Então,

$$F(x, y) = C(F_1(x), F_2(y)), \quad (x, y) \in \mathbb{R}^2, \quad (11.14)$$

define uma f.d. com marginais  $F_1$  e  $F_2$ .

(b) Reciprocamente, para uma f.d. bivariada  $F$ , com marginais  $F_1$  e  $F_2$ , existe uma cópula  $C$  satisfazendo (11.14) e esta é única se  $F_1$  e  $F_2$  são contínuas e, então, para todo  $(u, v) \in I^2$ , temos

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), \quad (11.15)$$

$F_i^{-1}$  denota a inversa generalizada de  $F_i$ ,  $i = 1, 2$ , isto é,  $F_i^{-1}(u) = \sup_z \{F_i(z) \leq u\}$ .

De modo conciso, cópulas são f.d.'s bivariadas (em geral, multivariadas) com marginais uniformes univariadas. No que segue, iremos supor que  $F_1$  e  $F_2$  sejam contínuas.

Do ponto de vista estatístico, o Teorema de Sklar sugere que a construção de um modelo multivariado seja feita em duas etapas:

- (1) ajuste das distribuições marginais;
- (2) calibração de uma cópulapropriada.

Cópulas são invariantes sob transformações estritamente crescentes. Vamos usar uma notação alternativa: se  $X$  e  $Y$  são v.a.'s contínuas, com f.d.  $F$  e cópula  $C$ , escreveremos também  $C_{XY}(u, v)$  para denotar tal cópula.

Se  $\psi, \varphi$  são funções estritamente crescentes, então

$$C_{\psi(X)\varphi(Y)}(u, v) = C_{XY}(u, v), \quad u, v \in I.$$

Veja o Problema 14 para outras propriedades similares.

**Exemplo 11.1.** As três cópulas a seguir são importantes.

- (a) *Cópula produto*:  $\Pi(u, v) = uv$ ,  $u, v \in [0, 1]$ .
- (b) *Cópula comonotônica*:  $M(u, v) = \min(u, v)$ ,  $u, v \in [0, 1]$ .
- (c) *Cópula contramonitorônica*:  $W(u, v) = \max(u + v - 1, 0)$ ,  $u, v \in [0, 1]$ .

Os gráficos de dispersão e f.d.'s (com as respectivas curvas de nível) de  $\Pi$ ,  $M$  e  $W$  encontram-se na Figura 11.1. As cópulas  $W$  e  $M$  representam, respectivamente, a dependência de quadrante negativa máxima e dependência de quadrante positiva

máxima entre  $X$  e  $Y$ . Por sua vez,  $\Pi$  representa a independência entre  $X$  e  $Y$ . Geometricamente,  $X$  e  $Y$  são PQD se e somente se o gráfico de  $z = C(u, v)$  está acima do gráfico da cópula produto  $z = \Pi(u, v)$ . Observe, também, que  $M$  é a f.d. do vetor  $(U, U)$  e  $W$  é a f.d. do vetor  $(U, 1 - U)$ , onde  $U \sim \mathcal{U}(0, 1)$ .

O pacote **copula** do software R foi usado para fazer os gráficos. Os comandos necessários são dados a seguir.

```
> library(copula) # Carrega o pacote de cópulas
> ic <- indCopula # Gera a cópula independente
> persp(ic,pcopula) # Faz o gráfico da densidade de cópula
> persp(ic,dcopula) # Faz gráfico da (f.d.) cópula
> contour(ic,pcopula) # Faz o gráfico de curvas de níveis
> contour(ic,dcopula) # idem
> u <- runif(1000) # simula 1.000 valores de uma uniforme
> plot(u,u) # Faz o gráfico de dispersão de M
> plot(u,-u) # Faz o gráfico de dispersão de W
> v <- runif(1000)
> plot(u,v) # Faz o gráfico de dispersão de \Pi
```

Para gerar as cópulas  $W(u, v)$  e  $M(u, v)$  pode-se programar, em S ou R, as funções  $\min(u, v)$  e  $\max(u + v - 1, 0)$ .

O seguinte resultado é fundamental.

**Proposição 11.1.** Para toda cópula  $C$  temos a desigualdade

$$W(u, v) \leq C(u, v) \leq M(u, v), \quad (u, v) \in I^2. \quad (11.16)$$

Dizemos que  $W(u, v)$  e  $M(u, v)$  são os *limites de Fréchet-Hoeffding*, em razão dos trabalhos de Hoeffding (1940, 1941) e Fréchet (1951). O resultado nos diz que o gráfico de qualquer cópula é uma superfície contínua dentro do cubo unitário  $I^3$ , com vértices  $(0, 0, 0)$ ,  $(1, 0, 0)$ ,  $(0, 1, 0)$  e  $(1, 1, 1)$ , e este gráfico está entre os gráficos das superfícies  $z = W(u, v)$  e  $z = M(u, v)$ . Uma observação importante é que  $W$  é uma cópula somente se  $n = 2$ . Para  $n \geq 3$ ,  $W(u_1, u_2, \dots, u_n) = \max(u_1 + u_2 + \dots + u_n - n + 1, 0)$  não é uma cópula, mas uma quasecópula. Podemos obter limites mais precisos se tivermos informações adicionais sobre a f.d.  $F$ . Veja Nelsen (2006) para mais informações.

Uma cópula tem outras propriedades importantes, que resumiremos na seguinte proposição.

**Proposição 11.2.** Seja  $C$  uma cópula. Então:

---

(a)  $C$  é uniformemente contínua em seu domínio.

(b) vale a condição

$$|C(u_1, v_1) - C(u_2, v_2)| \leq |u_2 - u_1| + |v_2 - v_1|,$$

ou seja,  $C$  é Lipschitz.

- (c) Para qualquer  $u$  ( $v$ ) em  $I$ , a derivada parcial  $\partial_1 = \frac{\partial C}{\partial u}$  ( $\partial_2 = \frac{\partial C}{\partial v}$ ) existe para quase todo  $u(v)$  de  $I$ , e  $0 \leq \partial_1 \leq 1$ ,  $0 \leq \partial_2 \leq 1$ .
- (d) As funções  $v \mapsto \frac{\partial C}{\partial u}$  e  $u \mapsto \frac{\partial C}{\partial v}$  são bem definidas e não decrescentes em quase toda a parte em  $I$ .

Vemos que, como definida,  $C$  é uma função de distribuição bivariada. Podemos, então, considerar a *densidade de cópula*  $c$ , definida por

$$c(F_1(x), F_2(y)) = \frac{\partial}{\partial x \partial y} C(F_1(x), F_2(y)), \quad (11.17)$$

se a derivada existir (neste caso dizemos que  $C$  é *absolutamente contínua*). De (11.14) obtemos que a densidade conjunta de  $X$  e  $Y$  será dada por

$$f(x, y) = c(F_1(x), F_2(y))f_1(x)f_2(y),$$

onde  $f_i$  é a f.d.p. de  $F_i$ ,  $i = 1, 2$ . Logo,

$$c(F_1(x), F_2(y)) = \frac{f(x, y)}{f_1(x)f_2(y)}. \quad (11.18)$$

Os coeficientes de associação de Kendall e Spearman podem ser relacionados com cópulas. Se  $(X, Y)$  tem f.d. conjunta  $F$  e  $C$  é a cópula associada a  $F$ , então pode-se provar que

$$\tau_K = 4 \int_0^1 \int_0^1 C(u, v)c(u, v)dudv - 1 \quad (11.19)$$

e

$$\begin{aligned} \rho_S &= 12 \int_0^1 \int_0^1 uv dC(u, v) - 3, \\ &= 12 \int_0^1 \int_0^1 C(u, v)dudv - 3, \\ &= 12 \int_0^1 \int_0^1 [C(u, v) - uv]dudv. \end{aligned} \quad (11.20)$$

A primeira expressão em (11.20) (veja o Problema 8) implica a expressão (11.9) e a segunda e terceira expressões fornecem as seguintes interpretações geométricas de  $\rho_S$  (Nelsen, 2003):

- (i) o volume sob o gráfico de  $z = C(u, v)$  sobre  $I^2$ , reescalado para estar em  $[-1, 1]$ ;
- (ii) o volume (sinalizado) entre os gráficos de  $z = C(u, v)$  e  $z = \Pi(u, v)$ , reescalado como em (i).

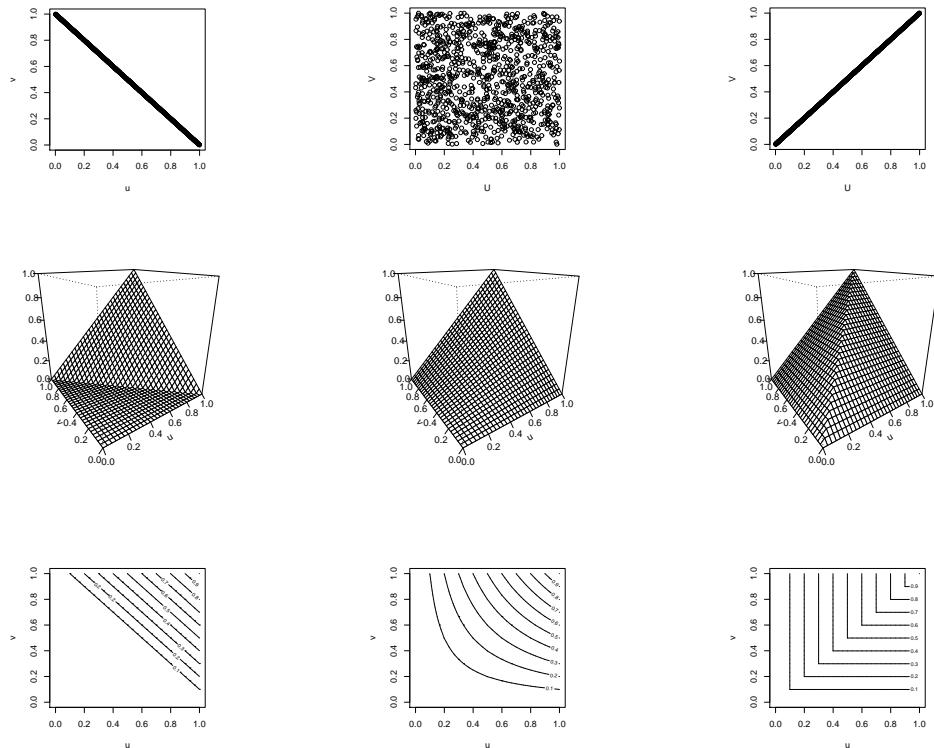


Figura 11.1: Cópulas contramontônica (esquerda), independente (centro) e comonotônica (direita).

Também é possível demonstrar as seguintes desigualdades:

$$-1 \leq 3\tau - 2\rho_S \leq 1,$$

$$\frac{3\tau - 1}{2} \leq \rho_S \leq \frac{1 + 2\tau - \tau^2}{2}, \quad \tau \geq 0,$$

e

$$\frac{\tau^2 + 2\tau - 1}{2} \leq \rho_S \leq \frac{1 + 3\tau}{2}, \quad \tau \leq 0.$$

Para detalhes, veja Nelsen (2006). Também pode-se provar que se  $F_1$  e  $F_2$  são contínuas, então

$$\lambda_u = \lim_{u \rightarrow 1^-} \frac{\bar{C}(u, u)}{1 - u},$$

$$\lambda_\ell = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u},$$

se esses limites existirem, onde  $\bar{C}(u, v)$  é a *cópula de sobrevivência*, dada por  $\bar{C}(u, v) = u + v - 1 + C(1-u, 1-v)$ , que está relacionada com a função de sobrevivência conjunta  $\bar{F}(u, v) = P(U > u, V > v)$ , com  $U, V$  uniformes em  $(0, 1)$ , ou seja,

$$\bar{F}(u, v) = \bar{C}(\bar{F}_1(u), \bar{F}_2(v)).$$

## 11.4 Famílias Paramétricas de Cópulas

Nesta seção apresentaremos algumas cópulas que são frequentemente utilizadas na prática. Cada uma delas depende de um ou mais parâmetros, que chamaremos  $\delta$ , que caracterizam a dependência entre as variáveis. Como antes, vamos nos restringir ao caso bivariado.

(a) **Cópula Gaussiana.** É a cópula de uma distribuição normal bivariada com parâmetro de correlação  $\delta$ , dada por

$$\begin{aligned} C_N(u, v; \delta) &= \int_{-\infty}^{\Phi^{-1}(u)} dx \int_{-\infty}^{\Phi^{-1}(v)} dy \frac{1}{2\pi\sqrt{1-\delta^2}} \exp\left\{-\frac{x^2 - 2\delta xy + y^2}{2(1-\delta^2)}\right\} \\ &= \Phi_\delta(\Phi^{-1}(u), \Phi^{-1}(v)), \end{aligned} \tag{11.21}$$

onde  $\Phi_\delta$  é a f.d. conjunta de uma normal padrão bivariada com coeficiente de correlação  $\delta$ .

A cópula Gaussiana gera distribuição normal padrão bivariada se, e somente se, as marginais são normais padrões. Veja o Exemplo 11.5 abaixo.

Se  $\delta = 1$ , então  $C_N = M$ , se  $\delta = -1$ , então  $C_N = W$  e se  $\delta = 0$ , então  $C_N = \Pi$ .

Para esta cópula temos que

$$\tau_K = \frac{2}{\pi} \text{arcsen}\delta, \quad \rho_S = \frac{6}{\pi} \text{arcsen}\frac{\delta}{2},$$

$$\lambda_\ell = \lambda_u = \begin{cases} 0, & \delta < 1 \\ 1, & \delta = 1 \end{cases}$$

(b) Cópulas de Valores Extremos (VE). Estas têm a propriedade

$$C(u^t, v^t) = (C(u, v))^t, \quad t > 0,$$

e podem ser representadas na forma

$$C(u, v) = \exp \left\{ \log(uv) A \left( \frac{\log(u)}{\log(uv)} \right) \right\},$$

onde  $A : [0, 1] \rightarrow [1/2, 1]$  é uma função convexa, com  $\max(t, t - 1) \leq A(t) \leq 1$ , chamada *função de dependência*.

Alguns casos particulares são:

(i) Cúpula de Gumbel, com f.d.

$$C_G(u, v; \delta) = \exp \left\{ -[(-\log(u))^\delta + (-\log(v))^\delta]^{1/\delta} \right\}, \quad \delta > 1. \quad (11.22)$$

A função de dependência desta cúpula é dada por

$$A(t) = (t^\delta + (1-t)^\delta)^{1/\delta}.$$

O parâmetro  $\delta$  mede o grau de dependência entre as variáveis e se  $\delta = 1$  não há dependência. Para esta cúpula,  $\tau_K = 1 - 1/\delta$  e  $\lambda_u = 2 - 2^{1/\delta}$ .

Outras cúpulas de VE são (veja Joe, 1997, e Nelsen, 2006, para detalhes):

(ii) Cúpula de Galambos;

(iii) Cúpula de Husler-Reiss;

(iv) Cúpula BB5.

(c) Cúpulas Arquimedians. Estas podem ser escritas na forma

$$C(u, v) = \phi^{-1}[\phi(u) + \phi(v)], \quad (11.23)$$

para uma função  $\phi : [0, 1] \rightarrow R^+$ , contínua, estritamente decrescente, convexa,  $\phi(1) = 0$ , chamada *geradora*.

A cúpula de Gumbel é Arquimeditana, com  $\phi(t) = (-\log t)^\delta$ . O coeficiente  $\tau_K$  de Kendall para essas cúpulas é dado por

$$\tau_K = 4 \int \frac{\phi(v)}{\phi'(v)} dv + 1.$$

Dois outros exemplos de cópulas Arquimedianas são a cópula de Frank, dada por

$$C_F(u, v; \delta) = -\delta^{-1} \log \left( \frac{\eta - (1 - e^{-\delta u})(1 - e^{-\delta v})}{\eta} \right), \quad (11.24)$$

onde  $\delta \neq 0$ ,  $\eta = 1 - e^{-\delta}$ ,  $\lambda_u = \lambda_\ell = 0$  e a cópula de Clayton (ou Kimeldorf-Sampson), dada por

$$C_{KS}(u, v; \delta) = [u^{-\delta} + v^{-\delta} - 1]^{-1/\delta}, \quad (11.25)$$

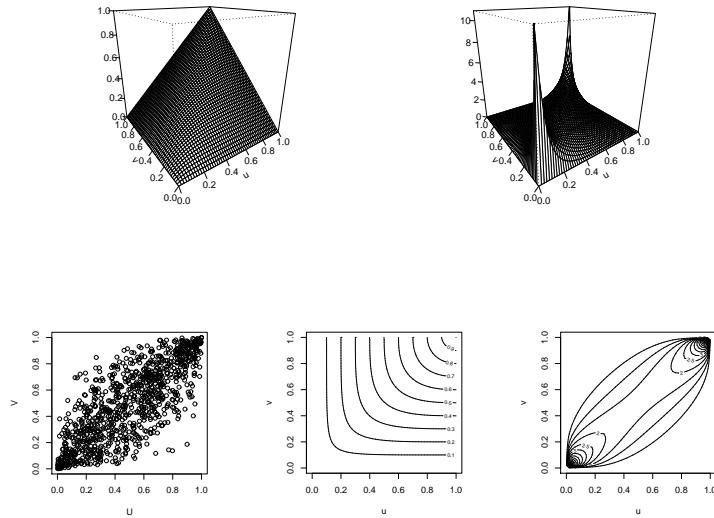
com  $\delta > 0$ ,  $\tau_K = \delta/(\delta + 2)$  e  $\lambda_\ell = 2^{-1/\delta}$ . Outras cópulas Arquimedianas são a cópula de Joe e as cópulas BB1, BB2, BB3, BB6 e BB7. Veja Joe (1997) para outros detalhes sobre essas e outras cópulas.

Nos exemplos a seguir, ilustramos os diversos tipos de dependência que podem ser captados por algumas das cópulas descritas acima. Em cada gráfico, mostramos o gráfico de postos padronizados (ver a seção seguinte para a definição desses postos), a cópula (f.d.), a densidade de cópula e as respectivas curvas de nível. Os comandos do software R encontram-se a seguir, para a cópula Gaussiana, com  $\delta = 0,8$ . Para as demais, os comandos são similares, substituindo `normalCopula` por `tCopula`, `claytonCopula`, `frankCopula` e `gumbelCopula`.

```
> library(copula)
> nc <- normalCopula(0.8)
> persp(nc,pcopula)
> persp(nc,dcopula)
> contour(nc,pcopula)
> contour(nc,dcopula)
> u <- rcopula(nc,1000)
> plot(u)
```

**Exemplo 11.2.** Na Figura 11.2, temos os gráficos para a cópula Gaussiana, com o coeficiente de dependência ( $\rho_P$ , neste caso)  $\delta = 0,8$ . Notamos uma dependência positiva e simétrica.

**Exemplo 11.3.** A Figura 11.3 traz a cópula  $t$  de Student, com  $\rho_P = 0,8$  e número de graus de liberdade  $\nu = 2$ . Obtemos gráficos semelhantes ao caso anterior, mas com maior concentração nos extremos.

Figura 11.2: Cúpula Gaussiana,  $\delta = 0, 8$ .

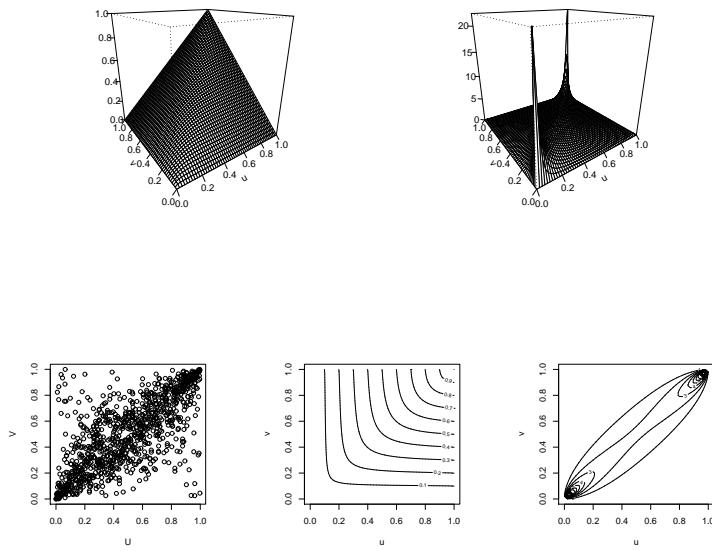
A cúpula  $t$  bivariada é definida como

$$C_{\nu,\Sigma}^t(u,v) = t_{\nu,\mathbf{0},\Sigma}(t_{\nu}^{-1}(u), t_{\nu}^{-1}(v)),$$

onde  $t_{\nu,\mathbf{0},\Sigma}$  é a f.d. de uma distribuição  $t$  bivariada de média  $\mathbf{0}$  e matriz de correlação  $\Sigma$ , e  $t_{\nu}$  é a f.d. de uma  $t$  univariada, com  $\nu$  graus de liberdade. Temos, pois, que  $\delta = (\nu, \Sigma)$ .

**Exemplo 11.4.** Neste exemplo vamos considerar três cúpulas Arquimedianas, Clayton, Frank e Gumbel, com parâmetros 2, 8 e 2, respectivamente. Observamos, na Figura 11.4, que no caso das cúpulas de Clayton e Gumbel há dependência positiva e assimétrica, enquanto a cúpula de Frank apresenta dependência positiva e simétrica, com intensidade maior nos extremos. Compare com os comportamentos das cúpulas Gaussiana e  $t$ .

**Exemplo 11.5.** Retomemos o Exemplo 11.2 de cúpula Gaussiana, com  $\delta = 0, 8$ , mas vejamos o que acontece se especificarmos marginais diferentes. No painel da esquerda da Figura 11.5, temos ambas as marginais normais padrões, no painel do meio, ambas as marginais são uniformes em  $(0, 1)$  e, no painel da direita, uma marginal é normal padrão e a outra é  $t(2)$ . No primeiro caso temos curvas de nível típicas de uma distribuição normal bivariada, enquanto os dois outros casos geram estruturas de dependência com padrões diferentes da normal bivariada.

Figura 11.3: Cúpula  $t$  de Student,  $\delta = 0, 8$ ,  $\nu = 2$ .

**Exemplo 11.6.** Metadistribuições. Este exemplo é baseado em Kojadinovic (2008). Vimos, pela equação (11.14), que o teorema de Sklar fornece um meio para construir distribuições bivariadas (em geral, multivariadas) com f.d.'s marginais e cópula arbitrárias. Consideremos, por exemplo, construir uma distribuição com cópula normal e marginais arbitrárias. Obteremos uma distribuição *metanormal*. Esta terminologia pode ser estendida a outras distribuições.

Como exemplo, vamos obter uma f.d. bivariada, sendo que cada marginal é normal padrão e tendo como cópula uma das seguintes:

- (i) cópula Gaussiana, com  $\delta = 0, 7$ ;
- (ii) cópula de Gumbel, com  $\delta = 2$ ;
- iii) cópula  $t$ , com  $\delta = \rho_p = 0, 71$  e  $\nu = 4$  graus de liberdade.
- (iv) cópula de Clayton, com  $\delta = 2, 2$ ;

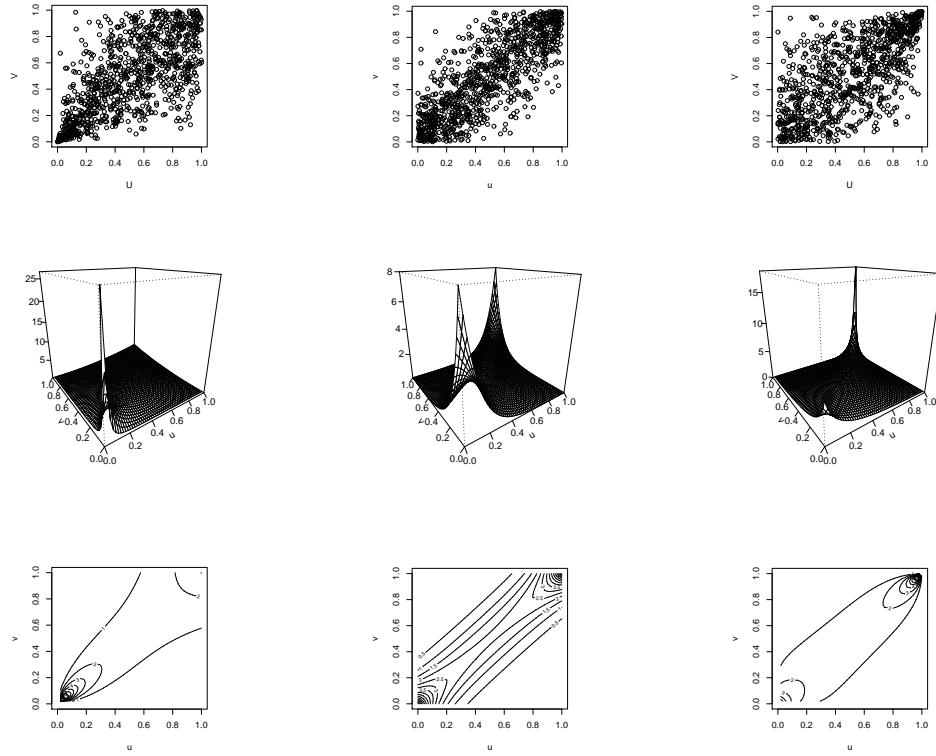


Figura 11.4: Cópulas de Clayton (esquerda), Frank (centro) e Gumbel (direita).

Os parâmetros foram escolhidos de modo que todas tenham o coeficiente de correlação linear aproximadamente igual a 0,7.

Na Figura 11.6, temos as cópulas, e na Figura 11.7, as respectivas metadistribuições.

Observamos que a metadistribuição de Gumbel (assim como sua cópula) apresenta dependência na cauda superior, a metadistribuição de Clayton tem dependência na cauda inferior, a  $t$  tem dependência em ambas as caudas, enquanto a normal não apresenta dependência nas caudas.

Já vimos como gerar as cópulas usando o R. Para gerar metadistribuições usamos as funções `mvdc` e `rmvdc`. Supondo que `nc`, `gc`, `tc` e `cc` representem as cópulas geradas, teremos os comandos para o caso normal (os demais são análogos):

```
> mnc <- mvdc(nc,c("norm","norm"),list(list(mean=0, sd=1),list(mean=0,
+ sd=1)))
> x <- rmvdc(mnc,2000)
> plot(x)
```

Nos Problemas 9, 10 e 11 o leitor é convidado a simular cópulas com outras estruturas de dependência e a gerar metadistribuições com marginais especificadas.

## 11.5 Ajuste de Cópulas Paramétricas

Dada uma amostra  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  de uma v.a. bivariada  $(X, Y)$ , com f.d. conjunta  $F$ , f.d.'s marginais  $F_1, F_2$  e cópula  $C$ , queremos estimar  $C$ . Podemos usar estimadores *paramétricos*, estimadores *não paramétricos* e estimadores *semiparamétricos*. No primeiro caso, usamos estimadores de Máxima Verossimilhança, no segundo caso, podemos usar cópulas empíricas (baseadas em postos) e estimadores suavizados (via kernels, ondaletas etc.). No terceiro caso, usamos estimadores de Pseudo Máxima Verossimilhança.

As funções de distribuição empíricas (f.d.e.) correspondentes a  $F, F_1$  e  $F_2$  são dadas por

$$F_n(x, y) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x, Y_i \leq y\}, \quad -\infty < x, y < \infty, \quad (11.26)$$

$$F_{1n}(x) = \frac{1}{n} \sum_{i=1}^n I\{X_i \leq x\}, \quad -\infty < x < \infty, \quad (11.27)$$

$$F_{2n}(x) = \frac{1}{n} \sum_{i=1}^n I\{Y_i \leq y\}, \quad -\infty < y < \infty. \quad (11.28)$$

Algumas vezes, usa-se o denominador  $n+1$  nestas fórmulas, para garantir que as funções estejam no intervalo  $[0, 1]$ . Pelo lema de Glivenko-Cantelli,  $F_{in}$  aproxima-se de  $F_i$ , quando  $n \rightarrow \infty$ ,  $i = 1, 2$ . Similarmente, pela transformada de probabilidade, espera-se que  $F_{in}(X_j)$ ,  $j = 1, \dots, n$  sejam aproximadamente uniformes,  $i = 1, 2$ .

Como vimos, a cópula  $C_{XY}$  é também a f.d. de  $(U, V) = (F_1(X), F_2(Y))$ , se  $F_1, F_2$  são contínuas, de maneira que  $(U_i, V_i) = (F_1(X_i), F_2(Y_i))$ ,  $i = 1, \dots, n$  formam uma amostra aleatória de  $C_{XY} = C$ .

Logo, se considerarmos  $(\hat{U}_i, \hat{V}_i) = (F_{1n}(X_i), F_{2n}(Y_i))$ ,  $i = 1, \dots, n$ , teremos uma boa ideia de como é a verdadeira cópula  $C$ , ou seja, podemos dizer que esses pares formam uma *pseudoamostra* de  $C$ .

Note, também, que  $(F_{1n}(X_i), F_{2n}(Y_i)) = (\frac{R_i}{n}, \frac{S_i}{n})$ ,  $i = 1, \dots, n$ , onde  $R_i$  é o posto de  $X_i$  entre  $X_1, \dots, X_n$  e  $S_i$  é o posto de  $Y_i$  entre  $Y_1, \dots, Y_n$ . Portanto, o gráfico dos postos padronizados  $(R_i/n, S_i/n)$ ,  $i = 1, \dots, n$  é útil para ressaltar a estrutura de dependência entre  $X$  e  $Y$ .

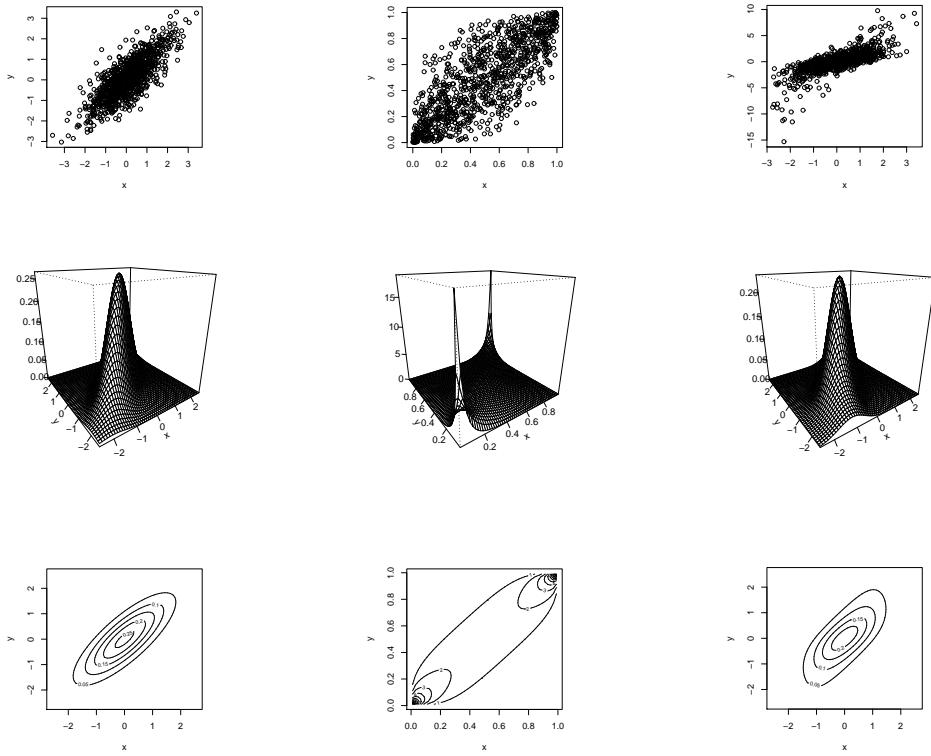


Figura 11.5: Cópulas Gaussianas com marginais: normais padrões (esquerda), uniformes em  $[0, 1]$  (centro) e normal padrão e  $t(2)$  (direita).

### 11.5.1 Estimadores de Máxima Verossimilhança

Dada a amostra  $(x_i, y_i), i = 1, \dots, n$ , de  $F$  bivariada, com marginais  $F_1$  e  $F_2$ , cópula  $C$  e densidade de cópula  $c$ , usando (11.18) a densidade conjunta fica

$$f(x_i, y_i, \eta) = c(F_1(x_i, \alpha_1), F_2(y_i, \alpha_2); \theta) f_1(x_i, \alpha_1) f_2(y_i, \alpha_2), \quad (11.29)$$

onde  $\alpha_1$  contém os parâmetros de  $F_1$ ,  $\alpha_2$  os parâmetros de  $F_2$  e  $\theta$  os parâmetros de  $c$  e seja  $\eta = (\alpha_1, \alpha_2, \theta)$ .

A log-verossimilhança é

$$\begin{aligned} \ell(\eta; \mathbf{x}, \mathbf{y}) &= \sum_{i=1}^n \log c(F_1(x_i, \alpha_1), F_2(y_i, \alpha_2); \theta) \\ &\quad + \sum_{i=1}^n \log f_1(x_i, \alpha_1) + \sum_{i=1}^n \log f_2(y_i, \alpha_2) \end{aligned} \quad (11.30)$$

e os estimadores de máxima verossimilhança (EMV) são obtidos maximizando-se esta função. Normalmente, esta é uma tarefa difícil se há muitos parâmetros.

Um procedimento em dois estágios (chamado *inference function for margins*, IFM) pode ser utilizado, onde, no passo 1, obtemos os estimadores dos parâmetros das marginais,

$$\hat{\alpha}_i = \arg \max \sum_{i=1}^n \log f_i(x_i; \alpha_i), \quad i = 1, 2,$$

e no passo 2 obtemos os estimadores dos parâmetros da cópula,

$$\hat{\theta} = \arg \max \sum_{i=1}^n \log c(F_1(x_i, \hat{\alpha}_1), F_2(y_i, \hat{\alpha}_2); \theta).$$

Este procedimento conduz a estimadores consistentes e assintoticamente normais. Veja Joe e Xu (1996) para detalhes.

### 11.5.2 Cópulas Empíricas

A cópula empírica  $C_n$  fornece uma aproximação (descontínua), ou uma estimativa não paramétrica de  $C$ , baseada em postos padronizados.

**Definição 11.2.** Sejam  $(R_i/n, S_i/n)$  os postos padronizados associados à amostra aleatória  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ . A *cópula empírica*  $C_n$  é definida por

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n I\left\{ \frac{R_i}{n} \leq u, \frac{S_i}{n} \leq v \right\}, \quad u, v \in I. \quad (11.31)$$

Deheuvels (1979) provou que

$$\sup_u \sqrt{n} |C_n(u) - C(u)| = o(1), a.s. \quad (11.32)$$

Os coeficientes  $\hat{\tau}_K$  e  $\hat{\rho}_S$  podem ser expressos em termos de  $C_n$ . Veja Nelsen (2006). A cópula empírica pode também ser expressa como

$$C_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v)), \quad 0 \leq u, v \leq 1. \quad (11.33)$$

Para obter estimadores mais suaves, podemos considerar estimadores suavizados a partir da cópula empírica. Para detalhes, veja Fermanian et al. (2004) e Morettin et al. (2010b).

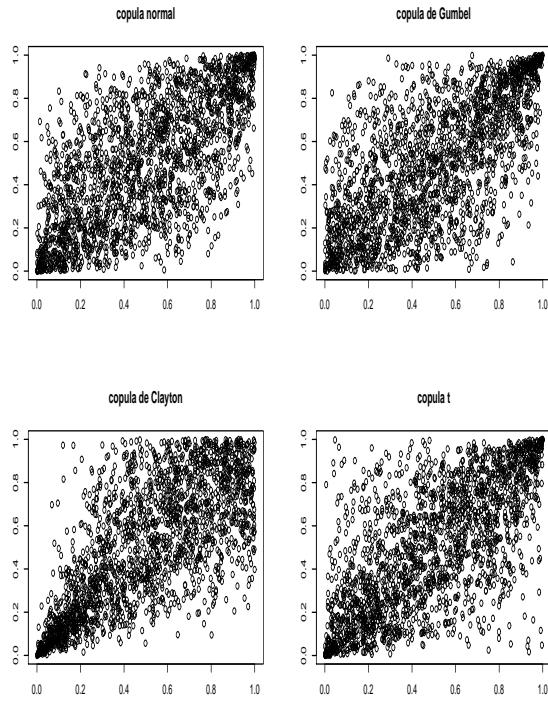


Figura 11.6: Cópulas Gaussiana, Gumbel, Clayton e  $t(4)$ .

### 11.5.3 Estimadores de Pseudo-MV

Neste enfoque,  $F_1$  e  $F_2$  são estimadas, usando modelos paramétricos, f.d.e. ou uma combinação de f.d.e e ajuste de uma distribuição de valores extremos para as caudas das distribuições, por exemplo a distribuição generalizada de Pareto. Para detalhes, ver Carmona (2004) e Zivot e Wang (2006). Seguimos os passos:

(ii) obtemos pseudoamostras para a cópula:

$$(\hat{u}_i, \hat{v}_i) = (\hat{F}_1(x_i), \hat{F}_2(y_i)), i = 1, \dots, n;$$

(iii) formamos a log-verossimilhança:

$$\ell(\theta, \hat{\mathbf{u}}, \hat{\mathbf{v}}) = \sum_{i=1}^n \log c(\hat{u}_i, \hat{v}_i; \theta),$$

e a maximizamos com respeito a  $\theta$  por métodos numéricos. Este método é também chamado de MV canônica.

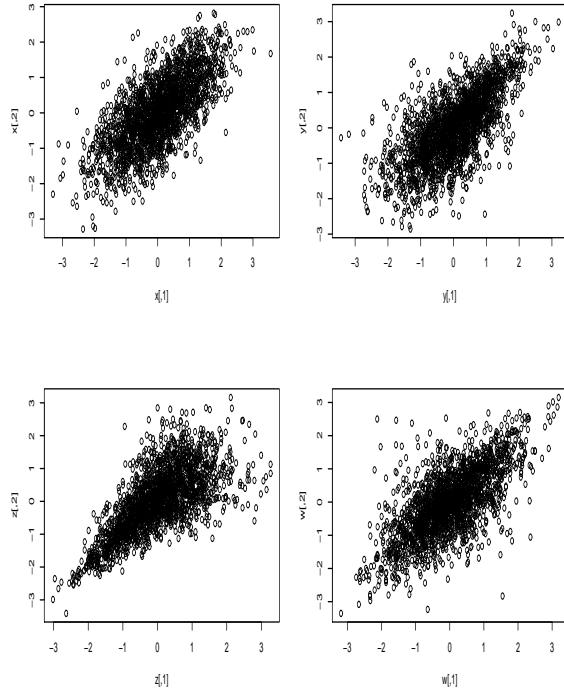


Figura 11.7: Metadistribuições Normal, Gumbel, Clayton e  $t(4)$ .

**Exemplo 11.7.** Vamos aplicar o método de pseudo-MV (PMV) para estimar dados provenientes de uma cópula de Clayton com parâmetro  $\delta = 2$ , supondo que as marginais  $F_1$  e  $F_2$  sejam normais padrões. Os comandos necessários no R estão a seguir, usando a função `fitCopula`, que admite os métodos `ml` (máxima verossimilhança), `mpl` (pseudomáxima verossimilhança), `itau` (inversão do coeficiente de Kendall), e `irho` (inversão do coeficiente de Spearman).

Os três últimos métodos pressupõem que `data` seja composto por pseudo-observações (postos padronizados e reescalados), enquanto o primeiro método pressupõe que “`data`” contenha as observações da cópula desconhecida. Os dois últimos correspondem ao método dos momentos: obtemos estimativa do  $\tau_K$  ou  $\rho_S$  e, conhecida a relação entre o coeficiente e a cópula, obtemos a estimativa do parâmetro que a caracteriza.

```

> library(copula)
> cc <- claytonCopula(2)
> mcc <- mvdc(cc,c("norm","norm"),list(list(mean=0, sd=1),list(mean=0,
  sd=1)))
> x <- rmvdc(mcc,1000)
> u <- cbind(rank(x[,1])/1001, rank(x[,2])/1001) # obtemos pseudo-
  observações;
> fitCopula(u,cc, method="mpl", 1) # estimação por PMV; parâmetro 1 é
  valor inicial;
> f # o resultado.

```

No Quadro 11.1 apresentamos os resultados. A estimativa de PMV é  $\hat{\delta} = 2,072$ , valor significativo (p-valor zero).

Quadro 11.1 : Estimador de PMV para o Exemplo 11.7

The method is Maximum Pseudo-Likelihood based on 1000 observations.
Estimate, Std. Error, z value, $Pr(> z )$
param 2.071874 0.1214130 17.06468 0
The maximized loglikelihood is 445.8937

Usando o método dos momentos, obtemos como estimativas 2,152(itau) e 2,156(irho), que são menos acuradas que a estimativa de PMV. A estimativa da cópula seria obtida por meio de uma cópula paramétrica de Clayton com  $\hat{\delta} = 2,072$ . Na prática, temos apenas os valores observados  $(x_i, y_i), i = 1, \dots, n$ , e o que fazemos é ajustar algumas cópulas paramétricas e escolher aquela que melhor se ajusta aos dados. No exemplo, supusemos que os dados  $x_i, y_i$  provêm de normais padrões. Daremos um exemplo para séries financeiras na seção seguinte.

## 11.6 Cópulas para Séries Temporais

Tudo o que vimos anteriormente foi desenvolvido para o caso que temos um par de v.a.'s  $(X, Y)$  e  $n$  pares independentes obtidos da correspondente distribuição bivariada. Contudo, quando temos duas séries temporais  $(X_t, Y_t), t = 1, \dots, T$ , a amostra  $(X_1, Y_1), (X_2, Y_2), \dots, (X_T, Y_T)$  não é composta de pares independentes: há correlação serial dentro de cada série e correlação cruzada entre as séries, de modo que, por exemplo, os pares  $(X_1, Y_1)$  e  $(X_2, Y_2)$  não são independentes.

Para contornar este problema, uma sugestão é ajustar modelos ARMA-GARCH a cada série e obter, então, os resíduos padronizados do ajuste. Para esses resíduos, ajustamos alguma cópula paramétrica. Este procedimento não produz, obviamente, amostras i.i.d., mas pelo menos a autocorrelação e a hetero-

roscedasticidade de cada série são removidas. Veja Dias e Embrechts (2009, 2010) e Patton (2006). Podemos, também, ajustar um modelo VAR-GARCH bivariado às duas séries.

Uma solução mais apropriada é considerar cópulas para séries temporais, pressupondo, por exemplo, que o vetor  $(X_t, Y_t)$  seja estacionário e alguma forma de independência assintótica (condições *mixing*) seja válida. Métodos não paramétricos, usando kernels e ondaletas, foram utilizados por Fermanian e Scaillet (2003) e Mordrettin et al. (2010a, 2010b). Não trataremos desse desenvolvimento aqui e nos limitaremos a considerar o procedimento citado aqui em um exemplo.

**Exemplo 11.8.** Na Figura 11.8, temos o gráfico de dispersão dos retornos diários dos índices SP500 e DJIA, de 03/01/1994 a 07/07/2000, com  $T = 1.700$  observações. Há uma alta correlação entre as séries, especificamente  $\rho_P = 0,933$ . Os coeficientes de Kendall e Spearman são  $\tau = 0,7341$  e  $\rho_S = 0,9009$ , respectivamente. Ajustamos um modelo AR(3)-GARCH(1,1) com erros  $t$  à série SP500 e um modelo AR(10)-GARCH(1,1) também com erros  $t$  à série DJIA.

Estes modelos foram considerados adequados, usando os testes de diagnóstico usuais. Após o ajuste, os resíduos padronizados apresentaram um coeficiente de correlação de 0,926. Na Figura 11.9, temos o gráfico de curvas de nível de uma cópula Gaussiana ajustada a esses resíduos, com  $\delta = 0,926$ : à esquerda a f.d. e à direita a densidade de cópula. Esses dois gráficos evidenciam a alta dependência entre as séries: as curvas de nível da f.d. apresentam cantos inferiores esquerdos próximos aos de uma cópula monotônica e as curvas da densidade apresentam uma concentração nos dois cantos, inferior esquerdo e superior direito (dependência positiva).

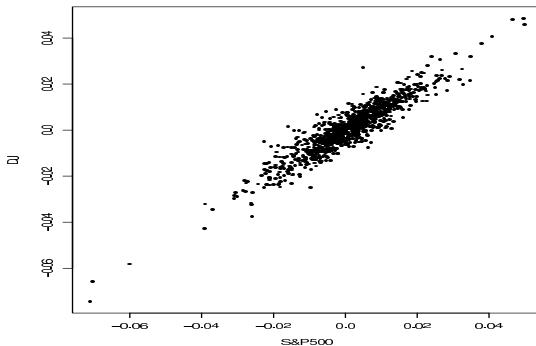


Figura 11.8: Gráfico de dispersão dos retornos de SP500 e DJIA.

Vamos considerar, agora, um exemplo com três séries, ou seja, teremos que ajustar cópulas trivariadas. Embora sejam séries temporais financeiras e teríamos que ajustar antes modelos ARMA-GARCH, vamos fazer o ajuste para as séries originais, a título de ilustração.

**Exemplo 11.9.** Consideremos as séries de retornos diárias das ações da Intel (INTC), Microsoft (MSFT) e General Electric (GE), no período de 01/01/1996 a 31/12/2000 (cinco anos), num total de  $T = 1.262$  observações. Estas séries constam da base de dados DJ do pacote QMRLib do R. Na Figura 11.10, apresentamos os diagramas de dispersão dos dados e, na Figura 11.11, temos os postos padronizados. Notamos que há dependência positiva mais acentuada entre INTC e MSFT e entre MSFT e GE.

Ajustamos as cópulas de Gumbel, Clayton e Frank, todas com parâmetro  $\delta = 2$  e as cópulas Gaussiana e t(5), com média zero e matriz de correlações a estimar pelo método de PMV. No Quadro 11.2 apresentamos os resultados.

Pelos valores da verossimilhança máxima, o melhor ajuste é dado pela cópula t, com  $\nu = 6, 5$ .

Contudo, uma abordagem mais rigorosa para comparar ajustes de cópulas diferentes ao mesmo conjunto de dados consiste em usar testes de adaptação (*goodness-of-fit tests*). Para detalhes, veja Genest et al. (2009) para uma resenha de possíveis testes. A função `gofCopula` do pacote `copula` do R pode ser usado para efetuar esses testes.

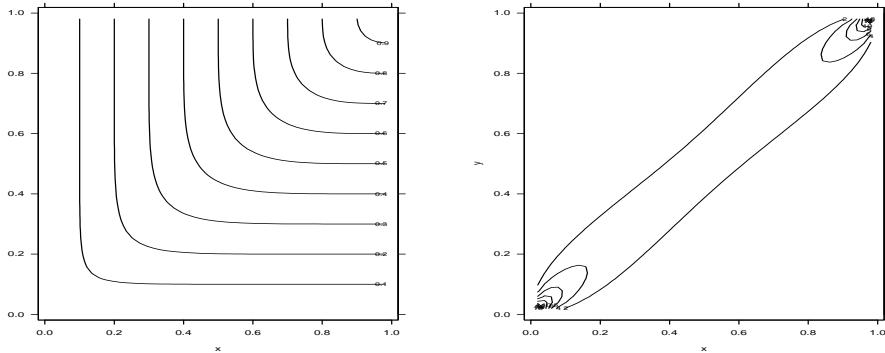


Figura 11.9: Curvas de nível da distribuição (esquerda) e densidade (direita) de uma cópula Gaussiana ajustada aos resíduos de SP500 e DJIA.

## 11.7 Valor em Risco e Cópulas

Já vimos que um problema importante é calcular o VaR (ou outra medida de risco) de uma carteira de ativos. Supondo que esta seja composta de  $N$  instrumentos  $A_1, \dots, A_N$ , com pesos  $w_1, \dots, w_N$  ( $\sum_{i=1}^N w_i = 1$ ), então o log-retorno da carteira no instante  $t$  é dado por (1.16), ou seja,

$$r_{c,t} = \log \sum_{i=1}^N w_i e^{r_{i,t}}. \quad (11.34)$$

Vamos considerar o caso simples de dois retornos, de modo que (11.34) fica

$$r_{c,t} = \log(w_1 e^{r_{1,t}} + w_2 e^{r_{2,t}}). \quad (11.35)$$

Como vimos no Capítulo 7, o  $\text{VaR}_p$  é o  $p$ -quantil da distribuição de  $r_{c,t}$ , isto é,

$$p = P(r_{c,t} \leq \text{VaR}_p) = P(\log(w_1 \exp\{r_{1,t}\} + w_2 \exp\{r_{2,t}\}) \leq \text{VaR}_p). \quad (11.36)$$

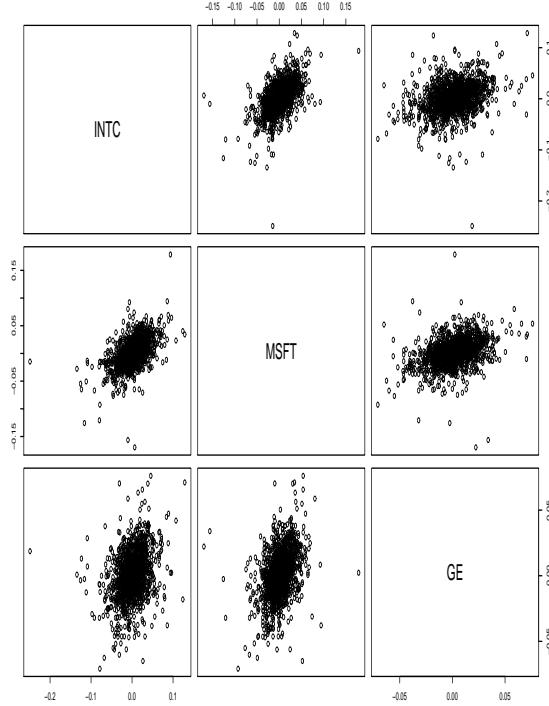


Figura 11.10: Gráficos de dispersão dos retornos de INTC, MSFT e GE.

Chamemos de  $F(r_1)$  e  $F(r_2)$  as distribuições dos retornos e de  $F(r_1, r_2)$  a distribuição conjunta. Por (11.14),  $F(r_1, r_2)$  é especificada por suas marginais e pela cópula  $C$ .

Sejam  $r_1, \dots, r_T$  valores simulados de  $r_{c,t}$ , baseados em  $T$  valores simulados de  $r_{1,t}$  e  $r_{2,t}$ . Então, o valor em risco em (11.36) pode ser estimado pelo  $p$ -quantil empírico dos valores  $r_1, \dots, r_T$ , como vimos na Seção 7.5. A DGP é apropriada para estimar o valor em risco. Podemos, então, ajustar DGP às marginais dos retornos, estimar uma cópula paramétrica, por exemplo, e obter os valores simulados para, finalmente, calcular o  $\text{VaR}$ . Vejamos o procedimento a seguir, por meio de um exemplo.

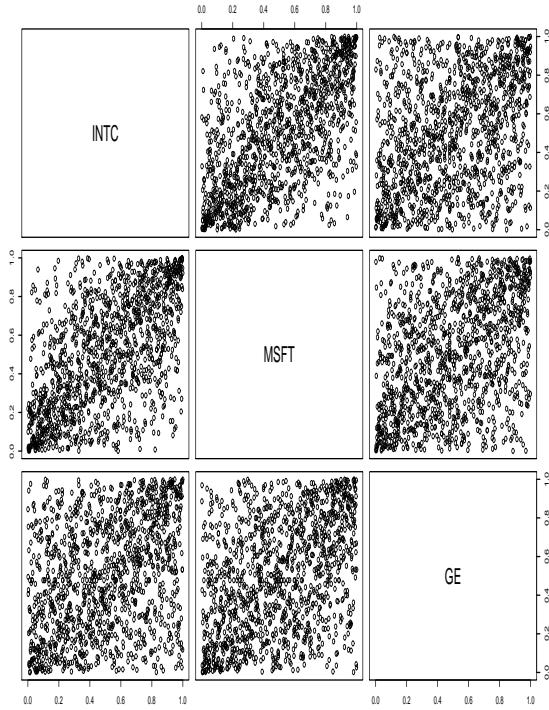


Figure 11.11: Gráficos dos postos padronizados dos retornos de INTC, MSFT e GE.

**Exemplo 11.10.** Vamos considerar uma carteira fictícia, com 70% de ações da Petrobras e 30% de ações da Vale e, para calcular o VaR da carteira, usaremos os retornos dessas ações no período de 31/08/1998 a 29/09/2010, com  $T = 2.990$  observações. Vamos ajustar distribuições generalizadas de Pareto a ambas as séries de retornos, usando a função `gpd.tail` do `S+FinMetrics`. A função `gpd.fit`, usada no Capítulo 7, é outra opção. Lembremos que, como no Exemplo 7.7, temos que selecionar os valores dos limiares para as séries, antes de estimar os parâmetros da DGP. Designando os retornos da Petrobras e da Vale por `rpetro` e `rvale`, respectivamente, os comando utilizados são:

```
> petro.est=gpd.tail(rpetro, lower=-0.02, upper=0.02) : # estima os parâmetros
   do modelo DGP; comando similar para rvale;
> tailplot(petro.est, tail="lower") : # é um comando similar com tail="upper",
   para verificar se o ajuste foi bom;
> petro.sim=gpd.1q(runif(2990),petro.est): # gera 2.990 valores dos retornos
   da Petrobras; comando similar para a Vale;
```

Com os valores simulados `petro.sim` e `vale.sim` podemos simular  $r_{c,t}$  por meio de

---

(11.35) e, em seguida, obter o VaR por meio do  $p$ -quantil empírico. Aqui, não é necessário usar cópulas. Suponha, agora, que ajustemos uma cópula normal, com  $\rho_P = 0,4$  (a correlação entre os retornos é 0,346). A função do S+FinMetrics a ser usada é a `VaR.exp.sim`. Os comandos necessários encontram-se a seguir.

```
> ncop=normal.copula(delta=0.4): # Ajusta uma cópula normal,  $\rho_P = 0,4$ ;
> set.seed(123);
> u.petro=gpd.2p(rpetro, petro.est): # gera os pseudo-valores  $u_i$ ;
> v.vale=gpd.2p(rvale, vale.est): # gera os pseudovalores  $v_i$ ;
> empcop.pv=empirical.copula(x=u.petro,y=v.vale): # calcula a cópula empírica;
> cop.normal.fit=fit.copula(empcop.pv, family="normal", plot=T): # ajusta
uma cópula normal, a partir da cópula empírica;
> VaR=VaR.exp.sim(n=10000,Q(c(0.01,0.05), copula=cop.normal.fit$copula,
+ x.est=petro.est, y.est=vale.est, lambda1=0.7, lambda2=0.3): # calcula o
VaR e o CVaR a partir de  $n = 10000$  simulações das marginais e da cópula.
```

A última função fornece  $\text{VaR}_{0.05} = 0,033$  e  $\text{VaR}_{0.01} = 0,056$ . Os valores do  $\text{CVaR}=ES$  são, respectivamente, 0,049 e 0,078, que são valores maiores do que os do VaR, como esperado.

## 11.8 Comentários Adicionais

Na Seção 11.2, introduzimos algumas medidas de dependência global, como o coeficiente de correlação de Pearson e os coeficientes de correlação de postos de Kendall e Spearman. Todos fornecem um único número, que descreve a dependência entre duas variáveis. A cópula fornece uma medida de dependência local, ou seja, para todo par  $(u, v)$  do quadrado unitário  $I^2$ , temos um valor em  $I = [0, 1]$ .

Várias medidas de dependência local foram definidas na literatura. Bairamov et al. (2003) propuseram a seguinte versão local do coeficiente de correlação de Pearson:

$$H(x, y) = \frac{E[(X - E(X|Y = y))(Y - E(Y|X = x))]}{\sqrt{E[(X - E(X|Y = y))^2]}\sqrt{E[(Y - E(Y|X = x))^2]}}, \quad (11.37)$$

para todo  $(x, y)$  pertencente ao suporte  $S$  de  $(X, Y)$ . Esta medida pode ainda ser escrita na forma

Quadro 11.2 : Estimadores de PMV para o Exemplo 11.9					
Cópula de Gumbel					
The method is Maximum Pseudo-Likelihood based on 1262 observations.					
Estimate, Std. Error, z value, $Pr(> z )$					
param	1.368256	0.01619893	84.46584	0	
The maximized loglikelihood is 294.5982					
Cópula de Clayton					
The method is Maximum Pseudo-Likelihood based on 1262 observations.					
Estimate, Std. Error, z value, $Pr(> z )$					
param	0.5856541	0.02356333	24.85447	0	
The maximized loglikelihood is 273.9629					
Cópula de Frank					
The method is Maximum Pseudo-Likelihood based on 1262 observations.					
Estimate, Std. Error, z value, $Pr(> z )$					
param	2.866457	0.1145294	25.02813	0	
The maximized loglikelihood is 323.3839					
Cópula Gaussiana					
The method is Maximum Pseudo-Likelihood based on 1262 observations.					
Estimate, Std. Error, z value, $Pr(> z )$					
rho.1	0.5781237	0.01568472	36.85904	0	
rho.2	0.3400171	0.01996401	17.03150	0	
rho.3	0.4016913	0.01787757	22.46901	0	
The maximized loglikelihood is 375.7089					
Cópula $t$					
The method is Maximum Pseudo-Likelihood based on 1262 observations.					
Estimate, Std. Error, z value, $Pr(> z )$					
rho.1	0.5877097	NA	NA	NA	
rho.2	0.3593434	NA	NA	NA	
rho.3	0.4224842	NA	NA	NA	
df	6.5015810	NA	NA	NA	
The maximized loglikelihood is 419.2701					

$$H(x, y) = \frac{\text{Cov}(X, Y) + \lambda_X(y)\lambda_Y(x)}{\sqrt{\text{Var}(X) + \lambda_X^2(y)}\sqrt{\text{Var}(Y) + \lambda_Y^2(x)}}, \quad (11.38)$$

na qual  $\lambda_X(y) = E(X) - E(X|Y = y)$ ,  $\lambda_Y(x) = E(Y) - E(Y|X = x)$  e  $\text{Cov}(X, Y)$  é a covariância entre  $X$  e  $Y$ .

Observa-se que, se  $X$  e  $Y$  são independentes, então  $H(x, y) = 0$ .

Note, ainda, que dividindo-se o numerador e o denominador de  $H(x, y)$  pelo produto do desvio padrão de  $X$  pelo desvio padrão de  $Y$ , obtemos

$$H(x, y) = \frac{\rho_P(X, Y) + \varphi_X(y)\varphi_Y(x)}{\sqrt{1 + \varphi_X^2(y)}\sqrt{1 + \varphi_Y^2(x)}}, \quad (11.39)$$

onde  $\varphi_X(y) = \lambda_X(y)/\sqrt{\text{Var}(X)}$ ,  $\varphi_Y(x) = \lambda_Y(x)/\sqrt{\text{Var}(Y)}$  e  $\rho_P(X, Y)$  é o coeficiente de correlação de Pearson.

Exemplos de  $H(x, y)$  para distribuições simétricas elípticas podem ser encontrados em Kotz e Nadarajah (2003), e exemplos para distribuições de valores extremos em Nadarajah et al. (2003). A estimação de  $H$  e extensões para o caso de processos estacionários podem ser vistas em Latif e Morettin (2013).

Outra medida local da associação entre  $Y$  e  $X$ , próximo de  $X = x$ , é a curva de correlação

$$\rho(x) = \frac{\beta(x)\sigma_X}{\sqrt{\beta^2(x)\sigma_X^2 + \sigma_\varepsilon^2(x)}}, \quad x \in S, \quad (11.40)$$

em que  $\beta(x) = \frac{\partial}{\partial x}E(Y|X = x)$  e  $\sigma_\varepsilon^2(x) = \text{Var}(Y|X = x)$ , sendo  $S$  o domínio da variável  $S$ . Essa medida foi proposta por Bjerve e Doksum (1993). Uma crítica a essa medida é que ela é função da variável  $X$  somente.

Bjerve e Doksum (1993) sugerem que  $\rho(x)$  seja estimada via regressão linear ponderada (veja Fan, 1993). Latif (2008) considera a função de correlação para um processo estacionário  $\{(X_t, Y_t), t \in \mathbb{Z}\}$ , supondo o modelo  $Y_t = m(X_t) + \sigma(X_t)\varepsilon_t$ , com  $\varepsilon_t$  i.i.d.  $(0,1)$ , sendo  $m(\cdot)$  o *drift* e  $\sigma(\cdot)$  a volatilidade. A estimação é feita por meio de ajuste de polinômios locais.

Para mais detalhes sobre medidas de dependência local, veja Latif (2008).

## 11.9 Problemas

1. Mostre que, se  $\rho_P(X, Y) = 0$  e se  $(X, Y)$  tiver distribuição normal bivariada, então  $X$  e  $Y$  são independentes.
2. Mostre que  $\rho_P(X, Y)$  é invariante sob transformações lineares estritamente crescentes de  $X$  e  $Y$ .
3. Prove as relações (11.3)-(11.5).
4. Prove que, se  $X$  e  $Y$  são PQD e se  $C$  é a cópula associada a  $F$ , então  $C(u, v) \geq uv$ , para todo par  $(u, v)$  de  $[0, 1]^2$ . A recíproca também vale. Dizemos também que  $F$  ou  $C$  é PQD.
5. **Função de concordância.** Sejam  $(X_1, Y_1)$  e  $(X_2, Y_2)$  dois vetores aleatórios com f.d. conjuntas  $H_1$  e  $H_2$ , mas marginais comuns,  $F$  (de  $X_1$  e  $X_2$ ) e  $G$  (de  $Y_1$  e  $Y_2$ ). Sejam  $C_1$  e  $C_2$  as cópulas de  $(X_1, Y_1)$  e  $(X_2, Y_2)$ , respectivamente.

Então,  $H_1(x, y) = C_1(F(x), G(y))$  e  $H_2(x, y) = C_2(F(x), g(y))$ . Denotemos por  $K$  a diferença entre as probabilidades de concordância e discordância de  $(X_1, Y_1)$  e  $(X_2, Y_2)$

$$K = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0].$$

Então, temos que (Nelsen, 2006):

$$K = K(C_1, C_2) = 4 \int_0^1 \int_0^1 C_2(u, v) dC_1(u, v) - 1.$$

Prove que (11.19) decorre desta relação, ou seja,  $\tau = K(C, C)$ .

6. Prove que  $\rho_S = 3K(C, \Pi)$ , onde  $C$  é a cópula de  $(X_1, Y_1)$ .
7. Prove que a função de concordância  $K$  satisfaz as seguintes propriedades:
  - (i)  $K$  é simétrica em seus argumentos;
  - (ii)  $K$  é não decrescente em seus argumentos;
  - (iii)  $K(M, M) = 1$ ,  $K(W, W) = -1$  e  $K(\Pi, \Pi) = 0$ .
8. Prove que (11.9) decorre da primeira relação em (11.20), se  $X$  e  $Y$  são contínuas, com cópula  $C$ .
9. Gere uma cópula de Clayton com parâmetro  $\delta = 2$  e marginais normais padrões.
10. Idem, com distribuições marginais normal padrão e  $t(2)$ .
11. Gere uma distribuição metaClayton, com marginais  $t$ . Escolha os parâmetros de forma apropriada.
12. **Coeficiente de Gini.** Se  $X$  e  $Y$  são v.a.'s com cópula  $C$ , o coeficiente de cograduação de Gini é definido por

$$\gamma_C = \gamma_{X,Y} = \int_0^1 \int_0^1 (|u + v - 1| - |u + v|) dC(u, v).$$

Prove que  $\gamma_C = K(C, M) + K(C, W)$ , ou seja,  $\gamma_C$  mede uma relação de concordância entre  $C$  e dependência monotônica (dada por  $M$  e  $W$ ).

13. Mostre que, quando  $\delta \rightarrow 0$ , a cópula de Frank se aproxima da cópula independente.

14. Sejam  $X$  e  $Y$  v.a.'s contínuas, com cópula  $C_{XY}$  e sejam  $\psi$ ,  $\varphi$  funções estritamente monotônicas. Prove que:
- se  $\psi$  e  $\varphi$  são ambas crescentes, então  $C_{\psi(X)\varphi(Y)}(u, v) = C_{XY}(u, v)$ ;
  - se  $\psi$  é crescente e  $\varphi$  é decrescente, então  $C_{\psi(X)\varphi(Y)}(u, v) = u - C_{XY}(u, 1 - v)$ ;
  - se  $\psi$  é decrescente e  $\varphi$  crescente, então  $C_{\psi(X),\varphi(Y)}(u, v) = v - C_{XY}(1 - u, v)$ ;
  - se  $\psi$  e  $\varphi$  são ambas decrescentes, então  $C_{\psi(X)\varphi(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v)$ .
15. Prove que (11.6) pode ser escrita como  $\tau_K = E[\text{sinal}((X_1 - X_2)(Y_1 - Y_2))]$ .
16. Prove que  $\hat{\tau}_K = 4/(n(n - 1))\text{PC}_n - 1$ , onde  $\text{PC}_n$  fornece o número de pares concordantes.
17. Considere os índices Ibovespa(Brasil) e Merval (Argentina) do arquivo d-indices.dat. Ajuste cópulas de Gumbel, Clayton e Frank e escolha a que melhor se ajusta, segundo o critério da máxima verossimilhança. Use os métodos de PMV e MV.
18. Agregue às séries do problema anterior, o índice IPC (do México). Ajuste cópulas (trivariadas) Gaussiana e  $t$ . Escolha a que melhor se ajuste, usando os mesmos métodos.
19. Use o método dos momentos (itau e irho) para ajustar as cópulas do problema anterior. Compare com os métodos utilizados anteriormente (PMV e MV).
20. Calcule o  $\text{VaR}_{0.05}$  de uma carteira formada por 60% de ações da Petrobras e 40% das ações da Cemig, no período de 02/01/95 a 27/12/2000.



# Modelos GARCH Multivariados

## 12.1 Introdução

Modelos multivariados podem ser úteis em áreas como seleção de carteiras (*portfolios*), apreçamento de opções, *hedging* e gestão de riscos. Neste capítulo, introduzimos modelos GARCH multivariados para descrever relações dinâmicas entre processos de volatilidade de um mesmo mercado ou entre vários mercados.

Aplicações de modelos multivariados podem ajudar a responder questões como (Bauwens et al., 2006):

- (a) a volatilidade de um mercado pode influenciar as volatilidades de outros mercados?
- (b) a volatilidade de um ativo pode ser transmitida para a volatilidade de outro ativo?
- (c) correlações entre retornos de ativos mudam com o tempo?

Um problema com os modelos GARCH multivariados é que o número de parâmetros a estimar aumenta com o número de séries envolvidas. Nesse sentido, a ideia é considerar modelos parcimoniosos que permitam uma estimação relativamente fácil dos parâmetros, mas que capturem a dinâmica relevante presente na estrutura de covariâncias. Por outro lado, ao especificar um modelo multivariado, é necessário impor que a matriz de covariâncias dos retornos seja positiva definida, uma tarefa que pode ser muito difícil ou mesmo impossível.

Suponha que temos um vetor de retornos  $\mathbf{r}_t$ , de orden  $n \times 1$ , e consideramos o modelo

$$\mathbf{r}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad (12.1)$$

em que  $\boldsymbol{\mu}_t = E(\mathbf{r}_t | \mathcal{F}_{t-1})$  é o vetor de médias condicionais, dada a informação passada, normalmente suposto constante e igual a  $\mu$ , e  $\boldsymbol{\varepsilon}_t$  é um vetor de inovações

da série no instante  $t$ . Podemos supor, também, que  $\mu_t$  siga um modelo VARMA multivariado, como no Capítulo 9, ou seja,

$$\mu_t = \sum_{i=1}^P \Phi_i \mathbf{r}_{t-i} - \sum_{i=1}^Q \Theta_i \varepsilon_{t-i}, \quad (12.2)$$

podendo-se acrescentar a essa equação um vetor de covariáveis (ou variáveis exógenas ou, ainda, explicativas).

Definamos  $\mathbf{H}_t$  como a matriz de covariâncias condicionais, de ordem  $n \times n$ , dada a informação passada  $\mathcal{F}_{t-1}$ , ou seja,

$$\mathbf{H}_t = \text{Cov}(\mathbf{r}_t | \mathcal{F}_{t-1}) = \text{Cov}(\varepsilon_t | \mathcal{F}_{t-1}). \quad (12.3)$$

Então, um modelo de volatilidade para a série de retornos  $\mathbf{r}_t$  será dado pela equação (12.1) mais a equação

$$\varepsilon_t = \mathbf{H}_t^{1/2} \mathbf{z}_t, \quad (12.4)$$

na qual  $E(\mathbf{z}_t) = \mathbf{0}$  e  $\text{Cov}(\mathbf{z}_t) = \mathbf{I}_n$ .

Há muitas possibilidades de generalizações de modelos univariados de volatilidade para o caso de  $n$  séries. Contudo, para  $n$  grande teremos uma quantidade grande de parâmetros a estimar, pois  $\mathbf{H}_t$  terá  $n(n+1)/2$  parâmetros desconhecidos. Um dos objetivos é buscar modelos que sejam parcimoniosos.

Há três abordagens paramétricas para a construção de modelos da forma GARCH multivariados(abreviadamente MGARCH) (Bauwens et al., 2006):

- (i) generalização direta de modelos GARCH univariados; nesta categoria, estão os modelos VEC e BEKK;
- (ii) combinações lineares de modelos GARCH univariados; nesta classe, estão os modelos fatoriais e ortogonais;
- (iii) combinações não lineares de modelos GARCH univariados; nesta classe, estão os modelos com correlações condicionais constantes (CCC) e modelos dinâmicos.

Silvennoinen e Teräsvirta (2009) fazem uma resenha de modelos GARCH multivariados e sugerem as seguintes linhas: (a) como em (i) acima,  $\mathbf{H}_t$  é modelada diretamente, como nos modelos VEC e BEKK; (b) considerar o caso em que  $\mathbf{r}_t$  pode ser representado por um número pequeno de fatores não observados (modelos fatoriais); (c) modelar variâncias e correlações condicionais diretamente, caso dos modelos CCC e extensões; (d) considerar modelos não paramétricos e semiparamétricos.

Podemos, também, considerar uma extensão para o caso multivariado do modelo EWMA. Aqui,

$$\mathbf{H}_t = (1 - \lambda)\mathbf{r}_{t-1}\mathbf{r}'_{t-1} + \lambda\mathbf{H}_{t-1}. \quad (12.5)$$

Dados  $\lambda$  e  $\mathbf{H}_1$ , as matrizes de covariâncias estimadas podem ser obtidas recursivamente. As funções `EWMA.cov` ou `mgarch` do `S+FinMetrics` podem ser usadas para implementar esse procedimento.

**Exemplo 12.1.** Consideremos os retornos diários do Ibovespa e da Petrobras (Petro4), de 02/01/1995 a 30/07/2010, com  $T = 3.847$  observações. Vamos estimar o modelo (12.5). Nesse caso,  $n = 2$  e utilizando a função `mgarch` obtemos as componentes do vetor de média e o valor de  $\lambda = 0,947$ . A seguir, usamos a função `EWMA.cov` com esse valor de  $\lambda$  para obter as variâncias condicionais dos dois retornos e a covariância entre eles, que estão mostradas na Figura 12.1. Em todas as figuras estão presentes picos correspondentes às crises que ocorreram no período.

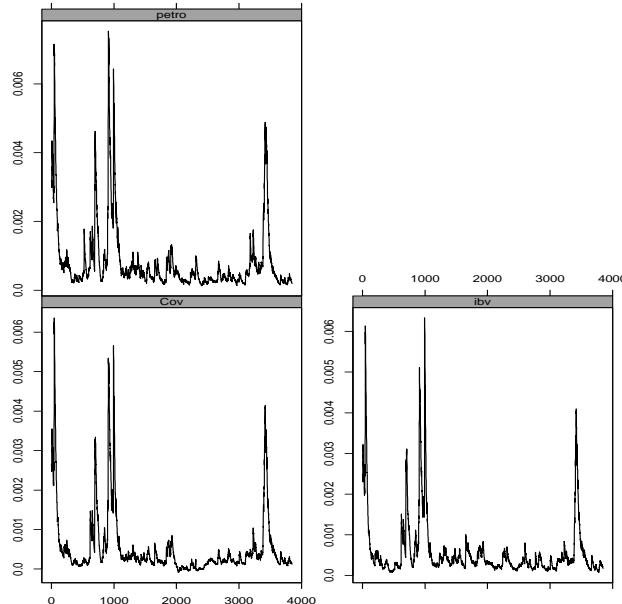


Figura 12.1: Variâncias condicionais (paineis superior e inferior direito) e covariância (paineil inferior esquerdo) para os retornos do Ibovespa e Petrobras.

## 12.2 Generalizações do Modelo GARCH Univariado

Como vimos, quando considerarmos modelos MGARCH, o número de parâmetros cresce rapidamente com  $n$  (dimensão de  $\mathbf{r}_t$ ). Para tornar o modelo tratável, pode-

mos impor estruturas mais simples, por exemplo, supor que matrizes dos coeficientes sejam diagonais. Além disso, temos que garantir que  $\mathbf{H}_t \geq \mathbf{0}$ .

### 12.2.1 Modelos VEC

O modelo VEC geral foi proposto por Bollerslev et al. (1988) e supõe que cada elemento de  $\mathbf{H}_t$  seja uma função linear de erros quadráticos e produtos de erros defasados e de valores defasados de elementos de  $\mathbf{H}_t$ . A partir de agora, vamos considerar somente o caso (1,1) para facilidade de notação.

**Definição 12.1.** O modelo VEC (1,1) é definido por:

$$\mathbf{h}_t = \mathbf{A}_0 + \mathbf{A}\eta_{t-1} + \mathbf{B}\mathbf{h}_{t-1}, \quad (12.6)$$

onde

$$\mathbf{h}_t = \text{vech}(\mathbf{H}_t), \quad (12.7)$$

$$\eta_t = \text{vech}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'). \quad (12.8)$$

Aqui,  $\text{vech}(\cdot)$  denota o operador que transforma a parte triangular inferior de uma matriz  $n \times n$  em um vetor  $n(n+1)/2 \times 1$ ,  $\mathbf{A}$  e  $\mathbf{B}$  são matrizes quadradas de ordem  $n(n+1)/2$  e  $\mathbf{A}_0$  é um vetor de ordem  $n(n+1)/2 \times 1$ . Por exemplo, se  $\mathbf{H}_t = [h_{ij,t}]$  é uma matriz  $2 \times 2$ , então  $\mathbf{h}_t = (h_{11,t}, h_{21,t}, h_{22,t})'$ . O número de parâmetros é  $n(n+1)(n(n+1)+1)/2$ , se  $n = 3$ , temos 78 parâmetros.

Existem somente condições suficientes para que  $\mathbf{H}_t \geq \mathbf{0}$ , para todo  $t$ , veja Gouriéroux (1997).

Uma simplificação é o VEC diagonal, onde cada  $h_{ij,t}$  depende somente de sua defasagem e de valores prévios de  $\boldsymbol{\varepsilon}_{it}\boldsymbol{\varepsilon}_{jt}$ . O número de parâmetros reduz-se a  $n(n+5)/2$ , de modo que se  $n = 3$  temos 12 parâmetros.

**Definição 12.2.** O modelo VEC diagonal (1,1), ou DVEC(1,1), é definido por

$$\mathbf{H}_t = \mathbf{A}_0 + \mathbf{A} \odot (\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}') + \mathbf{B} \odot \mathbf{H}_{t-1}, \quad (12.9)$$

onde  $\mathbf{A}_0$ ,  $\mathbf{A}$  e  $\mathbf{B}$  são matrizes simétricas e  $\odot$  denota o produto de Hadamard, ou seja, multiplicação elemento por elemento.

As extensões para VEC(p,q) ou DVEC(p,q) são imediatas.

**Exemplo 12.2.** Vamos escrever o modelo DVEC (1,1) bivariado na forma matricial

$$\begin{bmatrix} h_{11,t} & \\ h_{21,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} A_{11,0} & \\ A_{21,0} & A_{22,0} \end{bmatrix} + \begin{bmatrix} A_{11,1} & \\ A_{21,1} & A_{22,1} \end{bmatrix} \odot \begin{bmatrix} \varepsilon_{1,t-1}^2 & \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} +$$

$$\begin{bmatrix} B_{11,1} & \\ B_{21,1} & B_{22,1} \end{bmatrix} \odot \begin{bmatrix} h_{11,t-1} & \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix}.$$

Aqui, temos que as volatilidades são dadas por

$$h_{ii,t} = A_{ii,0} + A_{ii,1}\varepsilon_{i,t-1}^2 + B_{ii,1}h_{ii,t-1}, \quad i = 1, 2$$

e a covariância é dada por

$$h_{21,t} = A_{21,0} + A_{21,1}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + B_{21,1}h_{21,t-1}.$$

A matriz  $\mathbf{H}_t$  pode não ser positiva definida. Ela o será, para todo  $t$ , se  $\mathbf{A}_0, \mathbf{A}, \mathbf{B}$  e  $\mathbf{H}_0$  forem positivas definidas (Attanasio, 1991). Na prática, é suficiente tratar  $\mathbf{H}_t$  como simétrica e considerar somente a parte triangular inferior do sistema (12.9).

**Exemplo 12.3.** Consideremos os dados do Exemplo 12.1 e vamos ajustar um modelo DVEC(1,1) com erros gaussianos. No Quadro 12.1, temos a saída resultante da aplicação da função `mgarch` do S-Plus e a função `summary`:

```
> summary(ibv.petro.dvec, method="qmle")
```

Nesse quadro,  $C(1)$  e  $C(2)$  são os elementos de  $\mu$  na equação (12.1);  $A(1, 1)$ ,  $A(2, 1)$  e  $A(2, 2)$  são os elementos da matriz  $\mathbf{A}_0$ ;  $\text{ARCH}(1; i, j)$  são os elementos da matriz  $\mathbf{A}$  e  $\text{GARCH}(1; i, j)$  são os elementos da matriz  $\mathbf{B}$ . Vemos que todos os coeficientes são significativos e, portanto, obtemos que as volatilidades são dadas por

$$h_{11,t} = 0,00001292 + 0,1006\varepsilon_{1,t-1}^2 + 0,872h_{11,t-1}, \quad (12.10)$$

$$h_{22,t} = 0,00001464 + 0,09434\varepsilon_{2,t-1}^2 + 0,8851h_{22,t-1}, \quad (12.11)$$

$$h_{21,t} = 0,00000959 + 0,08447\varepsilon_{1,t-1}\varepsilon_{2,t-1} + 0,8897h_{21,t-1}. \quad (12.12)$$

Para obter a matriz de covariâncias dos estimadores, podemos usar a função `vcov` do S-Plus, que é baseada nos produtos dos gradientes. A função `op` calcula a matriz de covariâncias baseada na inversa da matriz Hessiana numérica e a função `qmle` dá a matriz de covariância robusta (estimadores de quase-verossimilhança). Assim, o comando

```
> sqrt(diag(vcov(ibv.petro.dvec,method="qmle"))))
```

produz desvios padrões das estimativas dos coeficientes. De modo similar, as funções `residuals` e `sigma.t` produzem os resíduos padronizados e volatilidades estimadas, respectivamente.

Quadro 12.1: Estimação do modelo DVEC(1,1) do Exemplo 12.3				
Call:				
<code>mgarch(formula.mean = ibv.petro ~ 1, formula.var = ~ dvec(1, 1), trace = FALSE)</code>				
Mean Equation: <code>structure(.Data = ibv.petro ~ 1, class = "formula")</code>				
)				
Conditional Variance Equation: <code>structure(.Data = ~ dvec(1, 1)</code>				
, <code>class = "formula"</code> )				
Conditional Distribution: gaussian				
Estimated Coefficients				
Coefficient	Value	Std.Error	t value	$Pr(> t )$
C(1)	1.762e-003	2.932e-004	6.009	2.035e-009
C(2)	2.002e-003	3.457e-004	5.790	7.583e-009
A(1, 1)	1.292e-005	3.308e-006	3.907	9.524e-005
A(2, 1)	9.587e-006	2.363e-006	4.058	5.048e-005
A(2, 2)	1.464e-005	3.813e-006	3.839	1.257e-004
ARCH(1; 1, 1)	1.006e-001	1.566e-002	6.427	1.463e-010
ARCH(1; 2, 1)	8.447e-002	1.121e-002	7.537	5.951e-014
ARCH(1; 2, 2)	9.434e-002	1.417e-002	6.655	3.225e-011
GARCH(1; 1, 1)	8.720e-001	1.963e-002	44.429	0.000e+000
GARCH(1; 2, 1)	8.897e-001	1.524e-002	58.388	0.000e+000
GARCH(1; 2, 2)	8.851e-001	1.681e-002	52.658	0.000e+000

No Quadro 12.2, temos as estatísticas usuais para o diagnóstico do modelo ajustado. Usando a função `plot` do **S+FinMetrics** obtemos os gráficos das Figuras 12.2, 12.3 e 12.4. Na Figura 12.2 temos as autocorrelações dos quadrados dos resíduos. Vemos que não há autocorrelações significativas para as séries individuais, mas algumas presentes nas correlações cruzadas, indicando que o modelo não capta toda a heteroscedasticidade presente nas séries.

Isso também é indicado pelos p-valores dos testes Ljung-Box, multiplicadores de Lagrange e F do Quadro 12.2. Na Figura 12.3 temos os gráficos  $Q \times Q$  para normalidade, indicando que a distribuição normal não é apropriada, o que também é mostrado pelo teste de Jarque-Bera no Quadro 12.2.

Quadro 12.2: Diagnóstico do modelo DVEC(1,1) do Exemplo 12.3

AIC(11) = -39774.87 , BIC(11) = -39706.07
---

Normality Test:			
Jarque-Bera	P-value	Shapiro-Wilk	P-value
394.2	0	0.9877	8.311e-018
758.4	0	0.9834	7.592e-021

Ljung-Box test for standardized residuals
---

Statistic	P-value	Chi <sup>2</sup> -d.f.
20.62	0.05629	12
21.04	0.04987	12

Ljung-Box test for squared standardized residuals
---

Statistic	P-value	Chi <sup>2</sup> -d.f.
22.54	0.03187	12
12.67	0.39350	12

TR <sup>2</sup>	P-value	F-stat	P-value
20.02	0.06677	1.829	0.1277
12.25	0.42552	1.118	0.4553

ACF of Squared Std. Residuals

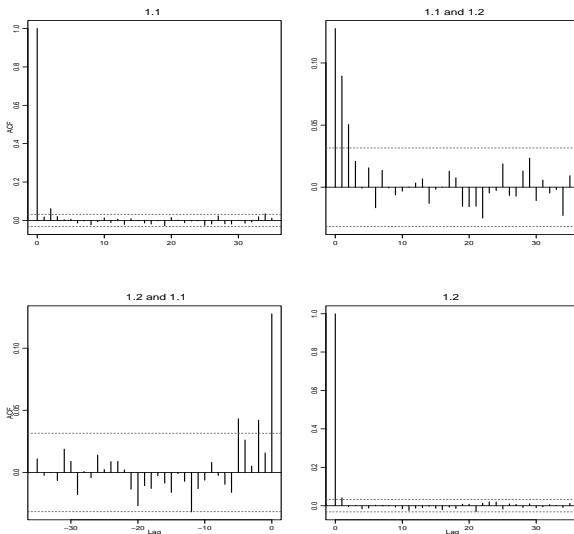


Figura 12.2: FAC amostral dos quadrados dos resíduos para o Exemplo 12.3.

Finalmente, na Figura 12.4, temos os dois gráficos das volatilidades estimadas pelo modelo e na Figura 12.5 a correlação cruzada condicional entre as duas séries. Um teste de Ljung-Box multivariado (*portmanteau*) também pode ser usado, o qual indica que a hipótese de não existência de correlação nos resíduos é rejeitada. Veja Zivot e Wang (2006) para detalhes.

Se analisarmos as autocorrelações e correlações cruzadas das séries, bem como as correlações parciais, poderemos tentar identificar possíveis modelos VAR que podem ser ajustados antes que modelos DVEC (ou outros) sejam considerados. Veja o Problema 1.

### 12.2.2 Modelos BEKK

Esta classe de modelos foi introduzida por Engle e Kroner (1995), em virtude do fato que é difícil garantir que  $\mathbf{H}_t \geq \mathbf{0}$  sem impor condições fortes sobre os parâmetros. Além disso, no modelo DVEC as volatilidades e covariâncias condicionais dependem somente de seus próprios valores defasados e do produto dos erros. Ou seja, um choque em uma série não afetará a volatilidade da outra série diretamente. Os modelos a seguir corrigem esse fato, à custa de um número maior de parâmetros.

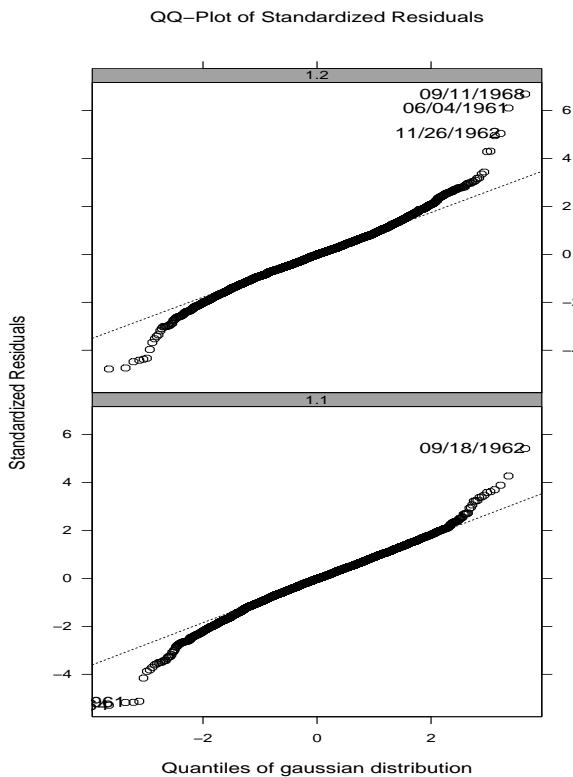


Figura 12.3:  $Q \times Q$  plot dos resíduos do modelo DVEC(1,1) ajustado.

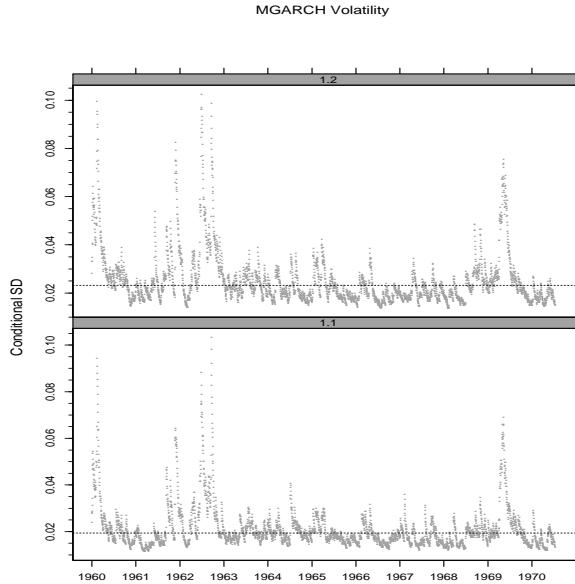


Figura 12.4: Volatilidades estimadas pelo modelo DVEC(1,1) ajustado.

**Definição 12.3.** O modelo BEKK(1,1) é definido por

$$\mathbf{H}_t = \mathbf{A}_0 \mathbf{A}_0' + \mathbf{A}' \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}' \mathbf{A} + \mathbf{B}' \mathbf{H}_{t-1} \mathbf{B}, \quad (12.13)$$

onde  $\mathbf{A}_0$  é triangular inferior e  $\mathbf{A}, \mathbf{B}$  são matrizes  $n \times n$  irrestritas.

Para reduzir o número de parâmetros, podemos supor que  $\mathbf{A}$  e  $\mathbf{B}$  sejam diagonais. Para que o modelo seja estacionário, os autovalores de  $\mathbf{A} + \mathbf{B}$  devem ser menores que um, em módulo.

Este modelo tem  $n(5n + 1)/2$  parâmetros. Por exemplo, se  $n = 3$ , teremos 24 parâmetros, comparado com 12 do modelo DVEC(1,1). A extensão para o modelo BEKK(1,K) e BEKK(1,p,q) é imediata. Esses modelos raramente são usados se tivermos mais do que três ou quatro séries. Os modelos fatoriais, descritos a seguir, contêm menos parâmetros.

**Exemplo 12.4.** Vamos ajustar um modelo BEKK(1,1) para os dados diários de retornos do Ibovespa e da Petrobras (Petro 4), considerados anteriormente no Exemplo 12.1. Os comandos usados do S+FinMetrics são:

```
> ibv.petro.bekk=mgarch(ibv.petro ~ 1, ~ bekk(1,1))
> summary(ibv.petro.bekk)
```

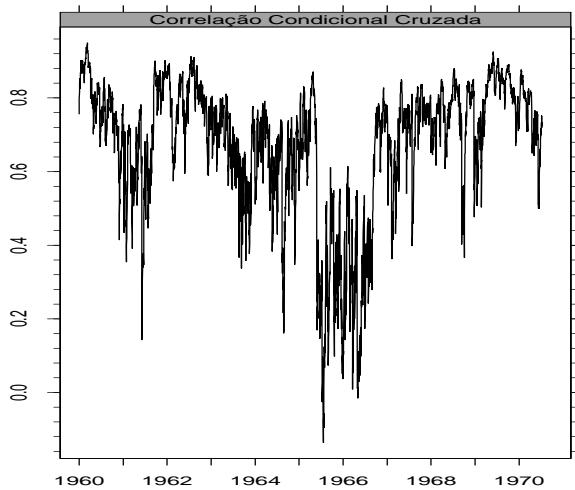


Figura 12.5: Correlações cruzadas condicionais estimadas pelo modelo DVEC(1,1) ajustado.

No Quadro 12.3, temos a saída do programa com coeficientes estimados e estatísticas apropriadas para avaliar se os verdadeiros coeficientes do modelo são significativos. Verificamos que todos são.

A equação da volatilidade é dada por

$$\begin{aligned}
 \begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} = & \begin{bmatrix} 0,0035 & 0 \\ 0,0033 & 0,0018 \end{bmatrix} \begin{bmatrix} 0,0035 & 0,0033 \\ 0 & 0,0018 \end{bmatrix} + \\
 & + \begin{bmatrix} 0,287 & 0,056 \\ 0,052 & 0,298 \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} 0,287 & 0,052 \\ 0,056 & 0,298 \end{bmatrix} + \\
 & + \begin{bmatrix} 0,946 & -0,019 \\ -0,021 & 0,948 \end{bmatrix} \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} 0,946 & -0,021 \\ -0,019 & 0,948 \end{bmatrix}.
 \end{aligned}$$

Quadro 12.3. Estimação do modelo BEKK(1,1) para o Exemplo 12.4

Call:

mgarch(formula.mean = ibv.petro ~ 1, formula.var = " ~ bekk(1, 1))

Mean Equation: structure(.Data = ibv.petro ~ 1

, class = "formula"

)

Conditional Variance Equation: structure(.Data = ~ bekk(1, 1)

, class = "formula"

)

Conditional Distribution: gaussian

Estimated Coefficients:

coefficient	value	Std. error	t value	$Pr(> t )$
C(1)	0.001724	0.0002829	6.092	1.224e-009
C(2)	0.001863	0.0003384	5.504	3.955e-008
A(1, 1)	0.003489	0.0002286	15.263	0.000e+000
A(2, 1)	0.003330	0.0003096	10.756	0.000e+000
A(2, 2)	0.001788	0.0001806	9.900	0.000e+000
ARCH(1; 1, 1)	0.287454	0.0151304	18.999	0.000e+000
ARCH(1; 2, 1)	0.052127	0.0153461	3.397	6.888e-004
ARCH(1; 1, 2)	0.056300	0.0129424	4.350	1.396e-005
ARCH(1; 2, 2)	0.298428	0.0141715	21.058	0.000e+000
GARCH(1; 1, 1)	0.946104	0.0057013	165.946	0.000e+000
GARCH(1; 2, 1)	-0.021364	0.0059473	-3.592	3.321e-004
GARCH(1; 1, 2)	-0.019165	0.0045394	-4.222	2.478e-005
GARCH(1; 2, 2)	0.948118	0.0048058	197.288	0.000e+000

AIC(13) = -39698.03

BIC(13) = -39616.72

No Quadro 12.4, temos as estatísticas para o diagnóstico do modelo ajustado. Verificamos que a suposição de normalidade dos resíduos não é verificada, bem como vemos que os resíduos não se comportam como ruído branco. As causas são as mesmas apontadas no ajuste do modelo DVEC.

Para comparar as correlações condicionais para os dois modelos ajustados até agora, usamos o comando:

```
> seriesPlot(cbind(ibv.petro.dvec$R.t[1,2],ibv.petro.bekk$R.t[1,2],
+ strip=c("DVEC Corr","BEKK Corr"), one.plot=F,layout=c(1,2,1))
```

Na Figura 12.6, apresentamos o resultado. Vemos que as figuras são similares, mas aquela correspondente ao modelo BEKK é mais volátil.

Quadro 12.4. Diagnóstico do modelo BEKK(1,1) para o Exemplo 12.4

Normality Test:

Jarque-Bera	P-value	Shapiro-Wilk	P-value
367.9	0	0.9883	2.382e-017
1279.6	0	0.9782	7.786e-024

Ljung-Box test for standardized residuals:

Statistic	P-value	Chi <sup>2</sup> -d.f.
20.61	0.05637	12
24.62	0.01671	12

Ljung-Box test for squared standardized residuals:

Statistic	P-value	Chi <sup>2</sup> -d.f.
18.27	1.078e-001	12
52.92	4.258e-007	12

Lagrange multiplier test:

TR <sup>2</sup>	P-value	F-stat	P-value
16.68	1.621e-001	1.523	0.219463
48.18	2.906e-006	4.436	0.004042

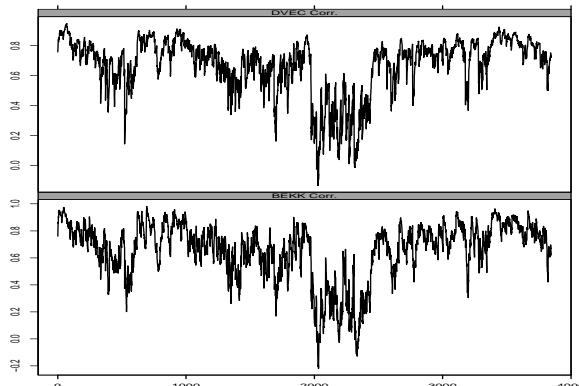


Figura 12.6: Correlações cruzadas condicionais estimadas pelos modelos DVEC(1,1) e BEKK(1,1) ajustados.

Nas Figuras 12.7-12.9, apresentamos as volatilidades, autocorrelações dos quadrados dos resíduos e  $Q \times Q$  plots dos resíduos, respectivamente, obtidas do ajuste do modelo BEKK(1,1).

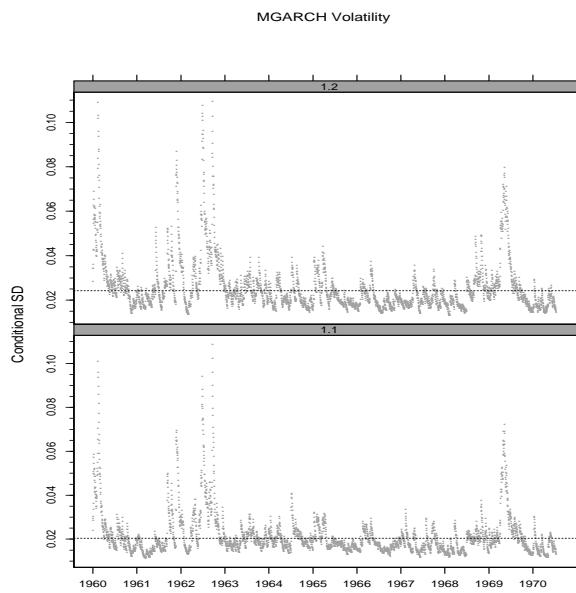


Figura 12.7: Volatilidades estimadas pelo modelo BEKK(1,1) ajustado.

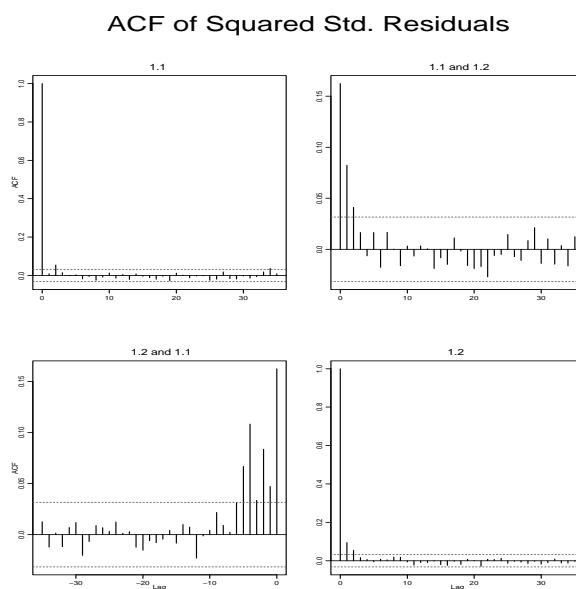


Figura 12.8: Autocorrelações dos quadrados dos resíduos estimadas pelos modelo BEKK(1,1) ajustado.

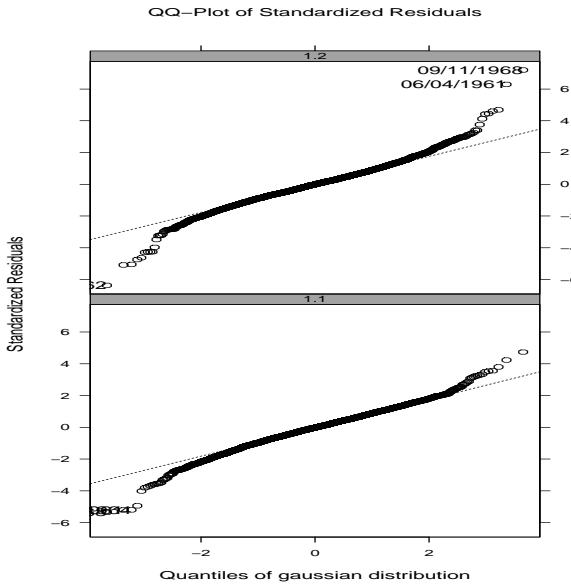


Figura 12.9:  $Q \times Q$  plot dos resíduos do modelo BEKK(1,1) ajustado.

### 12.3 Combinações Lineares de Modelos GARCH Univariados

Nesta seção, trataremos dos modelos fatoriais e, dentre estes, os modelos ortogonais. Há um número grande de modelos fatoriais propostos na literatura. Nesta seção abordaremos somente alguns. Para outros, veja os dois artigos de revisão, já citados anteriormente.

#### 12.3.1 Modelos Fatoriais

Modelos fatoriais (MF) são motivados por alguma teoria econômica, por exemplo, na teoria de apreçamento com arbitragem (Ross, 1976), os retornos são gerados por um pequeno número de fatores não observados.

O primeiro MF foi introduzido por Engle et al. (1990), no qual supõe-se que as observações sejam geradas por fatores condicionalmente heteroscedásticos, seguindo uma estrutura GARCH.

**Definição 12.4.** O modelo  $F$ -GARCH é tal que a matriz  $\mathbf{H}_t$  segue

$$\mathbf{H}_t = \boldsymbol{\Omega} + \sum_{k=1}^r \boldsymbol{\omega}_k \boldsymbol{\omega}_k' f_{k,t}, \quad (12.14)$$

onde os  $f_{k,t}$  são os fatores, que podem ser correlacionados,  $\boldsymbol{\Omega} \geq \mathbf{0}$  é uma matriz  $n \times n$  e  $\boldsymbol{\omega}_k$ ,  $k = 1, \dots, r$  são vetores  $n \times 1$  linearmente independentes (pesos fatoriais).

Supomos que os fatores seguem um modelo GARCH (1,1):

$$f_{k,t} = \omega_k + \alpha_k (\boldsymbol{\gamma}'_k \boldsymbol{\varepsilon}_{t-1})^2 + \beta_k f_{k,t-1}, \quad (12.15)$$

sendo os  $\omega_k, \alpha_k, \beta_k$  escalares e  $\boldsymbol{\gamma}_k$  um vetor  $n \times 1$  de pesos.

Engle et al. (1990) apresentam um método de estimação consistente (mas não eficiente) em dois estágios, usando MV. Diebold e Nerlove (1989) propuseram um modelo similar, mas que é, na realidade, um MVE.

A suposição de que os fatores sejam correlacionados não é desejável, pois pode acontecer que vários fatores captarem características similares dos dados. Vários modelos fatoriais foram, então, propostos, supondo que os fatores sejam não correlacionados. Nesses modelos, supomos que

$$\mathbf{r}_t = \mathbf{W}\mathbf{f}_t, \quad (12.16)$$

na qual  $\mathbf{W}$  é uma matriz  $n \times n$  não singular e os fatores  $\mathbf{f}_t$ , não correlacionados, tipicamente seguem um modelo GARCH. Diferentes especificações para  $\mathbf{W}$  resultam nos diversos MF propostos na literatura.

### 12.3.2 Modelos Ortogonais

Aqui, os dados observados são uma transformação ortogonal de  $n$  (ou um número menor de) processos GARCH univariados. Para detalhes sobre modelos ortogonais, consultar Alexander e Chibumba (1997) e Alexander (2001).

**Definição 12.5.** O modelo O-GARCH (1,1,m) é definido por

$$\mathbf{V}^{-1/2} \boldsymbol{\varepsilon}_t = \mathbf{u}_t = \boldsymbol{\Lambda}_m \mathbf{f}_t, \quad (12.17)$$

onde:

- (i)  $\mathbf{V} = \text{diag}\{v_1, \dots, v_n\}$ ,  $v_i = \text{Var}(\varepsilon_{it})$ ;
- (ii)  $\boldsymbol{\Lambda}_m$  é uma matriz de dimensão  $n \times m$ , dada por

$$\boldsymbol{\Lambda}_m = \mathbf{P}_m \text{diag}\{\ell_1^{1/2}, \dots, \ell_m^{1/2}\}, \quad (12.18)$$

sendo  $\ell_1 \geq \ell_2 \geq \dots \ell_m > 0$  os  $m$  maiores autovalores da matriz de correlação de  $\mathbf{u}_t$  e  $\mathbf{P}_m$  é a matriz  $n \times n$  dos autovetores associados.

- (iii)  $\mathbf{f}_t = (f_{1t}, \dots, f_{mt})'$  é um processo estocástico tal que

$$E(\mathbf{f}_t | \mathcal{F}_{t-1}) = 0, \quad \text{Var}(\mathbf{f}_t | \mathcal{F}_{t-1}) = \mathbf{Q}_t = \text{diag}\{\sigma_{f_{1t}}^2, \dots, \sigma_{f_{mt}}^2\}, \quad (12.19)$$

$$\sigma_{f_{it}}^2 = (1 - \alpha_i - \beta_i) + \alpha_i f_{i,t-1}^2 + \beta_i \sigma_{f_{i,t-1}}^2, \quad i = 1, \dots, m. \quad (12.20)$$

Segue-se que a matriz  $\mathbf{H}_t$ ,  $n \times n$ , é gerada por  $m$  modelos GARCH univariados e

$$\mathbf{H}_t = \text{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = \mathbf{V}_t^{1/2} \mathbf{Q}_t \mathbf{Q}_t' \mathbf{V}_t^{1/2}, \quad (12.21)$$

com  $\mathbf{V}_t = \text{Var}(\mathbf{u}_t | \mathcal{F}_{t-1}) = \mathbf{\Lambda}_m \mathbf{Q}_t \mathbf{\Lambda}_m'$ .

Os parâmetros populacionais a estimar são  $(\mathbf{V}, \mathbf{\Lambda}_m)$  e os coeficientes  $\alpha_i, \beta_i$  dos modelos GARCH univariados. O valor de  $m$  pode ser escolhido por uma análise de CP aplicada aos resíduos  $\hat{u}_t$  e deve ser menor que o número de ativos. Outra possibilidade é usar ICA (*independent components analysis*), veja adiante.

O modelo proposto por van der Weide (2002) supõe que a matriz  $\mathbf{\Lambda}$  seja quadrada e inversível, no lugar de ortogonal.

**Definição 12.6.** O modelo GO-GARCH(1,1) é como na definição anterior, com  $m = n$  e  $\mathbf{\Lambda}$  uma matriz não singular de parâmetros. A matriz de correlações condicionais de  $\varepsilon_t$  fica

$$\mathbf{R}_t = \mathbf{J}_t^{-1} \mathbf{V}_t \mathbf{J}_t^{-1}, \quad (12.22)$$

com  $\mathbf{J}_t = (\mathbf{V}_t \odot \mathbf{I}_n)^{1/2}$ ,  $\mathbf{V}_t = \mathbf{\Lambda} \mathbf{Q}_t \mathbf{\Lambda}'$ .

Os modelos ortogonais são estacionários se os processos GARCH univariados forem estacionários.

### 12.3.3 Modelo Fatorial via Componentes Principais

Uma outra possibilidade, sugerida por Ding (1994) e Alexander (1998) é usar um modelo GARCH baseado em componentes principais. Na análise de Componentes Principais (ACP), para qualquer matriz de covariâncias  $\mathbf{H}$  podemos encontrar uma matriz ortogonal  $\mathbf{\Lambda}$ , cujas colunas são os autovetores de  $\mathbf{H}$ , e uma matriz diagonal  $\mathbf{\Delta}$ , que tem na diagonal principal os autovalores de  $\mathbf{H}$ , tal que  $\mathbf{\Lambda} \mathbf{\Delta} \mathbf{\Lambda}' = \mathbf{H}$ . Se  $s_t = \mathbf{\Lambda}' \mathbf{r}_t$  são as CP de  $\mathbf{r}_t$ , esta tem matriz de covariâncias diagonal.

Nesse modelo, modelamos cada CP de  $s_t$  como um modelo GARCH univariado. O programa S+FinMetrics usa a função `mgarch`, com a especificação `pr-comp.garch(1,1)` como a fórmula da variância condicional.

**Exemplo 12.5.** Vamos considerar os retornos diários das ações da HP e da IBM, de 02/02/1984 a 31/12/1992, constantes da base de dados do S+FinMetrics, e lá denominadas `hp.s` e `ibm.s`, respectivamente. Cada série tem  $T = 2.000$  observações. Os gráficos dos retornos está na Figura 12.10. Vamos ajustar um modelo usando componentes principais.

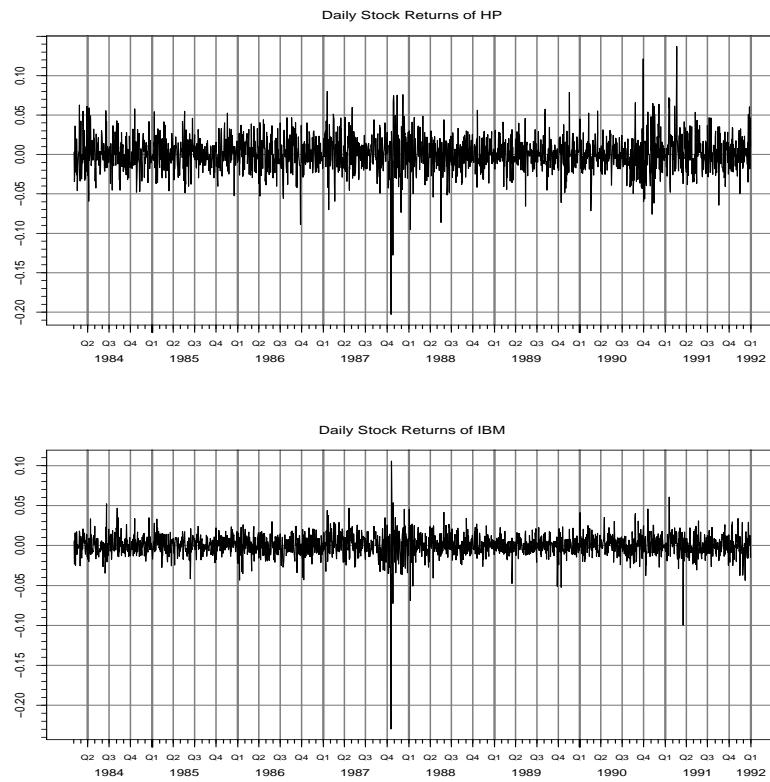


Figura 12.10: Retornos diários da HP e da IBM.

Os comandos necessários estão descritos a seguir, usando-se a distribuição  $t$  como erro no modelo.

```
> hp.ibm=seriesMerge(hp.s,ibm.s)
> plot(hp.s)
> hp.ibm.mod=mgarch(hp.ibm~ 1, ~prcomp.garch(1,1), cond.dist="t", trace=F)
> summary(hp.ibm.mod)
```

O modelo estimado está no Quadro 12.5 e o diagnóstico no Quadro 12.6. Vemos que o termo constante do modelo 12.1 é nulo, os outros coeficientes são significativos. Os testes usuais mostram que o modelo é adequado. Veja o Problema 12.

**Exemplo 12.6.** Vamos, agora, ajustar um modelo GO-GARCH (1,1) aos retornos do Ibovespa e Petrobras, de 02/01/2002 a 31/08/2009, com  $T = 1.901$  observações. Usamos a função `gogarchfit` do pacote `rmgarch` do repositório R. Lembremos que, nesse caso, o número de fatores é igual ao número de séries.

Quadro 12.5. Estimação do modelo com CP-GARCH para o Exemplo 12.5

Call:

```
mgarch(formula.mean = hp.ibm ~ 1, formula.var = ~ prcomp.garch(1, 1),
+ trace = FALSE, cond.dist = "t")
```

Mean Equation: structure(.Data = hp.ibm ~ 1  
, class = "formula")  
)

Conditional Variance Equation: structure(.Data = ~ prcomp.garch(1, 1)  
, class = "formula")  
)

Conditional Distribution: t

with estimated parameter 6.473744 and standard error 0.4896983

Estimated Coefficients:

	Value	Std.Error	t value	$Pr(> t )$
C(1)	-0.00037982	4.662e-004	-0.8148	0.41529501
C(2)	0.00006064	2.090e-004	0.2902	0.77171520
A(1, 1)	0.00002366	8.347e-006	2.8348	0.00463161
A(2, 2)	0.00004772	1.179e-005	4.0473	0.00005379
ARCH(1; 1, 1)	0.04303688	1.221e-002	3.5256	0.00043212
ARCH(1; 2, 2)	0.13065363	3.353e-002	3.8970	0.00010060
GARCH(1; 1, 1)	0.91427044	2.303e-002	39.7055	0.00000000
GARCH(1; 2, 2)	0.43375105	1.222e-001	3.5506	0.00039330

Eigenvectors: (orthonormal transform matrix):

HP	IBM
HP	-0.9054
IBM	-0.4245

Eigenvalues:

---

[1] 0.0006002 0.0001222

---

AIC(9) = -22211.32

BIC(9) = -22160.91

---

Os comandos necessários estão apresentados a seguir (veja, também, a documentação do programa):

Quadro 12.6. Diagnóstico do modelo CP-GARCH para o Exemplo 12.5

Normality Test:

	Jarque-Bera	P-value	Shapiro-Wilk	P-value
PRCOMP1	2390	0	0.9718	2.133e-019
PRCOMP2	7667	0	0.9504	1.773e-025

Ljung-Box test for standardized residuals:

	Statistic	P-value	Chi <sup>2</sup> -d.f.
PRCOMP1	17.66	0.1264	12
PRCOMP2	16.96	0.1509	12

Ljung-Box test for squared standardized residuals:

	Statistic	P-value	Chi <sup>2</sup> -d.f.
PRCOMP1	11.229	0.5094	12
PRCOMP2	3.963	0.9841	12

Lagrange multiplier test:

	TR <sup>2</sup>	P-value	F-stat	P-value
PRCOMP1	11.371	0.4974	1.0397	0.5205
PRCOMP2	3.845	0.9861	0.3503	0.9990

```
> spec.n=gogarchspec(mean.model=list(demean="constant"),
  variance.model=list(model="sGARCH", garchOrder=c(1,1), submodel=NULL),
  distribution.model=list(distribution="'manig'"), ica="fastica")
> fit=gogarchfit(spec=spec.n, data=mydata, gfun="tanh")
> coef(fit).
```

Obtemos as seguintes equações para as volatilidades dos fatores:

$$\sigma_{f_{1t}}^2 = 0,00086 + 0,04324f_{1,t-1}^2 + 0,95576\sigma_{f_{1,t-1}}^2,$$

$$\sigma_{f_{2t}}^2 = 0,00237 + 0,05573f_{2,t-1}^2 + 0,94327\sigma_{f_{2,t-1}}^2.$$

A função também retorna a matriz de rotação que foi utilizada. Na estimação, o pacote **rmgarch** usa a análise de componentes independentes (ICA) para obter os fatores, por meio do algoritmo FastICA. Veja a seção seguinte.

### 12.3.4 Modelo GICA-GARCH

Esse modelo foi proposto por García-Ferrer et al. (2012) e é também baseado no fato de que mercados financeiros dependem de alguns fatores latentes que representam os comovimentos de variáveis financeiras. Para estimar esses fatores, em vez de usar-se, por exemplo, ACP, é utilizada a Análise de Componentes Independentes (ICA, em inglês). Podemos dizer que a ICA busca fatores que sejam mutuamente independentes e não gaussianos. Dado que dados financeiros são, em geral, não gaussianos, a ICA parece ser mais apropriada do que a ACP para explicar esse comportamento.

Depois que as componentes independentes (CIs) são estimadas, elas são ordenadas em termos da variância total explicada e algumas delas são, então, usadas no modelo GICA-GARCH. Nesse modelo, um modelo ARMA-GARCH é ajustado a cada uma das CIs, de modo que a matriz de covariância condicional seja diagonal. Nesse sentido, o modelo supõe uma estrutura fatorial para a distribuição incondicional dos dados e também para a sua distribuição condicional, já que aproxima a matriz de covariância condicional dos dados por uma combinação linear das variâncias condicionais de um pequeno número de fatores.

A seguir, usaremos a notação do artigo citado acima. Seja  $\mathbf{X}_t = (X_{1t}, \dots, X_{nt})'$  um processo estacionário, de média zero e matriz de covariâncias  $\boldsymbol{\Gamma}_X(\tau)$ , positiva definida. Então,

$$\mathbf{X}_t = \mathbf{A}\mathbf{s}_t, \quad t = 1, \dots, T, \tag{12.23}$$

onde  $\mathbf{A}$  é uma matriz  $n \times r$  de posto completo e  $\mathbf{s}_t = (s_{1t}, \dots, s_{rt})'$  é um vetor  $r \times 1$  de fatores latentes. Cada elemento  $a_{ij}$  de  $\mathbf{A}$  mede o efeito de  $s_{jt}$  sobre  $X_{it}$ . Como as  $s_{it}$  são as CIs, com média zero,  $\mathbf{s}_t$  tem como matriz de covariâncias contemporâneas  $\boldsymbol{\Gamma}_S(0) = \mathbf{I}_r$ .

Dadas as observações  $\mathbf{X}_1, \dots, \mathbf{X}_T$ , o objetivo é estimar  $\mathbf{A}$  e  $\mathbf{s}_t$ . De outro modo, a ICA busca uma matriz  $\mathbf{W}$ , de ordem  $r \times n$ , tal que

$$\hat{\mathbf{s}}_t = \mathbf{W}\mathbf{X}_t, \quad t = 1, \dots, T, \tag{12.24}$$

de modo que as  $s_{it}$  sejam tão independentes quanto possível. De (12.23) segue que  $\boldsymbol{\Gamma}_X(0) = E(\mathbf{X}_t\mathbf{X}_t') = \mathbf{A}\mathbf{A}'$  e  $\boldsymbol{\Gamma}_X(\tau) = \mathbf{A}\boldsymbol{\Gamma}_S(\tau)\mathbf{A}'$ ,  $\tau \geq 1$ . Como as componentes são independentes, a matriz  $\boldsymbol{\Gamma}_S(\tau)$  é diagonal.

Para obter as CIs, três algoritmos foram propostos: JADE, FastICA e SOBI. Os dois primeiros são baseados na não gaussianidade das CIs e, o terceiro, na falta de correlação temporal entre as componentes. Para detalhes sobre esses algoritmos, ver Hyvärinen et al. (2001).

Para decidir quais são as CIs mais importantes, é necessário ponderá-las. Back e Weigend (1997) sugerem ordenar as CIs em termos de sua variabilidade explicada. Sabemos que

$$X_{it} = \sum_{j=1}^n a_{ij} s_{jt}, \quad \text{Var}(X_{it}) = \sum_{j=1}^n a_{ij}^2, \quad i = 1, \dots, n. \quad (12.25)$$

Para cada  $X_{it}$ , defina CIs ponderadas por

$$\mathbf{s}_t^{(i)} = \text{diag}(a_{i1}, \dots, a_{in}) \mathbf{s}_t,$$

de modo que

$$s_{jt}^{(i)} = a_{ij} s_{jt}, \quad j = 1, \dots, n, \quad \text{Var}(s_{jt}^{(i)}) = a_{ij}^2, \quad i, j = 1, \dots, n. \quad (12.26)$$

De (12.25) e (12.26) a variância de  $X_{it}$ , que é explicada por  $s_{jt}^{(i)}$ , é calculada por

$$\nu_j^{(i)} = \frac{a_{ij}^2}{\sum_{j=1}^n a_{ij}^2}, \quad i, j = 1, \dots, n, \quad (12.27)$$

e a variância total de  $\mathbf{X}_t$ , explicada pela  $j$ -ésima CI, é dada por

$$\vartheta_j = \frac{\sum_{i=1}^n \nu_j^{(i)}}{\sum_{j=1}^n (\sum_{i=1}^n \nu_j^{(i)})}, \quad j = 1, \dots, n.$$

Obtidas  $\{\vartheta_1, \dots, \vartheta_n\}$ , podemos dizer quanto da variância total é explicada por cada CI e, então, ordená-las em termos de variabilidade.

Obtidas as CIs, García-Ferrer et al. (2912) propõem ajustar primeiro um modelo VARMA (p,q)  $r$ -dimensional, como em (12.2), a  $\mathbf{s}_t$ , mas com  $\Phi_i$  e  $\Theta_\ell$  matrizes diagonais. A seguir, um modelo GARCH (m,n) é ajustado aos erros  $\varepsilon_t$ . Ou seja, temos (12.3) e (12.4), mas agora  $\mathbf{H}_t$ , a matriz de covariâncias condicionais de  $\mathbf{s}_t$ , é diagonal, de ordem  $r \times r$ , ou seja, supomos que as correlações condicionais das CIs sejam nulas. Temos, então, que

$$h_{jt} = \alpha_0^{(j)} + \sum_{i=1}^m \alpha_i^{(j)} \varepsilon_{j,t-i}^2 + \sum_{\ell=1}^n \beta_\ell^{(j)} h_{j,t-\ell}, \quad j = 1, \dots, r, \quad (12.28)$$

representa a variância condicional da  $j$ -ésima CI. Para que  $h_{jt} > 0$ , para todo  $j$ , suponha que  $\alpha_0^{(j)} > 0$ ,  $\alpha_i^{(j)} \geq 0$  e  $\beta_i^{(j)} \geq 0$ . Chamando  $q = \max(m, n)$ , também vale, para todo  $j$ ,  $\sum_{i=1}^q (\alpha_i^{(j)} + \beta_i^{(j)}) < 1$ .

O modelo é ajustado em duas etapas. Na primeira, usamos ICA para obter as CIs e a matriz de pesos (*loadings*). Todos os algoritmos mencionados acima padronizam os dados inicialmente. Na segunda etapa, os modelos GARCH univariados são ajustados às CIs.

**Exemplo 12.7.** Vamos reproduzir aqui a aplicação feita por Gegembauer (2010). Foram usados os preços diários de fechamento de 20 ações que compõem o índice

Ibovespa, de 02/01/2002 a 28/12/2007, totalizando  $T = 1.488$  observações. Foram consideradas ações de empresas de oito setores e diversos segmentos. Por exemplo, no setor Matérias Básicas, foram consideradas três empresas do segmento de Siderurgia, uma do segmento Petroquímico, uma do segmento de Papel e Celulose e uma do segmento de Mineração. Os retornos foram normalizados de modo a terem média zero. O autor usou um algoritmo de ponto fixo, FastICA. Os CIs foram ordenados segundo a variabilidade explicada, como explicado anteriormente. A Tabela 12.1 mostra essa ordenação.

Tabela 12.1: Componentes ordenadas em função da variabilidade explicada.

$\vartheta_j$	%	$\vartheta_j$	%	$\vartheta_j$	%	$\vartheta_j$	%
$\vartheta_{11}$	14,20	$\vartheta_5$	5,24	$\vartheta_4$	3,51	$\vartheta_1$	3,07
$\vartheta_{12}$	12,32	$\vartheta_{20}$	5,09	$\vartheta_6$	3,48	$\vartheta_{14}$	2,88
$\vartheta_{18}$	8,23	$\vartheta_7$	4,13	$\vartheta_2$	3,46	$\vartheta_{13}$	2,87
$\vartheta_3$	6,06	$\vartheta_{10}$	3,82	$\vartheta_{19}$	3,22	$\vartheta_{15}$	2,83
$\vartheta_{16}$	5,96	$\vartheta_{17}$	3,82	$\vartheta_9$	3,07	$\vartheta_8$	2,73

Optamos por considerar os quatro primeiros componentes da Tabela 12.1, que respondem por 40,81% do total da variabilidade explicada. A Tabela 12.2 apresenta os maiores coeficientes da matriz  $\mathbf{W}$  de (12.24) para as CIs escolhidas. Observamos que não é possível fazer uma associação clara com algum setor. Ou seja, as CIs não representam algum índice de mercado.

Os modelos ARMA-GARCH ajustados às quatro componentes estão na Tabela 12.3, da qual obtemos a matriz  $\mathbf{H}_t = \text{diag}\{h_{1t}, \dots, h_{4t}\}$ .

## 12.4 Combinações Não Lineares de Modelos GARCH Univariados

Nessa classe de modelos, podemos especificar separadamente as variâncias condicionais individuais e a matriz de correlação condicional.

### 12.4.1 Modelos com Correlações Condicionais

Inicialmente, escolhemos um modelo para cada variância condicional, que pode ser qualquer um da família GARCH. Depois, modela-se a matriz de correlações condicionais.

Um primeiro modelo, proposto por Bollerslev (1990), considera que a matriz de correlações condicionais é constante.

**Definição 12.7.** O modelo CCC (correlações condicionais constantes) é definido por

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R} \mathbf{D}_t = [\rho_{ij} \sqrt{h_{iit} h_{jxt}}]_{i,j}, \quad (12.29)$$

onde

$$\mathbf{D}_t = \text{diag}\{h_{11t}^{1/2}, \dots, h_{nnt}^{1/2}\}, \quad (12.30)$$

sendo  $h_{iit}$  definida como em qualquer modelo GARCH univariado e  $\mathbf{R} = [\rho_{ij}]$  é uma matriz simétrica, positiva definida,  $\rho_{ii} = 1$ , para todo  $i$ .

O modelo CCC tem  $n(n + 5)/2$  parâmetros,  $\mathbf{H}_t \geq \mathbf{0}$  se, e somente se, todas as  $n$  variâncias condicionais são positivas e  $\mathbf{R} \geq \mathbf{0}$ . O modelo CCC original tem a equação da variância dada por aquela de um modelo GARCH,

$$h_{iit} = w_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i h_{ii,t-1}, \quad i = 1, \dots, n. \quad (12.31)$$

Resultados teóricos sobre estacionariedade, ergodicidade e momentos não são fáceis de obter.

Tabela 12.2: Coeficientes das componentes escolhidas

<b>CP11</b>	<b>CP12</b>	<b>CP18</b>	<b>CP03</b>
Braskem - 25,151	Aracruz - 24,179	Vale - 37,124	Gerdau - 28,624
Itausa - 25,127	Itausa - 19,477	Embraer - 16,423	Itausa - 24,649
Gerdau - 22,867	Telemar - 18,811	Cemig - 15,012	Vale - 20,979
TIM - 11,122	TIM - 16,938	Aracruz - 12,406	Aracruz - 18,873
Telemar - 8,6327	Petrobras - 13,941	Souza Cruz - 12,384	CSN - 14,963
Bradesco - 7,0596	Braskem - 11,554	Usiminas - 12,092	Tim - 12,421
Souza Cruz - 6,9894	Comgs - 10,944	Gerdau - 11,006	Ambev - 8,7932
Cemig - 4,8665	Gerdau - 10,122	Comgs - 10,459	Sadia - 7,8157
Sadia - 4,5928	Vale - 10,092	Bradesco - 10,351	Braskem - 6,7779
Aracruz - 3,8861	Ambev - 9,8743	Eletrobras - 9,1754	Cemig - 6,4464
BB - 3,6095	Eletrobras - 7,9189	Tim - 7,8748	BB - 5,2088
Petrobras - 3,4416	Bradesco - 6,8929	Ambev - 6,1275	Souza Cruz - 5,0397
Ambev - 3,0161	CSN - 5,5489	BB - 2,7889	Usiminas - 3,3905
Vale - 2,8131	Usiminas - 5,2032	Petrobras - 1,9334	Embraer - 3,2827
Comgs - 1,5811	Embraer - 4,907	Sadia - 1,3859	Petrobras - 3,1551
CSN - 1,508	BB - 4,8098	Braskem - 1,0562	Lojas Ame. - 2,6693
Eletrobras - 0,7778	Lojas Ame. - 3,7955	CSN - 0,96037	Bradesco - 1,8108
Usiminas - 0,37879	Sadia - 2,936	Itausa - 0,83702	Comgs - 1,5549
Lojas Ame. - 0,35448	Cemig - 1,7648	Telemar - 0,5409	Eletrobras - 1,3869
Embraer - 0,23515	Souza Cruz - 1,6436	Lojas Ame. - 0,2064	Telemar 0,15594

Tabela 12.3: Modelos ajustados e seus coeficientes

Componente	Modelo	Coeficientes
CP11	GARCH(1,1)	$\alpha_0 = 0,072339$ $\alpha_1 = 0,065365$ $\beta_1 = 0,861581$
CP12	AR(4)-GARCH(1,1)	$\phi_3 = -0,06065; \phi_4 = -0,07035$ $\alpha_0 = 0,008099; \alpha_1 = 0,020639$ $\beta_1 = 0,970502$
CP18	GARCH(1,1)	$\alpha_0 = 0,059991$ $\alpha_1 = 0,04817$ $\beta_1 = 0,892026$
CP03	AR(5)-GARCH(1,1)	$\phi_1 = 0,065112; \phi_2 = -0,057627$ $\phi_5 = -0,054906; \alpha_0 = 0,020139$ $\alpha_1 = 0,046746; \beta_1 = 0,930999$

**Exemplo 12.8.** Vamos considerar os retornos diárias das ações da HP e da IBM, de 02/02/1984 a 31/12/1992, do Exemplo 12.5, e ajustar um modelo CCC (1,1).

Os comandos necessários estão descritos a seguir:

```
> hp.ibm.model=mgarch(hp.ibm~ 1, ~ccc(1,1), trace=F)
> summary(hp.ibm.model)
```

No Quadro 12.7 temos o resultado da aplicação da função `mgarch` do `S+FinMetrics`, para um modelo CCC(1,1) e, no Quadro 12.8, os resultados do diagnóstico do modelo ajustado. Vemos que, com exceção dos coeficientes constantes, os demais têm valor-*p* pequenos, mostrando que os correspondentes coeficientes verdadeiros são diferentes de zero.

Os testes de diagnóstico para a f.a.c dos resíduos e dos quadrados dos resíduos mostram que a hipótese de ruído branco para os erros do modelo não pode ser rejeitada.

Finalmente, os testes  $TR^2$  e  $F$  nos dizem que o modelo captou toda a heteroscedasticidade presente nos dados. O teste de Jarque-Bera rejeita a hipótese de normalidade dos retornos, o que também é mostrado na Figura 12.13, embora as caudas não sejam tão pesadas.

Nas Figuras 12.11, 12.12 e 12.14 temos, respectivamente, as f.a.c dos resíduos, dos quadrados dos resíduos e as volatilidades estimadas.

As equações para as volatilidades ficam

$$\begin{aligned} h_{11,t} &= 0,0000485 + 0,08217\varepsilon_{1,t-1}^2 + 0,82067h_{11,t-1}, \\ h_{22,t} &= 0,0000271 + 0,11974\varepsilon_{2,t-1}^2 + 0,74001h_{22,t-1}, \end{aligned}$$

ao passo que a matriz de correlações contantes estimada fica

$$\mathbf{H}_t = \mathbf{H} = \begin{bmatrix} 1,0000 & 0,5421 \\ 0,5421 & 1,0000 \end{bmatrix}.$$

Quadro 12.7. Estimação do modelo CCC(1,1) para o Exemplo 12.8

Call:

```
mgarch(formula.mean = hp.ibm ~ 1, formula.var = ~ ccc(1, 1), trace = FALSE)
Mean Equation: structure(.Data = hp.ibm ~ 1
, class = "formula"
)
Conditional Variance Equation: structure(.Data = ~ ccc(1, 1)
, class = "formula"
)
Conditional Distribution: gaussian
```

Estimated Coefficients

Coefficient	Value	Std.Error	t value	Pr(> t )
C(1)	0.00066463	4.607e-004	1.443	1.493e-001
C(2)	0.00031644	2.849e-004	1.111	2.668e-001
A(1, 1)	0.00004848	8.910e-006	5.441	5.960e-008
A(2, 2)	0.00002712	3.817e-006	7.107	1.643e-012
ARCH(1; 1, 1)	0.08217097	1.249e-002	6.578	6.066e-011
ARCH(1; 2, 2)	0.11973859	7.401e-003	16.178	0.000e+000
GARCH(1; 1, 1)	0.82066580	2.493e-002	32.916	0.000e+000
GARCH(1; 2, 2)	0.74000839	2.582e-002	28.658	0.000e+000

Estimated Conditional Constant Correlation Matrix:

	HP	IBM
HP	1.0000	0.5421
IBM	0.5421	1.0000

Standard Errors:

	[, 1]	[, 2]
[1, ]	NA	0.01441
[2, ]	0.01441	NA

AIC(8) = -21902.74

BIC(8) = -21857.93

Quadro 12.8. Diagnóstico do modelo CCC(1,1) para o Exemplo 12.8.

Normality Test:

	Jarque-Bera	P-value	Shapiro-Wilk	P-value
HP	644.3	0	0.9806	7.782e-016
IBM	2116.9	0	0.9657	2.099e-021

Ljung-Box test for standardized residuals:

	Statistic	P-value	Chi <sup>2</sup> -d.f.
HP	18.21	0.1094	12
IBM	12.01	0.4451	12

Ljung-Box test for squared standardized residuals:

	Statistic	P-value	Chi <sup>2</sup> -d.f.
HP	9.755	0.6374	12
IBM	4.171	0.9801	12

Lagrange multiplier test:

	TR <sup>2</sup>	P-value	F-stat	P-value
HP	10.135	0.6042	0.9261	0.6265
IBM	4.122	0.9811	0.3755	0.9978

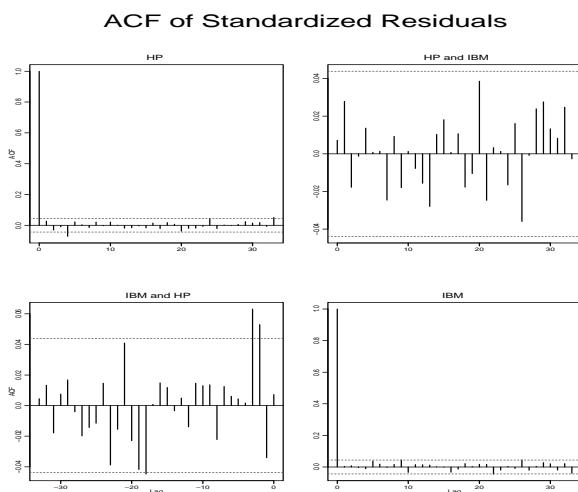


Figura 12.11: F.a.c. dos resíduos para o modelo CCC.

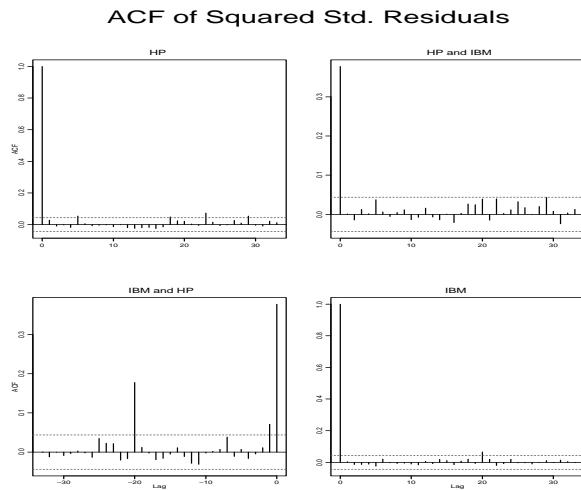


Figura 12.12: F.a.c. dos quadrados dos resíduos para o modelo CCC.

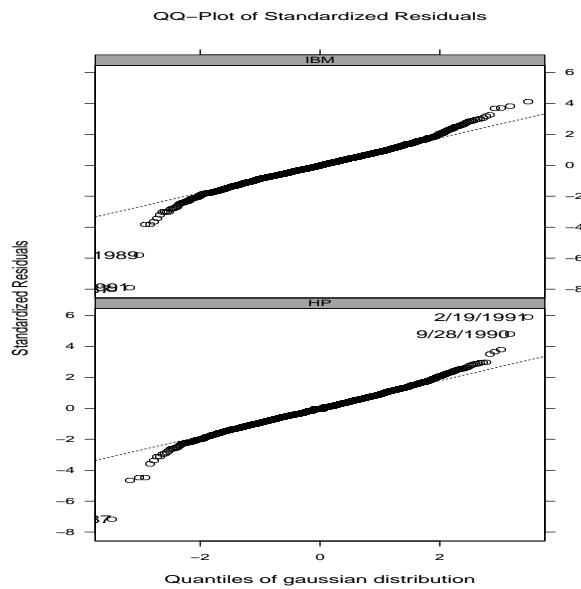


Figura 12.13: Gráfico  $Q \times Q$  para os resíduos do modelo CCC.

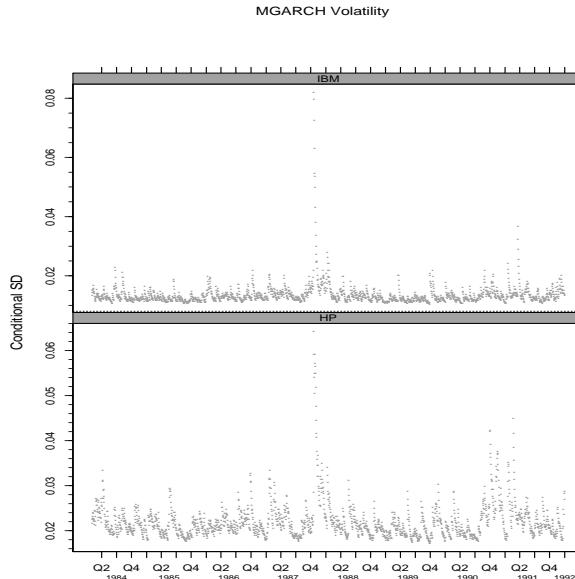


Figura 12.14: Volatilidades estimadas para o modelo CCC.

A suposição de correlação constante pode não ser realista para alguma aplicação. Engle (2002) e Tse e Tsui (2002), dentre outros, propuseram generalizações do modelo CCC, supondo a matriz de correlações condicionais dependendo do tempo. O modelo passa a ser chamado de modelo dinâmico com correlações condicionais (DCC).

**Definição 12.8.** (Engle, 2002). *O modelo DCC é definido por:*

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t, \quad (12.32)$$

onde  $\mathbf{D}_t$  é como em (12.30),  $h_{it}$  definida por qualquer GARCH univariado e

$$\mathbf{R}_t = (\text{diag}(\mathbf{Q}_t))^{-1/2} \mathbf{Q}_t (\text{diag}(\mathbf{Q}_t))^{-1/2}, \quad (12.33)$$

na qual  $\mathbf{Q}_t \geq \mathbf{0}$ , de orden  $n \times n$ , dada por

$$\mathbf{Q}_t = (1 - \alpha - \beta) \bar{\mathbf{Q}} + \alpha (\text{diag}(\mathbf{Q}_t))^{1/2} \mathbf{u}_{t-1} \mathbf{u}'_{t-1} (\text{diag}(\mathbf{Q}_t))^{1/2} + \beta \mathbf{Q}_{t-1}, \quad (12.34)$$

$\mathbf{u}_t = \mathbf{D}_t^{-1} \varepsilon_t$ ,  $\varepsilon_t$  como em (12.4),  $\bar{\mathbf{Q}}$  é a matriz de correlações incondicional de  $(\text{diag}(\mathbf{Q}_t))^{1/2} \mathbf{u}_t$ , de ordem  $n \times n$ ,  $\alpha \geq 0$ ,  $\beta \geq 0$ , escalares com  $\alpha + \beta < 1$ .

Tanto o modelo CCC como o modelo DCC podem ser estimados usando-se o pacote `rmgarch` do R. Veja o Problema 17. A seguir consideramos um exemplo utilizando a função `dccfit` do `rmgarch`.

**Exemplo 12.9.** Vamos considerar os retornos diários da Petrobras e do Ibovespa, de 02/01/1995 a 30/07/2010, com  $T = 3.847$  observações, e ajustemos um modelo DCC(1,1). As equações de volatilidade são dadas por:

$$\begin{aligned} h_{11,t} &= 0,000017 + 0,106085\varepsilon_{1,t-1}^2 + 0,869946h_{11,t-1}, \\ h_{22,t} &= 0,000012 + 0,101875\varepsilon_{2,t-1}^2 + 0,874577h_{22,t-1}. \end{aligned}$$

Os valores estimados de  $\alpha$  e  $\beta$  foram 0,042 e 0,953, respectivamente. A matriz de correlações condicionais estimadas é dada por

```
petro      ibv
petro 1.0000000 0.6539683
ibv   0.6539683 1.0000000
```

Os comandos necessários são, supondo normalidade:

```
> spec1=ugarchspec(distribution="std")
> mspec=multispec(c(spec1,spec1))
> fitspec=dccspec(mspec,VAR=TRUE,lag=1,dccOrder=c(1,1), model="DCC", distribution="mvt")
> dfit=dccfit(fitspec, petro.ibv)
```

Consulte o manual do pacote `rmgarch` no repositório R para detalhes.

## 12.4.2 Modelo Dinâmico Geral

Apresentamos, a seguir, uma extensão do modelo dinâmico geral de covariância (GDC), proposto por Kroner e Ng (1998), e que contém modelos de correlações condicionais dinâmicos.

**Definição 12.9.** O modelo GDC é definido por:

$$\mathbf{H}_t = \mathbf{D}_t \mathbf{R}_t \mathbf{D}_t + \boldsymbol{\Phi} \odot \boldsymbol{\Theta}_t, \quad (12.35)$$

onde:

- (a)  $\mathbf{D}_t = [d_{ijt}]$ ,  $d_{iit} = \sqrt{\theta_{iit}}$ , para todo  $i$ , e  $d_{ijt} = 0$ , para todo  $i \neq j$ ;
- (b)  $\boldsymbol{\Theta}_t = [\theta_{ijt}]$ , com

$$\theta_{ijt} = \omega_{ij} + \mathbf{a}' \boldsymbol{\varepsilon}_{t-1}' \mathbf{a}_j + \mathbf{g}_i' \mathbf{H}_{t-1} \mathbf{g}_j,$$

para todo  $i, j$ ,  $\mathbf{a}_i$ ,  $\mathbf{g}_i$ ,  $i = 1, \dots, n$  vetores paramétricos  $n \times 1$ ,  $\Omega = [\omega_{ij}] \geq \mathbf{0}$ , simétrica.

(c)  $\mathbf{R}_t$  especificada como no modelo DCC.

Elemento a elemento temos:

$$\begin{aligned} h_{iit} &= \theta_{iit}, \quad \forall i, \\ h_{ijt} &= \rho_{ijt} \sqrt{\theta_{iit} \theta_{jjt}} + \phi_{ij} \theta_{ijt}, \quad \text{para todo } i \neq j. \end{aligned} \quad (12.36)$$

O modelo GDC contém vários modelos MGARCH como casos especiais. Veja Bauwens et al. (2006) para detalhes.

## 12.5 Tópicos Adicionais

Nesta seção, discutiremos brevemente os tópicos de estimação, diagnóstico e previsão para modelos GARCH multivariados e faremos uma discussão breve sobre modelos localmente estacionários.

### [1] Estimação

Em (12.1) e (12.3), a média condicional  $\boldsymbol{\mu}_t$  e a matriz de covariâncias condicional  $\mathbf{H}_t$  dependem de um vetor de parâmetros  $\boldsymbol{\theta}_0$ . Dada uma amostra  $\mathbf{r}_t$ ,  $t = 1, \dots, T$ , queremos estimar o vetor  $\boldsymbol{\eta}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\xi}_0)$ , de ordem  $s \times 1$ , sendo que  $\boldsymbol{\xi}_0$  contém os parâmetros da distribuição dos erros  $\mathbf{z}_t$  em (12.4). Chamemos de  $f(\mathbf{r}_t | \boldsymbol{\eta}_0, \mathcal{F}_{t-1})$  a distribuição condicional do processo.

Estimadores de máxima verossimilhança são obtidos maximizando-se a função de verossimilhança (fv), supondo-se os  $\mathbf{z}_t$  independentes e identicamente distribuídos, que pode ser vista como uma função de quase-verossimilhança (QV) e obtemos estimadores de QMV. Escolhendo-se uma densidade  $g(\mathbf{z}_t | \boldsymbol{\theta}), \boldsymbol{\beta}$ , sendo  $\boldsymbol{\beta}$  um vetor de parâmetros *nuisance*, a fv fica

$$\mathcal{L}_T(\boldsymbol{\eta}) = \sum_{t=1}^T \log f(\mathbf{r}_t | \boldsymbol{\eta}, \mathcal{F}_{t-1}), \quad (12.37)$$

com

$$f(\mathbf{r}_t | \boldsymbol{\eta}, \mathcal{F}_{t-1}) = |\mathbf{H}_t|^{-1/2} g(\mathbf{H}_t^{-1/2}(\mathbf{r}_t - \boldsymbol{\mu}_t) | \boldsymbol{\eta}). \quad (12.38)$$

Em (12.41)  $|\mathbf{H}_t|^{-1/2}$  é o Jacobiano da transformação que leva  $\mathbf{z}_t$  em  $\mathbf{r}_t$ .

A distribuição mais empregada é a normal multivariada e (12.40) fica

$$\mathcal{L}_T(\boldsymbol{\eta}) = -\frac{1}{2} \sum_{t=1}^T \log |\mathbf{H}_t| - \frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}_t)' \mathbf{H}_t^{-1} (\mathbf{y}_t - \boldsymbol{\mu}_t), \quad (12.39)$$

embora saibamos que essa suposição de normalidade não seja adquada. Bollerslev e Wooldridge (1992) mostraram que estimadores consistentes de  $\boldsymbol{\theta}_0$  podem ser obtidos maximizando-se (12.42) com respeito a  $\boldsymbol{\theta}$ , mesmo que o processo não seja gaussiano. Esse estimador é chamado estimador de quase-máxima verossimilhança (QMV) e é consistente, desde que a média e a variância condicional sejam corretamente especificadas. Para outros detalhes, veja Jeantheau (1998) e Gourieroux (1997). Uma alternativa é usar uma densidade  $t$  de Student multivariada ou, então, misturas de normais multivariadas.

## [2] Diagnóstico

Os erros padronizados  $\mathbf{z}_t = \mathbf{H}_t^{-1/2} \boldsymbol{\varepsilon}_t$  devem satisfazer às condições (veja (12.4)):

- (a)  $\text{Cov}(\mathbf{z}_t) = \mathbf{I}_n$ ;
- (b)  $\text{Cov}(z_{it}^2, z_{jt}^2) = 0, \forall i \neq j$ ;
- (c)  $\text{Cov}(z_t^2, z_{j,t-k}^2) = 0, k > 0$ .

Segue-se que os resíduos padronizados devem satisfazer as mesmas condições, aproximadamente.

Podemos usar a estatística (*portmanteau*) de Box-Ljung-Pierce, dada por (9.33), que, sob a hipótese nula de que não haja efeitos ARCH, tem uma distribuição assintótica  $\chi^2(n^2 m)$ .

Testes de multiplicadores de Lagrange também podem ser usados, e esses têm poder maior que o teste portmanteau (quando a hipótese alternativa for correta). Veja Bollerslev et al. (1988) e Engle e Kroner (1995) para detalhes.

## [3] Previsão

Podemos obter previsões para a série multivariada  $\mathbf{r}_t$ , bem como para a sua matriz de covariâncias  $\mathbf{H}_t$ . Vamos considerar um exemplo, a saber, o modelo DVEC (1,1):

$$\mathbf{H}_t = \mathbf{A}_0 + \mathbf{A} \odot (\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}') + \mathbf{B} \odot \mathbf{H}_{t-1}. \quad (12.40)$$

Para a origem  $T$  e o horizonte de previsão  $h = 1$ , e sendo  $\mathcal{F} = \{\mathbf{H}_T, \mathbf{H}_{T-1}, \dots\}$ , temos

$$\hat{\mathbf{H}}_T(1) = E(\mathbf{H}_{T+1} | \mathcal{F}) = \mathbf{A}_0 + \mathbf{A} \odot E(\boldsymbol{\varepsilon}_T \boldsymbol{\varepsilon}_T') + \mathbf{B} \odot \mathbf{H}_T =$$

$$= \mathbf{A}_0 + \mathbf{A} \odot \boldsymbol{\varepsilon}_T \boldsymbol{\varepsilon}_T' + \mathbf{B} \odot \mathbf{H}_T,$$

e obtemos estimadores de  $\boldsymbol{\varepsilon}_T$  e  $\mathbf{H}_T$ , ao estimar o modelo. Para  $h = 2$ ,

$$\hat{\mathbf{H}}_T(2) = E(\mathbf{H}_{T+2}|\mathcal{F}) = \mathbf{A}_0 + (\mathbf{A} + \mathbf{B}) \odot E(\mathbf{H}_{T+1}),$$

e  $E(\mathbf{H}_{T+1})$  é obtida na previsão para  $h = 1$ . Iterando, obtemos  $\hat{\mathbf{H}}_T(h)$ , para  $h > 2$ .

A função predict para mgarch implementa o procedimento no S+FinMetrics.

#### [4] Modelagem via Processos Localmente Estacionários

A suposição de estacionariedade para a distribuição marginal da volatilidade é questionada, devido à presença de instabilidade de parâmetros, mudanças estruturais e mudanças de regimes em séries financeiras. Recentemente, tem havido tentativas para modelar a dependência de segunda ordem, variante no tempo, da volatilidade dessas séries.

A ideia é modelar essa dinâmica variante no tempo por meio de processos localmente estacionários. Veja Van Bellegem (2011) para uma resenha dessas especificações, tanto no caso univariado como no multivariado.

## 12.6 Problemas

1. Considere os dados (IBV e Petro) do Exemplo 12.3 e obtenha as correlações e correlações parciais das séries e ajuste algum modelo VAR(p). A seguir, ajuste um modelo DVEC(1,1). Faça o diagnóstico do modelo.
2. Mesma situação, só que agora ajuste um modelo BEKK(1,1) após ajustar o modelo VAR(p).
3. Ajuste modelos DVEC e BEKK às séries de retornos diários da Petrobrás e da Vale, de 31/08/1998 a 29/09/2010 ( $T = 2.991$ ).
4. Considere, agora, os retornos diários do Ibovespa, Petrobras e Vale, no período de 31/08/1998 a 29/09/2010 ( $T = 2.991$ ) e ajuste um modelo DVEC.
5. Para os mesmo dados do problema anterior, ajuste um modelo BEKK.
6. Para as séries de retornos diários da HP e da IBM, do Exemplo 12.6, ajuste modelos DVEC e BEKK.
7. Para as séries do Problema 3, ajuste um modelo fatorial baseado em componentes principais.
8. Mesmo problema, para os dados do Problema 4.

9. Para os dados do Problema 1, ajuste um modelo CCC (1,1).
  10. Ajuste um modelo CCC (1,1) para os dados da Petrobras e Vale do Problema 3.
  11. Mesmo problema para os dados do Ibovespa, da Petrobras e da Vale do Problema 4.
  12. Escreva o modelo ajustado no Exemplo 12.5. Faça os gráficos apropriados (f.a.c dos resíduos e quadrado dos resíduos), gráfico  $Q \times Q$  e gráfico das volatilidades estimadas. Comente.
  13. Para os modelos ajustados nos Exemplos 12.3-12.6, obtenha as previsões para  $h = 1, 2, 5$ .
  14. Considere modelos CCC-EGARCH e CCC-TGARCH para os dados dos Problemas 1, 3 e 6.
  15. Obtenha a Tabela 12.1. Use o algoritmo FastICA. Escreva explicitamente os modelos AR-GARCH obtidos na Tabela 12.3.
  16. Ajuste um modelo GO-GARCH (1,1) aos retornos da HP e da IBM, do Exemplo 12.6.
  17. Use o pacote `rmgarch` do R para ajustar um modelo DCC para os dados de retornos diários do Ibovespa e da Petrobras do Exemplo 12.3
-



---

# Referências

Abanto-Valle, C. A., Bandyopadhyayb, D., Lachos, V. H. and Guzman, I. R. E. (2010). Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions. *Computational Statistics and Data Analysis*, **54**, 2884–2893.

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. O. (1998). Volatility and correlation: Methods, models and applications. In ) *Risk Management and Analysis: Measuring and Modeling Financial Risk*, (C. O. Alexander, Ed.). New York: Wiley.

Alexander, C. O. (2001). *Market Models*. New York: Wiley.

Alexander, C. O. and Chibumba, A. M. (1997). Multivariate orthogonal factor GARCH. University of Sussex Discussion Papers in Mathematics.

Andersen, T. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, **39**, 885–905.

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P. (2000a). Exchange rates returns standardized by realized volatility are (nearly) Gaussian. *Multinational Finance Journal*, **4**, 159–179.

Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P. (2000b). Great realizati-

- ons. *Risk*, **13**, 105–108.
- Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, 42–55.
- Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 579–626.
- 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.
- Attanasio, O. (1991). Risk, time-varying second moments and market efficiency. *Review of Economic Studies*, **58**, 479–494.
- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, 5–59.
- Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **74**, 3–30.
- Bairamov, I., Kotz, S. and Kozubowski, T.J. (2003). A new measure of linear local dependence. *Statistics*, **37**, 243–258.
- Back, A. D. and Weigend, A. (1997). A first application of independent component analysis to extracting structure from stock returns. *International Journal of Neural Systems*, **8**, 473–484.
- Bandi, F. M. and Russell, J. R. (2003). Microstructure noise, realized volatility, and optimal sampling. Working paper.
- Bandi, F. M. and Russell, J. R. (2006). Volatility. Working paper, GSB, University of Chicago.
- 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.

Barndorff-Nielsen, O. E. and Shephard, N. (2002a). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, **17**, 457–477.

Barndorff-Nielsen, O. E. and Shephard, N. (2002b). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B*, **64**, 253–280.

Barndorff-Nielsen, O. E. and Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, **73**, 885–926.

Barndorff-Nielsen, O. E. and Shephard, N. (2004b). How accurate is the asymptotic approximation to the dsitribution of realized volatility? In *Identification and Inference for Econometric Models. A Festschrift in Honour of T.J. Rothenberg*. ( D.W. K. Andrews, J. Powell, P. A. Ruud, and J. H. Stock, Eds.). Cambridge: Cambridge University Press.

Bauwens, L. , Lubrano, M. and Richard, J. F. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford: Oxford University Press.

Bauwens, L. and Giot, P. (2000). The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks. *Annales d'Economie et de Statistique*, **60**, 117–150.

Bauwens, L. and Vereda, D. (2000). The stochastic conditional duration model: A latent factor model for the analysis of financial durations. *Journal of Econometrics*, **119**, 381–412.

Bauwens, L. and Giot, P. (2001). *Econometric Modelling of Stock Market Intraday Activity*. Boston: Kluwer Academic Publishers.

Bauwens, L. and Giot, P. (2003). Asymmetric ACD models: Introducing price information in ACD models. *Empirical Economics*, **28**, 709–731.

Bauwens, L., Laurent, S. and Rombouts, J. V. K. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, **21**, 79–109.

Bera, A.K. and Jarque, C.M. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters*, **7**, 313–318.

Bera, A. K. and Jarque, C. M. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, **55**, 163–172.

---

- 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.
- Berti, A. F. (2005). Estimação da Volatilidade Diária com Dados de Alta Frequência: Aplicações ao Cálculo do Valor em Risco do IBOVESPA. Dissertação de Mestrado, Instituto de Matemática e Estatística da Universidade de São Paulo.
- Beveridge, W. H. (1921). Weather and harvest cycles. *Economics Journal*, **31**, 429–452.
- Bierens, H. J. and Martins, L. F. (2010). Time-varying cointegration. *Econometric Theory*, **26**, 1453–1490.
- Bjerve, S. and Doksum, K. (1993). Correlation curves: Measures of association as functions of covariate values. *Annals of Statistics*, **21**, 890–902.
- 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. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate ARCH model. *Review of Economics and Statistics*, **72**, 498–505.
- Bollerslev, T. (2008). Glossary to ARCH (GARCH). <http://ssrn.com/abstract=1263250>.
- Bollerslev, T., Engle, R. F. and Wooldridge, J. M. (1988). A capital asset pricing

- model with time varying covariances. *Journal of Political Economy*, **96**, 116–131.
- Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**, 143–172.
- 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 ( R. F. Engle and D. L. McFadden, Eds.), 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.
- Bruscato, A., Morettin, P. A. and Toloi, C. M. C. (2010). Time-varying autoregressive conditional duration models. *Journal of Applied Statistics*, **37**, 847–864.
- 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.
- Cathy, W., Chen, S., Liu, F., Mike, K. and So, P. (2008). Heavy-tailed distributed

- threshold stochastic volatility models in financial time series. *Australian and New Zealand Journal of Statistics*, **50**, 29–51.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extremes*. London: Springer.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, **7**, 174–196.
- 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., Gençay, R., Muller, U. A., Olsen, R. B. and Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.
- Danielsson, J. and de Vries, C. G. (1997). Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, **4**, 241–257.
- Deheuvels, P. (1979). La fonction de dépendance empirique et ses propriétés, un test non paramétrique d'indépendance. *Bulletin de l'Académie Royale de Belgique, Classe des Sciences*, **65**, 274–292.
- DeJong, D. N. and Whiteman, C. H. (1991). Reconsidering trends and random walks in macroeconomic time series. *Journal of Monetary Analysis*, **28**, 221–254.
- De Vries, C. (1991), On the relation between GARCH and stable processes. *Journal of Econometrics*, **48**, 313–324.
- Dias, A. and Embrechts, P. (2009). Testing for structural changes in exchange rates dependence beyond linear correlation. *European Journal of Finance*, **15**, 619–637.
- Dias, A. and Embrechts, P. (2010). Modeling exchange rate dependence dynamics at different time horizons. *Journal of International Money and Finance*, **29**, 1687–1705.
- 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.

- Diebold, F. X. and Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4**, 1–21.
- Ding, Z. (1994). Time Series Analysis of Speculative Returns. Ph.D. Thesis, Department of Economics, University of California, San Diego.
- 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.
- Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, Ch. 8, 329–384.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, 987–1008.
- Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*,

**68**, 1–22.

Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, **20**, 339–350.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous GARCH. *Econometric Theory*, **11**, 122–150.

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., Ng, V. K, and Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. *Journal of Econometrics*, **45**, 213–238.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–1162.

Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, **22**, 367–381.

Engle, R. F. and Patton, A. J. (2001). What good is a volatility model. *Quantitative Finance*, **1**, 237–245.

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**, 383–417.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *The Annals of Statistics*, **21**, 196–216.

Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series. *Journal of Risk*, **5**, 25–54.

Fermanian, J.-D., Radulovic, D. and Wegkamp, M. (2004). Weak convergence of empirical copula processes. *Bernoulli*, **10**, 847–860.

Fernandes, M. and Gramming, J. (2005). A family of autoregressive conditional duration models. Preprint.

Fernandez, C. and Stell, M. (1998). On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.

Fonseca, E.L. (2017). Modelo de Cointegração Variando com o Tempo: Aborda-

---

gem via Ondaletas. Tese de Doutorado, Instituto de Matemática e Estatística da Universidade de São Paulo.

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.

Fréchet, M. (1951). Sur les tableaux de corrélation dont les margens son données. *Annales Université de Lyon, Section A*, **9**, 53–77.

Fuller, W.A. (1996). *Introduction to Statistical Time Series*. Second Edition. New York: Wiley.

García-Ferrer, A., González-Prieto, E. and Peña, D. (2012). A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting*, **28**, 70–93.

Garman, M. B. and Klass, M. J. (1980). On the estimation of security price volatility from historical data. *Journal of Business*, **53**, 67–78.

Gegembauer, H. V. (2010). Análise de Componentes Independentes com Aplicações em Séries Temporais Financeiras. Dissertação de Mestrado, IME-USP.

Genest, C., Masiello, E. and Tribouley, K. (2009). Estimating copula densities through wavelets. *Insurance: Mathematics and Economics*, **44**, 170–181.

Genest, C., Rémillard, B. and Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics*, **44**, 199–213.

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.

Giot, P. and Laurent, S. (2004). Modelling daily value-at-risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, **11**, 379–398.

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.

Gouriéroux, C. (1997). *ARCH Models and Financial Applications*. New York: Springer-Verlag.

---

- Gouriéroux, C. and Monfort, A. (1996). *Simulated-Based Econometric Methods*. Oxford: Oxford University Press.
- Granger, C. W. J. (1969). Investigating causal relationships by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.
- Granger, C. W. 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.
- Gouriéroux, C., Monfort, A. and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, **8**, 85–118.
- 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*, 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.
- Hansen, H. and Johansen, S. (1999). Some tests for parameter constancy in cointegrated VAR models. *Econometrics Journal*, **2**, 306–333.
- Harvey, A .C. (1998). Long memory in stochastic volatility. In *Forecasting Volatility in Financial Markets* (J. Knight and S. Satchell, Eds.), 307–320. Oxford: Butterworth-Heinemann.
- 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.
- Hoeffding, W. (1940). Masstabvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Matematik de Universität Berlin*, **5**, 179–233.

- Hoeffding, W. (1941). Masstabvariante Korrelationstheorie für diskontinuierliche Verteilungen. *Arkiv fr matematischen Wirtschaften un Sozialforschung*, **7**, 49–70.
- Hol, E. and Koopman, S. J. (2002). Stock index volatility forecasting with high frequency data. Discussion Paper N. 2002-068/4, Tinbergen Institute.
- 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.
- Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*. New York: Wiley.
- 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.
- Jaquier, E., Polson, N. G. and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat tails and correlated errors. *Journal of Econometrics*, **122**, 185–212.
- Jeantheau, T. (1998). Strong consistency of estimators of Multivariate ARCH models. *Econometric Theory*, **14**, 70–86.
- 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.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman and Hall.
- Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models. TR 166, Dept. Statistics, Univ. British Columbia.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Econo-*
-

- mic 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.
- Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package *stochvol*. *Journal of Statistical Software*, **69**, 1–30..
- 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, Y. S., Rachev, S. T., Bianchi, M. L. and Fabozzi, F. J. (2010). Tempered stable and tempered infinitely divisible GARCH models. *Journal of Banking and Finance*, **34**, 2096–2109.
- Kim, Y. S., Rachev, S. T., Bianchi, M. L. and Fabozzi, F. J. (2011). Time series analysis for financial market meltdowns. *Journal of Banking and Finance*, **35**, 1879–1891.
- Koenker, R. W. and Bassett, 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.
- Kojadinovic, I. (2008). *Copulas: Applications*. First Summer School on Copulas. Linz, Austria.
- Koopman, S. J., Harvey, A. C., Doornick, J. A. and Shephard, N. (2000). STAMP 5.0: Strucutural Time Series Analyser, Modeller and Predictor. London: Timberlake.
- Koutrouvelis, I. (1981). An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics - Simulation and Computation*, **10**, 17–28.
- Kotz, S. and Nadarajah, S. (2003). Local dependence functionsfor elliptically symmetric distributions. *Sankhya*, **65**, 207–223.
- Kroner, F. K. and Ng, V. K. (1998). Modelling asymmetric comovements of asset returns. *The Review of Financial Studies*, **11**, 817–844.
- Kruskal, W. H. (1958). Ordinal measures of association. *Journal of the American*
-

- Statistical Association*, **53**, 814–861.
- Kuan, C., Yeh, J. and Hsu, Y. (2009). Assessing Value at Risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics*, **150**, 261–270.
- 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.
- Latif, S. A. (2008). Medidas de Dependência Local para Séries Temporais. Tese de Doutorado, IME-USP, São Paulo.
- Latif, S. A. and Morettin, P. A. (2013). Estimation of a measure of correlation for independent samples and time series data. *Sankhyā, Series B*, **75**, 42–64.
- 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.
- Laurent, S., Bauwens, L. and Rombouts, J. V. K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, **21**, 79–109.
- Layton, A. P. (1984). A further note on the detection of Granger instantaneous causality. *Journal of Time Series Analysis*, **5**, 15–18.
- Lehmann, E. L. (1966). Some concepts of dependence. *Annals of Mathematical Statistics*, **37**, 1137–1153.
- 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.
- Lombardi, M. and Calzolari, G. (2008). Indirect estimation of  $\alpha$ -stable distributions and processes. *Econometrics Journal*, **11**, 1–15.
- Longstaey, 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.
- Lubrano, M. (1995). Testing for unit roots in a Bayesian framework. *Journal of Econometrics*, **69**, 81–109.
- 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.

Martens, M., Chang, Y.-C. and Taylor, S. J. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. *The Journal of Financial Research*, **25**, 283–299.

McCulloch, J. (1986). Simple consistent estimators of stable distributions parameters. *Communications in Statistics - Simulation and Computation*, **15**, 1109–1136.

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.

Menn C. and Rachev, S. T. (2009): Smoothly truncated stable distributions, GARCH-models, and option pricing. *Mathematical Methods in Operations Research*, **69**, 411-438.

Merton, R. C. (1973). Theory of option pricing. *Bell Journal of Economics and Management Science* , **4**, 141–183.

Miller, K. S. (1969). *An Introduction to the Calculus of Finite Differences and Difference Equations*. Dover.

Mills, T. C. (1999). *The Econometric Modelling of Financial Time Series*. Second Edition. Cambridge University Press.

Mittnik, S., Rachev, S. T. and Paolella, 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, Eds.), p. 79–110. Boston: Birkhäuser.

Mittnik, S., Rachev, S., Doganoglu, T. and Chenyao, D. (1999). Maximum likelihood

---

estimation of stable Paretian models. *Mathematical and Computer Modelling*, **29**, 275–293.

Morettin, P. A. (1984). The Levinson algorithm and its applications in time series analysis. *International Statistical Revue*, **52**, 83–92.

Morettin, P. A. and Toloi, C. M. C. (2006). *Análise de Séries Temporais*. Segunda Edição. São Paulo: Blucher-Associação Brasileira de Estatística.

Morettin, P.A., Toloi, C.M.C., Chiann, C. and de Miranda, J.C.S. (2010). Wavelet smoothed empirical copula estimators. *Brazilian Review of Finance*, **8**, 263–281.

Morettin, P. A., Toloi, C. M. C., Chiann, C. and de Miranda, J. C. S. (2011). Wavelet estimation of copulas for time series. *Journal of Time Series Econometrics*, **3**, Issue 3, Article 4.

Motta, A.C.O. (2001). Modelos de Espaço de Estados Não-Gaussianos e o Modelo de Volatilidade Estocástica. Dissertação de mestrado, Imecc-Unicamp.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions - Analyzing the dynamics of market components. *Journal of Empirical Finance*, **4**, 213–289.

Nadarajah, S., Mitov, K. and Kotz, S. (2003). Local dependence functions for extreme value distributions. *Journal of Applied Statistics*, **10**, 1081–1100.

Nelsen, R. B. (2003). Properties and applications of copulas: A brief survey. Proceedings of the First Brazilian Conference on Statistical Modelling in Insurance and Finance (J. Dahene, N. Kolev and P. A. Morettin, Eds.), 10–28. Institute of Mathematics and Statistics, University of São Paulo.

Nelsen, R. B. (2006). *An Introduction to Copulas*. Second Edition. New York: Springer.

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.

---

- Nowicka-Zagrajek, J. and Weron, A. (2001), Dependence structure of stable R-GARCH processes. *Probability and Mathematical Statistics* - Wroclaw University, Vol 21; Part 2, 371-380.
- Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics*, **140**, 425–449.
- Oomen, R. (2001). Using high frequency stock market index data to calculate, model and forecast realized return volatility. Working Paper 2001/6, European University Institute.
- 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.
- Park, J. Y. and Hahn, S. B. (1999). Cointegration regressions with time varying coefficients. *Econometric Theory*, **15**, 664–703.
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, **47**, 527–556.
- 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. (1991a). Optimal inference in cointegrated systems. *Econometrica*, **59**, 283–306.
- Phillips, P. C. B. (1991b). To criticize the critics: An objective Bayesian analysis of stochastic trends. *Journal of Applied Econometrics*, **6**, 333–364.
- 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 nonstationary 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.
- Quintos, C.E. (1997). Stability tests in error correction models. *Journal of Econometrics*, **82**, 289–315.
- Reiss, R.-D. and Thomas, M. (2001). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Second Edition. Birkhäuser.
- Rissanen, J. (1978). Modelling by shortest data description. *Automatica*, **14**, 465–471.
- 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.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, **39**, 1127–1140.
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13**, 341–360.
- Ruilova, J. C. (2007). Modelos ARCH Heterogêneos e Aplicações à Análise de Dados de Alta Frequência. Tese de Doutorado, IME-USP, São Paulo.
- 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.
- Sampaio, J. M. and Morettin, P. A. (2015). Indirect estimation of randomized generalized autoregressive conditional heteroskedastic models. *Journal of Statistical*
-

- Simulation and Computation*, **85**, 2702–2717.
- Sampaio, J. M. and Morettin, P. A. (2016). Stable randomized generalized autoregressive conditional heteroskedastic models. Submetido.
- Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models. In *Handbook of Financial Time Series*. (T. G. Anderson, R. A. Davis, J.-P. Kreiss, T. Mikosch (Eds.), 201–232, Springer.
- Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of The American Statistical Association*, **83**, 1097–1107.
- Stoll, H. and Whaley, R. (1990). Stock market structure and volatility. *Review of Financial Studies*, **3**, 37–71.
- Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, **6**, 231–252
- 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. (1982). Financial returns modelled by the product of two stochastic processes- a study of daily sugar prices 1961–79. In *Time Series Analysis: Theory and Practice 1*, (O. D. Anderson (Ed.), 203–226. Amsterdam: North Holland.
- Taylor, S. J. (2008). *Modeling Financial Time Series*. Second Edition. New York: Wiley.
- Taylor, S. J. and Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, **4**, 317–340.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. Second Edition. New York: Wiley.
- Tse, Y. K. and Tsui, A. K. C. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, **20**, 679–691.
- 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.
- Van Bellegem, S. (2011). Locally stationary volatility modelling. Working paper, CORE, Belgium.
- van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics*, **17**, 549–564.
- Venables, W. N. and Ripley, B. D. (2001). *Modern Applied Statistics with S-Plus*.
-

- Second Edition. New York: Springer.
- Weron, A. and Weron, R. (1995). *Lectures of Lévy  $\alpha$ -stable Variables and Processes*. Lectures Notes in Physics, 457, Springer-Verlag, pp. 379–392
- 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.
- Zhang, M. Y., Russell, J. R. and Tsay, R. S. (2001). A nonlinear autoregressive conditional duration model with applications to financial transaction data. *Journal of Econometrics*, **104**, 179–207.
- Zivot, E. (2005). *Analysis of High Frequency Financial Data: Methods, Models and Software*. Minicurso para a 11<sup>a</sup> Escola de Séries Temporais e Econometria. Vila Velha, ES, agosto de 2005.
- Zivot, E. and Wang, J. (2006). *Modelling Financial Time Series With SPLUS*. Second Edition. New York: Springer.



---

## Séries Usadas no Texto

As seguintes séries podem ser obtidas de <http://www.ime.usp.br/~pam/ef.html>.

CEMIG-D: Preços diários das ações da CEMIG, de 02/01/95 a 27/12/2000 ( $T = 1.499$ ).

d-cemig95.00.dat

DOW-D: Índices diários do Dow Jones Industrial Average, de 03/01/95 a 26/11/02 ( $T = 1.992$ ).

d-dow95.02.dat

HP-D: Log-retornos diários da HP de 02/02/1984 a 31/12/1992 ( $T = 2.000$ ).

base de dados do S+FinMetrics, hp.s.

IBM-D: Log-retornos da IBM de 02/02/1984 a 31/12/1992 ( $T = 2.000$ ).

base de dados do S+FinMetrics, ibm.s.

IBM-D: Log-retornos diários da IBM de 03/07/1962 a 31/12/1999 ( $T = 9.442$ ).

d-ibm62.99.dat

IBV-D: Índices diários do Ibovespa, de 02/01/95 a 27/12/2000 ( $T = 1.499$ ).

d-ibv95.00.dat

IBV2-D: Índices diários do Ibovespa, de 04/07/1994 a 29/09/2010 ( $T = 4.019$ ).

d-ibv94.10.dat

PETRO-D: Preços diários das ações da Petrobras, de 02/01/95 a 27/12/2000 ( $T = 1.499$ ).

d-petro95.00.dat

PETRO2-D: Preços diários das ações da Petrobras, de 18/08/1998 a 29/09/210 ( $T = 2.999$ ).

d-petro98.10

US/REAL-D: Taxas de câmbio diárias USD/Real, de 30/06/1994 a 01/07/1998 ( $T = 997$ ).  
d-usre94.98.dat

VALE-D: Preços diários das ações da VALE, de 31/08/1998 a 29/09/2010 ( $T = 2.991$ ).  
d-vale98.10.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

EUDO-ID: Taxas de câmbio Euro/Dolar, a cada 5 minutos, de 01/01/1999 a 31/12/2002 ( $T = 288.860$ ).  
id-eudo99.02.dat

IBV-ID: Índices intradiários do Ibovespa, a cada 15 minutos, de 06/04/1998 a 13/8/2003 ( $T = 37.961$ ).  
id-ibv98.03.dat

TLM-ID: Preços intradiários da Telemar PN, a cada 15 minutos, de 02/01/2002 a 31/3/2005 ( $T = 21.429$ ).  
id-tel02.05.dat

TLIBM-ID: Preços intradiários da Telemar PN, de 08/09/2004 a 10/09/2004 ( $T = 4096$ ) e preços intradiários da IBM, de 01/11/1990 a 08/11/1990 ( $T = 4.096$ ).

---

id-tlibm.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, 11
  - temporal, 11
- AIC, 82
  - correção, 83
  - de modelo AR, 82
  - de modelo ARMA, 82
- aleatório
  - passeio, 52
- Alta frequência, 209
- AR(p)
  - f.a.c., 71
  - f.a.c.p., 73
  - vetorial, 253
- ARCH, 1
  - estimação, 128
  - identificação, 127
  - modelos, 124
  - momentos, 125
  - previsão, 130
  - verificação, 129
- ARFIMA
  - estacionariedade, 165
  - estimação, 169
  - f.a.c., 166
  - modelos, 165
- ARIMA
  - identificação, 70
  - modelos, 69
- ARMA, 1
  - modelos, 69
- ARMA(p,q), 49
  - autocovariâncias, 50
  - estacionariedade, 50
  - f.a.c., 71
  - f.a.c.p., 73
  - invertibilidade, 50
- ask price, 210
- assimetria, 16
- auto-covariância
  - continuidade, 38
- auto-regressivos
  - AR(1), 42
- autocovariância
  - propriedades, 36
- autoregressivos
  - AR(p), 42
- autorregressivos
  - processos, 42
  - voriais, 253
- BIC, 83
  - de modelo ARMA, 83
- bid price, 210
- bid-ask spread, 210
- Bovespa, 217
- Browniano
  - movimento, 58
  - padrão, 58

- cópula, 309  
     $t$ , 321  
    Arquimediana, 320  
    de Clayton, 321  
    de Frank, 321  
    de Gumbel, 320, 321  
    definição, 314  
    densidade, 317  
    empírica, 327  
    Gaussiana, 319, 321  
    produto, 315  
cópulas, 309  
    para séries temporais, 330  
    paramétricas, 325  
    valor em risco, 332
- causalidade  
    de Granger, 266  
    instantânea, 269  
    unidirecional, 267
- Cholesky  
    decomposição, 279
- cointegração  
    dinâmica, 305  
    vetor de, 284
- cointegrados  
    processos, 283
- comonotônicas  
    cópulas, 315  
    variáveis, 310
- componentes  
    independentes, 360
- contramonomotônicas  
    cópulas, 315  
    variáveis, 310
- correlação  
    coeficiente, 310
- correlações  
    matriz de, 249
- covariâncias  
    estimação, 251  
    matriz de, 247
- curtose, 16
- curva  
    de impacto de informação, 139
- CVaR, 201
- dados  
    não sincronizados, 210  
    tick-by-tick, 209
- DAF, 209
- DF  
    teste, 75
- diagnóstico  
    de modelos ACD, 230  
    de modelos ARIMA, 90  
    de modelos MGARCH, 371
- distribuição  
    de erro generalizada, 155  
    de Pareto generalizada, 207  
    de valores extremos, 194, 204  
    estável, 25  
     $t$  assimétrica, 155
- distribuições  
    especiais, 155  
    finito-dimensionais, 33
- durações, 212
- Durbin-Levinson  
    algoritmo, 73
- ergodicidade, 13
- erro  
    de previsão, 94  
    de previsão, variância, 94
- estáveis  
    distribuições, 25
- estacionário  
    estrito, 35  
    fraco, 36  
    processo, 31
- estacionários  
    lineares, 40
- estacionariedade  
    condição, 44
- estatística

- R/S, 161
- estimação, 84
  - coeficiente de correlação, 311
  - coeficiente de Kendall, 312
  - de modelos ACD, 229
  - de modelos AR, 85
  - de modelos MA, 86
  - de modelos MGARCH, 370
  - de modelos VAR, 259
- estimadores
  - de MV, 326
  - de MV exatos, 87
  - de pseudo-MV, 328
  - variância dos, 88
- EViews, 22, 89, 129, 259
- Eviews, 23
- EWMA, 20
- FIEGARCH
  - modelos, 175
- FIGARCH
  - modelos, 175
- Fréchet, 204
- função
  - de autocorrelação, 38
  - de autocorrelação parcial, 71
  - de autocovariância, 34
- funcional
  - teorema limite central, 63
- GARCH
  - estimação, 135
  - extensões, 139
  - identificação, 135
  - modelos, 134
  - previsão, 135
- GARCH(1,1), 135
- Garman-Klass
  - volatilidade, 179
- GPH
  - estimador, 164
  - procedimento, 163
- Gumbel, 204
- GVE, 204
  - inferência, 206
- heteroscedasticidade, 18
- Hurst
  - coeficiente, 157
- Ibovespa
  - dados intradiários, 218
- ICA, 360
- identificação
  - de modelos ACD, 229
  - de modelos ARIMA, 70
  - procedimento de, 74
- IFM, 327
- índice
  - de estabilidade, 26
- inferência
  - indireta, 236
- infinitamente
  - divisível, 28
- infinitamente divisível
  - distribuição, 28
- Johansen
  - teste de, 295
- Kendall
  - coeficiente, 312
- limites
  - de Fréchet-Hoeffding, 316
- liquidez, 244
- log-retorno, 8
  - da carteira, 10
- longa
  - memória, 157
- média
  - de um processo, 34
- MA(q), 46
  - f.a.c., 71

- f.a.c.p., 73
- invertibilidade, 48
- martingales, 64
  - convergência, 65
  - diferença de, 65
- matrizes, 273
  - autovalores, 276
  - autovetores, 276
  - decomposição, 276
  - determinante de, 273
  - particionadas, 274
  - ponto de, 273
  - produto de Kronecker, 275
  - traço de, 273
  - vectrização, 275
- medidas de dependência, 309
  - de caudas, 313
  - globais, 310
  - locais, 335
- memória
  - curta, 157
  - longa, 157
- memória longa, 157
  - testes, 160
- mercado
  - de taxa de câmbio, 245
  - order driven, 245
  - price driven, 244
- metadistribuições, 323
- Minitab, 89
- modelo
  - ACD, 227
  - ACD(1,1), 229
  - auxiliar, 236
  - CCC, 362
  - DCC, 368
  - de correção de erros, 288
  - de duração condicional, 226
  - de interesse, 236
  - demandas por moeda, 306
  - dinâmico geral, 369
  - GICA-GARCH, 360
- modelos
  - ARCH, 124
  - ARFIMA, 165
  - ARIMA, 55
  - ARMA vetoriais, 264
  - autorregressivos vetoriais, 253
  - BEKK, 348
  - com CP, 356
  - de volatilidade estocástica, 143
  - EGARCH, 139
  - estruturais, 279
  - EWMA multivariados, 342
  - fatoriais, 354
  - FIEGARCH, 175
  - FIGARCH, 175
  - GARCH, 134
  - GARCH multivariados, 341
  - HAR, 243
  - HARCH, 243
  - IGARCH, 136
  - MGARCH, 342
  - mistas, 87
  - multivariados, 247
  - ortogonais, 355
  - ortogonais generalizados, 356
  - R-GARCH, 233
  - sazonais, 97
  - sazonais multiplicativos, 99
  - TGARCH, 142
  - VARMA, 265
  - VEC, 344
  - VEC diagonais, 344
- momentos, 34
  - média, 34
- Movimento Browniano
  - geométrico, 61
  - geral, 60
- MVE, 143
  - estimação, 146
  - forma canônica, 143
  - formulação de Jaquier, 143
  - propriedades, 144

- paridade
  - poder de compra, 285
- posto
  - de cointegração, 291
- postos
  - padronizados, 327
- preliminares, 1
- previsão
  - com modelos ARIMA, 92
  - de modelos MGARCH, 371
  - erro, 94
  - intervalo de confiança, 95
- processo
  - autossimilar, 60
  - com memória longa, 157
  - complexo, 39
  - contínuo em mq, 38
  - especificação de um, 33
  - estocástico, 31
  - gaussiano, 36
  - integrado, 55
  - linear, 40
  - linear geral, 51
  - linear multivariado, 251
  - não estacionário, 52
  - puramente aleatório, 41
  - realização, 32
  - trajetória, 32
  - trend-stationary, 53
- processos
  - de Lévy, 68
  - estacionários, 31
  - estocsticos, 31
  - localmente estacionários, 372
- R, 23
- raiz unitária, 105
  - teste ADF, 109
  - teste DF, 108
  - teste PP, 114
- resíduos, 90
- retorno
  - intradiário, 218
  - overnight, 218
  - retornos, 7
    - agregação, 10
    - anualização, 8
    - brutos simples, 7
    - compostos, 8
    - distribuição, 12
    - fatos estilizados, 18
    - intradiários, 209
    - líquidos simples, 7
    - multiperíodos, 8
  - risco, 181
    - medidas coerentes, 201
    - valor em, 181
  - RiskMetrics, 20
  - rmgarch, 368
  - rpubs, 23
  - ruído branco, 41
    - discreto, 41
  - S+FinMetrics, 22, 23, 129
  - SARIMA
    - modelos, 99
  - sazonais
    - modelos, 97
  - sazonal
    - diferença, 97
  - sazonalidade
    - de retornos, 224
  - sequência
    - i.i.d.*, 41
    - aleatória, 40
  - simulação
    - de v.a estável, 28
  - Spearman
    - coeficiente, 313
  - SPlus, 22
  - STAMP, 22, 147
  - stochvol
    - pacote do R, 147
  - tendências

- comuns, 286
- Teorema
  - de Sklar, 315
- teste
  - ADF, 109
  - de autocorrelação residual, 90
  - de Box-Pierce-Ljung, 91
  - de Engle e Granger, 293
  - de normalidade, 29
  - DF, 106
  - DF-extensões, 110
  - Phillips-Perron, 114
- testes
  - de cointegração, 292
- unitárias
  - raízes, 75, 105
- valor
  - em risco, 181
- valores extremos
  - teoria, 203
- VAR
  - construção, 258
  - diagnóstico, 260
  - equações de Yule-Walker, 257
  - estacionário, 257
  - estimação, 259
  - identificação, 258
  - modelos, 253
  - previsão, 260
- VaR, 183
  - condicional, 201
  - definição, 183
  - RiskMetrics, 184
  - usando modelos GARCH, 189
  - usando POT, 196
  - usando quantis empíricos, 193
  - usando TVE, 193
- variância
  - realizada, 223
- variação