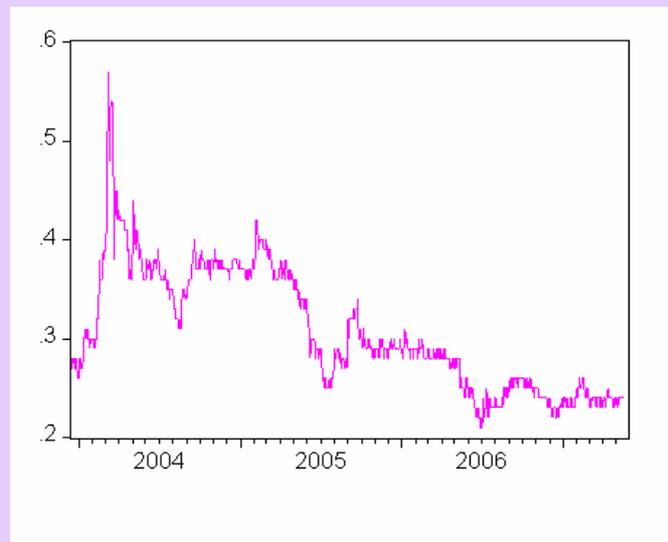


Esmeralda Gonçalves

Nazaré Mendes Lopes

Séries Temporais



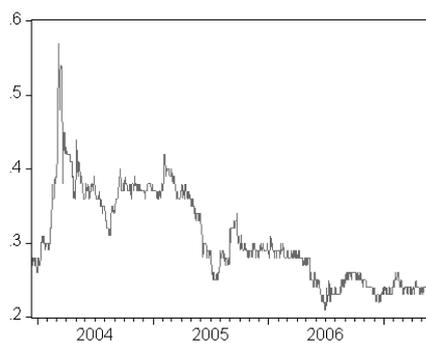
Modelações lineares e não lineares

2ª edição

Sociedade Portuguesa de Estatística

Esmeralda Gonçalves Nazaré Mendes Lopes

Séries Temporais



Modelações lineares e não lineares

2ª edição

Sociedade Portuguesa de Estatística

Séries temporais
Modelações lineares e não lineares

2^a Edição

E. Gonçalves, N. Mendes Lopes

16 de Janeiro de 2008

FICHA TÉCNICA:

Título: Séries Temporais. Modelações lineares e não lineares (2ª Edição)

Autores: Esmeralda Gonçalves e Nazaré Mendes Lopes

Editora: Sociedade Portuguesa de Estatística

Produção gráfica, Impressão e Distribuição: Instituto Nacional de Estatística

Tiragem: 500 exemplares

ISBN: 978-972-8890-16-2

Depósito Legal: 269869/08

Para

Tó-Zé, Emanuel e Rodrigo

Nando, Leo e Ro

NOTA DA SPE

Os manuais que a Sociedade Portuguesa de Estatística (SPE) tem vindo a editar anualmente para suporte do mini-curso que decorre no início de cada congresso Anual da SPE têm sido muito solicitados, quer por docentes dos ensinos universitário e politécnico para serem usados como textos de apoio de várias unidades curriculares de estudos graduados e pós-graduados, quer por investigadores teóricos e aplicados para apoio ao seu trabalho de investigação, quer ainda por estatísticos profissionais para utilização no exercício da sua actividade. Temos mesmo tido solicitações provenientes de outros países de língua portuguesa ou de colegas estrangeiros que entendem o português. Os poucos exemplares que sobram dos Congressos rapidamente se têm esgotado, pelo que a Direcção da SPE sentiu a necessidade de proceder a segundas edições para melhor servir a comunidade. É também a oportunidade para os autores, escolhidos de entre os mais reputados especialistas nacionais da respectiva área, poderem corrigir as inevitáveis gralhas da primeira edição e introduzirem algumas melhorias e actualizações que considerem justificadas.

O Instituto Nacional de Estatística (INE) é uma instituição com a qual temos tido intensa e frutuosa colaboração e que, nos últimos anos, tem impresso os manuais dos mini-cursos. O INE aceitou ser nosso parceiro nesta iniciativa editorial, encarregando-se da impressão e distribuição destas segundas edições, o que nos dá a garantia de um fácil acesso por parte dos interessados. Estamos muito gratos ao INE por mais esta valiosa colaboração.

Note-se que se trata de manuais sobre temas especializados que dificilmente poderiam ser editados por editoras comerciais (ou só poderiam sê-lo a um custo unitário elevado), uma vez que o público a que se

dirigem é relativamente limitado. Graças à generosidade dos autores, que prescindem dos seus direitos de autor, ao formato de bolso (que, além de cómodo, permite menores custos de produção) e ao facto de a SPE e o INE apoiarem estas edições, é possível elas chegarem ao leitor a preços muito moderados.

Com esta iniciativa, a SPE e o INE põem ao dispor da comunidade académica e da comunidade estatística e dos utilizadores de metodologias estatísticas textos em língua portuguesa de elevada qualidade científica e pedagógica cobrindo temas importantes e actuais das Probabilidades e da Estatística, que podem funcionar como manuais escolares para o ensino superior e como apoio à investigação e à prática profissional.

Esta é a primeira das segundas edições que nos propomos disponibilizar. Não houve para isso qualquer razão especial, a não ser o facto de, logo a seguir à decisão da Direcção da SPE de lançar esta iniciativa, terem, por mera coincidência, as autoras solicitado autorização para lançarem uma segunda edição da obra (esgotada há já algum tempo), que aliás têm usado há vários anos nas suas aulas e daí colhido uma preciosa experiência muito útil no trabalho de revisão. Prontamente aceitaram que essa segunda edição fosse também editada pela SPE. Por isso lhes estamos gratos, bem como pela sua paciência já que, tendo esta obra servido de cobaia para esta iniciativa, houve alguns atrasos decorrentes da necessidade de definir e acordar questões organizativas e processuais. Trata-se, naturalmente, de uma importante obra que nos orgulhamos de editar e que esperamos possa ser muito útil ao leitor.

O Presidente da SPE
Carlos A. Braumann

Prefácio à segunda edição

... Todo o mundo é composto de mudança,
Tomando sempre novas qualidades ...

Luís de Camões, Sonetos.

A monografia que agora se reedita foi inicialmente publicada no âmbito do XI Congresso Anual da Sociedade Portuguesa de Estatística, como texto de suporte ao mini-curso que ministramos no referido Congresso.

O trabalho de ensino e investigação que temos continuado a desenvolver no contexto dos temas aqui tratados levou-nos a propôr a sua reedição numa forma mais completa e melhorada sobretudo no que respeita às aplicações. Neste sentido, foram acrescentados, nomeadamente nos capítulos 3 e 5, vários exemplos que permitem não só evidenciar aspectos práticos de aplicação da teoria mas também abordar, em contextos particulares, novos desenvolvimentos teóricos. A teoria tratada nesta obra é ainda ilustrada no final dos capítulos 4 e 6, relativos à análise estatística de séries temporais, através da modelação de dados reais recorrendo ao software estatístico EViews particularmente dirigido para a análise de tal tipo de dados.

Mantendo um dos objectivos já referido na primeira edição, concretamente que este texto fosse suficientemente auto-contido, complementamos alguns capítulos com anexos sobre temas particularmente relevantes para a sua compreensão e desenvolvimento. Destes anexos salientamos o relativo à Esperança Condicional onde procuramos resumir a informação fundamental sobre um tema de enorme utilidade não só no contexto das séries temporais mas dos processos estocásticos em geral.

À Sociedade Portuguesa de Estatística, que nos deu a honra de aceitar editar de novo esta obra, reafirmamos o nosso propósito de

apresentar um trabalho que vá ao encontro da exigência de qualidade dos textos da SPE e expressamos o nosso vivo e reconhecido agradecimento.

Coimbra, Janeiro de 2008

Prefácio

... Corre o tempo e não se cansa
em cada tempo outro tempo
todo o tempo é de mudança ...

Manuel Alegre, Guitarras do meu país

A modelação matemática de conjuntos de dados evoluindo de forma aleatória com o tempo, habitualmente designados por séries temporais ou cronológicas, preocupa os estudiosos da descrição matemática do *não determinismo* desde o advento, no século XVII, da ciência probabilista. Contudo, só em 1927 surgem, devidos a Yule, os primeiros modelos concretos cumprindo tal objectivo, designadamente os modelos auto-regressivos. Daí em diante a busca de modelos que melhor descrevem esta ou aquela realidade temporal tem sido permanente, tendo tais estudos sido dominados, até finais dos anos 70, pelos modelos lineares, os modelos auto-regressivos médias móveis. Só a partir desta data surgem na literatura os primeiros modelos estocásticos capazes de ter em conta certas características de não linearidade presentes nos sistemas reais a descrever, nomeadamente os modelos bilineares e os condicionalmente heteroscedásticos.

No presente livro caminha-se, de acordo com tal evolução histórico-científica, do linear para o não linear tendo o ênfase sido posto, em qualquer dos casos, no estudo de questões de natureza probabilista. De facto, só o conhecimento correcto das propriedades probabilistas dos modelos permitirá encontrar algum que venha a confirmar uma boa adequação aos dados em análise.

Assim, de forma detalhada e sistemática, são estudadas ao longo deste texto as principais questões associadas ao tratamento probabilista de modelos lineares e não lineares de séries temporais, com particular ênfase para os problemas de estacionaridade, forte e fraca, e de inversibi-

lidade. Os capítulos 3 e 5, relativos respectivamente aos modelos lineares e não lineares, têm este objectivo.

Sem descurar a análise estatística de tais modelos, damos-lhe, no entanto, menor destaque centrando-nos, quer no contexto linear quer não linear, na apresentação das estatísticas e metodologias de estimação mais usuais e correspondentes propriedades. No que concerne a análise estatística dos modelos lineares, a que dedicamos o capítulo 4 deste livro, abordamos ainda de forma breve questões ligadas à previsão estatística. Por outro lado, no âmbito da análise estatística de modelos não lineares de séries temporais, desenvolvida ao longo do capítulo 6, é apresentado um teste estatístico não clássico com grandes potencialidades de aplicação à generalidade dos modelos de séries temporais.

A necessidade, por nós sentida, de elaborar um trabalho tão auto-contido quanto possível levou-nos a iniciá-lo com dois capítulos generalistas onde são estudadas as definições e propriedades básicas dos processos estocásticos, isto é, dos modelos probabilistas gerais subjacentes aos sistemas dinâmicos não deterministas.

Este trabalho vem na sequência de várias disciplinas de Licenciatura e Mestrado, do foro dos Processos Estocásticos, que temos vindo a leccionar no Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra. Mais concretamente, resumem-se aqui os principais desenvolvimentos teóricos de um curso de Mestrado recentemente leccionado. Não lhe é por isso alheia a dialéctica ensino-aprendizagem que desenvolvemos com os nossos alunos.

O nosso agradecimento será assim sempre, primordialmente, para os que nos ouvem e também, agora, para os que nos lêem ...

Coimbra, Julho de 2003

Índice

Introdução	xi
Considerações gerais	xi
Objectivos da análise de séries temporais	xviii
1 - Processos Estocásticos	1
1.1 Conceitos básicos	1
1.2 Teorema de Kolmogorov	4
1.3 Processos de segunda ordem	7
1.3.1 Definição. Propriedades	7
1.3.2 Exemplos de processos de segunda ordem	10
1.4 Processos estacionários	13
1.4.1 Processos fortemente estacionários	14
1.4.2 Processos estacionários no sentido fraco	16
1.5 Espaços lineares	23
1.5.1 Operadores atraso e avanço	24
1.5.2 Melhor previsão linear baseada no passado	29
1.6 Anexo - Espaços de Hilbert	32
2 - Resumo das ligações temporais	41
2.1 Autocovariância e autocorrelação	41
2.2 Medida e densidade espectrais	50
2.3 Autocorrelações parciais	63
2.4 Anexo - Esperança condicional	74
3 - Processos ARMA unidimensionais	93
3.1 Processos estocásticos regulares	93
3.2 Processos auto-regressivos	99
3.2.1 Definição e existência	99
3.2.2 Processo AR(p)	100
3.2.3 Resumo das ligações temporais	114
3.2.4 Caracterização dos processos AR	120
3.3 Processos médias móveis	126
3.3.1 Definição	126
3.3.2 Multiplicidade de representações	127
3.3.3 Representação AR de um processo MA(q)	127

3.3.4	Resumo das ligações temporais	131
3.3.5	Caracterização dos processos MA	136
3.4	Processos auto-regressivos médias móveis	137
3.4.1	Definição	137
3.4.2	Resumo das ligações temporais	143
3.4.3	Caracterização dos processos ARMA	146
3.4.4	Exemplos de aplicação	148
3.4.5	Generalização dos modelos ARMA	157
3.4.6	Representação de espaços de estado de um modelo ARMA	162
3.5	Anexo - Inversibilidade de $1-\lambda L$	164
3.6	Anexo - Equações às diferenças	165
4	Inferência em modelos lineares	169
4.1	Propriedades dos estimadores	170
4.1.1	Média empírica	170
4.1.2	Autocovariâncias empíricas	172
4.1.3	Periodograma	175
4.2	Previsão em modelos ARIMA	182
4.3	Modelação de uma série real	184
5	Modelações não lineares	191
5.1	Heteroscedasticidade condicional	192
5.1.1	Definição geral e propriedades básicas	194
5.1.2	Diversas modelações condicionalmente heteroscedásticas	200
5.1.3	Modelos com erros condicionalmente heteroscedásticos	225
5.2	Modelos bilineares	238
5.2.1	Definição e considerações gerais	239
5.2.2	Modelo bilinear simples	240
6	Inferência em modelos não lineares	255
6.1	Estimação num modelo com erros ARCH	256
6.2	Um teste para modelos bilineares	261
6.2.1	Ergodicidade de uma série temporal	262
6.2.2	O modelo: hipóteses e propriedades gerais	264

6.2.3 Um procedimento de decisão	265
6.2.4 Implementação do procedimento de decisão como um teste	270
6.2.5 Estudo de simulação	272
6.3 Modelação de uma série real	282
Bibliografia	293

Introdução

Considerações gerais

São poucas as áreas científicas que não se vêm hoje confrontadas com o estudo de fenómenos evoluindo com o tempo.

A descrição, explicação, controlo ou previsão de fenómenos como, entre outros, a evolução de um índice bolsreiro, a variação da inflação ou da taxa de desemprego, a variação da temperatura ou da pluviometria numa determinada região do globo, a variação da população (segundo a idade, o sexo, ...) num certo país, o movimento de partículas, sujeitas a impactos aleatórios, num fluido ou a emissão de fótons gerados em desintegração radioactiva têm vindo a interessar especialistas de áreas tão diversas como a Economia, a Meteorologia, as Ciências Exactas ou as Ciências Experimentais.

Originalmente as análises estatísticas diziam apenas respeito à descrição de conjuntos numerosos de dados, designados genericamente por informação. Com o nascimento das Probabilidades, ao longo dos séculos XVII, XVIII e XIX, foram dados os primeiros passos no domínio da Estatística Matemática cujo objectivo último é o de conhecer o modelo aleatório subjacente a tais observações.

Os probabilistas construíram uma importante teoria matemática, a teoria dos processos estocásticos, os quais permitem analisar as propriedades de um conjunto de variáveis indexadas pelo tempo.

A ideia natural de repetição de observações independentes e idênticas, presente na Estatística clássica, não é compatível, em geral, com fenómenos evoluindo com o tempo. De facto, a hipótese

de independência das observações perde sentido neste contexto pois estamos perante variáveis cujo comportamento pode depender de um segmento de observações entre dois instantes conhecidos.

De facto, uma das principais questões que se levantam aquando da análise de um fenómeno que evolui ao longo do tempo é a da identificação do tipo de evolução. Outro aspecto fulcral é o da previsão de valores futuros. Tais estudos só se tornam possíveis se a evolução for regrada, isto é, se não for totalmente anárquica. Assim, a hipótese de independência da estatística clássica é, em particular, substituída por hipóteses de estabilidade no tempo (estacionaridade) ou de precisão da tendência de evolução (linear, quadrática, crescente ou decrescente, ...).

O estudo da evolução ao longo do tempo das variáveis que descrevem tais fenómenos pressupõe a obtenção de um certo número de observações dessas mesmas variáveis. Tais observações são obtidas, por exemplo, por meio de inquéritos ou registos.

Em domínios como a meteorologia ou a física, certas características (temperatura, higrometria, movimento de uma partícula no seio de um fluido) são observadas por meio de aparelhos físicos. Nestes casos, o índice dos instantes de observação percorre um intervalo real pois estamos perante uma infinidade de observações.

Noutros domínios as observações são recolhidas em datas em geral equidistantes, isto é, estão disponíveis por exemplo uma vez por mês, uma vez por trimestre, etc. Identificamos então as datas de observação com os inteiros $1, 2, \dots, T$.

A sucessão de observações de uma variável é chamada série temporal ou série cronológica.

Uma série temporal é, assim, um conjunto de dados evoluindo com o tempo onde a ordem de recolha desempenha um papel importante. Este facto é salientado quando se utilizam representações gráficas das séries observadas.

A representação gráfica de uma série temporal é um auxiliar precioso e muito utilizado na análise da sua evolução.

A figura 1, que representa a evolução, entre 1990 e 2001, e previsão para os anos 2002 e 2003, do Produto Interno Bruto (PIB) em Portugal e na União Europeia (UE) (Dinheiro&Direitos, n^o 51, EdiDeco, Abril / Maio 2002), é disso um exemplo.

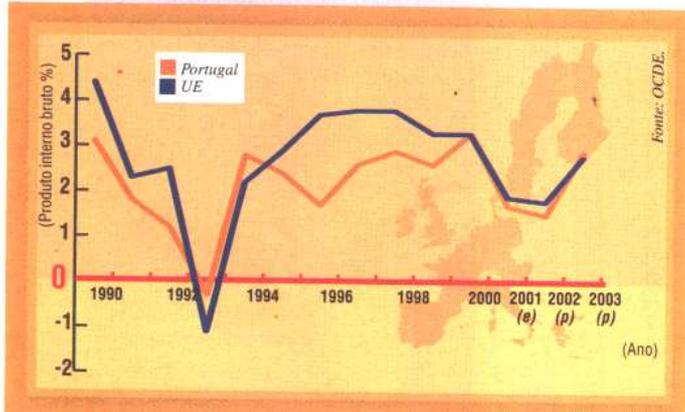


Fig. 1 - Evolução e previsão do PIB em Portugal e na UE entre 1990 e 2003

Os primeiros registos conhecidos de representações gráficas de observações temporais estão ligados à Astronomia. Em particular, o gráfico da figura 2, representando a inclinação das órbitas dos planetas em função do tempo, data do século X ou XI e é considerado, segundo Kendall (1973), o mais antigo diagrama temporal do mundo ocidental.

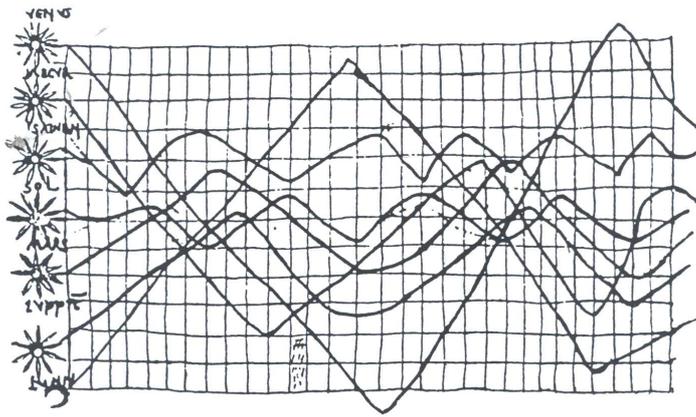


Fig. 2 - Inclinação das órbitas dos planetas do sistema solar (séc. X ou XI)

Foi também na Astronomia que surgiram as primeiras séries temporais constituídas expressamente para fins de análise.

Tal facto vem ao encontro do entendimento comumente aceite da Astronomia ter sido, quer na Antiguidade Clássica quer no período do Renascimento, fonte de importantes descobertas. Referir-se-á a propósito que foi um conjunto de dados sobre o movimento dos planetas obtido por Brahe (1546-1601) que levou à formulação das bem conhecidas leis de Kepler (1571-1630).

Só nos séculos XVIII e XIX, e com o desenvolvimento dos conhecimentos matemáticos, foi possível ultrapassar a visualização gráfica. Esta descrição terá sido complementada inicialmente com uma análise frequencial e, posteriormente, com a abordagem no domínio dos tempos.

O exemplo seguinte (Gouriéroux, Monfort, 1990) diz respeito a uma série dos nossos dias e pretende ilustrar alguns aspectos ou ideias gerais que podem surgir quando somos confrontados com a análise de uma série temporal.

Dispomos do valor do índice mensal dos preços no consumidor relativos ao período Janeiro de 1970 a Dezembro de 1978 (com base 100 em Julho de 1970), apresentado no quadro 1.

	1970	1971	1972	1973	1974	1975	1976	1977	1978
Jan	97.9	102.5	108.3	115.5	127.4	145.9	159.9	174.3	190.3
Fev	98.2	103.0	108.9	115.8	129.1	147.0	161.0	175.5	191.7
Mar	98.5	103.4	109.4	116.4	130.6	148.2	162.4	177.1	193.4
Abr	99.0	104.0	109.8	117.2	132.7	149.5	163.8	179.4	195.5
Mai	99.4	104.7	110.4	118.3	134.3	150.6	164.9	181.1	197.4
Jun	99.8	105.1	111.0	119.2	135.8	151.7	165.6	182.5	198.9
Jul	100	105.6	111.9	120.2	137.5	152.8	167.2	184.1	201.5
Ag	100.4	106.0	112.5	121.0	138.6	153.8	168.4	185.1	202.5
Set	100.8	106.5	113.2	122.1	140.1	155.1	170.2	186.7	203.8
Out	101.2	107.1	114.2	123.4	141.8	156.3	171.8	188.2	208.7
Nov	101.6	107.5	114.9	124.5	149.1	157.3	173.2	188.9	206.8
Dez	101.9	108.0	115.5	125.3	144.3	158.2	173.8	189.4	207.8

Quadro 1 - Índice mensal dos preços no consumidor

A representação gráfica destes valores é apresentada na figura 3, onde I_t representa o índice mensal dos preços no consumidor na data t ($t = 1, 2, \dots, 108$). A escolha de uma escala reduzida para os tempos permite pôr em evidência a evolução da série a médio prazo.

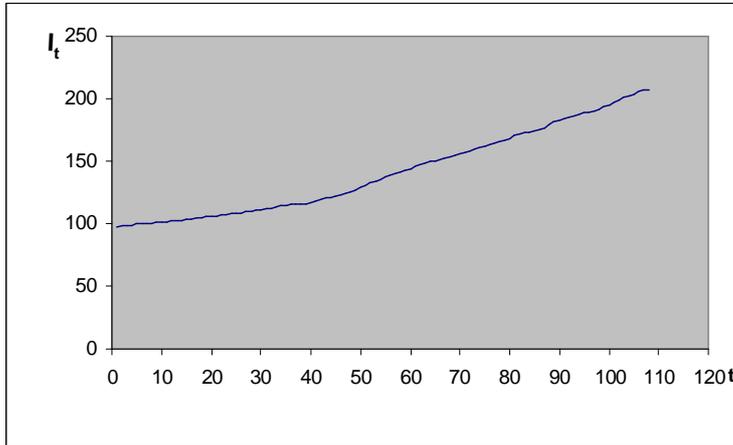


Fig. 3 - Evolução do índice mensal dos preços no consumidor entre 1970 e 1978

A representação da série dada revela a presença de uma ruptura nas vizinhanças da observação 50, o que poderá estar associado ao primeiro grande aumento do preço do petróleo no final de 1973.

Constata-se a presença de uma evolução crescente pelo que talvez se revele interessante o estudo do crescimento relativo mensal daquele índice de preços, isto é, da série deduzida da inicial pela transformação

$$\delta_{I_t} = \frac{I_t - I_{t-1}}{I_{t-1}}.$$

Os valores em percentagem da série transformada, composta por 107 observações, encontram-se no quadro 2.

	1970	1971	1972	1973	1974	1975	1976	1977	1978
Jan		0.59	0.28	0.0	1.68	1.11	1.07	0.29	0.48
Fev	0.31	0.49	0.55	0.26	1.33	0.75	0.69	0.69	0.74
Mar	0.31	0.39	0.46	0.52	1.16	0.82	0.87	0.91	0.89
Abr	0.51	0.58	0.37	0.69	1.61	0.88	0.86	1.30	1.09
Mai	0.4	0.67	0.55	0.94	1.21	0.74	0.67	0.95	0.97
Jun	0.4	0.38	0.54	0.76	1.12	0.73	0.42	0.77	0.76
Jul	0.2	0.48	0.81	0.84	1.25	0.73	0.97	0.88	1.21
Ag	0.4	0.38	0.54	0.67	0.8	0.65	0.72	0.54	0.60
Set	0.4	0.47	0.62	0.91	1.08	0.85	1.07	0.86	0.64
Out	0.4	0.56	0.88	1.06	1.21	0.77	0.94	0.8	0.93
Nov	0.4	0.37	0.61	0.89	0.92	0.064	0.81	0.37	0.53
Dez	0.3	0.47	0.52	0.64	0.54	0.57	0.35	0.26	0.43

Quadro 2 - Crescimento relativo do índice mensal dos preços no consumidor

A representação gráfica desta nova série, na figura 4, mostra-nos que esta apresenta mais mudanças do que a série inicial. Além disso é visível que um acréscimo brusco deste índice é muitas vezes seguido por um acréscimo mais lento, ou mesmo diminuição, nos meses seguintes.

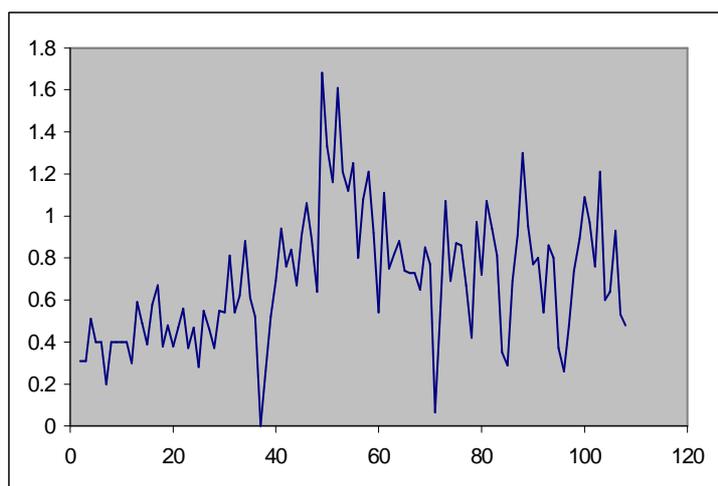


Fig. 4 - Crescimento relativo do índice mensal dos preços no consumidor

Poderia também ser interessante estudar a evolução a médio prazo do crescimento relativo do índice de preços ou ainda a existência de semelhanças de forma, de periodicidades, ao longo dos vários anos. Este último aspecto é habitualmente denominado sazonalidade e está ilustrado na figura 5.

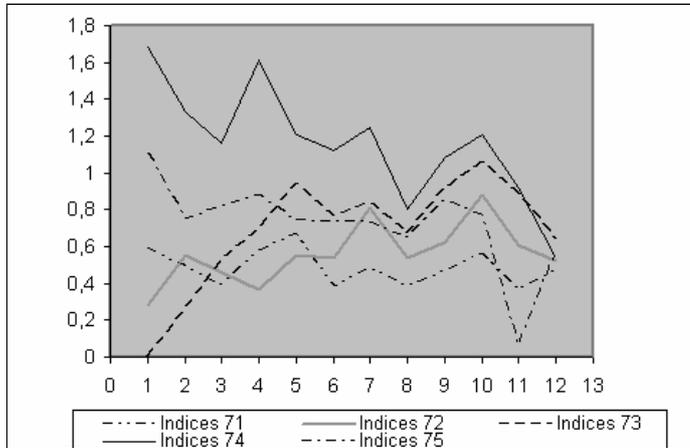


Fig. 5 Comportamentos sazonais do índice mensal dos preços

A representação gráfica da série dos acréscimos do índice dos preços em diferentes anos (1971, 1972, 1973, 1974, 1975) mostram-nos que o acréscimo relativo ao mês de Agosto (mês 8) é sempre inferior aos de Julho (mês 7) e Setembro (mês 9), o que é explicado pelo "adormecimento" da economia no mês de férias. Assim, o efeito do mês sobre o valor do índice representa aqui o fenómeno da sazonalidade.

A identificação da sazonalidade e sua correcção são imprescindíveis na análise de qualquer série temporal. Suponhamos, por exemplo, que foi tomada uma certa medida económica no mês de Julho com vista à redução do aumento de preços; será que a constatação da baixa de preços em Agosto permitirá tirar conclusões sobre a eficácia de tal medida? Com efeito, só com a análise da série corrigida de variações sazonais se poderá responder a tal questão.

Objectivos da análise de séries temporais

Na análise de séries temporais destacam-se dois objectivos fundamentais designados por modelação e previsão.

A base da modelação corresponde à construção de um modelo matemático que permita descrever, o mais adequadamente possível, a evolução da série.

São inúmeras as famílias de modelos propostas na literatura para a análise de uma série temporal. Podemos destacar, quer pela simplicidade de utilização quer pelo conhecimento das propriedades subjacentes, os modelos de ajustamento, os modelos autoprojectivos e também os modelos explicativos.

Os modelos de ajustamento pressupõem que a variável em estudo, X , é, no instante t , uma função de t e de uma perturbação aleatória u_t :

$$X_t = f(t, u_t).$$

Os modelos autoprojectivos pressupõem a existência de uma dinâmica endógena, isto é, supõe-se que a variável em estudo, X , é, no instante t , uma função dos seus valores passados e de uma perturbação aleatória u_t :

$$X_t = f(X_{t-1}, X_{t-2}, \dots, u_t)$$

Nos modelos explicativos tem-se

$$X_t = f(Y_t, u_t)$$

onde Y_t é uma família de variáveis aleatórias exógenas.

A *perturbação aleatória* ou *ruído* $u = (u_t)$ é habitualmente uma família de variáveis aleatórias centradas, de mesma variância e não correlacionadas.

Uma vez seleccionada uma certa família de modelos, torna-se necessário identificar o modelo que melhor se adequa às observações e estimar os parâmetros que o compõem. Esta estimação é feita, obviamente, à custa da série observada.

O modelo obtido é então validado por meio de testes apropriados.

Se o modelo obtido é considerado válido, podemos então avaliar os valores futuros da série temporal, isto é, entrar na fase fundamental de previsão.

É claro que esta previsão será tanto melhor quanto mais regular for, ao longo do tempo, a função que descreve a série temporal e quanto menor for o horizonte de previsão. As previsões propostas são, tal como no caso da estatística clássica, apresentadas sob a forma de intervalos, os intervalos de previsão, também aqui caracterizados pela respectiva confiança e correspondente precisão.

Neste livro daremos particular realce à família de modelos auto-projectivos. A definição e o estudo deste tipo de processos enquadra-se no contexto geral da modelação matemática de séries temporais que, desde a introdução dos processos auto-regressivos por Yule, em 1927, tem despertado grande interesse e conhecido importante desenvolvimento.

Até ao início dos anos 80 estes estudos foram dominados pelos modelos lineares, com particular ênfase para a já clássica classe dos processos auto-regressivos médias móveis (ARMA), situando-se nos anos 70, e na sequência do aparecimento do livro de Box e Jenkins (1970, ed. revista 1976), a época de ouro do seu desenvolvimento. A formulação retida para tais processos estocásticos permite exprimir o valor presente da série como função linear dos seus valores passados e dos valores presente e passados de um ruído que se interpreta como a inovação da série. A importância destes processos no estudo e desenvolvimento da modelação geral de séries temporais leva-nos a dedicar-lhes, neste texto, particular atenção.

Estando a evolução dos modelos ARMA limitada a uma formulação linear torna-se evidente o carácter restritivo do seu campo de aplicação, com vista à obtenção de bons ajustamentos para os sistemas reais a descrever. Os processos lineares revelam-se, em particular, insuficientes quer para o tratamento de certos problemas financeiros e monetários (índices da bolsa de valores, taxas de juro e de câmbio, inflação, ...) nos quais a variabilidade instantânea (ou volatilidade) das séries de valores associados depende de modo significativo do passado, quer para o estudo de fenómenos

naturais susceptíveis de alterações bruscas e violentas como, por exemplo, fenómenos sísmológicos onde podem surgir sucessões de registos com picos repentinos seguidos de baixos valores em termos absolutos.

Assim, têm surgido na literatura várias classes de modelos que permitem ter em conta diversos aspectos de não-linearidade de que destacamos a classe dos modelos condicionalmente heteroscedásticos, devida a Engle (1982), e a dos modelos bilineares, introduzida por Granger e Andersen (1978). O estudo destas classes de modelos não-lineares será também objecto deste livro.

Capítulo 1

Processos Estocásticos

1.1 Conceitos básicos

Considere-se um espaço de probabilidade (Ω, \mathcal{A}, P) . Sejam (E, \mathcal{E}) um espaço mensurável e \mathcal{T} um conjunto qualquer de índices.

Definição. Um *processo estocástico* indexado por \mathcal{T} é uma família de variáveis aleatórias (v.a.) $(X_t, t \in \mathcal{T})$ definidas sobre (Ω, \mathcal{A}, P) e com valores em (E, \mathcal{E}) .

O conjunto \mathcal{T} é chamado *espaço dos tempos*.

O conjunto E é denominado *espaço dos estados* do processo.

Para cada $\omega \in \Omega$, fixo, $(X_t(\omega), t \in \mathcal{T})$ define uma *realização* ou *trajectória do processo*.

Os processos estocásticos com que vamos trabalhar são, em geral, famílias de variáveis aleatórias em que $(E, \mathcal{E}) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ com $n \geq 1$ ⁽¹⁾. Assim, os processos estocásticos poderão ser unidimensionais, se $X_t \in \mathbb{R}$, ou multidimensionais, se $X_t \in \mathbb{R}^n$, $n > 1$. Neste segundo caso escrevemos

$$X_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \dots \\ X_{n,t} \end{bmatrix}$$

¹ $\mathcal{B}(\mathbb{R}^n)$ denota a σ -álgebra de Borel de \mathbb{R}^n .

Genericamente, tem-se que

$$\forall \omega \in \Omega, \quad (X_t(\omega), t \in \mathcal{T}) \in (\mathbb{R}^n)^{\mathcal{T}}$$

onde $(\mathbb{R}^n)^{\mathcal{T}}$ é o conjunto das funções definidas sobre \mathcal{T} com valores em \mathbb{R}^n .

As v.a. X_t são chamadas *margens* do processo estocástico e as respectivas leis de probabilidade, P_{X_t} , denominam-se *leis marginais*.

Observações.

1. Suporemos que a dimensão do processo é independente do instante de observação. Convém, no entanto, observar que na prática isto nem sempre se verifica. Para ilustrar esta afirmação consideremos o exemplo dos inquéritos sobre o emprego num determinado país. Tais inquéritos são geralmente feitos com uma periodicidade regular e, semestralmente por exemplo, dispomos de dados relativos a diversos grupos de indivíduos:

$$\begin{aligned} \text{n}^\circ \text{ de agricultores: } & X_{1,t} \\ \text{n}^\circ \text{ de empregados de escritório: } & X_{2,t}. \end{aligned}$$

Por razões várias, que poderão ter a ver com o custo do inquérito, trimestralmente temos apenas acesso ao dado global $X_{1,t} + X_{2,t}$. Então, ao longo do tempo, teremos a seguinte informação:

1º trimestre	2º trimestre	3º trimestre	4º trimestre
$X_{1,t} + X_{2,t}$	$X_{1,t+1}$ $X_{2,t+1}$	$X_{1,t+2} + X_{2,t+2}$	$X_{1,t+3}$ $X_{2,t+3}$

2. Suporemos que o tempo é **discreto**. Neste caso diz-se que temos um *processo estocástico de tempo discreto* ou uma *série temporal*, sendo habitual considerar $\mathcal{T} = \mathbb{N}$ ou $\mathcal{T} = \mathbb{Z}$.

Nas séries económicas esta hipótese de trabalho é em geral verificada já que dispomos de dados em instantes isolados. Há no entanto casos excepcionais como, por exemplo, o das cotações da bolsa. Como sabemos, tais cotações são apresentadas diariamente ou mesmo várias vezes ao dia, isto é, os dados são recolhidos em

instantes tão próximos que poderá revelar-se útil o seu tratamento em tempo contínuo.

Contrariamente às séries económicas, as séries físicas aparecem geralmente em tempo contínuo. Assim acontece, em particular, com dados relativos à temperatura, à pluviosidade e aos batimentos cardíacos.

3. Se $\mathcal{T} = \mathbb{N}$ ou $\mathcal{T} = \mathbb{Z}$ temos observações supostas equidistantes. Também esta hipótese de trabalho é por vezes violada na prática. Por exemplo, se pretendermos efectuar um inquérito sobre os gastos familiares, nas habitações, duas vezes por ano, que meses deveriam ser escolhidos para proceder ao inquérito? Deveremos, naturalmente, escolher os meses em que é mais provável encontrar as pessoas em suas casas. Se eliminarmos os correspondentes a férias ou os mais propensos a saídas poderíamos ser conduzidos, com argumentos do tipo a seguir apresentados, aos meses de Março e Outubro, que não são equidistantes.

Janeiro	Fevereiro	Março
—	—	—
Abril	Maio	Junho
Saídas pela Páscoa	Saídas pelos dias festivos	Saídas pelos feriados
Julho	Agosto	Setembro
Períodos de Férias	Períodos de Férias	Saídas pelas compras escolares
Out	Nov	Dezembro
—	—	Saídas pelo Natal

4. Poderíamos pensar em considerar \mathcal{T} igual ao período de observação, digamos, $\mathcal{T} = \{1, 2, \dots, T\}$. Há no entanto alguns inconvenientes para esta abordagem.

i) Reparemos que a previsão para os instantes $T + 1, T + 2, \dots$ implicaria a necessidade de considerar um período mais alargado, isto é,

$$\mathcal{T}^* = \{1, 2, \dots, T, T + 1, T + 2, \dots\}.$$

ii) Por outro lado, havendo uma observação suplementar em $T + 1$, impor-se-á a conseqüente previsão para os instantes $T + 2, T + 3, \dots$

Assim, parece claro que não podemos pensar em limitar superiormente \mathcal{T} . Prolongamos então os tempos para $+\infty$ e consideramos $\mathcal{T} = \mathbb{N}$.

O interesse do prolongamento dos tempos para $-\infty$, isto é, tomar $\mathcal{T} = \mathbb{Z}$, é principalmente matemático; permite obter em certos casos resultados mais simples. Pode ainda justificar-se com a previsão no passado longínquo, necessária por exemplo em séries do domínio da Arqueologia ou, também, da Economia.

1.2 Teorema de Kolmogorov

Seja $X = (X_t, t \in \mathcal{T})$ um processo estocástico definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) e com valores em $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Como já vimos, o processo estocástico X assume valores no espaço de dimensão infinita $(\mathbb{R}^n)^{\mathcal{T}}$. O estudo da lei do processo estocástico X , no sentido da lei imagem de P por X , impõe a construção de uma σ -álgebra sobre $(\mathbb{R}^n)^{\mathcal{T}}$ e a caracterização das probabilidades sobre tal σ -álgebra.

Tal estudo, que não vamos desenvolver aqui, é devido a Kolmogorov e pode encontrar-se, por exemplo, em Neveu (1970) ou Métivier (1972). De facto, Kolmogorov constrói a σ -álgebra de Borel sobre $(\mathbb{R}^n)^{\mathcal{T}}$, estuda as probabilidades nesta σ -álgebra e a partir daí caracteriza a lei de um processo estocástico.

Ora, dado um qualquer processo estocástico $X = (X_t, t \in \mathcal{T})$ podemos fazer-lhe corresponder o conjunto dos vectores aleatórios de dimensão finita $(X_{t_1}, \dots, X_{t_k})$, com t_1, \dots, t_k quaisquer em \mathcal{T} , com os quais sabemos trabalhar.

Esta constatação levou Kolmogorov a analisar a seguinte questão: em que medida é que o conhecimento da família de vectores aleatórios de dimensão finita, associados a um processo estocástico, e respectivas leis permite caracterizar esse processo e sua lei?

Designa-se então por P_X a lei de um processo estocástico com valores em $(\mathbb{R}^n)^{\mathcal{T}}$ munido da sua tribo de Borel $\mathcal{B}[(\mathbb{R}^n)^{\mathcal{T}}]$. Considere-se a família das leis em $(\mathbb{R}^{nk}, \mathcal{B}(\mathbb{R}^{nk}))$ associadas a qualquer subfamília finita de X , $(X_{t_1}, \dots, X_{t_k})$, t_1, \dots, t_k quaisquer em \mathcal{T} ,

$$\left\{ P_{(X_{t_1}, \dots, X_{t_k})} \right\}$$

chamadas *leis de dimensão finita* ou *leis temporais* do processo X .

As leis $\left\{ P_{(X_{t_1}, \dots, X_{t_k})} \right\}$ satisfazem a chamada *condição de projectividade*, isto é, verifica-se a compatibilidade entre vectores e respectivas leis se, designando por π_J^I as projecções coordenadas⁽²⁾, se tem

$$P_{(X_{t_1}, \dots, X_{t_k})} = \left(P_{(X_{t_1}, \dots, X_{t_k}, \dots, X_{t_{k+m}})} \right)_{\pi_{t_1, \dots, t_k}^{t_1, \dots, t_{k+m}}}.$$

Assim, a lei de $(X_{t_1}, \dots, X_{t_k})$ é a lei imagem de qualquer vector de dimensão finita que inclua aquelas componentes pela projecção coordenada adequada.

Esta condição de coerência da família de leis de probabilidade de dimensão finita assegura a existência e unicidade da lei P_X do processo estocástico X sobre $\left((\mathbb{R}^n)^{\mathcal{T}}, \mathcal{B} \left[(\mathbb{R}^n)^{\mathcal{T}} \right] \right)$, conforme está expresso no teorema seguinte.

Teorema de Kolmogorov (Neveu, 1970) Se dispusermos de uma família de leis de probabilidade

$$\left\{ P_{(X_{t_1}, \dots, X_{t_k})}, \forall k, \forall t_1, \dots, t_k \text{ em } \mathcal{T} \right\}$$

satisfazendo as respectivas condições de projectividade, então existe uma lei única P_X tal que

$$P_{(X_{t_1}, \dots, X_{t_k})} = (P_X)_{\pi_{t_1, \dots, t_k}}, \forall k, \forall t_1, \dots, t_k \text{ em } \mathcal{T}.$$

Ilustremos este teorema com o caso particular dos processos estocásticos Gaussianos, que passamos a definir.

²A todo o subconjunto finito I do conjunto \mathcal{T} associamos o espaço produto $(\mathbb{R}^n)^I$, isto é, o conjunto das sequências finitas $x_i, i \in I, x_i \in \mathbb{R}^n$. Para I, J partes finitas de \mathcal{T} , com $J \subset I$, designamos por $\pi_J^I : (\mathbb{R}^n)^I \rightarrow (\mathbb{R}^n)^J$ a projecção que consiste em reter apenas as coordenadas com índices pertencentes a J .

Definição. Um processo estocástico real $X = (X_t, t \in \mathcal{T})$ é *Gaussiano* se a lei de qualquer subfamília finita é Gaussiana.⁽³⁾

Verifiquemos que a definição é de facto coerente, isto é, prove-mos que a família das leis normais

$$P_{T_k} = N(m_{T_k}, \Gamma_{T_k}),$$

onde $T_k, k \in \mathbb{N}$, descreve as partes finitas de \mathcal{T} , verifica a condição de projectividade. Designam-se por m_{T_k} e Γ_{T_k} , respectivamente, o vector média e a matriz de variâncias-covariâncias do vector $(X_t, t \in T_k)$.

Sejam $T_1 = \{t_1, \dots, t_r\}$ e $T_2 = \{t'_1, \dots, t'_l\}$, com $T_2 \subset T_1$. Prove-mos que

$$P_{T_2}(B) = P_{T_1} \left(\left(\pi_{T_2}^{T_1} \right)^{-1}(B) \right), B \in \mathcal{B}(\mathbb{R}^l).$$

Ora,

$$P_{T_2}(B) = N(m_{T_2}, \Gamma_{T_2})(B).$$

Por outro lado,

$$\left(\pi_{T_2}^{T_1} \right)^{-1}(B) = \left\{ (x_t)_{t \in T_1} \in \mathbb{R}^r : (x_{t'_1}, \dots, x_{t'_l}) \in B \right\} = A.$$

Então,,

$$\begin{aligned} P_{T_1} \left(\left(\pi_{T_2}^{T_1} \right)^{-1}(B) \right) &= N(m_{T_1}, \Gamma_{T_1})(A) \\ &= P((X_{t_1}, \dots, X_{t_r}) \in A) \\ &= P \left((X_{t'_1}, \dots, X_{t'_l}) \in B \right), \text{ da definição de } A \\ &= P_{(X_{t'_1}, \dots, X_{t'_l})}(B) \\ &= N(m_{T_2}, \Gamma_{T_2})(B) \end{aligned}$$

³Para o estudo de vectores aleatórios normais sugere-se, por exemplo, E. Gonçalves e N. Mendes Lopes (2000), cap. IV.

uma vez que um vector marginal de um vector aleatório normal também é normal.

1.3 Processos de segunda ordem

1.3.1 Definição. Propriedades

Definição. Um processo estocástico $X = (X_t, t \in \mathcal{T})$ é de segunda ordem se

$$E \|X_t\|^2 < +\infty, \forall t \in \mathcal{T},$$

onde $\|x\|$ é a notação usada para norma do elemento x . ⁽⁴⁾

Se \mathbb{R}^n estiver munido da norma euclideana tem-se, representando por X_t' o transposto do vector X_t ,

$$\|X_t\|^2 = X_t' X_t = X_{1,t}^2 + X_{2,t}^2 + \dots + X_{n,t}^2.$$

Assim,

$$E \|X_t\|^2 < +\infty \iff E [X_{1,t}^2 + X_{2,t}^2 + \dots + X_{n,t}^2] < +\infty,$$

condição que nos vai permitir estabelecer o resultado seguinte que relaciona a segunda ordem do processo com a segunda ordem dos "processos componentes".

Propriedade. Um processo multivariado

$$X = ([X_{1,t} \ X_{2,t} \ \dots \ X_{n,t}], t \in \mathcal{T})$$

é de segunda ordem se e só se $X_j = (X_{j,t}, t \in \mathcal{T})$ é de segunda ordem, para todo $j = 1, 2, \dots, n$.

Prova. É evidente, pela linearidade da esperança matemática, que

$$E [X_{j,t}^2] < +\infty, j = 1, 2, \dots, n \Rightarrow E [X_{1,t}^2 + X_{2,t}^2 + \dots + X_{n,t}^2] < +\infty.$$

⁴Ao longo do texto será sempre esta a notação usada para norma, independentemente do espaço e da norma que se considerar. Se necessário, será explicitada qual a norma a usar num determinado contexto.

Inversamente, para provar que

$$E [X_{1,t}^2 + X_{2,t}^2 + \dots + X_{n,t}^2] < +\infty \Rightarrow E [X_{j,t}^2] < +\infty, j = 1, 2, \dots, n$$

racionamos por absurdo e usamos o facto de $X_{j,t}^2$ ser de valores positivos. ■

Se um processo é de segunda ordem podemos concluir de imediato os seguintes factos sobre os seus "processos componentes":

1. $X_{j,t}$ possui valor médio: $E [X_{j,t}] = m_j (t)$.

O valor médio de X_t é o vector, com n linhas, formado por estes valores. Representá-lo-emos por $E [X_t] = m (t)$.

2. $X_{j,t}$ possui variância: $V [X_{j,t}]$.

3. Existe covariância entre $X_{i,s}$ e $X_{j,t}$ pois, pela desigualdade de Schwartz em $L^2 (\Omega, A, P)$ ⁽⁵⁾, $E (XY) \leq \sqrt{E (X^2) E (Y^2)}$.

Podemos assim definir a família de matrizes quadradas $\Gamma = [\Gamma (s, t), s, t \in \mathcal{T}]$ com

$$\Gamma (s, t) = Cov (X_s, X_t) = E [(X_s - m(s)) (X_t - m(t))'] .$$

A lei de um processo de segunda ordem pode então ser parcialmente resumida pelas seguintes funções:

i) função média: $m = (m (t), t \in \mathcal{T})$

ii) função das covariâncias: $\Gamma = (\Gamma (s, t), s, t \in \mathcal{T})$.

Propriedade. A função $\Gamma = (\Gamma (s, t), s, t \in \mathcal{T})$ é a função de covariância dum processo de segunda ordem com valores em \mathbb{R}^n se e só se Γ é um operador real, hermitico e semi-definido positivo ⁽⁶⁾, isto é, verificando:

i) a simetria: $\Gamma (s, t) = \Gamma (t, s)', \forall s, t \in \mathcal{T}$

⁵Ver Anexo, parágrafo 1.6.

⁶Uma matriz Γ quadrada de ordem k é hermitiana se $\Gamma = \Gamma^*$ (* designa o transposto-conjugado).

Γ é semi-definida positiva (ou definida não-negativa) se $u^* \Gamma u \geq 0, \forall u \in \mathbb{C}^k$.

Γ é definida positiva se $u^* \Gamma u > 0$, para todo $u \neq 0$.

ii) a positividade:

$$\forall k \in \mathbb{N}, \forall h_1, \dots, h_k \in \mathbb{R}^n, \forall t_1, \dots, t_k \in \mathcal{T}, \sum_{i=1}^k \sum_{j=1}^k h_i' \Gamma(t_i, t_j) h_j \geq 0.$$

Prova. Considerando o elemento genérico da matriz de covariância $\Gamma(t, s)'$, tem-se

$$\begin{aligned} [\Gamma(t, s)']_{i,j} &= [\Gamma(t, s)]_{j,i} = \text{Cov}(X_{t,j}, X_{s,i}) \\ &= \text{Cov}(X_{s,i}, X_{t,j}) = [\Gamma(s, t)]_{i,j}. \end{aligned}$$

Por outro lado,

$$\forall k \in \mathbb{N}, \forall h_1, \dots, h_k \in \mathbb{R}^n, \forall t_1, \dots, t_k \in \mathcal{T}, \forall i, j \in \{1, \dots, k\},$$

$$h_i' \Gamma(t_i, t_j) h_j = \sum_{p=1}^n \sum_{q=1}^n h_{p,i} h_{q,j} \text{Cov}(X_{p,t_i}, X_{q,t_j}),$$

com

$$\text{Cov}(X_{p,t_i}, X_{q,t_j}) = E[(X_{p,t_i} - E(X_{p,t_i}))(X_{q,t_j} - E(X_{q,t_j}))].$$

Então

$$\begin{aligned} & \sum_{i=1}^k \sum_{j=1}^k h_i' \Gamma(t_i, t_j) h_j \\ &= \sum_{i=1}^k \sum_{j=1}^k \sum_{p=1}^n \sum_{q=1}^n h_{p,i} h_{q,j} E[(X_{p,t_i} - E(X_{p,t_i}))(X_{q,t_j} - E(X_{q,t_j}))] \\ &= E \left[\sum_{i=1}^k \sum_{p=1}^n h_{p,i} (X_{p,t_i} - E(X_{p,t_i})) \right]^2 \geq 0. \end{aligned}$$

Inversamente, para todo o operador Γ real hermítico positivo, existe pelo menos um processo de segunda ordem admitindo Γ como função de covariância. Basta considerar um processo Gaussiano de função de covariâncias Γ e função média nula pois, da definição de processo Gaussiano é fácil concluir que tal processo é caracterizado por aquelas duas funções.

1.3.2 Exemplos de processos de segunda ordem

1. Sejam f e g duas funções deterministas de $t, t \in \mathcal{T}$, e A e B duas variáveis aleatórias reais de quadrado integrável. O processo

$$X_t = Af(t) + Bg(t)$$

é um processo de segunda ordem.

Notemos que se conhecermos $X_t(\omega) = A(\omega)f(t) + B(\omega)g(t)$ para um ω fixo e se existirem t_0 e t_1 tais que $f(t_0)$ e $g(t_0)$ sejam linearmente independentes de $f(t_1)$ e $g(t_1)$ então, conhecidos $X_{t_0}(\omega)$ e $X_{t_1}(\omega)$, podemos conhecer A e B e automaticamente $X_t(\omega)$, para todo o t (isto é, toda a trajectória do processo).

2. Ruído branco

Chamamos *ruído branco* a uma sucessão de vectores aleatórios $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ que são

- centrados: $m = 0_{\mathbb{R}^n}$
- não correlacionados: $\Gamma(s, t) = 0, \forall s \neq t$
- com igual matriz de variâncias-covariâncias:
 $\Gamma(t, t) = V(\varepsilon_t) = \Sigma$, independente de t .

Assim, num ruído branco não há correlações entre 2 instantes diferentes. Em contrapartida, pode haver correlações instantâneas porque, por exemplo, $Cov(\varepsilon_{1,t}, \varepsilon_{2,t}) = \Sigma_{1,2}$ poderá ser não nula. A ausência de correlações instantâneas verifica-se no denominado *ruído branco reduzido* no qual se considera $\Sigma = I_n$, com I_n a matriz identidade de ordem n .

No caso unidimensional, a definição anterior resume-se ao seguinte: $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco se $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ e $E(\varepsilon_t \varepsilon_s) = 0$, se $t \neq s, t \in \mathbb{Z}, s \in \mathbb{Z}$.

3. Média móvel infinita (bidireccional)

Seja ε um ruído branco de \mathbb{R}^n . Seja (A_{jt}) uma sucessão de matrizes de ordens m e n , duplamente indexada, com $j \in \mathbb{Z}$, e tal que

$$\sum_{j=-\infty}^{+\infty} \|A_{jt}\| < +\infty, \forall t. \quad (7)$$

Consideremos

$$X_t = \sum_{j=-\infty}^{+\infty} A_{jt}\varepsilon_{t-j}.$$

Esta igualdade tem sentido porque

$$\|A_{jt}\varepsilon_{t-j}\| \leq \|A_{jt}\| \|\varepsilon_{t-j}\|$$

e como

$$\|\varepsilon_{t-j}\|^2 = E(\varepsilon'_{t-j}\varepsilon_{t-j}) = E\left(\sum_{i=1}^n \varepsilon_{i,t-j}^2\right) = \sum_{i=1}^n V(\varepsilon_{i,t-j}) = Tr(\Sigma),$$

onde Tr designa o traço de uma matriz, isto é, a soma dos elementos da sua diagonal principal, então

$$\|A_{jt}\varepsilon_{t-j}\| \leq \|A_{jt}\| \sqrt{Tr(\Sigma)}.$$

Assim $(A_{jt}\varepsilon_{t-j})$ é o termo geral de uma série convergente em norma uma vez que, por hipótese, $\sum_{j=-\infty}^{+\infty} \|A_{jt}\| < +\infty, \forall t$. Dizemos, pois, que $(A_{jt}\varepsilon_{t-j})$ é o termo geral de uma série absolutamente convergente (em L_m^2) o que implica que a série converge (em L_m^2).

Assim $X_t = \sum_{j=-\infty}^{+\infty} A_{jt}\varepsilon_{t-j}$ está bem definido (em L_m^2).

O processo $X = \left(X_t = \sum_{j=-\infty}^{+\infty} A_{jt}\varepsilon_{t-j}, t \in \mathbb{Z} \right)$ denomina-se *média móvel infinita bidireccional*.

Podemos calcular o valor médio de X_t e a covariância entre X_t e X_s . Para isso, comecemos por estabelecer um resultado (Gouriéroux, Monfort, 1990, Cap. V, Anexo), que nos mostrará

⁷Sem perda de generalidade, consideramos

$$\|A\|^2 = \sup_{x \neq 0} \frac{x'A'Ax}{x'x} = \text{maior valor próprio de } A'A.$$

que é lícito, nas condições dadas, permutar os símbolos de integral e somatório.

Consideremos uma sucessão de v.a. U_k de L_m^2 e suponhamos que U_k converge para U no sentido de L_m^2 , isto é,

$$E \left(\|U_k - U\|^2 \right) \xrightarrow{k \rightarrow +\infty} 0.$$

Da desigualdade de Schwartz,

$$\forall i \in \{1, \dots, m\}, E \left[(U_k - U)' e_i \right] \leq E \left(\|U_k - U\|^2 \right),$$

onde (e_1, \dots, e_m) é a base canónica de \mathbb{R}^m .

Assim,

$$\forall i \in \{1, \dots, m\}, E (U_{i,k} - U_i) \xrightarrow{k \rightarrow +\infty} 0$$

ou seja

$$E (U_k) \xrightarrow{k \rightarrow +\infty} E (U).$$

Apliquemos este resultado ao cálculo de $E(X_t)$.

Por um lado $\sum_{j=-\infty}^{+\infty} A_{jt} \varepsilon_{t-j}$ converge em L_m^2 . Assim, basta aplicar

o resultado anterior a $U_{k,k'} = \sum_{j=-k'}^k A_{jt} \varepsilon_{t-j}$ e $U = \sum_{j=-\infty}^{+\infty} A_{jt} \varepsilon_{t-j}$.

Pela linearidade da esperança matemática,

$$\lim_{\substack{k \rightarrow +\infty \\ k' \rightarrow +\infty}} E \left(\sum_{j=-k'}^k A_{jt} \varepsilon_{t-j} \right) = \lim_{\substack{k \rightarrow +\infty \\ k' \rightarrow +\infty}} \sum_{j=-k'}^k E (A_{jt} \varepsilon_{t-j}).$$

Mas, usando o resultado anterior,

$$\lim_{\substack{k \rightarrow +\infty \\ k' \rightarrow +\infty}} E \left(\sum_{j=-k'}^k A_{jt} \varepsilon_{t-j} \right) = E \left(\sum_{j=-\infty}^{+\infty} A_{jt} \varepsilon_{t-j} \right).$$

Assim, da definição de série convergente e da unicidade do limite decorre

$$E \left(\sum_{j=-\infty}^{+\infty} A_{jt} \varepsilon_{t-j} \right) = \lim_{\substack{k \rightarrow +\infty \\ k' \rightarrow +\infty}} \sum_{j=-k'}^k E(A_{jt} \varepsilon_{t-j}) = \sum_{j=-\infty}^{+\infty} E(A_{jt} \varepsilon_{t-j}).$$

Obtemos, então,

$$E(X_t) = \sum_{j=-\infty}^{+\infty} A_{jt} E(\varepsilon_{t-j}) = 0$$

$$\begin{aligned} \Gamma_X(s, t) &= \text{Cov}(X_s, X_t) \\ &= \text{Cov} \left(\sum_{j=-\infty}^{+\infty} A_{js} \varepsilon_{s-j}, \sum_{i=-\infty}^{+\infty} A_{it} \varepsilon_{t-i} \right) \\ &= \sum_{j=-\infty}^{+\infty} \sum_{i=-\infty}^{+\infty} A_{js} \text{Cov}(\varepsilon_{s-j}, \varepsilon_{t-i}) A'_{it}. \end{aligned}$$

Como $\text{Cov}(\varepsilon_{s-j}, \varepsilon_{t-i}) \neq 0$ se e só se $s - j = t - i$, obtemos

$$\Gamma_X(s, t) = \sum_{i=-\infty}^{+\infty} A_{s-t+i, s} \Sigma A'_{it}$$

Podemos concluir que o processo média móvel infinita apresentado é de segunda ordem.

1.4 Processos estacionários

Se o comportamento de um processo ao longo do tempo fosse totalmente anárquico seria difícil ou mesmo impossível desenvolver estudos sobre o mesmo, designadamente estudos com vista à previsão de valores futuros do processo. Em geral os fenómenos observados permitem considerar um certo número de hipóteses de invariância no tempo, genericamente designadas por estacionariedade.

1.4.1 Processos fortemente estacionários

Definição. Um processo X é *fortemente estacionário* (ou *estritamente estacionário*) se

$\forall k \in \mathbb{N}, \forall t_1, \dots, t_k \in \mathbb{Z}, \forall h \in \mathbb{Z}, (X_{t_1}, \dots, X_{t_k})$ e $(X_{t_1+h}, \dots, X_{t_k+h})$ têm a mesma lei.

Em particular, a lei de $X = (X_t, t \in \mathbb{Z})$ coincide com a lei de $\tilde{X} = (X_{t+r}, t \in \mathbb{Z})$, r arbitrariamente fixo em \mathbb{Z} , ou seja, há invariância da lei do processo por translação no tempo.

Notemos que nesta definição não se exige que o processo seja de segunda ordem.

Exemplos.

1. Se A é uma variável aleatória então $X_t = A, \forall t$, é um processo fortemente estacionário já que a lei de $(X_{t_1}, \dots, X_{t_k})$ é a lei do vector (A, \dots, A) , de dimensão k , a qual é independente de t_1, \dots, t_k .

2. Se $(X_t, t \in \mathbb{Z})$ é uma sucessão de variáveis aleatórias independentes de mesma lei P , então o processo $X = (X_t, t \in \mathbb{Z})$ é fortemente estacionário porque $P_{(X_{t_1}, \dots, X_{t_k})} = P^{\otimes k}$, que é independente de t_1, \dots, t_k .

A propriedade seguinte, cuja prova é imediata, relaciona a estacionaridade forte de um processo multivariado com a estacionaridade forte das suas componentes.

Propriedade. Se $X = ([X_{1,t} \ X_{2,t} \ \dots \ X_{n,t}]', t \in \mathbb{Z})$ é fortemente estacionário então $X_j = (X_{j,t}, t \in \mathbb{Z})$ é fortemente estacionário, $\forall j = 1, \dots, n$.

O recíproco desta propriedade pode não ser verdadeiro. Para ilustrar esta afirmação consideremos o processo estocástico

$$\left(Z_t = \begin{pmatrix} X_t \\ Y_t \end{pmatrix}, t \in \mathbb{Z} \right)$$

tal que

i) $(X_t, t \in \mathbb{Z})$ é um processo Gaussiano real, independente, centrado e fortemente estacionário, isto é,

- $\forall t \in \mathbb{Z}, X_t$ segue a lei $\mathcal{N}(0, \sigma)$
- $(X_t, t \in \mathbb{Z})$ é uma família de variáveis aleatórias independentes.

$$\text{ii) } (Y_t, t \in \mathbb{Z}) = ((-1)^t X_t, t \in \mathbb{Z}).$$

O processo $(Y_t, t \in \mathbb{Z})$ é também uma família de variáveis aleatórias independentes e seguindo todas a lei $\mathcal{N}(0, \sigma)$. Assim, $(Y_t, t \in \mathbb{Z})$ é fortemente estacionário.

Provemos que $(Z_t, t \in \mathbb{Z})$ não é fortemente estacionário. Basta ver que as variáveis aleatórias Z_t não são identicamente distribuídas.

De facto, para qualquer $(x, y) \in \mathbb{R}^2$ tem-se

- se t é par

$$\begin{aligned} F_{Z_t}(x, y) &= P(X_t \leq x, Y_t \leq y) \\ &= P(X_t \leq \min(x, y)) \\ &= F_{N(0, \sigma)}(\min(x, y)) > 0 \end{aligned}$$

- se t é ímpar

$$\begin{aligned} F_{Z_t}(x, y) &= P(X_t \leq x, -X_t \leq y) \\ &= \begin{cases} P(-y \leq X_t \leq x), & -y < x \\ 0, & -y \geq x \end{cases} \end{aligned}$$

Por exemplo, $F_{Z_2}(1, -2) = F_{N(0, \sigma)}(-2) > 0$ e $F_{Z_1}(1, -2) = 0$.

Observação. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico de segunda ordem e fortemente estacionário.

Tem-se

$$\begin{aligned} E(X_t) &= \int x dP_{X_t}(x) \\ E(X_{t+h}) &= \int x dP_{X_{t+h}}(x) = \int x dP_{X_t}(x) = E(X_t) \end{aligned}$$

ou seja, $E(X_t) = m_t = m$, isto é, é independente de t .

Do mesmo modo, a variância é também independente de t .

Quanto à covariância, sabemos que a lei de (X_{t_1}, X_{t_2}) é a lei de (X_{t_1+h}, X_{t_2+h}) , $\forall h$. Então

$$\text{Cov}(X_{t_1}, X_{t_2}) = \text{Cov}(X_{t_1+h}, X_{t_2+h})$$

isto é,

$$\forall t_1, t_2, h, \Gamma(t_1, t_2) = \Gamma(t_1 + h, t_2 + h);$$

em particular, com $h = -t_1$ obtemos

$$\Gamma(t_1, t_2) = \Gamma(0, t_2 - t_1)$$

ou seja, a covariância é apenas função da diferença entre os índices de X .

Podemos assim enunciar o resultado seguinte:

Propriedade. Se $X = (X_t, t \in \mathbb{Z})$ é um processo estocástico fortemente estacionário e de segunda ordem então $E(X_t)$ e $V(X_t)$ são independentes de t e $\text{Cov}(X_t, X_s) = \gamma(t - s)$.

Vamos estudar processos estocásticos que, até à ordem dois, apresentam estas características de regularidade, ou seja, de acordo com a definição seguinte, processos fracamente estacionários.

1.4.2 Processos estacionários no sentido fraco

Definição. Um processo de segunda ordem $X = (X_t, t \in \mathbb{Z})$ é *estacionário no sentido fraco* se

- $E(X_t) = m, \forall t \in \mathbb{Z}$
- $\text{Cov}(X_t, X_{t+h}) = \gamma(h), \forall t \in \mathbb{Z}$.

Alguns autores adoptam a terminologia *estacionário à ordem dois*⁽⁸⁾ ou *estacionário em covariância*. No que se segue e sempre

⁸Mais geralmente, o processo real X é estacionário à ordem j se todos os momentos conjuntos de $(X_{t_1}, \dots, X_{t_n})$ de ordem inferior ou igual a j existem e são iguais aos momentos conjuntos correspondentes de $(X_{t_1+h}, \dots, X_{t_n+h})$,

isto é

$$E[(X_{t_1})^{j_1} \dots (X_{t_n})^{j_n}] = E[(X_{t_1+h})^{j_1} \dots (X_{t_n+h})^{j_n}]$$

com $j_1 \geq 0, \dots, j_n \geq 0$ e $j_1 + \dots + j_n \leq j$, para todos $(t_1, \dots, t_n) \in \mathbb{Z}^n$ e $h \in \mathbb{Z}$.

que estivermos a referir-nos a este tipo de estacionaridade, diremos apenas X é *estacionário*.

A função $\gamma(h) = \Gamma(0, h)$, $h \in \mathbb{Z}$, chama-se *função de autocovariância*.

A propriedade seguinte, cuja prova é imediata, relaciona a estacionaridade de um processo multivariado com a das suas componentes.

Propriedade. Se $X = ([X_{1,t} \ X_{2,t} \ \dots \ X_{n,t}]', t \in \mathbb{Z})$ é estacionário então $X_j = (X_{j,t}, t \in \mathbb{Z})$ é estacionário, $\forall j = 1, \dots, n$.

De modo semelhante ao observado na estacionaridade forte, o facto de os processos componentes serem fracamente estacionários nada nos permite concluir sobre a estacionaridade fraca do processo que os admite como componentes. Retomando o exemplo usado no caso fortemente estacionário, vemos que $(X_t, t \in \mathbb{Z})$ e $(-X_t, t \in \mathbb{Z})$ são fracamente estacionários. No entanto $(Z_t, t \in \mathbb{Z})$ não é fracamente estacionário pois

$$\begin{aligned} \text{Cov}(Z_t, Z_t) &= E(Z_t Z_t') = E\left(\begin{pmatrix} X_t \\ Y_t \end{pmatrix} (X_t \ Y_t)\right) \\ &= E\begin{bmatrix} X_t^2 & (-1)^t X_t^2 \\ (-1)^t X_t^2 & X_t^2 \end{bmatrix}, \end{aligned}$$

isto é,

$$\text{Cov}(Z_t, Z_t) = \begin{cases} \begin{bmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{bmatrix} & , t \text{ par} \\ \begin{bmatrix} \sigma^2 & -\sigma^2 \\ -\sigma^2 & \sigma^2 \end{bmatrix} & , t \text{ ímpar.} \end{cases}$$

Os dois tipos de estacionaridade que foram introduzidos não estão directamente relacionados. De facto, pode ocorrer uma qualquer das situações a seguir descritas.

a) Um processo pode ser fortemente estacionário e não ser estacionário.

Basta considerar o processo $X = (X_t, t \in \mathbb{Z})$ constituído por uma sucessão de variáveis aleatórias independentes e identicamente distribuídas (v.a.r.i.i.d.) com a lei de Cauchy (relembremos que, neste caso, a esperança de X_t não existe).

b) Um processo pode ser estacionário e não ser fortemente estacionário.

Este caso é ilustrado considerando o processo $X = (X_t, t \in \mathbb{Z})$ constituído por uma sucessão de v.a.r. independentes tais que

$$P(X_t = x) = \begin{cases} 1 - \frac{1}{|t|} & , x = 0 \\ \frac{1}{2|t|} & , x = \sqrt{|t|} \\ \frac{1}{2|t|} & , x = -\sqrt{|t|} \end{cases}, \quad \forall t \in \mathbb{Z} - \{0\},$$

e

$$P(X_0 = x) = \begin{cases} \frac{1}{2}, & x = -1 \\ \frac{1}{2}, & x = 1. \end{cases}$$

c) Um processo fortemente estacionário de segunda ordem é estacionário.

d) Um processo estacionário tal que a lei de $(X_{t_1}, \dots, X_{t_k})$, $P_{(X_{t_1}, \dots, X_{t_k})}$, depende apenas de $E(X_{t_1}), \dots, E(X_{t_k})$ e de $Cov(X_{t_i}, X_{t_j})$ é fortemente estacionário.

Em particular

i) se um processo é Gaussiano e estacionário então é fortemente estacionário.

ii) se as v.a.r. X_t , $t \in \mathbb{Z}$, são independentes e X_t segue a lei de Poisson de parâmetro λ_t e se o processo $X = (X_t, t \in \mathbb{Z})$ é estacionário (o que implica que $\lambda_t = \lambda$) então X é fortemente estacionário.

Debrucemo-nos mais um pouco sobre a definição de estacionariedade fraca.

Embora sobre m não seja feita qualquer restrição, já o mesmo não acontece com Γ .

De facto, sendo $Cov(X_t, X_{t+h}) = Cov(X_{t+h}, X_t)'$ devemos ter $\gamma(h) = \gamma(-h)'$, isto é, γ é simétrica (ou par, se X é unidimensional).

Por outro lado, sendo $\Gamma(t_i, t_j)$ uma matriz de variâncias-covariâncias, tem-se

$$\sum_{i=1}^k \sum_{j=1}^k h_i' \Gamma(t_i, t_j) h_j \geq 0, \forall k \in \mathbb{N}, \forall h_1, \dots, h_k \in \mathbb{R}^n, \forall t_i, t_j \in \mathbb{Z},$$

isto é,

$$\sum_{i=1}^k \sum_{j=1}^k h_i' \gamma(t_j - t_i) h_j \geq 0$$

o que implica, considerando $t_i = i$,

$$\sum_{i=1}^k \sum_{j=1}^k h_i' \gamma(j - i) h_j \geq 0$$

ou seja, γ é um operador semi-definido positivo.

Notemos ainda que ao considerar, nos processos de segunda ordem, a sub-classe dos processos estacionários estamos a introduzir uma forte simplificação ao nível dos momentos. De facto, deixamos de ter uma função $m(t)$ para passarmos a uma constante m e deixamos de considerar uma função de 2 argumentos, $\Gamma(s, t)$, para passarmos a ter apenas $\gamma(h)$.

Exemplo. Seja ε um ruído branco real definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) , de variância σ^2 estritamente positiva. Seja X o processo estocástico real definido por:

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}, \forall t \in \mathbb{Z},$$

com $(a_j, j \in \mathbb{N}_0)$ tal que $\sum_{j=0}^{+\infty} |a_j| < +\infty$.

Sabendo que a convergência envolvida na definição de X_t é no sentido de $L^2(\Omega, \mathcal{A}, P)$, provemos que X é um processo estocástico fracamente estacionário.

A prova de que X é um processo de segunda ordem resume-se a

$$\begin{aligned} \|X_t\| &= \left\| \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \right\| \leq \sum_{j=0}^{+\infty} |a_j| \|\varepsilon_{t-j}\| \\ &\leq \sum_{j=0}^{+\infty} |a_j| [E(\varepsilon_{t-j}^2)]^{\frac{1}{2}} = \sigma \sum_{j=0}^{+\infty} |a_j| < +\infty. \end{aligned}$$

Provemos que $E(X_t)$ é independente de t . Tem-se

$$E(X_t) = E\left(\sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}\right) = E\left(\lim_{N \rightarrow +\infty} \sum_{j=0}^N a_j \varepsilon_{t-j}\right).$$

Já vimos que $\sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$ existe em L^2 .

Vejamos agora que $\sum_{j=0}^N a_j \varepsilon_{t-j} \xrightarrow{L^2, N \rightarrow +\infty} \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$.

$$\left\| \sum_{j=0}^N a_j \varepsilon_{t-j} - \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \right\| = \left\| \sum_{j=N+1}^{+\infty} a_j \varepsilon_{t-j} \right\| \leq \sigma \sum_{j=N+1}^{+\infty} |a_j|$$

que tende para 0 quando $N \rightarrow +\infty$, por ser o resto de uma série convergente.

Assim,

$$\lim_{N \rightarrow +\infty} E\left(\sum_{j=0}^N a_j \varepsilon_{t-j} - \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}\right)^2 = 0.$$

Mas, da desigualdade de Cauchy-Schwartz,

$$E \left(\sum_{j=0}^N a_j \varepsilon_{t-j} - \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \right) \leq \left[E \left(\sum_{j=0}^N a_j \varepsilon_{t-j} - \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \right)^2 \right]^{\frac{1}{2}}$$

Assim,

$$\lim_{N \rightarrow +\infty} E \left(\sum_{j=0}^N a_j \varepsilon_{t-j} \right) = E \left(\sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \right)$$

ou seja,

$$\lim_{N \rightarrow +\infty} \sum_{j=0}^N a_j E(\varepsilon_{t-j}) = 0 = E \left(\sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \right),$$

isto é, $E(X_t) = 0, \forall t$.

Provemos agora que $\Gamma(t, s)$ depende apenas de $t - s$.

$$\begin{aligned} \Gamma(t, s) &= E(X_t X_s) = E \left(\sum_{j=0}^{+\infty} a_j \varepsilon_{t-j} \sum_{k=0}^{+\infty} a_k \varepsilon_{s-k} \right) \\ &= E \left(\sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} a_j a_k \varepsilon_{t-j} \varepsilon_{s-k} \right) = \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} a_j a_k E(\varepsilon_{t-j} \varepsilon_{s-k}), \end{aligned}$$

onde a troca da esperança com o limite se justifica de modo análogo ao anterior.

Consideremos $t > s$.

$$E(\varepsilon_{t-j} \varepsilon_{s-k}) = \begin{cases} \sigma^2, & t-j = s-k \\ 0, & \text{se não} \end{cases} = \begin{cases} \sigma^2, & j = t-s+k \\ 0, & \text{se não} \end{cases}.$$

Então

$$\Gamma(t, s) = \sigma^2 \sum_{k=0}^{+\infty} a_{t-s+k} a_k = \sigma^2 \sum_{k=0}^{+\infty} a_{|t-s|+k} a_k.$$

Se $t < s$, obtemos, de modo análogo,

$$E(\varepsilon_{t-j}\varepsilon_{s-k}) = \begin{cases} \sigma^2, & t-j = s-k \\ 0, & \text{se não} \end{cases} = \begin{cases} \sigma^2, & k = s-t+j \\ 0, & \text{se não} \end{cases}.$$

$$\text{e então } \Gamma(t, s) = \sigma^2 \sum_{j=0}^{+\infty} a_j a_{s-t+j} = \sigma^2 \sum_{j=0}^{+\infty} a_{|t-s|+j} a_j.$$

$$\text{Assim } \Gamma(t, s) = \sigma^2 \sum_{j=0}^{+\infty} a_j a_{|t-s|+j}, \forall t, s \in \mathbb{Z}.$$

Podemos pois concluir que X é um processo estocástico fracamente estacionário.

Consideremos agora o processo Y definido por

$$Y_t = \sum_{j=0}^m \theta_j \varepsilon_{t-j}, \forall t \in \mathbb{Z},$$

com $\theta_0 = 1$ e $\theta_1, \dots, \theta_m \in \mathbb{R}$.

É claro que Y é fracamente estacionário pois é um caso particular de X .

Mostremos que se ε é um processo Gaussiano real, então Y é fortemente estacionário.

Comecemos por provar que Y é um processo Gaussiano.

Sejam $t_1, \dots, t_n \in \mathbb{Z}$, ($t_1 < \dots < t_n$), onde, sem perda de generalidade, $t_i > t_{i-1} + m$, $i = 2, \dots, n$. Temos

$$\begin{aligned} (Y_{t_1}, \dots, Y_{t_n})' &= \left(\sum_{j=0}^m \theta_j \varepsilon_{t_1-j}, \dots, \sum_{j=0}^m \theta_j \varepsilon_{t_n-j} \right)' \\ &= \begin{bmatrix} \theta_0 & \dots & \theta_m & 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & 0 & \theta_0 & \dots & \theta_m & \dots & 0 \\ \dots & \dots \\ 0 & \dots & 0 & 0 & \dots & \dots & \dots & \theta_m \end{bmatrix} \begin{bmatrix} \varepsilon_{t_1} \\ \varepsilon_{t_1-1} \\ \dots \\ \varepsilon_{t_1-m} \\ \dots \\ \varepsilon_{t_n} \\ \dots \\ \varepsilon_{t_n-m} \end{bmatrix} \end{aligned}$$

onde as matrizes envolvidas são de ordens $(n, n(m+1))$ e $(n(m+1), 1)$, respectivamente.

Como $(\varepsilon_{t_1}, \dots, \varepsilon_{t_1-m}, \dots, \varepsilon_{t_n}, \dots, \varepsilon_{t_n-m})$ é Gaussiano, tem-se que $(Y_{t_1}, \dots, Y_{t_n})$ é Gaussiano por ser uma transformação linear de um vector aleatório Gaussiano.

Assim todas as leis de dimensão finita de Y são normais, logo Y é um processo Gaussiano.

Sendo Y Gaussiano e fracamente estacionário tem-se que Y é fortemente estacionário.

1.5 Espaços lineares

Tal como no caso das médias móveis infinitas, os modelos estocásticos que vamos estudar são combinações lineares dos valores sucessivos de um processo de segunda ordem. Este facto leva-nos à introdução de vários subespaços de L^2 .

Representamos por H ⁽⁹⁾ o menor subespaço fechado de L^2 gerado por uma família de variáveis aleatórias $X_\tau, \tau \in \mathbb{Z}$, e pelas constantes.

Assim, H é constituído pelas combinações lineares afins das variáveis aleatórias $X_\tau, \tau \in \mathbb{Z}$, e pelos limites destas combinações, no sentido de L^2 . Nota-se que H é um subespaço de Hilbert.

Representamos por \underline{H}_t o menor subespaço fechado de L^2 gerado por $X_\tau, \tau \leq t$, e pelas constantes. Assim, \underline{H}_t é constituído pelos limites, no sentido de L^2 , das combinações lineares afins $\sum_{i=1}^k A_i X_{\tau_i} + b, \tau_i \leq t$, com A_1, A_2, \dots, A_k matrizes e b um vector. Diremos que \underline{H}_t é o subespaço linear gerado pelo passado do processo no instante t .

De modo análogo, representamos por \overline{H}^t o menor subespaço fechado de L^2 gerado por $X_\tau, \tau \geq t$, e pelas constantes. Diremos que \overline{H}^t é o subespaço linear gerado pelo futuro do processo no instante t .

Todos estes subespaços estão contidos em H . Verifica-se facilmente que a sucessão de subespaços associada aos passados é cres-

⁹Ou H^X , se for necessário explicitar o processo X .

cente e que a sucessão de subespaços associada aos futuros é decrescente.

1.5.1 Operadores atraso e avanço

Seja $X = (X_t, t \in \mathbb{Z})$ um processo estacionário.

Sendo a noção de estacionaridade estabelecida à custa da comparação entre as propriedades de segunda ordem do processo X e as do processo X^* definido por $X_t^* = X_{t+h}$, $h \in \mathbb{Z}$, será natural estudar a aplicação que a X_t faz corresponder X_{t+1} ou aquela que a X_t faz corresponder X_{t-1} , $\forall t \in \mathbb{Z}$.

Suponhamos, sem perda de generalidade, que X é centrado e seja H o subespaço de L^2 associado a X , anteriormente definido.

Começemos por recordar o resultado seguinte, cuja prova pode ser encontrada em Azencott, Dacunha-Castelle (1984, p. 42):

Teorema. Sejam $(H, \langle \cdot, \cdot \rangle)$ e $(H^*, \langle \cdot, \cdot \rangle)$ dois espaços de Hilbert e U uma aplicação de $M \subset H$ em H^* tal que

$$\langle Ug, Uh \rangle = \langle g, h \rangle, \forall g, h \in M.$$

Então U prolonga-se de modo único, como aplicação isométrica, sobre o menor subespaço fechado contendo M .

Propriedade.

1. Há um operador isométrico único de H em H , que representamos por L (ou B , de 'backward'), que é tal que

$$LX_t = X_{t-1}, \forall t \in \mathbb{Z}.$$

2. Há um operador isométrico único de H em H , que representamos por F (de 'forward'), que é tal que

$$FX_t = X_{t+1}, \forall t \in \mathbb{Z}.$$

L é denominado operador *atraso* e F operador *avanço*.

Prova. Seja $M = \{X_t, t \in \mathbb{Z}\}$. Obviamente $M \subset H$.

Seja $H^* = H$. Notemos que esta escolha é possível uma vez que $LX_t = X_{t-1} \in H^*$.

Mostremos que L é isometria sobre M , isto é, que

$$\langle LX_t, LX_{t+h} \rangle = \langle X_t, X_{t+h} \rangle .$$

Ora, sendo X centrado (e supondo-o real para simplificar) obtemos

$$\langle LX_t, LX_{t+h} \rangle = E(LX_t LX_{t+h}) = E(X_{t-1} X_{t+h-1}) = \Gamma(h)$$

e

$$\langle X_t, X_{t+h} \rangle = E(X_t X_{t+h}) = \Gamma(h) .$$

O teorema enunciado garante-nos então que L se prolonga de modo único, e como isometria, sobre H .

A demonstração para F é análoga.

Observações.

1. Os operadores L e F conservam as constantes.

De facto, consideremos um processo $X = (X_t, t \in \mathbb{Z})$ tal que

$$X_t = a, \quad \forall t \in \mathbb{Z},$$

com $a \in \mathbb{R}^n$.

O processo assim definido está em H . Então

$$\forall t \in \mathbb{Z}, LX_t = X_{t-1} = a.$$

Analogamente para F .

2. L e F são operadores lineares.

A prova é imediata a partir da definição.

3. Estes operadores dependem obviamente do processo considerado. Representá-los-emos simplesmente por F e L para não sobrecarregar as notações.

4. Cada um destes operadores pode ser composto consigo mesmo. Por exemplo, sendo h um natural, a aplicação L^h é tal que

$$\forall t \in \mathbb{Z}, L^h(X_t) = L^{h-1}(X_{t-1}) = \dots = L(X_{t-h+1}) = X_{t-h}.$$

$$\text{Analogamente, } F^h(X_t) = X_{t+h}, \forall t \in \mathbb{Z}.$$

5. Os operadores L e F são inversos um do outro:

$$L \circ F = F \circ L = I.$$

As propriedades de isometria dos operadores L e F são extremamente úteis para estabelecer a estacionaridade de processos no espaço linear $H = H^X$.

De facto, se $Y = (Y_t, t \in \mathbb{Z})$ é um processo unidimensional tal que

$$\forall t, Y_t \in H^X, \quad \text{e} \quad \forall t, LY_t = Y_{t-1},$$

então Y é estacionário.

Com efeito,

$$\langle LY_t, 1 \rangle = \langle Y_t, 1 \rangle \iff E(Y_{t-1}) = E(Y_t) \iff E(Y_t) = m_Y.$$

Atendendo ainda à isometria de L , as covariâncias de Y são também independentes de t .

Mais geralmente, temos o resultado seguinte:

Propriedade. Seja Y um processo m -dimensional ($m \neq n$) cujas componentes $Y_{j,t}$ pertencem a H^X e tal que $LY_t = Y_{t-1}, \forall t \in \mathbb{Z}$.

Este processo é estacionário.

Prova. O resultado decorre da aplicação, a cada uma das componentes de Y , do facto de L ser uma isometria.

De facto, detalhando cada uma das componentes de Y tem-se, por L ser isométrico, para todos $t \in \mathbb{Z}, j \in \{1, \dots, m\}$,

$$\langle LY_{j,t}, L1 \rangle = \langle Y_{j,t}, 1 \rangle \iff E(Y_{j,t-1}) = E(Y_{j,t})$$

o que implica que o valor médio é constante.

Por outro lado, $\forall t \in \mathbb{Z}, \forall j, k \in \{1, \dots, m\}, \forall h \in \mathbb{Z}$, tem-se

$$\langle LY_{j,t}, LY_{k,t+h} \rangle = \langle Y_{j,t}, Y_{k,t+h} \rangle$$

isto é

$$\langle Y_{j,t-1}, Y_{k,t+h-1} \rangle = \langle Y_{j,t}, Y_{k,t+h} \rangle$$

ou seja,

$$E(Y_{j,t-1}Y_{k,t+h-1}) = E(Y_{j,t}Y_{k,t+h})$$

o que implica que as covariâncias não dependem senão da diferença h entre os índices temporais.

Tem-se naturalmente um resultado análogo usando o operador F .

Exemplos.

1. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estacionário. Defina-se o processo estocástico Y tal que $Y_t = X_{t-1}$.

Como $Y_t \in H^X$ e $LY_t = LX_{t-1} = X_{t-2} = Y_{t-1}$, podemos afirmar, tendo em conta a última propriedade, que Y é estacionário.

2. Consideremos o processo estocástico bidimensional estacionário

$X = \left(\begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix}, t \in \mathbb{Z} \right)$ e seja $Y = (Y_t = a_1X_{1,t} + a_2X_{2,t}, t \in \mathbb{Z})$

onde $a_1, a_2 \in \mathbb{R}$. O processo Y assim definido é estacionário. De facto, $Y \in H^X$ e, além disso, $\forall t \in \mathbb{Z}$,

$$LY_t = L(a_1X_{1,t} + a_2X_{2,t}) = a_1X_{1,t-1} + a_2X_{2,t-1} = Y_{t-1}.$$

3. Seja $X = (X_t, t \in \mathbb{Z})$ um processo unidimensional estacionário.

a) $Y_t = \sum_{p=0}^N a_p X_{t-p}$, $t \in \mathbb{Z}$, é um elemento de H^X e verifica

$LY_t = Y_{t-1}$. Consequentemente, o processo $Y = (Y_t, t \in \mathbb{Z})$ é um processo estacionário.

b) Seja $(a_i, i \in \mathbb{Z})$ uma sucessão de números reais tais que $\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$. O processo $Y = (Y_t, t \in \mathbb{Z})$ tal que

$$Y_t = \sum_{i=-\infty}^{+\infty} a_i X_{t-i}$$

é estacionário.

De facto basta reparar que $Y \in H^X$ e que $LY_t = Y_{t-1}$.

Observemos, no entanto, que há processos com valores em H^X estacionários e que não verificam a condição $LY_t = Y_{t-1}, \forall t \in \mathbb{Z}$.

Seja, por exemplo, $X = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco unidimensional de variância σ^2 e defina-se o processo Y por

$$Y_t = \varepsilon_{2t}, t \in \mathbb{Z}.$$

Este processo está em H^X e é estacionário pois

$$EY_t = 0, V(Y_t) = V(\varepsilon_{2t}) = \sigma^2, Cov(Y_t, Y_{t+h}) = 0, \forall h \neq 0.$$

Mas $LY_t = L\varepsilon_{2t} = \varepsilon_{2t-1}$ e $Y_{t-1} = \varepsilon_{2t-2}$.

A partir do processo X , podemos ainda considerar outro exemplo. Seja agora $Y = (Y_t, t \in \mathbb{Z})$ o processo definido por

$$Y_t = (-1)^t \varepsilon_t, t \in \mathbb{Z}.$$

Este processo está em H^X e é estacionário pois

$$EY_t = 0, V(Y_t) = V(\varepsilon_t) = \sigma^2, Cov(Y_t, Y_{t+h}) = 0, \forall h \neq 0.$$

Mas

$$\begin{aligned} LY_t &= \begin{cases} \varepsilon_{t-1} & , t \text{ par} \\ -\varepsilon_{t-1} & , t \text{ ímpar} \end{cases} \\ &= -Y_{t-1}. \end{aligned}$$

Este exemplo permite ainda mostrar que a soma de dois processos estacionários não é necessariamente um processo estacionário. De facto, sendo $X_t = \varepsilon_t$ e $Y_t = (-1)^t \varepsilon_t$, com ε ruído branco, tem-se

$$X_t + Y_t = \begin{cases} 2\varepsilon_t & , t \text{ par} \\ 0 & , t \text{ ímpar} \end{cases} .$$

Então

$$V(X_t + Y_t) = \begin{cases} 4\sigma^2 & , t \text{ par} \\ 0 & , t \text{ ímpar} \end{cases}$$

pelo que $(X_t + Y_t, t \in \mathbb{Z})$ não é um processo estacionário.

1.5.2 Melhor previsão linear baseada no passado

Seja $X = (X_t, t \in \mathbb{Z})$ um processo n -dimensional estacionário. A melhor previsão linear de X_{t+1} sabendo o passado \underline{H}_t , ou regressão linear de X_{t+1} sobre o seu passado, é, por definição, a projecção ortogonal de X_{t+1} sobre \underline{H}_t .

Representa-se por

$$E_L(X_{t+1}/\underline{H}_t) = E_L(X_{t+1}/X_t, X_{t-1}, \dots)$$

e é denominada esperança linear de X_{t+1} dado o seu passado. ⁽¹⁰⁾

Assim

$$\|X_{t+1} - E_L(X_{t+1}/\underline{H}_t)\| = \min_{Z \in \underline{H}_t} \|X_{t+1} - Z\| .$$

O processo $\hat{X} = (E_L(X_{t+1}/\underline{H}_t), t \in \mathbb{Z})$, formado pelas sucessivas previsões (ao horizonte 1), é estacionário.

De facto,

¹⁰ A esperança linear é uma esperança condicional (ver Anexo, parágrafo 2.4) particular, pelo que goza do mesmo tipo de propriedades. Para um estudo mais detalhado ver Gouriéroux, Monfort (1989), Vol. 2, (Annexe).

- $\forall t \in \mathbb{Z}, E_L(X_{t+1}/\underline{H}_t) \in \underline{H}_t \subset H$
- $L(E_L(X_{t+1}/\underline{H}_t)) = E_L(X_t/\underline{H}_{t-1}),$

por definição de esperança linear e porque L é um operador linear.

Definição. Chamamos *inovação do processo* X ao processo estocástico $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ formado pelos erros das previsões ao horizonte 1, isto é, tal que

$$\varepsilon_t = X_t - E_L(X_t/\underline{H}_{t-1}), \quad \forall t \in \mathbb{Z}.$$

O erro de previsão no instante t ao horizonte 1 é então $\varepsilon_{t+1} = X_{t+1} - E_L(X_{t+1}/\underline{H}_t)$; é, pois, o que não conseguimos prever, o que se produz 'de novo'. É esta razão que justifica a designação de *inovação*.

Tem-se

$$L\varepsilon_{t+1} = L(X_{t+1} - E_L(X_{t+1}/\underline{H}_t)) = X_t - E_L(X_t/\underline{H}_{t-1}) = \varepsilon_t$$

o que permite afirmar que a sucessão dos erros das previsões sucessivas ao horizonte 1 de um processo estacionário é um processo estacionário.

Podemos também observar que a inovação de X no instante $t+1$ é a parte de X_{t+1} não correlacionada com \underline{H}_t , ou seja, tem-se

$$\forall Y \in \underline{H}_t, \text{Cov}(\varepsilon_{t+1}, Y) = 0.$$

De facto, $\varepsilon_{t+1} = X_{t+1} - E_L(X_{t+1}/\underline{H}_t)$ é ortogonal a \underline{H}_t (basta ter em conta que $E_L(X_{t+1}/\underline{H}_t)$ é a projecção ortogonal de X_{t+1} sobre \underline{H}_t , como é ilustrado na figura 1.1).

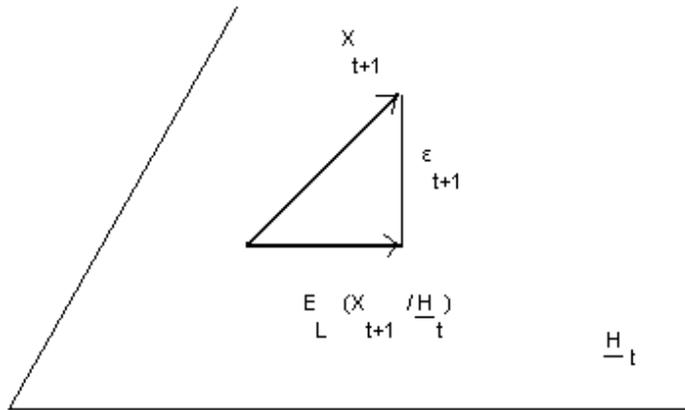


Fig. 1.1 - Vector inovação

Então

$$\forall Y \in \underline{H}_t, \langle Y, \varepsilon_{t+1} \rangle = E(Y' \varepsilon_{t+1}) = 0.$$

Além disso, $E(\varepsilon_{t+1}) = E(X_{t+1} - E_L(X_{t+1}/\underline{H}_t)) = 0$.

De facto,

$$\forall i \in \{1, \dots, n\}, e_i = (0, \dots, 0, 1, 0, \dots, 0)' \in \underline{H}_t$$

e então

$$\forall i \in \{1, \dots, n\}, E[(X_{t+1} - E_L(X_{t+1}/\underline{H}_t))' e_i] = 0 = E(\varepsilon_{i,t+1}).$$

Finalmente,

$$\forall Y \in \underline{H}_t, \text{Cov}(Y, \varepsilon_{t+1}) = E(Y \varepsilon'_{t+1}) - E(Y)E(\varepsilon'_{t+1}) = 0,$$

pois $Y \varepsilon'_{t+1}$ é uma matriz quadrada cujo elemento (i, j) é, com notações simplificadas,

$$Y_i \varepsilon_j = Y' e_i e_j' \varepsilon$$

e portanto

$$E(Y_i \varepsilon_j) = E\left(Y' e_i e_j' \varepsilon\right) = E\left[(e_j e_i' Y)' \varepsilon\right] = 0,$$

porque $e_j e_i' Y \in \underline{H}_t$.

Propriedade. O processo inovação ε é um ruído branco.

Prova. O processo $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é de segunda ordem pois $\varepsilon_t \in H^X$.

Por outro lado, como vimos na prova anterior, tem-se $\forall t \in \mathbb{Z}$, $E(\varepsilon_t) = 0$.

Além disso, $Cov(\varepsilon_t, \varepsilon_t) = V(\varepsilon_t) = E(\varepsilon_t \varepsilon_t') = \Sigma$, independente de t , pois já mostramos que ε é estacionário.

Finalmente, $\forall h > 0$,

$$\begin{aligned} Cov(\varepsilon_t, \varepsilon_{t+h}) &= E(\varepsilon_t \varepsilon_{t+h}') \\ &= E\left(\varepsilon_t (X_{t+h} - E_L(X_{t+h}/\underline{H}_{t+h-1}))'\right) = 0 \end{aligned}$$

pois $\varepsilon_t \in \underline{H}_t \subset \underline{H}_{t+h-1}$ e $X_{t+h} - E_L(X_{t+h}/\underline{H}_{t+h-1})$ é ortogonal a \underline{H}_{t+h-1} .

Esta propriedade mostra-nos como construir um ruído branco à custa de um qualquer processo estacionário.

1.6 Anexo - Espaços de Hilbert

Espaços com produto interno

Seja H um espaço vectorial sobre \mathbb{C} (resp., sobre \mathbb{R}).

Definição. H é um *espaço com produto interno* se para quaisquer $x, y \in H$ existe um número complexo (resp., real), que representamos por $\langle x, y \rangle$, tal que

- (a) $\langle x, y \rangle = \overline{\langle y, x \rangle}$, (resp., $\langle y, x \rangle$).
- (b) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$, $\forall x, y, z \in H$
- (c) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$, $\forall x, y \in H$, $\forall \alpha \in \mathbb{C}$ (resp., \mathbb{R})
- (d) $\langle x, x \rangle \geq 0$, $\forall x \in H$
- (e) $\langle x, x \rangle = 0$ se e só se $x = 0$.

Exemplos.

1. \mathbb{R}^n é um espaço com produto interno definindo

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i,$$

onde $x = (x_1, \dots, x_n)'$, $y = (y_1, \dots, y_n)'$, $x_i \in \mathbb{R}$, $y_i \in \mathbb{R}$, $i = 1, \dots, n$.

2. \mathbb{C}^n é um espaço com produto interno definindo

$$\langle w, z \rangle = \sum_{i=1}^n w_i \bar{z}_i,$$

onde $w = (w_1, \dots, w_n)'$, $z = (z_1, \dots, z_n)'$, $z_i \in \mathbb{C}$, $w_i \in \mathbb{C}$, $i = 1, \dots, n$.

Seja H um espaço com produto interno e consideremos a aplicação

$$\begin{aligned} \|\cdot\| &: H \rightarrow \mathbb{R}^+ \\ x &\mapsto \|x\| = \sqrt{\langle x, x \rangle}. \end{aligned}$$

Desigualdade de Cauchy-Schwartz. Tem-se

$$|\langle x, y \rangle| \leq \|x\| \|y\|, \forall x, y \in H$$

e

$$|\langle x, y \rangle| = \|x\| \|y\| \text{ se e só se } x = y \frac{\langle x, y \rangle}{\langle y, y \rangle}.$$

Definição. Os elementos x e y de H dizem-se ortogonais se $\langle x, y \rangle$ é nulo.

Propriedade. Se H é um espaço vectorial sobre \mathbb{C} (resp., \mathbb{R}) com produto interno, então

- (a) $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in H$.
- (b) $\|\alpha x\| = |\alpha| \|x\|$, $\forall x \in H, \forall \alpha \in \mathbb{C}$ (resp., \mathbb{R}).
- (c) $\|x\| \geq 0$, $\forall x \in H$
- (d) $\|x\| = 0$ se e só se $x = 0$.

A propriedade (a) é designada por *desigualdade triangular*.

Definição. A norma de um elemento x de H é $\|x\| = \sqrt{\langle x, x \rangle}$.

Definição. Uma sucessão $(x_n, n \in \mathbb{N})$ de elementos de H converge em norma para $x \in H$ se $\lim_{n \rightarrow +\infty} \|x_n - x\| = 0$.

Propriedade (Continuidade do produto interno). Se $(x_n, n \in \mathbb{N})$ e $(y_n, n \in \mathbb{N})$ são sucessões de elementos de H tais que $\lim_{n \rightarrow +\infty} \|x_n - x\| = 0$ e $\lim_{n \rightarrow +\infty} \|y_n - y\| = 0$, então

- (a) $\lim_{n \rightarrow +\infty} \|x_n\| = \|x\|$
- (b) $\lim_{n \rightarrow +\infty} \langle x_n, y_n \rangle = \langle x, y \rangle$.

Espaços de Hilbert

Sabemos que em \mathbb{R} ou \mathbb{C} a noção de sucessão convergente é equivalente à de sucessão de Cauchy. Este facto não é, no entanto, generalizável a qualquer espaço vectorial normado. De facto, uma sucessão de um tal espaço pode ser de Cauchy e não ser convergente. Os espaços de Hilbert são espaços com produto interno que mantêm tal equivalência.

Seja H um espaço com produto interno.

Definição. Uma sucessão $(x_n, n \in \mathbb{N})$ de elementos de H é uma *sucessão de Cauchy* se

$$\lim_{\substack{n \rightarrow +\infty \\ m \rightarrow +\infty}} \|x_n - x_m\| \longrightarrow 0.$$

De modo equivalente, $(x_n, n \in \mathbb{N})$ é uma sucessão de Cauchy se para cada $\epsilon > 0$, existe um número natural $N(\epsilon) > 0$ tal que

$$\|x_n - x_m\| < \epsilon, \text{ para todos os } m > n > N(\epsilon).$$

Definição. Um espaço de Hilbert H é um espaço vectorial com produto interno que é completo, i.e., em que toda a sucessão de Cauchy $(x_n, n \in \mathbb{N})$ converge em norma para um (único) elemento $x \in H$, ou seja,

$$\exists x \in H : \lim_{n \rightarrow +\infty} \|x_n - x\| = 0$$

A propriedade da unicidade do limite mantém-se neste contexto geral.

Propriedade. (Critério de Cauchy). Se $(x_n, n \in \mathbb{N})$ é uma sucessão de elementos de um espaço de Hilbert H então (x_n) converge em norma se e só se

$$\lim_{\substack{n \rightarrow +\infty \\ m \rightarrow +\infty}} \|x_n - x_m\| = 0.$$

Prova. A suficiência do critério de Cauchy é a reafirmação de H ser completo.

A necessidade é uma consequência elementar da desigualdade triangular. Assim, se $\|x_n - x\| \rightarrow 0$, então

$$\|x_n - x_m\| \leq \|x_n - x\| + \|x - x_m\| \rightarrow 0,$$

quando m e n tendem para $+\infty$. ■

Exemplo. O critério de Cauchy é essencialmente usado para verificar a convergência em norma de uma sucessão cujo limite não é especificado. Consideremos por exemplo a sucessão

$$S_n = \sum_{i=1}^n a_i X_i$$

onde $(X_n, n \in \mathbb{N})$ é uma sucessão de v.a. de lei $N(0,1)$, independentes. Considerando a norma em L^2 , é fácil verificar que, se $m > n$,

$$\|S_m - S_n\|^2 = \sum_{i=n+1}^m a_i^2$$

e então, pelo critério de Cauchy, (S_n) converge em *m.q.* se e só se para cada $\epsilon > 0$, existe $N(\epsilon) > 0$ tal que

$$\sum_{i=n+1}^m a_i^2 < \epsilon, \text{ para } m > n > N(\epsilon).$$

Assim, (S_n) converge em *m.q.* se e só se $\sum_{i=1}^{+\infty} a_i^2 < +\infty$.

Exemplos de Espaços de Hilbert

1. Verifiquemos que \mathbb{R}^k munido do produto interno $\langle x, y \rangle = \sum_{i=1}^k x_i y_i$ é completo.

Se $x_n = (x_{1,n}, \dots, x_{k,n})' \in \mathbb{R}^k, n \in \mathbb{N}$, é tal que

$$\|x_n - x_m\|^2 = \sum_{i=1}^k |x_{i,n} - x_{i,m}|^2 \longrightarrow 0,$$

quando m e n tendem para $+\infty$, então cada uma das componentes verifica

$$|x_{i,n} - x_{i,m}| \longrightarrow 0,$$

quando m e n tendem para $+\infty$.

Como \mathbb{R} é completo, existe $x_i \in \mathbb{R}$ tal que

$$|x_{i,n} - x_i| \longrightarrow 0, \text{ quando } n \longrightarrow +\infty$$

e então se $x = (x_1, \dots, x_k)$ tem-se

$$\|x_n - x\| \longrightarrow 0, \text{ quando } n \longrightarrow +\infty.$$

A prova de que o espaço vectorial \mathbb{C}^k sobre \mathbb{C} é completo é análoga.

Assim \mathbb{R}^k e \mathbb{C}^k são ambos espaços de Hilbert.

2. O espaço $L^2(\Omega, A, P)$

Consideremos um espaço de probabilidade (Ω, A, P) e a família $\mathcal{L}^2(\Omega, A, P)$ de todas as v.a.r. X definidas sobre (Ω, A, P) e tais que $E(X^2) < +\infty$.

O conjunto $\mathcal{L}^2(\Omega, A, P)$ munido das operações usuais de multiplicação por um escalar real e adição de v.a. é um espaço vectorial verificando-se, em particular,

$$E[(aX)^2] = a^2 E(X^2) < +\infty, \forall a \in \mathbb{R}, \forall X \in \mathcal{L}^2(\Omega, A, P)$$

$$E(X + Y)^2 \leq 2E(X^2) + 2E(Y^2) < +\infty, \forall X, Y \in \mathcal{L}^2(\Omega, A, P),$$

pois $(X + Y)^2 \leq 2X^2 + 2Y^2$.

Notemos também que o elemento zero de $\mathcal{L}^2(\Omega, A, P)$ é a v.a. identicamente nula sobre Ω .

Para quaisquer dois elementos $X, Y \in \mathcal{L}^2(\Omega, A, P)$ definamos

$$\langle X, Y \rangle = E(XY).$$

Esta aplicação verifica todas as propriedades de um produto interno com excepção da última. De facto, se $\langle X, X \rangle = 0$ não podemos deduzir que $X(\omega) = 0$ para todo $\omega \in \Omega$. Apenas podemos concluir que $P(X = 0) = 1$.

Esta dificuldade é ultrapassada introduzindo uma relação de equivalência em $\mathcal{L}^2(\Omega, A, P)$. Dizemos que $X, Y \in \mathcal{L}^2(\Omega, A, P)$ são equivalentes se $P(X = Y) = 1$.

O conjunto quociente por esta relação de equivalência, denotado $L^2(\Omega, A, P)$ ou $L^2_{\mathbb{R}}(\Omega, A, P)$, é assim composto por classes de v.a. tais que duas v.a. da mesma classe são iguais com probabilidade 1. Neste novo espaço o operador anterior é já um produto interno.

A convergência em norma de uma sucessão (X_n) de elementos de L^2 para um limite X equivale então a

$$\lim_{n \rightarrow +\infty} \|X_n - X\|^2 = \lim_{n \rightarrow +\infty} E(|X_n - X|^2) = 0.$$

A convergência em norma de X_n para X em L^2 é, de modo natural, chamada convergência em média quadrática (m.q.).

L^2 é um espaço de Hilbert. De facto, falta apenas provar que é completo, isto é, se (X_n) é uma sucessão de elementos de L^2 e tal que $\lim_{\substack{n \rightarrow +\infty \\ m \rightarrow +\infty}} \|X_m - X_n\|^2 = 0$, então existe $X \in L^2$ tal que

$$X_n \xrightarrow[n \rightarrow +\infty]{} X \quad (m.q.).$$

Começemos por estabelecer o seguinte resultado auxiliar:

Lema. Se $(X_n) \in L^2$ e $\|X_{n+1} - X_n\| \leq 2^{-n}$, $n = 1, 2, \dots$, então existe uma v.a. X em L^2 tal que $X_n \rightarrow X$ com probabilidade 1.

Prova. Seja $X_0 = 0$. Então $X_n = \sum_{j=1}^n (X_j - X_{j-1})$.

Pelo teorema da convergência monótona e pela desigualdade de Cauchy-Schwarz tem-se

$$\begin{aligned} E \left(\sum_{j=1}^{+\infty} |X_j - X_{j-1}| \right) &= \sum_{j=1}^{+\infty} E(|X_j - X_{j-1}|) \leq \sum_{j=1}^{+\infty} \|X_j - X_{j-1}\| \\ &\leq \|X_1\| + \sum_{j=1}^{+\infty} 2^{-j} < +\infty. \end{aligned}$$

Concluimos então que $\lim_{n \rightarrow +\infty} \sum_{j=1}^n |X_j - X_{j-1}|$ existe e é finito com probabilidade 1; conseqüentemente o mesmo acontece a

$$\lim_{n \rightarrow +\infty} \sum_{j=1}^n (X_j - X_{j-1}) = \lim_{n \rightarrow +\infty} X_n.$$

Proposição. $L^2(\Omega, A, P)$ é completo.

Prova. Se (X_n) é uma sucessão de Cauchy em L^2 então existem inteiros n_1, n_2, \dots tais que $n_1 < n_2 < \dots$ e

$$\|X_n - X_m\| \leq 2^{-k}, \text{ para } m, n > n_k.$$

Pela proposição anterior existe uma v.a. X tal que $X_{n_k} \rightarrow X$ com probabilidade 1, quando $k \rightarrow +\infty$.

Por outro lado.

$$\|X_n - X\|^2 = \int |X_n - X|^2 dP = \int \liminf_k |X_n - X_{n_k}|^2 dP$$

e, pelo lema de Fatou

$$\|X_n - X\|^2 \leq \liminf_k \|X_n - X_{n_k}\|^2.$$

Sendo (X_n) uma sucessão de Cauchy conclui-se que

$$\lim_{n \rightarrow +\infty} \|X_n - X\|^2 = 0.$$

O facto de $E|X|^2 < +\infty$ decorre da desigualdade triangular

$$\|X\| \leq \|X_n - X\| + \|X_n\|$$

e do facto de (X_n) ser uma sucessão de L^2 convergente para X .

3. Espaços L^2 complexos: $L^2_{\mathbb{C}}(\Omega, \mathcal{A}, P)$

O espaço das v.a. com valores complexos X definidas sobre (Ω, \mathcal{A}, P) e tais que $E|X|^2 < +\infty$ é um espaço de Hilbert complexo se definirmos um produto interno por

$$\langle X, Y \rangle = E(X\bar{Y}),$$

(identificando as v.a. X e Y tais que $P(X = Y) = 1$).

4. Espaços $L^2_{n, \mathbb{R}}(\Omega, \mathcal{A}, P)$

Um vector aleatório real de dimensão n , X , definido sobre (Ω, \mathcal{A}, P) está em $\mathcal{L}^2_n(\Omega, \mathcal{A}, P)$ se $E\|X\|^2 < +\infty$.

Como num espaço de dimensão finita todas as normas são equivalentes (Kolmogorov & Fomin, 1975, p. 152), esta definição é independente da norma que estiver definida em \mathbb{R}^n .

A relação de igualdade $P - q.c.$

$$X \sim Y \iff P(X = Y) = 1$$

é também uma relação de equivalência sobre o espaço vectorial dos vectores aleatórios reais definidos sobre (Ω, \mathcal{A}, P) .

Analogamente aos casos anteriores vamos considerar o conjunto quociente de $\mathcal{L}_n^2(\Omega, \mathcal{A}, P)$ por esta relação de equivalência, nomeadamente o espaço $L_n^2(\Omega, \mathcal{A}, P)$.

$L_n^2(\Omega, \mathcal{A}, P)$ é um espaço vectorial normado. De facto, a aplicação definida sobre $L_n^2(\Omega, \mathcal{A}, P)$ por

$$\forall X, Y \in L_n^2(\Omega, \mathcal{A}, P), \langle X, Y \rangle = E(X'Y)$$

tem as propriedades de produto interno. A norma correspondente é então

$$\forall X \in L_n^2(\Omega, \mathcal{A}, P), \|X\| = \left(E\|X\|^2\right)^{\frac{1}{2}}$$

A convergência no sentido de L_n^2 é, conseqüentemente, definida por

$$X_n \xrightarrow[n \rightarrow +\infty]{L_n^2} X \iff E\left(\|X_n - X\|^2\right) \xrightarrow[n \rightarrow +\infty]{} 0. \blacksquare$$

Outros aspectos interessantes sobre os espaços de Hilbert podem ser encontrados em Rudin (1987).

Capítulo 2

Resumo das ligações temporais

Neste capítulo, estudaremos alguns resumos de segunda ordem de processos estocásticos estacionários. Limitar-nos-emos aos processos unidimensionais.

2.1 Autocovariância e autocorrelação

Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico estacionário com valores em \mathbb{R} . A sua *função de autocovariância* é, como já vimos, a aplicação $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ tal que $\gamma(h) = \text{Cov}(X_t, X_{t+h})$.

A proposição seguinte estabelece algumas propriedades básicas da função de autocovariância.

Proposição. Se γ é a função de autocovariância de um processo estacionário $X = (X_t, t \in \mathbb{Z})$ então

- i) $\gamma(0) \geq 0$
- ii) $|\gamma(h)| \leq \gamma(0), \forall h \in \mathbb{Z}$
- iii) γ é uma função par: $\gamma(h) = \gamma(-h), \forall h \in \mathbb{Z}$.

Prova. A primeira propriedade é consequência óbvia de $V(X_t) \geq 0$.

A segunda propriedade é consequência imediata da desigualdade de Cauchy-Schwartz:

$$|Cov(X_t, X_{t+h})| \leq \sqrt{V(X_t)}\sqrt{V(X_{t+h})}.$$

A terceira propriedade é estabelecida tendo em conta que

$$\gamma(h) = Cov(X_t, X_{t+h}) = Cov(X_{t+h}, X_t) = \gamma(-h). \quad \blacksquare$$

Como consequência do teorema de caracterização das funções de covariância, apresentado no capítulo anterior, decorre um resultado equivalente de caracterização para as funções de autocovariância.

Comecemos por apresentar a seguinte definição:

Definição. Uma função $k : \mathbb{Z} \rightarrow \mathbb{R}$ diz-se *definida não-negativa*, ou *de tipo positivo* ⁽¹⁾ quando

$$\sum_{i=1}^n \sum_{j=1}^n a_i k(i-j) a_j \geq 0, \quad \forall n \in \mathbb{N}, \quad \forall a = (a_1, \dots, a_n)' \in \mathbb{R}^n.$$

Teorema. (Caracterização de funções de autocovariância)

Uma função $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ interpreta-se como a função de autocovariância de um processo estacionário se e só se

i) γ é uma função par: $\gamma(h) = \gamma(-h), \forall h \in \mathbb{Z}$.

ii) γ é de tipo positivo:

$$\forall n \in \mathbb{N}, \forall (u_1, \dots, u_n)' \in \mathbb{R}^n, \sum_{i=1}^n \sum_{j=1}^n u_i u_j \gamma(i-j) \geq 0.$$

Prova. Para provar que uma função γ par e de tipo positivo se interpreta como função de autocovariância de um processo estacionário, basta associar-lhe um processo Gaussiano centrado e cuja função de autocovariância seja tal que $\Gamma(s, t) = \gamma(t-s), t, s \in \mathbb{Z}$.

¹Referindo-nos a uma função preferiremos a designação “de tipo positivo”.

Inversamente, se γ é a função de autocovariância de um processo estacionário, γ é obviamente par.

Por outro lado, se $u = (u_1, \dots, u_n)' \in \mathbb{R}^n$, $t = (t_1, \dots, t_n)' \in \mathbb{Z}^n$ e

$$Z_t = (X_{t_1} - EX_{t_1}, \dots, X_{t_n} - EX_{t_n})'$$

então

$$0 \leq V(u'Z_t) = u'E(Z_t Z_t')u = u'\Gamma_n u = \sum_{i=1}^n \sum_{j=1}^n u_i u_j \gamma(t_i - t_j)$$

com Γ_n a matriz de autocovariância de $(X_{t_1}, \dots, X_{t_n})'$. ■

Exemplo. Seja $k : \mathbb{Z} \rightarrow \mathbb{R}$ a função definida por

$$k(h) = \begin{cases} 1 & , h = 0 \\ \rho & , h = -1, h = 1 \\ 0 & , \text{outros valores de } h. \end{cases}$$

Prova-se que k não é função de autocovariância se $|\rho| > \frac{1}{2}$.

Para tal basta considerar a matriz $K_n = [k(i-j)]_{i,j=1,\dots,n}$ e provar que se $\rho > \frac{1}{2}$ e $a = (1, -1, 1, \dots)' \in \mathbb{R}^n$, então

$$a'K_n a = n - 2(n-1)\rho < 0 \text{ para } n < \frac{2\rho}{2\rho-1}.$$

A mesma conclusão pode ser obtida quando $\rho < -\frac{1}{2}$ e considerando $a = (1, 1, \dots, 1)' \in \mathbb{R}^n$.

A função de autocovariância fornece-nos dois tipos de informação:

a) informação sobre a variabilidade da série (dispersão), por intermédio de $\gamma(0) = V(X_t)$.

b) informação sobre as ligações temporais, através das autocovariâncias $\gamma(1), \gamma(2), \dots$, e consequentemente sobre as correlações.

Definição. Se $\gamma(0) > 0$, chama-se função de autocorrelação à aplicação $\rho : \mathbb{Z} \rightarrow \mathbb{R}$ definida por $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.

Observemos que a função ρ mede a correlação existente entre X_t e X_{t+h} ; de facto,

$$\text{corr}(X_t, X_{t+h}) = \frac{\text{Cov}(X_t, X_{t+h})}{\sqrt{V(X_t)}\sqrt{V(X_{t+h})}} = \frac{\gamma(h)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \rho(h).$$

O teorema anterior tem então a seguinte forma em termos da função de autocorrelação.

Propriedade. Uma função $\rho : \mathbb{Z} \rightarrow \mathbb{R}$ interpreta-se como a função de autocorrelação de um processo estacionário se e só se

- i) ρ é par
- ii) ρ é de tipo positivo
- iii) $\rho(0) = 1$.

Ao gráfico da função de autocorrelação chamamos *correlograma*. Este gráfico, devido à paridade da função, restringe-se muitas vezes a \mathbb{N}_0 (por simetria obtém-se o gráfico completo).

Na figura seguinte ilustramos o aspecto usual de um correlograma.

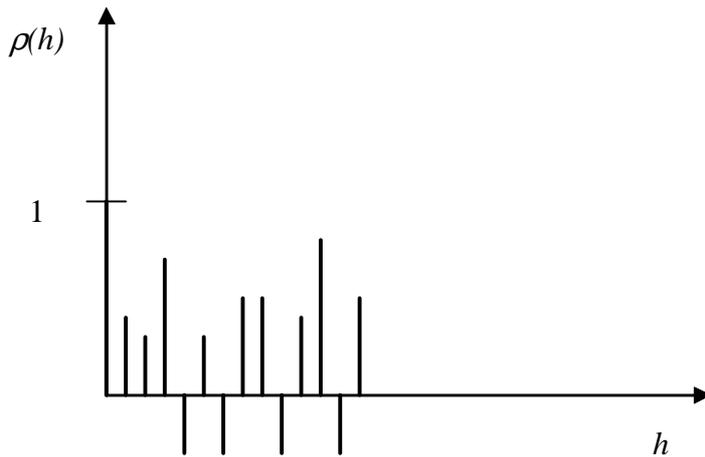


Fig. 2.1 - Correlograma

Exemplos.

1. Se $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância σ^2 , a sua função de autocorrelação é

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 0, & h \neq 0 \end{cases}, \quad h \in \mathbb{Z}.$$

2. Seja $X_t = \varepsilon_t - \varepsilon_{t-12}$, $t \in \mathbb{Z}$, onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância σ^2 .

Tem-se

$$E(X_t) = 0, V(X_t) = 2\sigma^2$$

e

$$E(X_t X_{t+h}) = \begin{cases} 2\sigma^2, & h = 0 \\ -\sigma^2, & h = -12, h = 12 \\ 0, & h \in \mathbb{Z} \setminus \{-12, 0, 12\} \end{cases}$$

Assim, a função de autocorrelação de X é

$$\rho(h) = \begin{cases} 1, & h = 0 \\ -\frac{1}{2}, & h = -12, h = 12 \\ 0, & h \in \mathbb{Z} \setminus \{-12, 0, 12\} \end{cases}.$$

3. Seja $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância σ^2 e $X = (X_t, t \in \mathbb{Z})$ o processo estocástico definido por

$$X_t = \varepsilon_t + \varphi\varepsilon_{t-1} + \dots + \varphi^i\varepsilon_{t-i} + \dots = \sum_{i=0}^{+\infty} \varphi^i\varepsilon_{t-i}$$

com $|\varphi| < 1$.

Sabemos que X é um processo estacionário. Tem-se

$$\gamma(h) = \sigma^2 \frac{\varphi^{|h|}}{1 - \varphi^2}, \quad h \in \mathbb{Z}.$$

A função de autocorrelação de X é, pois,

$$\rho(h) = \varphi^{|h|}, \quad h \in \mathbb{Z}.$$

O correlograma tem então as formas seguintes, consoante $\varphi > 0$ ou $\varphi < 0$:

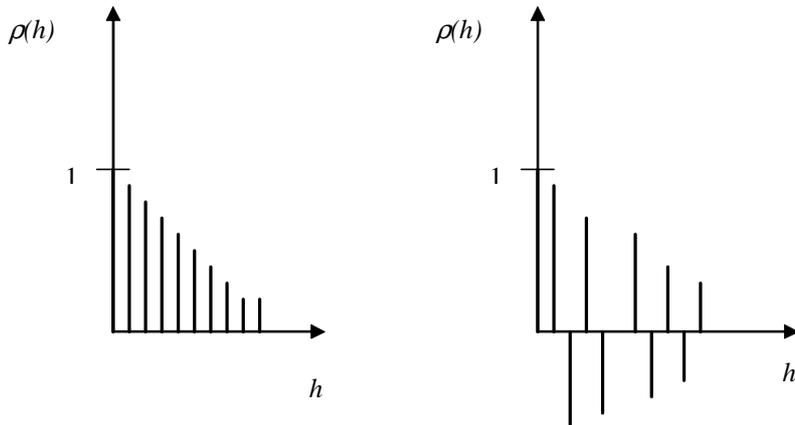


Fig. 2.2 - Correlograma de X

Matriz de correlação de m componentes consecutivas do processo X

A matriz de correlação de m componentes consecutivas do processo, $X_t, X_{t+1}, \dots, X_{t+m-1}$, é a matriz, que representaremos por $R(m)$, cujos elementos são as correlações entre as sucessivas variáveis, isto é,

$$\begin{aligned}
 & \left[\begin{array}{cccc} \text{corr}(X_t, X_t) & \text{corr}(X_t, X_{t+1}) & \dots & \dots \\ \text{corr}(X_{t+1}, X_t) & \text{corr}(X_{t+1}, X_{t+1}) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \text{corr}(X_{t+m-1}, X_{t+m-1}) \end{array} \right] \\
 & = \left[\begin{array}{ccccc} 1 & \rho(1) & \rho(2) & \dots & \rho(m-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(m-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(m-3) \\ \dots & \dots & \dots & \dots & \dots \\ \rho(m-1) & \rho(m-2) & \rho(m-3) & \dots & 1 \end{array} \right]
 \end{aligned}$$

Notemos que os elementos de uma ‘diagonal’ paralela à diagonal principal são iguais. Além disso, é uma matriz simétrica. Uma

matriz com estas características é designada *matriz de Toeplitz*.

Como consequência do facto de ρ ser do tipo positivo, decorre que a matriz de autocorrelação $R(m)$ é uma matriz semi-definida positiva para todo o $m \in \mathbb{N}$.

Assim, para todo $m \in \mathbb{N}$, $R(m)$ tem todos os menores principais não negativos.

A proposição seguinte é então uma consequência da última propriedade sobre a função ρ .

Propriedade. Uma função ρ pode interpretar-se como uma função de autocorrelação se e só se, para todo $m \in \mathbb{N}$, as matrizes $R(m)$ associadas são matrizes de Toeplitz tais que todos os seus menores principais são não negativos (*i.e.*, matrizes de Toeplitz semi-definidas positivas).

Notemos que as autocorrelações sucessivas estão, assim, sujeitas a uma infinidade de restrições. Detalhemos as três primeiras condições.

i) $\det R(1) = 1$.

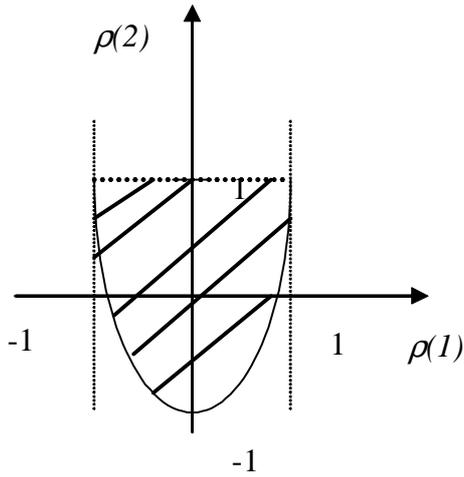
ii) $\det R(2) = \begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix} = 1 - \rho^2(1) \geq 0 \iff |\rho(1)| \leq 1$
(propriedade clássica de um coeficiente de correlação).

iii) $\det R(3) = \begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}$
 $= [1 - \rho(2)] [1 - 2\rho^2(1) + \rho(2)] \geq 0,$

o que é equivalente a

$$\rho(2) < 1 \text{ e } \rho(2) \geq 2\rho^2(1) - 1.$$

O domínio de variação de $(\rho(1), \rho(2))$ está ilustrado na figura seguinte.

Fig. 2.3 - Variação de $(\rho(1), \rho(2))$

Assim, se a correlação de ordem 1 é muito elevada (isto é, muito próxima de 1) o mesmo acontece à correlação de ordem 2; ou seja, não pode haver uma grande diferença entre $\rho(1)$ e $\rho(2)$ quando $\rho(1)$ é elevado.

Se $\rho(2) = 0$, isto é, se há ausência de correlação à ordem 2, então

$$2\rho^2(1) - 1 \leq 0 \Rightarrow \rho^2(1) \leq \frac{1}{2} \Rightarrow |\rho(1)| \leq \frac{1}{\sqrt{2}}$$

ou seja, só pode haver correlação à ordem 2 nula se $|\rho(1)| \leq \frac{1}{\sqrt{2}}$.

Exemplos.

1. Seja $X = (X_t, t \in \mathbb{Z})$ o processo estocástico dado por $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância σ^2 e $\theta \in \mathbb{R}$.

Para $h \geq 0$, tem-se

$$\rho(h) = \begin{cases} 1, & h = 0 \\ -\frac{\theta}{1 + \theta^2}, & h = 1 \\ 0, & h \geq 2. \end{cases}$$

O estudo de $\rho(1) = \rho_\theta(1)$ mostra que

- ρ_θ é uma função ímpar
- ρ_θ é decrescente se $\theta \in [-1, 1]$
- ρ_θ é crescente se $\theta \in \mathbb{R} \setminus [-1, 1]$
- $\rho_0 = 0$; $\rho_1 = -\frac{1}{2}$.
- $\lim_{\theta \rightarrow +\infty} \rho_\theta = 0$.

Na figura 2.4 esboçamos o gráfico desta função.

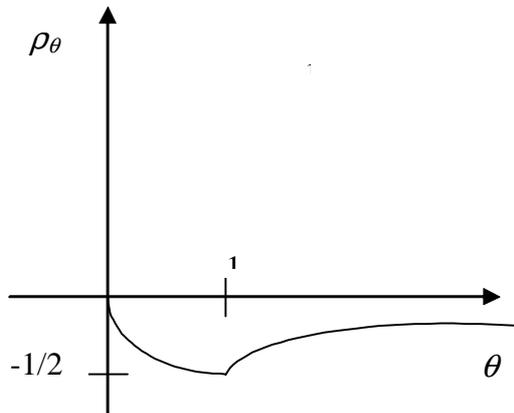


Fig. 2.4 - Gráfico de ρ_θ

Assim, o domínio de variação de $\rho(1) = \rho_\theta(1)$, $\theta \in \mathbb{R}$, é $[-\frac{1}{2}, \frac{1}{2}]$. Isto é, $\rho(1)$ tem, de facto, um domínio de variação muito limitado. Notemos ainda que $[-\frac{1}{2}, \frac{1}{2}] \subset [-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ o que vem ao encontro da ausência de correlações de ordem superior.

2. Seja $X = (X_t, t \in \mathbb{Z})$ o processo estocástico tal que $X_t = \sum_{i=0}^{+\infty} \varphi^i \varepsilon_{t-i}$, com $|\varphi| < 1$ e onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância σ^2 .

Vimos já que $\rho(h) = \varphi^h$, para $h \geq 0$.

Estudemos o domínio de variação de $\rho(1)$ e $\rho(2)$, isto é, de (φ, φ^2) , $|\varphi| < 1$.

A condição vista atrás, $\rho(2) - 2\rho^2(1) + 1 \geq 0$, é equivalente a $|\varphi| \leq 1$.

Então, o domínio de variação é o arco de parábola apresentado na figura seguinte.

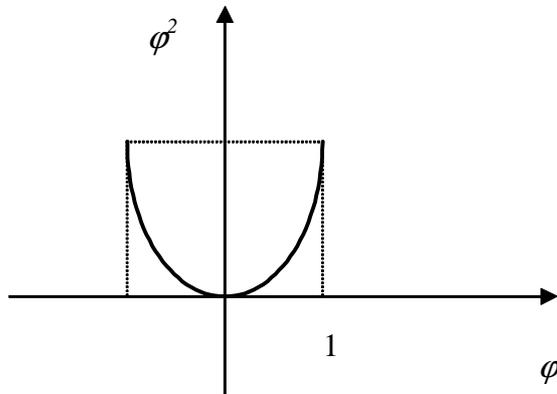


Fig. 2.5 - Domínio de variação (φ, φ^2)

2.2 Medida e densidade espectrais

Utilizando os resultados gerais da transformada de Fourier obtém-se um resumo das ligações temporais dum processo estocástico estacionário equivalente à função de autocovariância. O teorema de Herglotz-Bochner, que apresentamos de seguida, está na base deste estudo.

Teorema. (Herglotz-Bochner) Uma função $g : \mathbb{Z} \rightarrow \mathbb{C}$ é de tipo positivo, isto é, verifica

$$\forall n \in \mathbb{N}, \forall (u_1, \dots, u_n) \in \mathbb{C}^n, \sum_{j=1}^n \sum_{k=1}^n u_j \bar{u}_k g(j-k) \geq 0,$$

se e só se pode escrever-se na forma

$$g(h) = \int_{-\pi}^{+\pi} \exp(i\omega h) d\mu(\omega)$$

onde μ é uma medida positiva limitada sobre $[-\pi, \pi[$. A medida μ é única e denomina-se *medida espectral* associada a g .

Prova. i) Consideremos u_j da forma

$$u_j = \exp(-ij\omega), \quad j = 1, \dots, n.$$

Então

$$\sum_{j=1}^n \sum_{k=1}^n u_j \bar{u}_k g(j-k) \geq 0 \iff \sum_{j=1}^n \sum_{k=1}^n \exp[-i(j-k)\omega] g(j-k) \geq 0.$$

Seja

$$\begin{aligned} G_n(\omega) &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{k=1}^n \exp[-i(j-k)\omega] g(j-k) \\ &= \frac{1}{2\pi n} \sum_{k=1-n}^{n-1} \exp(-ik\omega) g(k) [n - |k|] \\ &= \frac{1}{2\pi} \sum_{k=1-n}^{n-1} \exp(-ik\omega) g(k) \left[1 - \frac{|k|}{n}\right] \end{aligned}$$

Como sabemos $\{\exp(-ik\omega), k \in \mathbb{Z}\}$ é uma base ortogonal de $L^2_{\mathbb{C}}([-\pi, \pi[, \mathcal{B}_{[-\pi, \pi[}, \lambda)$, onde λ é a medida de Lebesgue (cf. Koopmans, 1974, p.19); consequentemente

$$\int_{-\pi}^{+\pi} \exp[i(h-k)\omega] d\omega = \begin{cases} 0 & h \neq k \\ 2\pi & h = k \end{cases}.$$

Assim, $I = \int_{-\pi}^{+\pi} G_n(\omega) \exp(i\omega h) d\omega$ é tal que

$$\begin{aligned} I &= \frac{1}{2\pi} \sum_{k=1-n}^{n-1} g(k) \left[1 - \frac{|k|}{n}\right] \int_{-\pi}^{+\pi} \exp[i(h-k)\omega] d\omega \\ &= g(h) \left(1 - \frac{|h|}{n}\right). \end{aligned}$$

Então, considerando $\mu_n = G_n \lambda_{[-\pi, \pi[}$, vem

$$\int_{-\pi}^{+\pi} \exp(ih\omega) d\mu_n(\omega) = g(h) \left(1 - \frac{|h|}{n}\right).$$

Notemos que μ_n é uma medida positiva (pois $G_n \geq 0$) de massa total igual a

$$\mu_n([- \pi, \pi[) = \int_{-\pi}^{+\pi} d\mu_n(\omega) = g(0)$$

(> 0 , dada a positividade da medida) independente de n .

Usando um resultado de compacidade relativa das medidas sobre $[-\pi, \pi[$ (Billingsley, 1968, p.35) pode extrair-se da sucessão μ_n uma subsucessão μ_{n_k} que converge fracamente para uma medida μ com a mesma massa total. Sendo a convergência fraca equivalente à convergência dos integrais de funções contínuas com suporte compacto deduz-se, uma vez que a aplicação $\omega \mapsto \exp(ih\omega)$ é contínua, pelo lema de Helly-Bray (Loève, 1977, p.182), que

$$\begin{aligned} \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu(\omega) &= \lim_{n_k \rightarrow +\infty} \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu_{n_k}(\omega) \\ &= \lim_{n_k \rightarrow +\infty} g(h) \left(1 - \frac{|h|}{n_k}\right) = g(h), \end{aligned}$$

ou seja, g admite a representação integral anunciada para uma medida μ nas condições do enunciado. Admite-se a unicidade desta medida.

ii) Inversamente, seja g uma função definida por

$$g(h) = \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu(\omega)$$

onde μ é uma medida positiva limitada sobre $[-\pi, \pi[$. Tem-se

$$\begin{aligned} \sum_{j=1}^n \sum_{k=1}^n u_j \bar{u}_k g(j-k) &= \sum_{j=1}^n \sum_{k=1}^n u_j \bar{u}_k \int_{-\pi}^{+\pi} \exp[i(j-k)\omega] d\mu(\omega) \\ &= \int_{-\pi}^{+\pi} \sum_{j=1}^n u_j \exp(ij\omega) \sum_{k=1}^n \bar{u}_k \exp(-ik\omega) d\mu(\omega) \\ &= \int_{-\pi}^{+\pi} \left| \sum_{j=1}^n u_j \exp(ij\omega) \right|^2 d\mu(\omega) \geq 0. \blacksquare \end{aligned}$$

Medida espectral

As funções de autocovariância e de autocorrelação de um processo estacionário estão, como já vimos, nas condições do teorema de Herglotz-Bochner. Assim podemos associar-lhes, de maneira única, uma representação integral para uma medida positiva e limitada sobre $[-\pi, \pi[$.

Definição. Seja $X = (X_t, t \in \mathbb{Z})$ um processo real estacionário. Chamamos *medida espectral de X* à medida μ associada à sua função de autocovariância.

Assim, a medida espectral do processo X é a única medida simétrica, positiva, limitada sobre $[-\pi, \pi[$ tal que

$$\gamma(h) = \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu(\omega), h \in \mathbb{Z}.$$

A condição suplementar de simetria provém do facto de γ ser real; com efeito, a condição $\gamma(h) = \overline{\gamma(h)}$, $\forall h \in \mathbb{Z}$, é equivalente a

$$\begin{aligned} \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu(\omega) &= \int_{-\pi}^{+\pi} \exp(-ih\omega) d\mu(\omega) \\ &= \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu^s(\omega) \end{aligned}$$

onde μ^s designa a medida simétrica de μ , isto é, tal que

$$\forall A \in \mathcal{B}_{[-\pi, \pi[}, \mu^s(A) = \mu(-A).$$

Como $\{\exp(-ik\omega), k \in \mathbb{Z}\}$ é uma base do espaço $L^2_{\mathbb{C}}([-\pi, \pi[, \mathcal{B}_{[-\pi, \pi[}, \lambda)$, obtemos $\mu = \mu^s$ (da unicidade de μ).

O teorema de Herglotz-Bochner e a equivalência anterior permitem-nos ainda afirmar que se μ é uma medida positiva, simétrica, limitada sobre $[-\pi, \pi[$ a função definida por

$$\gamma(h) = \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu(\omega)$$

é real, par, de tipo positivo e pode interpretar-se como a função de autocovariância de um processo estacionário; é então possível exprimi-la na forma

$$\gamma(h) = \int_{-\pi}^{+\pi} \cos(h\omega) d\mu(\omega).$$

É, assim, equivalente dar uma função de autocovariância de um processo estocástico real ou uma medida positiva, simétrica e limitada sobre $[-\pi, \pi[$.

Densidade espectral

A eventual continuidade absoluta de μ relativamente à medida de Lebesgue conduz-nos à definição seguinte.

Definição. Um processo estacionário $X = (X_t, t \in \mathbb{Z})$ admite uma *densidade espectral* f se a sua medida espectral μ é absolutamente contínua relativamente à medida de Lebesgue λ sobre $[-\pi, \pi[$, de densidade f .

Exemplo. Seja $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância σ^2 .

Sabemos que

$$\gamma(h) = \begin{cases} \sigma^2, & h = 0 \\ 0, & h \neq 0 \end{cases}.$$

Como $\int_{-\pi}^{+\pi} \exp(ih\omega) d\omega = \begin{cases} 2\pi, & h = 0 \\ 0, & h \neq 0 \end{cases}$ podemos escrever

$$\begin{aligned} \gamma(h) &= \frac{\sigma^2}{2\pi} \int_{-\pi}^{+\pi} \exp(ih\omega) d\omega = \int_{-\pi}^{+\pi} \exp(ih\omega) \frac{\sigma^2}{2\pi} d\omega \\ &= \int_{-\pi}^{+\pi} \exp(ih\omega) d\mu(\omega), \end{aligned}$$

com $\mu = \frac{\sigma^2}{2\pi} \lambda_{[-\pi, \pi[}$.

Assim, a densidade espectral do ruído branco é

$$f(\omega) = \frac{\sigma^2}{2\pi}, \quad \omega \in [-\pi, \pi[.$$

Inversamente, todo o processo centrado admitindo uma densidade espectral constante $f(\omega) = k > 0$, $\omega \in [-\pi, \pi[$, é um ruído branco.

Com efeito,

$$\gamma(h) = \int_{-\pi}^{+\pi} \exp(ih\omega) k d\omega = \begin{cases} 2k\pi, & h = 0 \\ 0, & h \neq 0 \end{cases}$$

ou seja, $\gamma(h) = 0$ para $h \neq 0$, e $\gamma(0) = k$ (constante), isto é, o processo é um ruído branco.

O resultado seguinte estabelece uma condição necessária e suficiente para que exista densidade espectral.

Teorema. O processo estacionário real $X = (X_t, t \in \mathbb{Z})$ tem uma densidade espectral de quadrado integrável em relação à medida de Lebesgue sobre $[-\pi, \pi[$ se e só se as autocovariâncias são tais que

$$\sum_{h=-\infty}^{+\infty} [\gamma(h)]^2 < +\infty.$$

Prova. Relembremos que se $(H, \langle \cdot, \cdot \rangle)$ é um espaço de Hilbert com uma base ortonormada (e_h) , os elementos de H escrevem-se na forma

$$\sum \alpha_h e_h, \text{ com } \sum |\alpha_h|^2 < +\infty$$

ou seja,

$$H = \left\{ e = \sum_h \alpha_h e_h : \sum_h |\alpha_h|^2 < +\infty \right\}$$

onde $\alpha_h = \langle e, e_h \rangle$ e $\|e\|^2 = \sum_h |\alpha_h|^2$.

i) Consideremos, então, o espaço de Hilbert

$$H = L_{\mathbb{C}}^2([-\pi, \pi[, B_{[-\pi, \pi[}, \lambda_{[-\pi, \pi[}).$$

O produto interno neste espaço é, como sabemos, definido por

$$\forall f, g \in H, \langle f, g \rangle = \int_{-\pi}^{+\pi} f \bar{g} d\lambda$$

e uma base ortonormal de H é $\left\{ \frac{1}{\sqrt{2\pi}} \exp(i\omega h), h \in \mathbb{Z} \right\}$ (Koopmans, 1974, p.19).

Se f , densidade espectral, é de quadrado integrável, então $f \in L^2_{\mathbb{C}}([- \pi, \pi[, B_{[- \pi, \pi[, \lambda_{[- \pi, \pi[})$ pelo que

$$f = \sum_h \alpha_h \frac{1}{\sqrt{2\pi}} \exp(i\omega h)$$

com $\sum_h |\alpha_h|^2 < +\infty$ e

$$\begin{aligned} \alpha_h &= \left\langle f, \frac{1}{\sqrt{2\pi}} \exp(i\omega h) \right\rangle \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} f(\omega) \overline{\exp(i\omega h)} d\omega \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} f(\omega) \exp(-i\omega h) d\omega. \end{aligned}$$

Mas f é par, uma vez que μ é simétrica relativamente à origem. Então, da representação integral de $\gamma(h)$, obtém-se

$$\alpha_h = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} f(\omega) \exp(i\omega h) d\omega = \frac{1}{\sqrt{2\pi}} \gamma(h). \quad (2.2.1)$$

É pois evidente a partir desta forma que

$$\sum_h |\alpha_h|^2 < +\infty \iff \sum_h |\gamma(h)|^2 < +\infty.$$

ii) Inversamente, se $(\gamma(h), h \in \mathbb{Z})$ é tal que $\sum_h [\gamma(h)]^2 < +\infty$, podemos considerar

$$\alpha_h = \frac{1}{\sqrt{2\pi}} \gamma(h)$$

e escrever $f(\omega) = \frac{1}{2\pi} \sum_h \gamma(h) \exp(i\omega h)$. Esta soma existe no sentido de L^2 , *i.e.*, f é de quadrado integrável, porque $\sum_h [\gamma(h)]^2 < +\infty$.

Além disso, da igualdade (2.2.1),

$$\gamma(h) = \int_{-\pi}^{+\pi} f(\omega) \exp(i\omega h) d\omega$$

o que permite dizer que a medida espectral é $f d\lambda$. ■

Corolário. Sob as hipóteses precedentes,

$$\sum_h [\gamma(h)]^2 < +\infty \Rightarrow f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) \exp(i\omega h),$$

sendo a igualdade no sentido de $L^2_{\mathbb{C}}([- \pi, \pi[, B_{[- \pi, \pi[, \lambda_{[- \pi, \pi[})$.

Teorema. Se a função de autocovariância de um processo estacionário é absolutamente somável, isto é, se

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty$$

o processo admite uma densidade espectral sendo uma versão dada por

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) \exp(i\omega h)$$

onde a soma é no sentido da convergência uniforme sobre $[-\pi, \pi[$. Esta versão da densidade é contínua.

Prova. i) A existência resulta do facto de

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty \Rightarrow \sum_h [\gamma(h)]^2 < +\infty$$

(pois para h suficientemente grande, tem-se $|\gamma(h)| < 1$ e então $|\gamma(h)|^2 < |\gamma(h)|$).

ii) A convergência é uniforme uma vez que

$$|\gamma(h) \exp(i\omega h)| = |\gamma(h)|,$$

pelo que o resto da série de termo geral $\gamma(h) \exp(i\omega h)$ é majorado, em módulo, pelo resto da série convergente de termo geral $\gamma(h)$ independente de ω .

A função f é contínua porque é limite uniforme duma sucessão de funções contínuas, $f_n(\omega) = \frac{1}{2\pi} \sum_{h=0}^n \gamma(h) \exp(i\omega h)$. ■

A prova do penúltimo teorema mostra que, sob certas condições, o conhecimento de γ é equivalente ao da densidade espectral. Se a análise de uma série temporal é feita à custa da função de autocovariância (resp., à custa da densidade espectral) denomina-se "análise no domínio dos tempos" (resp., "análise no domínio das frequências").

Esta última designação está associada ao facto de todo o processo estacionário de segunda ordem $X_t, t \in \mathbb{Z}$, poder ser aproximado, no sentido de L_2 , por uma soma finita da forma

$$X_{n,t} = \sum_{j=1}^n \exp(it\lambda_j) \Delta Z(\lambda_j),$$

onde as diferenças $\Delta Z(\lambda_j) = Z(\lambda_j) - Z(\lambda_{j-1})$ são v.a. ortogonais em $L_2(\Omega, A, P)$. Assim, aproximamos X_t por uma soma de exponenciais complexas com frequências λ_j e amplitudes complexas $\Delta Z(\lambda_j)$, $\lambda_j \in [-\pi, \pi[$ (Ash e Gardner, 1975, Azencott e Dacunha-Castelle, 1984, Koopmans, 1974).

Observação. A paridade da função γ permite-nos obter diversas expressões para a densidade espectral. Assim,

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) \exp(i\omega h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) \exp(-i\omega h) \\ &= \frac{1}{2\pi} \gamma(0) + \frac{1}{2\pi} \sum_{h=1}^{+\infty} \gamma(h) [\exp(i\omega h) + \exp(-i\omega h)] \\ &= \frac{1}{2\pi} \gamma(0) + \frac{1}{\pi} \sum_{h=1}^{+\infty} \gamma(h) \cos(\omega h). \end{aligned}$$

Exemplos.

Analisemos, para alguns processos estocásticos, a existência de densidade espectral.

a) Seja $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$, $t \in \mathbb{Z}$, onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância $\sigma^2 > 0$ e $\theta_1, \dots, \theta_q \in \mathbb{R}$.

Este processo X é denominado *média móvel de ordem q* e designa-se por $MA(q)$.

O processo $(X_t, t \in \mathbb{Z})$ é estacionário pois $LX_t = X_{t-1}, \forall t$, $X \in H^\varepsilon$.

Tem-se

$$m_t = E(X_t) = 0, \forall t \in \mathbb{Z}$$

e

$$\gamma(h) = E(X_t X_{t+h})$$

$$= E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q})(\varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \dots + \theta_q \varepsilon_{t+h-q})].$$

Se $h = 0$,

$$\begin{aligned} \gamma(0) &= E \left[\varepsilon_t^2 + \sum_{i=1}^q \theta_i^2 \varepsilon_{t-i}^2 + \sum_{i=1}^q \sum_{\substack{j=1 \\ j \neq i}}^q \theta_i \theta_j \varepsilon_{t-i} \varepsilon_{t-j} + 2 \sum_{i=1}^q \varepsilon_t \varepsilon_{t-i} \right] \\ &= \sigma^2 + \sum_{i=1}^q \theta_i^2 \sigma^2 = \sigma^2 \left(1 + \sum_{i=1}^q \theta_i^2 \right). \end{aligned}$$

Analisemos agora o caso $h \neq 0$. Obtemos, com $\theta_0 = 1$,

$$\gamma(h) = E \left(\sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \varepsilon_{t-i} \varepsilon_{t+h-j} \right) = \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j E(\varepsilon_{t-i} \varepsilon_{t+h-j}).$$

- Se $h > 0$,

$$E(\varepsilon_{t-i} \varepsilon_{t+h-j}) = \begin{cases} 0 & , j \neq i+h \\ \sigma^2 & , j = i+h \end{cases}, i, j = 0, 1, \dots, q$$

e então

$$\gamma(h) = \sigma^2 \sum_{i=0}^q \theta_i \theta_{i+h} = \begin{cases} \sigma^2 \sum_{i=0}^{q-h} \theta_i \theta_{i+h}, & h \leq q \\ 0 & h > q. \end{cases}$$

porque para $i+h > q$, $\theta_{i+h} = 0$.

- Se $h < 0$,

$$E(\varepsilon_{t-i} \varepsilon_{t+h-j}) = E(\varepsilon_{t-i} \varepsilon_{t-(j+|h|)}) = \begin{cases} 0 & , j \neq i + |h| \\ \sigma^2 & , j = i + |h| \end{cases},$$

com $i, j = 0, 1, \dots, q$. Então

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

Em resumo,

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & -q \leq h \leq q, \theta_0 = 1. \\ 0, & \text{se não} \end{cases}$$

É claro que

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| = \sigma^2 \sum_{h=-q}^{+q} \left| \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|} \right| < +\infty.$$

Então, todo o processo $MA(q)$ admite densidade espectral e uma versão dessa densidade é da forma

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \gamma(0) + \frac{1}{\pi} \sum_{h=1}^{+\infty} \gamma(h) \cos(\omega h) \\ &= \frac{1}{2\pi} \left[\sigma^2 \left(1 + \sum_{i=1}^q \theta_i^2 \right) \right] + \frac{\sigma^2}{\pi} \sum_{h=1}^{+\infty} \cos(\omega h) \left(\sum_{i=0}^{q-h} \theta_i \theta_{i+h} \right). \end{aligned}$$

para todo $\omega \in [-\pi, \pi[$.

Se, por exemplo, $q = 1$, obtemos

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} [\sigma^2 (1 + \theta_1^2)] + \frac{\sigma^2}{\pi} \theta_1 \cos \omega \\ &= \frac{\sigma^2}{2\pi} (1 + \theta_1^2 + 2\theta_1 \cos \omega), \quad \omega \in [-\pi, \pi[. \end{aligned}$$

b) Se considerarmos agora o processo definido por $X_t = \sum_{i=-\infty}^{+\infty} \alpha_i \varepsilon_{t-i}$,

$t \in \mathbb{Z}$, onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância $\sigma^2 > 0$ e $(\alpha_i, i \in \mathbb{Z})$ é uma sucessão real absolutamente somável (*i.e.*, tal que $\sum_{i=-\infty}^{+\infty} |\alpha_i| < +\infty$) é fácil concluir a existência de densidade espectral e mostrar que uma sua versão é a função

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=-\infty}^{+\infty} \alpha_j \exp(i\omega j) \right|^2, \quad \forall \omega \in [-\pi, \pi[.$$

c) Seja agora $Y_t = \sum_{j=-\infty}^{+\infty} \beta_j X_{t-j}$, $t \in \mathbb{Z}$, onde $X = (X_t, t \in \mathbb{Z})$ é o processo definido na alínea anterior e $(\beta_i, i \in \mathbb{Z})$ é uma sucessão real absolutamente somável. A função de autocovariância do processo Y é dada por

$$\begin{aligned} \gamma_Y(h) &= \text{cov}(Y_t, Y_{t+h}) = E \left(\sum_{j=-\infty}^{+\infty} \beta_j X_{t-j} \sum_{k=-\infty}^{+\infty} \beta_k X_{t+h-k} \right) \\ &= \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \beta_j \beta_k \gamma_X(h+j-k). \end{aligned}$$

Esta igualdade é válida porque

$$\left\| \sum_{j=-\infty}^{+\infty} \beta_j X_{t-j} \sum_{k=-\infty}^{+\infty} \beta_k X_{t+h-k} \right\| \leq \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \|\beta_j \beta_k X_{t-j} X_{t+h-k}\|$$

e como

$$\begin{aligned} \|\beta_j \beta_k X_{t-j} X_{t+h-k}\| &\leq |\beta_j| |\beta_k| [E(X_{t-j}^2)]^{\frac{1}{2}} [E(X_{t-k}^2)]^{\frac{1}{2}} \\ &= |\beta_j| |\beta_k| \gamma_X^2(0) \end{aligned}$$

então

$$\sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \|\beta_j \beta_k X_{t-j} X_{t+h-k}\| \leq \gamma_X^2(0) \sum_{j=-\infty}^{+\infty} |\beta_j| \sum_{k=-\infty}^{+\infty} |\beta_k| < +\infty.$$

Por outro lado, $\sum_{h=-\infty}^{+\infty} |\gamma_Y(h)| < +\infty$. Com efeito,

$$\sum_{h=-\infty}^{+\infty} |\gamma_Y(h)| \leq \sum_{h=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} |\beta_j| |\beta_k| |\gamma_X(h+j-k)| < +\infty$$

tendo em conta que $\sum_{j=-\infty}^{+\infty} |\beta_j|$ é convergente e que

$$\begin{aligned} \sum_{h=-\infty}^{+\infty} |\gamma_X(h)| &= \sum_{h=-\infty}^{+\infty} \left| \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} a_j a_k \gamma_\varepsilon(h+j-k) \right| \\ &\leq \sigma^2 \sum_{h=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} |a_j| |a_{h+j}| < +\infty. \end{aligned}$$

O processo Y admite então uma densidade espectral, sendo uma versão dada por

$$\begin{aligned} f_Y(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma_Y(h) \exp(i\omega h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \beta_j \beta_k \gamma_X(h+j-k) \exp(i\omega h) \\ &= \sum_{j=-\infty}^{+\infty} \beta_j e^{-i\omega j} \sum_{k=-\infty}^{+\infty} \beta_k e^{i\omega k} \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma_X(h+j-k) e^{i\omega(h+j-k)} \end{aligned}$$

onde a troca dos somatórios foi possível uma vez que a série é absolutamente convergente (toda a série absolutamente convergente é comutativamente convergente).

Assim,

$$\forall \omega \in [-\pi, \pi[, \quad f_Y(\omega) = \left| \sum_{j=-\infty}^{+\infty} \beta_j \exp(i\omega j) \right|^2 f_X(\omega).$$

d) Seja $X = (X_t, t \in \mathbb{Z})$ um processo estacionário com função de autocovariância γ_X tal que $\sum_{h=-\infty}^{+\infty} |\gamma_X(h)| < +\infty$ e densidade espectral f_X .

Defina-se um novo processo estacionário Y por

$$Y_t = \sum_{j=-\infty}^{+\infty} \beta_j X_{t-j}, \quad t \in \mathbb{Z}, \quad \text{com} \quad \sum_{j=-\infty}^{+\infty} |\beta_j| < +\infty.$$

Prova-se que Y admite uma densidade espectral da forma

$$f_Y(\omega) = f_X(\omega) |B(\omega)|^2, \quad \forall \omega \in [-\pi, \pi[$$

onde $B(\omega) = \sum_{j=-\infty}^{+\infty} \beta_j \exp(i\omega j)$.

2.3 Autocorrelações parciais

Previsões lineares com base num passado de dimensão finita

Seja $X = (X_t, t \in \mathbb{Z})$ um processo real, univariado, centrado e estacionário, definido sobre (Ω, \mathcal{A}, P) .

Interessamo-nos pela melhor *previsão linear* afim de X_t conhecendo X_{t-1}, \dots, X_{t-k} , com k finito. Chamamos a k a "dimensão da memória".

Assim, interessamo-nos pela projecção ortogonal de X_t sobre o sub-espaço de L^2 das funções lineares afins de X_{t-1}, \dots, X_{t-k} , isto é, sobre

$$\text{Lin}(X_{t-1}, \dots, X_{t-k}) = \{b_0 + b_1 X_{t-1} + \dots + b_k X_{t-k}, b_i \in \mathbb{R}\}.$$

A resposta para este problema em $L^2(\Omega, \mathcal{A}, P)$ é a solução de

$$\min_{a_0, a_1, \dots, a_k} E |X_t - (a_0 + a_1 X_{t-1} + \dots + a_k X_{t-k})|^2,$$

que representaremos por $E_L(X_t/X_{t-1}, \dots, X_{t-k})$. Assim,

$$E_L(X_t/X_{t-1}, \dots, X_{t-k}) = a_1(k) X_{t-1} + \dots + a_k(k) X_{t-k},$$

onde $a_1(k), \dots, a_k(k)$ são chamados *coeficientes de regressão linear* sobre o passado de dimensão k .

Notemos que $a_0 = 0$ uma vez que o processo é centrado e

$$E[E_L(X_t/X_{t-1}, \dots, X_{t-k})] = E(X_t) = 0.$$

Observação. Quando k aumenta, o espaço sobre o qual se projecta aumenta e, portanto, obtém-se uma melhor previsão linear de X_t ; assim, o erro de previsão

$$E[X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k})]^2$$

diminui quando o nosso conhecimento do tempo passado aumenta, o que vai obviamente ao encontro da nossa intuição.

De facto,

$$\begin{aligned} e_{t,k} &= E[X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k})]^2 \\ &= E[(X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k-1})) + \\ &\quad + (E_L(X_t/X_{t-1}, \dots, X_{t-k-1}) - E_L(X_t/X_{t-1}, \dots, X_{t-k}))]^2 \\ &= e_{t,k+1} + E[E_L(X_t/X_{t-1}, \dots, X_{t-k-1}) - E_L(X_t/X_{t-1}, \dots, X_{t-k})]^2 + \\ &\quad + 2E[(X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k-1})) \\ &\quad (E_L(X_t/X_{t-1}, \dots, X_{t-k-1}) - (E_L(X_t/X_{t-1}, \dots, X_{t-k})))] \\ &= e_{t,k+1} + E[E_L(X_t/X_{t-1}, \dots, X_{t-k-1}) - E_L(X_t/X_{t-1}, \dots, X_{t-k})]^2, \end{aligned}$$

pois

$$E_L(X_t/X_{t-1}, \dots, X_{t-k-1}) - E_L(X_t/X_{t-1}, \dots, X_{t-k})$$

pertence ao espaço $Lin(X_{t-1}, \dots, X_{t-k-1})$ e

$$X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k-1})$$

é ortogonal a este espaço.

Determinação dos coeficientes de regressão linear

Vamos determinar os coeficientes de regressão linear supondo que o processo estocástico X é tal que todas as matrizes de autocorrelação $R(m)$, $m \in \mathbb{N}$, são invertíveis.

Tem-se, $\forall Z \in \text{Lin}(X_{t-1}, \dots, X_{t-k})$,

$$E\{[X_t - (a_1(k)X_{t-1} + \dots + a_k(k)X_{t-k})]Z\} = 0.$$

Em particular,

$$\forall i \in \{1, \dots, k\}, E\{[X_t - a_1(k)X_{t-1} - \dots - a_k(k)X_{t-k}]X_{t-i}\} = 0$$

ou seja,

$$\forall i \in \{1, \dots, k\}, E(X_t X_{t-i}) = \sum_{j=1}^k a_j(k) E(X_{t-j} X_{t-i})$$

isto é,

$$\forall i \in \{1, \dots, k\}, \text{Cov}(X_t, X_{t-i}) = \sum_{j=1}^k a_j(k) \text{Cov}(X_{t-j}, X_{t-i}).$$

Temos, pois,

$$\begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \dots \\ \gamma(k) \end{bmatrix} = V(\vec{X}) \begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix}$$

onde $V(\vec{X})$ é a matriz de variâncias-covariâncias de $\vec{X} = [X_{t-1} \dots X_{t-k}]'$. Então

$$\begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix} = [V(\vec{X})]^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \dots \\ \gamma(k) \end{bmatrix}$$

$$\begin{aligned}
&= \begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \dots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \dots & \dots & \gamma(k-2) \\ \dots & \dots & \gamma(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \gamma(k-1) & \gamma(k-2) & \dots & \dots & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \dots \\ \dots \\ \gamma(k) \end{bmatrix} \\
&= \begin{bmatrix} 1 & \rho(1) & \dots & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \dots & \rho(k-2) \\ \dots & \dots & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \rho(k-1) & \rho(k-2) & \dots & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \dots \\ \rho(k) \end{bmatrix} \\
&= [R(k)]^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(k) \end{bmatrix}. \blacksquare
\end{aligned}$$

Do ponto de vista numérico, uma inversão pode ser muito longa. Daí a procura de um algoritmo de cálculo recursivo dos $a_j(k)$, que nos vai conduzir ao denominado algoritmo de Durbin. Com tal objectivo, comecemos por estabelecer a propriedade seguinte:

Propriedade. Os coeficientes de regressão linear sobre o passado de dimensão k coincidem com os coeficientes de regressão linear sobre o futuro de dimensão k , isto é,

$$E_L(X_t/X_{t+1}, \dots, X_{t+k}) = a_1(k)X_{t+1} + \dots + a_k(k)X_{t+k}.$$

Prova. De modo análogo ao anterior concluímos que o vector dos coeficientes de regressão é dado por

$$[R(k)]^{-1} \begin{bmatrix} \rho(-1) \\ \rho(-2) \\ \dots \\ \rho(-k) \end{bmatrix} = \begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix}.$$

A paridade de ρ permite concluir. ■

Note-se que, embora se tenha a igualdade dos coeficientes, é óbvio que os valores das previsões, isto é,

$$E_L(X_t/X_{t+1}, \dots, X_{t+k}) \quad \text{e} \quad E_L(X_t/X_{t-1}, \dots, X_{t-k})$$

são em geral diferentes.

Vamos usar esta propriedade para estudar a variação dos coeficientes de regressão com a dimensão da memória, o que nos vai permitir deduzir o algoritmo de Durbin.

Algoritmo de Durbin

Projectemos $E_L(X_t/X_{t-1}, \dots, X_{t-k})$ sobre o passado linear de dimensão $k-1$, isto é, sobre $Lin(X_{t-1}, \dots, X_{t-k+1})$.

$$\begin{aligned} E_L[E_L(X_t/X_{t-1}, \dots, X_{t-k})/X_{t-1}, \dots, X_{t-k+1}] \\ = a_1(k) E_L(X_{t-1}/X_{t-1}, \dots, X_{t-k+1}) + \\ + \dots + a_{k-1}(k) E_L(X_{t-k+1}/X_{t-1}, \dots, X_{t-k+1}) + \\ + a_k(k) E_L(X_{t-k}/X_{t-1}, \dots, X_{t-k+1}) \end{aligned}$$

ou seja, uma vez que $Lin(X_{t-1}, \dots, X_{t-k+1}) \subset Lin(X_{t-1}, \dots, X_{t-k})$,

$$\begin{aligned} E_L(X_t/X_{t-1}, \dots, X_{t-k+1}) = a_1(k) X_{t-1} + \dots + a_{k-1}(k) X_{t-k+1} + \\ + a_k(k) E_L(X_{t-k}/X_{t-1}, \dots, X_{t-k+1}). \end{aligned}$$

Na última parcela projectamos sobre o futuro de dimensão $k-1$ (relativamente a $t-k$). Da propriedade anterior decorre

$$\begin{aligned} a_1(k-1) X_{t-1} + \dots + a_{k-1}(k-1) X_{t-k+1} \\ = a_1(k) X_{t-1} + \dots + a_{k-1}(k) X_{t-k+1} + \\ + a_k(k) [a_1(k-1) X_{t-k+1} + \dots + a_{k-1}(k-1) X_{t-1}] \end{aligned}$$

ou seja,

$$\begin{cases} a_1(k-1) = a_1(k) + a_k(k) a_{k-1}(k-1) \\ \dots \\ a_{k-1}(k-1) = a_{k-1}(k) + a_k(k) a_1(k-1) \end{cases} .$$

De forma resumida

$$a_j(k) = a_j(k-1) - a_k(k) a_{k-j}(k-1), j = 1, \dots, k-1. \quad (2.3.1)$$

Podemos, pois, calcular $a_j(k)$, $j = 1, \dots, k-1$, em função de $a_j(k-1)$, $j = 1, \dots, k-1$, e de $a_k(k)$.

Falta-nos agora o cálculo recursivo de $a_k(k)$.

Retomemos o sistema

$$R(k) \begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(k) \end{bmatrix}.$$

A última linha deste sistema é

$$\sum_{j=1}^{k-1} \rho(k-j) a_j(k) + a_k(k) = \rho(k).$$

Substituindo os $a_j(k)$, $j = 1, \dots, k-1$ obtidos anteriormente obtemos

$$\sum_{j=1}^{k-1} \rho(k-j) a_j(k-1) - a_k(k) \left[\sum_{j=1}^{k-1} \rho(k-j) a_{k-j}(k-1) - 1 \right] = \rho(k)$$

de onde retiramos

$$a_k(k) = \frac{\rho(k) - \sum_{j=1}^{k-1} \rho(k-j) a_j(k-1)}{1 - \sum_{j=1}^{k-1} \rho(k-j) a_{k-j}(k-1)}. \quad (2.3.2)$$

Obtivemos assim o *algoritmo de Durbin* que podemos resumir da seguinte forma:

- Para o cálculo de $a_j(k)$, $j = 1, \dots, k-1$, na etapa de ordem k ($k \geq 2$),

i) dispomos de $a_1(k-1)$, $a_2(k-1)$, ..., $a_{k-1}(k-1)$

ii) calculamos $a_k(k)$ utilizando (2.3.2)

iii) calculamos $a_j(k)$, $j = 1, \dots, k-1$, utilizando (2.3.1).

- O procedimento recursivo é iniciado com o valor $a_1(1) = \rho(1)$.

Exemplo. Seja $X = (X_t, t \in \mathbb{Z})$ um processo média móvel de ordem um, $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$, onde ε é um ruído branco de variância $\sigma^2 > 0$ e θ um número real. Tem-se

$$\rho(h) = \begin{cases} 1, & h = 0 \\ -\frac{\theta}{1+\theta^2}, & h = -1, +1 \\ 0, & |h| \geq 2 \end{cases}.$$

Então,

$$a_1(1) = \rho(1) = -\frac{\theta}{1+\theta^2} = r(1)$$

e como para $k \geq 2$ se tem

$$\begin{cases} a_k(k) = \frac{-\rho(1)a_{k-1}(k-1)}{1-\rho(1)a_1(k-1)} \\ a_j(k) = a_j(k-1) - a_k(k)a_{k-j}(k-1), \quad j = 1, \dots, k-1, \end{cases}$$

obtemos

$$r(2) = a_2(2) = -\frac{\rho(1)a_1(1)}{1-\rho(1)a_1(1)} = -\frac{[\rho(1)]^2}{1-[\rho(1)]^2},$$

$$\text{e } a_1(2) = a_1(1) - a_2(2)a_1(1) = \rho(1) + \frac{[\rho(1)]^2}{1-[\rho(1)]^2}\rho(1) = \frac{\rho(1)}{1-[\rho(1)]^2}.$$

Continuando

$$r(3) = a_3(3) = \frac{-\rho(1)a_2(2)}{1-\rho(1)a_1(2)} = \frac{\frac{[\rho(1)]^3}{1-[\rho(1)]^2}}{1-\frac{[\rho(1)]^2}{1-[\rho(1)]^2}} = \frac{[\rho(1)]^3}{1-2[\rho(1)]^2}$$

$$a_1(3) = \dots, a_2(3) = \dots,$$

$$r(4) = a_4(4) = \dots$$

e assim sucessivamente.

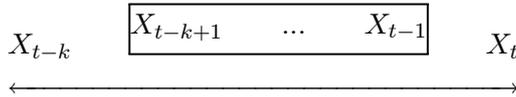
Autocorrelações parciais

O resumo de segunda ordem que vamos agora introduzir permite-nos medir a ligação (directa) que existe entre duas margens não consecutivas do processo, isto é, não vai ter em conta a correlação que passa pelas margens intermédias.

Definição. Sendo $X = (X_t, t \in \mathbb{Z})$ estacionário chamamos *autocorrelação parcial de ordem k* à correlação entre X_t e X_{t-k} uma vez retiradas as ligações lineares que passam por intermédio das margens $X_{t-1}, \dots, X_{t-k+1}$.

Assim, enquanto que a autocorrelação de ordem k mede a dependência linear entre X_{t-k} e X_t , a autocorrelação parcial de ordem k é a correlação existente entre os resíduos de X_{t-k} e X_t depois de efectuar a regressão linear de cada uma destas variáveis sobre $X_{t-1}, \dots, X_{t-k+1}$.

Pretendemos medir a influência directa entre duas quaisquer margens. Esquemáticamente



Note-se que

– a "explicação" de X_t em termos de $(X_{t-k+1}, \dots, X_{t-1})$, como combinação linear afim, é dada por $E_L(X_t/X_{t-1}, \dots, X_{t-k+1})$

– a "explicação" de X_{t-k} em termos de $(X_{t-k+1}, \dots, X_{t-1})$, como combinação linear afim, é dada por $E_L(X_{t-k}/X_{t-1}, \dots, X_{t-k+1})$.

Assim, retirando as ligações que passam por intermédio de $X_{t-k+1}, \dots, X_{t-1}$, obtém-se:

$$\begin{aligned}
 Z_1 &= X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k+1}) \\
 Z_2 &= X_{t-k} - E_L(X_{t-k}/X_{t-1}, \dots, X_{t-k+1})
 \end{aligned}$$

Então, a autocorrelação parcial de ordem k é, por definição,

$$r(k) = \text{corr}(Z_1, Z_2) = \frac{\text{cov}(Z_1, Z_2)}{\sqrt{V(Z_1)}\sqrt{V(Z_2)}}.$$

Como veremos, devido à estacionaridade do processo, este coeficiente depende apenas de k .

Propriedade. $r(k) = a_k(k)$.

Prova. Partamos de

$$E_L(X_t/X_{t-1}, \dots, X_{t-k}) = a_1(k)X_{t-1} + \dots + a_k(k)X_{t-k}$$

e projectemos este elemento sobre o espaço linear gerado por $X_{t-1}, \dots, X_{t-k+1}$ (memória de dimensão $k-1$). Vimos já que

$$E_L(X_t/X_{t-1}, \dots, X_{t-k+1}) = a_1(k)X_{t-1} + \dots + a_{k-1}(k)X_{t-k+1} + a_k(k)E_L(X_{t-k}/X_{t-1}, \dots, X_{t-k+1}).$$

Por subtracção destas igualdades obtemos

$$E_L(X_t/X_{t-1}, \dots, X_{t-k}) - E_L(X_t/X_{t-1}, \dots, X_{t-k+1}) = a_k(k)Z_2.$$

Seja $Z = E_L(X_t/X_{t-1}, \dots, X_{t-k}) - E_L(X_t/X_{t-1}, \dots, X_{t-k+1})$.

Temos pois

$$Z = a_k(k)Z_2,$$

o que implica

$$a_k(k) = \frac{\text{Cov}(Z, Z_2)}{V(Z_2)}.$$

Por outro lado,

$$\begin{aligned} Z_1 &= X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k+1}) \\ &= E_L(X_t/X_{t-1}, \dots, X_{t-k}) - E_L(X_t/X_{t-1}, \dots, X_{t-k+1}) + \\ &\quad + X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k}) \\ &= Z + Z^* \end{aligned}$$

com $Z^* = X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k})$.

Então

$$a_k(k) = \frac{\text{Cov}(Z_1 - Z^*, Z_2)}{V(Z_2)}.$$

Mas Z^* é ortogonal ao subespaço linear gerado por X_{t-1}, \dots, X_{t-k} e Z_2 pertence a esse espaço.

Então

$$\text{Cov}(Z^*, Z_2) = 0.$$

Assim

$$a_k(k) = \frac{\text{Cov}(Z_1, Z_2)}{V(Z_2)} = \frac{\text{Cov}(Z_1, Z_2)}{\sqrt{V(Z_1)}\sqrt{V(Z_2)}} = r(k),$$

uma vez que $V(Z_2) = V(Z_1)$, já que o processo é estacionário e a regressão sobre o futuro é, como vimos, equivalente à regressão sobre o passado de igual dimensão. ■

Propriedade. É equivalente conhecer a função de autocorrelação $(\rho(h), h \in \mathbb{N})$ ou a função de autocorrelação parcial $(r(h), h \in \mathbb{N})$.

Prova. Por recorrência obtém-se facilmente tal equivalência.

i) Suponhamos conhecida a função $(\rho(h), h \in \mathbb{N})$.

Se $h = 1$,

$$r(1) = a_1(1) = \frac{\text{Cov}(X_t, X_{t-1})}{V(X_{t-1})} = \frac{\gamma(1)}{\gamma(0)} = \rho(1).$$

Supondo conhecidos $r(1), \dots, r(h)$, a prova para $h + 1$ resulta da utilização do algoritmo de Durbin uma vez que

$$\begin{bmatrix} a_1(h+1) \\ a_2(h+1) \\ \dots \\ a_{h+1}(h+1) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(h) \\ \rho(1) & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(h+1) \end{bmatrix}.$$

ii) Inversamente vejamos que, para cada $h \in \mathbb{N}$, $\rho(h)$ é função de $r(1), \dots, r(h)$.

Vimos já que $\rho(1) = r(1)$.

Suponhamos a propriedade verificada para $h - 1$ e provemo-la para h . Consideremos o sistema

$$R(h) \begin{bmatrix} a_1(h) \\ a_2(h) \\ \dots \\ a_h(h) \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(h) \end{bmatrix}$$

isto é,

$$\begin{bmatrix} 1 & \rho(1) & \dots & \rho(h-1) \\ \rho(1) & 1 & \dots & \rho(h-2) \\ \dots & \dots & \dots & \dots \\ \rho(h-1) & \rho(h-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1(h) \\ a_2(h) \\ \dots \\ a_h(h) \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(h) \end{bmatrix}$$

que podemos escrever na forma

$$\begin{bmatrix} R(h-1) & \vec{\rho}_{(h-1),1} \\ (\vec{\rho}_{(h-1),1})' & 1 \end{bmatrix} \begin{bmatrix} \vec{a}_{1,h-1}(h) \\ a_h(h) \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \dots \\ \rho(h) \end{bmatrix}$$

ou de modo equivalente

$$\begin{cases} R(h-1) \vec{a}_{1,h-1}(h) + \vec{\rho}_{(h-1),1} a_h(h) = \begin{bmatrix} \rho(1) \\ \dots \\ \rho(h-1) \end{bmatrix} \\ (\vec{\rho}_{(h-1),1})' \vec{a}_{1,h-1}(h) + a_h(h) = \rho(h) \end{cases} .$$

Usando o primeiro sistema e a hipótese de recorrência vemos que $\vec{a}_{1,h-1}(h) = (a_1(h), \dots, a_{h-1}(h))'$ é função de $r(1), \dots, r(h-1)$ (pela indução) e de $r(h)$ (pois $a_h(h) = r(h)$). Então, $\rho(h)$ é, pelo segundo sistema, função de $r(1), \dots, r(h)$. ■

Observações.

1. Para cada $h \in \mathbb{N}$, é equivalente conhecer $(\rho(1), \dots, \rho(h))$ ou $(r(1), \dots, r(h))$.
2. Tem-se $|r(h)| \leq 1$, porque são autocorrelações.
3. As autocorrelações parciais desempenham um papel análogo ao das autocorrelações. A sua representação gráfica toma o nome de *correlograma parcial*.

2.4 Anexo - Esperança condicional

1. Introdução

Seja (X, Y) um par de v.a. definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) e no qual apenas Y é observado. Se X e Y forem correlacionadas, é natural que o conhecimento do valor de Y nos permita obter alguma inferência sobre os valores da variável não observada X . Quando procuramos saber que informação Y traz a X deparamo-nos com o chamado problema de filtragem, isto é, o problema da estimação duma v.a. não observada X a partir duma v.a. observada Y . Para o ilustrar consideremos a seguinte situação:

Um economista observa a evolução do preço de um certo produto no intervalo de tempo $[t_1, t_2]$ com o objectivo de prever o preço deste mesmo produto no instante t_3 ($t_3 > t_2$). Para modelar tal fenómeno aleatório consideremos uma família de v.a.r. $(X_t, t \geq t_1)$ onde X_t representa o preço do produto no instante t . Conhecidas as realizações das variáveis $X_t, t_1 \leq t \leq t_2$, pretende-se então prever da melhor maneira possível o valor de X_{t_3} .

O problema da interpolação é de natureza análoga. De facto, consiste em determinar da melhor forma possível X_{t_0} a partir de v.a. X_t , com $t \in [t_1, t_2] \cup [t_3, t_4]$ com $t_2 < t_0 < t_3$.

A previsão e a interpolação são dois casos particulares do problema geral da filtragem.

Considerem-se então duas v.a. definidas sobre (Ω, \mathcal{A}, P)

$$\begin{aligned} Y &: (\Omega, \mathcal{A}, P) \longrightarrow (E, \mathcal{E}) \\ X &: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}), \end{aligned}$$

onde (E, \mathcal{E}) é um espaço mensurável qualquer.

Conhecendo $Y(\omega) = y \in E$ queremos fazer um prognóstico sobre $X(\omega)$ por um processo válido para todo o ω .

Põe-se assim o problema de procurar uma função

$$f : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

de tal modo que a função $f(Y)$ aproxime convenientemente X . A função $f(Y)$ é denominada estimador de X . Diz-se que um

estimador $f^* = f^*(Y)$ é melhor no sentido do erro quadrático médio se

$$E[X - f^*(Y)]^2 = \inf_f E[X - f(Y)]^2,$$

ou seja, escolhemos a distância em média quadrática para julgar a qualidade da aproximação. Neste caso, impõe-se supor X de quadrado integrável.

Ilustremos a resolução do problema da procura do melhor estimador, no sentido do erro quadrático médio, na classe dos estimadores lineares, isto é, da forma $f(Y) = a_1 + a_2Y$, $a_1, a_2 \in \mathbb{R}$, onde Y é uma v.a. real.

Considere-se a função

$$g(a_1, a_2) = E[X - (a_1 + a_2Y)]^2.$$

Os pontos a_1^* e a_2^* que anulam as derivadas parciais de g em ordem a a_1 e a a_2 são dados por

$$\begin{cases} a_1^* = E(X) - a_2^*E(Y) \\ a_2^* = \frac{Cov(X, Y)}{V(Y)} \end{cases}.$$

A análise das condições de segunda ordem permite concluir que o melhor estimador linear de X no sentido do erro quadrático médio é

$$f^*(Y) = a_1^* + a_2^*Y = E(X) + \frac{Cov(X, Y)}{V(Y)}(Y - E(Y)).$$

Ao valor $E[X - f^*(Y)]^2$ chamamos erro quadrático médio associado ao estimador f^* .

Tem-se, neste caso,

$$E[X - f^*(Y)]^2 = V(X) - \frac{[Cov(X, Y)]^2}{V(Y)} = V(X) [1 - \rho_{X, Y}^2] = e^*.$$

Podemos então concluir que:

- quanto maior, em valor absoluto, for o valor do coeficiente de regressão $\rho_{X, Y}$, menor é o erro e^* ; em particular, se $|\rho| = 1$ tem-se $e^* = 0$.

- se X e Y forem não correlacionadas então $f^*(Y) = E(X)$, isto é, quando há ausência de correlação entre X e Y a melhor estimativa linear de X em termos de Y é $E(X)$.

É claro que esta abordagem só responde parcialmente ao problema inicial pois limitamos a escolha de f à classe dos estimadores lineares. Vejamos como abordar o problema de forma mais geral.

Sabemos que se (Ω, \mathcal{A}, P) é um espaço de probabilidade e B um acontecimento de probabilidade positiva, a probabilidade condicionada por B é definida por

$$\forall A \in \mathcal{A}, \quad P(A / B) = \frac{P(A \cap B)}{P(B)}.$$

Em particular, se X é uma variável aleatória definida sobre (Ω, \mathcal{A}, P) com valores em $(\mathbb{R}, \mathcal{B})$, a lei de X condicionada por $B \in \mathcal{A}$, com $P(B) > 0$, é definida por

$$\forall D \in \mathcal{B}, \quad P_{X,B}(D) = P(X \in D / B) = \frac{P(\{X \in D\} \cap B)}{P(B)}.$$

Se existe, a esperança matemática de X condicionada por B pode ser então definida por

$$\begin{aligned} E(X/B) &= \int x \, dP_{X,B}(x) \\ &= \frac{1}{P(B)} \int_B x \, dP_X(x) = \frac{1}{P(B)} E(X \mathbb{I}_B). \end{aligned}$$

No entanto, em geral, não é nem um só acontecimento nem um acontecimento de probabilidade positiva que influencia a média de X . Suponhamos para fixar ideias que X e Y são v.a.r. com densidade representando, respectivamente, o peso e a altura dos indivíduos de uma população. Se se observa a altura de um indivíduo da população, um acontecimento do tipo $\{Y = y\}$ realiza-se; pretende-se então, por exemplo, dar uma estimacção do peso do indivíduo, isto é, dar um sentido a $E(X/Y = y)$. Mas $\{Y = y\}$ tem probabilidade nula; por outro lado será desejável exprimir

$E(X/Y = y)$ como a realização de uma variável aleatória não dependendo só de $Y = y$ mas dos acontecimentos da tribo gerada por Y , $\sigma(Y)$ ($\sigma(Y) \subset \mathcal{A}$).

À função

$$f(y) = E(X/Y = y), \quad \forall y \in E,$$

chamamos *regressão de X em Y* .

Então o problema inicial pode agora pôr-se do seguinte modo: sendo X uma v.a.r. definida sobre (Ω, \mathcal{A}, P) e \mathcal{A}_0 uma tribo contida em \mathcal{A} , pretendemos definir, por um processo válido, uma v.a. T sobre $(\Omega, \mathcal{A}_0, P)$ que aproxime X para todo $\omega \in \Omega$.

Começamos por considerar X de quadrado integrável.

2. Aproximação em $L^2(\Omega, \mathcal{A}, P)$

Seja \mathcal{A}_0 uma sub-tribo de \mathcal{A} e defina-se a aplicação

$$i : L^2(\Omega, \mathcal{A}_0, P) \longrightarrow L^2(\Omega, \mathcal{A}, P)$$

que à classe de X em $L^2(\Omega, \mathcal{A}_0, P)$ associa a classe de X em $L^2(\Omega, \mathcal{A}, P)$.

Sendo esta aplicação injectiva, podemos identificar $L^2(\Omega, \mathcal{A}_0, P)$ com a sua imagem por i em $L^2(\Omega, \mathcal{A}, P)$. Tem-se então

$$L^2(\Omega, \mathcal{A}_0, P) \subset L^2(\Omega, \mathcal{A}, P).$$

Além disso, sendo $L^2(\Omega, \mathcal{A}_0, P)$ completo, é um subespaço de Hilbert de $L^2(\Omega, \mathcal{A}, P)$ ⁽²⁾. Assim, em particular, o espaço $L^2(\Omega, X^{-1}(\mathcal{B}), P)$, com \mathcal{B} tribo de Borel de \mathbb{R} e X v.a.r. sobre (Ω, \mathcal{A}, P) , é um subespaço de Hilbert de $L^2(\Omega, \mathcal{A}, P)$.

Lema. Seja X uma v.a.r. definida sobre (Ω, \mathcal{A}, P) e Y uma v.a. definida sobre (Ω, \mathcal{A}, P) com valores em (E, \mathcal{E}) . Para que X seja $\sigma(Y)$ -mensurável (*i.e.*, $X^{-1}(\mathcal{B}) \subset \sigma(Y)$) é necessário e suficiente que exista

² Todo o subespaço vectorial H_0 dum espaço de Hilbert H é ainda um espaço de Hilbert - para o produto escalar de H - se e só se ele é completo (Rudin, Real and complex analysis, McGraw-Hill International editions, 1987).

$$f : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B}) \text{ mensurável}$$

tal que $X = f(Y) = f \circ Y$.

Prova.

1. Se $X = f \circ Y$ então X é $\sigma(Y)$ -mensurável porque é a composta de duas funções mensuráveis. De facto,

$$X^{-1}(\mathcal{B}) = Y^{-1}(f^{-1}(\mathcal{B})) \subset Y^{-1}(\mathcal{E}) = \sigma(Y).$$

2. Suponhamos agora que X é $\sigma(Y)$ -mensurável.

i) Se X é uma v.a.r. em patamar, tem-se

$$X = \sum_{i=1}^n y_i \mathbb{I}_{A_i}$$

onde $\{A_i, 1 \leq i \leq n\}$ é uma partição de Ω $\sigma(Y)$ -mensurável e $y_i \in \mathbb{R}$ para todo i .

Como, para todo o i , $A_i \in \sigma(Y)$, então

$$\exists B_i \in \mathcal{E} : A_i = Y^{-1}(B_i).$$

Consideremos então $f = \sum_{i=1}^n y_i \mathbb{I}_{B_i}$.

A função f é mensurável e tem-se

$$\forall \omega \in \Omega, \quad f \circ Y(\omega) = \sum_{i=1}^n y_i \mathbb{I}_{B_i}(Y(\omega)) = \sum_{i=1}^n y_i \mathbb{I}_{A_i}(\omega) = X(\omega).$$

ii) Se X é uma v.a.r. positiva, existe uma sucessão de v.a.r. em patamar, $(X_n)_{n \in \mathbb{N}}$, convergindo para X ⁽³⁾.

Além disso sendo X_n $\sigma(Y)$ -mensurável, decorre de i) que existe $f_n : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, \mathcal{B})$ tal que

$$f_n \circ Y = X_n, \text{ para cada } n.$$

Seja D o domínio de convergência da sucessão $(f_n)_{n \in \mathbb{N}}$, isto é,

³Foata, D., Fuchs, A., Calcul des probabilités, 2e édition, Dunod, 1998.

$$D = \{y \in E : (f_n(y)) \text{ é convergente}\}.$$

Este conjunto pertence a \mathcal{E} uma vez que

$$D = \{\liminf f_n = \limsup f_n\} \cap \{|\limsup f_n| < +\infty\}.$$

Além disso $Y(\Omega) \subset D$. De facto,

$$y \in Y(\Omega) \implies f_n(y) = X_n(Y^{-1}(y)) = X_n(\omega)$$

e então

$$\lim_{n \rightarrow +\infty} f_n(y) = \lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega) \implies y \in D.$$

Consideremos então

$$f = \lim_{n \rightarrow +\infty} f_n \mathbb{I}_D.$$

Esta função é \mathcal{E} -mensurável. Além disso,

$$\forall \omega \in \Omega, \lim_{n \rightarrow +\infty} f_n \circ Y(\omega) = \lim_{n \rightarrow +\infty} [f_n(Y(\omega))] = f \circ Y(\omega) = X(\omega).$$

Assim, $f \circ Y = X$.

iii) Se X é uma v.a.r. qualquer, decompomos X nas suas partes positiva e negativa, $X = X^+ - X^-$ ⁽⁴⁾, as quais são v.a.r. positivas, e aplicamos o resultado do ponto anterior. ■

Teorema. Sejam Y e X duas v.a. definidas sobre o mesmo espaço (Ω, \mathcal{A}, P) :

$$\begin{aligned} Y & : (\Omega, \mathcal{A}, P) \longrightarrow (E, \mathcal{E}) \\ X & : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}). \end{aligned}$$

Seja X de quadrado integrável. Então existe em $L^2(\Omega, \sigma(Y), P)$ uma v.a.r. T e uma só (P -q.c., *i.e.*, mod. P) verificando as condições equivalentes

⁴ $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$,

i) $\forall g : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, B)$ limitada, $E(g \circ Y.X) = E(g \circ Y.T)$.

ii) $\forall g : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, B)$ limitada, $E(X - T)^2 \leq E(X - g \circ Y)^2$.

Prova. Mostremos que para toda a v.a.r. T definida sobre $L^2(\Omega, \sigma(Y), P)$ se tem **i)** \iff **iii)** e **ii)** \iff **iv)**, com

iii) $\forall Z \in L^2(\Omega, \sigma(Y), P)$, $E(ZX) = E(ZT)$

iv) $\forall Z \in L^2(\Omega, \sigma(Y), P)$, $E(X - T)^2 \leq E(X - Z)^2$.

i) \Rightarrow **iii)**

Seja $Z \in L^2(\Omega, \sigma(Y), P)$. Do lema anterior existe $g : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ tal que $Z = g \circ Y$.

Como g^+ e g^- são v.a.r. positivas, existem sucessões de funções mensuráveis em patamar, positivas definidas sobre (E, \mathcal{E}) tais que $(\varphi_n)_n \nearrow g^+$ e $(\psi_n)_n \nearrow g^-$.⁽⁵⁾

A sucessão $(g_n)_{n \in \mathbb{N}}$ definida por

$$\forall n \geq 1, \quad g_n = \varphi_n - \psi_n$$

tem as propriedades seguintes

- $g_n \xrightarrow{n \rightarrow +\infty} g$
- $|g_n| \leq \varphi_n + \psi_n (\leq g^+ + g^- \leq |g|)$
- $\forall n \geq 1, E(g_n \circ Y.X) = E(g_n \circ Y.T)$ de i), pois g_n é limitada
- $g_n \circ Y \xrightarrow{n \rightarrow +\infty} g \circ Y$ o que implica

$$g_n \circ Y.X \xrightarrow{n \rightarrow +\infty} g \circ Y.X \quad \text{e} \quad g_n \circ Y.T \xrightarrow{n \rightarrow +\infty} g \circ Y.T.$$

Além disso, estas duas últimas convergências são dominadas pois

$$|g_n \circ Y.X| = |g_n \circ Y| |X| \leq |g(Y)| |X| = |ZX| \leq \frac{1}{2} (Z^2 + X^2),$$

que é integrável. Do mesmo modo

$$|g_n \circ Y.T| \leq \frac{1}{2} (Z^2 + T^2).$$

Do teorema da convergência dominada⁽⁶⁾ tem-se

⁵Foata, D., Fuchs, A., Calcul des probabilités, 2e édition, Dunod, 1998.

⁶ Foata, D., Fuchs, A., Calcul des probabilités, 2e édition, Dunod, 1998.

$$E(g_n \circ Y.X) = \int (g_n \circ Y.X) dP \xrightarrow{n \rightarrow +\infty} \int ZX dP = E(ZX)$$

e

$$E(g_n \circ Y.T) = \int (g_n \circ Y.T) dP \xrightarrow{n \rightarrow +\infty} \int g \circ Y.T dP = E(ZT).$$

De i) e da unicidade do limite vem $E(ZX) = E(ZT)$.

iii) \implies i)

Seja $g : (E, \mathcal{E}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ limitada. Então

$$\int (g \circ Y)^2 dP \leq M^2 < +\infty,$$

onde M é um majorante de g . Então $g(Y) \in L^2(\Omega, \sigma(Y), P)$.
Aplicando iii) a $g(Y)$ deduzimos i).

ii) \implies iv)

Seja $Z \in L^2(\Omega, \sigma(Y), P)$; do lema anterior $Z = g \circ Y$.

Decompondo $g = g^+ - g^-$ sabemos que existem sucessões de funções mensuráveis em patamar, positivas definidas sobre (E, \mathcal{E}) tais que $(\varphi_n)_n \nearrow g^+$ e $(\psi_n)_n \nearrow g^-$.

Seja $(g_n)_{n \in \mathbb{N}}$ definida por

$$\forall n \geq 1, \quad g_n = \varphi_n - \psi_n.$$

Por ii) tem-se, $\forall n \in \mathbb{N}$,

$$E(X - T)^2 \leq E(X - g_n \circ Y)^2 \text{ pois os } g_n \text{ são limitados.} \quad (2.1)$$

Além disso,

- $g_n \circ Y \xrightarrow{n \rightarrow +\infty} Z$
- $|X - g_n \circ Y|^2 \leq (|X| + |g_n(Y)|)^2$
 $\leq (|X| + |g(Y)|)^2 \leq (|X| + |Z|)^2 \leq 2(|X|^2 + |Z|^2)$

que é uma função integrável.

O teorema da convergência dominada permite então escrever

$$E(X - g_n \circ Y)^2 \xrightarrow{n \rightarrow +\infty} E(X - Z)^2.$$

Logo da relação (2.1) e da definição de limite

$$E(X - T)^2 \leq E(X - Z)^2.$$

iv) \implies ii)

Por hipótese, $\forall Z \in L^2(\Omega, \sigma(Y), P)$, $E(X - T)^2 \leq E(X - Z)^2$.

Em particular, se g é limitada então $g(Y) \in L^2(\Omega, \sigma(Y), P)$ e tem-se

$$E(X - T)^2 \leq E(X - g(Y))^2.$$

Pelo teorema da projecção ⁽⁷⁾ existe em $L^2(\Omega, \sigma(Y), P)$ um elemento T e um só (mod. P) verificando as condições equivalentes *iii*) e *iv*) pelo que existe em $L^2(\Omega, \sigma(Y), P)$ uma v.a.r. T única, módulo P , verificando as condições equivalentes *i*) e *ii*). ■

Notas.

1. Pelo lema anterior, T é evidentemente da forma $f(Y)$ com f uma função real, definida sobre E e $(\mathcal{E}, \mathcal{B})$ -mensurável.

2. f é de quadrado integrável relativamente a P_Y pois, pelo teorema da transferência ⁽⁸⁾,

$$+\infty > E(T^2) = \int_{\Omega} f^2(Y) dP = \int_E f^2(x) dP_Y(x).$$

3. Esperança condicionada por uma sub-tribo

O teorema anterior tem a seguinte extensão para v.a.r. integráveis:

⁷ Seja H_0 um subespaço de Hilbert dum espaço de Hilbert H . Para todo $x \in H$, existe um único elemento de H_0 , denotado x_0 e chamado projecção de x sobre H_0 , verificando uma das duas condições equivalentes:

i) $\|x - x_0\| = \inf_{y \in H_0} \|x - y\|$

ii) $\forall y \in H_0, \langle x, y \rangle = \langle x_0, y \rangle$ (Rudin, W., Real and Complex Analysis, McGraw-Hill International editions, third edition, 1987, p. 80).

⁸Foata, D., Fuchs, A., Calcul des probabilités, 2e édition, Dunod, 1998.

Teorema. Sejam Y e X duas v.a. definidas sobre o mesmo espaço de probabilidade

$$\begin{aligned} Y &: (\Omega, \mathcal{A}, P) \longrightarrow (E, \mathcal{E}) \\ X &: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, B). \end{aligned}$$

Supomos X integrável. Então existe em $L(\Omega, \sigma(Y), P)$ uma v.a.r. T única (mod. P), denotada $E(X/Y)$, verificando a propriedade seguinte

$$\forall g : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, B_{\mathbb{R}}) \text{ limitada, } E[g(Y) X] = E[g(Y) E(X/Y)].$$

Omite-se a demonstração deste teorema.

Observa-se no entanto que, pelo teorema anterior, se $X \in L^2(\Omega, \mathcal{A}, P)$ então existe em $L^2(\Omega, \sigma(Y), P)$ uma v.a.r. T única (mod. P) verificando a propriedade seguinte

$$\forall g : (E, \mathcal{E}) \longrightarrow (\mathbb{R}, B_{\mathbb{R}}) \text{ limitada, } E[g(Y) X] = E[g(Y) T].$$

Assim $T = E(X/Y)$, P -q.c..

Atendendo à condição equivalente ii) presente no teorema do parágrafo 2, $E(X/Y)$ é então a melhor aproximação de X no sentido do erro quadrático médio.

Definição. A v.a.r. $E(X/Y)$ definida (mod. P) pelo teorema anterior chama-se *esperança condicional de X conhecendo $\sigma(Y)$* (tribo gerada por Y) ou, simplesmente, conhecendo Y .

Mais geralmente, podemos falar de condicionamento por uma qualquer sub-tribo \mathcal{A}_0 de \mathcal{A} .

Definição. (Condicionamento por uma sub-tribo) Sejam (Ω, \mathcal{A}, P) um espaço de probabilidade e \mathcal{A}_0 uma sub-tribo de \mathcal{A} . Se Y designa a aplicação identidade de Ω , Y é $(\mathcal{A}, \mathcal{A}_0)$ -mensurável e $E(X/Y)$ é denotada $E(X/\mathcal{A}_0)$.

Tem-se, pois,

$\forall g : (\Omega, \mathcal{A}_0) \longrightarrow (\mathbb{R}, \mathcal{B})$ limitada, $E[g(Id)X] = E[g(Id)E(X/\mathcal{A}_0)]$.

Note-se que se X é de quadrado integrável podemos concluir que $E(X/\mathcal{A}_0)$ é a projecção ortogonal de X sobre $L^2(\Omega, \mathcal{A}_0, P)$. Logo

$$E(X - E(X/\mathcal{A}_0))^2 \leq E(X - Z)^2$$

para todo $Z \in L^2(\Omega, \mathcal{A}_0, P)$.

Teorema. Seja X uma v.a.r. definida sobre (Ω, \mathcal{A}, P) P -integrável e \mathcal{A}_0 uma sub-tribu de \mathcal{A} . Para toda a v.a.r. $T \in L(\Omega, \mathcal{A}_0, P)$ as proposições seguintes são equivalentes:

i) T é uma versão de $E(X/\mathcal{A}_0)$, isto é, $T = E(X/\mathcal{A}_0)$, P -q.c..

ii) $\forall A \in \mathcal{A}_0, \int X \mathbb{I}_A dP = \int T \mathbb{I}_A dP$.

Demonstração.

$i) \Rightarrow ii)$

Se $A \in \mathcal{A}_0$, \mathbb{I}_A é uma função real \mathcal{A}_0 -mensurável e limitada então, como $T = E(X/\mathcal{A}_0)$, P -q.c.,

$$E(X\mathbb{I}_A) = E(T\mathbb{I}_A) \iff \int_A X dP = \int_A T dP.$$

$ii) \Rightarrow i)$

Suponhamos que $g : (\Omega, \mathcal{A}_0) \rightarrow (\mathbb{R}, \mathcal{B})$ é uma função em pata-mar da forma

$$g = \sum_{i=1}^n y_i \mathbb{I}_{B_i}, B_i \in \mathcal{A}_0.$$

Então

$$\begin{aligned} E(X.g) &= E\left(X \cdot \sum_{i=1}^n y_i \mathbb{I}_{B_i}\right) = \sum_{i=1}^n y_i E(X\mathbb{I}_{B_i}) \\ &= \sum_{i=1}^n y_i \int_{B_i} X dP = \sum_{i=1}^n y_i \int_{B_i} T dP \\ &= \sum_{i=1}^n y_i E(T\mathbb{I}_{B_i}) = E(T.g). \end{aligned}$$

Então $T = E(X/\mathcal{A}_0), P-$ q.c..

Se g é uma função mensurável limitada positiva, $g = \lim_{n \rightarrow +\infty} \nearrow g_n$ em que (g_n) é uma sucessão crescente de funções em patamar positivas e tem-se

$$\forall n \in \mathbb{N}, E(Xg_n) = E(Tg_n)$$

Além disso, $|Xg_n| \leq |X||g| \leq |X|M$, e $|Tg_n| \leq |T|M$.

Então, como $Xg_n \xrightarrow{n \rightarrow +\infty} Xg$, $Tg_n \xrightarrow{n \rightarrow +\infty} Tg$ e

$\lim_{n \rightarrow +\infty} E(Xg_n) \rightarrow \lim_{n \rightarrow +\infty} E(Tg_n)$ tem-se, utilizando o teorema da convergência dominada e a unicidade do limite,

$$E(Xg) = E(Tg)$$

e portanto $T = E(X/\mathcal{A}_0), P-$ q.c..

Se g é uma função mensurável limitada qualquer, $g = g^+ - g^-$, com g^+ e g^- limitadas, e

$$\begin{aligned} E(X.g) &= E(X.g^+) - E(X.g^-) \\ &= E(T.g^+) - E(T.g^-) \\ &= E(T.(g^+ - g^-)), \end{aligned}$$

concluindo-se então que $T = E(X/\mathcal{A}_0), P-$ q.c..

Nota. Se \mathcal{A}_0 é a tribo gerada por uma v.a. Y definida sobre (Ω, \mathcal{A}, P) com valores em (E, \mathcal{E}) , isto é, $\mathcal{A}_0 = \sigma(Y)$, tem-se

$$\begin{aligned} T &= E(X/Y), P- \text{ q.c.}, \text{ se e só se} \\ \forall A \in \sigma(Y), \int X \mathbb{I}_A dP &= \int T \mathbb{I}_A dP. \end{aligned}$$

Exemplo. Seja Y uma v.a.r. discreta de suporte $S = \{y_1, y_2, \dots, y_n\}$. Verifiquemos que

$$T = \sum_{i=1}^n \left[\frac{1}{P(Y=y_i)} \left(\int_{\{Y=y_i\}} X dP \right) \mathbb{I}_{\{Y=y_i\}} \right]$$

é uma versão de $E(X/Y)$.

Queremos provar que

$$\int_A X dP = \int_A T dP, \quad \forall A \in \sigma(Y).$$

Se $A \in \sigma(Y)$ então $A \cap Y^{-1}(S) = \bigcup_{j \in J} \{Y = y_j\}$.

Então

$$\begin{aligned} & \int_A \sum_{i=1}^n \left[\frac{1}{P(Y=y_i)} \left(\int_{\{Y=y_i\}} X dP \right) \mathbb{I}_{\{Y=y_i\}} \right] dP \\ &= \int_{A \cap Y^{-1}(S)} \sum_{i=1}^n \left[\frac{1}{P(Y=y_i)} \left(\int_{\{Y=y_i\}} X dP \right) \mathbb{I}_{\{Y=y_i\}} \right] dP \\ &= \sum_{i=1}^n \frac{1}{P(Y=y_i)} \left(\int_{\{Y=y_i\}} X dP \right) \int_{A \cap Y^{-1}(S)} \mathbb{I}_{\{Y=y_i\}} dP \\ &= \sum_{i=1}^n \frac{1}{P(Y=y_i)} \left(\int_{\{Y=y_i\}} X dP \right) P(A \cap \{Y=y_i\}) \\ &= \sum_{j \in J} \frac{1}{P(Y=y_j)} \left(\int_{\{Y=y_j\}} X dP \right) P(\{Y=y_j\}) = \int_A X dP. \end{aligned}$$

Além disso, T é $\sigma(Y)$ -mensurável porque T é da forma $g(Y)$ com

$$g = \sum_{i=1}^n \left(\frac{1}{P(Y=y_i)} \int_{\{Y=y_i\}} X dP \right) \mathbb{I}_{\{y_i\}}.$$

T é integrável porque

$$\begin{aligned} \int |T| dP &\leq \int \bigcup_{j=1}^n \{Y=y_j\} \sum_{i=1}^n \left[\frac{1}{P(Y=y_i)} \left(\int_{\{Y=y_i\}} |X| dP \right) \mathbb{I}_{\{Y=y_i\}} \right] dP \\ &\leq \int_{\Omega} |X| dP < +\infty. \end{aligned}$$

Assim $T = E(X/Y)$, P -q.c.. ■

4. Propriedades da esperança condicional

O teorema seguinte resume as principais propriedades da esperança condicional.

Teorema. Sejam (Ω, \mathcal{A}, P) um espaço de probabilidade e \mathcal{A}_0 uma sub-tribu de \mathcal{A} . Valem as seguintes propriedades:

a) Se $X \in L(\Omega, \mathcal{A}_0, P)$, $E(X/\mathcal{A}_0) = X$, P -q.c..

Em particular, sendo $\mathcal{A}_0 = \sigma(Y)$ com $Y : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{E})$, tem-se

$$X \in L(\Omega, \sigma(Y), P) \Rightarrow E(X/Y) = X, P - \text{q.c..}$$

b) $\forall a_1, a_2 \in \mathbb{R}, \forall X_1, X_2 \in L(\Omega, \mathcal{A}, P)$

$$E(a_1X_1 + a_2X_2/\mathcal{A}_0) = a_1E(X_1/\mathcal{A}_0) + a_2E(X_2/\mathcal{A}_0), P - \text{q.c..}$$

c) $\forall X \in L(\Omega, \mathcal{A}, P)$, $E(E(X/\mathcal{A}_0)) = E(X)$.

d) $\forall X \in L(\Omega, \mathcal{A}, P)$, $E(h.X/\mathcal{A}_0) = hE(X/\mathcal{A}_0)$, para toda a função $h : (\Omega, \mathcal{A}_0) \rightarrow (\mathbb{R}, \mathcal{B})$ limitada.

e) $\forall X \in L(\Omega, \mathcal{A}, P)$, $X \geq 0 \Rightarrow E(X/\mathcal{A}_0) \geq 0$, P -q.c..

f) $\forall (X_n)_{n \in \mathbb{N}} \in L(\Omega, \mathcal{A}, P)$, $\forall X \in L(\Omega, \mathcal{A}, P)$,

- $(X_n \geq 0 \text{ e } X_n \nearrow X) \Rightarrow E(X_n/\mathcal{A}_0) \nearrow E(X/\mathcal{A}_0)$, P -q.c..
- $X_n \nearrow X \Rightarrow E(X_n/\mathcal{A}_0) \rightarrow E(X/\mathcal{A}_0)$, P -q.c..

g) $\forall X \in L(\Omega, \mathcal{A}, P)$, $\forall h : (\Omega, \mathcal{A}_0) \rightarrow (\mathbb{R}, \mathcal{B})$,

$$E(E(X/\mathcal{A}_0)/\sigma(h)) = E(X/\sigma(h)), P - \text{q.c..}$$

Demonstração.

a) Uma vez que, por hipótese, $X \in L(\Omega, \mathcal{A}_0, P)$ a prova é trivial pois resume-se à identidade $\forall A \in \mathcal{A}_0, \int_A X dP = \int_A X dP$.

b) $a_1E(X_1/\mathcal{A}_0) + a_2E(X_2/\mathcal{A}_0) \in L(\Omega, \mathcal{A}_0, P)$ e
 $\forall A \in \mathcal{A}_0$,

$$\begin{aligned} & \int_A [a_1E(X_1/\mathcal{A}_0) + a_2E(X_2/\mathcal{A}_0)] dP \\ &= a_1 \int_A E(X_1/\mathcal{A}_0) dP + a_2 \int_A E(X_2/\mathcal{A}_0) dP \\ &= a_1 \int_A X_1 dP + a_2 \int_A X_2 dP = \int_A (a_1X_1 + a_2X_2) dP. \end{aligned}$$

c) Por definição $\forall A \in \mathcal{A}_0$, $\int_A X dP = \int_A E(X/\mathcal{A}_0) dP$.
 Logo, se $A = \Omega$,

$$\int_{\Omega} X dP = \int_{\Omega} E(X/\mathcal{A}_0) dP.$$

d) $hE(X/\mathcal{A}_0)$ é \mathcal{A}_0 -mensurável e integrável. Para toda a função
 $g : (\Omega, \mathcal{A}_0) \longrightarrow (\mathbb{R}, \mathcal{B})$ limitada temos

$$E(ghE(X/\mathcal{A}_0)) = E((gh)E(X/\mathcal{A}_0)) = E(g(hX)).$$

Concluimos que $E(hX/\mathcal{A}_0) = hE(X/\mathcal{A}_0)$, q.c..

e) Se $X \in L(\Omega, \mathcal{A}, P)$ e $X \geq 0$, temos

$$\forall A \in \mathcal{A}_0, 0 \leq \int_A X dP = \int_A E(X/\mathcal{A}_0) dP.$$

Se $A = \{E(X/\mathcal{A}_0) < 0\}$, este integral é negativo ou nulo, pelo
 que só pode ser nulo. Então,

$$E(X/\mathcal{A}_0)\mathbb{I}_{\{E(X/\mathcal{A}_0) < 0\}} = 0, \text{ q.c.},$$

o que implica $E(X/\mathcal{A}_0) \geq 0$, q.c..

f) Se $(X_n)_{n \in \mathbb{N}}$ é crescente então, por e),

$$\begin{aligned} \forall m, n \in \mathbb{N}, m > n &\implies X_m - X_n \geq 0 \\ &\implies E(X_m - X_n/\mathcal{A}_0) \geq 0, \text{ q.c..} \end{aligned}$$

Logo, a sucessão $E(X_n/\mathcal{A}_0)$ é crescente q.c. e cada um dos
 seus termos é quase certamente não negativo porque $X_n \geq 0$.

Então, existe $\bar{Z} : (\Omega, \mathcal{A}_0) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B}_{\overline{\mathbb{R}}})$ tal que

$$\bar{Z} = \lim \nearrow E(X_n/\mathcal{A}_0), \text{ q.c.},$$

e tem-se, pela propriedade de Beppo-Levi ⁽⁹⁾,

$$\begin{aligned} E(\bar{Z}) &= E(\lim \nearrow E(X_n/\mathcal{A}_0)) \\ &= \lim \nearrow E(E(X_n/\mathcal{A}_0)) \\ &= \lim \nearrow E(X_n) = E(\lim \nearrow X_n) = E(X) < +\infty. \end{aligned}$$

Concluimos que \bar{Z} é uma função q.c. finita. Podemos substituí-la por uma v.a. Z , \mathcal{A}_0 -mensurável e

$$\lim \nearrow E(X_n/\mathcal{A}_0) = Z, \text{ q.c.}$$

Então, $\forall A \in \mathcal{A}_0$,

$$\begin{aligned} 0 \leq \int_A Z dP &= \int_A \lim \nearrow E(X_n/\mathcal{A}_0) dP = \lim \nearrow \int_A E(X_n/\mathcal{A}_0) dP \\ &= \lim \nearrow \int_A X_n dP = \int_A \lim X_n dP = \int_A X dP \end{aligned}$$

e portanto $Z = E(X/\mathcal{A}_0)$, q.c..

Omitimos a demonstração da segunda parte.

g) Seja $\mathcal{C} = \sigma(h)$. Tem-se $\mathcal{C} \subset \mathcal{A}_0$. Pretende-se então provar que

$$E(E(X/\mathcal{A}_0)/\mathcal{C}) = E(X/\mathcal{C}), \text{ q.c.}, \text{ onde } \mathcal{C} = \sigma(h),$$

isto é, que

$$\forall A \in \mathcal{C}, \int_A E(X/\mathcal{A}_0) dP = \int_A E(X/\mathcal{C}) dP.$$

De facto, uma vez que se $A \in \mathcal{C}$ então $A \in \mathcal{A}_0$, tem-se

$$\forall A \in \mathcal{C}, \int_A E(X/\mathcal{A}_0) dP = \int_A X dP.$$

Assim, usando a definição de esperança condicional

⁹Foata, D., Fuchs, A., Calcul des probabilités, 2e édition, Dunod, 1998.

$$E(X/\mathcal{C}) = E(E(X/\mathcal{A}_0)/\mathcal{C}). \blacksquare$$

Seguindo os mesmos passos que em teoria clássica da integração obtém-se o teorema da convergência dominada para as esperanças condicionais.

Teorema. Seja (X_n) uma sucessão de v.a.r. definidas sobre um mesmo espaço de probabilidade (Ω, \mathcal{A}, P) e X uma v.a.r., definida também sobre este espaço, integrável. Consideremos uma v.a.r. Y definida sobre (Ω, \mathcal{A}, P) e possuindo esperança matemática finita. Supõe-se que

- $\forall n \in \mathbb{N}, |X_n| \leq Y, P$ - q.c.
- $X_n \rightarrow X, P$ - q.c.

Então $E(X_n/\mathcal{A}_0) \rightarrow E(X/\mathcal{A}_0), P$ - q.c., para toda a sub-tribo \mathcal{A}_0 de \mathcal{A} .

Teorema. Se X é uma v.a.r. e Y uma v.a. com valores num espaço mensurável (E, \mathcal{E}) , ambas definidas sobre o mesmo espaço de probabilidade (Ω, \mathcal{A}, P) e independentes, e se X possui esperança matemática então

$$E(X/Y) = E(X), P\text{- q.c.}$$

Prova. Seja $\mathcal{C} = \sigma(Y)$ e seja $B \in \sigma(Y)$.

A v.a.r. \mathbb{I}_B e X são independentes pois $\sigma(Y)$ é independente de $\sigma(X)$. Logo, $\sigma(\mathbb{I}_B)$ é independente de $\sigma(X)$. Consequentemente,

$$\begin{aligned} \int_B E(X/Y) dP &= \int_B X dP = \int \mathbb{I}_B X dP \\ &= E(\mathbb{I}_B X) = E(\mathbb{I}_B) E(X) = P(B) E(X) \\ &= \int_B E(X) dP. \blacksquare \end{aligned}$$

A noção de esperança condicional generaliza-se facilmente ao caso em que X assume valores em $(\mathbb{R}^p, B_{\mathbb{R}^p})$, $p > 1$.

Consideremos então as variáveis aleatórias

$$\begin{aligned} Y & : (\Omega, \mathcal{A}, P) \longrightarrow (E, \mathcal{E}) \\ X & : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^p, B_{\mathbb{R}^p}). \end{aligned}$$

Supondo que X é P -integrável, isto é, que X_1, \dots, X_p são elementos de $L(\Omega, \mathcal{A}, P)$ então, por definição, a *esperança condicional de X sabendo Y* é o ve.a. de $(\mathbb{R}^p, B_{\mathbb{R}^p})$ definido quase certamente por

$$E(X/Y) = (E(X_1/Y), E(X_2/Y), \dots, E(X_p/Y)).$$

No caso do condicionamento por uma sub-tribo a definição é análoga.

Capítulo 3

Processos ARMA unidimensionais

Neste capítulo vamos estudar uma importante classe de modelos auto-projectivos de séries temporais: os modelos auto-regressivos médias móveis, ou simplesmente *ARMA* (do inglês *auto regressive moving average*). Esta classe de modelos pode identificar-se com os modelos lineares de séries temporais e foi introduzida por Box e Jenkins, em 1972, tendo desde aí revelado grandes potencialidades nas aplicações.

Veremos adiante que esta família de modelos permite aproximar, no sentido de L^2 , uma classe bastante geral de processos estocásticos, a dos processos (linearmente) regulares. Tal facto leva-nos a iniciar este capítulo com uma breve referência a estes últimos processos.

3.1 Processos estocásticos regulares

Seja H um espaço de Hilbert e $(H_n, n \in \mathbb{Z})$ uma sucessão de sub-espacos de Hilbert (contida em H), crescente (no sentido da relação de inclusão). Neste caso diz-se, de modo abreviado, que $(H_n, n \in \mathbb{Z})$ é uma sucessão de espacos encaixados.

Representamos por $H_{-\infty}$ a intersecção de tal sucessão de espacos, $\bigcap_{n \in \mathbb{Z}} H_n$, e por $H_{+\infty}$ o subespaco de Hilbert gerado pela sua

união, $\bigcup_{n \in \mathbb{Z}} H_n$. Assim, $H_{+\infty}$ é o menor subespaço de Hilbert contendo aquela união.

Para todo o $n \in \mathbb{Z}$, seja P_n a projecção ortogonal sobre H_n .

O lema seguinte, cuja prova pode ser encontrada em Dacunha-Castelle, Duffo (1983, vol. 2, p. 14 - 15), diz-nos que a projecção no espaço limite da sucessão H_n , quando n tende para $+\infty$ (resp., $-\infty$), é o limite da sucessão das funções projecção (P_n) sobre tal sucessão de espaços; denote-se tal limite por $P_{+\infty}$ (resp., $P_{-\infty}$).

Lema. Para qualquer $x \in H$ tem-se

$$P_{+\infty}(x) = \lim_{n \rightarrow +\infty} P_n(x) \quad \text{e} \quad P_{-\infty}(x) = \lim_{n \rightarrow -\infty} P_n(x),$$

sendo as igualdades no sentido da norma definida em H .

Vejamos como poderemos utilizar este resultado na classe dos processos de segunda ordem.

Seja X um processo estocástico (eventualmente multidimensional) de segunda ordem, isto é, tal que

$$\forall t \in \mathbb{Z}, \quad X_t \in L_2.$$

Considere-se o subespaço linear gerado por $\{X_t, X_{t-1}, \dots\}$, definido no primeiro capítulo, \underline{H}_t^X , $t \in \mathbb{Z}$ ⁽¹⁾.

A sucessão $(\underline{H}_t, t \in \mathbb{Z})$ é composta por subespaços de L_2 encaixados, pois, como vimos, $\underline{H}_{t-1} \subset \underline{H}_t$.

Defina-se $\underline{H}_{-\infty} = \bigcap_{t \in \mathbb{Z}} \underline{H}_t$.

Sendo $E_L(X_t/\underline{H}_s)$ a projecção ortogonal de $X_t \in L_2$ sobre \underline{H}_s , concluimos, pelo lema anterior, que $E_L(X_t/\underline{H}_s)$ converge em L_2 quando $s \rightarrow -\infty$, tendo-se

$$\lim_{s \rightarrow -\infty} E_L(X_t/\underline{H}_s) = E_L(X_t/\underline{H}_{-\infty}).$$

Definição. Seja X um processo de segunda ordem.

- a) X é (linearmente) *singular* se $E_L(X_t/\underline{H}_s) = X_t, \forall s \leq t$.
- b) X é (linearmente) *regular* se

¹Se não for necessário explicitar o processo envolvido, escreveremos apenas \underline{H}_t .

$$\lim_{s \rightarrow -\infty} E_L(X_t/\underline{H}_s) = E_L(X_t/\underline{H}_{-\infty}) = E(X_t), \quad \forall t.$$

Podemos pois dizer que um processo X é singular se as diversas previsões, $E_L(X_t/\underline{H}_s) = X_t, s \leq t$, são previsões perfeitas.

A definição de regularidade, por seu lado, admite a seguinte leitura: diremos que um processo X é regular se o seu passado infinitamente longínquo não traz qualquer informação para prever X_t .

O teorema seguinte é usado como definição de processo singular e regular por diversos autores (cf. Dacunha-Castelle, Duffo, 1983, Vol. 2, p. 15).

Seja \underline{H} o subespaço de Hilbert gerado pela união $\bigcup_{t \in \mathbb{Z}} \underline{H}_t$. Este espaço coincide com o subespaço de Hilbert de L_2 gerado por $(X_t, t \in \mathbb{Z})$ e pelas constantes; representá-lo-emos, tal como no capítulo anterior, por H .

Teorema. a) O processo de segunda ordem centrado X é (linearmente) singular se e somente se

$$\underline{H}_{-\infty} = \underline{H}_{+\infty} = H$$

b) O processo de segunda ordem centrado X é (linearmente) regular se e somente se

$$\underline{H}_{-\infty} = \{0\}.$$

Prova. a) O processo X é (linearmente) singular se e só se

$$\begin{aligned} E_L(X_t/\underline{H}_s) &= X_t, \quad \forall s \leq t \\ \iff X_t &\in \underline{H}_s, \quad \forall s \leq t \\ \iff \forall t \in \mathbb{Z}, &\quad \underline{H}_t = \underline{H}_{t-1}, \end{aligned}$$

tendo em conta a definição de \underline{H}_t e porque os espaços são encaixados.

A última relação é, então, equivalente a

$$\underline{H}_{-\infty} = \underline{H}_{+\infty}.$$

b) O processo X é (linearmente) regular se e só se para todo $t \in \mathbb{Z}$, $E_L(X_t/\underline{H}_{-\infty}) = 0$, porque X é centrado; de modo equivalente tem-se

$$E_L(\underline{H}_{+\infty}/\underline{H}_{-\infty}) = \{0\},$$

isto é, a projecção ortogonal de qualquer elemento de $\underline{H}_{+\infty}$ sobre $\underline{H}_{-\infty}$ é 0.

Esta relação é ainda equivalente a $\underline{H}_{-\infty} = \{0\}$, porque $\underline{H}_{-\infty} \subset \underline{H}_{+\infty}$ (Gouriéroux, Monfort, 1989, Vol 2, Propriedade R23, p. 475). ■

Os dois teoremas seguintes, cujas provas podem ser encontradas em Azencott e Dacunha Castelle (1984, p. 39, 68), ilustram a importância da noção de processo regular pois mostram, em particular, que tais processos são caracterizados por uma representação média móvel infinita unidireccional.

Teorema. Seja X um processo estocástico real de segunda ordem, centrado e estacionário. As condições seguintes são equivalentes:

- a) X é (linearmente) regular.
 b) Existe um ruído branco $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ tal que

$$\forall t \in \mathbb{Z}, \underline{H}_t^X = \underline{H}_t^\varepsilon.$$

Um tal ruído é único a menos de uma constante multiplicativa. Podemos, em particular, escolher

$$\varepsilon_t = X_t - E_L(X_t/\underline{H}_{t-1})$$

isto é, ε_t é a inovação de X_t . Além disso, existe uma sucessão de números reais $(c_k, k \in \mathbb{N}_0)$ tal que $\sum_{k=0}^{+\infty} c_k^2 < +\infty$ e

$$X_t = \sum_{k=0}^{+\infty} c_k \varepsilon_{t-k}, \text{ para todo } t \in \mathbb{Z},$$

isto é, X admite uma representação média móvel infinita unidireccional. ■

Teorema. Seja X um processo estocástico real de segunda ordem, centrado e estacionário. As condições seguintes são equivalentes:

a) X é (linearmente) regular.

b) X admite uma densidade espectral da forma $f(\lambda) = \left| \sum_{k=0}^{+\infty} c_k e^{-ik\lambda} \right|^2$,

$\lambda \in [-\pi, \pi[$, onde a sucessão de números reais $(c_k, k \in \mathbb{N}_0)$ verifica

$$\sum_{k=0}^{+\infty} c_k^2 < +\infty.$$

c) Existe uma sucessão de reais $(c_k, k \in \mathbb{N}_0)$ tal que $\sum_{k=0}^{+\infty} c_k^2 < +\infty$

e um ruído branco $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ tal que $X_t = \sum_{k=0}^{+\infty} c_k \varepsilon_{t-k}$. ■

Exemplos.

1. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico definido por

$$X_t = \varepsilon_t + \Theta_1 \varepsilon_{t-1} + \dots + \Theta_q \varepsilon_{t-q}$$

onde $\Theta_1, \dots, \Theta_q$ são matrizes deterministas e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um

ruído branco de função de autocovariância $\gamma(h) = \begin{cases} \Sigma, & h = 0 \\ 0, & h \neq 0 \end{cases}$,

sendo Σ invertível. Dizemos que X é um processo *média móvel de ordem q , vectorial*.

Mostremos que X é regular. Ora,

$$\forall t, \forall s < t - q, E_L(X_t / \underline{H}_s) = 0$$

onde $\underline{H}_s = \underline{H}_s^X$ é o espaço gerado por $(X_j, j \leq s)$ e pelas constantes.

De facto, para todo t , X_t é combinação linear de $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$ e então

$$\underline{H}_t^X \subset \underline{H}_t^\varepsilon.$$

Assim,

$$\begin{aligned} E_L(X_t / \underline{H}_s) &= E_L(X_t / \underline{H}_s^X) = E_L(E_L(X_t / \underline{H}_s^\varepsilon) / \underline{H}_s^X) \\ &= E_L(0 / \underline{H}_s^X) = 0 \end{aligned}$$

pois, sendo X_t uma função linear de $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$, X_t é ortogonal a $\underline{H}_s^\varepsilon$ para $s < t - q$.

Finalmente, uma vez que, para todo s , $\underline{H}_s \supset \underline{H}_{-\infty}$, obtemos

$$E_L(X_t / \underline{H}_{-\infty}) = E_L(E_L(X_t / \underline{H}_s) / \underline{H}_{-\infty}) = 0$$

e, como $E(X_t) = 0$, concluímos que X é um processo regular.

2. Sejam A e B duas variáveis aleatórias reais de quadrado integrável não correlacionadas. O processo estocástico $X = (X_t, t \in \mathbb{Z})$ definido por

$$X_t = \begin{cases} A, & t \text{ par} \\ B, & t \text{ ímpar} \end{cases}$$

é singular.

De facto, tem-se $\underline{H}_t = H(A, B) = H$, para todo $t \in \mathbb{Z}$. Logo, o processo X é singular.

Observação. Vimos que os processos estacionários (linearmente) regulares podem ser identificados aos processos admitindo uma representação média móvel unidireccional:

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}, V(\varepsilon_t) = 1.$$

As propriedades de segunda ordem dependem, pois, da sucessão de coeficientes $(a_0, a_1, \dots, a_j, \dots)$. Esta representação natural dos processos regulares apresenta o inconveniente claro de envolver uma infinidade de parâmetros $a_j, j \in \mathbb{N}_0$.

Vamos agora introduzir alguns modelos, ou classes de modelos, estacionários que permitem aproximar de modo tão preciso quanto queiramos, no sentido de L_2 , um processo regular, mas apresentando a vantagem de depender apenas de um número finito de parâmetros.

3.2 Processos auto-regressivos

3.2.1 Definição e existência

Definição. a) Um processo estocástico centrado $X = (X_t, t \in \mathbb{Z})$ de segunda ordem (estacionário ou não) admite uma representação auto-regressiva de ordem p , denotada $AR(p)$, se existem números reais $\varphi_1, \dots, \varphi_p$ e um ruído branco $(\varepsilon_t, t \in \mathbb{Z})$ de variância σ^2 , ($\sigma^2 > 0$), tais que

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t,$$

com $\varphi_p \neq 0$.

b) Diz-se que tal representação é *canônica* se ε_t é ortogonal a \underline{H}_{t-1}^X , isto é, se ε_t é não correlacionado com o passado de X à data t .

Se introduzirmos o operador atraso L associado a X , tal que $LX_t = X_{t-1}$, e designarmos o polinômio auto-regressivo por $\Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, vemos que a relação de recorrência presente na definição anterior se escreve na forma

$$\Phi(L) X_t = \varepsilon_t.$$

A definição anterior suscita, entre outras, as seguintes questões:

- Dados $\varphi_1, \dots, \varphi_p$ e ε , existe um processo X que admita tal representação? Mais, existirá um que seja *estacionário*?
- Existindo um processo verificando tal representação, será ele único?
- Se um processo X admite uma tal representação, poderá ele verificar outras representações AR ? Ou seja, que podemos dizer sobre a unicidade da representação?
- Se um processo X admite várias representações, existirá sempre uma representação canônica?

Exemplo. Consideremos um processo estocástico $X = (X_t, t \in \mathbb{Z})$ tal que $X_t = \varepsilon_t, t \in \mathbb{Z}$, com $(\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância $\sigma^2 > 0$. Neste caso dizemos, abusivamente, que X admite uma representação $AR(0)$.

As respostas às primeira e quarta questões são claramente afirmativas. As restantes são, neste momento, menos óbvias.

A análise das várias questões apresentadas será feita recorrendo ao caso particular de um processo auto-regressivo de ordem 1, a partir do qual se generaliza o estudo ao caso geral.

3.2.2 Processo AR(p)

1. Caso particular p=1

Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico real admitindo a seguinte representação

$$X_t = \varphi X_{t-1} + \varepsilon_t, \quad (3.2.2.1)$$

com $\varphi \in \mathbb{R} \setminus \{0\}$ e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância $\sigma^2 > 0$.

Analisemos a existência de processos admitindo tal representação, dados φ e ε .

Temos

$$\begin{aligned} X_t &= \varphi X_{t-1} + \varepsilon_t \\ &= \varphi(\varphi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \dots \\ &= \varepsilon_t + \varphi \varepsilon_{t-1} + \dots + \varphi^J \varepsilon_{t-J} + \varphi^{J+1} X_{t-J-1} \\ &= \sum_{n=0}^J \varphi^n \varepsilon_{t-n} + \varphi^{J+1} X_{t-J-1}, \forall J \end{aligned}$$

Se $|\varphi| < 1$, a parcela $\sum_{n=0}^J \varphi^n \varepsilon_{t-n}$ converge, quando J tende para $+\infty$, no sentido de L_2 . Quanto a $\varphi^{J+1} X_{t-J-1}$, converge se X for, por exemplo, limitado em L_2 , tendo-se

$$\begin{aligned} \|\varphi^{J+1} X_{t-J-1}\| &= \left[E (\varphi^{J+1} X_{t-J-1})^2 \right]^{\frac{1}{2}} \\ &= |\varphi|^{J+1} \left[E (X_{t-J-1}^2) \right]^{\frac{1}{2}} \xrightarrow{J \rightarrow +\infty} 0. \end{aligned}$$

Nestas condições, obtivemos uma solução para o modelo (3.2.2.1); poderão existir outras soluções (não limitadas, por exemplo).

Confirmemos a existência de outras soluções particulares.

Fixemos uma data inicial $t_0 = 0$ e escolhamos arbitrariamente uma v.a. X_0 que interpretamos como o valor do processo nessa data. Este procedimento é motivado pelo facto de estar presente no modelo (3.2.2.1) uma equação às diferenças, sendo então necessário impôr uma condição inicial para fixar a solução.

Fixado X_0 obtemos, sem ambiguidade, todos os valores futuros do processo:

$$\begin{aligned} X_1 &= \varphi X_0 + \varepsilon_1 \\ X_2 &= \varepsilon_2 + \varphi \varepsilon_1 + \varphi^2 X_0 \\ &\dots \\ X_t &= \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \dots + \varphi^{t-1} \varepsilon_1 + \varphi^t X_0, \text{ para } \forall t > 0. \end{aligned}$$

Da mesma forma obteremos os valores do processo correspondentes a $t < 0$ notando que, sendo $\varphi \neq 0$,

$$X_t = \varphi X_{t-1} + \varepsilon_t \iff X_{t-1} = \frac{1}{\varphi} X_t - \frac{1}{\varphi} \varepsilon_t.$$

Assim,

$$\begin{aligned} X_{-1} &= \frac{1}{\varphi} X_0 - \frac{1}{\varphi} \varepsilon_0 \\ X_{-2} &= \frac{1}{\varphi^2} X_0 - \frac{1}{\varphi} \varepsilon_{-1} - \frac{1}{\varphi^2} \varepsilon_0 \\ &\dots \\ X_{-t} &= \frac{1}{\varphi^t} X_0 - \frac{1}{\varphi} \varepsilon_{-t+1} - \frac{1}{\varphi^2} \varepsilon_{-t+2} - \dots - \frac{1}{\varphi^t} \varepsilon_0. \end{aligned}$$

Definimos, pois, sem ambiguidade, o processo $(X_t, t \in \mathbb{Z})$ fixando X_0 .

Como escolher X_0 ? Se X_0 for de quadrado integrável, é claro que, para todo $t \in \mathbb{Z}$, $X_t \in L_2$.

Podemos pois afirmar que, qualquer que seja $\varphi \neq 0$ e qualquer que seja o ruído ε , existem processos de segunda ordem verificando

a relação (3.2.2.1). Estes processos dependem da escolha da v.a. X_0 .

Assim a equação (3.2.2.1) admite uma infinidade de processos solução de segunda ordem.

Estudamos soluções particulares do modelo (3.2.2.1). Procuremos agora a solução geral.

Sejam X_t uma solução particular da equação e \tilde{X}_t qualquer outra solução da equação. O processo diferença

$$Y_t = \tilde{X}_t - X_t, \quad t \in \mathbb{Z},$$

verifica

$$\begin{aligned} Y_t - \varphi Y_{t-1} &= \tilde{X}_t - X_t - \varphi(\tilde{X}_{t-1} - X_{t-1}) \\ &= \tilde{X}_t - \varphi\tilde{X}_{t-1} - (X_t - \varphi X_{t-1}) = 0 \end{aligned}$$

ou seja, $Y = (Y_t, t \in \mathbb{Z})$ é solução da equação homogénea.

A solução geral da equação é, pois,

$$\tilde{X}_t = X_t + Y_t,$$

isto é, a solução geral da equação é obtida somando uma solução particular da equação à solução geral da equação homogénea.

Analisando Y obtemos

$$Y_t = \varphi Y_{t-1} = \varphi^2 Y_{t-2} = \dots = \varphi^t Y_0, \quad t \in \mathbb{Z},$$

pelo que a solução geral do modelo (3.2.2.1) é

$$\tilde{X}_t = X_t + \varphi^t Y_0, \quad t \in \mathbb{Z},$$

onde $X = (X_t, t \in \mathbb{Z})$ é uma solução particular.

Analisemos agora a existência de soluções estacionárias.

a) Começemos por estudar o caso em que $|\varphi| < 1$.

Suponhamos que há uma solução estacionária, X . Esta solução verifica, tal como vimos,

$$X_t = \sum_{n=0}^J \varphi^n \varepsilon_{t-n} + \varphi^{J+1} X_{t-J-1}, \quad \forall J.$$

Ora $\sum_{n=0}^J \varphi^n \varepsilon_{t-n}$ tende (em L_2) para $\sum_{n=0}^{+\infty} \varphi^n \varepsilon_{t-n}$ e $\varphi^{J+1} X_{t-J-1}$ tende (em L_2) para zero, quando $J \rightarrow +\infty$, uma vez que sendo X estacionário, é limitado em L_2 ($\forall t \in \mathbb{Z}, E(X_t^2) = \text{constante} < +\infty$).

Assim, um processo estacionário solução de (3.2.2.1) é necessariamente da forma

$$X_t = \sum_{n=0}^{+\infty} \varphi^n \varepsilon_{t-n}.$$

Inversamente, sabemos que um processo que admite uma representação média móvel infinita com coeficientes $a_j = \varphi^j$ (absolutamente somáveis) é estacionário.

Assim, se $|\varphi| < 1$ há uma única solução estacionária.

Observação. Todas as outras soluções são da forma

$$\tilde{X}_t = \sum_{n=0}^{+\infty} \varphi^n \varepsilon_{t-n} + \varphi^t Y_0.$$

Notemos que para t suficientemente grande se tem $\tilde{X}_t \cong X_t$ ⁽²⁾ já que $\varphi^t Y_0 \rightarrow 0$ (em L_2), ou seja, as soluções \tilde{X}_t são *assintoticamente estacionárias*.

Vejamos que a solução estacionária obtida corresponde a uma representação canónica.

Ora,

$$X_t = \sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j} \in \underline{H}_t^\varepsilon$$

² A notação $\tilde{X}_t \cong X_t$ significa $E(\tilde{X}_t - X_t)^2 \rightarrow 0, t \rightarrow +\infty$.

de onde deduzimos que $\underline{H}_t^X \subset \underline{H}_t^\varepsilon$.

Mas ε_t é ortogonal a $\underline{H}_{t-1}^\varepsilon$. Então ε_t é ortogonal a \underline{H}_{t-1}^X .

Observemos que sempre que temos uma representação canónica de X , ε_t é a inovação no instante t . De facto,

$$\begin{aligned} E_L(X_t / \underline{H}_{t-1}^X) &= E_L(\varphi X_{t-1} + \varepsilon_t / \underline{H}_{t-1}^X) \\ &= \varphi X_{t-1} \end{aligned}$$

pois ε_t é ortogonal a \underline{H}_{t-1}^X .

Assim $\varepsilon_t = X_t - \varphi X_{t-1} = X_t - E_L(X_t / \underline{H}_{t-1}^X)$ é a inovação de X no instante t .

b) Estudemos agora o caso em que $|\varphi| > 1$.

Suponhamos que há uma solução estacionária.

De $X_t = \varphi X_{t-1} + \varepsilon_t$ podemos escrever $X_{t-1} = \frac{1}{\varphi} X_t - \frac{1}{\varphi} \varepsilon_t$, isto é, para todo $t \in \mathbb{Z}$,

$$\begin{aligned} X_t &= \frac{1}{\varphi} X_{t+1} - \frac{1}{\varphi} \varepsilon_{t+1} \\ &= \dots \\ &= -\sum_{j=1}^J \frac{1}{\varphi^j} \varepsilon_{t+j} + \frac{1}{\varphi^J} X_{t+J}. \end{aligned}$$

Quando $J \rightarrow +\infty$, a primeira parcela converge para $-\sum_{j=1}^{+\infty} \frac{1}{\varphi^j} \varepsilon_{t+j}$

e a segunda converge para zero (convergências em L_2).

Seguindo o mesmo raciocínio da parte anterior, obtemos uma única solução estacionária:

$$X_t = -\sum_{j=1}^{+\infty} \frac{1}{\varphi^j} \varepsilon_{t+j}.$$

Observações.

1. Esta solução faz intervir os valores futuros do ruído.

2. Todas as outras soluções são da forma

$$\tilde{X}_t = -\sum_{j=1}^{+\infty} \frac{1}{\varphi^j} \varepsilon_{t+j} + \varphi^t Y_0,$$

pois a equação é a mesma de há pouco. Notemos que, quando $t \rightarrow +\infty$, a parcela $|\varphi^t Y_0|$ tende para $+\infty$. Assim, todas as soluções não estacionárias são processos explosivos.

Retomando a solução estacionária, podemos analisar se a representação obtida é canónica, isto é, se o ruído interveniente ε_t é ortogonal ao passado \underline{H}_{t-1}^X .

Ora

$$\begin{aligned} \text{Cov}(\varepsilon_t, X_{t-1}) &= E(\varepsilon_t X_{t-1}) \\ &= E\left[\varepsilon_t \left(-\frac{1}{\varphi} \varepsilon_t - \frac{1}{\varphi^2} \varepsilon_{t+1} - \dots\right)\right] \\ &= -\frac{1}{\varphi} \sigma^2. \end{aligned}$$

Sendo $\text{Cov}(\varepsilon_t, X_{t-1}) \neq 0$, ε_t não é ortogonal a \underline{H}_{t-1}^X pelo que a representação não é canónica.

Vejamos que o processo X admite uma representação canónica $AR(1)$.

De $X_t - \varphi X_{t-1} = \varepsilon_t$ deduzimos, como já vimos, que a densidade espectral de X é dada por

$$f_X(\omega) |1 - \varphi e^{i\omega}|^2 = f_\varepsilon(\omega), \quad \omega \in [-\pi, \pi[.$$

Mas $f_\varepsilon(\omega) = \frac{\sigma^2}{2\pi}$ pelo que

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{i\omega}|^2}.$$

Consideremos o processo η definido por

$$\eta_t = X_t - \frac{1}{\varphi} X_{t-1}, \quad t \in \mathbb{Z}.$$

A sua densidade espectral é

$$\begin{aligned}
 f_{\eta}(\omega) &= \left|1 - \frac{1}{\varphi}e^{i\omega}\right|^2 f_X(\omega) \\
 &= \frac{\sigma^2 \left|1 - \frac{1}{\varphi}e^{i\omega}\right|^2}{2\pi \left|1 - \varphi e^{i\omega}\right|^2} \\
 &= \frac{\sigma^2}{2\pi} \frac{1}{|\varphi|^2} \frac{\left|1 - \frac{1}{\varphi}e^{i\omega}\right|^2}{\left|\frac{1}{\varphi} - e^{i\omega}\right|^2} \\
 &= \frac{\sigma^2}{2\pi} \frac{1}{|\varphi|^2}, \quad \omega \in [-\pi, \pi[.
 \end{aligned}$$

Então η é um ruído branco de variância $\frac{\sigma^2}{|\varphi|^2}$, tendo em conta a caracterização que vimos do ruído branco através da densidade espectral.

O processo X admite, pois, duas representações

$$\begin{aligned}
 X_t - \varphi X_{t-1} &= \varepsilon_t, \quad V(\varepsilon_t) = \sigma^2 \\
 X_t - \frac{1}{\varphi} X_{t-1} &= \eta_t, \quad V(\eta_t) = \frac{\sigma^2}{|\varphi|^2}.
 \end{aligned}$$

Esta segunda representação é canónica, já que $\left|\frac{1}{\varphi}\right| < 1$.

c) Analisemos finalmente os casos em que $\varphi = 1$ ou $\varphi = -1$.

Suponhamos que há uma solução X estacionária. Podemos sempre escrever

$$X_t - \varphi^J X_{t-J} = \varepsilon_t + \varphi \varepsilon_{t-1} + \dots + \varphi^{J-1} \varepsilon_{t-(J-1)}.$$

Vejamos quanto vale a variância de cada um dos membros.

$$\begin{aligned}
 V(X_t - \varphi^J X_{t-J}) &= V(X_t) + V(X_{t-J}) - 2\varphi^J \text{Cov}(X_t, X_{t-J}) \\
 &= 2V(X_t) - 2\varphi^J \text{Cov}(X_t, X_{t-J}) \\
 &\leq 2V(X_t) + 2|\text{Cov}(X_t, X_{t-J})| \\
 &\leq 4V(X_t)
 \end{aligned}$$

pois, pela desigualdade de Schwartz, $[Cov(X, Y)]^2 \leq V(X) V(Y)$.

Assim $X_t - \varphi^J X_{t-J}$ é majorado, em L_2 , por $4\gamma(0)$.

Do lado direito

$$V(\varepsilon_t + \varphi\varepsilon_{t-1} + \dots + \varphi^{J-1}\varepsilon_{t-(J-1)}) = J\sigma^2$$

que tende para $+\infty$ quando $J \rightarrow +\infty$.

Encontramos pois uma contradição pelo que, quando $\varphi = \pm 1$, não há soluções estacionárias.

Este resultado depende obviamente da hipótese $\sigma \neq 0$. Se $\sigma = 0$ então $\varepsilon = 0$ quase certamente (porque $E(\varepsilon_t) = 0$) e então se, por exemplo, $\varphi = 1$ obtém-se

$$X_t - X_{t-1} = 0 \Rightarrow X_t = Y, \text{ independente de } t.$$

Este processo $X_t = Y, \forall t \in \mathbb{Z}$, é estacionário desde que Y seja uma v.a.r. de L_2 .

2. Caso geral

O que fizemos para um modelo $AR(1)$ pode ser feito mais geralmente para um auto-regressivo de ordem p , $p \geq 1$. As propriedades que a seguir estabelecemos resumem esse estudo.

Começemos então por assegurar a existência de soluções de segunda ordem para o modelo auto-regressivo.

Propriedade. Qualquer que seja o polinómio $\Phi(L)$ de grau p e o ruído branco ε , existem processos de segunda ordem que verificam a equação

$$\Phi(L)X_t = \varepsilon_t, \forall t \in \mathbb{Z}.$$

Prova. A demonstração é semelhante à que seguimos com os processos $AR(1)$. É, no entanto, necessário fixar p condições iniciais de segunda ordem escolhidas arbitrariamente. Há, pois, uma infinidade de processos solução.

A existência de solução estacionária é agora assegurada pelo resultado seguinte:

Propriedade.

a) Entre as soluções X de $\Phi(L)X_t = \varepsilon_t, t \in \mathbb{Z}$, há uma estacionária se e só se todas as raízes do polinómio $\Phi(L)$ forem de módulos diferentes de 1.

b) Além disso, uma tal solução estacionária é única.

Prova. A prova da condição suficiente de estacionaridade passa pela inversão do polinómio $\Phi(L)$.

Suponhamos então que as raízes z_1, \dots, z_p do polinómio $\Phi(L)$ são de módulos diferentes de 1. Sejam, sem perda de generalidade, as primeiras r de módulos superiores a 1 e as restantes $p - r$ de módulos inferiores a 1 (onde $r \in \{0, 1, \dots, p\}$).

Podemos escrever

$$\begin{aligned} \Phi(L) &= 1 - \sum_{i=1}^p \varphi_i L^i = \prod_{j=1}^r \left(1 - \frac{L}{z_j}\right) \prod_{j=r+1}^p \left(1 - \frac{L}{z_j}\right) \\ &= \prod_{j=1}^r \left(1 - \frac{L}{z_j}\right) \prod_{j=r+1}^p (z_j - L) \frac{1}{\prod_{j=r+1}^p z_j} \\ &= \prod_{j=1}^r \left(1 - \frac{L}{z_j}\right) (-L)^{p-r} \prod_{j=r+1}^p (1 - z_j F) \frac{1}{\prod_{j=r+1}^p z_j} \end{aligned}$$

onde $F = L^{-1}$ designa o operador avanço.

$\Phi(L)$ será invertível se

$$\begin{cases} \forall j \in \{1, \dots, r\}, \left(1 - \frac{L}{z_j}\right) \text{ for invertível} \\ \forall j \in \{r+1, \dots, p\}, (1 - z_j F) \text{ for invertível} \end{cases}$$

já que o outro factor verifica

$$\begin{aligned} \left[(-L)^{p-r} \frac{1}{\prod_{j=r+1}^p z_j} \right]^{-1} &= \left[\prod_{j=r+1}^p \left(-\frac{L}{z_j} \right) \right]^{-1} \\ &= \prod_{j=r+1}^p (-z_j) F = \left[\prod_{j=r+1}^p (-z_j) \right] F^{p-r}. \end{aligned}$$

Ora, a aplicação $1-\lambda L$ é invertível se e só se $|\lambda| \neq 1$ (Gouriéroux, Monfort (1990)).

Então, como $\forall j \in \{1, 2, \dots, r\}, |z_j| > 1$, tem-se que $\left(1 - \frac{L}{z_j}\right)$ é invertível; por outro lado, $\forall j \in \{r+1, \dots, p\}, |z_j| < 1$, e consequentemente $(1 - z_j F)$ é invertível. Logo $\Phi(L)$ é invertível.

Ter-se-á então

$$X_t = \left[\prod_{j=1}^r \left(\sum_{i=0}^{+\infty} \left(\frac{1}{z_j} \right)^i L^i \right) \prod_{j=r+1}^p \left(\sum_{i=0}^{+\infty} z_j^i F^i \right) \prod_{j=r+1}^p (-z_j) F^{p-r} \right] \varepsilon_t$$

ou seja, é esta a forma do único processo estacionário X que verifica $\Phi(L) X_t = \varepsilon_t$.

Notemos que $\Phi(L)$ e $[\Phi(L)]^{-1}$ são séries em L uma vez que o produto (*i.e.*, composição) de séries em L (e em F) é uma série em L .

Podemos pois concluir que o processo estacionário solução $X = (X_t, t \in \mathbb{Z})$ é uma média móvel infinita,

$$X_t = \sum_{j=-\infty}^{+\infty} a_j \varepsilon_{t-j},$$

em geral bidireccional.

Será unidireccional se todas as raízes forem de módulos superiores a um, tendo-se

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}$$

ou se todas as raízes forem de módulos inferiores a um tendo-se, neste caso,

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t+j}.$$

Acabamos então de provar que, se todas as raízes de $\Phi(L)$ são de módulos diferentes de um, há uma solução estacionária única para $\Phi(L)X_t = \varepsilon_t$.

Para provar a necessidade da condição de estacionaridade, suponhamos agora que há uma solução estacionária única para $\Phi(L)X_t = \varepsilon_t$ e mostremos que todas as raízes de $\Phi(L)$ são de módulos diferentes de um.

Procederemos por absurdo, supondo que uma das raízes, z_l , é de módulo igual a 1, isto é, $z_l = e^{i\omega_0}$ (e de multiplicidade 1 para não sobrecarregar as notações). Mostraremos que não há soluções médias-móveis infinitas estacionárias (o que é mais restritivo do que a simples estacionaridade).

Seja então X uma solução estacionária de densidade espectral $f_X(\omega)$, $\omega \in [-\pi, \pi[$.

Da relação $\Phi(L)X_t = \varepsilon_t$ tem-se

$$\frac{\sigma^2}{2\pi} = f_X(\omega) |\Phi(e^{i\omega})|^2.$$

Então, sendo z_j , $j = 1, \dots, p$, as raízes do polinómio em $e^{i\omega}$, $\Phi(e^{i\omega})$, podemos escrever

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi} \frac{1}{\left| \prod_{j=1}^p (z_j - e^{i\omega}) \right|^2 |\varphi_p|^2} \\ &= \frac{\sigma^2}{2\pi |\varphi_p|^2} \frac{1}{\left| \prod_{j=1, j \neq l}^p (z_j - e^{i\omega}) \right|^2} \frac{1}{|z_l - e^{i\omega}|^2} \\ &= \frac{\sigma^2}{2\pi |\varphi_p|^2} \frac{1}{\left| \prod_{j=1, j \neq l}^p (z_j - e^{i\omega}) \right|^2} \frac{1}{|e^{i\omega_0} - e^{i\omega}|^2}. \end{aligned}$$

Quando ω tende para ω_0 obtém-se $f_X(\omega) \simeq k \frac{1}{(\omega - \omega_0)^2}$.

Ora

$$\begin{aligned} \int_{-\pi}^{+\pi} \frac{1}{(\omega - \omega_0)^2} d\omega &= \int_{-\pi}^{\omega_0} \frac{1}{(\omega - \omega_0)^2} d\omega + \int_{\omega_0}^{\pi} \frac{1}{(\omega - \omega_0)^2} d\omega \\ &= \lim_{\delta \rightarrow 0^+} \left\{ -\frac{1}{\omega - \omega_0} \Big|_{-\pi}^{\omega_0 - \delta} - \frac{1}{\omega - \omega_0} \Big|_{\omega_0 + \delta}^{\pi} \right\} \\ &= +\infty \end{aligned}$$

pele que a análise da integrabilidade desta função contraria a existência de densidade espectral. ■

Observações. Do teorema anterior decorrem as seguintes conclusões sobre a solução estacionária X do modelo $AR(p)$:

1. Tem-se $X_t = \sum_{i=0}^{+\infty} a_i L^i \varepsilon_t$ se e só se todas as raízes de $\Phi(L)$ são de módulos superiores a 1.
2. Tem-se $X_t = \sum_{i=0}^{+\infty} a_i F^i \varepsilon_t$ se e só se todas as raízes de $\Phi(L)$ são de módulos inferiores a 1.
3. Quando $X_t = \sum_{i=0}^{+\infty} a_i L^i \varepsilon_t$ tem-se $\underline{H}_t^X \subset \underline{H}_t^\varepsilon$. Mas, sendo

$$\varepsilon_t = X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p}$$

tem-se também a inclusão inversa. Ou seja, $\underline{H}_t^\varepsilon = \underline{H}_t^X$.

Assim esta representação é canónica uma vez que ε_t é ortogonal a $\underline{H}_{t-1}^\varepsilon = \underline{H}_{t-1}^X$.

Vejamus que, neste caso, ε_t é a inovação de X_t . Temos

$$\begin{aligned} E_L(X_t / \underline{H}_{t-1}^X) &= E_L(\varepsilon_t + \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} / \underline{H}_{t-1}^X) \\ &= \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} \end{aligned}$$

pele que $\varepsilon_t = X_t - E_L(X_t / \underline{H}_{t-1}^X)$.

A propriedade seguinte estabelece a existência de múltiplas representações.

Propriedade. Sejam $\Phi(L)$ um polinómio de grau p de raízes de módulos diferentes de 1 e X a solução estacionária de $\Phi(L)X_t = \varepsilon_t$, $t \in \mathbb{Z}$. O processo X admite outras representações auto-regressivas.

Prova. Uma vez que $\Phi(L)X_t = \varepsilon_t$ e que X tem uma representação $MA(+\infty)$, a densidade espectral de X verifica

$$\frac{\sigma^2}{2\pi} = f_X(\omega) \left| 1 - \sum_{j=1}^p \varphi_j e^{i\omega j} \right|^2, \quad \omega \in [-\pi, \pi[,$$

isto é,

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{\left| 1 - \sum_{j=1}^p \varphi_j e^{i\omega j} \right|^2} = \frac{\sigma^2}{2\pi} \frac{1}{|\Phi(e^{i\omega})|^2}$$

ou ainda, designando por z_j as raízes de Φ ,

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{\left| \prod_{j=1}^p \left(1 - \frac{e^{i\omega}}{z_j} \right) \right|^2}.$$

Suponhamos que há r raízes de módulos superiores a 1, z_1, \dots, z_r , e $p-r$ de módulos inferiores a 1, z_{r+1}, \dots, z_p (com $r \in \{0, 1, \dots, p\}$).

Consideremos o polinómio seguinte

$$\Phi^*(z) = \prod_{i=1}^r \left(1 - \frac{L}{z_i} \right) \prod_{i=r+1}^p (1 - z_i L)$$

obtido a partir de Φ substituindo as raízes de módulos inferiores a 1 pelas suas inversas.

Seja $(\eta_t, t \in \mathbb{Z})$ o processo estacionário definido por

$$\eta_t = \Phi^*(L)X_t.$$

Mostremos que η é um ruído branco analisando a sua densidade espectral.

$$\begin{aligned} f_\eta(\omega) &= |\Phi^*(e^{i\omega})|^2 f_X(\omega) = \frac{\sigma^2 |\Phi^*(e^{i\omega})|^2}{2\pi |\Phi(e^{i\omega})|^2} \\ &= \frac{\sigma^2}{2\pi} \prod_{j=r+1}^p \frac{|1 - z_j e^{i\omega}|^2}{\left|1 - \frac{e^{i\omega}}{z_j}\right|^2} \\ &= \frac{\sigma^2}{2\pi} \prod_{j=r+1}^p |z_j|^2, \quad \omega \in [-\pi, \pi[. \end{aligned}$$

Sendo a densidade espectral independente de ω , concluímos que $(\eta_t, t \in \mathbb{Z})$ é um ruído branco de variância $\sigma^2 \prod_{j=r+1}^p |z_j|^2$, que é inferior à variância de ε_t .

Obtivemos assim outra representação $AR(p)$ para o processo X .

Observação. A demonstração anterior permite concluir que tomando, consoante os casos, uma raiz ou a sua inversa, podemos obter múltiplas representações de $(X_t, t \in \mathbb{Z})$; por exemplo, se as raízes são reais e distintas obtemos 2^p representações distintas.

Além disso, podemos sempre reduzir-nos a um polinómio auto-regressivo de raízes em módulo superior a 1; claro que só neste caso a representação de $(X_t, t \in \mathbb{Z})$ faz intervir um ruído branco que se interpreta como a inovação. É a única representação canónica; designa-se então por representação canónica de $(X_t, t \in \mathbb{Z})$.

A propriedade seguinte resume toda esta informação. ⁽³⁾

Teorema. Sejam $\Phi(L)$ um polinómio de grau p de raízes de módulos diferentes de 1 e X a solução estacionária de $\Phi(L)X_t = \varepsilon_t$.

X admite outras representações auto-regressivas que se obtêm a partir de uma das representações substituindo algumas das raízes pelos seus inversos (e mantendo as restantes).

³Para demonstrar esta propriedade faltaria apenas provar que todas as representações são de mesmo grau p e que só podem ser obtidas substituindo um subconjunto qualquer de raízes de Φ pelos seus inversos.

Em particular:

- a) todas as representações são do mesmo grau, p ;
- b) se todas as raízes são reais e distintas há 2^p representações diferentes;
- c) tal processo estacionário auto-regressivo admite uma representação canônica única.

Note-se que no caso de existência de raízes complexas, a substituição deverá ser feita por pares de raízes conjugadas para que o polinômio seja de coeficientes reais. Assim, se todas as raízes são complexas e distintas, haverá $2^{\frac{p}{2}}$ representações diferentes.

3.2.3 Resumo das ligações temporais

Seja X um processo estacionário verificando

$$\Phi(L) X_t = \varepsilon_t$$

com $V(\varepsilon_t) = \sigma^2 > 0$ e onde todas as raízes do polinômio Φ , de grau p , são de módulo superior a 1. Partimos, assim, da representação canônica do processo auto-regressivo de ordem p , $X = (X_t, t \in \mathbb{Z})$.

Determinemos a expressão da função de autocorrelação, da função densidade espectral e das autocorrelações parciais do processo X .

Função de autocovariância, função de autocorrelação

Começemos por determinar a expressão de $V(X_t) = \gamma(0)$.

A partir de

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + \varepsilon_t$$

obtemos

$$\begin{aligned} \text{Cov}(X_t, \varepsilon_t) &= \text{Cov}(X_t, X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p}) \\ &= \gamma(0) - \varphi_1 \gamma(1) - \dots - \varphi_p \gamma(p). \end{aligned}$$

Por outro lado, uma vez que partimos da representação canônica, temos

$$X_t = \sum_{i=0}^{+\infty} a_i \varepsilon_{t-i}, \quad \text{com } a_0 = 1,$$

e então

$$\text{Cov}(X_t, \varepsilon_t) = \text{Cov}(\varepsilon_t + a_1\varepsilon_{t-1} + \dots, \varepsilon_t) = \sigma^2.$$

Consequentemente

$$\sigma^2 = \gamma(0) - \varphi_1\gamma(1) - \dots - \varphi_p\gamma(p) = \gamma(0) [1 - \varphi_1\rho(1) - \dots - \varphi_p\rho(p)]$$

e portanto a variância de X_t é

$$\gamma(0) = \frac{\sigma^2}{1 - \varphi_1\rho(1) - \dots - \varphi_p\rho(p)},$$

onde ρ é a função de autocorrelação de X .

Analisemos agora as autocorrelações, $\rho(h)$, $h \geq 1$.

Como ε_t é ortogonal a $\underline{H}_{t-1}(X)$, tem-se

$$\text{Cov}(X_{t-h}, \varepsilon_t) = 0, h \geq 1,$$

ou seja, para todo $h \geq 1$,

$$\text{Cov}(X_{t-h}, X_t - \varphi_1X_{t-1} - \dots - \varphi_pX_{t-p}) = 0,$$

isto é,

$$\gamma(h) - \varphi_1\gamma(h-1) - \dots - \varphi_p\gamma(h-p) = 0$$

ou de modo equivalente,

$$\rho(h) - \varphi_1\rho(h-1) - \dots - \varphi_p\rho(h-p) = 0.$$

Podemos então enunciar o resultado seguinte:

Propriedade. A sucessão das autocorrelações de um processo $AR(p)$ estacionário verifica a equação de recorrência linear

$$\rho(h) - \varphi_1\rho(h-1) - \dots - \varphi_p\rho(h-p) = 0, \quad \forall h \geq 1.$$

A função de autocorrelação é, pois, completamente determinada a partir do conhecimento de p condições iniciais $\rho(1)$, $\rho(2)$, ..., $\rho(p)$. Tais condições obtêm-se resolvendo o sistema seguinte, que decorre da propriedade anterior e onde se usa o facto de $\rho(0) = 1$ e ρ ser uma função par:

$$\begin{cases} \rho(1) - \varphi_1 - \varphi_2\rho(1) - \dots - \varphi_p\rho(1-p) = 0 \\ \rho(2) - \varphi_1\rho(1) - \varphi_2 - \dots - \varphi_p\rho(2-p) = 0 \\ \dots \\ \rho(p) - \varphi_1\rho(p-1) - \varphi_2\rho(p-2) - \dots - \varphi_p = 0 \end{cases}.$$

Notemos que podemos interpretá-lo como um sistema linear em $\varphi_1, \varphi_2, \dots, \varphi_p$, pois é equivalente ao seguinte, dito *sistema de Yule-Walker*,

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(p) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(1-p) \\ \rho(1) & 1 & \dots & \rho(2-p) \\ \dots & \dots & \dots & \dots \\ \rho(p-1) & \rho(p-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \dots \\ \varphi_p \end{bmatrix}.$$

É também um sistema linear em $\rho(1), \rho(2), \dots, \rho(p)$ admitindo a seguinte forma equivalente

$$\begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \dots \\ \varphi_p \end{bmatrix} = \begin{bmatrix} 1-\varphi_2 & -\varphi_3 & \dots & 0 \\ -\varphi_1-\varphi_3 & 1-\varphi_4 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(p) \end{bmatrix}.$$

Desta última forma obtemos então $\rho(1), \rho(2), \dots, \rho(p)$; os valores de $\rho(h)$, $h \geq p+1$, são depois obtidos através da equação de recorrência.

Observação. O sistema de Yule-Walker permite calcular os coeficientes $\varphi_1, \varphi_2, \dots, \varphi_p$ em função de $\rho(1), \rho(2), \dots, \rho(p)$ o que é interessante para a fase de estimação. A ideia subjacente consiste, em linhas gerais, em calcular as autocorrelações empíricas (de um processo suposto estacionário e centrado) associadas a X_1, X_2, \dots, X_T ,

$$\hat{\rho}(h) = \frac{\frac{1}{T-h} \sum_{t=1}^{T-h} X_t X_{t+h}}{\frac{1}{T} \sum_{t=1}^T X_t^2}, h = 1, \dots, T-1,$$

e em aceitar que, quando T é elevado, as autocorrelações empíricas observadas não diferem muito dos correspondentes resumos teóricos, isto é,

$$\frac{\frac{1}{T-h} \sum_{t=1}^{T-h} x_t x_{t+h}}{\frac{1}{T} \sum_{t=1}^T x_t^2} \approx \rho(h).$$

Usando notações evidentes vem, do sistema de Yule-Walker,

$$\vec{\varphi} = [R(p-1)]^{-1} \vec{\rho}.$$

Assim, esperamos obter bons estimadores para $\varphi_1, \varphi_2, \dots, \varphi_p$ considerando

$$\widehat{\vec{\varphi}} = [\widehat{R}(p-1)]^{-1} \widehat{\vec{\rho}}. \blacksquare$$

Da última propriedade decorre o seguinte corolário:

Corolário. Se o polinómio Φ tem todas as suas raízes z_j , $j = 1, \dots, p$, distintas, as autocorrelações são da forma

$$\rho(h) = \sum_{j=1}^p \left(\frac{1}{z_j} \right)^h a_j, \quad a_j \in \mathbb{C},$$

onde a_1, \dots, a_p são determinados à custa das condições iniciais $\rho(1), \dots, \rho(p)$.

Prova. A sucessão das autocorrelações $\rho(h)$, $h \geq 1$, verifica uma equação de recorrência linear homogénea de ordem p cujos coeficientes estão directamente ligados aos dos polinómio Φ . O polinómio característico associado à equação de recorrência é

$$z^p - \varphi_1 z^{p-1} - \dots - \varphi_p z^0 = z^p \left(1 - \varphi_1 \frac{1}{z} - \dots - \varphi_p \frac{1}{z^p} \right) = z^p \Phi \left(\frac{1}{z} \right)$$

cujas raízes são $\frac{1}{z_j}$, $j = 1, \dots, p$.

A solução geral é, pois, da forma anunciada. \blacksquare

Notemos que, sendo $|z_j| > 1$, se tem $\left(\frac{1}{z_j}\right)^h \rightarrow 0$, quando $h \rightarrow +\infty$. Assim, se as raízes são distintas as autocorrelações tendem para zero de modo exponencial (com uma velocidade de convergência superior ou igual a $\left(\max \frac{1}{|z_j|}\right)^h = \left(\frac{1}{\min |z_j|}\right)^h$).

Se há raízes múltiplas, as autocorrelações aparecem como combinações das potências $\left(\frac{1}{z_j}\right)^h$ multiplicadas por polinômios em h :

$$\rho(h) = \sum \left(\frac{1}{z_j}\right)^h [a_j + b_j h + c_j h^2 + \dots].$$

Densidade espectral

Da equação auto-regressiva verificada pelo processo estocástico X , $\Phi(L)X_t = \varepsilon_t$, decorre a seguinte expressão para a sua densidade espectral:

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{1}{|\Phi(e^{i\omega})|^2}, \quad \omega \in [-\pi, \pi].$$

Exemplo. Seja $X_t = \varphi X_{t-1} + \varepsilon_t$, com $|\varphi| < 1$, $\forall t \in \mathbb{Z}$.

A densidade espectral de $X = (X_t, t \in \mathbb{Z})$ é

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi e^{i\omega}|^2} \\ &= \frac{\sigma^2}{2\pi} \frac{1}{|1 - \varphi \cos \omega - i \sin \omega|^2} \\ &= \frac{\sigma^2}{2\pi} \frac{1}{(1 - \varphi \cos \omega)^2 + \varphi^2 \sin^2 \omega} \\ &= \frac{\sigma^2}{2\pi} \frac{1}{1 + \varphi^2 - 2\varphi \cos \omega}. \end{aligned}$$

Sendo f_X uma função par, esboçemos o seu gráfico para $\omega \in [0, \pi]$. Temos, em particular,

$$f_X(0) = \frac{\sigma^2}{2\pi} \frac{1}{(1-\varphi)^2}, \quad f_X(\pi) = \frac{\sigma^2}{2\pi} \frac{1}{(1+\varphi)^2}.$$

Se $0 < \varphi < 1$, tem-se $f_X(\pi) < f_X(0)$ e a análise das derivadas de primeira e segunda ordem conduz-nos ao seguinte gráfico, onde está também esboçado o gráfico da densidade espectral quando $\varphi \rightarrow 1$.

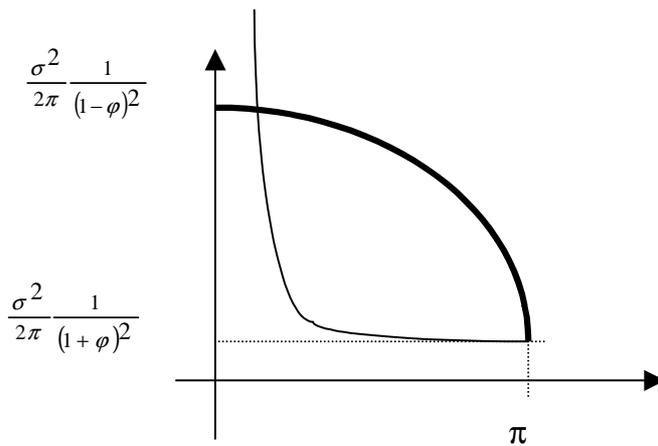


Fig. 3.1 - Densidade espectral para $0 < \varphi < 1$ e $\varphi \rightarrow 1$

Notemos que os valores de ω mais próximos de zero (mais pequenos) são os mais "carregados", o que se torna ainda mais acentuado quando $\varphi \rightarrow 1$.

Pequenos valores de ω correspondem a pequenas frequências que, por sua vez, correspondem a longos períodos. É este facto que leva à designação de *processos de memória longa* adoptada para uma classe mais geral do que a dos *AR* estacionários na qual se inclui a possibilidade de o "polinómio" auto-regressivo ter raízes de módulos iguais a 1.

Autocorrelações parciais

Relembremos que a autocorrelação parcial de ordem k , $r(k)$, é o coeficiente de X_{t-k} quando efectuamos a regressão linear de X_t sobre o espaço linear gerado por X_{t-1}, \dots, X_{t-k} .

Sendo $X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p} + \varepsilon_t$, $t \in \mathbb{Z}$, e uma vez que suposemos que estamos a trabalhar com a representação

canônica, tem-se que ε_t é ortogonal a \underline{H}_{t-1}^X e portanto

$$E_L(X_t/X_{t-1}, \dots, X_{t-k}, \dots) = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p}.$$

Então, para todo o $k \geq p$ tem-se

$$\begin{aligned} E_L(X_t/X_{t-1}, \dots, X_{t-k}) &= E_L(E_L(X_t/X_{t-1}, \dots, X_{t-k}, \dots)/X_{t-1}, \dots, X_{t-k}) \\ &= E_L(\varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}/X_{t-1}, \dots, X_{t-k}) \\ &= \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p}. \end{aligned}$$

Assim, considerando $k = p, p+1, \dots$, obtemos

$$\begin{aligned} r(p) &= \varphi_p \\ r(p+1) &= 0 \\ &\dots \\ r(k) &= 0, \text{ se } k > p. \end{aligned}$$

Podemos então enunciar o seguinte resultado:

Teorema. A sucessão das autocorrelações parciais de um processo $AR(p)$ anula-se a partir da ordem $p+1$.

A expressão de $r(k)$ para $k < p$ é obtida pelo algoritmo de Durbin.

3.2.4 Caracterização dos processos AR

Neste parágrafo vamos estudar condições necessárias e suficientes para que um processo estacionário admita uma representação auto-regressiva. Veremos que tais condições recaem sobre as sucessões das autocorrelações $(\rho(h), h \in \mathbb{Z})$ e das autocorrelações parciais $(r(h), h \in \mathbb{Z})$. Tais teoremas de caracterização permitir-nos-ão “reconhecer” os processos AR a partir daquelas sucessões.

Teorema. Uma condição necessária e suficiente para que um processo estacionário centrado admita uma representação AR é que a

sucessão das autocorrelações $(\rho(h), h \in \mathbb{Z})$ verifique uma equação de recorrência linear homogénea para $h \geq 1$.

Prova. A condição necessária foi estabelecida no parágrafo 3.2.3.

Quanto à condição suficiente, suponhamos que $\rho(h)$ é tal que

$$\rho(h) - \psi_1 \rho(h-1) - \dots - \psi_{p^*} \rho(h-p^*) = 0, \forall h \geq 1$$

onde ψ_i são reais quaisquer com $\psi_{p^*} \neq 0$.

Temos então, para todo $h \geq 1$,

$$\begin{aligned} Cov(X_t, X_{t-h}) - \psi_1 Cov(X_{t-1}, X_{t-h}) - \dots - \\ - \dots - \psi_{p^*} Cov(X_{t-p^*}, X_{t-h}) = 0, \end{aligned}$$

o que é equivalente a

$$Cov(X_t - \psi_1 X_{t-1} - \dots - \psi_{p^*} X_{t-p^*}, X_{t-h}) = 0, \forall h \geq 1$$

ou ainda

$$X_t - (\psi_1 X_{t-1} + \dots + \psi_{p^*} X_{t-p^*}) \text{ ortogonal a } X_{t-1}, X_{t-2}, \dots$$

Então

$$E_L(X_t - (\psi_1 X_{t-1} + \dots + \psi_{p^*} X_{t-p^*}) / \underline{H}_{t-1}^X) = 0,$$

ou seja

$$E_L(X_t / \underline{H}_{t-1}^X) = \psi_1 X_{t-1} + \dots + \psi_{p^*} X_{t-p^*}.$$

Introduzindo então a inovação de X no instante t

$$\varepsilon_t = X_t - E_L(X_t / \underline{H}_{t-1}^X)$$

tem-se, de facto,

$$X_t = \sum_{j=1}^{p^*} \psi_j X_{t-j} + \varepsilon_t$$

ou seja, X admite uma representação AR , uma vez que foi provado no primeiro capítulo que $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ assim definido é um ruído branco. ■

Se o processo X não é centrado tem-se o mesmo resultado para o processo centrado $(X_t - m, t \in \mathbb{Z})$, onde $m = E(X_t), t \in \mathbb{Z}$.

Observação. A ordem da equação de recorrência referida na propriedade anterior não é única.

Suponhamos, por exemplo, que $X_t - \varphi X_{t-1} = \varepsilon_t, \forall t \in \mathbb{Z}$.

Obtemos $\rho(h) - \varphi\rho(h-1) = 0$ o que nos conduz a $\rho(h) = \varphi^h, h \geq 0$.

Verificamos que uma das equações de recorrência é de ordem um:

$$\rho(h) - \varphi\rho(h-1) = 0.$$

Mas como $\rho(h-1) - \varphi\rho(h-2) = 0$, obtemos também

$$\rho(h) + (1 - \varphi)\rho(h-1) - \varphi\rho(h-2) = 0,$$

que é de ordem dois.

Ainda neste caso, notemos que

$$1 + (1 - \varphi)L - \varphi L^2 = (1 + L)(1 - \varphi L)$$

e desde que o factor $1 - \varphi L$ esteja presente os outros factores, bem como o seu número, são arbitrários. Assim podemos concluir que a ordem p da representação AR é a menor ordem das equações de recorrência lineares, homogéneas, verificadas pela sucessão $(\rho(h), h \in \mathbb{Z})$.

Teorema. Uma condição necessária e suficiente para que um processo estacionário centrado X admita uma representação AR é que as autocorrelações, $\rho(h), h \geq 0$, se escrevam como combinações de funções exponenciais de h , da forma

$$\rho(h) = \sum_{j=1}^{p^*} A_j \frac{1}{(z_j)^h}, A_j \in \mathbb{C}.$$

Prova. A condição necessária foi estabelecida no parágrafo 3.2.3.

Para estabelecer a condição suficiente, basta provar que as autocorrelações com esta forma verificam uma equação de recorrência

pois, pela propriedade anterior, poderemos concluir que X admite uma representação AR .

Podemos escrever

$$\begin{aligned} \rho(h) - \frac{1}{z_1} \rho(h-1) &= \sum_{j=1}^{p^*} A_j \frac{1}{(z_j)^h} - \frac{1}{z_1} \sum_{j=1}^{p^*} A_j \frac{1}{(z_j)^{h-1}} \\ &= A_1 \frac{1}{(z_1)^h} - A_1 \frac{1}{(z_1)^h} \\ &\quad + \sum_{j=2}^{p^*} A_j \left[\frac{1}{(z_j)^h} - \frac{1}{z_1} \frac{1}{(z_j)^{h-1}} \right] \\ &= \left(1 - \frac{1}{z_1} L \right) \sum_{j=2}^{p^*} A_j \frac{1}{(z_j)^h} \end{aligned}$$

onde, com notação simbólica, L actua sobre h .

Assim $1 - \frac{1}{z_1} L$ permite eliminar o termo correspondente a $j = 1$. Para eliminar os outros bastará aplicar, a ambos os membros, $\prod_{j=1}^{p^*} \left(1 - \frac{1}{z_j} L \right)$ obtendo-se

$$\left(1 - \frac{1}{z_1} L \right) \left(1 - \frac{1}{z_2} L \right) \dots \left(1 - \frac{1}{z_{p^*}} L \right) \rho(h) = 0,$$

ou seja, a sucessão de autocorrelações verifica uma equação de recorrência linear de ordem p^* . ■

Teorema. Uma condição necessária e suficiente para que um processo estacionário centrado admita uma representação AR é que a sucessão das autocorrelações parciais se anule a partir de uma certa ordem.

Prova. A condição necessária foi estabelecida no parágrafo 3.2.3.

Para provar a suficiência da condição começemos por verificar que

$$E_L (X_t / X_{t-1}, \dots, X_{t-h}) =$$

$$E_L(X_t/X_{t-1}, \dots, X_{t-h+1}, X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})).$$

A projecção está a ser feita sobre o mesmo espaço. De facto, seja $\underline{H}_{t-1}^{t-h}$ o espaço linear gerado por $X_{t-1}, \dots, X_{t-h+1}, X_{t-h}$.

Ora, o espaço linear gerado por

$$X_{t-1}, \dots, X_{t-h+1}, X_{t-h} - \sum_{j=1}^{h-1} \alpha_j X_{t-j}$$

é também $\underline{H}_{t-1}^{t-h}$.

Mas os elementos $X_{t-1}, \dots, X_{t-h+1}$ são ortogonais a $X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})$.

Como $Proj_{E \oplus F} X = Proj_E X + Proj_F X$, vem

$$E_L(X_t/X_{t-1}, \dots, X_{t-h}) = E_L(X_t/X_{t-1}, \dots, X_{t-h+1}) + E_L(X_t/(X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})))$$

A segunda parcela do lado direito pode escrever-se na forma

$$E_L(X_t/(X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1}))) = r(h) [X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})]$$

onde $r(h)$ é o coeficiente de regressão linear

$$r(h) = \frac{Cov(X_t, X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1}))}{V[X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})]}.$$

Obtivemos assim uma decomposição de $E_L(X_t/X_{t-1}, \dots, X_{t-h})$ em duas componentes ortogonais:

$$E_L(X_t/X_{t-1}, \dots, X_{t-h}) = E_L(X_t/X_{t-1}, \dots, X_{t-h+1}) + r(h) [X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})].$$

Sejam

$$\eta_h^2 = V[E_L(X_t/X_{t-1}, \dots, X_{t-h})]$$

$$\sigma_h^2 = V[X_{t-h} - E_L(X_{t-h}/X_{t-1}, \dots, X_{t-h+1})].$$

Atendendo à ortogonalidade das variáveis aleatórias da decomposição tem-se

$$\eta_h^2 = \eta_{h-1}^2 + [r(h)]^2 \sigma_h^2$$

ou ainda,

$$\begin{aligned} \eta_{h+1}^2 &= \eta_h^2 + [r(h+1)]^2 \sigma_{h+1}^2 \\ &= \eta_{h-1}^2 + [r(h)]^2 \sigma_h^2 + [r(h+1)]^2 \sigma_{h+1}^2 \\ &\quad \dots \\ \eta_{h+k}^2 &= \eta_{h+k-1}^2 + [r(h+k)]^2 \sigma_{h+k}^2 \\ &= \eta_{h+k-2}^2 + [r(h+k-1)]^2 \sigma_{h+k-1}^2 + [r(h+k)]^2 \sigma_{h+k}^2 \\ &= \dots \\ &= \eta_{h-1}^2 + [r(h)]^2 \sigma_h^2 + [r(h+1)]^2 \sigma_{h+1}^2 + \dots + [r(h+k)]^2 \sigma_{h+k}^2. \end{aligned}$$

Da hipótese, existe $r \in \mathbb{N}$ tal que $\forall h > r, r(h) = 0$.

Então

$$\forall h > r, \quad \eta_h^2 = \eta_r^2.$$

Ora $E_L(X_t/X_{t-1}, \dots, X_{t-h})$ tende em L_2 , quando $h \rightarrow +\infty$, para $E_L(X_t/X_{t-1}, \dots) = E_L(X_t/\underline{H}_{t-1}^X)$.

Logo,

$$\eta_h^2 \rightarrow V[E_L(X_t/\underline{H}_{t-1}^X)], \quad \text{quando } h \rightarrow +\infty.$$

Sendo η_h^2 constante a partir duma certa ordem e igual a η_r^2 tem-se

$$\begin{aligned} \eta_r^2 &= V[E_L(X_t/X_{t-1}, \dots, X_{t-r})] \\ &= V[E_L(X_t/\underline{H}_{t-1}^X)] = \eta^2. \end{aligned} \quad (3.2.4.1)$$

Ora, o espaço linear gerado por X_{t-1}, \dots, X_{t-r} está contido em \underline{H}_{t-1}^X . Então, devemos ter

$$E_L(X_t/\underline{H}_{t-1}^X) = E_L(X_t/X_{t-1}, \dots, X_{t-r}) + Z \quad (3.2.4.2)$$

com Z pertencente ao ortogonal do espaço linear gerado por X_{t-1}, \dots, X_{t-r} , e então

$$V[E_L(X_t/\underline{H}_{t-1}^X)] = V[E_L(X_t/X_{t-1}, \dots, X_{t-r})] + V(Z).$$

Concluimos então de (3.2.4.1) que $V(Z) = 0$ o que implica $Z = 0$ quase certamente e, conseqüentemente, de (3.2.4.2),

$$E_L(X_t/\underline{H}_{t-1}^X) = E_L(X_t/X_{t-1}, \dots, X_{t-r}), q.c.,$$

ou seja,

$$X_t - E_L(X_t/X_{t-1}, \dots, X_{t-r}) = X_t - E_L(X_t/\underline{H}_{t-1}^X).$$

Reconhecemos no segundo membro a componente de ordem t de um ruído branco e então

$$X_t - E_L(X_t/X_{t-1}, \dots, X_{t-r}) = \varepsilon_t$$

com ε_t ortogonal a \underline{H}_{t-1}^X .

Assim, X admite uma representação AR de ordem r . ■

3.3 Processos médias móveis

3.3.1 Definição

Definição. Um processo de segunda ordem $X = (X_t, t \in \mathbb{Z})$ admite uma *representação média móvel de ordem q* , de modo abreviado $MA(q)$, quando

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}, \quad \forall t \in \mathbb{Z},$$

onde $\theta_1, \dots, \theta_q \in \mathbb{R}$, com $\theta_q \neq 0$, e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância $\sigma^2 > 0$.

Considerando o operador atraso L associado ao processo ε , podemos escrever a igualdade anterior na forma

$$X_t = \Theta(L)\varepsilon_t$$

com $\Theta(L) = 1 - \theta_1L - \theta_2L^2 - \dots - \theta_qL^q$.

Notemos que o processo X é, dados ε e $\theta_1, \dots, \theta_q$, definido sem ambigüidade, isto é, é único. Por outro lado, sendo uma média móvel finita, é estacionário e regular.

3.3.2 Multiplicidade de representações

Tal como no caso dos modelos AR também os modelos médias móveis admitem várias representações. De facto, tem-se o resultado seguinte cuja prova segue os mesmos passos da proposição correspondente estabelecida para os modelos AR .

Teorema. Seja X um processo $MA(q)$ definido por $X_t = \Theta(L)\varepsilon_t$, $\forall t \in \mathbb{Z}$. Este processo admite outras representações média móvel. Tais representações obtêm-se a partir de uma qualquer delas substituindo uma (ou várias) raízes pelas suas inversas e mantendo as restantes. Em particular,

a) Todas as representações MA do processo X têm a mesma ordem q .

b) Existe uma representação MA na qual todas as raízes de Θ são de módulo superior ou igual a 1.

3.3.3 Representação AR de um processo $MA(q)$

A obtenção de uma forma AR infinita para um processo $MA(q)$, referida por certos autores como ‘inversibilidade do processo’, tem-se nas condições do teorema seguinte.

Teorema. Seja X um processo $MA(q)$ definido por $X_t = \Theta(L)\varepsilon_t$, $\forall t \in \mathbb{Z}$, onde Θ é um polinómio de grau q cujas raízes são todas de módulos estritamente superiores a 1. Nestas condições podemos escrever

$$\varepsilon_t = X_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \dots$$

isto é, podemos escrever X sob forma $AR(+\infty)$

$$\Pi(L)X_t = \varepsilon_t$$

com $\Pi(L) = 1 + \sum_{j=1}^{+\infty} \pi_j L^j = \frac{1}{\Theta(L)}$.

Além disso ε é o processo inovação.

Prova. Sendo todas as raízes de Θ de módulos estritamente superiores a 1, $\Theta(L)$ é invertível e $[\Theta(L)]^{-1}$ é uma série em L ,

$$\prod_{j=1}^q \left(\sum_{i=0}^{+\infty} \left(\frac{1}{z_j} \right)^i L^i \right) = \sum_{j=0}^{+\infty} \phi_j L^j,$$

com $\phi_0 = 1$. Então

$$\varepsilon_t = [\Theta(L)]^{-1} X_t = \left(\sum_{j=0}^{+\infty} \phi_j L^j \right) X_t = X_t + \sum_{j=1}^{+\infty} \pi_j X_{t-j}$$

com $\phi_j = \pi_j$.

Mostremos agora que ε é a inovação.

Da forma $AR(+\infty)$ concluímos que $\varepsilon_t \in \underline{H}_t^X$. Da forma $MA(q)$ concluímos que $X_t \in \underline{H}_t^\varepsilon$.

Consequentemente $\underline{H}_t^X = \underline{H}_t^\varepsilon$, $\forall t \in \mathbb{Z}$.

Então

$$\begin{aligned} E_L(X_t / \underline{H}_{t-1}^X) &= E_L(X_t / \underline{H}_{t-1}^\varepsilon) \\ &= E_L(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} / \underline{H}_{t-1}^\varepsilon) \\ &= -\theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \\ &= X_t - \varepsilon_t. \blacksquare \end{aligned}$$

Note-se que, sob as condições deste teorema, a representação é canónica. Também aqui se pode obter a representação canónica para todo o processo $MA(q)$ cujo polinómio média móvel tem todas as suas raízes de módulos diferentes de 1.

O exemplo seguinte ilustra a situação em que pelo menos uma das raízes do polinómio média móvel é de módulo 1. Como veremos, o ruído interveniente na representação é a inovação e a representação é canónica mas o processo X não é invertível.

Exemplo. Consideremos o processo $MA(1)$

$$X_t = \varepsilon_t - \varepsilon_{t-1}, \quad t \in \mathbb{Z},$$

com $(\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância $\sigma^2 > 0$.

Recorrendo ao operador L , tem-se $X_t = (1 - L)\varepsilon_t$.

Mas $1 - L$ não é invertível (Gouriéroux, Monfort, 1990, p.174). Podemos, no entanto, determinar directamente a inovação no instante t .

Do parágrafo inicial deste capítulo, sabemos que

$$E_L(X_t / \underline{H}_{t-1}^X) = \lim_{k \rightarrow +\infty} E_L(X_t / X_{t-1}, \dots, X_{t-k}).$$

Por outro lado,

$$E_L(X_t / X_{t-1}, \dots, X_{t-k}) = \sum_{r=1}^k a_r(k) X_{t-r}$$

com $a_1(k), \dots, a_k(k)$ verificando, de acordo com o parágrafo 2.3, o sistema

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(k-2) \\ \dots & \dots & \dots & \dots \\ \gamma(k-1) & \gamma(k-2) & \dots & \gamma(0) \end{bmatrix} \begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \dots \\ \gamma(k) \end{bmatrix}.$$

Mas

$$\gamma(h) = Cov(X_t, X_{t+h}) = E(X_t X_{t+h}) = \begin{cases} 2\sigma^2, & h = 0 \\ -\sigma^2, & h = -1, 1 \\ 0, & h \in \mathbb{Z} \setminus \{-1, 0, 1\} \end{cases}.$$

O sistema anterior é, assim, equivalente ao seguinte

$$\begin{bmatrix} 2 & -1 & \dots & 0 & 0 \\ -1 & 2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix} \begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

ou ainda, equivalente a

$$\begin{cases} 2a_1(k) - a_2(k) = -1 \\ -a_{j-1}(k) + 2a_j(k) - a_{j+1}(k) = 0, \quad j = 2, \dots, k-1 \\ -a_{k-1}(k) + 2a_k(k) = 0 \end{cases}.$$

Vemos que os coeficientes $a_j(k)$, $j = 2, \dots, k-1$, verificam a equação de recorrência linear homogênea

$$a_{j-1}(k) - 2a_j(k) + a_{j+1}(k) = 0,$$

cuja equação característica, $z^2 - 2z + 1 = 0$, tem $z = 1$ como raiz dupla. Então

$$a_j(k) = A + Bj, \quad j = 2, \dots, k-1.$$

Para determinar $A = A(k)$ e $B = B(k)$ (*i.e.*, eventualmente dependentes de k), utilizamos a primeira e a última equação do sistema. Assim, de

$$\begin{cases} 2(A+B) - (A+2B) = -1 \\ -[A+B(k-1)] + 2[A+Bk] = 0 \end{cases}$$

obtemos $A = -1$, $B = \frac{1}{k+1}$. Então,

$$a_j(k) = -1 + \frac{1}{k+1}j = \frac{-k-1+j}{k+1}, \quad j = 1, \dots, k.$$

Consequentemente,

$$E_L(X_t/X_{t-1}, \dots, X_{t-k}) = -\sum_{r=1}^k \frac{k+1-r}{k+1} X_{t-r}$$

ou seja, em função do ruído, temos

$$\begin{aligned} E_L(X_t/X_{t-1}, \dots, X_{t-k}) &= -\sum_{r=1}^k \frac{k+1-r}{k+1} (\varepsilon_{t-r} - \varepsilon_{t-r-1}) \\ &= -\frac{k}{k+1} \varepsilon_{t-1} + \frac{1}{k+1} \sum_{r=2}^{k+1} \varepsilon_{t-r} \\ &= -\varepsilon_{t-1} + \frac{1}{k+1} \sum_{r=1}^{k+1} \varepsilon_{t-r}. \end{aligned}$$

Então

$$X_t - E_L(X_t/X_{t-1}, \dots, X_{t-k}) = \varepsilon_t - \frac{1}{k+1} \sum_{r=1}^{k+1} \varepsilon_{t-r}.$$

A última parcela, tendo em conta a não correlação das variáveis $(\varepsilon_t, t \in \mathbb{Z})$, é tal que

$$E \left(\frac{1}{k+1} \sum_{r=1}^{k+1} \varepsilon_{t-r} \right)^2 = \left(\frac{1}{k+1} \right)^2 \sum_{r=1}^{k+1} E [(\varepsilon_{t-r})^2] = \frac{1}{k+1} \sigma^2 \longrightarrow 0,$$

quando k tende para $+\infty$.

Logo $-\frac{1}{k+1} \sum_{r=1}^{k+1} \varepsilon_{t-r}$ tende em L_2 para 0, quando $k \rightarrow +\infty$.

Como $E_L(X_t/X_{t-1}, \dots, X_{t-k})$ tende em L_2 para $E_L(X_t/\underline{H}_{t-1}^X)$, tem-se então que

$$\varepsilon_t = X_t - E_L(X_t/\underline{H}_{t-1}^X),$$

o que implica que ε_t é, de facto, a inovação na data t .

Apesar de escrita em função dos valores presente e passados de X , não estamos perante uma representação auto-regressiva infinita (a igualdade $\varepsilon_t = X_t + \lim_k \sum_{r=1}^k \frac{k+1-r}{k+1} X_{t-r}$ não vale em L_2 e os coeficientes dependem de k). Assim o processo X é estacionário, mas não é invertível. ■

3.3.4 Resumo das ligações temporais

Vamos agora estudar os resumos das ligações temporais de um processo X com representação $MA(q)$

$$X_t = \Theta(L) \varepsilon_t, \forall t \in \mathbb{Z}$$

onde supomos que as raízes do polinómio Θ são de módulos estritamente superiores a 1.

Funções de autocovariância e de autocorrelação

Calculemos a autocovariância de ordem h , $\gamma(h) = E(X_t X_{t+h})$, para $h \geq 0$.

$$\begin{aligned} & E(X_t X_{t+h}) \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q})(\varepsilon_{t+h} - \theta_1 \varepsilon_{t+h-1} - \dots - \theta_q \varepsilon_{t+h-q})] \\ &= \begin{cases} (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2, & h = 0 \\ (-\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q) \sigma^2, & 0 < h \leq q \\ 0, & h > q \end{cases} \end{aligned}$$

Podemos resumir esta forma à seguinte onde se toma, por convenção, $\theta_0 = -1$:

$$\gamma(h) = \begin{cases} (\theta_0 \theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q) \sigma^2, & 0 \leq h \leq q \\ 0, & h > q \end{cases}.$$

Então, a função de autocovariância de X é

$$\gamma(h) = \begin{cases} \sigma^2 \sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}, & h = -q, -q+1, \dots, q-1, q \\ 0, & \text{caso contrário} \end{cases},$$

considerando $\theta_0 = -1$.

A função de autocorrelação de X deduz-se imediatamente:

$$\rho(h) = \begin{cases} \frac{\sum_{j=0}^{q-|h|} \theta_j \theta_{j+|h|}}{\sum_{j=0}^q \theta_j^2}, & h = -q, -q+1, \dots, q-1, q, \theta_0 = -1. \\ 0, & \text{caso contrário} \end{cases}$$

Notemos que a escolha de um modelo $MA(1)$ não é compatível com a existência de fortes correlações à ordem 1. De facto, para um processo definido por

$$X_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

a autocorrelação de ordem 1 é $\rho(1) = -\frac{\theta}{1+\theta^2}$ que, como já vimos, é, para $|\theta| < 1$, em módulo menor do que $\frac{1}{2}$.

Densidade espectral

Sendo

$$\forall t \in \mathbb{Z}, X_t = \Theta(L) \varepsilon_t = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

a densidade espectral de X é dada por

$$f_X(\omega) = \frac{\sigma^2}{2\pi} |\Theta(e^{i\omega})|^2, \omega \in [-\pi, \pi[.$$

Coefficientes de regressão e autocorrelação parcial

Seja $X = (X_t, t \in \mathbb{Z})$ um processo $MA(q)$ verificando $X_t = \Theta(L) \varepsilon_t$ e $(\rho(h), h \in \mathbb{Z})$ a sua função de autocorrelação. Determinemos os coeficientes de regressão $a_j(k), j = 1, \dots, k$.

Sabemos que os coeficientes de regressão verificam

$$V \left(\begin{bmatrix} X_{t-1} \\ \dots \\ X_{t-k} \end{bmatrix} \right) \vec{a}_k = \begin{bmatrix} \gamma(1) \\ \dots \\ \gamma(k) \end{bmatrix}$$

com $\vec{a}_k = \begin{bmatrix} a_1(k) \\ \dots \\ a_k(k) \end{bmatrix}$, ou seja,

$$\begin{bmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ \dots & \dots & \dots & \dots \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{bmatrix} \vec{a}_k = \begin{bmatrix} \rho(1) \\ \dots \\ \dots \\ \rho(k) \end{bmatrix}.$$

Como $\rho(j) = 0$ se $j > q$, obtemos, se $k > q$, um sistema com ‘bandas nulas’, designadamente,

$$\begin{bmatrix} 1 & \rho(1) & \dots & \rho(q) & 0 & \dots & 0 & 0 \\ \rho(1) & 1 & \dots & \rho(q-1) & \rho(q) & \dots & 0 & 0 \\ \dots & \dots \\ \rho(q) & \dots & \dots & 1 & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & 1 & \rho(1) \\ 0 & \dots & \dots & \dots & \dots & \dots & \rho(1) & 1 \end{bmatrix} \vec{a}_k = \begin{bmatrix} \rho(1) \\ \dots \\ \rho(q) \\ 0 \\ \dots \\ \dots \\ \dots \\ 0 \end{bmatrix}.$$

Analisando a equação de ordem j , para $j \geq q + 1$ e $j \leq k - q$ (caso em que aparecem todos os termos $\rho(1), \dots, \rho(q)$ após $\rho(0) = 1$), obtemos:

$$\rho(q) a_{j-q}(k) + \dots + \rho(1) a_{j-1}(k) + a_j(k) + \rho(1) a_{j+1}(k) + \dots + \rho(q) a_{j+q}(k) = 0.$$

Estabelecemos, pois, o resultado seguinte:

Propriedade. Os coeficientes $(a_j(k), j = 1, \dots, k)$ verificam uma equação de recorrência linear homogênea de ordem $2q$ cujo polinómio característico é

$$\frac{z^q}{\gamma(0)} \sigma^2 \Theta(z) \Theta\left(\frac{1}{z}\right).$$

Prova. De facto, tem-se

$$\begin{aligned} \rho(q) + \rho(q-1)z + \dots + \rho(1)z^{q-1} + z^q + \rho(1)z^{q+1} + \dots + \rho(q)z^{2q} \\ = z^q [\rho(q)z^{-q} + \rho(q-1)z^{-q+1} + \dots + \rho(q)z^q] \\ = \frac{z^q}{\gamma(0)} [\gamma(q)z^{-q} + \dots + \gamma(q)z^q] \\ = \frac{z^q}{\gamma(0)} \sigma^2 \Theta(z) \Theta\left(\frac{1}{z}\right). \end{aligned}$$

Vemos que as raízes do polinómio característico são as q raízes do polinómio Θ , $z_l, l = 1, \dots, q$, bem como os respectivos inversos $\frac{1}{z_l}, l = 1, \dots, q$. Supondo que são distintas teremos

$$a_j(k) = \sum_{l=1}^q (z_l)^j A_l(k) + \sum_{l=1}^q \left(\frac{1}{z_l}\right)^j B_l(k).$$

As $2q$ incógnitas $A_l(k)$ e $B_l(k)$ ($l = 1, \dots, q$) são obtidas à custa das condições iniciais dadas pelas equações para $j = 1, \dots, q$ e para

$j = k - q + 1, \dots, k$, as quais conduzem a um sistema com $2q$ equações a $2q$ incógnitas.

Exemplo. Consideremos um processo $MA(1)$

$$X_t = \varepsilon_t - \theta\varepsilon_{t-1},$$

com ε um ruído branco de variância σ^2 e θ um real de módulo menor que 1.

Os coeficientes de regressão de X_t sobre o espaço linear gerado por X_{t-1}, \dots, X_{t-k} são

$$a_j(k) = -\frac{1}{1 - \theta^{2k+2}}\theta^j + \frac{\theta^{2k+2}}{(1 - \theta^{2k+2})\theta^j}.$$

De facto, tais coeficientes de regressão verificam o sistema

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \dots \\ \rho(k) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \dots & \rho(k-1) \\ \rho(1) & 1 & \dots & \rho(k-2) \\ \dots & \dots & \dots & \dots \\ \rho(k-1) & \rho(k-2) & \dots & 1 \end{bmatrix} \begin{bmatrix} a_1(k) \\ a_2(k) \\ \dots \\ a_k(k) \end{bmatrix}.$$

Ora $\rho(h)$ é apenas diferente de zero para $h = 0, 1, -1$, tendo-se

$$\gamma(h) = \begin{cases} (1 + \theta^2)\sigma^2, & h = 0 \\ -\theta\sigma^2, & h = -1, +1 \end{cases}$$

e conseqüentemente

$$\rho(h) = \begin{cases} 1, & h = 0 \\ -\frac{\theta}{1+\theta^2}, & h = -1, +1 \end{cases}.$$

Então, para $j = 2, \dots, k-1$, tem-se

$$-\frac{\theta}{1+\theta^2}a_{j-1}(k) + a_j(k) - \frac{\theta}{1+\theta^2}a_{j+1}(k) = 0$$

cuja equação característica é

$$\theta z^2 - (1 + \theta^2)z + \theta = 0$$

de raízes $z = \theta$, $z = \frac{1}{\theta}$, distintas pois $|\theta| < 1$.

Então

$$a_j(k) = A\theta^j + B \left(\frac{1}{\theta}\right)^j.$$

Utilizando a primeira e a última equação do sistema como condições iniciais,

$$a_1(k) - \frac{\theta}{1+\theta^2}a_2(k) = -\frac{\theta}{1+\theta^2} \quad \text{e} \quad -\frac{\theta}{1+\theta^2}a_{k-1}(k) + a_k(k) = 0,$$

obtemos o sistema

$$\begin{cases} A\theta + B\frac{1}{\theta} - \frac{\theta}{1+\theta^2} \left[A\theta^2 + B\left(\frac{1}{\theta}\right)^2 \right] = -\frac{\theta}{1+\theta^2} \\ -\frac{\theta}{1+\theta^2} \left[A\theta^{k-1} + B\left(\frac{1}{\theta}\right)^{k-1} \right] + \left[A\theta^k + B\left(\frac{1}{\theta}\right)^k \right] = 0 \end{cases}$$

de onde retiramos os valores

$$A = -\frac{1}{1-\theta^{2k+2}}, B = \frac{\theta^{2k+2}}{1-\theta^{2k+2}}. \blacksquare$$

3.3.5 Caracterização dos processos MA

Teorema. Um processo estacionário regular admite uma representação MA de ordem finita se e só se a sucessão das autocovariâncias se anula a partir de uma certa ordem. Se $q+1$ designa esta ordem mínima, q é a ordem da representação MA .

Prova. Já provamos que se um processo admite uma representação $MA(q)$ então $\gamma(h) = 0$, $h \geq q+1$.

Inversamente, seja H a ordem a partir da qual a sucessão das autocorrelações se anula.

Se $\rho(h) = 0$, para $h \geq H+1$, então X_t é ortogonal a X_{t-H-1} , X_{t-H-2} , ..., isto é, X_t é ortogonal a \underline{H}_{t-H-1}^X .

Vimos anteriormente que $\underline{H}_{t-H-1}^X = \underline{H}_{t-H-1}^\varepsilon$ uma vez que o processo X é regular.

Ainda da regularidade de X tem-se

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j},$$

pelo que

$$E_L (X_t / \underline{H}^X_{t-H-1}) = E_L (X_t / \underline{H}^\varepsilon_{t-H-1}) = \sum_{j=H+1}^{+\infty} a_j \varepsilon_{t-j} = 0$$

pela ortogonalidade existente. Isto permite concluir que

$$\forall j \geq H + 1, \quad a_j = 0.$$

Concluimos que o processo admite uma representação *MA* de ordem finita. ■

3.4 Processos auto-regressivos médias móveis

3.4.1 Definição

A modelação auto-regressiva média móvel, abreviadamente *ARMA*, generaliza simultaneamente os modelos auto-regressivos puros e as médias móveis puras que estudamos anteriormente. Esta modelação apresenta a vantagem de conduzir frequentemente a boas aproximações das séries reais com um número de parâmetros inferior aos dos modelos puros.

Os processos correspondentes a tal modelação, que iremos abordar neste parágrafo, são os processos $X = (X_t, t \in \mathbb{Z})$ verificando uma equação da forma

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

ou, de modo equivalente,

$$\Phi(L) X_t = \Theta(L) \varepsilon_t$$

com $\Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância não nula.

Definição. Um processo estacionário $X = (X_t, t \in \mathbb{Z})$ admite uma representação *ARMA*(p, q) *mínima* se verifica

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

onde

- i) $\varphi_p \neq 0, \theta_q \neq 0$,
- ii) os polinómios Φ e Θ não têm raízes em comum,
- iii) os polinómios Φ e Θ têm as suas raízes de módulos estritamente superiores a 1.

A definição anterior impõe algumas restrições aos polinómios Φ e Θ . A sua necessidade tornar-se-á mais evidente se analisarmos as consequências da não observância de tais restrições sobre uma formulação *ARMA*.

Em primeiro lugar nota-se que tais restrições não são necessárias à existência de processos solução. De facto, dados Φ , Θ e ε há sempre processos solução, pois bastará considerar p condições iniciais.

Por outro lado, uma tal representação admite uma solução estacionária quando Φ tem todas as suas raízes de módulos diferentes de 1. De facto, neste caso Φ é invertível e esta solução é necessariamente da forma

$$X_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t,$$

isto é, X_t pode escrever-se sob forma *MA*($+\infty$) na qual intervêm valores passados e futuros do ruído.

No entanto, a condição anterior para a existência de solução estacionária é apenas suficiente; se considerarmos, por exemplo,

$$X_t - X_{t-1} = \varepsilon_t - \varepsilon_{t-1} \Leftrightarrow (1 - L) X_t = (1 - L) \varepsilon_t,$$

vemos que há uma solução estacionária única, $X_t = \varepsilon_t$.

Ou seja, pode haver na parte *AR* raízes de módulo 1 e, no entanto, haver soluções estacionárias. Isto acontece se as raízes de módulo 1 de Φ forem também raízes de Θ .

Outro aspecto a ter em conta tem a ver com a existência de raízes comuns aos dois polinómios pois poder-se-á questionar a

possibilidade de as eliminar. De facto, se X é estacionário com representação $MA(+\infty)$ e tal que $\Phi(L)X_t = \Theta(L)\varepsilon_t$ então

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta(e^{i\omega})}{\Phi(e^{i\omega})} \right|^2, \quad \omega \in [-\pi, \pi[.$$

Se Φ e Θ têm raízes comuns, teremos

$$\Theta(e^{i\omega}) = \Theta'(e^{i\omega})H(e^{i\omega}), \quad \Phi(e^{i\omega}) = \Phi'(e^{i\omega})H(e^{i\omega})$$

e então

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta'(e^{i\omega})}{\Phi'(e^{i\omega})} \right|^2.$$

Constatamos que a presença de raízes comuns conduz a uma multiplicidade de representações. De facto, se Φ^* e Θ^* são polinómios sem raízes comuns e se $\Phi^*(L)X_t = \Theta^*(L)\varepsilon_t$ então, para qualquer polinómio H ,

$$H(L)\Phi^*(L)X_t = H(L)\Theta^*(L)\varepsilon_t.$$

Entre as várias representações será natural reter uma para a qual a densidade espectral tenha a forma de fracção irredutível (em $e^{i\omega}$), ou seja, tal que Φ e Θ não tenham raízes comuns ou, de modo equivalente, tal que as ordens p e q sejam tão pequenas quanto possível. Isto explica a terminologia ‘representação mínima’ presente na definição.

Esta causa de multiplicidade não está presente nos modelos puros (AR ou MA) de ordem finita, porque se $\Phi^*(L)X_t = \varepsilon_t$, por exemplo, a multiplicação por $H(L)$ levar-nos-ia a sair da classe dos AR puros.

Note-se finalmente que uma representação onde Φ e Θ não tenham raízes comuns, admite uma solução estacionária se e só se as raízes de Φ são em módulo diferentes de 1.

Relativamente à existência de raízes de módulos superiores a um para os polinómios Φ e Θ , comecemos por observar que se tal

acontecer para o polinómio Φ (de grau mínimo) então

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j},$$

isto é, apenas os valores presente e passados do ruído intervêm nesta representação.

Claro que podemos sempre colocar-nos nesta situação, substituindo algumas das raízes de Φ pelas suas inversas.

Por outro lado, se o processo estacionário X verifica $\Phi(L)X_t = \Theta(L)\varepsilon_t$, não tendo Φ e Θ raízes comuns, ele admite também outras representações (baseadas noutros ruídos)

$$\Phi(L)X_t = \tilde{\Theta}(L)\tilde{\varepsilon}_t$$

onde $\tilde{\Theta}$ se obtém a partir de Θ substituindo algumas das raízes pelas suas inversas. Em particular, podemos sempre colocar-nos na situação em que as raízes de Θ são de módulo superior ou igual a 1. Se exigirmos, adicionalmente, que Θ não tenha raízes de módulo 1, então Θ é invertível e temos

$$\varepsilon_t = \frac{\Phi(L)}{\Theta(L)}X_t = \sum_{j=0}^{+\infty} \pi_j X_{t-j}$$

ou seja,

- X_t admite uma representação $AR(+\infty)$
- ε_t interpreta-se como a inovação de X no instante t .

Assim, sob as condições da representação mínima presentes na definição anterior

- há uma e uma só solução estacionária;
- a representação mínima é única;
- a densidade espectral do processo é função dos polinómios da representação

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta(e^{i\omega})}{\Phi(e^{i\omega})} \right|^2 \quad \omega \in [-\pi, \pi[;$$

3.4. PROCESSOS AUTO-REGRESSIVOS MÉDIAS MÓVEIS 141

- o processo admite uma representação $MA(+\infty)$, $X_t = \frac{\Theta(L)}{\Phi(L)}\varepsilon_t$, e, em particular, X é regular;
- o processo admite uma representação $AR(+\infty)$, $\varepsilon_t = \frac{\Phi(L)}{\Theta(L)}X_t$, e ε é o processo inováção.

Exemplo. Consideremos um modelo $ARMA(1, 1)$

$$(1 - \varphi L) X_t = (1 - \theta L) \varepsilon_t.$$

Supondo que se trata da representação mínima, determinemos as representações $MA(+\infty)$ e $AR(+\infty)$ de $(X_t, t \in \mathbb{Z})$.

Se a representação considerada é a mínima então $\varphi \neq 0$, $\theta \neq 0$, $\varphi \neq \theta$, $|\varphi| < 1$ e $|\theta| < 1$.

A representação $MA(+\infty)$ de $(X_t, t \in \mathbb{Z})$ é então

$$\begin{aligned} X_t &= \frac{1}{1 - \varphi L} (1 - \theta L) \varepsilon_t = \sum_{j=0}^{+\infty} \varphi^j L^j (1 - \theta L) \varepsilon_t \\ &= \sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j} - \theta \sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j-1} = \varepsilon_t + \sum_{j=1}^{+\infty} (\varphi - \theta) \varphi^{j-1} \varepsilon_{t-j}. \end{aligned}$$

Quanto à representação $AR(+\infty)$ de $(X_t, t \in \mathbb{Z})$ tem-se

$$\varepsilon_t = \frac{1 - \varphi L}{1 - \theta L} X_t = X_t + \sum_{j=1}^{+\infty} (\theta - \varphi) \theta^{j-1} X_{t-j}.$$

Este exemplo permite ainda ilustrar a dificuldade da fase de identificação de um modelo $ARMA$, isto é, a fase da escolha das ordens p e q . De facto, os valores dos parâmetros do modelo $ARMA(1, 1)$ podem ser tais que ele seja bem aproximado por um AR ou MA eventualmente com menos parâmetros:

- a) se $\varphi = 0.001$ e $\theta = 0.3$, a representação $MA(+\infty)$ de X_t é

$$X_t = \varepsilon_t - 0.299\varepsilon_{t-1} - 0.299(0.0001)\varepsilon_{t-2} - \dots$$

a qual quase não se distingue, em termos práticos, da representação $MA(1)$ $X_t = \varepsilon_t - 0.299\varepsilon_{t-1}$.

b) se $\varphi = 0.5$ e $\theta = 0.45$ obtemos, por exemplo da representação $AR(+\infty)$ de X_t ,

$$\varepsilon_t = X_t - 0.05X_{t-1} + \dots$$

ou seja, a representação $ARMA(1, 1)$ é praticamente equivalente a uma representação $AR(1)$ ou mesmo $MA(0)$. ■

Exemplo. Consideremos um processo estocástico real $(X_t, t \in \mathbb{Z})$ de segunda ordem verificando a seguinte equação de recorrência:

$$X_t = \frac{1}{2}X_{t-1} + \varepsilon_t - \frac{1}{4}\varepsilon_{t-1}, \quad \forall t \in \mathbb{Z}$$

onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância σ^2 estritamente positiva.

Mostremos que se X é fracamente estacionário, então X é o processo definido, no sentido da convergência em média quadrática, por

$$X_t = \varepsilon_t + \sum_{j=1}^{+\infty} \left(\frac{1}{2}\right)^{j+1} \varepsilon_{t-j}, \quad \forall t \in \mathbb{Z}.$$

De facto, podemos escrever

$$\begin{aligned} X_t &= \frac{1}{2}X_{t-1} + \varepsilon_t - \frac{1}{4}\varepsilon_{t-1} \\ &= \frac{1}{2} \left(\frac{1}{2}X_{t-2} + \varepsilon_{t-1} - \frac{1}{4}\varepsilon_{t-2} \right) + \varepsilon_t - \frac{1}{4}\varepsilon_{t-1} \\ &= \frac{1}{4}X_{t-2} + \varepsilon_t + \frac{1}{4}\varepsilon_{t-1} - \frac{1}{8}\varepsilon_{t-2} \\ &= \dots \end{aligned}$$

ou seja, como facilmente se prova por recorrência,

$$X_t = \frac{1}{2^j}X_{t-j} + \varepsilon_t + \sum_{i=1}^{j-1} \frac{1}{2^{i+1}}\varepsilon_{t-i} - \frac{1}{2^{j+1}}\varepsilon_{t-j}.$$

Provemos que, quando $j \rightarrow +\infty$,

$$\left\| \frac{1}{2^j} X_{t-j} + \varepsilon_t + \sum_{i=1}^{j-1} \frac{1}{2^{i+1}} \varepsilon_{t-i} - \frac{1}{2^{j+1}} \varepsilon_{t-j} - \left(\varepsilon_t + \sum_{i=1}^{+\infty} \left(\frac{1}{2}\right)^{i+1} \varepsilon_{t-i} \right) \right\|$$

tende para zero.

Ora,

$$\begin{aligned} & \lim_{j \rightarrow +\infty} \left\| \frac{1}{2^j} X_{t-j} - \sum_{i=j}^{+\infty} \frac{1}{2^{i+1}} \varepsilon_{t-i} - \frac{1}{2^{j+1}} \varepsilon_{t-j} \right\| \\ & \leq \lim_{j \rightarrow +\infty} \left[\frac{1}{2^j} \|X_{t-j}\| + \sum_{i=j}^{+\infty} \frac{1}{2^{i+1}} \|\varepsilon_{t-i}\| + \frac{1}{2^{j+1}} \|\varepsilon_{t-j}\| \right] \\ & = \lim_{j \rightarrow +\infty} \left[\frac{1}{2^j} C + \sigma \sum_{i=j}^{+\infty} \frac{1}{2^{i+1}} + \sigma \frac{1}{2^{j+1}} \right] = 0, \end{aligned}$$

uma vez que da estacionaridade fraca de ε e X decorre

$$\|X_{t-j}\| = \left[E \left(X_{t-j}^2 \right) \right]^{\frac{1}{2}} = \sqrt{\gamma(0)} \text{ (constante), e } \|\varepsilon_{t-i}\| = \sigma, \forall i.$$

A unicidade do limite em L^2 permite então escrever

$$X_t \stackrel{L^2}{=} \varepsilon_t + \sum_{j=1}^{+\infty} \left(\frac{1}{2}\right)^{j+1} \varepsilon_{t-j}. \blacksquare$$

3.4.2 Resumo das ligações temporais

Função de autocovariância

Teorema. Se X é um processo tal que $\Phi(L)X_t = \Theta(L)\varepsilon_t$, com Φ e Θ polinómios de graus p e q , respectivamente, nas condições da definição anterior, então a sua função de autocovariância verifica a seguinte equação de recorrência linear homogénea de grau p , a partir da ordem $q + 1$:

$$\gamma(h) - \sum_{j=1}^p \varphi_j \gamma(h-j) = 0, \quad \forall h \geq q + 1.$$

Prova. A partir de

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

obtemos, para todo $h \geq 0$,

$$\begin{aligned} \text{Cov}(X_{t-h}, X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p}) &= \\ &= \text{Cov}(X_{t-h}, \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}). \end{aligned}$$

O segundo membro é nulo para todo o $h \geq q + 1$.

De facto, sob as hipóteses da última definição e tendo em conta a representação $MA(+\infty)$, X_{t-h} é função de $\varepsilon_{t-h}, \varepsilon_{t-h-1}, \dots$

Então

$$\forall h \geq q + 1, \gamma(h) - \sum_{j=1}^p \varphi_j \gamma(h-j) = 0. \blacksquare$$

Exemplo. Seja X um processo nas condições da propriedade anterior e tal que as raízes de Φ são distintas e iguais a $z_j, j = 1, \dots, p$. O polinómio característico da equação de recorrência verificada por $\gamma(h), h \geq q + 1$, é

$$z^p - \varphi_1 z^{p-1} - \dots - \varphi_p z^0 = z^p \left(1 - \frac{\varphi_1}{z} - \dots - \frac{\varphi_p}{z^p} \right) = z^p \Phi \left(\frac{1}{z} \right).$$

Assim, tendo em conta que as raízes da equação característica são $\frac{1}{z_j}, j = 1, \dots, p$, vem

$$\gamma(h) = A_1 \left(\frac{1}{z_1} \right)^h + A_2 \left(\frac{1}{z_2} \right)^h + \dots + A_p \left(\frac{1}{z_p} \right)^h.$$

Para determinar as p constantes A_1, A_2, \dots, A_p são necessárias p condições iniciais. Podemos obtê-las resolvendo o sistema que se obtém a partir da igualdade seguinte, válida qualquer que seja $h \geq 0$,

$$\begin{aligned} \text{Cov}(X_{t-h}, X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p}) \\ = \text{Cov}(X_{t-h}, \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}), \end{aligned}$$

e no qual usamos o desenvolvimento $MA(+\infty)$ de X_{t-h} ,

$$X_{t-h} = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j-h}, \text{ com } a_0 = 1. \text{ Assim,}$$

$$\begin{aligned} \gamma(h) - \sum_{j=1}^p \varphi_j \gamma(h-j) &= Cov(X_{t-h}, \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}) \\ &= Cov\left(\sum_{j=0}^{+\infty} a_j \varepsilon_{t-j-h}, \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}\right) \\ &= (-\theta_h - a_1 \theta_{h+1} - \dots - a_{q-h} \theta_q) \sigma^2, \end{aligned}$$

com $\theta_0 = -1$.

Utilizando estas relações, a equação de recorrência e a paridade da função γ , todos os valores das autocovariâncias podem ser obtidos. De facto,

a) se $p \geq q$, calculamos $\gamma(1), \dots, \gamma(p)$ e usamos a equação de recorrência com estas condições iniciais;

b) se $p < q$ é ainda necessário calcular $\gamma(p+1), \dots, \gamma(q)$ e, em seguida, usar $\gamma(q-p+1), \dots, \gamma(q)$ como valores iniciais.

$$\text{Teremos então } \gamma(h) = \sum_{j=1}^p A_j \left(\frac{1}{z_j}\right)^h, \quad h \geq \max(1, q-p+1).$$

Outra forma de determinar as condições iniciais consiste em utilizar a função geradora das autocovariâncias

$$G(z) = \sigma^2 A(z) A\left(\frac{1}{z}\right),$$

$$\text{com } A(z) = \frac{\Theta(z)}{\Phi(z)}. \blacksquare$$

Teorema. Se X é um processo estacionário centrado regular tal que a sua função de autocovariância γ verifica

$$\gamma(h) - \tilde{\varphi}_1 \gamma(h-1) - \dots - \tilde{\varphi}_{\tilde{p}} \gamma(h-\tilde{p}) = 0, \quad \forall h \geq \tilde{q} + 1,$$

então X admite uma representação *ARMA*. As ordens da representação *ARMA* mínima são p e q , com:

p igual à menor das ordens das equações de recorrência daquele tipo verificadas por γ a partir de uma certa ordem;

$q + 1$ igual ao menor inteiro a partir do qual γ verifica uma equação de recorrência de ordem p .

Prova. A condição verificada pela função de autocovariância implica que $X_t - \sum_{j=1}^{\tilde{p}} \tilde{\varphi}_j X_{t-j}$ é ortogonal a $H_{t-\tilde{q}-1}^X = H_{t-\tilde{q}-1}^\varepsilon$, onde ε designa a inovação de X .

Como X é regular isto implica que $X_t - \sum_{j=1}^{\tilde{p}} \tilde{\varphi}_j X_{t-j}$ é da forma

$$\varepsilon_t - \tilde{\theta}_1 \varepsilon_{t-1} - \dots - \tilde{\theta}_{\tilde{q}} \varepsilon_{t-\tilde{q}}.$$

A caracterização de p e q resulta da sua interpretação enquanto ordens mínimas. ■

Densidade espectral

Propriedade. Se um processo X estacionário regular admite uma representação *ARMA* mínima $\Phi(L)X_t = \Theta(L)\varepsilon_t$, a sua densidade espectral é da forma

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \frac{\Theta(e^{i\omega})}{\Phi(e^{i\omega})} \right|^2, \quad \omega \in [-\pi, \pi[.$$

Prova. Basta ter em conta que X admite uma representação *MA*($+\infty$).

3.4.3 Caracterização dos processos ARMA

Há caracterizações das ordens mínimas que se baseiam em determinantes envolvendo a função de autocovariância γ do processo. Consideremos as matrizes seguintes

$$A(i, j) = \begin{bmatrix} \gamma(j+1) & \gamma(j+2) & \dots & \gamma(j+i+1) \\ \gamma(j) & \gamma(j+1) & \dots & \gamma(j+i) \\ \dots & \dots & \dots & \dots \\ \gamma(j+1-i) & \gamma(j+2-i) & \dots & \gamma(j+1) \end{bmatrix}$$

e seja

$$\Delta(i, j) = \det A(i, j).$$

Este determinante surge de modo natural quando calculamos as correlações canônicas entre os vectores de dimensão $i + 1$

$$\begin{bmatrix} X_t & X_{t-1} & \dots & X_{t-i} \end{bmatrix}' \text{ e } \begin{bmatrix} X_{t-j-1} & X_{t-j-2} & \dots & X_{t-j-i-1} \end{bmatrix}',$$

pois tais correlações são, por definição, os valores próprios da matriz

$$[A(i, -1)]^{-1} [A(i, j)]' [A(i, -1)]^{-1} A(i, j).$$

Uma caracterização útil na prática consiste em ver as ordens (i, j) para as quais existe uma correlação canónica nula. A propriedade seguinte (Beguin, Gouriéroux e Monfort (1980), Gouriéroux (1987)) estabelece tal caracterização.

Propriedade. Seja X um processo regular. Este processo admite uma representação $ARMA(p, q)$ mínima se e só se

$$\begin{cases} \Delta(i, j) = 0, \forall i \geq p, \forall j \geq q \\ \Delta(p, q-1) \neq 0 \\ \Delta(p-1, q) \neq 0. \end{cases}$$

Além disso,

$$\Delta(i, q-1) \neq 0, \forall i \geq p-1 \text{ e } \Delta(p-1, j) \neq 0, \forall j \geq q-1. \blacksquare$$

Esta propriedade conduz à construção de uma tabela de dupla entrada, tendo em linha o índice i , em coluna o índice j e no cruzamento o valor do determinante $\Delta(i, j)$. O processo X admite uma representação $ARMA$ se e só se esta tabela contém um ‘canto’ inferior direito composto apenas por zeros. Isto explica o nome de *método do canto* atribuído a este método de identificação ⁽⁴⁾.

⁴Nesta tabela o símbolo * significa que $\Delta(i, j) \neq 0$.

$i \setminus j$	0	1	2	...	$q-1$	q
0									
1									
2									
...									
$p-1$					*	*	*	*	*
p					*	0	0	0	0
...					*	0	0	0	0
...					*	0	0	0	0
...					*	0	0	0	0

Um tal procedimento será fácil de executar se dispusermos de um algoritmo que permita calcular de modo expedito os valores dos vários determinantes. A propriedade seguinte (Glasbey (1982), Pham Dinh (1984)) fornece-nos um tal algoritmo pois estabelece uma fórmula que permite calcular uma linha (resp. coluna) de determinantes a partir das duas linhas (resp. colunas) anteriores.

Propriedade. Tem-se

$$[\Delta(i, j)]^2 = \Delta(i+1, j) \Delta(i-1, j) + \Delta(i, j-1) \Delta(i, j+1). \quad \blacksquare$$

3.4.4 Exemplos de aplicação

1. Seja X um processo $AR(1)$ definido por $X_t = \rho X_{t-1} + \varepsilon_t$ com $|\rho| < 1$ e onde ε é um ruído branco de variância σ^2 . Consideremos o processo Y definido por $Y_t = X_{kt}, t \in \mathbb{Z}$, com $k \in \mathbb{Z}$ fixo.

Mostremos que Y é estacionário e que admite uma representação $AR(1)$.

O processo Y é de segunda ordem porque $|\rho| < 1$ e tem-se

$$E(Y_t) = 0, \quad E(Y_t Y_{t+h}) = E(X_{kt} X_{kt+kh}) = \gamma_X(kh).$$

Mas γ_X verifica a equação de recorrência

$$\gamma_X(h) - \rho \gamma_X(h-1) = 0.$$

Consequentemente

$$\gamma_X(kh) = \rho^{kh} \gamma_X(0)$$

com $\gamma_X(0) = E(X_t^2) = \rho^2 \gamma_X(0) + \sigma^2$, isto é, $\gamma_X(0) = \frac{\sigma^2}{1-\rho^2}$.

Então

$$\gamma_Y(h) = \frac{\sigma^2}{1-\rho^2} \rho^{kh}, h \geq 0.$$

Esta sucessão verifica a equação de recorrência de ordem 1 seguinte

$$\gamma_Y(h) - \rho^k \gamma_Y(h-1) = 0.$$

Então o processo Y verifica

$$Y_t - \rho^k Y_{t-1} = \eta_t$$

ou seja,

$$\begin{aligned} X_{kt} - \rho^k X_{k(t-1)} &= \rho X_{kt-1} + \varepsilon_{kt} - \rho^k X_{kt-k} \\ &= \dots \\ &= \rho^{k-1} \varepsilon_{kt-k+1} + \dots + \varepsilon_{kt}. \end{aligned}$$

Sendo η o processo, função de ε , tal que

$$\eta_t = \rho^{k-1} \varepsilon_{kt-k+1} + \dots + \varepsilon_{kt} = \sum_{j=0}^{k-1} \rho^j \varepsilon_{kt-j}$$

prova-se facilmente que se trata de um ruído branco, o que permite concluir a existência da representação $AR(1)$ para Y .

2. Consideremos o processo estocástico real $X = (X_t, t \in \mathbb{Z})$ admitindo a seguinte representação $ARMA$

$$X_t - 0.3X_{t-1} = \varepsilon_t - 10\varepsilon_{t-1}, t \in \mathbb{Z}$$

onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância 1.

Determinemos uma representação $MA(+\infty)$ de X .

Temos

$$X_t - 0.3X_{t-1} = \varepsilon_t - 10\varepsilon_{t-1} \iff (1 - 0.3L)X_t = (1 - 10L)\varepsilon_t.$$

Assim,

$$X_t = \frac{1 - 10L}{1 - 0.3L} \varepsilon_t = \varepsilon_t + \sum_{i=1}^{+\infty} \left(-10(0.3)^{i-1} + (0.3)^i \right) \varepsilon_{t-i}.$$

Para obter uma representação $AR(+\infty)$ de X onde apenas intervenham os valores presente e passados do processo, defina-se o processo η tal que

$$\eta_t = \frac{1 - 0.3L}{1 - 0.1L} X_t.$$

O processo η é um ruído branco de variância 100 uma vez que a sua densidade espectral é dada por

$$\begin{aligned} f_\eta(w) &= \left| \frac{1 - 0.3e^{i\omega}}{1 - 0.1e^{i\omega}} \right|^2 f_X(w) \\ &= \left| \frac{1 - 0.3e^{i\omega}}{1 - 0.1e^{i\omega}} \right|^2 \frac{1}{2\pi} \left| \frac{1 - 10e^{i\omega}}{1 - 0.3e^{i\omega}} \right|^2 = \frac{100}{2\pi}. \end{aligned}$$

O processo X admite então a seguinte representação canónica

$$X_t - 0.3X_{t-1} = \eta_t - 0.1\eta_{t-1}$$

pelo que a representação $AR(+\infty)$ procurada é

$$\eta_t = \frac{1 - 0.3L}{1 - 0.1L} X_t = X_t - \sum_{i=1}^{+\infty} \left(0.3(0.1)^{i-1} - (0.1)^i \right) X_{t-i}.$$

O erro de previsão ao horizonte 1 no instante t é $X_{t+1} - E_L(X_{t+1}/\underline{H}_t)$ onde $E_L(X_{t+1}/\underline{H}_t)$ é a previsão linear de X_{t+1} dado o seu passado \underline{H}_t ; mostremos que a variância deste erro é igual a 100.

Ora

$$E_L(X_{t+1}/\underline{H}_t) = \sum_{i=1}^{+\infty} \left(0.3(0.1)^{i-1} - (0.1)^i \right) X_{t+1-i} + E_L(\eta_{t+1}/\underline{H}_t).$$

Mas $\underline{H}_t = \underline{H}_t^X = \underline{H}_t^\eta$ e η_{t+1} é ortogonal a \underline{H}_t^η .

Donde

$$E_L(X_{t+1}/\underline{H}_t) = \sum_{i=1}^{+\infty} \left(0.3(0.1)^{i-1} - (0.1)^i \right) X_{t+1-i}$$

e então

$$V[X_{t+1} - E_L(X_{t+1}/\underline{H}_t)] = V(\eta_{t+1}) = 100.$$

3. Consideremos o processo estocástico real $X = (X_t, t \in \mathbb{Z})$ tal que

$$(1 - \alpha L)^2 X_t = \varepsilon_t, \quad t \in \mathbb{Z},$$

onde L é o operador atraso associado a X , α um número real e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco de variância $\sigma_\varepsilon^2 > 0$.

As raízes do polinómio auto-regressivo $\Phi(L) = (1 - \alpha L)^2$ são $z_1 = z_2 = \frac{1}{\alpha}$. Mostremos que se $|\alpha| \neq 1$, o processo X é estacionário.

i) Se $|\alpha| < 1$, então $1 - \alpha L$ é invertível e

$$\begin{aligned} X_t &= \frac{1}{1 - \alpha L} \frac{1}{1 - \alpha L} \varepsilon_t = \sum_{j=0}^{+\infty} \alpha^j L^j \sum_{k=0}^{+\infty} \alpha^k L^k \varepsilon_t \\ &= \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \alpha^{j+k} L^{j+k} \varepsilon_t \\ &= \sum_{n=0}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n}. \end{aligned}$$

Mas

$$\sum_{n=0}^{+\infty} |(n+1) \alpha^n| = \sum_{n=0}^{+\infty} (n+1) |\alpha|^n < +\infty,$$

pois $\lim_n \frac{(n+2) |\alpha|^{n+1}}{(n+1) |\alpha|^n} = |\alpha| < 1$.

Sendo aquela série convergente, o processo X é fracamente estacionário pois ε é também fracamente estacionário.

ii) Se $|\alpha| > 1$, vem

$$(1 - \alpha L)^2 = \left[(-\alpha L) \left(1 - \frac{1}{\alpha L} \right) \right]^2 = \alpha^2 L^2 \left(1 - \frac{1}{\alpha} F \right)^2.$$

Então

$$\begin{aligned} (1 - \alpha L)^2 X_t &= \varepsilon_t \Leftrightarrow X_t = \left[\alpha^2 L^2 \left(1 - \frac{1}{\alpha} F \right)^2 \right]^{-1} \varepsilon_t \\ &\Leftrightarrow X_t = \frac{1}{\alpha^2} F^2 \frac{1}{\left(1 - \frac{1}{\alpha} F \right)} \frac{1}{\left(1 - \frac{1}{\alpha} F \right)} \varepsilon_t \\ &\Leftrightarrow X_t = \frac{1}{\alpha^2} F^2 \sum_{n=0}^{+\infty} (n+1) \left(\frac{1}{\alpha} \right)^n \varepsilon_{t+n} \end{aligned}$$

Como $\lim_n \frac{(n+2)|\alpha|^{n+1}}{(n+1)|\alpha|^{n+2}} = \frac{1}{|\alpha|} < 1$, a série $\sum_{n=0}^{+\infty} |(n+1) \left(\frac{1}{\alpha} \right)^n|$ é convergente pelo que X é fracamente estacionário.

Em resumo, se $|\alpha| < 1$, $X_t = \sum_{n=0}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n}$ e se $|\alpha| > 1$,
 $X_t = \frac{1}{\alpha^2} F^2 \sum_{n=0}^{+\infty} (n+1) \left(\frac{1}{\alpha} \right)^n \varepsilon_{t+n} = \sum_{n=0}^{+\infty} (n+1) \left(\frac{1}{\alpha} \right)^{n+2} \varepsilon_{t+n+2}$.

Suponhamos, a partir de agora, $|\alpha| < 1$ e determinemos a função de autocorrelação e a função de autocorrelação parcial de X .

Temos

$$E(X_t) = E \left(\varepsilon_t + \sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n} \right) = 0$$

pois é possível trocar o operador esperança matemática, E , com o limite. De facto, comecemos por estabelecer que, sendo

$$U_k = \sum_{n=1}^k (n+1) \alpha^n \varepsilon_{t-n} \text{ e } U = \sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n},$$

então U_k converge em L^2 para U quando k tende para $+\infty$, isto é,

$$\lim_k E(U_k - U)^2 = 0,$$

ou ainda que, para $m_1 > m_2$, $\|U_{m_1} - U_{m_2}\|^2$ tende para zero quando $m_1 \rightarrow +\infty, m_2 \rightarrow +\infty$. Ora,

$$\begin{aligned} \|U_{m_1} - U_{m_2}\|^2 &= \left\| \sum_{n=m_1}^{m_2} (n+1) \alpha^n \varepsilon_{t-n} \right\|^2 \\ &= E \left[\sum_{n=m_1}^{m_2} (n+1) \alpha^n \varepsilon_{t-n} \right]^2 \\ &= \sum_{n=m_1}^{m_2} (n+1)^2 \alpha^{2n} \sigma_\varepsilon^2 \end{aligned}$$

Como $\lim_n \frac{(n+2)^2 \alpha^{2n+2}}{(n+1)^2 \alpha^{2n}} = \alpha^2 < 1$, podemos então concluir que $\lim_k E(U_k - U)^2 = 0$.

Por outro lado,

$$E(U_k - U) = E[(U_k - U) 1] \leq [E(U_k - U)^2]^{\frac{1}{2}}$$

pelo que $\lim_k E(U_k - U) = 0 \Leftrightarrow \lim_k E(U_k) = E(U)$.

Assim,

$$\begin{aligned} \lim_k E \left[\sum_{n=1}^k (n+1) \alpha^n \varepsilon_{t-n} \right] &= \lim_k \sum_{n=1}^k (n+1) \alpha^n E(\varepsilon_{t-n}) \\ &= 0 = E \left[\sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n} \right]. \end{aligned}$$

Calculemos agora a variância de $X, \gamma_X(0)$.

$$\gamma_X(0) = E(X_t^2) = E \left[\left(\varepsilon_t + \sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n} \right)^2 \right]$$

$$\begin{aligned}
&= E[\varepsilon_t^2 + 2\varepsilon_t \sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n} + \sum_{n=1}^{+\infty} (n+1)^2 \alpha^{2n} \varepsilon_{t-n}^2 + \\
&\quad + \sum_{n=1}^{+\infty} \sum_{m \neq n}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n} (m+1) \alpha^m \varepsilon_{t-m}] \\
&= E(\varepsilon_t^2) + 2E \left[\varepsilon_t \sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n} \right] + E \left[\sum_{n=1}^{+\infty} (n+1)^2 \alpha^{2n} \varepsilon_{t-n}^2 \right] + \\
&\quad + \sum_{n=1}^{+\infty} \sum_{m \neq n}^{+\infty} (n+1) (m+1) \alpha^{n+m} E(\varepsilon_{t-n} \varepsilon_{t-m}).
\end{aligned}$$

A segunda parcela é nula pois ε_t é ortogonal a $\sum_{n=1}^{+\infty} (n+1) \alpha^n \varepsilon_{t-n}$ (elemento de $H(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$), assim como a última pois $E(\varepsilon_{t-n} \varepsilon_{t-m}) = 0$ para $n \neq m$. Então

$$\begin{aligned}
\gamma_X(0) &= E(\varepsilon_t^2) + E \left[\sum_{n=1}^{+\infty} (n+1)^2 \alpha^{2n} \varepsilon_{t-n}^2 \right] \\
&= E \left(\lim_k \sum_{n=1}^k (n+1)^2 \alpha^{2n} \varepsilon_{t-n}^2 \right) \\
&= \lim_k E \left(\sum_{n=1}^k (n+1)^2 \alpha^{2n} \varepsilon_{t-n}^2 \right)
\end{aligned}$$

pelo teorema da convergência monótona, pelo que

$$\gamma_X(0) = \lim_k \sum_{n=1}^k E \left[(n+1)^2 \alpha^{2n} \varepsilon_{t-n}^2 \right] = \sum_{n=1}^{+\infty} (n+1)^2 \alpha^{2n} \sigma_\varepsilon^2.$$

Este valor é finito pois $\lim_n \frac{(n+2)^2 \alpha^{2n+2}}{(n+1)^2 \alpha^{2n}} = \alpha^2 < 1$. Então,

$$\gamma_X(0) = \sum_{n=0}^{+\infty} (n+1)^2 (\alpha)^{2n} \sigma_\varepsilon^2.$$

Determinemos agora $\gamma_X(h) = Cov(X_t, X_{t-h})$, $h > 0$. Ora

$$(1 - \alpha L)^2 X_t = \varepsilon_t \Leftrightarrow X_t - 2\alpha X_{t-1} + \alpha^2 X_{t-2} = \varepsilon_t.$$

Como

$$\text{Cov}(X_{t-h}, \varepsilon_t) = 0, h > 0,$$

pois $X_{t-h} \in H(X_{t-h-1}, X_{t-h-2}, \varepsilon_{t-h}) \subset H(\varepsilon_{t-h}, \varepsilon_{t-h-1}, \dots)$, vem

$$\begin{aligned} \text{Cov}(X_{t-h}, X_t - 2\alpha X_{t-1} + \alpha^2 X_{t-2}) &= 0 \Leftrightarrow \\ \gamma_X(h) - 2\alpha\gamma_X(h-1) + \alpha^2\gamma_X(h-2) &= 0, \end{aligned}$$

para $h \geq 1$.

O polinómio característico desta equação às diferenças linear homogénea é $x^2 - 2\alpha x + \alpha^2 = 0$ cujas raízes são $x = \alpha$ ou $x = -\alpha$.

Assim, a solução geral da equação é

$$\gamma_X(h) = (A + Bh)\alpha^h, h \geq 1.$$

Para determinar A e B , usamos as condições iniciais correspondentes a $h = 0$ e $h = 1$.

Notemos que para $h = 1$ vem

$$\gamma_X(1) - 2\alpha\gamma_X(0) + \alpha^2\gamma_X(1) = 0 \Leftrightarrow \gamma_X(1) = \frac{2\alpha}{1 + \alpha^2}\gamma_X(0).$$

Então,

$$\begin{cases} \gamma_X(0) = A \\ \gamma_X(1) = (A + B)\alpha \end{cases} \Leftrightarrow \begin{cases} A = \gamma_X(0) \\ B = \frac{\gamma_X(1)}{\alpha} - \gamma_X(0) \end{cases}.$$

Finalmente, a autocovariância de ordem h de X é

$$\gamma_X(h) = \left(\gamma_X(0) + \frac{\gamma_X(1) - \alpha\gamma_X(0)}{\alpha} h \right) \alpha^h, h \geq 0,$$

e a autocorrelação de ordem h de X é

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}, h \geq 0.$$

A paridade destas funções completa a sua definição.

Quanto à função de autocorrelação parcial, vem

$$E_L(X_t/X_{t-1}, \dots, X_{t-k}, \dots) = 2\alpha X_{t-1} + \alpha^2 X_{t-2}, \forall k \geq 2.$$

Então $\forall k \geq 2$, $E_L(X_t/X_{t-1}, \dots, X_{t-k}) = 2\alpha X_{t-1} + \alpha^2 X_{t-2}$.

Deduzimos que $r(2) = \alpha^2$ e $r(k) = 0$, $k > 2$.

Finalmente $r(1) = \rho_X(1) = \frac{\gamma_X(1)}{\gamma_X(0)} = \frac{2\alpha}{1+\alpha^2}$.

Podemos ainda averiguar, com $|\alpha| < 1$, a existência de densidade espectral.

Sendo $X_t = \varepsilon_t + \sum_{n=1}^{+\infty} (n+1)\alpha^n \varepsilon_{t-n}$, com $\sum_{n=1}^{+\infty} (n+1)|\alpha|^n < +\infty$ pois $|\alpha| < 1$, então a densidade espectral de X existe sendo uma versão dada, para $\omega \in [-\pi, \pi[$, por

$$\begin{aligned} f_X(\omega) &= \left| \sum_{n=0}^{+\infty} (n+1)\alpha^n \exp(i\omega n) \right| \frac{\sigma_\varepsilon^2}{2\pi} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{n=0}^{+\infty} (n+1)[\alpha \exp(i\omega)]^n \right| \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{\left| [1 - \alpha \exp(i\omega)]^2 \right|^2} \\ &= \frac{\sigma_\varepsilon^2}{2\pi} \frac{1}{\left| 1 - 2\alpha \exp(i\omega) + \alpha^2 \exp(2i\omega) \right|^2} \end{aligned}$$

usando o facto de, sob condições aqui verificadas,

$$\sum_{n=0}^{+\infty} (n+1)x^n = \sum_{n=1}^{+\infty} nx^{n-1} = \left(\sum_{n=0}^{+\infty} x^n \right)' = \left(\frac{1}{1-x} \right)' = \frac{1}{(1-x)^2}.$$

Consideremos, finalmente, $\alpha = 2$ e determinemos a representação canónica do processo X .

Vimos que, neste caso, $X_t = \sum_{n=0}^{+\infty} \left(\frac{1}{2}\right)^{n+2} (n+1) \varepsilon_{t+n+2}$.

Esta representação não é canónica pois ε_t não é ortogonal a $H(X_{t-1}, X_{t-2}, \dots)$. De facto,

$$\begin{aligned} Cov(\varepsilon_t, X_{t-2}) &= Cov\left(\varepsilon_t, \sum_{n=0}^{+\infty} \left(\frac{1}{2}\right)^{n+2} (n+1) \varepsilon_{t+n}\right) \\ &= \left(\frac{1}{2}\right)^2 \sigma_\varepsilon^2 \neq 0. \end{aligned}$$

Consideremos então o processo $\eta = (\eta_t, t \in \mathbb{Z})$ tal que

$$\eta_t = \left(1 - \frac{1}{2}L\right)^2 X_t.$$

A densidade espectral deste novo processo é, para $\omega \in [-\pi, \pi[$,

$$\begin{aligned} f_\eta(\omega) &= \left| \left(1 - \frac{1}{2} \exp(i\omega)\right)^2 \right|^2 f_X(\omega) \\ &= \frac{\sigma_\varepsilon^2 \left| \left(1 - \frac{1}{2} \exp(i\omega)\right)^2 \right|^2}{2\pi \left| [1 - 2 \exp(i\omega)]^2 \right|^2} = \frac{\sigma_\varepsilon^2}{32\pi}. \end{aligned}$$

Sendo f_η constante, podemos concluir que η é um ruído branco.

Assim, a representação canônica de X quando $\alpha = 2$ é

$$X_t = X_{t-1} - \frac{1}{4}X_{t-2} + \eta_t,$$

com $\eta = (\eta_t, t \in \mathbb{Z})$ um ruído branco de variância $\frac{\sigma_\varepsilon^2}{16}$.

3.4.5 Representação de espaços de estado de um modelo ARMA

Os modelos de espaços de estado são definidos à custa de duas equações. A primeira equação relaciona variáveis observáveis X_t com variáveis, ditas de estado, Z_t , que podem ser ou não observáveis. A segunda equação descreve a evolução destas variáveis de estado. Uma formulação possível é então

$$\begin{cases} X_t = HZ_t + v_t \\ Z_t = GZ_{t-1} + u_t \end{cases}$$

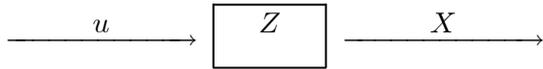
onde H e G são matrizes reais e v_t e u_t são vectores de ruídos, temporalmente independentes (isto é, independentes quando $t \neq t'$), independentes entre eles, de médias nulas e de matrizes de variâncias-covariâncias tais que $V(u_t) = P$, $V(v_t) = Q$.

Na terminologia da teoria dos sistemas, as variáveis $u = (u_t)$ são chamadas *input* e as variáveis X dizem-se *output* ou observações; as variáveis Z , que resumem o efeito dinâmico de u sobre X , dizem-se *variáveis de estado*. O sistema apresentado é composto por duas equações claramente diferentes:

- a segunda, dita equação de estado, explica como evolui a variável de estado em função do *input*;

- a primeira equação, dita equação de medida, dá-nos a determinação do *output* em função do estado do sistema.

A figura seguinte resume visualmente esta terminologia.



Os exemplos seguintes ilustram a definição de modelos de espaços de estado.

Exemplos.

1. Consideremos o processo unidimensional $AR(2)$

$$X_t - \varphi_1 X_{t-1} - \varphi_2 X_{t-2} = \varepsilon_t.$$

Podemos escrevê-lo na forma

$$\left\{ \begin{array}{l} X_t = [1 \quad 0] \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} \\ \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix} \end{array} \right. .$$

Neste caso

$$Z_t = \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix}.$$

2. Considere-se o processo unidimensional $X_t = \varepsilon_t - \theta\varepsilon_{t-1}$.

Definindo $Z_t = \begin{bmatrix} \varepsilon_t - \theta\varepsilon_{t-1} \\ -\theta\varepsilon_t \end{bmatrix}$, podemos escrever

$$\begin{cases} X_t = [1 & 0] Z_t \\ Z_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} - \theta\varepsilon_{t-2} \\ -\theta\varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ -\theta\varepsilon_t \end{bmatrix} \end{cases} \blacksquare$$

Passamos agora ao estudo do caso geral de uma formulação *ARMA* unidimensional. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico real tal que

$$X_t - \varphi_1 X_{t-1} - \dots - \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

onde os números reais φ_i e θ_j são tais que ε_t é a inovação de X no instante t :

$$\varepsilon_t = X_t - E_L(X_t/X_{t-1}, X_{t-2}, \dots).$$

Como sabemos, tal processo pode escrever-se sob forma média móvel infinita:

$$X_t = \sum_{j=0}^{+\infty} a_j \varepsilon_{t-j}.$$

Vamos então estabelecer o resultado seguinte, em cuja prova se efectua a construção de uma representação de espaços de estado para o processo, particularmente útil na fase de previsão.

Propriedade. Um processo estocástico X , verificando uma modelação *ARMA* nas condições acima consideradas, admite uma representação de espaços de estado.

Prova. Sejam

$$\begin{aligned} k &= \max(p, q + 1); \\ X_t(j) &= E_L(X_{t+j}/X_t, X_{t-1}, \dots), j = 0, 1, 2, \dots; \\ Z_t &= \begin{bmatrix} X_t(0) \\ X_t(1) \\ \dots \\ X_t(k-1) \end{bmatrix}. \end{aligned}$$

Como $X_t(0) = X_t$, a primeira equação da representação é

$$X_t = [1 \ 0 \ 0 \ \dots \ 0] Z_t.$$

Então, usando as notações consideradas no modelo geral de espaços de estado, $V(v_t) = Q = 0$.

Determinemos agora a equação de evolução das variáveis de estado, Z_t .

i) Por um lado, de $X_{t+j} = \sum_{i=0}^{+\infty} a_i \varepsilon_{t+j-i}$, obtém-se para a esperança linear relativamente a X_t, X_{t-1}, \dots ,

$$X_t(j) = \sum_{i=0}^{+\infty} a_i E_L(\varepsilon_{t+j-i}/X_t, X_{t-1}, \dots).$$

Mas ε_t é ortogonal a $\underline{H}_{t-1}^\varepsilon = \underline{H}_{t-1}^X$ e portanto

$$E_L(\varepsilon_{t+j-i}/X_t, X_{t-1}, \dots) = 0, \text{ se } j-i > 0.$$

Então

$$\begin{aligned} X_t(j) &= \sum_{i=j}^{+\infty} a_i E_L(\varepsilon_{t+j-i}/X_t, X_{t-1}, \dots) \\ &= \sum_{i=j}^{+\infty} a_i \varepsilon_{t+j-i}, \end{aligned}$$

uma vez que $\varepsilon_t, \varepsilon_{t-1}, \dots \in \underline{H}_t^X$.

$$\text{Assim } X_t(j) = a_j \varepsilon_t + \sum_{i=j+1}^{+\infty} a_i \varepsilon_{t+j-i}.$$

Mas a segunda parcela pode interpretar-se do modo seguinte

$$\sum_{i=j+1}^{+\infty} a_i \varepsilon_{t+j-i} = \sum_{i=j+1}^{+\infty} a_i \varepsilon_{(t-1)+j+1-i} = X_{t-1}(j+1).$$

A relação obtida

$$X_t(j) = a_j \varepsilon_t + X_{t-1}(j+1)$$

permite o cálculo de $X_t(j)$, para $j = 0, \dots, k-2$ (ou mesmo para $j > k-2$) a partir dos valores associados ao instante anterior.

ii) Por outro lado, da equação *ARMA* escrita para o índice $t+k-1$,

$$X_{t+k-1} - \sum_{i=1}^p \varphi_i X_{t+k-1-i} = \varepsilon_{t+k-1} - \sum_{i=1}^q \theta_i \varepsilon_{t+k-1-i},$$

decorre para a esperança linear dado X_{t-1}, X_{t-2}, \dots ,

$$\begin{aligned} E_L(X_{t+k-1}/X_{t-1}, X_{t-2}, \dots) - \sum_{i=1}^p \varphi_i E_L(X_{t+k-1-i}/X_{t-1}, X_{t-2}, \dots) = \\ = E_L(\varepsilon_{t+k-1}/X_{t-1}, X_{t-2}, \dots) - \sum_{i=1}^q \theta_i E_L(\varepsilon_{t+k-1-i}/X_{t-1}, X_{t-2}, \dots). \end{aligned}$$

Como $k \geq q+1$ e $E_L(\varepsilon_{t+j}/X_{t-1}, X_{t-2}, \dots) = 0, j \geq 0$, vem

$$X_{t-1}(k) - \sum_{i=1}^p \varphi_i X_{t-1}(k-i) = 0.$$

Se em i) considerarmos $j = k-1$ obtemos

$$X_t(k-1) = a_{k-1} \varepsilon_t + X_{t-1}(k)$$

o que, tendo em conta a expressão obtida para $X_{t-1}(k)$, pode ser escrito na forma

$$X_t(k-1) = \sum_{i=1}^p \varphi_i X_{t-1}(k-i) + a_{k-1} \varepsilon_t$$

e explicita $X_t(k-1)$ em função do vector $(X_{t-1}(k-1), X_{t-1}(k-2), \dots, X_{t-1}(k-p))$.

Decorre então de i) e de ii) que

$$\begin{bmatrix} X_t(0) \\ X_t(1) \\ \dots \\ X_t(k-1) \end{bmatrix} = G \begin{bmatrix} X_{t-1}(0) \\ X_{t-1}(1) \\ \dots \\ X_{t-1}(k-1) \end{bmatrix} + \begin{bmatrix} 1 \\ a_1 \\ \dots \\ a_{k-1} \end{bmatrix} \varepsilon_t$$

com $G = \begin{bmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & 1 \\ \dots & \dots & \dots & \dots & \varphi_2 & \varphi_1 \end{bmatrix}.$

Em resumo, o processo Z verifica uma equação de evolução da forma

$$Z_t = GZ_{t-1} + u_t. \blacksquare$$

Destacamos de novo que o interesse da representação obtida reside sobretudo na simplicidade da actualização das previsões.

3.4.6 Generalização dos modelos ARMA

Em algumas séries temporais a hipótese de estacionaridade não é fácil de aceitar. Pode, em contrapartida, ser mais razoável aceitar tal hipótese para a série das diferenças

$$X_t - X_{t-1} = (1 - L) X_t$$

ou, mais geralmente, para a série das diferenças de ordem d , isto é, para

$$(1 - L)^d X_t, \quad d \in \mathbb{N}.$$

Uma ideia natural é, então, considerar os processos cujas diferenças de uma certa ordem admitem uma modelação *ARMA*. Somos assim conduzidos ao estudo de modelos da forma

$$\Phi(L)(1 - L)^d X_t = \Theta(L)\varepsilon_t, \quad t \in \mathbb{Z},$$

onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco, Φ e Θ verificam as hipóteses usuais relativas à modelação *ARMA* e $d \in \mathbb{N}$.

Um processo X que admita aquela representação é denominado *ARIMA*(p, d, q), onde a letra *I* provém da palavra *integrated*.

Repare-se que X pode ser interpretado como verificando um modelo *ARMA*($p + d, q$); nota-se, no entanto, que tal não apresenta grande interesse pois estamos perante um processo não estacionário uma vez que existe, para o polinómio auto-regressivo, uma raiz de módulo igual a 1 de multiplicidade d .

Refira-se ainda que, em termos práticos, não é habitual ir além das diferenças de ordem $d = 1$ ou $d = 2$.

Finalmente notemos que não é o processo X que é definido mas sim as diferenças sucessivas. Assim, se tivermos um modelo $ARIMA(0, 1, 0)$

$$X_t - X_{t-1} = \varepsilon_t$$

são definidas as diferenças $X_t - X_{t-1}, X_{t-1} - X_{t-2}, \dots$. Para recuperarmos X_t precisaremos de um valor inicial X_{t_0} . Se, por outro lado, X é tal que

$$X_t - X_{t-1} = \varepsilon_t - \theta\varepsilon_{t-1}, t \in \mathbb{Z}, |\theta| < 1,$$

então tal processo ficará completamente definido dadas as condições iniciais ε_{t_0} e X_{t_0} .

Podemos assim formular a seguinte definição de modelo $ARIMA(p, d, q)$.

Definição. Um processo $X = (X_t, t \geq 0)$ verifica um modelo $ARIMA(p, d, q)$ se

$$\Phi(L)(1-L)^d X_t = \Theta(L)\varepsilon_t, t \geq 0,$$

onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco, $\Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$, $\varphi_p \neq 0$, $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, $\theta_q \neq 0$, são polinómios cujas raízes são de módulo superior a 1, as condições iniciais

$$(X_{-1}, X_{-2}, \dots, X_{-p-d}, \varepsilon_{-1}, \varepsilon_{-2}, \dots, \varepsilon_{-q})$$

são não correlacionadas com $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_t, \dots$ e $d \in \mathbb{N}$.

Vimos já que certas séries temporais apresentam um perfil sazonal nítido; por exemplo, para certas séries mensais os dados relativos a um mesmo mês em diferentes anos têm tendência a situar-se de maneira análoga em relação à média anual. Nestes casos parece conveniente fazer aparecer num modelo $ARIMA$ deslocamentos múltiplos de 12.

Na teoria, nada impede que se considerem valores de p e q suficientemente elevados para que tais atrasos sejam tidos em conta; no entanto, isto implicaria tal aumento no número de parâmetros que seria praticamente impossível estimá-los convenientemente e, assim, utilizar tais modelos na prática.

A fim de contornar este problema, Box e Jenkins introduziram uma classe de modelos, generalização dos anteriores, capazes de ter em conta tais sazonalidades. Tais modelos são denominados *SARIMA* (a letra *S* decorre do termo *sazonal*). As principais propriedades desta classe de processos, cujo estudo não será desenvolvido neste texto, encontram-se descritas em Gouriéroux e Monfort (1990, p. 249).

3.5 Anexo - Inversibilidade de $1 - \lambda L$

A aplicação $1 - \lambda L$ é uma aplicação do conjunto dos processos estacionários nele próprio invertível se e só se $|\lambda| \neq 1$ e tal que

$$(1 - \lambda L)^{-1} = \begin{cases} \sum_{i=0}^{+\infty} \lambda^i L^i, & |\lambda| < 1 \\ \sum_{i=0}^{+\infty} \frac{1}{\lambda^i} L^{-i}, & |\lambda| < 1. \end{cases}$$

1. Considere-se $|\lambda| < 1$

A série $\sum_{i=0}^{+\infty} \lambda^i$ é absolutamente convergente pelo que podemos definir a aplicação $\sum_{i=0}^{+\infty} \lambda^i L^i$.

Ora, a composição de $1 - \lambda L$ com $\sum_{i=0}^{+\infty} \lambda^i L^i$ conduz a

$$\sum_{i=0}^{+\infty} \lambda^i L^i (1 - \lambda L) = \sum_{i=0}^{+\infty} \lambda^i L^i - \sum_{i=0}^{+\infty} \lambda^{i+1} L^{i+1} = L^0 = Id. \quad (5)$$

Então, tendo em conta a comutatividade do produto das séries em L e em F , concluímos que $1 - \lambda L$ é, neste caso, invertível e a sua inversa é $\sum_{i=0}^{+\infty} \lambda^i L^i$.

2. Seja agora $|\lambda| < 1$

Podemos escrever

⁵Note-se que $1 - \lambda L$ é uma forma simplificada, habitualmente presente na literatura, de $Id - \lambda L$ onde Id representa o operador Identidade.

$$1 - \lambda L = -\lambda L \left(1 - \frac{1}{\lambda} F\right).$$

O primeiro factor é invertível de inversa $-\frac{1}{\lambda}F$. Por outro lado, sendo $|\frac{1}{\lambda}| < 1$, a série $\sum_{i=0}^{+\infty} \frac{1}{\lambda^i} L^{-i} = \sum_{i=0}^{+\infty} \frac{1}{\lambda^i} F^i$ existe e prova-se, de modo análogo ao feito anteriormente, que é a inversa de $1 - \frac{1}{\lambda}F$.

Assim $1 - \lambda L$ é, também neste caso, invertível e tem-se

$$\begin{aligned} (1 - \lambda L)^{-1} &= \left[-\lambda L \left(1 - \frac{1}{\lambda} F\right) \right]^{-1} = \left(1 - \frac{1}{\lambda} F\right)^{-1} (-\lambda L)^{-1} \\ &= \sum_{i=0}^{+\infty} \frac{F^i}{\lambda^i} \left(-\frac{1}{\lambda} F\right) = \left(-\frac{1}{\lambda} F\right) \sum_{i=0}^{+\infty} \frac{F^i}{\lambda^i}. \end{aligned}$$

3. Note-se que se $|\lambda| = 1$ a aplicação não é invertível.

Constatemos este facto considerando $\lambda = 1$. Neste caso todo o processo constante é transformado no processo identicamente nulo pois

$$(1 - L) X_t = k - k = 0$$

e portanto a aplicação não é injectiva.

Observemos também que não é sobrejectiva. De facto, todo o processo constante igual a $k \neq 0$ não pode ser imagem por $1 - L$ de nenhum processo estacionário. Se assim fosse, ter-se-ia

$$(1 - L) X_t = k \implies X_t - X_{t-1} = k \implies E(X_t) - E(X_{t-1}) = k \neq 0$$

e o processo X não seria estacionário.▲

3.6 Anexo - Equações às diferenças

Consideremos uma função y_x onde $x \in \mathbb{Z}$. Definamos o operador avanço, D :

$$\left\{ \begin{array}{l} Dy_x = y_{x+1} \\ D^2 y_x = D(Dy_x) = y_{x+2} \\ \dots \\ D^n y_x = D(D^{n-1} y_x) = y_{x+n} \end{array} \right. .$$

Definição. Sejam a_1, a_2, \dots, a_n números reais. Uma equação envolvendo $y_x, y_{x+1}, \dots, y_{x+n}$ da forma

$$y_{x+n} + a_1 y_{x+n-1} + \dots + a_{n-1} y_{x+1} + a_n y_x = g_x \quad (A)$$

diz-se uma *equação de diferenças linear de ordem n de coeficientes constantes*.

Se $g_x = 0$ a equação diz-se *homogénea*.

Prova-se que a solução geral de (A) é a soma de uma combinação linear de todas as soluções da equação homogénea e de uma solução particular de (A). Além disso, para uma equação de grau n existem n constantes arbitrárias associadas à equação homogénea que, em cada caso, são encontradas a partir de n condições iniciais.

Determinemos a solução geral da equação homogénea.

A equação homogénea é equivalente a

$$(D^n + a_1 D^{n-1} + \dots + a_{n-1} D + a_n) y_x = 0.$$

Mostremos que as soluções desta equação são da forma r^x , onde r é uma raiz da equação característica

$$D^n + a_1 D^{n-1} + \dots + a_{n-1} D + a_n = 0.$$

Com efeito, seja $y_x = r^x$. Obtemos

$$r^{x+n} + a_1 r^{x+n-1} + \dots + a_n r^x = r^x (r^n + a_1 r^{n-1} + \dots + a_n) = 0$$

uma vez que r é raiz da equação característica.

Sempre que a equação característica tem n raízes reais distintas, r_1, \dots, r_n , a solução geral da equação homogénea é então

$$y_x = c_1 r_1^x + c_2 r_2^x + \dots + c_n r_n^x.$$

No caso da equação característica ter raízes múltiplas, por exemplo uma raiz r de multiplicidade k , as k soluções correspondentes a esta raiz são da forma

$$r^x, xr^x, x(x-1)r^x, \dots, x(x-1)\dots(x-k+1)r^x$$

uma vez que as primeiras $(k-1)$ derivadas da equação característica relativamente a D têm r como raiz.

Uma solução particular da equação geral é determinada, por exemplo, usando o método dos coeficientes indeterminados.

Exemplo. Consideremos a seguinte equação de diferenças linear de ordem 2 de coeficientes constantes:

$$y_{x+2} + 6y_{x+1} + 9y_x = 16x^2.$$

A equação homogênea é

$$(D^2 + 6D + 9)y_x = 0.$$

A equação característica $D^2 + 6D + 9 = 0$ tem uma raiz dupla igual a -3 .

Então a solução geral da equação homogênea é

$$y_x = c_1(-3)^x + c_2x(-3)^x.$$

Seja agora \tilde{y}_x uma solução particular da equação geral. Será da forma

$$\tilde{y}_x = Ax^2 + Bx + C.$$

Então, como \tilde{y}_x verifica $y_{x+2} + 6y_{x+1} + 9y_x = 16x^2$, vem

$$A(x+2)^2 + B(x+2) + C + 6[A(x+1)^2 + B(x+1) + C] + 9[Ax^2 + Bx + C] = 16x^2$$

o que permite obter

$$A = 1, B = -1, C = -\frac{1}{8},$$

pelo que a solução geral da equação inicial é

$$y_x = c_1(-3)^x + c_2x(-3)^x + x^2 - x - \frac{1}{8}.$$

Duas condições iniciais permitir-nos-iam agora determinar c_1 e c_2 .▲

Capítulo 4

Inferência em modelos lineares

Na análise de uma série temporal deparamos simplesmente com uma série de valores observados a partir dos quais pretendemos obter informação sobre a lei de probabilidade que rege o processo subjacente a tais observações. Para tal recorreremos aos métodos e resultados da Estatística Matemática tendo, no entanto, em mente que saímos do quadro clássico da amostragem, i.e., deixamos de ter variáveis aleatórias reais independentes e identicamente distribuídas (v.a.r.i.i.d.).

Vamos interessar-nos aqui pela estimação dos resumos de ordem um e dois da lei de um processo estacionário: média, função de autocovariância, densidade espectral (quando existe), ... Vamos usar para tal os correspondentes estimadores empíricos, indicando as suas principais propriedades assintóticas. Constataremos que resultados limites como os do teorema do limite central e das leis dos grandes números se generalizam ao caso dos processos estacionários, o que nos permitirá assegurar a validade da estimação.

No que vai seguir-se vamos supôr que pretendemos estimar os resumos de ordem um e de ordem dois de uma série temporal estacionária $X = (X_t, t \in \mathbb{Z})$, para o que dispomos de observações x_1, x_2, \dots, x_T (isto é, $\{1, 2, \dots, T\}$ é o intervalo de observação) das variáveis aleatórias X_1, X_2, \dots, X_T do processo estocástico X .

4.1 Propriedades dos estimadores

4.1.1 Média empírica

O estimador habitual para a média teórica $m = E(X_t)$, $t \in \mathbb{Z}$, é a média empírica

$$\bar{X}_T = \frac{1}{T} \sum_{t=1}^T X_t.$$

O mesmo tipo de aproximação pode ser utilizado para estimar a esperança de uma função qualquer de X_t . Por exemplo, $E(X_t^2)$

pode ser estimada por $\frac{1}{T} \sum_{t=1}^T X_t^2$.

São, obviamente, estimadores cêntricos.

As propriedades de convergência dos estimadores empíricos podem, em geral, ser estabelecidas no sentido da convergência em média quadrática ou no sentido da convergência quase certa. Existem diversas versões de teoremas de convergência, dependendo das hipóteses de estacionaridade feitas sobre o processo (estacionaridade fraca ou forte). Temos ainda aqui propriedades de normalidade assintótica análogas ao caso particular da amostragem aleatória, mas sob hipóteses mais fortes que as conducentes às convergências funcionais.

Apresentaremos alguns resultados de convergência no caso de processos fracamente estacionários admitindo uma representação média móvel infinita.

Teorema. Seja $X = (X_t, t \in \mathbb{Z})$ um processo admitindo a representação média móvel infinita

$$X_t = \sum_{i=0}^{+\infty} a_i \varepsilon_{t-i}, \text{ com } \sum_{i=0}^{+\infty} |a_i| < +\infty,$$

onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de variância σ^2 . Nestas condições, \bar{X}_T converge em média quadrática para $E(X_t) = 0$, com a velocidade de $\frac{1}{T}$ (quando $T \rightarrow +\infty$).

Prova. Se o processo admite uma representação média móvel infinita, então admite densidade espectral da forma

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{+\infty} a_j \exp(-i\omega j) \right|^2, \omega \in [-\pi, \pi[.$$

Assim, a medida espectral μ de X é absolutamente contínua, logo $\mu(\{0\}) = 0$.

A proposição para processos regulares abaixo enunciada ⁽¹⁾ permite assim concluir que \overline{X}_T converge em média quadrática para $E(X_t) = 0$, quando $T \rightarrow +\infty$. Além disso,

$$\lim_T TE(\overline{X}_T^2) = 2\pi f_X(0) = \sigma^2 \left| \sum_{i=0}^{+\infty} a_i \right|^2. \blacksquare$$

Este resultado generaliza-se de modo imediato ao caso dos processos da forma $X_t = m + \sum_{i=0}^{+\infty} a_i \varepsilon_{t-i}$, uma vez que o podemos aplicar ao processo definido por $Y_t = X_t - m, t \in \mathbb{Z}$, obtendo-se

$$\frac{1}{T} \sum_{t=1}^T (X_t - m) \xrightarrow{L_2} 0$$

e, de modo equivalente, $\overline{X}_T \xrightarrow{L_2} m$.

O resultado seguinte (Anderson, 1971, p. 429) apresenta condições que garantem a normalidade assintótica de $\sqrt{T} \overline{X}_T$.

Teorema. Se $X_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}$ onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é uma sucessão de v.a.r.i.i.d. centradas de variância σ^2 e se $\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$,

¹Seja X um processo estacionário, de média M . Para que a média empírica associada a T observações do processo, \overline{X}_T , convirja para M em L^2 é necessário e suficiente que a medida espectral μ de X verifique $\mu(\{0\}) = 0$. Se além disso X tem densidade espectral f contínua em zero, tem-se

$\lim_{T \rightarrow +\infty} TE[(\overline{X}_T - M)^2] = 2\pi f(0)$. (Azencott, Dacunha-Castelle, 1984, p. 103).

então $\sqrt{T} \bar{X}_T$ tem uma distribuição limite normal de média nula e variância $\sum_{h=-\infty}^{+\infty} \gamma(h)$ com $\gamma(h) = \sigma^2 \sum_{k=-\infty}^{+\infty} a_k a_{k+h}$.

Note-se que $\sum_{h=-\infty}^{+\infty} \gamma(h) = 2\pi f_X(0)$.

Exemplo. Se X é um processo $ARMA(p, q)$ tal que

$$\Phi(L) X_t = \Theta(L) \varepsilon_t$$

com Φ e Θ sob as condições da representação mínima, a densidade espectral de X é

$$f(\omega) = \frac{\sigma^2 |\Theta(e^{i\omega})|^2}{2\pi |\Phi(e^{i\omega})|^2}.$$

Então $\sqrt{T} \bar{X}_T$ tem uma distribuição limite normal centrada de variância $\sigma^2 \frac{|\Theta(1)|^2}{|\Phi(1)|^2}$. ■

4.1.2 Autocovariâncias empíricas

Interessamo-nos agora pela estimação das autocovariâncias $\gamma(h) = Cov(X_t, X_{t+h})$, $h \geq 0$.

Uma questão que naturalmente se poderá colocar é a de saber se podemos estimar convenientemente uma infinidade de parâmetros $\gamma(h)$ à custa de um número finito de observações; a justificação de que dispomos é o facto de termos $\gamma(h) \rightarrow 0$, quando $h \rightarrow +\infty$.

1. Caso em que a média é conhecida

Se a média, m , é conhecida podemos supôr, sem perda de generalidade, que é nula. Neste caso, são dois os estimadores habitualmente propostos para $\gamma(h)$:

$$\text{i) } \hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-h} X_t X_{t+h}, h = 0, 1, \dots, T-1$$

$$\text{ii) } \tilde{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} X_t X_{t+h}, h = 0, 1, \dots, T-1.$$

Notemos que estes estimadores não nos permitem estimar $\gamma(h)$ para $h \geq T$, o que intuitivamente conduz a aproximações pouco precisas se a ordem de h não for pequena relativamente a T .

Por outro lado, as estimativas correspondentes poderão ser bastante diferentes.

Isto pode ser ilustrado considerando $h = T-1$. De facto temos

$$\begin{aligned} \hat{\gamma}_T(T-1) &= \frac{1}{T} X_1 X_T \\ \tilde{\gamma}_T(T-1) &= X_1 X_T; \end{aligned}$$

além disso, tendo em conta que apenas se baseiam em duas observações, não poderemos esperar uma grande precisão.

É fácil verificar que $\tilde{\gamma}_T$ é um estimador cêntrico e que $\hat{\gamma}_T$ é apenas assintoticamente cêntrico.

Podemos, no entanto, invocar três razões para preferir $\hat{\gamma}_T$.

Em primeiro lugar, a sua expressão, envolvendo a divisão por uma mesma quantidade, T , torna os cálculos mais simples.

Além disso, $\hat{\gamma}_T(T-1)$ é melhor do que $\tilde{\gamma}_T(T-1)$ quando T é grande porque $\hat{\gamma}_T(T-1) \rightarrow 0$, quando T tende para $+\infty$. Ou seja, para T grande, o comportamento de $\hat{\gamma}_T$ é mais próximo do de γ porque se um processo é estacionário a sua função de autocovariância verifica $\gamma(h) \rightarrow 0$ quando $h \rightarrow +\infty$.

Finalmente, $(\hat{\gamma}_T(h), h = 0, 1, \dots)$ é de tipo positivo, isto é, a matriz

$$\begin{bmatrix} \hat{\gamma}_T(0) & \hat{\gamma}_T(1) & \dots & \dots & \dots \\ \hat{\gamma}_T(1) & \hat{\gamma}_T(0) & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \hat{\gamma}_T(1) & \hat{\gamma}_T(0) \end{bmatrix}$$

é semi-definida positiva, verificando pois as mesmas restrições que as quantidades a estimar $(\gamma(h), h = 0, 1, \dots)$.

Para provar que aquela matriz é semi-definida positiva consideremos a matriz de ordem $(T+H, T)$, H qualquer:

$$Y = \begin{bmatrix} X_1 & 0 & 0 & \dots & 0 \\ X_2 & X_1 & 0 & \dots & \dots \\ X_3 & X_2 & X_1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ X_T & \dots & \dots & \dots & X_1 \\ 0 & X_T & \dots & \dots & X_2 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & X_T & \dots \\ 0 & \dots & \dots & 0 & X_T \end{bmatrix}$$

Então $Y'Y$, que é semi-definida positiva, é igual a

$$Y'Y = \begin{bmatrix} \sum_{t=1}^T X_t^2 & \sum_{t=1}^{T-1} X_t X_{t+1} & \dots & \dots \\ \dots & \sum_{t=1}^T X_t^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \sum_{t=1}^T X_t^2 \end{bmatrix}$$

$$= T \begin{bmatrix} \hat{\gamma}_T(0) & \hat{\gamma}_T(1) & \dots & \dots \\ \hat{\gamma}_T(1) & \hat{\gamma}_T(0) & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \hat{\gamma}_T(0) \end{bmatrix}.$$

2. Caso em que a média é desconhecida

Se m é desconhecida temos de utilizar o seu estimador; assim, os estimadores propostos para $\gamma(h)$ são

$$\text{i) } \hat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T), h = 0, 1, \dots, T-1$$

$$\text{ii) } \tilde{\gamma}_T(h) = \frac{1}{T-h} \sum_{t=1}^{T-h} (X_t - \bar{X}_T) (X_{t+h} - \bar{X}_T), h = 0, 1, \dots, T-1.$$

O primeiro é em geral preferido, por razões semelhantes às invocadas no caso anterior.

Os estimadores $\hat{\gamma}_T$ e $\tilde{\gamma}_T$, anteriormente definidos, são conhecidos genericamente por funções de autocovariância empírica. Aos

estimadores cêntricos $\tilde{\gamma}_T$ e $\tilde{\tilde{\gamma}}_T$ acrescenta-se usualmente a designação de corrigidos.

Relativamente ao comportamento assintótico das autocovariâncias empíricas temos o mesmo tipo de aproximação que no caso da média empírica. O resultado seguinte (Anderson, 1971, p. 481) apresenta condições que garantem a normalidade assintótica de

$$\sqrt{T}(\hat{\gamma}_T(0) - \gamma(0), \dots, \hat{\gamma}_T(k) - \gamma(k)).$$

Resultados da mesma natureza, obtidos sob outro jogo de condições, figuram, por exemplo, em Azencott, Dacunha-Castelle, (1984, p. 111).

Teorema. Seja $X_t = \sum_{i=-\infty}^{+\infty} a_i \varepsilon_{t-i}$ uma média móvel infinita para a qual $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é uma sucessão de v.a.r.i.i.d. centradas de variância σ^2 e momento de quarta ordem finito verificando $E(\varepsilon_t^4) = 3\sigma^4 + k_4 < +\infty$.

Se $\sum_{i=-\infty}^{+\infty} |a_i| < +\infty$, então a distribuição limite de

$$\sqrt{T}(\hat{\gamma}_T(0) - \gamma(0), \dots, \hat{\gamma}_T(k) - \gamma(k))$$

é normal de média nula e de matriz de variâncias-covariâncias $\Gamma = [\Gamma_{h,g}]$, $0 \leq h, g \leq k$, com

$$\begin{aligned} \Gamma_{h,g} &= \sum_{r=-\infty}^{+\infty} [\gamma(r+g)\gamma(r+h) + \gamma(r-g)\gamma(r+h)] + \\ &\quad + \frac{k_4}{\sigma^4} \gamma(h)\gamma(g) \\ &= 4\pi \int_{-\pi}^{+\pi} \cos(\omega g) \cos(\omega h) f^2(\omega) d\omega + \frac{k_4}{\sigma^4} \gamma(h)\gamma(g). \end{aligned}$$

4.1.3 Periodograma

Se desejamos estimar a densidade espectral f , é natural utilizar a relação que a liga à função de autocovariância, ou seja,

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma(h) e^{i\omega h}.$$

Utilizando então as autocovariâncias empíricas $\hat{\gamma}_T(h)$, para $h = 0, 1, \dots, T-1$, obtemos como estimador natural

$$\hat{f}_T(\omega) = \frac{1}{2\pi} \sum_{h=-(T-1)}^{T-1} \hat{\gamma}_T(h) e^{i\omega h}.$$

A função $I_T(\omega) = \sum_{h=-(T-1)}^{T-1} \hat{\gamma}_T(h) e^{i\omega h}$ é denominada *periodograma*.

Propriedade. $I_T(\omega) = \frac{1}{T} \left| \sum_{t=1}^T X_t e^{i\omega t} \right|^2.$

Prova. Tem-se

$$\frac{1}{T} \left| \sum_{t=1}^T X_t e^{i\omega t} \right|^2 = \frac{1}{T} \sum_{t=1}^T X_t e^{i\omega t} \sum_{\tau=1}^T X_\tau e^{-i\omega \tau} = \frac{1}{T} \sum_{t=1}^T \sum_{\tau=1}^T X_t X_\tau e^{i\omega(t-\tau)}.$$

Fazendo a mudança de índice de t para $h = t - \tau$, vem

$$\frac{1}{T} \sum_{t=1}^T \sum_{\tau=1}^T X_t X_\tau e^{i\omega(t-\tau)} = \frac{1}{T} \sum_{h=-(T-1)}^{T-1} \left(\sum_{\tau=1}^{T-h} X_{\tau+h} X_\tau \right) e^{i\omega h}. \blacksquare$$

O estimador \hat{f}_T não apresenta um bom comportamento limite. O resultado seguinte (Anderson, 1971, p. 474) explicita esse comportamento.

Teorema. Quando T tende para $+\infty$,

$$E(\hat{f}_T(\omega)) \longrightarrow f(\omega), \forall \omega$$

$$V(\hat{f}_T(\omega)) \longrightarrow \begin{cases} f^2(\omega), & \omega \notin \{-\pi, 0, \pi\} \\ 2f^2(\omega), & \omega \in \{-\pi, 0, \pi\} \end{cases}$$

$$Cov(\hat{f}_T(\omega), \hat{f}_T(\omega')) \longrightarrow 0, \text{ se } \omega \neq \omega'.$$

Prova. Analisemos apenas o comportamento assintótico em média de $\widehat{f}_T = \frac{1}{2\pi} I_T$.

$$\begin{aligned} E(I_T(\omega)) &= E \left[\sum_{h=-(T-1)}^{T-1} \widehat{\gamma}_T(h) e^{i\omega h} \right] = E \left[\sum_{h=-(T-1)}^{T-1} \widehat{\gamma}_T(|h|) e^{i\omega h} \right] \\ &= \sum_{h=-(T-1)}^{T-1} \frac{T-|h|}{T} \gamma(|h|) e^{i\omega h}. \end{aligned}$$

Suponhamos que $\sum_{h=-\infty}^{+\infty} |\gamma(h)| < +\infty$. Então podemos aplicar o teorema da convergência dominada de Lebesgue e deduzir que

$$\lim_{T \rightarrow +\infty} E(I_T(\omega)) = \sum_{h=-\infty}^{+\infty} \gamma(h) e^{i\omega h} = 2\pi f(\omega). \quad \blacksquare$$

Assim, embora seja assintoticamente cêntrico, \widehat{f}_T não é um estimador convergente. Isto acontece devido à presença, na sua expressão, de estimadores de $\gamma(h)$ pouco precisos, nomeadamente quando h é próximo de T .

Exemplo. Verifiquemos o resultado relativo a $V(\widehat{f}_T(\omega))$, presente no teorema anterior, no caso simples em que X é um ruído branco gaussiano centrado e reduzido, isto é,

$$X_t = \varepsilon_t, t \in \mathbb{Z},$$

com ε_t variáveis aleatórias reais, de lei $N(0, 1)$, independentes.

Tem-se

$$\begin{aligned} [I_T(\omega)]^2 &= \frac{1}{T^2} \left| \sum_{t=1}^T \varepsilon_t e^{i\omega t} \right|^4 \\ &= \frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T \sum_{u=1}^T \sum_{v=1}^T \varepsilon_t \varepsilon_s \varepsilon_u \varepsilon_v e^{i\omega(-s+t+v-u)}. \end{aligned}$$

Consequentemente

$$E([I_T(\omega)]^2) = \frac{1}{T^2} \sum_{s=1}^T \sum_{t=1}^T \sum_{u=1}^T \sum_{v=1}^T e^{i\omega(-s+t+v-u)} E(\varepsilon_t \varepsilon_s \varepsilon_u \varepsilon_v).$$

Calculemos então $E(\varepsilon_t \varepsilon_s \varepsilon_u \varepsilon_v)$.

- Se $s = t = u = v$, $E(\varepsilon_t^4) = 3$.

- Se $t = s \neq u = v$ ou $t = u \neq s = v$ ou $t = v \neq s = u$, aquela esperança reduz-se a uma da forma $E(\varepsilon_t^2 \varepsilon_u^2) = 1$.

Então

$$\begin{aligned} E\left([I_T(\omega)]^2\right) &= \frac{1}{T^2} \sum_{s=1}^T E(\varepsilon_t^4) + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T E(\varepsilon_t^2 \varepsilon_u^2) \\ &\quad + \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T E(\varepsilon_t^2 \varepsilon_s^2) \\ &\quad + \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T E(\varepsilon_t^2 \varepsilon_s^2) e^{i\omega(-2s+2t)}, \end{aligned}$$

isto é,

$$\begin{aligned} E\left([I_T(\omega)]^2\right) &= \frac{3}{T} + 2 + \frac{1}{T^2} \sum_{t=1}^T e^{2i\omega t} \sum_{s=1}^T e^{-2i\omega s} \quad (4.2.3.1) \\ &= \frac{3}{T} + 2 + \frac{1}{T^2} \frac{e^{2i\omega} - e^{2i\omega(T+1)}}{1 - e^{2i\omega}} \frac{e^{-2i\omega} - e^{-2i\omega(T+1)}}{1 - e^{-2i\omega}} \\ &= \frac{3}{T} + 2 + \frac{1}{T^2} \frac{1 - e^{2i\omega T}}{1 - e^{2i\omega}} \frac{1 - e^{-2i\omega T}}{1 - e^{-2i\omega}}. \end{aligned}$$

Assim,

$$\lim_{T \rightarrow +\infty} E\left([I_T(\omega)]^2\right) = \lim_{T \rightarrow +\infty} \frac{3}{T} + 2 + \frac{1}{T^2} \frac{1 - e^{2i\omega T}}{1 - e^{2i\omega}} \frac{1 - e^{-2i\omega T}}{1 - e^{-2i\omega}}.$$

Se $\omega \neq 0, +\pi, -\pi$, então

$$\lim_{T \rightarrow +\infty} E\left([I_T(\omega)]^2\right) = 2$$

e

$$\begin{aligned} \lim_{T \rightarrow +\infty} E \left(\left[\widehat{f}_T(\omega) \right]^2 \right) &= \lim_{T \rightarrow +\infty} E \left(\left[\frac{1}{2\pi} I_T(\omega) \right]^2 \right) \\ &= \frac{2}{(2\pi)^2} = 2 [f(\omega)]^2 \end{aligned}$$

uma vez que $f(\omega) = \frac{1}{2\pi}$. Então

$$\begin{aligned} \lim_{T \rightarrow +\infty} V \left(\widehat{f}_T(\omega) \right) &= \lim_{T \rightarrow +\infty} E \left(\left[\widehat{f}_T(\omega) \right]^2 - \left[E \left(\widehat{f}_T(\omega) \right) \right]^2 \right) \\ &= 2 [f(\omega)]^2 - [f(\omega)]^2 = [f(\omega)]^2. \end{aligned}$$

Se $\omega = 0, +\pi, -\pi$, temos que ter em conta o último termo. Vejamos apenas o caso em que $\omega = 0$.

De (4.2.3.1) obtemos

$$E \left([I_T(\omega)]^2 \right) = \frac{3}{T} + 3.$$

Então,

$$\lim_{T \rightarrow +\infty} E \left(\left[\widehat{f}_T(\omega) \right]^2 \right) = \lim_{T \rightarrow +\infty} \frac{1}{(2\pi)^2} \left[\frac{3}{T} + 3 \right] = \frac{3}{(2\pi)^2}$$

e conseqüentemente

$$\lim_{T \rightarrow +\infty} V \left(\widehat{f}_T(\omega) \right) = \frac{3}{(2\pi)^2} - \frac{1}{(2\pi)^2} = \frac{2}{(2\pi)^2} = 2 [f(\omega)]^2.$$

Para $\omega \neq \omega'$,

$$\begin{aligned} \text{Cov} \left(\widehat{f}_T(\omega), \widehat{f}_T(\omega') \right) &= \\ &= E \left(\widehat{f}_T(\omega) \widehat{f}_T(\omega') \right) - E \left(\widehat{f}_T(\omega) \right) E \left(\widehat{f}_T(\omega') \right) \\ &= \frac{1}{4\pi^2} E \left(I_T(\omega) I_T(\omega') \right) - E \left(\widehat{f}_T(\omega) \right) E \left(\widehat{f}_T(\omega') \right). \end{aligned}$$

Mas

$$\begin{aligned}
E(I_T(\omega) I_T(\omega')) &= \\
&= E\left(\frac{1}{T} \left| \sum_{t=1}^T \varepsilon_t e^{i\omega t} \right|^2 \frac{1}{T} \left| \sum_{s=1}^T \varepsilon_s e^{i\omega' s} \right|^2\right) \\
&= \frac{1}{T^2} E\left(\sum_{t=1}^T \varepsilon_t e^{i\omega t} \sum_{u=1}^T \varepsilon_u e^{-i\omega u} \sum_{s=1}^T \varepsilon_s e^{i\omega' s} \sum_{v=1}^T \varepsilon_v e^{-i\omega' v}\right) \\
&= \frac{1}{T^2} E\left(\sum_{t=1}^T \sum_{u=1}^T \sum_{s=1}^T \sum_{v=1}^T \varepsilon_t \varepsilon_u \varepsilon_s \varepsilon_v e^{i\omega(t-u)} e^{i\omega'(s-v)}\right) \\
&= \sum_{t=1}^T \frac{3}{T^2} + \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T E(\varepsilon_t^2) E(\varepsilon_s^2) + \\
&\quad + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T E(\varepsilon_t^2) E(\varepsilon_u^2) e^{i\omega(t-u)} e^{i\omega'(t-u)} + \\
&\quad + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T E(\varepsilon_t^2 \varepsilon_u^2) e^{i\omega(t-u)} e^{i\omega'(u-t)},
\end{aligned}$$

isto é,

$$\begin{aligned}
E(I_T(\omega) I_T(\omega')) &= \\
&= \frac{3}{T} + 1 + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T e^{i\omega(t-u)} e^{i\omega'(t-u)} + \\
&\quad + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T e^{i\omega(t-u)} e^{i\omega'(u-t)} \\
&= \frac{3}{T} + 1 + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T e^{i(\omega+\omega')(t-u)} + \frac{1}{T^2} \sum_{t=1}^T \sum_{u=1}^T e^{i(\omega-\omega')(u-t)} \\
&= \frac{3}{T} + 1 + \frac{1}{T^2} e^{i\omega} \frac{1 - e^{i\omega T}}{1 - e^{i\omega}} e^{-i\omega'} \frac{1 - e^{-i\omega' T}}{1 - e^{-i\omega'}} + \\
&\quad + \frac{1}{T^2} e^{i\omega} \frac{1 - e^{i\omega T}}{1 - e^{i\omega}} e^{-i\omega'} \frac{1 - e^{-i\omega' T}}{1 - e^{-i\omega'}} \\
&= \frac{3}{T} + 1 + \frac{2}{T^2} e^{i(\omega-\omega')} \frac{1 - e^{i\omega T}}{1 - e^{i\omega}} \frac{1 - e^{-i\omega' T}}{1 - e^{-i\omega'}}.
\end{aligned}$$

Quando $T \rightarrow +\infty$, esta expressão tende para 1, uma vez que $\omega \neq \omega'$.

Então

$$\lim_{T \rightarrow +\infty} Cov(\hat{f}_T(\omega), \hat{f}_T(\omega')) = \frac{1}{4\pi^2} - \frac{1}{2\pi} \frac{1}{2\pi} = 0.$$

O periodograma, apesar de ser o estimador natural da densidade espectral, tem um interesse prático limitado pelo seu mau comportamento assintótico. Este mau comportamento é consequência do facto de ser dada a mesma importância (peso) aos estimadores $\hat{\gamma}_T(h)$, quer para pequenos valores de h quer para valores de h próximos de T .

É então interessante considerar estimadores de forma semelhante, mas onde os vários termos são afectados de ponderações diferentes dando, em particular, maior importância a baixos valores de h . Somos conduzidos à estimação da densidade espectral por utilização de ‘janelas móveis’, também denominados ‘geradores de janelas’ ou ‘núcleos’.

Definição. Um gerador de janela é uma função real contínua Φ tal que

- a) $\Phi(x) = 0$ se $x \leq -1$ ou $x \geq 1$
- b) $\Phi(0) = 1$
- c) $\Phi(x) = \Phi(-x)$
- d) Φ decrescente no intervalo $[0, 1]$.

Para melhorar a estimação da densidade espectral retêm-se apenas os valores de $\hat{\gamma}_T(h)$ para h pequeno ($h \leq K_T$) e ponderam-se, usando Φ . Obtêm-se estimadores da forma

$$\hat{f}_T(\omega, K_T, \Phi) = \frac{1}{2\pi} \sum_{h=-K_T}^{K_T} \Phi\left(\frac{h}{K_T}\right) \hat{\gamma}_T(h) e^{i\omega h},$$

onde $K_T \xrightarrow{T \rightarrow +\infty} +\infty$.

Resultados sobre o comportamento destes estimadores, para certos tipos de janelas, podem encontrar-se em Anderson (1971, p. 526).

4.2 Previsão em modelos ARIMA

A previsão de uma série temporal consiste em determinar a sua evolução em períodos não observados. Para tal devemos dispôr de uma série de observações e de uma modelação do fenómeno estudado.

O recurso a um modelo estocástico do tipo *ARIMA* ou *SARIMA* com a finalidade de previsão necessita de um tratamento, cuja sistematização é conhecida por *metodologia de Box e Jenkins*, que engloba as fases seguintes:



A análise exploratória dos dados tem a ver com a observação, a partir da representação gráfica da série inicial, do comportamento geral da série (presença de sazonalidade, tendência, pontos aberrantes, ...). Esta observação preliminar induzir-nos-á à necessidade, ou não, de tornar a série estacionária (segunda fase).

A segunda fase consiste em aplicar à série temporal $\{x_t\}$ transformações geralmente lineares, tais como as anteriormente estudadas, $y_t = (1 - L)^d x_t$ ou $y_t = (1 - L)^d (1 - L^s)^D x_t$, que retirem à série inicial certas características como a tendência ou a sazonalidade.

Para a série estacionária precisamos agora de determinar as ordens p e q das partes auto-regressiva e média móvel, bem como as eventuais ordens de diferenciação. Uma metodologia possível para a obtenção destes parâmetros é dada pela análise do correlograma

e do correlograma parcial à custa do comportamento teórico destas funções em cada tipo de modelo. É a fase de *Identificação* do modelo.

Identificado o modelo torna-se necessário obter os valores estimados dos parâmetros $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ e σ^2 . Recorre-se para isso a métodos da máxima verosimilhança ou do tipo mínimos quadrados. Estamos na fase de *Estimação*. Os métodos usados envolvem as autocorrelações e as autocorrelações parciais empíricas da série.

Põe-se agora a questão de saber se o modelo escolhido na fase de identificação e cujos parâmetros foram estimados pode ser considerado válido, isto é, se se adequa aos dados. Tal verificação recai sobre

- a qualidade estatística do modelo estimado (mediante testes feitos aos parâmetros)

- a análise dos resíduos da estimação, baseada no facto de que, se o modelo é válido, os resíduos deverão apresentar propriedades compatíveis com a hipótese de ruído branco. Na análise dos resíduos um estudo do seu correlograma pode já dar uma resposta parcial (se, por exemplo, os valores das autocorrelações não são significativamente diferentes de zero). Podem também ser construídos testes para precisar esta resposta.

Desta fase de *Validação* podemos avançar directamente para a *previsão* se o modelo é considerado válido ou, no caso contrário, voltar às fases preliminares de análise dos dados e identificação, retomando, por exemplo, um modelo que tenha sido posto de parte na primeira abordagem.

Depois de constatada a validade do modelo estamos de posse de um modelo *ARIMA* (ou *SARIMA*) e podemos então obter previsões de qualidade, para um horizonte j , a partir do instante T , $X_T(j)$, segundo a estratégia apresentada em Gouriéroux, Monfort (1990, cap.VI).

No caso de um modelo *ARMA*, tal estratégia pode ser combinada com a representação de espaços de estado que fornece um critério particularmente simples para actualizar as previsões pois, como vimos, permite relacionar as previsões feitas no instante T (para os horizontes $1, \dots, k-1$) com as previsões feitas no instante

$T + 1$ (e mesmos horizontes).

4.3 Modelação de uma série real

Vamos agora ilustrar a aplicação dos resultados probabilistas e estatísticos associados aos modelos de séries temporais que foram estudados. Dispomos de uma série de dados reais e vamos procurar, de acordo com a metodologia proposta por Box e Jenkins desenvolvida no parágrafo precedente, o modelo (ou modelos) pertencente à classe dos modelos ARIMA que melhor se adequa a tais dados. Neste estudo utilizamos o *software* estatístico EViews, particularmente concebido para a modelação de séries temporais.

Consideremos a série temporal relativa às cotações diárias de fecho das acções da Brisa no mercado Euronext, durante o período de 2 de Janeiro de 2004 a 13 de Maio de 2007.

A trajectória desta série, presente na figura 4.1, mostra uma tendência claramente crescente. Será certamente aconselhável considerar a série das diferenças.

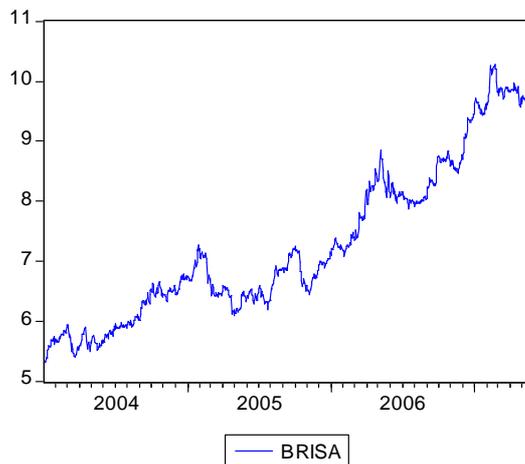


Fig. 4.1 Cotações diárias de fecho da Brisa de 2 de Janeiro de 2004 a 13 de Maio de 2007

A análise descritiva da série Brisa está resumida na figura 4.2 e revela, em particular, que estamos perante uma distribuição clara-

mente não Gaussiana (o p -valor do teste Jarque-Bera é nulo até à sexta casa decimal).

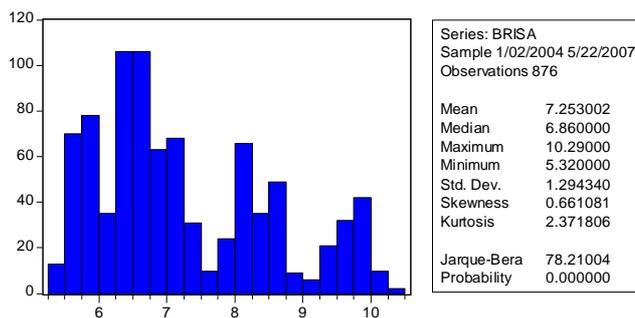


Fig. 4.2 Análise descritiva das cotações diárias de fecho da Brisa

A análise do correlograma da série Brisa, presente na figura 4.3, confirma a necessidade da diferenciação, pois a autocorrelação (parcial) de ordem 1 é muito próxima de 1; além disso, as autocorrelações decrescem muito lentamente.

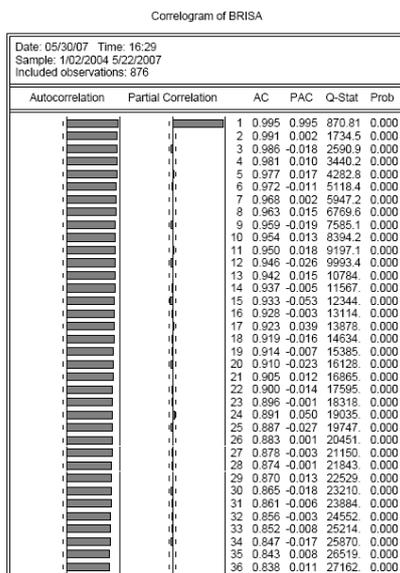


Fig 4.3 Correlograma da série Brisa

Vamos então considerar a série diferenciada. A análise da figura 4.4 mostra que as autocorrelações desta nova série não são significativas, pelo que estaremos na presença de um ruído branco, isto é, de um modelo ARMA com ordens $p = 0$ e $q = 0$.

Correlogram of D(BRISA)

Date: 05/30/07 Time: 16:30 Sample: 1/02/2004 5/22/2007 Included observations: 875						
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.037	-0.037	1.2280	0.268
		2	0.038	0.036	2.4764	0.290
		3	-0.054	-0.051	5.0130	0.171
		4	-0.036	-0.041	6.1623	0.187
		5	-0.008	-0.007	6.2260	0.285
		6	0.034	0.033	7.2210	0.301
		7	0.040	0.040	8.6607	0.278
		8	0.009	0.007	8.7304	0.366
		9	-0.048	-0.048	10.756	0.293
		10	0.008	0.011	10.815	0.372
		11	0.067	0.077	14.846	0.190
		12	-0.024	-0.024	15.355	0.223
		13	0.013	0.000	15.517	0.276
		14	-0.004	0.003	15.532	0.343
		15	-0.031	-0.026	16.369	0.358
		16	-0.051	-0.051	18.706	0.284
		17	0.028	0.022	19.387	0.307
		18	-0.025	-0.029	19.927	0.337
		19	0.005	-0.005	19.946	0.398
		20	0.010	0.017	20.034	0.456
		21	-0.033	-0.037	21.028	0.457
		22	-0.035	-0.039	22.127	0.452
		23	-0.075	-0.069	27.193	0.248
		24	0.031	0.023	28.068	0.257
		25	-0.016	-0.017	28.301	0.294
		26	0.010	0.004	28.397	0.339
		27	0.013	0.015	28.556	0.383
		28	-0.010	-0.010	28.639	0.431
		29	0.016	0.030	28.864	0.472
		30	0.048	0.053	30.944	0.418
		31	-0.040	-0.046	32.393	0.388
		32	0.045	0.040	34.252	0.360
		33	0.011	0.034	34.360	0.402
		34	-0.056	-0.058	37.234	0.322
		35	0.009	0.001	37.314	0.363
		36	-0.026	-0.017	37.944	0.381

Fig. 4.4 - Correlograma da série das diferenças das Cotações diárias de fecho da Brisa

Os parâmetros do modelo estimado encontram-se no quadro 4.5.

Dependent Variable: D(BRISA) Method: Least Squares Date: 05/30/07 Time: 15:57 Sample (adjusted): 1/05/2004 5/11/2007 Included observations: 875 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.004937	0.002354	2.097206	0.0363
R-squared	0.000000	Mean dependent var		0.004937
Adjusted R-squared	0.000000	S.D. dependent var		0.069637
S.E. of regression	0.069637	Akaike info criterion		-2.489906
Sum squared resid	4.238272	Schwarz criterion		-2.484449
Log likelihood	1090.334	Durbin-Watson stat		2.074313

Quadro 4.5 Modelo estimado

Note-se que a constante c pode considerar-se nula, mas apenas para níveis de significância menores que 0.03.

A análise do correlograma dos resíduos associados a este ajustamento, presente na figura 4.6, confirma que a sua estrutura é compatível com a de ruído branco.

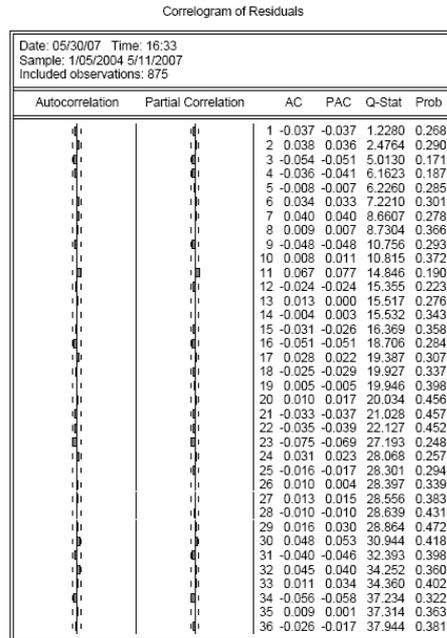


Fig. 4.6 Correlograma dos resíduos da série Brisa diferenciada.

A análise descritiva dos resíduos obtidos, patente na figura 4.7, mostra que a média é praticamente nula e que o desvio padrão é 0.07. A assimetria é ligeiramente positiva e o coeficiente de curtose é superior ao da lei Normal sendo a hipótese de normalidade rejeitada para valores de nível de significância superiores a 10^{-7} .

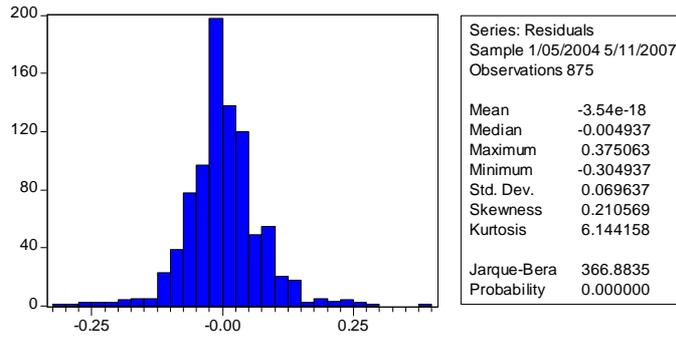


Fig. 4.7 Análise descritiva dos resíduos

O modelo considerado adequado para descrever a série Brisa, X , é então

$$X_t = X_{t-1} + 0.004937 + \varepsilon_t, \quad t \in \mathbb{Z},$$

com $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um ruído branco.

Podemos agora analisar as previsões que tal modelo forneceria para as 7 cotações diárias seguintes, isto é, para o período de 14 a 22 de Maio de 2007.

Na figura 4.8 apresentam-se as previsões obtidas e os correspondentes limites dos intervalos de confiança a 95%. Os valores destas previsões estão próximos de 9.65 com uma ligeira tendência de crescimento.

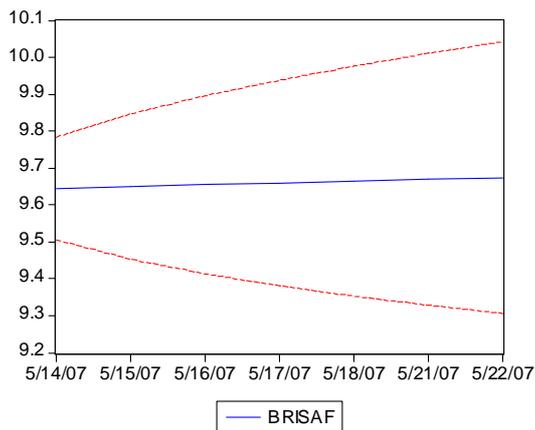


Fig. 4.8 Previsões da cotação de fecho da Brisa no período de 14 a 22/5/2007.

Como curiosidade, acrescentamos que os valores reais da série neste período foram os registados no quadro 4.9.

14/5/2007	15/5/2007	16/5/2007	17/5/2007	18/5/2007	21/5/2007	22/5/2007
9.66	9.66	9.64	9.64	9.65	9.64	9.66

Quadro 4.9 Cotações de fecho da Brisa de 14/5 a 22/5/2007.

Capítulo 5

Modelações não lineares

Os modelos estocásticos de séries temporais foram desenvolvidos com o objectivo de descrever, explicar, prever ou controlar sistemas dinâmicos evoluindo com o tempo. Até ao início dos anos 80 estes estudos foram dominados pelos modelos lineares, com particular ênfase nos modelos ARMA (auto-regressivos médias móveis), que estudamos nos capítulos anteriores. Vimos que a formulação retida para estes modelos permite exprimir o valor presente da série como função linear dos seus valores passados e dos valores presente e passados de um ruído que, sob certas condições de regularidade, se pode interpretar como a inovação da série.

A omissão da não-linearidade na descrição matemática de algumas séries temporais pode levar a consequências indesejáveis, tanto mais que séries com tais características podem encontrar-se em áreas científicas, hoje tão dependentes deste tipo de estudos, como a Econometria, a Matemática Financeira, a Geofísica ou a Astrofísica ... Por exemplo, em Sismologia aparecem por vezes sucessões de registos com picos repentinos seguidos de baixos valores. Também os dados financeiros apresentam frequentemente este tipo de irregularidades; por exemplo, em registos da Bolsa de Valores é frequente observar períodos de grandes oscilações seguidos de períodos de estabilidade.

Estas constatações empíricas levaram ao aparecimento na literatura, essencialmente a partir da segunda metade dos anos 70, de várias classes de modelos, ditos não-lineares, que permitem

descrever tais irregularidades.

Neste texto iremos considerar duas dessas classes, designadamente a dos *modelos condicionalmente heteroscedásticos* e a dos *modelos bilineares*.

5.1 Heteroscedasticidade condicional

Os processos lineares revelam-se, como já observamos, insuficientes para o tratamento de certos problemas financeiros e monetários (índices da bolsa de valores, taxas de juro e de câmbio, inflação, ...) nos quais a variabilidade instantânea (ou volatilidade) das séries de valores associados depende de modo significativo do passado. Esta característica, reconhecida pelo menos desde Mandelbrot (1963) que, sobre dados relativos à inflação e a taxas de juros, afirmou *...large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes ...*, é também constatada por Friedman que, na sua Nobel Lecture (1977), considera que as mais altas taxas de inflação estão geralmente associadas à sua maior variabilidade e, presumivelmente, a uma maior incerteza sobre as taxas futuras.

Assim revela-se indispensável introduzir modelos de séries temporais que tenham, em particular, especificada a evolução no tempo dos seus momentos condicionais.

A figura 5.1, que apresenta uma trajectória do processo $\varepsilon_t = Z_t \sqrt{0.3 + 0.7\varepsilon_{t-1}^2}$, com $(Z_t, t \in \mathbb{Z})$ uma sucessão de v.a.r.i.i.d. de lei $N(0, 1)$ e Z_t independente da σ -álgebra descrevendo o passado, $\underline{\varepsilon}_{t-1}$, evidencia a característica assinalada por Mandelbrot.

Este tipo de séries apresenta também características de não normalidade. De facto, as observações, não correlacionadas ao longo do tempo, são em geral bem descritas por uma distribuição simétrica, unimodal e com caudas mais pesadas que a normal (leptocúrtica).

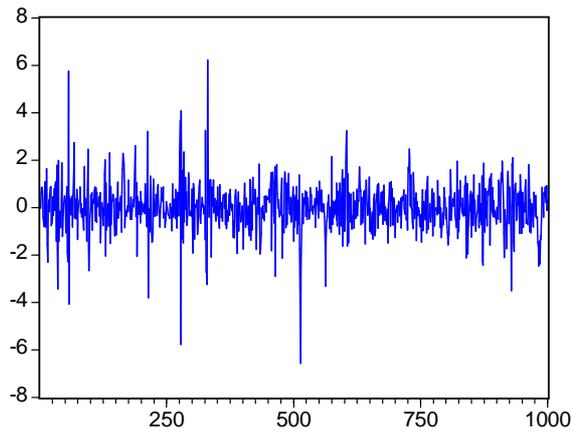


Fig. 5.1 - Simulação do processo $\varepsilon_t = Z_t \sqrt{0.3 + 0.7\varepsilon_{t-1}^2}$

Na figura 5.2, onde se apresenta a trajetória das cotações de fecho (valores diários) das ações da Pararede na Euronext Lisboa entre 2 de Janeiro de 2004 e 17 de Junho de 2005, visualiza-se a presença da referida volatilidade instantânea.

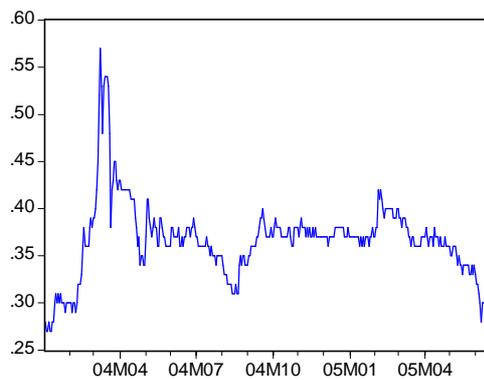


Fig. 5.2 - Valores de fecho diário da cotação das ações da Pararede

Por outro lado, na figura 5.3 ilustra-se a característica de não normalidade de dados desta natureza, com o histograma, associado à distribuição empírica daquela variável, acompanhado dos

valores dos resumos de tal distribuição e do p -valor quase nulo do teste de Jarque-Bera à normalidade do modelo.

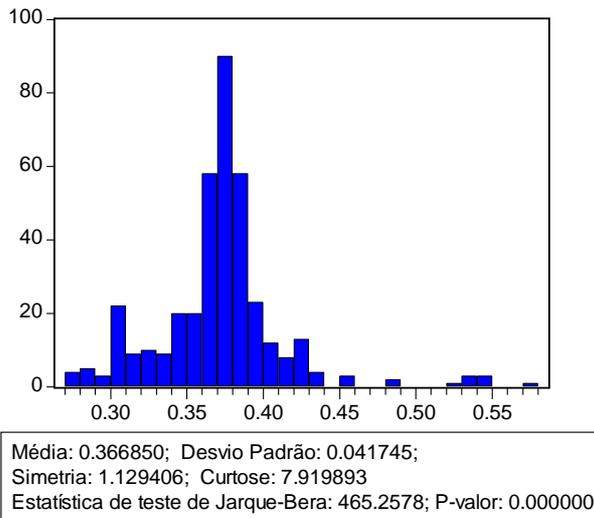


Fig. 5.3 - Histograma das cotações de fecho das acções da Pararede

Com o objectivo de ter em conta tais características na modelação de tal tipo de dados, Engle propõe, em 1982, uma classe de modelos nos quais a variância condicional ao passado do processo no instante t (volatilidade) é uma função linear dos quadrados dos valores do processo anteriores àquele instante t . A heteroscedasticidade (variância não constante no tempo) condicional auto-regressiva presente em tais modelos — facto que justifica o nome que lhes foi atribuído, ARCH — tornou-os particularmente úteis na análise de séries apresentando o comportamento referido; em particular, apresentam propriedades de não correlação ao longo do tempo, de leptocurtose e pequenos (resp. grandes) valores são, em geral, seguidos de pequenos (resp. grandes) valores.

5.1.1 Definição geral e propriedades básicas

As formulações da variância condicional que, para os processos em estudo, se vêm multiplicando na literatura desde Engle (1982),

permitted-nos apresentar uma definição geral de processo de erro condicionalmente heteroscedástico (CH).

Seja $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um processo estocástico real e designemos por $\underline{\varepsilon}_t$ a σ -álgebra gerada por $(\varepsilon_t, \varepsilon_{t-1}, \dots)$, i.e., $\underline{\varepsilon}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, \dots)$.

Definição.

1. Diz-se que $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ segue um *modelo condicionalmente heteroscedástico* de ordens p e q , $\text{CH}(p, q)$, se

$$\begin{cases} E(\varepsilon_t | \underline{\varepsilon}_{t-1}) = 0 \\ V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = H(\varepsilon_{t-1}, \dots, \varepsilon_{t-q}, h_{t-1}, \dots, h_{t-p}) \end{cases}, \forall t \in \mathbb{Z},$$

onde H é uma função real definida sobre \mathbb{R}^{p+q} , estritamente positiva e Borel-mensurável.

2. O processo ε diz-se *condicionalmente Gaussiano* se a lei condicional de ε_t relativamente à tribo $\underline{\varepsilon}_{t-1}$, $\mathcal{L}(\varepsilon_t | \underline{\varepsilon}_{t-1})$, é a lei normal centrada de variância h_t , $N(0, \sqrt{h_t})$.

Começemos por analisar a coerência da definição, mostrando que existem processos compatíveis com a definição anterior.

Tal existência é, em particular, assegurada pelo seguinte resultado:

Teorema. Seja ε um processo estocástico real tal que

$$\forall t \in \mathbb{Z}, \varepsilon_t = \sqrt{h_t} Z_t, \quad (5.1.1.1)$$

com $h_t = H(\varepsilon_{t-1}, \dots, \varepsilon_{t-q}, h_{t-1}, \dots, h_{t-p}) = G(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$, sendo G uma função real, estritamente positiva e mensurável, e $(Z_t, t \in \mathbb{Z})$ uma sucessão de v.a.r.i.i.d., centradas e reduzidas e tais que Z_t é independente de $\underline{\varepsilon}_{t-1}$. O processo ε é solução de um modelo $\text{CH}(p, q)$.

Além disso, ε é um processo condicionalmente Gaussiano se e somente se ε é solução da equação (5.1.1.1), com Z_t v.a. reais cuja lei comum é normal.

Prova. Como $h_t = G(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ e Z_t é independente de $\underline{\varepsilon}_{t-1}$, obtém-se

$$E(\varepsilon_t | \underline{\varepsilon}_{t-1}) = \sqrt{h_t} E(Z_t | \underline{\varepsilon}_{t-1}) = \sqrt{h_t} E(Z_t) = 0.$$

Além disso,

$$\begin{aligned} V(\varepsilon_t | \underline{\varepsilon}_{t-1}) &= E\left(\left(\sqrt{h_t} Z_t\right)^2 \middle| \underline{\varepsilon}_{t-1}\right) = h_t E(Z_t^2 | \underline{\varepsilon}_{t-1}) \\ &= h_t V(Z_t) = h_t. \end{aligned}$$

Relativamente à segunda parte do teorema é óbvio que se a lei das v.a.r. Z_t é normal então $\mathcal{L}(\varepsilon_t | \underline{\varepsilon}_{t-1})$ é normal, centrada e de variância h_t , pois é obtida da lei de Z_t por uma transformação linear (note-se que $\sqrt{h_t}$ é constante para a lei condicional).

Se, por outro lado, ε é um processo condicionalmente Gaussiano então necessariamente ε_t pode explicitar-se na forma indicada no teorema.

De facto, definam-se as v.a.r. $Y_t = \frac{\varepsilon_t}{\sqrt{h_t}}, t \in \mathbb{Z}$, e provemos que estão nas condições do teorema.

i) Começemos por provar que $\mathcal{L}(Y_t | \underline{\varepsilon}_{t-1})$ é normal centrada e reduzida.

$$\begin{aligned} P(Y_t \leq x | \underline{\varepsilon}_{t-1}) &= P\left(Y_t \leq x \middle| \underline{\varepsilon}_{t-1}\right), \text{ com } \underline{\varepsilon}_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \\ &= P\left(\varepsilon_t \leq x \sqrt{h_t} \left(\underline{\varepsilon}_{t-1}\right) \middle| \underline{\varepsilon}_{t-1} = \underline{x}_{t-1}\right), P_{\underline{\varepsilon}_{t-1}} \text{- q.c.} \\ &= F_t\left(x \sqrt{h_t} \left(\underline{x}_{t-1}\right)\right) = F(x), P_{\underline{\varepsilon}_{t-1}} \text{- q.c.}, x \in \mathbb{R}, \end{aligned}$$

onde F_t e F são, respectivamente, as funções de distribuição das leis $N\left(0, \sqrt{h_t} \left(\underline{x}_{t-1}\right)\right)$ e $N(0, 1)$.

ii) Provemos agora que Y_t é independente de $\underline{\varepsilon}_{t-1}, \forall t \in \mathbb{Z}$.

Defina-se, para todos $l \geq q$ e $r \geq p$, o vector aleatório

$$U_{t-1} = (\varepsilon_{t-1}, \dots, \varepsilon_{t-q}, \dots, \varepsilon_{t-l}, h_{t-1}, \dots, h_{t-p}, \dots, h_{t-r}).$$

Quaisquer que sejam $B \in \mathcal{B}(\mathbb{R})$ e $C \in \mathcal{B}(\mathbb{R}^{l+r})$,

$$\begin{aligned} P(Y_t \in B, U_{t-1} \in C) &= \int_{B \times C} P_{Y_t}^{U_{t-1}}(dx) P_{U_{t-1}}(dy) \\ &= \int_{B \times C} N(0, 1)(dx) P_{U_{t-1}}(dy) \\ &= N(0, 1)(B) P_{U_{t-1}}(C), \end{aligned}$$

do teorema de Fubini.

Assim Y_t é independente de U_{t-1} . Mas $\sigma(U_{t-1}) = \underline{\varepsilon}_{t-1}$, uma vez que h_t , sendo a variância de uma lei Gaussiana centrada, caracteriza essa lei, isto é, caracteriza a lei de $\varepsilon_t | \underline{\varepsilon}_{t-1}$. Y_t é, assim, independente de $\underline{\varepsilon}_{t-1}$.

iii) Vejamos finalmente que as v.a.r. (Y_t) são independentes.

Sabemos já que Y_t é independente de $\underline{\varepsilon}_{t-1}$. Além disso, por construção, $Y_t \in \underline{\varepsilon}_t$; logo $\underline{Y}_{t-1} \subset \underline{\varepsilon}_{t-1}, \forall t \in \mathbb{Z}$.

Assim, para todo o $t \in \mathbb{Z}$, Y_t é independente de \underline{Y}_{t-1} , o que implica a independência das v.a.r. (Y_t). ■

Algumas características interessantes destes modelos podem também ser de imediato obtidas como consequência da sua definição:

1. A condição $E(\varepsilon_t | \underline{\varepsilon}_{t-1}) = 0$, que se pode interpretar como uma condição de ortogonalidade ao passado do processo, permite-nos concluir ainda a ortogonalidade e a inexistência de correlações condicionais a qualquer passado do processo. Nomeadamente,

i) $\forall h \in \mathbb{N}, E(\varepsilon_t | \underline{\varepsilon}_{t-h}) = 0$;

ii) $\forall h \in \mathbb{N}, \forall k \in \mathbb{N}, Cov(\varepsilon_t, \varepsilon_{t+h} | \underline{\varepsilon}_{t-k}) = 0$.

De facto, como $\underline{\varepsilon}_{t-h} \subset \underline{\varepsilon}_{t-1}$, para todo o $h \in \mathbb{N}$, vem das propriedades da esperança condicional

$$E(\varepsilon_t | \underline{\varepsilon}_{t-h}) = E(E(\varepsilon_t | \underline{\varepsilon}_{t-1}) | \underline{\varepsilon}_{t-h}) = 0.$$

Quanto à propriedade ii), tem-se

$$\begin{aligned} \text{Cov}(\varepsilon_t, \varepsilon_{t+h} | \underline{\varepsilon}_{t-k}) &= E(\varepsilon_t \varepsilon_{t+h} | \underline{\varepsilon}_{t-k}) \\ &= E(E(\varepsilon_t \varepsilon_{t+h} | \underline{\varepsilon}_{t+h-1}) | \underline{\varepsilon}_{t-k}), \end{aligned}$$

atendendo a que $\underline{\varepsilon}_{t-k} \subset \underline{\varepsilon}_{t+h-1}$, $\forall h \in \mathbb{N}$, $\forall k \in \mathbb{N}$.

Assim, como ε_t é $\underline{\varepsilon}_{t+h-1}$ -mensurável, para todo $h \in \mathbb{N}$, obtém-se

$$\text{Cov}(\varepsilon_t, \varepsilon_{t+h} | \underline{\varepsilon}_{t-k}) = E\left(\underbrace{\varepsilon_t E(\varepsilon_{t+h} | \underline{\varepsilon}_{t+h-1})}_{=0} \middle| \underline{\varepsilon}_{t-k}\right) = 0. \blacksquare$$

2. Note-se que, relativamente à lei condicional, estes processos são centrados e não correlacionados; não são no entanto, em geral, de variância constante (*i.e.*, homoscedásticos), pelo que não são ruídos brancos no sentido clássico. Chamar-lhe-emos genericamente *processos de erro*.

Vejamos agora que tais processos são ruídos brancos, desde que sejam processos estacionários (no sentido fraco).

De facto, por hipótese, $V(\varepsilon_t)$ existe e é independente de t .

Além disso,

$$\begin{aligned} E(\varepsilon_t) &= E(E(\varepsilon_t | \underline{\varepsilon}_{t-1})) = 0 \\ \text{Cov}(\varepsilon_t, \varepsilon_{t+h}) &= E(\varepsilon_t \varepsilon_{t+h}) = E(\varepsilon_t E(\varepsilon_{t+h} | \underline{\varepsilon}_{t+h-1})) = 0, \quad h \in \mathbb{N}. \end{aligned}$$

Finalmente, $\text{Cov}(\varepsilon_t, \varepsilon_{t-h}) = \text{Cov}(\varepsilon_t, \varepsilon_{t+h})$, da hipótese de estacionaridade. \blacksquare

3. Esta família de processos permite modelar processos de erro não gaussianos. Tem-se, em particular, a propriedade seguinte:

Um processo de 2ª ordem condicionalmente Gaussiano não é necessariamente marginalmente Gaussiano.

Com efeito, seja ε um modelo CH(0, 1) condicionalmente Gaussiano de 2ª ordem, isto é,

$$\mathcal{L}(\varepsilon_t | \underline{\varepsilon}_{t-1}) = N(0, \sqrt{h_t}), \quad \text{com } h_t = H(\varepsilon_{t-1}).$$

Suponhamos, por absurdo, que ε_t segue a lei normal centrada de variância σ_t^2 .

Nestas condições o vector aleatório $(\varepsilon_t, \dots, \varepsilon_{t+n})$, $n \in \mathbb{N}$, é um vector absolutamente contínuo de densidade

$$\begin{aligned} f_{\varepsilon_t, \dots, \varepsilon_{t+n}}(x_1, \dots, x_{n+1}) &= f_{\varepsilon_t}(x_1) f_{\varepsilon_{t+1}|\varepsilon_t}(x_2) \dots f_{\varepsilon_{t+n}|\varepsilon_{t+n-1}, \dots, \varepsilon_t}(x_{n+1}) \\ &= \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{1}{2\sigma_t^2}x_1^2} \frac{1}{\sqrt{2\pi h_{t+1}}} e^{-\frac{1}{2h_{t+1}}x_2^2} \dots \frac{1}{\sqrt{2\pi h_{t+n}}} e^{-\frac{1}{2h_{t+n}}x_{n+1}^2} \\ &= \frac{1}{(2\pi)^{\frac{n+1}{2}} \sigma_t \sqrt{h_{t+1}} \dots \sqrt{h_{t+n}}} e^{-\frac{1}{2} \left(\frac{x_1^2}{\sigma_t^2} + \sum_{i=1}^n \frac{x_{i+1}^2}{h_{t+i}} \right)}, \end{aligned}$$

que é a função densidade da lei normal em $(\mathbb{R}^{n+1}, \mathcal{B}(\mathbb{R}^{n+1}))$ centrada e tendo por matriz de covariância a matriz diagonal

$$\Gamma_{t,n} = \begin{bmatrix} \sigma_t^2 & 0 & \dots & 0 \\ 0 & h_{t+1} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & h_{t+n} \end{bmatrix}.$$

O processo ε é, assim, de componentes não correlacionadas pelo que, integrando um vector aleatório Gaussiano, seriam independentes. A variância condicional seria assim independente dos valores passados do processo, o que contradiz a hipótese. ■

4. O aspecto leptocúrtico da distribuição marginal é outra característica destes modelos a qual é, em particular, compatível com as distribuições empíricas de dados de natureza financeira sobretudo obtidos em alta frequência. De facto, a curtose associada a estes processos é, em geral, superior à da lei normal centrada e reduzida.

Suponhamos para o efeito que o processo é condicionalmente Gaussiano e admite momentos até à ordem 4. Então, da normali-

dade da lei condicional e da desigualdade de Schwartz vem

$$\begin{aligned} E(\varepsilon_t^4) &= E(E(\varepsilon_t^4 | \underline{\varepsilon}_{t-1})) = 3E\left[\left(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})\right)^2\right] \\ &\geq 3\left[E(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}))\right]^2 = 3[E(\varepsilon_t^2)]^2, \end{aligned}$$

o que permite concluir que o coeficiente de curtose $K = \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2}$ é superior ou igual ao valor 3 correspondente à curtose da lei normal.

Além disso, podemos escrever K na forma

$$\begin{aligned} K &= \frac{3E\left(\left(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})\right)^2\right)}{[E(\varepsilon_t^2)]^2} \\ &= \frac{3\left[\left(E(\varepsilon_t^2)\right)^2 + E\left(\left(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})\right)^2\right) - \left(E\left(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})\right)\right)^2\right]}{[E(\varepsilon_t^2)]^2} \\ &= 3 + \frac{V[E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})]}{[E(\varepsilon_t^2)]^2} = 3 + \frac{V[h_t]}{[E(\varepsilon_t^2)]^2}, \end{aligned}$$

pelo que estes modelos são leptocúrticos sempre que a função h_t for aleatória, isto é, sempre que estamos perante modelos condicionalmente heteroscedásticos no sentido estrito da definição. ■

A escrita anterior permite ainda observar que a curtose está directamente ligada a uma medida natural da heteroscedasticidade condicional.

5.1.2 Diversas modelações condicionalmente heteroscedásticas

Desde os modelos pioneiros, introduzidos por Engle em 1982 e designados por ARCH, a permanente procura de uma cada vez melhor descrição da realidade levou ao aparecimento, na literatura, de múltiplas generalizações de tais modelos.

A definição destes modelos passa obviamente pela especificação da forma funcional H . Tal especificação permitirá ainda deduzir características particulares de cada um de tais modelos.

Passamos de seguida à apresentação de uma selecção de tais funções que, em nosso entender, ilustram os objectivos subjacentes a tal estudo. Seja então $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um processo estocástico real seguindo um modelo condicionalmente heteroscedástico.

Formulações auto-regressivas

1. Modelos ARCH(q) (Engle, 1982)

Diz-se que ε segue um modelo com *heteroscedasticidade condicional auto-regressiva de ordem q* (ARCH(q)) se,

$$\forall t \in \mathbb{Z}, V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2,$$

com $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$.

Uma consequência imediata desta formulação quadrática é a facilidade com que garantimos a sua positividade.

Outra consequência interessante, e que justifica a designação auto-regressiva associada a estes modelos, é o facto de tal formulação ser equivalente a uma representação auto-regressiva de ordem q para o processo $\varepsilon^2 = (\varepsilon_t^2, t \in \mathbb{Z})$.

De facto, introduzindo a inovação do processo ε^2 , $u_t = \varepsilon_t^2 - E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}^2)$, vem das propriedades da esperança condicional

$$\begin{aligned} u_t &= \varepsilon_t^2 - E[E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}) | \underline{\varepsilon}_{t-1}^2], \text{ pois } \underline{\varepsilon}_{t-1}^2 \subset \underline{\varepsilon}_{t-1}, \\ &= \varepsilon_t^2 - E[h_t | \underline{\varepsilon}_{t-1}^2] = \varepsilon_t^2 - h_t. \end{aligned}$$

Assim

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + u_t,$$

com

$$\begin{aligned} E(u_t) &= E(E(\varepsilon_t^2 - h_t | \underline{\varepsilon}_{t-1})) = E(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}) - E(h_t | \underline{\varepsilon}_{t-1})) \\ &= E(h_t - h_t) = 0. \end{aligned}$$

Note-se, no entanto, que a inovação $(u_t, t \in \mathbb{Z})$, embora sendo centrada e mesmo eventualmente de variância constante, é, em geral, condicionalmente heteroscedástica.

De facto, se o modelo é condicionalmente Gaussiano, temos

$$\begin{aligned} E(u_t^2 | \underline{\varepsilon}_{t-1}) &= E(\varepsilon_t^4 + h_t^2 - 2\varepsilon_t^2 h_t | \underline{\varepsilon}_{t-1}) \\ &= E(\varepsilon_t^4 | \underline{\varepsilon}_{t-1}) + h_t^2 - 2h_t E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}) \\ &= E(\varepsilon_t^4 | \underline{\varepsilon}_{t-1}) - h_t^2 = 2h_t^2, \end{aligned}$$

tendo em conta que, sendo a lei condicional Gaussiana centrada e de variância h_t , $E(\varepsilon_t^4 | \underline{\varepsilon}_{t-1}) = 3(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}))^2 = 3h_t^2$.

Estes modelos foram estudados por Engle, sob a hipótese de normalidade da lei condicional, e utilizados para estimar a média e a variância da taxa de inflação do Reino Unido (Engle (1982)) e dos Estados Unidos (Engle (1983)).

Tais estudos, em particular a representação auto-regressiva anterior, conduziram a resultados compatíveis com a já referida conjectura de Mandelbrot conforme se deduz da frase, presente em Engle (1982), *the ARCH effect is found to be significant and the estimated variances increase substantially during the chaotic seventies*.

Vejamos agora algumas propriedades particulares dos modelos ARCH.

A variância condicional a qualquer passado de um processo ε verificando um modelo ARCH(q) pode escrever-se na forma

$$\begin{aligned} V(\varepsilon_t | \underline{\varepsilon}_{t-h}) &= E(\varepsilon_t^2 | \underline{\varepsilon}_{t-h}) \\ &= f(\varepsilon_{t-h}^2, \varepsilon_{t-h-1}^2, \dots, \varepsilon_{t-h-q+1}^2) + \\ &\quad + g(u_t, u_{t-1}, \dots, u_{t-h+1}), \end{aligned}$$

onde f e g são transformações lineares afins de $\varepsilon_{t-h}^2, \varepsilon_{t-h-1}^2, \dots, \varepsilon_{t-h-q+1}^2$ e de $u_t, u_{t-1}, \dots, u_{t-h+1}$, respectivamente. As expressões destas funções f e g , de complexidade crescente com q , podem obter-se substituindo sucessivamente $\varepsilon_{t-1}^2, \dots, \varepsilon_{t-h+1}^2$.

Deduzamos essas expressões no caso particular em que $q = 1$.

Propriedade. As variâncias condicionais a qualquer passado de um processo ε verificando um modelo ARCH(1) são dadas por

$$\forall h \in \mathbb{N}, V(\varepsilon_t | \underline{\varepsilon}_{t-h}) = \alpha_0 \frac{1 - \alpha_1^h}{1 - \alpha_1} + \alpha_1^h \varepsilon_{t-h}^2.$$

Prova. Usando recursivamente a forma AR, $\varepsilon_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + u_t$, decorre

$$\begin{aligned} \varepsilon_t^2 &= \alpha_0 + \alpha_1 (\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + u_{t-1}) + u_t \\ &= \alpha_0 (1 + \alpha_1) + \alpha_1^2 (\alpha_0 + \alpha_1 \varepsilon_{t-3}^2 + u_{t-2}) + \alpha_1 u_{t-1} + u_t \\ &= \alpha_0 (1 + \alpha_1 + \alpha_1^2) + \alpha_1^3 \varepsilon_{t-3}^2 + \alpha_1^2 u_{t-2} + \alpha_1 u_{t-1} + u_t \\ &= \dots \\ &= \alpha_0 (1 + \alpha_1 + \dots + \alpha_1^{h-1}) + \alpha_1^h \varepsilon_{t-h}^2 + \sum_{j=0}^{h-1} \alpha_1^j u_{t-j} \\ &= \alpha_0 \frac{1 - \alpha_1^h}{1 - \alpha_1} + \alpha_1^h \varepsilon_{t-h}^2 + \sum_{j=0}^{h-1} \alpha_1^j u_{t-j}. \end{aligned}$$

Assim, das propriedades da esperança condicional

$$V(\varepsilon_t | \underline{\varepsilon}_{t-h}) = E(\varepsilon_t^2 | \underline{\varepsilon}_{t-h}) = \alpha_0 \frac{1 - \alpha_1^h}{1 - \alpha_1} + \alpha_1^h \varepsilon_{t-h}^2,$$

já que, para $j \in \{0, 1, \dots, h-1\}$,

$$E(u_{t-j} | \underline{\varepsilon}_{t-h}) = E(E(\varepsilon_{t-j}^2 - h_{t-j} | \underline{\varepsilon}_{t-j-1}) | \underline{\varepsilon}_{t-h}) = 0. \blacksquare$$

O teorema seguinte analisa a propriedade de estacionaridade fraca de um modelo ARCH(q).

Teorema.

1. Se ε é um processo estocástico real estacionário satisfazendo um modelo ARCH(q) então

$$\sum_{i=1}^q \alpha_i < 1.$$

2. Se ε é um processo satisfazendo um modelo ARCH(q) tal que $\sum_{i=1}^q \alpha_i < 1$, então ε é assintoticamente estacionário de 2ª ordem.

Prova.

1. Se o processo é estacionário então é de 2ª ordem e tem-se

$$E(\varepsilon_t^2) = E(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})) = \alpha_0 + \sum_{i=1}^q \alpha_i E(\varepsilon_{t-i}^2);$$

da estacionaridade decorre então

$$\left(1 - \sum_{i=1}^q \alpha_i\right) E(\varepsilon_t^2) = \alpha_0,$$

o que, tendo em conta a positividade da variância e de α_0 , implica $\sum_{i=1}^q \alpha_i < 1$.

2. Suponhamos agora que ε satisfaz um modelo ARCH(q), *i.e.*,

$$V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \quad \text{e} \quad \sum_{i=1}^q \alpha_i < 1.$$

Então, da forma AR obtida para ε^2 e tendo em conta que $E(u_t) = 0$ vem, integrando ambos os membros da igualdade,

$$E(\varepsilon_t^2) = \alpha_0 + \sum_{i=1}^q \alpha_i E(\varepsilon_{t-i}^2).$$

Note-se que os integrais envolvidos existem, podendo não ser finitos, uma vez que temos funções mensuráveis positivas.

Assim, se as raízes do polinómio característico associado à equação homogénea, $\alpha(L) = 1 - \sum_{i=1}^q \alpha_i L^i$, estão no exterior do círculo unitário, a sucessão $E(\varepsilon_t^2)$ converge e então o processo ε é de 2ª ordem e assintoticamente estacionário.

Vejamos então que as raízes de $\alpha(L)$ estão no exterior do círculo unitário se e somente se $\sum_{i=1}^q \alpha_i < 1$.

i) De facto se $\sum_{i=1}^q \alpha_i \geq 1$ então $\alpha(1) \leq 0$.

Mas como $\alpha(0) = 1$, existiria, nestas condições, uma raiz real no intervalo $]0, 1]$, o que contradiz a hipótese.

ii) Por outro lado, se $\sum_{i=1}^q \alpha_i < 1$, suponhamos que existe uma raiz z de $\alpha(L)$ de módulo inferior ou igual a 1.

Nestas condições,

$$1 = \sum_{i=1}^q \alpha_i z^i = \left| \sum_{i=1}^q \alpha_i z^i \right| \leq \sum_{i=1}^q \alpha_i |z|^i \leq \sum_{i=1}^q \alpha_i < 1,$$

o que é uma contradição. ■

Como consequência do teorema anterior deduzimos que um processo estocástico estacionário verificando um modelo ARCH(q) é centrado e tem variância dada por

$$V(\varepsilon_t) = \frac{\alpha_0}{1 - \sum_{i=1}^q \alpha_i}, \quad \text{com } \sum_{i=1}^q \alpha_i < 1.$$

O aspecto leptocúrtico dos modelos ARCH é analisado de modo simples nos modelos condicionalmente Gaussianos de ordem 1.

Propriedade. Se ε é um processo estocástico real satisfazendo um modelo ARCH(1) condicionalmente Gaussiano tal que

$$3\alpha_1^2 < 1$$

então o seu coeficiente de curtose K é dado por

$$K = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2} > 3.$$

Prova. Se $3\alpha_1^2 < 1$ então $\alpha_1 < 1$, pelo que o processo é assintoticamente estacionário de 2ª ordem tendo-se

$$V(\varepsilon_t) = \frac{\alpha_0}{1 - \alpha_1}.$$

Por outro lado, como o modelo é condicionalmente Gaussiano,

$$\begin{aligned} E(\varepsilon_t^4 | \underline{\varepsilon}_{t-1}) &= 3(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}))^2 \\ &= 3(\alpha_0^2 + \alpha_1^2 \varepsilon_{t-1}^4 + 2\alpha_0 \alpha_1 \varepsilon_{t-1}^2). \end{aligned}$$

Integrando ambos os membros da igualdade anterior de funções mensuráveis não negativas, vem

$$E(\varepsilon_t^4) = 3\alpha_0^2 + 3\alpha_1^2 E(\varepsilon_{t-1}^4) + \frac{6\alpha_0^2 \alpha_1}{1 - \alpha_1}.$$

Assim, se as raízes do polinómio característico associado $\alpha(L) = 1 - 3\alpha_1^2 L$ estão no exterior do círculo unitário, *i.e.*, se $3\alpha_1^2 < 1$, a sucessão $E(\varepsilon_t^4)$ converge e o processo ε tem momento de 4ª ordem constante.

Assim

$$E(\varepsilon_t^4) = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)},$$

o que permite concluir. ■

A representação auto-regressiva de ordem q obtida para $\varepsilon^2 = (\varepsilon_t^2, t \in \mathbb{Z})$ permite ainda obter, quando este processo é estacionário, uma equação de recorrência linear para a respectiva função de autocovariância, designadamente,

$$\forall k \in \mathbb{N}, \gamma(k) = Cov(\varepsilon_t^2, \varepsilon_{t-k}^2) = \sum_{i=1}^q \alpha_i \gamma(k - i).$$

Este sistema de equações pode ser usado na identificação do modelo e, em particular, na determinação da ordem de atrasos q .

2. Modelos ARCH(q) exponenciais e em valor absoluto (Engle, 1982)

No trabalho pioneiro de Engle (1982) surgem ainda outras formulações para a equação de evolução da variância condicional. Destas destacamos os modelos exponenciais e os de valor absoluto.

Diz-se que ε segue um modelo ARCH(q) *exponencial* se

$$\forall t \in \mathbb{Z}, V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \exp \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right),$$

onde $\alpha_0, \alpha_1, \dots, \alpha_q$ são reais arbitrários.

Este modelo apresenta a vantagem da variância condicional ser positiva para todos os valores dos parâmetros, mas conduz a processos que não têm variância finita.

Em contrapartida os modelos de valor absoluto, que vamos definir de seguida, exigem tal como os ARCH originais a positividade dos coeficientes, mas apresentam, segundo Engle, a vantagem de conduzir a processos de variância finita quaisquer que sejam os valores admissíveis de tais coeficientes.

Diz-se que ε segue um modelo ARCH(q) *de valor absoluto* se

$$\forall t \in \mathbb{Z}, V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i |\varepsilon_{t-i}|,$$

onde $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$.

Uma generalização dos modelos ARCH(q) exponenciais foi apresentada por Nelson (1989). De facto, este autor introduz os processos ARCH *exponenciais* (E-ARCH) nos quais a variância condicional h_t é tal que

$$\forall t \in \mathbb{Z}, \log h_t = \alpha_t + \sum_{i=1}^{+\infty} \beta_i g(Z_{t-i}),$$

onde as variáveis aleatórias Z_t são independentes e identicamente distribuídas e Z_t é independente de ε_{t-1} , para todo $t \in \mathbb{Z}$, e tal que $\varepsilon_t = \sqrt{h_t}Z_t$; além disso $(\alpha_t, t \in \mathbb{Z})$ e $(\beta_t, t \in \mathbb{Z})$ são sucessões de números reais e g uma função que permita ter em conta quer a grandeza quer o sinal de Z_t .

Nelson introduz estes modelos a fim de dar uma resposta a críticas, por ele próprio levantadas, às principais formulações condicionalmente heteroscedásticas surgidas até àquela data na literatura. Com efeito, tais formulações, sendo quadráticas, não permitem ter em conta o sinal do processo na evolução futura da sua variância condicional; por outro lado, as restrições de positividade dos coeficientes, que em geral devem ser consideradas, provocam problemas numéricos na fase da correspondente estimação.

3. Modelos ARCH não lineares de ordem q (NARCH(q)) (Higgins and Bera, 1990)

Diz-se que ε segue um modelo ARCH *não linear de ordem q* (NARCH(q)) se

$$\forall t \in \mathbb{Z}, h_t = \left[\alpha_0 (\sigma^2)^\delta + \sum_{i=1}^q \alpha_i (\varepsilon_{t-i}^2)^\delta \right]^{\frac{1}{\delta}},$$

onde $\sigma^2 > 0$, $\alpha_i \geq 0, i = 0, \dots, q, \delta > 0$ e $\sum_{i=1}^q \alpha_i < 1$.

É interessante notar que a função presente nesta formulação da variância condicional é considerada em Economia a função de produção com elasticidade de substituição constante (Arrow and al, 1961). Nota-se que a elasticidade de substituição, que vale $\frac{1}{1-\delta}$, toma valores no intervalo $]1, +\infty[$ desde que $\delta \in]0, 1[$.

Ora, quando δ é igual a 1, reencontramos uma formulação ARCH(q). Além disso, se ε é estacionário

$$V(\varepsilon_t) = \frac{\alpha_0 \sigma^2}{1 - \sum_{i=1}^q \alpha_i},$$

o que, substituindo na definição de h_t , conduz a

$$V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \left(1 - \sum_{i=1}^q \alpha_i\right) V(\varepsilon_t) + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2$$

e permite interpretar a variância condicional como uma média ponderada da variância “global” e das variâncias “locais” $\varepsilon_{t-1}^2, \dots, \varepsilon_{t-q}^2$.

Assim se, de forma heurística, a variância condicional for encarada como *output* e as variâncias “global” e “locais” como *input*, então o modelo ARCH(q) tem a forma de uma função de produção linear com elasticidade de substituição infinita.

4. Modelos threshold ARCH de ordem q (TARCH(q))

Como já observamos, os modelos considerados apresentam uma importante limitação resultante da formulação quadrática retida para a expressão da variância condicional. Com efeito, a influência dos valores passados do processo na variabilidade instantânea da série não tem em conta os seus sinais, mas apenas os seus valores absolutos. Ora, em particular no caso das séries financeiras, surgem frequentemente assimetrias na variabilidade; por exemplo, a variabilidade é em geral maior depois de um decréscimo do que de um acréscimo de igual valor absoluto.

Os modelos threshold ARCH (ARCH com limiares), introduzidos em Zakoian (1990), pretendem dar uma resposta a este problema.

Diz-se que ε segue um modelo *threshold* ARCH de ordem q ou ARCH com limiares de ordem q (TARCH(q)) se,

$$\forall t \in \mathbb{Z}, V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \left(\alpha_0 + \sum_{i=1}^q \alpha_{i,+} \varepsilon_{t-i}^+ - \sum_{i=1}^q \alpha_{i,-} \varepsilon_{t-i}^- \right)^2,$$

com $\alpha_0 > 0$, $\alpha_{i,+} \geq 0$, $\alpha_{i,-} \geq 0$, $i = 1, \dots, q$, e onde $\varepsilon_l^+ = \varepsilon_l \mathbb{I}_{\{\varepsilon_l > 0\}}$ e $\varepsilon_l^- = \varepsilon_l \mathbb{I}_{\{\varepsilon_l < 0\}}$, $l \in \mathbb{Z}$.

Estes modelos apresentam características idênticas às dos ARCH já que pequenos (resp. grandes) valores do processo são geralmente

seguidos por pequenos (resp. grandes) valores. Têm no entanto a vantagem de permitir ter em conta o sinal do processo.

Pode também obter-se uma representação auto-regressiva para uma função do processo (vector de espaço de estados). Com efeito, em Zakoian (1990) são dadas condições para que o processo estocástico de \mathbb{R}^{4q} , $w = (w_t, t \in \mathbb{Z})$, tal que

$$w'_t = [(\varepsilon_t^+)^2, (\varepsilon_t^-)^2, \varepsilon_t^+ \varepsilon_{t-1}^+, \varepsilon_t^+ \varepsilon_{t-1}^-, \varepsilon_t^- \varepsilon_{t-1}^+, \varepsilon_t^- \varepsilon_{t-1}^-, \dots, \varepsilon_t^+ \varepsilon_{t-q+1}^+, \varepsilon_t^+ \varepsilon_{t-q+1}^-, \varepsilon_t^- \varepsilon_{t-q+1}^+, \varepsilon_t^- \varepsilon_{t-q+1}^-, \varepsilon_t^+, \varepsilon_t^-]$$

verifique a equação de recorrência

$$w_t = b + \sum_{i=1}^q B_i w_{t-i} + \eta_t$$

onde $(\eta_t, t \in \mathbb{Z})$ é um processo vectorial tal que $E(\eta_t | \underline{w}_{t-1}) = 0$, b é um vector de dimensão $4q$ e B_i ($i = 1, \dots, q$) são matrizes quadradas de ordem $4q$.

Analisemos tal representação e algumas das suas implicações no caso particular de um modelo TARCH(1) condicionalmente Gaussiano.

Consideremos então um processo estocástico ε tal que

$$\begin{cases} \varepsilon_t = \sqrt{h_t} Z_t \\ h_t = (\alpha_0 + \alpha \varepsilon_{t-1}^+ - \beta \varepsilon_{t-1}^-)^2 \end{cases},$$

onde $(Z_t, t \in \mathbb{Z})$ é uma família de variáveis aleatórias independentes, identicamente distribuídas, normais, centradas e reduzidas tais que, para todo o $t \in \mathbb{Z}$, Z_t é independente de $\underline{\varepsilon}_{t-1}$.

O processo w é agora tal que

$$w_t = [(\varepsilon_t^+)^2, (\varepsilon_t^-)^2, \varepsilon_t^+, \varepsilon_t^-]'$$

Ora, nas condições acima consideradas, e tendo em conta a simetria da lei de $\varepsilon_t | \underline{\varepsilon}_{t-1}$, vem

$$E\left((\varepsilon_t^+)^2 | \underline{\varepsilon}_{t-1}\right) = E\left((\varepsilon_t^-)^2 | \underline{\varepsilon}_{t-1}\right) = \frac{1}{2} E\left(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}\right)$$

$$\begin{aligned}
&= \frac{1}{2} (\alpha_0 + \alpha \varepsilon_{t-1}^+ - \beta \varepsilon_{t-1}^-)^2 \\
&= \frac{1}{2} \alpha_0^2 + \alpha_0 \alpha \varepsilon_{t-1}^+ - \alpha_0 \beta \varepsilon_{t-1}^- + \frac{1}{2} \alpha^2 (\varepsilon_{t-1}^+)^2 + \frac{1}{2} \beta^2 (\varepsilon_{t-1}^-)^2
\end{aligned}$$

pois $\mathcal{L}(\varepsilon_t^+ | \underline{\varepsilon}_{t-1}) = \mathcal{L}(-\varepsilon_t^- | \underline{\varepsilon}_{t-1})$, $\varepsilon_t = \varepsilon_t^+ + \varepsilon_t^-$ e $\varepsilon_t^+ \varepsilon_t^- = 0$.
Além disso,

$$\begin{aligned}
E(\varepsilon_t^+ | \underline{\varepsilon}_{t-1}) &= -E(\varepsilon_t^- | \underline{\varepsilon}_{t-1}) = \frac{1}{\sqrt{2\pi}} \sqrt{h_t} \\
&= \frac{1}{\sqrt{2\pi}} (\alpha_0 + \alpha \varepsilon_{t-1}^+ - \beta \varepsilon_{t-1}^-).
\end{aligned}$$

Por outro lado, tendo em conta que $\underline{\varepsilon}_{t-1} = \underline{w}_{t-1}$, tem-se

$$E(w_t | \underline{w}_{t-1}) = \begin{pmatrix} E((\varepsilon_t^+)^2 | \underline{w}_{t-1}) \\ E((\varepsilon_t^-)^2 | \underline{w}_{t-1}) \\ E(\varepsilon_t^+ | \underline{w}_{t-1}) \\ E(\varepsilon_t^- | \underline{w}_{t-1}) \end{pmatrix} = \begin{pmatrix} E((\varepsilon_t^+)^2 | \underline{\varepsilon}_{t-1}) \\ E((\varepsilon_t^-)^2 | \underline{\varepsilon}_{t-1}) \\ E(\varepsilon_t^+ | \underline{\varepsilon}_{t-1}) \\ E(\varepsilon_t^- | \underline{\varepsilon}_{t-1}) \end{pmatrix}.$$

Assim, obtém-se facilmente

$$E(w_t | \underline{w}_{t-1}) = \begin{bmatrix} \frac{1}{2} \alpha^2 & \frac{1}{2} \beta^2 & \alpha_0 \alpha & -\alpha_0 \beta \\ \frac{1}{2} \alpha^2 & \frac{1}{2} \beta^2 & \alpha_0 \alpha & -\alpha_0 \beta \\ 0 & 0 & \frac{\alpha}{\sqrt{2\pi}} & -\frac{\beta}{\sqrt{2\pi}} \\ 0 & 0 & -\frac{\alpha}{\sqrt{2\pi}} & \frac{\beta}{\sqrt{2\pi}} \end{bmatrix} w_{t-1} + \begin{bmatrix} \frac{1}{2} \alpha_0^2 \\ \frac{1}{2} \alpha_0^2 \\ \frac{1}{\sqrt{2\pi}} \alpha_0 \\ -\frac{1}{\sqrt{2\pi}} \alpha_0 \end{bmatrix},$$

de onde decorre a representação auto-regressiva anunciada considerando o processo de inovação η , tal que

$$\forall t \in \mathbb{Z}, \eta_t = w_t - E(w_t | \underline{w}_{t-1}). \blacksquare$$

A representação auto-regressiva anterior permite-nos analisar as condições de estacionaridade de 2ª ordem de um tal modelo.

Teorema.

1. Se ε é um processo estocástico real estacionário satisfazendo um modelo TARCH(1) condicionalmente Gaussiano então

$$\alpha^2 + \beta^2 < 2.$$

2. Seja ε um processo estocástico real seguindo um modelo TARCH(1) condicionalmente Gaussiano tal que

$$\alpha^2 + \beta^2 < 2.$$

Nestas condições ε é assintoticamente estacionário de 2ª ordem.

Prova.

1. Se o processo é estacionário então é de 2ª ordem e tem-se

$$\begin{aligned} E(\varepsilon_t^2) &= E(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})) = E(E(h_t Z_t^2 | \underline{\varepsilon}_{t-1})) \\ &= E(h_t E(Z_t^2 | \underline{\varepsilon}_{t-1})) = E(h_t), \end{aligned}$$

uma vez que $E(Z_t^2 | \underline{\varepsilon}_{t-1}) = E(Z_t^2) = 1$.

Mas

$$h_t = (\alpha_0 + \alpha \varepsilon_{t-1}^+ - \beta \varepsilon_{t-1}^-)^2 = \left(\alpha_0 + (\alpha Z_{t-1}^+ - \beta Z_{t-1}^-) \sqrt{h_{t-1}} \right)^2,$$

pelo que

$$\begin{aligned} E(h_t) &= \alpha_0^2 + E \left[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)^2 h_{t-1} \right] \\ &\quad + 2\alpha_0 E \left[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-) \sqrt{h_{t-1}} \right] \\ &= \alpha_0^2 + E \left[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)^2 \right] E[h_{t-1}] \\ &\quad + 2\alpha_0 E \left[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-) \right] E \left[\sqrt{h_{t-1}} \right], \end{aligned}$$

tendo em conta que, para todo o $l \in \mathbb{Z}$, Z_l é independente de $\underline{\varepsilon}_{l-1}$.

Da estacionaridade de ε , decorre então

$$\left(1 - E \left[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)^2 \right] \right) E(h_t)$$

$$= \alpha_0^2 + 2\alpha_0 E[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)] E[\sqrt{h_{t-1}}],$$

o que implica, tendo em conta a positividade de $\alpha_0^2 + 2\alpha_0 E[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)] E[\sqrt{h_{t-1}}]$,

$$E[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)^2] < 1.$$

Mas, prova-se facilmente que

$$E[(\alpha Z_{t-1}^+ - \beta Z_{t-1}^-)^2] = (\alpha^2 + \beta^2) E(Z_{t-1}^+)^2 = \frac{(\alpha^2 + \beta^2)}{2}.$$

2. Suponhamos agora que ε satisfaz um modelo TARCH(1) condicionalmente Gaussiano tal que $\alpha^2 + \beta^2 < 2$.

Da expressão obtida anteriormente para $E(w_t | \underline{w}_{t-1})$, onde w é o processo vectorial de espaço de estados associado a ε , deduz-se que para todo $t \in \mathbb{Z}$,

$$E(w_t) = \begin{bmatrix} \frac{1}{2}\alpha^2 & \frac{1}{2}\beta^2 & \alpha_0\alpha & -\alpha_0\beta \\ \frac{1}{2}\alpha^2 & \frac{1}{2}\beta^2 & \alpha_0\alpha & -\alpha_0\beta \\ 0 & 0 & \frac{\alpha}{\sqrt{2\pi}} & -\frac{\beta}{\sqrt{2\pi}} \\ 0 & 0 & -\frac{\alpha}{\sqrt{2\pi}} & \frac{\beta}{\sqrt{2\pi}} \end{bmatrix} E(w_{t-1}) + \begin{bmatrix} \frac{1}{2}\alpha_0^2 \\ \frac{1}{2}\alpha_0^2 \\ \frac{1}{\sqrt{2\pi}}\alpha_0 \\ -\frac{1}{\sqrt{2\pi}}\alpha_0 \end{bmatrix},$$

pelo que a sucessão $(E(w_t), t \in \mathbb{Z})$ satisfaz uma equação de recorrência linear vectorial.

Designando por A a matriz factor de $E(w_{t-1})$, temos que se as raízes do polinómio $\tilde{A}(z) = \det(I - Az)$, associado à equação homogénea, estão todas no exterior do disco unitário, então a sucessão $E(w_t)$ converge (Gouriéroux, Monfort, 1990, cap. VIII).

Da expressão do vector $E(w_t)$ decorre então que, naquelas condições, o processo ε é de 2^α ordem e assintoticamente estacionário.

Mas, da definição de valor próprio da matriz A decorre que as raízes do polinómio $\tilde{A}(z)$ estão todas no exterior do disco unitário se e só se os valores próprios da matriz A forem em módulo estritamente menores do que 1.

Tais valores próprios são as soluções da equação $\det(A - \lambda I) = 0$. Ora a matriz $A - \lambda I$ é a seguinte matriz tri-

angular por blocos

$$A - \lambda I = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

com

$$A_{11} = \begin{bmatrix} \frac{1}{2}\alpha^2 - \lambda & \frac{1}{2}\beta^2 \\ \frac{1}{2}\alpha^2 & \frac{1}{2}\beta^2 - \lambda \end{bmatrix}; \quad A_{22} = \begin{bmatrix} \frac{\alpha}{\sqrt{2\pi}} - \lambda & -\frac{\beta}{\sqrt{2\pi}} \\ -\frac{\alpha}{\sqrt{2\pi}} & \frac{\beta}{\sqrt{2\pi}} - \lambda \end{bmatrix};$$

$$A_{12} = \begin{bmatrix} \alpha_0\alpha & -\alpha_0\beta \\ \alpha_0\alpha & -\alpha_0\beta \end{bmatrix}; \quad A_{21} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Assim

$$\det(A - \lambda I) = 0 \iff \det A_{11} \det A_{22} = 0.$$

A resolução desta equação conduz aos seguintes valores próprios

$$\left(0, 0, \frac{\alpha^2 + \beta^2}{2}, \frac{\alpha + \beta}{\sqrt{2\pi}} \right).$$

A condição de estabilidade de solução é então equivalente à verificação do sistema

$$\begin{cases} \alpha^2 + \beta^2 < 2 \\ \alpha + \beta < \sqrt{2\pi} \end{cases},$$

isto é, equivalente a $\alpha^2 + \beta^2 < 2$, pois a segunda desigualdade decorre como consequência. ■

Do teorema anterior podemos concluir que se $\alpha^2 + \beta^2 < 2$ então ε é centrado e tem variância dada por

$$V(\varepsilon_t) = \frac{2\alpha_0^2}{2 - (\alpha^2 + \beta^2)} \frac{\sqrt{2\pi} + \alpha + \beta}{\sqrt{2\pi} - (\alpha + \beta)}.$$

A dedução dessa variância envolve cálculos matriciais pesados que não vamos aqui apresentar em detalhe. Indiquemos apenas o processo de cálculo que deverá seguir-se.

Se $\alpha^2 + \beta^2 < 2$, decorre da equação de recorrência linear para a sucessão $(E(w_t), t \in \mathbb{Z})$

$$(I - AL)E(w_t) = b \iff E(w_t) = (I - AL)^{-1}b,$$

onde $b = \left(\frac{1}{2}\alpha_0^2 \quad \frac{1}{2}\alpha_0^2 \quad \frac{1}{\sqrt{2\pi}}\alpha_0 \quad -\frac{1}{\sqrt{2\pi}}\alpha_0 \right)'$ e L é o operador atraso actuando sobre t .

Mas como L é invariante para funções constantes vem que

$$(I - AL)^{-1} b = \left(\sum_{i=0}^{+\infty} \Phi_i L^i \right) b = \sum_{i=0}^{+\infty} \Phi_i b = (I - A)^{-1} b,$$

isto é,

$$E(w_t) = \begin{pmatrix} E\left((\varepsilon_t^+)^2\right) \\ E\left((\varepsilon_t^-)^2\right) \\ E\left(\varepsilon_t^+\right) \\ E\left(\varepsilon_t^-\right) \end{pmatrix} = (I - A)^{-1} b.$$

Assim tendo em conta que

$$V(\varepsilon_t) = E(\varepsilon_t^2) = E\left((\varepsilon_t^+)^2\right) + E\left((\varepsilon_t^-)^2\right),$$

o cálculo da variância decorre da inversão da matriz

$$I - A = \begin{bmatrix} 1 - \frac{1}{2}\alpha^2 & \frac{1}{2}\beta^2 & \alpha_0\alpha & -\alpha_0\beta \\ \frac{1}{2}\alpha^2 & 1 - \frac{1}{2}\beta^2 & \alpha_0\alpha & -\alpha_0\beta \\ 0 & 0 & 1 - \frac{\alpha}{\sqrt{2\pi}} & -\frac{\beta}{\sqrt{2\pi}} \\ 0 & 0 & -\frac{\alpha}{\sqrt{2\pi}} & 1 - \frac{\beta}{\sqrt{2\pi}} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

que representamos por blocos com notações evidentes.

Usando o resultado de inversão de uma matriz por blocos (Gouriéroux, Monfort, 1989, vol.II, p.459), e tendo em conta a forma triangular obtida, vem

$$(I - A)^{-1} = \begin{bmatrix} B_{11}^{-1} & -B_{11}^{-1}B_{12}B_{22}^{-1} \\ 0 & B_{22}^{-1} \end{bmatrix},$$

cujos cálculos conduzem ao valor pretendido.

Note-se que nos resultados anteriormente apresentados tem papel determinante a caracterização obtida para o processo no caso de lei condicional Gaussiana. É fácil concluir que podem ser obtidos resultados análogos com leis condicionais mais gerais desde

que se assuma que o processo é solução de uma equação do tipo $\varepsilon_t = \sqrt{h_t}Z_t$, onde $(Z_t, t \in \mathbb{Z})$ é uma família de variáveis aleatórias independentes, identicamente distribuídas, centradas e reduzidas tais que, para todo o $t \in \mathbb{Z}$, Z_t é independente de $\underline{\varepsilon}_{t-1}$.

Mediante as condições de existência dos momentos de ordem quatro da lei marginal de ε_t , pode avaliar-se o valor da curtose de tais distribuições bem como a sua evolução em função dos parâmetros de heteroscedasticidade. Zakoian (1990) detalha este estudo no caso de um modelo TARCH(1) condicionalmente Gaussiano.

Formulações auto-regressivas generalizadas

1. Modelos ARCH generalizados de ordens p e q (GARCH(p, q)) (Bollerslev, 1986)

As formulações da variância condicional que temos vindo a apresentar sendo baseadas em relações auto-regressivas podem conduzir, na prática, a uma ordem de atrasos bastante grande. Para minimizar tal problema, Bollerslev (1986) propõe uma generalização dos modelos ARCH onde é considerada uma dinâmica auto-regressiva média móvel. De facto, Bollerslev faz depender a variância condicional não só do passado do processo mas também do seu próprio passado; os modelos consequentes foram designados por ARCH generalizados ou, de forma abreviada, por GARCH.

Diz-se que $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ segue um modelo ARCH *generalizado de ordens p e q* (GARCH(p, q)) se a sua variância condicional é tal que

$$\forall t \in \mathbb{Z}, V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j},$$

onde $\alpha_0 > 0, \alpha_i \geq 0, i = 1, \dots, q, \beta_j \geq 0, j = 1, \dots, p$.

Esta formulação permite-nos reescrever a variância condicional de modo a obter uma representação que descreva a evolução do processo no tempo.

De facto, considerando a inovação correspondente ao quadrado do processo, isto é, o processo

$$u_t = \varepsilon_t^2 - E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}) = \varepsilon_t^2 - E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1}) = \varepsilon_t^2 - h_t, t \in \mathbb{Z},$$

vem

$$\begin{aligned} \varepsilon_t^2 &= \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \varepsilon_{t-j}^2 - \sum_{j=1}^p \beta_j u_{t-j} + u_t \\ &= \alpha_0 + \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 - \sum_{j=1}^p \beta_j u_{t-j} + u_t, \end{aligned}$$

onde $\alpha_i = 0$, $i > q$ e $\beta_i = 0$, $i > p$.

A representação encontrada é, para o processo ε^2 , do tipo ARMA($\max(p, q), p$); a inovação u não é, tal como vimos no caso dos modelos ARCH, necessariamente homoscedástica.

Esta representação permite estudar a estacionaridade de 2ª ordem de tais modelos.

Analogamente ao que vimos no caso dos modelos ARCH mostra-se, (Gouriéroux (1992), p.44), que este modelo é assintoticamente estacionário no sentido fraco se os parâmetros envolvidos na definição são tais que

$$\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j < 1.$$

Além disso, também de modo análogo ao visto no caso dos modelos ARCH e TARCH, pode mostrar-se a necessidade desta condição para a estacionaridade de um processo seguindo um modelo GARCH.

A representação ARMA obtida conduz ainda a uma equação de recorrência para a função de autocovariância de ε^2 , quando ε é estacionário,

$$\gamma(k) = \sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) \gamma(k-i), k \geq p+1,$$

que foi usada por Bollerslev(1986, 1988) na identificação do modelo e, em particular, na determinação das ordens p e q se $q \geq p$ e somente p no caso contrário.

2. Modelos ARCH em média de ordem q (ARCH-M(q))
(Engle, Lilien and Robbins, 1987)

Os modelos ARCH em média ou ARCH-M (Engle, Lilien and Robins (1987)), nos quais a variância condicional aparece como variável explicativa da média condicional, poderão estar mais adaptados à descrição da influência da volatilidade no modelo.

Seja ε um processo estocástico real seguindo um modelo ARCH(q). Diz-se que um processo estocástico $Y = (Y_t, t \in \mathbb{Z})$ segue um *modelo ARCH em média de ordem q* (ARCH-M(q)) se

$$\begin{cases} Y_t = X_t b + \delta h_t + \varepsilon_t \\ h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \end{cases} ,$$

onde $X_t, t \in \mathbb{Z}$, são variáveis aleatórias reais explicativas.

Uma generalização imediata de tais modelos são aqueles em que o processo ε segue um modelo GARCH(p, q).

3. Modelos TARCH generalizados de ordens p e q
(GTARCH(p, q)) (Rabemananjara and Zakoian, 1992)

Seguindo a ideia de Bollerslev (1986), é proposta em Rabemananjara e Zakoian (1993) uma generalização dos modelos TARCH, os modelos GTARCH, nos quais a variância condicional tem também prevista uma dinâmica endógena.

Os processos GARCH com limiares de ordens p e q , ou GTARCH(p, q), têm como principal objectivo modelar assimetrias na variabilidade tendo em conta, de modo distinto, a influência de valores simétricos do processo subjacente.

Diz-se então que um processo ε segue um *modelo TARCH generalizado de ordens p e q* (GTARCH(p, q)) se admite como variân-

cia condicional a função

$$\forall t \in \mathbb{Z}, h_t = \left(\alpha_0 + \sum_{i=1}^q \alpha_{i,+} \varepsilon_{t-i}^+ - \sum_{i=1}^q \alpha_{i,-} \varepsilon_{t-i}^- + \sum_{j=1}^p \beta_j h_{t-j} \right)^2,$$

com $\alpha_0 > 0$, $\alpha_{i,+} \geq 0$, $\alpha_{i,-} \geq 0$, $i = 1, \dots, q$, e $\beta_j \geq 0$, $j = 1, \dots, p$.

Constatamos, aquando do estudo dos modelos TARCH, que a introdução simultânea, na definição da variabilidade, dos processos ε^+ e ε^- complica fortemente a obtenção de uma dinâmica linear, do tipo ARMA, para uma função do processo. Obviamente, neste caso mais geral, tal estudo será ainda mais pesado sob o ponto de vista calculatório. Houve assim necessidade de procurar outra metodologia para, em particular, analisar as questões de estacionaridade do processo subjacente.

Tal metodologia baseia-se fundamentalmente na obtenção de uma representação Markoviana multivariada do tipo auto-regressivo com coeficientes aleatórios da qual se deduz a solução estacionária para o processo ε . Nomeadamente, em Gonçalves e Mendes-Lopes (1994, 1996) são obtidas condições necessárias e suficientes de estacionaridade forte e fraca para processos ε , seguindo um modelo TARCH(q) ou GTARCH(p, q), solução de uma equação do tipo $\varepsilon_t = \sqrt{h_t} Z_t$, onde $(Z_t, t \in \mathbb{Z})$ é uma família de variáveis aleatórias independentes, identicamente distribuídas, centradas e reduzidas tais que, para todo o $t \in \mathbb{Z}$, Z_t é independente de ε_{t-1} . Estes resultados generalizam o estudo efectuado no trabalho pioneiro de Zakoian (1990), para analisar a estacionaridade forte de um modelo TARCH(1) condicionalmente Gaussiano, que vamos detalhar neste texto.

Seja $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ o processo estocástico real (p.e.r.) definido por

$$\begin{cases} \varepsilon_t = h_t Z_t \\ h_t = \alpha + \beta \varepsilon_{t-1}^+ - \gamma \varepsilon_{t-1}^- \end{cases}, \quad (5.1.2.1)$$

onde $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$ e $Z = (Z_t, t \in \mathbb{Z})$ é um ruído branco gaussiano de variância unitária tal que Z_t é independente de ε_{t-1} . Além disso, verifica-se a seguinte hipótese:

\mathcal{H} : As v.a.r. $A_t = \beta Z_t^+ - \gamma Z_t^-$, $t \in \mathbb{Z}$, são tais que $E(\ln A_t)$ existe e é estritamente negativa.

a) Começemos por mostrar que $(h_t, t \in \mathbb{Z})$ verifica a seguinte equação de recorrência:

$$h_t = \alpha \left(1 + \sum_{j=1}^p A_{t-1} A_{t-2} \dots A_{t-j} \right) + A_{t-1} \dots A_{t-p-1} h_{t-p-1}. \quad (5.1.2.2)$$

De facto,

$$\begin{aligned} h_t &= \alpha + \beta \varepsilon_{t-1}^+ - \gamma \varepsilon_{t-1}^- \\ &= \alpha + \beta h_{t-1} Z_{t-1}^+ - \gamma h_{t-1} Z_{t-1}^- \\ &= \alpha + A_{t-1} h_{t-1} = \alpha + A_{t-1} (\alpha + A_{t-2} h_{t-2}), \end{aligned}$$

pelo que repetindo recursivamente este procedimento se obtém a equação referida.

b) Provemos agora que p.e.r. $Y = (Y_t, t \in \mathbb{Z})$ tal que

$$Y_t = \alpha \left(1 + \lim_{p \rightarrow +\infty} (q.c.) \sum_{j=1}^p A_{t-1} A_{t-2} \dots A_{t-j} \right)$$

existe e é uma solução fortemente estacionária da equação de recorrência (5.1.2.2).

Para ver que Y está bem definido basta verificar que a série

$$\sum_{j=1}^{+\infty} A_{t-1} A_{t-2} \dots A_{t-j}$$

é quase certamente convergente. De facto, usando o critério de Cauchy tem-se

$$\begin{aligned} \lim_{l \rightarrow +\infty} (q.c.) (A_{t-1} A_{t-2} \dots A_{t-l})^{\frac{1}{l}} &= e^{\lim_{l \rightarrow +\infty} (q.c.) \frac{1}{l} \sum_{i=1}^l \log A_{t-i}} \\ &= e^{\delta} < 1, \end{aligned}$$

pois, como as v.a.r. $\log A_{t-i}$, $i = 1, 2, \dots, l$, são i.i.d. e têm valor médio, decorre do teorema de Kolmogorov que

$$\lim_{l \rightarrow +\infty} (q.c.) \frac{1}{l} \sum_{i=1}^l \log A_{t-i} \xrightarrow{l \rightarrow +\infty} E(\log A_t) = \delta < 0.$$

Y é solução da referida equação de recorrência pois é fácil verificar que $Y_t = \alpha + A_{t-1}Y_{t-1}$.

Finalmente, a estacionaridade forte de Y decorre do facto deste ser uma função mensurável de Z , que é um processo fortemente estacionário por se tratar de uma sucessão de v.a.i.i.d..

c) O processo Y é, q.c., a única solução fortemente estacionária de (5.1.2.2).

De facto, sendo $V = (V_t, t \in \mathbb{Z})$ uma outra solução fortemente estacionária de (5.1.2.2) tem-se

$$\begin{aligned} |Y_t - V_t| &= A_{t-1} \dots A_{t-p-1} |Y_{t-p-1} - V_{t-p-1}| \\ &\leq A_{t-1} \dots A_{t-p-1} (|Y_{t-p-1}| + |V_{t-p-1}|), \quad \forall p \in \mathbb{N}. \end{aligned}$$

Então

$$\begin{aligned} \forall \varepsilon > 0, P(|Y_t - V_t| > \varepsilon) &\leq P\left(A_{t-1} \dots A_{t-p-1} |Y_{t-p-1}| > \frac{\varepsilon}{2}\right) \\ &+ P\left(A_{t-1} \dots A_{t-p-1} |V_{t-p-1}| > \frac{\varepsilon}{2}\right). \end{aligned}$$

Seja δ um real positivo arbitrariamente fixo e escolhamos M e M' tais que $P(|Y_t| > M) < \frac{\delta}{4}$ e $P(|V_t| > M') < \frac{\delta}{4}$. Note-se que esta escolha é independente de t atendendo à estacionaridade forte dos processos. Assim

$$\begin{aligned} P(|Y_t - V_t| > \varepsilon) &\leq P\left(A_{t-1} \dots A_{t-p-1} > \frac{\varepsilon}{2M}\right) + P(|Y_{t-p-1}| > M) \\ &+ P\left(A_{t-1} \dots A_{t-p-1} > \frac{\varepsilon}{2M'}\right) + P(|V_{t-p-1}| > M'). \end{aligned}$$

Mas, da hipótese \mathcal{H} e do teorema da convergência de Kolmogorov, tem-se

$$\begin{aligned} \lim_{p \rightarrow +\infty} (q.c.) A_{t-1} \dots A_{t-p-1} &= \lim_{p \rightarrow +\infty} (q.c.) e^{\sum_{i=1}^{p+1} \ln A_{t-i}} \\ &= \lim_{p \rightarrow +\infty} (q.c.) e^{p \left(\frac{1}{p} \sum_{i=1}^p \ln A_{t-i} \right)} = 0, \end{aligned}$$

pelo que $P(|Y_t - V_t| > \varepsilon) \leq \delta$, com δ qualquer arbitrariamente pequeno, isto é,

$$Y_t = V_t, \text{ P- q.c. .}$$

d) Se os parâmetros β e γ são tais que $\beta^2 + \gamma^2 < 2$, então o processo Y é forte e fracamente estacionário.

Para provar que Y é fortemente estacionário basta provar que, nestas condições, $E(\log A_t) < 0$.

Da desigualdade de Jensen

$$E(\log A_t) \leq \log(E(A_t)) \leq \log(E(A_t^2))^{\frac{1}{2}} < 0,$$

pois

$$\begin{aligned} E(A_t^2) &= \beta^2 E(Z_t^+)^2 + \gamma^2 E(Z_t^-)^2 \\ &= (\beta^2 + \gamma^2) E(Z_t^+)^2 = \frac{(\beta^2 + \gamma^2)}{2}. \end{aligned}$$

Para concluir que Y é fracamente estacionário basta agora provar que

$$E(Y_t^2) = \alpha^2 E\left[\left(1 + \sum_{j=1}^{+\infty} A_{t-1}A_{t-2}\dots A_{t-j}\right)^2\right] < +\infty.$$

Ora

$$\begin{aligned} E(Y_t^2) &= \alpha^2\{1 + 2\sum_{j=1}^{+\infty} E(A_{t-1}A_{t-2}\dots A_{t-j}) + \\ &\quad + \sum_{j=1}^{+\infty} \sum_{k=1}^{+\infty} E[(A_{t-1}A_{t-2}\dots A_{t-j})(A_{t-1}A_{t-2}\dots A_{t-k})]\} \end{aligned}$$

A convergência destas séries decorre então das seguintes majorações, decorrentes da aplicação da desigualdade de Schwartz,

$$\begin{aligned} E(A_{t-1}A_{t-2}\dots A_{t-j}) &= E(A_1A_2\dots A_j) = [E(A_1)]^j \\ &\leq [E(A_1^2)]^{\frac{j}{2}} = \left(\sqrt{\frac{\beta^2 + \gamma^2}{2}}\right)^j. \end{aligned}$$

$$\begin{aligned} E[(A_{t-1}A_{t-2}\dots A_{t-j})(A_{t-1}A_{t-2}\dots A_{t-k})] \\ \leq E\left[(A_{t-1}A_{t-2}\dots A_{t-j})^2\right]^{\frac{1}{2}} E\left[(A_{t-1}A_{t-2}\dots A_{t-k})^2\right]^{\frac{1}{2}} \end{aligned}$$

$$= [E(A_1^2)]^{\frac{j+k}{2}}.$$

De facto, sendo $\theta = \sqrt{\frac{(\beta^2 + \gamma^2)}{2}}$,

$$\sum_{j=1}^{+\infty} E(A_{t-1}A_{t-2}\dots A_{t-j}) \leq \frac{\theta}{1-\theta}$$

e

$$\begin{aligned} & \sum_{j=1}^{+\infty} \sum_{k=1}^{+\infty} E[(A_{t-1}A_{t-2}\dots A_{t-j})(A_{t-1}A_{t-2}\dots A_{t-k})] \\ &= \sum_{j=1}^{+\infty} E[(A_{t-1}A_{t-2}\dots A_{t-j})^2] + \\ & \quad + 2 \sum_{j=1}^{+\infty} \sum_{k=j+1}^{+\infty} E[(A_{t-1}A_{t-2}\dots A_{t-j})(A_{t-1}A_{t-2}\dots A_{t-k})] \\ &\leq \frac{\theta^2}{1-\theta^2} + 2 \sum_{j=1}^{+\infty} \sum_{k=j+1}^{+\infty} \theta^{j+k} = \frac{\theta^2}{(1-\theta)^2}. \end{aligned}$$

e) Vejamos agora que Y é também a única solução fracamente estacionária de (5.1.2.2).

Seja $V = (V_t, t \in \mathbb{Z})$ uma outra solução fracamente estacionária do referido modelo. Então usando a independência das variáveis (A_t) bem como a estacionaridade fraca dos processos Y e V vem

$$\begin{aligned} E((Y_t - V_t)^2) &= E((A_{t-1}\dots A_{t-p-1}(Y_{t-p-1} - V_{t-p-1}))^2) \\ &\leq 2E((A_{t-1}\dots A_{t-p-1})^2(Y_{t-p-1}^2 + V_{t-p-1}^2)) \\ &\leq 2(E(A_1^2))^{p+1}(E(Y_t^2) + E(V_t^2)) \\ &\leq \left(\frac{(\beta^2 + \gamma^2)}{2}\right)^{p+1} \leq M, \end{aligned}$$

com M constante.

Assim $E \left[(Y_t - V_t)^2 \right] < \eta$ para todo o η positivo arbitrariamente pequeno, ou seja $Y_t = V_t$ no sentido de L^2 .

f) Em conclusão, quando $\beta^2 + \gamma^2 < 2$ existe uma solução forte e fracamente estacionária, X , de (5.1.2.1) única, definida a partir da solução única Y do modelo (5.1.2.2) por

$$X_t = Z_t Y_t, \quad t \in \mathbb{Z}.$$

O estudo deste caso particular permite exibir modelos fortemente estacionários que não são fracamente estacionários.

De facto, determinemos α, β, γ tais que $E(\log A_t) < 0$. Ora

$$\begin{aligned} \log A_t &= \log(\beta Z_t^+ - \gamma Z_t^-) \\ &= \log[(\beta \mathbb{I}_{\{Z_t > 0\}} + \gamma \mathbb{I}_{\{Z_t < 0\}}) |Z_t|] \\ &= (\log \beta) \mathbb{I}_{\{Z_t > 0\}} + (\log \gamma) \mathbb{I}_{\{Z_t < 0\}} + \log |Z_t|. \end{aligned}$$

Assim,

$$E(\log A_t) = \frac{\log(\beta\gamma)}{2} + E(\log |Z_t|).$$

De Nelson (1990, teorema 2.3) temos $E(\log Z_t^2) = \log 2 + \Psi(\frac{1}{2})$, onde Ψ é a função de Euler (Davis, 1965).

Assim

$$E(\log |Z_t|) = \frac{1}{2} \left(\log 2 + \Psi\left(\frac{1}{2}\right) \right).$$

Ora $E(\log A_t) < 0$ se e só se $\beta\gamma < \exp[-2E(\log |Z_t|)]$ ou, de modo equivalente,

$$\beta\gamma < \frac{1}{2} \exp\left(-\Psi\left(\frac{1}{2}\right)\right) \cong 3.5619\dots$$

Esta condição é mais fraca que a condição de estacionariedade fraca, $\beta^2 + \gamma^2 < 2$; assim, os modelos TAR(1) tais que $\beta\gamma < 3.5619$ e $\beta^2 + \gamma^2 \geq 2$ são fortemente estacionários mas não fracamente estacionários (basta tomar $2 \leq \beta\gamma < 3.5619$). A figura 5.4 ilustra estas regiões.

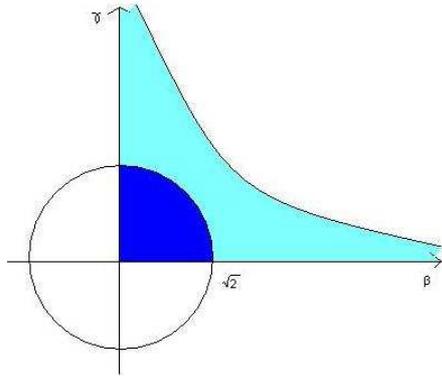


Fig. 5.4 Regiões de estacionaridade forte e fraca

5.1.3 Modelos com erros condicionalmente heteroscedásticos

A hipótese de marcha aleatória (ou de martingala) tem tido um papel importante na descrição e estudo das séries financeiras, conforme podemos constatar no trabalho pioneiro de Bachelier (1900) e, mais recentemente, nos trabalhos de Fama (1965) ou de Samuelson (1973). Este facto leva a que as formulações condicionalmente heteroscedásticas inicialmente propostas (Engle (1982), Bollerslev (1986), ...) correspondam a processos centrados e não correlacionados que permitem modelar directamente os rendimentos sob a forma de ruído branco (homoscedástico ou não).

Um processo estocástico real $Y = (Y_t, t \in \mathbb{Z})$ é uma *marcha aleatória* se e só se

$$E(Y_{t+1} | \underline{Y}_t) = Y_t.$$

Além disso, prova-se facilmente que esta condição é equivalente quer a

$$E(Y_{t+1} - Y_t | \underline{Y}_t) = 0,$$

quer a

$$E\left(\frac{Y_{t+1} - Y_t}{Y_t} \middle| \underline{Y}_t\right) = 0.$$

Assim, se Y corresponde a uma série de preços, a condição de marcha aleatória pode ser reescrita fazendo aparecer os rendimentos absolutos ou relativos. Torna-se portanto evidente a importância desta hipótese uma vez que, qualquer que seja a sofisticação técnica de venda, o ganho médio é sempre nestas condições nulo, não podendo portanto esperar-se lucro ou perda sistemáticos.

No entanto, a hipótese de marcha aleatória pode questionar-se ou, pelo menos, testar-se a partir dos dados (Gouriéroux (1992)). O processo, interpretado como série de rendimentos, pode agora depender de vários factores exógenos ou possuir por si só uma dinâmica endógena. Este facto leva a colocar a formulação condicionalmente heteroscedástica sobre o respectivo processo de erro. Há exemplos desta situação logo em Engle (1982), onde são considerados modelos de regressão com erros ARCH. Além disso, em 1984 e 1986, Weiss propõe uma modelação ARCH para o processo de erro de um modelo ARMA(p, q).

Mais geralmente, (Gouriéroux, 1992a, Guégan, 1994, Gonçalves, Mendes-Lopes, 1996a, 1998) podemos sugerir a utilização de modelações heteroscedásticas no processo de erro associado a um modelo linear geral. Tais processos mistos são designados por ARMAX com erros condicionalmente heteroscedásticos se têm a definição geral seguinte:

Definição. Um processo estocástico real $Y = (Y_t, t \in \mathbb{Z})$ segue um modelo ARMAX com erros condicionalmente heteroscedásticos (ARMAX-CH) se verifica a equação

$$\Phi(L)Y_t = A(L)X_t + \Theta(L)\varepsilon_t,$$

onde X_t são variáveis explicativas observáveis, $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um processo de erro condicionalmente heteroscedástico e L é o operador atraso, com

$$\begin{aligned}\Phi(L) &= 1 - \varphi_1 L - \dots - \varphi_p L^p, \quad \Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q, \\ A(L) &= 1 - a_1 L - \dots - a_r L^r.\end{aligned}$$

A análise destes modelos passa essencialmente pelo estudo das propriedades do processo de erro pois que, conhecidas estas, o estudo do processo em causa inclui-se no dos modelos ARMA, ou,

mais geralmente, no do modelo linear geral (Gallant, 1987 ou Gouriéroux, Monfort, 1989).

Exemplo. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico real (p.e.r.) definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) tal que

$$\left(1 - \frac{1}{3}L\right) (1 - \alpha L) X_t = \varepsilon_t, \quad t \in \mathbb{Z},$$

onde α é um número real tal que $|\alpha| < 1$ e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um p.e.r. definido por

$$\begin{cases} \varepsilon_t = h_t Z_t \\ h_t = \delta + \beta |\varepsilon_{t-1}| \end{cases},$$

com $\delta > 0$, $\beta \geq 0$ e $(Z_t, t \in \mathbb{Z})$ um ruído branco gaussiano de variância unitária tal que Z_t é independente de $\underline{\varepsilon}_{t-1} = \sigma(\varepsilon_t, \varepsilon_{t-1}, \dots)$.

a) O processo ε segue um modelo condicionalmente heteroscedástico e é condicionalmente gaussiano.

De facto,

$$E(\varepsilon_t | \underline{\varepsilon}_{t-1}) = E(h_t Z_t | \underline{\varepsilon}_{t-1}) = h_t E(Z_t | \underline{\varepsilon}_{t-1}) = h_t E(Z_t) = 0,$$

$$V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = E(h_t^2 Z_t^2 | \underline{\varepsilon}_{t-1}) = h_t^2 E(Z_t^2 | \underline{\varepsilon}_{t-1}) = h_t^2,$$

pois h_t é $\underline{\varepsilon}_{t-1}$ -mensurável e Z_t é independente de $\underline{\varepsilon}_{t-1}$.

Por outro lado, $\mathcal{L}(\varepsilon_t | \underline{\varepsilon}_{t-1}) \equiv N(0, h_t)$ pois

$$\begin{aligned} F_{\varepsilon_t}^{\underline{\varepsilon}_{t-1}}(x) &= P(h_t Z_t \leq x | \underline{\varepsilon}_{t-1}) = P\left(Z_t \leq \frac{x}{h_t} \middle| \underline{\varepsilon}_{t-1}\right) \\ &= F_{N(0,1)}\left(\frac{x}{h_t}\right) = F_{N(0, h_t)}(x), \end{aligned}$$

tendo de novo em conta que Z_t é independente de $\underline{\varepsilon}_{t-1}$, h_t é $\underline{\varepsilon}_{t-1}$ -mensurável e ainda a estabilidade da lei normal para as transformações afins.

b) Se ε é estacionário de 2^a ordem então $\beta < 1$.

De facto,

$$\begin{aligned} E(\varepsilon_t^2) &= E(E(\varepsilon_t^2 | \underline{\varepsilon}_{t-1})) = E(h_t^2) E(Z_t^2) \\ &= \delta^2 + \beta^2 E(\varepsilon_{t-1}^2) + 2\delta\beta E(|\varepsilon_{t-1}|) \end{aligned}$$

Da estacionaridade de ε decorre

$$(1 - \beta^2) E(\varepsilon_t^2) = \delta^2 + 2\delta\beta E(|\varepsilon_{t-1}|)$$

o que implica $\beta < 1$.

c) Por outro lado se $\beta < 1$ então ε é estacionário de 2ª ordem. Além disso, é um ruído branco de variância

$$V(\varepsilon_t) = \frac{(\sqrt{2\pi} + 2\beta)\delta^2}{(\sqrt{2\pi} - 2\beta)(1 - \beta^2)}$$

Seja $\omega = (\omega_t, t \in \mathbb{Z})$ o p.e. vectorial tal que $\omega_t = \begin{pmatrix} \varepsilon_t^2 \\ |\varepsilon_t| \end{pmatrix}$.

Mostremos que ω admite uma representação autorregressiva de ordem 1.

De facto, seja $u = (u_t, t \in \mathbb{Z})$ o processo inovação associado ao processo ω , isto é, $u_t = \omega_t - E(\omega_t | \underline{\omega}_{t-1})$.

$$E(\omega_t | \underline{\omega}_{t-1}) = \begin{pmatrix} E(\varepsilon_t^2 | \underline{\omega}_{t-1}) \\ E(|\varepsilon_t| | \underline{\omega}_{t-1}) \end{pmatrix},$$

Mas

$$\begin{aligned} E(\varepsilon_t^2 | \underline{\omega}_{t-1}) &= h_t^2 = \delta^2 + \beta^2 \varepsilon_{t-1}^2 + 2\delta\beta |\varepsilon_{t-1}| \\ E(|\varepsilon_t| | \underline{\omega}_{t-1}) &= h_t E(|Z_t|) = \frac{2h_t}{\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} (\delta + \beta |\varepsilon_{t-1}|). \end{aligned}$$

Assim

$$E(\omega_t | \underline{\omega}_{t-1}) = \begin{pmatrix} \delta^2 \\ \frac{2}{\sqrt{2\pi}}\delta \end{pmatrix} + \begin{pmatrix} \beta^2 & 2\delta\beta \\ 0 & \frac{2}{\sqrt{2\pi}}\beta \end{pmatrix} \begin{pmatrix} \varepsilon_{t-1}^2 \\ |\varepsilon_{t-1}| \end{pmatrix};$$

e consequentemente, o processo ω verifica um modelo auto-regressivo vectorial descentrado de ordem 1, a saber

$$\omega_t = \begin{pmatrix} \delta^2 \\ \frac{2}{\sqrt{2\pi}}\delta \end{pmatrix} + \begin{pmatrix} \beta^2 & 2\delta\beta \\ 0 & \frac{2}{\sqrt{2\pi}}\beta \end{pmatrix} \omega_{t-1} + u_t,$$

Desta representação deduz-se então

$$E(\omega_t) = \begin{pmatrix} \delta^2 \\ \frac{2}{\sqrt{2\pi}}\delta \end{pmatrix} + \begin{pmatrix} \beta^2 & 2\delta\beta \\ 0 & \frac{2}{\sqrt{2\pi}}\beta \end{pmatrix} E(\omega_{t-1}).$$

Assim, sendo $A = \begin{pmatrix} \beta^2 & 2\delta\beta \\ 0 & \frac{2}{\sqrt{2\pi}}\beta \end{pmatrix}$, se as raízes do polinómio $\det(I - Az)$ estão no exterior do disco unitário então a equação de recorrência anterior tem uma solução independente de t , de onde decorre que $E(\varepsilon_t^2)$ será, nestas condições independente de t . Ora as raízes daquele polinómio estão no exterior do disco unitário se e somente se os valores próprios de A forem em módulo menores do que 1. Mas

$$\begin{aligned} \det(A - \lambda I) = 0 &\iff (\beta^2 - \lambda) \left(\frac{2}{\sqrt{2\pi}}\beta - \lambda \right) = 0 \\ &\iff \lambda = \beta^2 \text{ ou } \lambda = \frac{2}{\sqrt{2\pi}}\beta \end{aligned}$$

que são ambos, em valor absoluto, menores que 1 porque, por hipótese, $\beta < 1$.

Por outro lado, sendo o processo ε condicionalmente heteroscedástico, é fácil concluir que $E(\varepsilon_t) = 0$ e $Cov(\varepsilon_t, \varepsilon_{t+h}) = 0$, $\forall h \in \mathbb{Z} - \{0\}$.

Em conclusão, ε é um processo estacionário de 2ª ordem desde que $\beta < 1$, sendo nomeadamente um ruído branco.

Calculemos a sua variância.

Da equação de recorrência vem

$$E \begin{pmatrix} \varepsilon_t^2 \\ |\varepsilon_t| \end{pmatrix} = \begin{pmatrix} \delta^2 \\ \frac{2}{\sqrt{2\pi}}\delta \end{pmatrix} + \begin{pmatrix} \beta^2 & 2\delta\beta \\ 0 & \frac{2}{\sqrt{2\pi}}\beta \end{pmatrix} E \begin{pmatrix} \varepsilon_{t-1}^2 \\ |\varepsilon_{t-1}| \end{pmatrix},$$

isto é,

$$\begin{cases} E(\varepsilon_t^2) = \delta^2 + \beta^2 E(\varepsilon_{t-1}^2) + 2\delta\beta E(|\varepsilon_{t-1}|) & \text{(i)} \\ E(|\varepsilon_t|) = \frac{2}{\sqrt{2\pi}}\delta + \frac{2}{\sqrt{2\pi}}\beta E(|\varepsilon_{t-1}|) & \text{(ii)} \end{cases}.$$

De (ii) vem, atendendo à estacionaridade de ω ,

$$E(|\varepsilon_t|) = \frac{2\delta}{\sqrt{2\pi-2\beta}},$$

deduzindo-se então de (i), de novo graças à estacionaridade de ω ,

$$E(\varepsilon_t^2) = \frac{(\sqrt{2\pi+2\beta})\delta^2}{(\sqrt{2\pi-2\beta})(1-\beta^2)}.$$

d) Supondo ε um processo estacionário de 2^a ordem, vamos provar que o processo X é estacionário de 2^a ordem e determinar os coeficientes da correspondente representação média móvel infinita. Verifica-se por fim que a representação obtida é canónica.

O polinómio auto-regressivo associado a X é

$$\Phi(L) = \left(1 - \frac{1}{3}L\right)(1 - \alpha L),$$

pelo que, sendo as suas raízes ($z_1 = 3$ e $z_2 = \frac{1}{\alpha}$) de módulos diferentes de 1, tem-se a estacionaridade de 2^a ordem de X .

Quanto à representação média móvel infinita tem-se

$$\begin{aligned} X_t &= \left(\frac{1}{1 - \frac{1}{3}L}\right) \left(\frac{1}{1 - \alpha L}\right) \varepsilon_t \\ &= \sum_{j=0}^{+\infty} \sum_{k=0}^{+\infty} \left(\frac{1}{3}L\right)^j (\alpha L)^k \varepsilon_t \\ &= \sum_{n=0}^{+\infty} \left(\sum_{j=0}^n \left(\frac{1}{3}\right)^j \alpha^{n-j}\right) \varepsilon_{t-n}, \end{aligned}$$

notando-se que

$$\sum_{n=0}^{+\infty} \left(\sum_{j=0}^n \left(\frac{1}{3}\right)^j |\alpha|^{n-j}\right) \leq \sum_{n=0}^{+\infty} (n+1) \left(\max\left(\frac{1}{3}, |\alpha|\right)\right)^n,$$

que é uma série convergente pois $|\alpha| < 1$.

A representação obtida é canónica porque, para todo o t , ε_t é ortogonal ao subespaço linear gerado por $(X_{t-1}, X_{t-2}, \dots)$, denotado por \underline{H}_{t-1}^X . De facto,

$$X_t = \sum_{n=0}^{+\infty} \left(\sum_{j=0}^n \left(\frac{1}{3} \right)^j \alpha^{n-j} \right) \varepsilon_{t-n} \in \underline{H}_t^\varepsilon,$$

pois é o limite de uma função linear de $(\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-M})$. Assim, $\underline{H}_{t-1}^X \subseteq \underline{H}_{t-1}^\varepsilon$, $t \in \mathbb{Z}$. Mas ε_t é ortogonal a $\underline{H}_{t-1}^\varepsilon$, logo é ortogonal a \underline{H}_{t-1}^X .

e) Vamos agora obter a função de autocorrelação de X . Seja $a_n = \sum_{j=0}^n \left(\frac{1}{3} \right)^j \alpha^{n-j}$, $n \in \mathbb{N}$.

$$\begin{aligned} \gamma_X(0) &= E(X_t^2) = E \left(\sum_{n=0}^{+\infty} a_n \varepsilon_{t-n} \right)^2 \\ &= E[\varepsilon_t^2 + 2\varepsilon_t \sum_{n=1}^{+\infty} a_n \varepsilon_{t-n} + \sum_{n=1}^{+\infty} a_n^2 \varepsilon_{t-n}^2 + \\ &\quad + \sum_{n=1}^{+\infty} \sum_{m \neq n}^{+\infty} a_n a_m \varepsilon_{t-n} \varepsilon_{t-m}] \\ &= \sigma_\varepsilon^2 + \sum_{n=1}^{+\infty} a_n^2 \sigma_\varepsilon^2 = \sum_{n=0}^{+\infty} \left(\sum_{j=0}^n \left(\frac{1}{3} \right)^j \alpha^{n-j} \right)^2 \sigma_\varepsilon^2, \end{aligned}$$

com $\sigma_\varepsilon^2 = V(\varepsilon_t)$, já calculada.

Por outro lado, para calcular $\gamma_X(h) = E(X_{t-h} X_t)$, $h > 0$, note-se que

$$\left(1 - \frac{1}{3}L \right) (1 - \alpha L) X_t = \varepsilon_t$$

é equivalente a

$$X_t = \left(\frac{1}{3} + \alpha \right) X_{t-1} - \frac{\alpha}{3} X_{t-2} + \varepsilon_t.$$

Assim, como $Cov(X_{t-h}, \varepsilon_t) = 0$, $\forall h > 0$, pois X_{t-h} é ortogonal a ε_t , vem

$$\begin{aligned} \text{Cov}(X_{t-h}, X_t) - \left(\frac{1}{3} + \alpha\right) \text{Cov}(X_{t-h}, X_{t-1}) + \\ + \frac{\alpha}{3} \text{Cov}(X_{t-h}, X_{t-2}) = 0, \end{aligned}$$

ou seja

$$\gamma_X(h) - \left(\frac{1}{3} + \alpha\right) \gamma_X(h-1) + \frac{\alpha}{3} \gamma_X(h-2) = 0, \quad \forall h > 0.$$

Se $h = 1$, obtém-se

$$\gamma_X(1) - \left(\frac{1}{3} + \alpha\right) \gamma_X(0) + \frac{\alpha}{3} \gamma_X(1) = 0 \Leftrightarrow \gamma_X(1) = \frac{1 + 3\alpha}{3 + \alpha} \gamma_X(0).$$

O polinómio característico da equação às diferenças linear homogénea $x^2 - \left(\frac{1}{3} + \alpha\right)x + \frac{\alpha}{3} = 0$ tem as raízes $x = \frac{1}{3}$ e $x = \alpha$.

Assim, para $\alpha = \frac{1}{3}$ a solução é

$$\gamma_X(h) = (A + Bh) \left(\frac{1}{3}\right)^h$$

com A e B tais que

$$\begin{cases} A = \gamma_X(0) \\ B = 3\gamma_X(1) - \gamma_X(0) \end{cases} .$$

Então

$$\gamma_X(h) = (\gamma_X(0) + (3\gamma_X(1) - \gamma_X(0))h) \left(\frac{1}{3}\right)^h, \quad \forall h \geq 0,$$

e portanto

$$\varrho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad \forall h \geq 0,$$

com o valor de $\gamma_X(0)$ correspondente a $\alpha = \frac{1}{3}$.

A paridade desta função completa a sua definição.

Se $\alpha \neq \frac{1}{3}$, então

$$\gamma_X(h) = A \left(\frac{1}{3}\right)^h + B\alpha^h,$$

de onde decorre

$$\begin{cases} A = \frac{3(\gamma_X(1) - \alpha\gamma_X(0))}{1 - 3\alpha} \\ B = \frac{\gamma_X(0) - 3\gamma_X(1)}{1 - 3\alpha} \end{cases}.$$

Assim, $\forall h \geq 0$,

$$\gamma_X(h) = \frac{3(\gamma_X(1) - \alpha\gamma_X(0))}{1 - 3\alpha} \left(\frac{1}{3}\right)^h + \frac{\gamma_X(0) - 3\gamma_X(1)}{1 - 3\alpha} \alpha^h,$$

de onde se deduz a expressão da autocorrelação de X .

f) Analisemos a existência de densidade espectral de X e apresentemos uma versão desta função.

Verifica-se facilmente que $\sum_{h=0}^{+\infty} |\gamma_X(h)| < +\infty$, pelo que a função densidade espectral, f_X , existe.

Além disso, da equação de evolução do processo X , decorre facilmente,

$$f_X(\omega) = \frac{\sigma_\varepsilon^2}{2\pi |\Phi(e^{i\omega})|^2} = \frac{\sigma_\varepsilon^2}{2\pi \left|1 - \left(\alpha + \frac{1}{3}\right)e^{i\omega} + \frac{\alpha}{3}e^{2i\omega}\right|^2},$$

com $\Phi(L) = \left(1 - \frac{1}{3}L\right)(1 - \alpha L)$.

g) Supondo, a partir de agora, $\alpha = \frac{1}{3}$, vamos obter as previsões deste modelo no instante t a um horizonte s ($s \in \mathbb{N}$) e deduzir, em função do desvio padrão condicional do processo ε , $(h_t, t \in \mathbb{Z})$, a variância condicional do erro da previsão, ao horizonte 2, deste modelo no instante t .

As previsões deste modelo no instante t a um horizonte s são dadas por $E(X_{t+s} | \underline{X}_t)$.

De forma recursiva obtém-se

$$\begin{aligned}
 X_{t+s} &= \frac{2}{3}X_{t+s-1} - \frac{1}{9}X_{t+s-2} + \varepsilon_{t+s} \\
 &= \frac{3}{9}X_{t+s-2} - \frac{2}{27}X_{t+s-3} + \frac{2}{3}\varepsilon_{t+s-1} + \varepsilon_{t+s} \\
 &= \frac{4}{27}X_{t+s-3} - \frac{3}{81}X_{t+s-4} + \frac{3}{9}\varepsilon_{t+s-2} + \frac{2}{3}\varepsilon_{t+s-1} + \varepsilon_{t+s} \\
 &= \dots \\
 &= \frac{s+1}{3^s}X_t - \frac{s}{3^{s+1}}X_{t-1} + \sum_{j=0}^{s-1} \frac{j+1}{3^j}\varepsilon_{t+s-j},
 \end{aligned}$$

pelo que

$$E(X_{t+s} | \underline{X}_t) = \frac{s+1}{3^s}X_t - \frac{s}{3^{s+1}}X_{t-1}.$$

Calculemos agora a variância condicional do erro da previsão, ao horizonte 2, deste modelo no instante t .

$$V[(X_{t+2} - E(X_{t+2} | \underline{X}_t)) | \underline{X}_t] = V(X_{t+2} | \underline{X}_t)$$

Ora

$$X_{t+2} = \frac{3}{9}X_t - \frac{2}{27}X_{t-1} + \frac{2}{3}\varepsilon_{t+1} + \varepsilon_{t+2},$$

pelo que

$$E(X_{t+2} | \underline{X}_t) = \frac{1}{3}X_t - \frac{2}{27}X_{t-1}.$$

Assim,

$$\begin{aligned}
 V(X_{t+2} | \underline{X}_t) &= E\left(\left(\frac{2}{3}\varepsilon_{t+1} + \varepsilon_{t+2}\right)^2 \middle| \underline{X}_t\right) \\
 &= \frac{4}{9}E(E(\varepsilon_{t+1}^2 | \underline{\varepsilon}_t) | \underline{X}_t) + E(E(\varepsilon_{t+2}^2 | \underline{\varepsilon}_{t+1}) | \underline{X}_t) \\
 &\quad + \frac{4}{3}E(E(\varepsilon_{t+1}\varepsilon_{t+2} | \underline{\varepsilon}_{t+1}) | \underline{X}_t) \\
 &= \frac{4}{9}E(h_{t+1}^2 | \underline{X}_t) + E(h_{t+2}^2 | \underline{X}_t).
 \end{aligned}$$

Como a representação de X é canônica, tem-se $\underline{X}_t = \underline{\varepsilon}_t$. Assim

$$\begin{aligned} E(h_{t+2}^2 | \underline{\varepsilon}_t) &= E((\delta^2 + \beta^2 \varepsilon_{t+1}^2 + 2\delta\beta |\varepsilon_{t+1}|) | \underline{\varepsilon}_t) \\ &= \delta^2 + \beta^2 h_{t+1}^2 + 2\delta\beta E(|h_{t+1} Z_{t+1}| | \underline{\varepsilon}_t) \\ &= \delta^2 + \beta^2 h_{t+1}^2 + 2\delta\beta h_{t+1} E(|Z_{t+1}| | \underline{\varepsilon}_t) \end{aligned}$$

e, como Z é um ruído branco Gaussiano reduzido tem-se

$$E(|Z_{t+1}| | \underline{\varepsilon}_t) = \frac{2}{\sqrt{2\pi}}.$$

Finalmente,

$$V(X_{t+2} | \underline{X}_t) = \frac{4}{9} h_{t+1}^2 + \delta^2 + \beta^2 h_{t+1}^2 + \frac{4}{\sqrt{2\pi}} \delta \beta h_{t+1},$$

com $h_{t+1} = \delta + \beta |\varepsilon_t|$. ■

O modelo do exemplo anterior, tendo especificada a evolução da variância condicional ao passado, permite melhorar a precisão das previsões, uma vez que podemos utilizar a variância condicional (função da informação passada), em vez da variância não condicional, na análise do erro de previsão.

No exemplo seguinte vamos analisar esse erro a qualquer horizonte para o modelo AR(1)-ARCH(1).

Exemplo. Seja $Y = (Y_t, t \in \mathbb{Z})$ o processo tal que

$$Y_t - \varphi Y_{t-1} = \varepsilon_t, \quad (5.1.3.1)$$

onde ε segue um modelo ARCH(1), *i.e.*,

$$\begin{cases} E(\varepsilon_t | \underline{\varepsilon}_{t-1}) = 0 \\ V(\varepsilon_t | \underline{\varepsilon}_{t-1}) = h_t = \alpha_0 + \alpha \varepsilon_{t-1}^2 \end{cases},$$

com $\alpha_0 > 0$, $\alpha \geq 0$.

O modelo (5.1.3.1) admite solução estacionária de 2ª ordem se e somente se

$$|\varphi| \neq 1 \text{ e } \alpha < 1$$

Além disso, se

$$|\varphi| < 1 \text{ e } \alpha < 1,$$

então

- i) Y admite a representação $\text{MA}(\infty)$: $Y_t = \sum_{j=0}^{+\infty} \varphi^j \varepsilon_{t-j}$.
- ii) ε é o processo de inovação de Y .

As previsões deste modelo no instante t , para um horizonte s , $s \in \mathbb{N}$, são da forma

$$E(Y_{t+s} | \underline{Y}_t) = \varphi^s Y_t.$$

De facto, usando recursivamente a expressão de Y_t , vem

$$Y_{t+s} = \varphi^s Y_t + \sum_{j=0}^{s-1} \varphi^j \varepsilon_{t+s-j},$$

pelo que

$$\begin{aligned} E(Y_{t+s} | \underline{Y}_t) &= \varphi^s Y_t + \sum_{j=0}^{s-1} \varphi^j E(\varepsilon_{t+s-j} | \underline{Y}_t) \\ &= \varphi^s Y_t + \sum_{j=0}^{s-1} \varphi^j E(E(\varepsilon_{t+s-j} | \underline{\varepsilon}_{t+s-j-1}) | \underline{Y}_t) \\ &= \varphi^s Y_t, \end{aligned}$$

tendo em conta que $\underline{Y}_t \subset \underline{\varepsilon}_{t+k}$, $\forall k \in \mathbb{N}_0$.

A variância condicional do erro de previsão é dada por

$$\begin{aligned} V(Y_{t+s} | \underline{Y}_t) &= E\left[(Y_{t+s} - E(Y_{t+s} | \underline{Y}_t))^2 | \underline{Y}_t\right] \\ &= E\left[\left(\sum_{i=1}^s \varphi^{s-i} \varepsilon_{t+i}\right)^2 | \underline{Y}_t\right] \\ &= \sum_{i=1}^s \varphi^{2(s-i)} E\left(E\left((\varepsilon_{t+i})^2 | \underline{\varepsilon}_{t+i-1}\right) | \underline{Y}_t\right), \end{aligned}$$

usando a não correlação condicional do processo de erro, as relações entre as σ -álgebras $\underline{\varepsilon}_t$ e \underline{Y}_t e o facto de

$$\sum_{j=0}^{s-1} \varphi^j \varepsilon_{t+s-j} = \sum_{i=1}^s \varphi^{s-i} \varepsilon_{t+i}.$$

Assim,

$$V(Y_{t+s} | \underline{Y}_t) = \sum_{i=1}^s \varphi^{2(s-i)} E(h_{t+i} | \underline{\varepsilon}_t).$$

Mas

$$\begin{aligned} E(h_{t+i} | \underline{\varepsilon}_t) &= E(\alpha_0 + \alpha \varepsilon_{t+i-1}^2 | \underline{\varepsilon}_t) = \alpha_0 + \alpha E(\varepsilon_{t+i-1}^2 | \underline{\varepsilon}_t) \\ &= \alpha_0 + \alpha E(\alpha_0 + \alpha \varepsilon_{t+i-2}^2 + u_{t+i-1} | \underline{\varepsilon}_t) \\ &= \alpha_0 + \alpha E(h_{t+i-1} | \underline{\varepsilon}_t) + \alpha E(u_{t+i-1} | \underline{\varepsilon}_t) \\ &= \alpha_0 + \alpha E(h_{t+i-1} | \underline{\varepsilon}_t), \end{aligned}$$

pois $E(u_{t+i-1} | \underline{\varepsilon}_t) = E(E(u_{t+i-1} | \underline{\varepsilon}_{t+i-2}) | \underline{\varepsilon}_t) = 0$, uma vez que $\underline{\varepsilon}_t \subset \underline{\varepsilon}_{t+i-2}$.

Então, usando esta relação recursivamente, vem

$$\begin{aligned} E(h_{t+i} | \underline{\varepsilon}_t) &= \alpha_0 + \alpha_0 \alpha + \alpha^2 E(h_{t+i-2} | \underline{\varepsilon}_t) \\ &= \dots = \alpha_0 (1 + \alpha + \dots + \alpha^{i-2}) + \alpha^{i-1} E(h_{t+1} | \underline{\varepsilon}_t) \\ &= \alpha_0 (1 + \alpha + \dots + \alpha^{i-2}) + \alpha^{i-1} h_{t+1} \\ &= \alpha_0 \frac{1 - \alpha^{i-1}}{1 - \alpha} + \alpha^{i-1} h_{t+1} \end{aligned}$$

Assim

$$\begin{aligned} V(Y_{t+s} | \underline{Y}_t) &= \sum_{i=1}^s \varphi^{2(s-i)} E(h_{t+i} | \underline{\varepsilon}_t) \\ &= \frac{\alpha_0}{1 - \alpha} \varphi^{2s} \sum_{i=1}^s \varphi^{-2i} (1 - \alpha^{i-1}) \\ &\quad + \varphi^{2s} h_{t+1} \sum_{i=1}^s \varphi^{-2i} \alpha^{i-1}, \end{aligned}$$

ou seja, é igual a

$$\begin{aligned} & \frac{\alpha_0}{1-\alpha} \varphi^{2s} \sum_{i=1}^s \varphi^{-2i} + \left(h_{t+1} - \frac{\alpha_0}{1-\alpha} \right) \varphi^{2s} \sum_{i=1}^s \varphi^{-2i} \alpha^{i-1} = \\ & = \frac{\alpha_0}{1-\alpha} \frac{\varphi^{-2+2s} - \varphi^{-2}}{1-\varphi^{-2}} + \left(h_{t+1} - \frac{\alpha_0}{1-\alpha} \right) \frac{\varphi^{-2+2s} - \alpha^s \varphi^{-2}}{1-\alpha \varphi^{-2}}. \end{aligned}$$

Note-se que esta variância depende do passado através de h_{t+1} .

Analisemos a evolução com s desta variância condicional.

Atendendo à estacionaridade e à regularidade do processo

$$\begin{aligned} \lim_{s \rightarrow +\infty} V(Y_{t+s} | \underline{Y}_t) &= \lim_{s \rightarrow \infty} V(Y_t | \underline{Y}_{t-s}) = V(Y_t | \underline{Y}_{-\infty}) \\ &= V(Y_t | \{0\}) = V(Y_t). \end{aligned}$$

Assim

$$\lim_{s \rightarrow +\infty} V(Y_{t+s} | \underline{Y}_t) = \frac{\alpha_0}{1-\alpha} \frac{1}{1-\varphi^2} = V(Y_t).$$

Observe-se que, desta expressão e do valor da variância de um processo AR(1) estacionário, deduzimos ainda o valor (já conhecido) da variância de um modelo ARCH(1) : $V(\varepsilon_t) = \frac{\alpha_0}{1-\alpha}$. ■

5.2 Modelos bilineares

Os modelos bilineares surgiram inicialmente, num contexto determinista, no domínio da teoria de controlo, conforme pode constatar-se nos trabalhos de, entre outros, Mohler (1973) e Brockett (1976).

A presença de não-linearidade em séries temporais não deterministas provenientes de áreas como a Sismologia, a Astrofísica ou a Econometria levaram à consideração de tais modelos, mas agora como funções de processos aleatórios. Em Granger e Andersen (1978) surge pela primeira vez a análise probabilista de modelos bilineares para séries temporais.

5.2.1 Definição e considerações gerais

Seja $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ um processo estocástico real com estrutura de *processo de erro*.

Definição. Um processo estocástico real $X = (X_t, t \in \mathbb{Z})$ verifica um *modelo bilinear* de ordens p, q, k, l ($BL(p, q, k, l)$) se

$$X_t = \sum_{i=1}^p \phi_i X_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j} + \sum_{r=1}^k \sum_{s=1}^l \beta_{rs} X_{t-r} \varepsilon_{t-s},$$

onde $\theta_0 = 1$, ϕ_i ($i = 1, \dots, p$), θ_j ($j = 1, \dots, q$), β_{rs} ($r = 1, \dots, k$, $s = 1, \dots, l$), são constantes reais.

Note-se que estes modelos são claramente uma generalização dos modelos lineares $ARMA(p, q)$.

Granger e Andersen concentraram os seus estudos numa classe particular destes modelos, a dos *modelos bilineares simples*.

Definição.

1. Um processo estocástico real $X = (X_t, t \in \mathbb{Z})$ verifica um *modelo bilinear simples* de ordens k, l ($BL_S(k, l)$) se X_t verifica a seguinte equação auto-regressiva

$$X_t = \beta X_{t-k} \varepsilon_{t-l} + \varepsilon_t, \quad k > 0, l > 0,$$

onde β é um número real.

2. Se $k > l$, o modelo bilinear simples diz-se *superdiagonal*. Dir-se-á *diagonal* se $k = l$ e *subdiagonal* se $k < l$.

O estudo destes modelos começou por ser desenvolvido supondo $(\varepsilon_t, t \in \mathbb{Z})$ uma sucessão de v.a. independentes e identicamente distribuídas centradas e de variância constante.

A análise probabilista destes modelos, em particular no que diz respeito às questões de estacionaridade, levou ao estudo da função de autocorrelação de X que se concluiu ser nula na maior parte dos casos e, como tal, muito próxima da de um ruído branco ou de outros modelos lineares do tipo média móvel.

Este aspecto pode obviamente acarretar problemas aquando da fase de identificação em termos dos momentos de segunda ordem. Assim, com o objectivo de obter uma caracterização diferente para X , foram desenvolvidos estudos sobre o comportamento dos momentos de ordens superiores a 2, em particular das autocorrelações do processo $X^2 = (X_t^2, t \in \mathbb{Z})$.

Nos casos dos modelos superdiagonal e diagonal, foram obtidas, em Li (1984) as primeiras $k - 1$ autocorrelações do processo X^2 , quando o processo ε é uma sucessão de v.a. normais independentes de média zero e variância σ^2 , $\sigma > 0$. Sob as mesmas hipóteses para ε , Gabr (1988) obtém os momentos de terceira ordem de um qualquer modelo bilinear simples.

Como já foi observado, as hipóteses de independência e de normalidade do processo de erro são questionáveis por não serem compatíveis com a descrição empírica de muitas séries reais, em particular as séries financeiras. Neste sentido, encontramos desenvolvimentos importantes dos estudos anteriormente referidos, sob hipóteses de dependência para o processo de erro, em Martins (1997, 1999, 1999a, 2000).

5.2.2 Modelo bilinear simples

Consideremos o processo bilinear simples, $X = (X_t, t \in \mathbb{Z})$, definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) e obedecendo à equação de evolução

$$X_t = \beta X_{t-k} \varepsilon_{t-l} + \varepsilon_t, \quad k > 0, l > 0 \quad (5.2.2.1)$$

onde β é um parâmetro real e $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ uma sucessão de v.a. independentes e identicamente distribuídas com média zero e variância σ^2 , $\sigma > 0$.

Note-se que nestas condições o processo de erro é forte e fracamente estacionário.

Consideremos a seguinte hipótese sobre tal processo de erro:

$$\mathcal{H}_B: E(|\ln |\varepsilon_t||) < +\infty; \quad \ln |\beta| + E(\ln |\varepsilon_t|) < 0.$$

Estacionaridade forte e inversibilidade

No estudo do modelo bilinear simples vamos começar por assegurar a existência e unicidade de uma solução fortemente estacionária.

Teorema. Seja X um processo estocástico real verificando o modelo bilinear simples acima definido. Sob as hipóteses gerais consideradas para o processo de erro, existe uma solução fortemente estacionária para o modelo (5.2.2.1) dada por

$$X_t = \varepsilon_t + \sum_{n=1}^{+\infty} \beta^n \varepsilon_{t-nk} \prod_{j=1}^n \varepsilon_{t-l-(j-1)k}, \quad P - q.c., \quad t \in \mathbb{Z}.$$

Mais, o processo X assim definido é, $P - q.c.$, a única solução fortemente estacionária de (5.2.2.1).

Prova. Usando recursivamente a relação (5.2.2.1), obtém-se

$$\begin{aligned} X_t &= \beta X_{t-k} \varepsilon_{t-l} + \varepsilon_t \\ &= \beta^2 X_{t-2k} \varepsilon_{t-k-l} \varepsilon_{t-l} + \beta \varepsilon_{t-k} \varepsilon_{t-l} + \varepsilon_t \\ &= \dots \\ &= \varepsilon_t + \sum_{i=1}^n T_i(t) + \beta^{n+1} X_{t-(n+1)k} \prod_{j=0}^n \varepsilon_{t-l-jk}, \quad n \in \mathbb{N}_0, \end{aligned} \quad (5.2.2.2)$$

onde, para cada $t \in \mathbb{Z}$, definimos $T_i(t) = \beta^i \varepsilon_{t-ik} \prod_{j=1}^i \varepsilon_{t-l-(j-1)k}$, $i = 1, \dots, n$, convencionando $\sum_{i=1}^0 T_i(t) = 0$.

Ora, como ε é um processo de componentes independentes e identicamente distribuídas, decorre do teorema de Kolmogorov (Gonçalves, Mendes-Lopes, 2000, p.207) e da hipótese \mathcal{H}_B ,

$$\begin{aligned} &\lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \ln |\beta^n \prod_{j=1}^n \varepsilon_{t-l-(j-1)k}| \\ &= \lim_{n \rightarrow +\infty} (q.c.) \left(\frac{1}{n} \ln |\beta^n| + \frac{1}{n} \sum_{j=1}^n \ln |\varepsilon_{t-l-(j-1)k}| \right) \\ &= \ln |\beta| + \lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \sum_{j=1}^n \ln |\varepsilon_{t-l-(j-1)k}| \end{aligned}$$

$$= \ln |\beta| + E(\ln |\varepsilon_t|) = \gamma < 0.$$

Repare-se agora que

$$\frac{1}{n} \ln |T_n(t)| = \frac{1}{n} \ln |\beta^n \prod_{j=1}^n \varepsilon_{t-l-(j-1)k}| + \frac{1}{n} \ln |\varepsilon_{t-nk}|.$$

Além disso, $\lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \ln |\varepsilon_{t-nk}| = 0$. De facto, a estacionaridade forte de (ε_t) implica

$$P(\{\omega : \lim_{n \rightarrow +\infty} \frac{1}{n} \ln |\varepsilon_{t-nk}(\omega)| = 0\}) \\ = P(\{\omega : \lim_{n \rightarrow +\infty} \frac{1}{n} \ln |\varepsilon_t(\omega)| = 0\}) = 1,$$

pois $\ln |\varepsilon_t(\omega)|$ não depende de n .

Concluimos assim que

$$\lim_{n \rightarrow \infty} (q.c.) \frac{1}{n} \ln |T_n(t)| = \gamma < 0,$$

pelo que

$$\lim_{n \rightarrow \infty} (q.c.) |T_n(t)|^{1/n} = e^\gamma < 1.$$

O critério de Cauchy para séries numéricas de termos positivos permite então afirmar que a série $\sum_{n=1}^{+\infty} T_n(t)$, $t \in \mathbb{Z}$, é *q.c.* convergente, definindo assim um processo fortemente estacionário uma vez que é função mensurável de $\varepsilon_t, \varepsilon_{t-1}, \dots$ e ε é um processo fortemente estacionário ⁽¹⁾. Definindo

$$Y_t = \varepsilon_t + \sum_{n=1}^{+\infty} T_n(t), \quad t \in \mathbb{Z},$$

é claro que o processo $Y = (Y_t, t \in \mathbb{Z})$ é também fortemente estacionário.

¹A estacionaridade forte de uma série temporal $(X_t, t \in \mathbb{Z})$ implica a estacionaridade forte da série temporal $(Z_t, t \in \mathbb{Z})$ com

$$Z_t = f(\dots, X_{t-1}, X_t, X_{t+1}, \dots),$$

para toda a função mensurável f definida em $\mathbb{R}^{\mathbb{Z}}$ e com valores em \mathbb{R} (Azencott, Dacunha-Castelle, 1984, p.30).

Torna-se ainda simples verificar que este processo satisfaz a equação de recorrência (5.2.2.1). De facto,

$$\beta Y_{t-k} \varepsilon_{t-l} + \varepsilon_t = \beta \left(\varepsilon_{t-k} + \sum_{n=1}^{+\infty} \beta^n \varepsilon_{t-k-nk} \prod_{j=1}^n \varepsilon_{t-k-l-(j-1)k} \right) \varepsilon_{t-l} + \varepsilon_t$$

pelo que

$$\begin{aligned} \beta Y_{t-k} \varepsilon_{t-l} + \varepsilon_t &= \varepsilon_t + \beta \varepsilon_{t-k} \varepsilon_{t-l} + \sum_{n=1}^{+\infty} \beta^{n+1} \varepsilon_{t-(n+1)k} \prod_{j=0}^n \varepsilon_{t-l-jk} \\ &= \varepsilon_t + \sum_{n=0}^{+\infty} \beta^{n+1} \varepsilon_{t-(n+1)k} \prod_{j=0}^n \varepsilon_{t-l-jk} \\ &= \varepsilon_t + \sum_{n=1}^{+\infty} \beta^n \varepsilon_{t-nk} \prod_{j=0}^{n-1} \varepsilon_{t-l-jk} = Y_t. \end{aligned}$$

Vejamos agora que a solução encontrada é a única solução fortemente estacionária. Tomando limites em (5.2.2.2), observamos que qualquer solução X de (5.2.2.1), em particular uma solução X fortemente estacionária, verifica

$$X_t = Y_t + \lim_{n \rightarrow \infty} (q.c.) \beta^{n+1} X_{t-(n+1)k} \prod_{j=0}^n \varepsilon_{t-l-jk},$$

desde que o limite exista.

Seja $Z_n(t) = \beta^n X_{t-nk} \prod_{j=0}^{n-1} \varepsilon_{t-l-jk}$. Tem-se então

$$\begin{aligned} \lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \ln |Z_n(t)| &= \lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \ln |X_{t-nk}| \\ &\quad + \lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \ln |\beta^n \prod_{j=0}^{n-1} \varepsilon_{t-l-jk}| \\ &= \gamma < 0, \end{aligned}$$

uma vez que $\lim_{n \rightarrow +\infty} (q.c.) \frac{1}{n} \ln |X_{t-nk}| = 0$, pelo facto de X ser fortemente estacionário.

Então

$$\lim_{n \rightarrow +\infty} (q.c.) |Z_n(t)| = \lim_{n \rightarrow +\infty} (q.c.) e^{n(\frac{1}{n} \ln |Z_n(t)|)} = 0.$$

Assim, se X é uma qualquer solução fortemente estacionária de (5.2.2.1) tem-se $X_t = Y_t$, P - $q.c.$, o que assegura a unicidade $q.c.$ da solução fortemente estacionária. ■

Notemos que, sob a hipótese \mathcal{H}_B , qualquer função real mensurável de $((X_t, \varepsilon_t), t \in \mathbb{Z})$ é um processo fortemente estacionário uma vez que é função mensurável de $(\varepsilon_t, t \in \mathbb{Z})$.

O teorema seguinte, cuja prova omitiremos, estabelece uma condição suficiente para a inversibilidade de um qualquer modelo bilinear simples.

Teorema. Seja X um processo estocástico real verificando o modelo bilinear simples acima definido. Sob as hipóteses gerais consideradas para o processo de erro e supondo ainda que a solução fortemente estacionária é tal que $E(|\ln |X_t||) < +\infty$ e $\ln |\beta| + E(\ln |X_t|) < 0$, então o modelo (5.2.2.1) é invertível sendo

$$\varepsilon_t = X_t + \sum_{n=1}^{\infty} (-\beta)^n X_{t-nk} \prod_{j=1}^n X_{t-l-(j-1)k}, \quad P - q.c., \quad t \in \mathbb{Z}.$$

Comentário sobre a prova. De (5.2.2.1) tem-se

$$\varepsilon_t = -\beta X_{t-k} \varepsilon_{t-l} + X_t.$$

Então a prova deste resultado decorre de modo imediato de uma versão mais geral do teorema anterior correspondente aos modelos bilineares simples em que o processo de erro é apenas fortemente estacionário e não necessariamente constituído por v.a. independentes (Martins, 2000, p. 27).

Estrutura de 2ª ordem

Vamos agora estudar a estrutura de 2ª ordem de um modelo bilinear simples, isto é, analisar as condições de estacionaridade

fraca e determinar as correspondentes funções média e de covariância.

Nota-se que este estudo depende da relação entre k e l . Assim, deverão ser estudados separadamente os casos superdiagonal, diagonal e subdiagonal. Ilustraremos tais cálculos no caso diagonal. O estudo dos outros casos pode ser encontrado em Martins (2000), sendo de referir a sua maior complexidade no caso do modelo subdiagonal.

Iniciamos este estudo com a apresentação de uma condição necessária e suficiente para a estacionaridade forte e fraca de um modelo bilinear simples diagonal $X = (X_t, t \in \mathbb{Z})$, isto é, tal que

$$X_t = \beta X_{t-k} \varepsilon_{t-k} + \varepsilon_t, \quad k > 0.$$

Teorema. Seja X um processo estocástico real verificando o modelo bilinear simples diagonal tal que o processo de erro $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é uma sucessão de v.a. independentes e identicamente distribuídas com média e momento de 3ª ordem nulos, variância σ^2 , momento de 4ª ordem μ_4 ($\sigma > 0, \mu_4 > 0$) e tal que $E(|\ln |\varepsilon_t||) < +\infty$. Nestas condições, o processo X é forte e fracamente estacionário se e somente se $\beta^2 \sigma^2 < 1$.

Prova. Suponhamos em primeiro lugar que $\beta^2 \sigma^2 < 1$; nestas condições,

$$2 \ln |\beta| < -\ln E(\varepsilon_t^2).$$

Tendo em conta a convexidade da função $g(x) = -\ln x$, $x \in \mathbb{R}^+$, e o facto de $E(|\ln |\varepsilon_t||) < +\infty$, podemos aplicar a desigualdade de Jensen (Rudin (1987), p.62) à v.a. ε_t^2 . Assim

$$-\ln E(\varepsilon_t^2) \leq E(-\ln \varepsilon_t^2) = -2E(\ln |\varepsilon_t|),$$

e então $\ln |\beta| + E(\ln |\varepsilon_t|) < 0$, o que assegura a estacionaridade forte de X .

Vamos agora provar que a condição $\beta^2 \sigma^2 < 1$ assegura ainda a estacionaridade fraca de X , isto é, a existência de momento de 2ª ordem de X_t , $\forall t \in \mathbb{Z}$.

Consideremos então o modelo $X_t = \beta X_{t-k} \varepsilon_{t-k} + \varepsilon_t$, $k > 0$.

Para cada $i \in \mathbb{N}$, consideremos $T_i = T_i(t)$ o processo definido anteriormente no caso particular em que $l = k$, i.e.,

$$T_i(t) = \beta^i \varepsilon_{t-ik}^2 \prod_{j=1}^{i-1} \varepsilon_{t-jk}, \quad t \in \mathbb{Z},$$

e seja $Y_t = \varepsilon_t + \sum_{i=1}^{+\infty} T_i(t)$, $t \in \mathbb{Z}$. Da estacionaridade forte de X temos

$$Y_t = X_t, \quad P - q.c..$$

Assim a tese decorre de $E(Y_t^2) < +\infty, \forall t \in \mathbb{Z}$.

Ora,

$$Y_t^2 \leq \varepsilon_t^2 + 2|\varepsilon_t| \sum_{i=1}^{+\infty} |T_i| + \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |T_i T_j|,$$

pelo que

$$E(Y_t^2) \leq E(\varepsilon_t^2) + 2 \sum_{i=1}^{+\infty} E(|\varepsilon_t| |T_i|) + \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} E(|T_i T_j|),$$

uma vez que temos integrais de funções mensuráveis positivas.

Vamos agora analisar a convergência de cada uma das séries presentes no segundo membro da desigualdade.

Relativamente a $\sum_{i=1}^{+\infty} E(|\varepsilon_t| |T_i|)$ temos, usando a desigualdade de Schwartz, para cada $i \in \mathbb{N}$,

$$\begin{aligned} E(|\varepsilon_t| |T_i|) &= |\beta^i| E[(|\varepsilon_t| |\varepsilon_{t-ik}|) (|\varepsilon_{t-k}| |\varepsilon_{t-2k}| \cdots |\varepsilon_{t-ik}|)] \\ &\leq |\beta^i| [E(\varepsilon_t^2 \varepsilon_{t-ik}^2)]^{1/2} [E(\varepsilon_{t-k}^2 \varepsilon_{t-2k}^2 \cdots \varepsilon_{t-ik}^2)]^{1/2} \\ &= \sigma^2 \left[(\beta^2 \sigma^2)^{1/2} \right]^i, \end{aligned}$$

tendo em conta o facto das v.a. ε_t serem independentes e identicamente distribuídas.

Como $(\beta^2 \sigma^2)^{1/2} < 1$, a série de termo geral $E(|\varepsilon_t| |T_i|)$ é convergente.

Quanto à série $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} E(|T_i T_j|)$, tem-se, analogamente ao caso anterior, para cada $i, j \in \mathbb{N}$,

$$\begin{aligned} E(|T_i T_j|) &= |\beta|^{i+j} E \left[(\varepsilon_{t-ik}^2 |\varepsilon_{t-k}| |\varepsilon_{t-2k}| \cdots |\varepsilon_{t-(i-1)k}|) \right. \\ &\quad \left. (\varepsilon_{t-jk}^2 |\varepsilon_{t-k}| |\varepsilon_{t-2k}| \cdots |\varepsilon_{t-(j-1)k}|) \right] \\ &\leq |\beta|^{i+j} \left[E \left(\varepsilon_{t-ik}^4 \varepsilon_{t-k}^2 \varepsilon_{t-2k}^2 \cdots \varepsilon_{t-(i-1)k}^2 \right) \right]^{1/2} \\ &\quad \left[E \left(\varepsilon_{t-jk}^4 \varepsilon_{t-k}^2 \varepsilon_{t-2k}^2 \cdots \varepsilon_{t-(j-1)k}^2 \right) \right]^{1/2} \\ &= \frac{\mu_4}{\sigma^2} \left[(\beta^2 \sigma^2)^{1/2} \right]^{i+j}, \end{aligned}$$

o que assegura a convergência da série de termo geral $E(|T_i T_j|)$.

Como $E(\varepsilon_t^2) = \sigma^2 < +\infty$ obtém-se $E(Y_t^2) < +\infty$.

Suponhamos agora que X é forte e fracamente estacionário. Então existe $E(X_t^2)$ e o seu valor é independente de t . Partindo directamente da relação de recorrência que define X para calcular $E(X_t^2)$, obtém-se

$$\begin{aligned} E(X_t^2) &= \beta^2 E(X_t^2 \varepsilon_t^2) + E(\varepsilon_t^2) + 2\beta E(X_{t-k} \varepsilon_{t-k} \varepsilon_t) \\ &= \beta^2 E(X_t^2 \varepsilon_t^2) + \sigma^2, \end{aligned}$$

tendo em conta a estacionaridade forte dos produtos envolvidos e a independência das variáveis aleatórias ε_t . Como $E(X_t^2) < +\infty$, a igualdade anterior permite concluir que $E(X_t^2 \varepsilon_t^2) < +\infty$ e

$$\begin{aligned} E(X_t^2 \varepsilon_t^2) &= \beta^2 E(X_{t-k}^2 \varepsilon_{t-k}^2 \varepsilon_t^2) + E(\varepsilon_t^4) + 2\beta E(X_{t-k} \varepsilon_{t-k} \varepsilon_t^3) \\ &= \beta^2 \sigma^2 E(X_t^2 \varepsilon_t^2) + \mu_4, \end{aligned}$$

que é equivalente a

$$(1 - \beta^2 \sigma^2) E(X_t^2 \varepsilon_t^2) = \mu_4,$$

de onde decorre $\beta^2 \sigma^2 < 1$, pois $E(X_t^2 \varepsilon_t^2) < +\infty$ e $\mu_4 > 0$. ■

Passemos agora ao cálculo das funções média e de autocorrelação de um modelo diagonal verificando as hipóteses gerais do teorema anterior e ainda tal que $\beta^2 \sigma^2 < 1$.

Note-se que a hipótese $E(\varepsilon_t^3) = 0$, que tem a vantagem de simplificar bastante os cálculos, não é muito restritiva no contexto dos processos de erro que seguem frequentemente leis simétricas.

O valor médio de X_t é, neste caso, dado por

$$\begin{aligned} E(X_t) &= \beta E(X_{t-k} \varepsilon_{t-k}) + E(\varepsilon_t) \\ &= \beta E(X_t \varepsilon_t) \\ &= \beta^2 E(X_{t-k} \varepsilon_{t-k} \varepsilon_t) + \beta E(\varepsilon_t^2) \\ &= \beta \sigma^2, \end{aligned}$$

uma vez que da independência das v.a. (ε_t) decorre

$$E(X_{t-k} \varepsilon_{t-k} \varepsilon_t) = E(X_{t-k} \varepsilon_{t-k}) E(\varepsilon_t) = 0,$$

pois $X_{t-k} = f(\varepsilon_{t-k}, \varepsilon_{t-k-1}, \dots)$.

Relativamente ao cálculo da função de autocorrelação, decorre da prova do teorema anterior que

$$E(X_t^2) = \beta^2 E(X_t^2 \varepsilon_t^2) + \sigma^2,$$

com

$$E(X_t^2 \varepsilon_t^2) = \frac{\mu_4}{1 - \beta^2 \sigma^2}.$$

Consequentemente,

$$E(X_t^2) = \frac{\beta^2 \mu_4}{1 - \beta^2 \sigma^2} + \sigma^2.$$

Quanto ao valor de $E(X_t X_{t-j})$, $j \in \mathbb{N}$, temos, seguindo um procedimento análogo aos anteriores,

$$E(X_t X_{t-j}) = \beta E(X_{t-k} \varepsilon_{t-k} X_{t-j}).$$

O valor da esperança do segundo membro desta igualdade depende da relação entre j e k . Assim, se $j = k$,

$$\begin{aligned} E(X_t X_{t-k}) &= \beta E(X_t^2 \varepsilon_t) \\ &= \beta^3 E(X_{t-k}^2 \varepsilon_{t-k}^2 \varepsilon_t) + \beta E(\varepsilon_t^3) + 2\beta^2 E(X_{t-k} \varepsilon_{t-k} \varepsilon_t^2) \\ &= 2\beta^2 \sigma^2 E(X_t \varepsilon_t) = 2\beta^2 \sigma^4. \end{aligned}$$

Então

$$\text{Cov}(X_t, X_{t-k}) = E(X_t X_{t-k}) - [E(X_t)]^2 = 2\beta^2 \sigma^4 - (\beta \sigma^2)^2 = \beta^2 \sigma^2.$$

Quando $j < k$, a estacionaridade dos processos envolvidos na igualdade $E(X_t X_{t-j}) = \beta E(X_{t-k} \varepsilon_{t-k} X_{t-j})$, permite-nos escrevê-la na forma

$$E(X_t X_{t-j}) = \beta E(X_{t-k+j} \varepsilon_{t-k+j} X_t).$$

Assim,

$$\begin{aligned} E(X_t X_{t-j}) &= \beta^2 E(X_{t-k+j} \varepsilon_{t-k+j} X_{t-k} \varepsilon_{t-k}) + \\ &\quad + \beta E(X_{t-k+j} \varepsilon_{t-k+j} \varepsilon_t) \\ &= \beta^2 E(X_t \varepsilon_t X_{t-j} \varepsilon_{t-j}) \\ &= \beta^3 E(X_{t-k} \varepsilon_{t-k} \varepsilon_t X_{t-j} \varepsilon_{t-j}) + \beta^2 E(\varepsilon_t^2 X_{t-j} \varepsilon_{t-j}) \\ &= \beta^2 \sigma^2 E(X_t \varepsilon_t) = \beta^2 \sigma^4, \end{aligned}$$

pois $E(X_t \varepsilon_t) = \sigma^2$, como se pode observar no cálculo de $E(X_t)$. Então, $\rho_X(j) = 0$, $0 < j < k$.

De forma análoga prova-se que $\rho_X(j) = 0$, $j > k$.

A função de autocorrelação de (X_t) no caso diagonal é então dada por

$$\rho_X(j) = \begin{cases} 1 & \text{se } j = 0 \\ \beta^2 \sigma^4 / V(X_t) & \text{se } j = k, -k \\ 0 & \text{nos outros casos, } j \in \mathbb{Z}, \end{cases}$$

com $V(X_t) = \sigma^2(1 - \beta^2 \sigma^2) + \beta^2 \mu_4 / (1 - \beta^2 \sigma^2)$.

Nota-se que esta é também a função de autocorrelação de um modelo $MA(k)$, o que vem ao encontro da observação anteriormente feita sobre a identificação de um modelo bilinear com base nos momentos de 2ª ordem.

Observemos finalmente que o estudo correspondente para os modelos superdiagonais, bem como alguns dos subdiagonais, apenas exige que sejam verificadas hipóteses análogas sobre o processo

de erro até à ordem 2. Além disso, a condição necessária e suficiente de estacionaridade forte e fraca de tais modelos é sempre $\beta^2\sigma^2 < 1$.

Exemplo. Consideremos um processo estocástico $X = (X_t, t \in \mathbb{Z})$ definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) tal que

$$X_t = \beta X_{t-2}\varepsilon_{t-1} + \varepsilon_t,$$

onde $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco de componentes independentes, de variância finita σ^2 , tal que $E(|\ln|\varepsilon_t||) < +\infty$ e $\beta^2\sigma^2 < 1$.

Vejamos que X é fracamente estacionário.

De facto, comecemos por observar que, nas condições do enunciado, X verifica um modelo bilinear simples superdiagonal de ordens $(2, 1)$. Além disso, X é fortemente estacionário. Por outro lado, definindo para cada $i \in \mathbb{N}$, o processo $T_i = T_i(t)$ tal que

$$T_i(t) = \beta^i \varepsilon_{t-2i} \prod_{j=1}^i \varepsilon_{t-2j+1}, \quad t \in \mathbb{Z},$$

provamos que $Y_t = \varepsilon_t + \sum_{i=1}^{+\infty} T_i(t)$, $t \in \mathbb{Z}$, é uma solução fortemente estacionária do modelo anterior. Assim, a unicidade da solução fortemente estacionária permite concluir que

$$Y_t = X_t, \quad P - q.c..$$

Para estabelecer a estacionaridade fraca de X , basta mostrar que $E(Y_t^2) < +\infty$, $\forall t \in \mathbb{Z}$.

Ora,

$$Y_t^2 \leq \varepsilon_t^2 + 2|\varepsilon_t| \sum_{i=1}^{+\infty} |T_i| + \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} |T_i T_j|,$$

pelo que

$$E(Y_t^2) \leq E(\varepsilon_t^2) + 2 \sum_{i=1}^{+\infty} E(|\varepsilon_t| |T_i|) + \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} E(|T_i T_j|),$$

uma vez que temos o integral de funções mensuráveis positivas.

Vamos agora analisar a convergência de cada uma das séries presentes no segundo membro da desigualdade.

Relativamente a $\sum_{i=1}^{+\infty} E(|\varepsilon_t||T_i|)$ temos para cada $i \in \mathbb{N}$,

$$\begin{aligned} E(|\varepsilon_t||T_i|) &= |\beta|^i |E[|\varepsilon_t||\varepsilon_{t-2i}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2i+1}|]| \\ &= |\beta|^i |E(|\varepsilon_t|^2)| [E(|\varepsilon_t|)]^i \\ &\leq |\beta|^i \sigma^2 (\sigma^2)^{\frac{i}{2}} = \sigma^2 (\beta^2 \sigma^2)^{\frac{i}{2}} \end{aligned}$$

tendo em conta o facto das v.a. (ε_t) serem independentes e identicamente distribuídas e usando a desigualdade de Shwartz.

A série de termo geral $E(|\varepsilon_t||T_i|)$ é convergente pois $(\beta^2 \sigma^2)^{1/2} < 1$.

Quanto à série $\sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} E(|T_i T_j|)$, tem-se, usando o mesmo tipo de argumentos,

$$\begin{aligned} &E(|T_i T_j|) \\ &= |\beta|^{i+j} E[(|\varepsilon_{t-2i}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2i+1}|)(|\varepsilon_{t-2j}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2j+1}|)] \\ &= \begin{cases} |\beta|^{2i} E(\varepsilon_{t-2i}^2 |\varepsilon_{t-1}|^2 \dots |\varepsilon_{t-2i+1}|^2), & \text{se } i = j \\ |\beta|^{i+j} E[(|\varepsilon_{t-2i}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2i+1}|)(|\varepsilon_{t-2j}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2j+1}|)], & \text{se } i \neq j \end{cases} \\ &\leq \begin{cases} \sigma^2 (\beta^2 \sigma^2)^i, & i = j \\ |\beta|^{i+j} \left(E[(|\varepsilon_{t-2i}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2i+1}|)^2] \right)^{\frac{1}{2}} \times \\ \quad \times \left(E[(|\varepsilon_{t-2j}||\varepsilon_{t-1}| \dots |\varepsilon_{t-2j+1}|)^2] \right)^{\frac{1}{2}}, & i \neq j \end{cases} \\ &= \begin{cases} \sigma^2 (\beta^2 \sigma^2)^i, & i = j \\ \sigma^2 |\beta|^{i+j} (\sigma^2)^{\frac{i+j}{2}}, & i \neq j \end{cases} \\ &= \sigma^2 (\beta^2 \sigma^2)^{\frac{i+j}{2}}, \end{aligned}$$

o que assegura a sua convergência.

Podemos finalmente concluir que, nas condições do enunciado, $E(Y_t^2) < +\infty$.

Determinemos agora as funções média e de autocorrelação de X .

O valor médio de X_t é, neste caso, dado por

$$E(X_t) = \beta E(X_{t-2}\varepsilon_{t-1}) + E(\varepsilon_t) = 0,$$

uma vez que da independência das v.a. (ε_t) decorre

$$E(X_{t-2}\varepsilon_{t-1}) = E(X_{t-2})E(\varepsilon_{t-1}) = 0,$$

pois $X_{t-2} = f(\varepsilon_{t-2}, \varepsilon_{t-3}, \dots)$.

Quanto à função de autocorrelação tem-se

$$\begin{aligned} E(X_t^2) &= \beta^2 E(X_{t-2}^2 \varepsilon_{t-1}^2) + E(\varepsilon_t^2) + 2\beta E(X_{t-2}\varepsilon_{t-1}\varepsilon_t) \\ &= \beta^2 E(X_{t-2}^2)E(\varepsilon_{t-1}^2) + \sigma^2 \\ &= \beta^2 E(X_t^2)\sigma^2 + \sigma^2, \end{aligned}$$

pelo que

$$E(X_t^2) = \frac{\sigma^2}{1 - \beta^2\sigma^2}.$$

Por outro lado, $\forall j \in \mathbb{N}$,

$$E(X_t X_{t-j}) = \beta E(X_{t-2}\varepsilon_{t-1} X_{t-j}) + E(\varepsilon_t X_{t-j})$$

$$= \beta E(X_{t-2}\varepsilon_{t-1} X_{t-j})$$

$$= \begin{cases} 0, & j \geq 2 \\ \beta E(X_{t-2}\varepsilon_{t-1} (\beta X_{t-3}\varepsilon_{t-2} + \varepsilon_{t-1})), & j = 1 \end{cases}$$

$$= \begin{cases} 0, & j \geq 2 \\ \beta^2 E(X_{t-2} X_{t-3} \varepsilon_{t-1} \varepsilon_{t-2}) + E(X_{t-2} \varepsilon_{t-1}^2), & j = 1 \end{cases}$$

$$= \begin{cases} 0, & j \geq 2 \\ \beta^2 E(X_{t-2} X_{t-3} \varepsilon_{t-2}) E(\varepsilon_{t-1}) + E(X_{t-2}) E(\varepsilon_{t-1}^2), & j = 1 \end{cases}$$

$$= 0.$$

A função de autocorrelação de (X_t) é então dada por

$$\rho_X(j) = \begin{cases} 1, & \text{se } j = 0 \\ 0, & \text{se } j \in \mathbb{Z} - \{0\} \end{cases}$$

com $V(X_t) = \frac{\sigma^2}{1-\beta^2\sigma^2}$.

Nota-se que esta função é também a função de autocorrelação de um ruído branco com aquela variância.

Estas conclusões sobre os momentos de ordem dois dos modelos bilineares levam-nos a reafirmar a importância da análise do comportamento dos momentos de ordens superiores.

Um estudo detalhado e completo dos momentos de ordem 3 e das autocorrelações do quadrado do processo em modelos bilineares simples com processos de erro bastante gerais pode ser encontrado em Martins (2000) onde foi também efectuada uma análise empírica sobre a aplicação de tais estudos à identificação das ordens dos modelos.

Capítulo 6

Inferência em modelos não lineares

Na análise estatística dos modelos não lineares são utilizadas as metodologias usuais de estimação e testes em séries temporais. Assim, o tratamento estatístico de tais modelos, no que concerne, por exemplo, à estimação dos parâmetros envolvidos, aos testes aos parâmetros ou à previsão, é abordado de modo clássico segundo os métodos de tratamento estatístico do modelo linear geral (Gouriéroux, Monfort, 1989). As principais técnicas de inferência desenvolvidas assentam então na determinação do estimador da máxima verosimilhança (ou pseudo-máxima verosimilhança) dos parâmetros presentes no modelo. Este estudo será aqui ilustrado no contexto dos modelos condicionalmente heteroscedásticos.

As primeiras técnicas estatísticas em modelos condicionalmente heteroscedásticos tiveram por base um modelo de regressão com erros ARCH condicionalmente Gaussianos (Engle, 1982, 1983). Em Weiss (1986) e Bollerslev (1986), entre outros, foram considerados modelos de regressão com erros heteroscedásticos mais complexos; todos estes modelos estão, no entanto, incluídos numa classe geral de processos $(Y_t, t \in \mathbb{Z})$ verificando as seguintes condições (Gouriéroux, Monfort, 1992a):

$$\begin{cases} E(Y_t | \underline{Y}_{t-1}, \underline{X}_t) = m_t(\theta) = m(\underline{Y}_{t-1}, \underline{X}_t; \theta) \\ V(Y_t | \underline{Y}_{t-1}, \underline{X}_t) = h_t^2(\theta) = h^2(\underline{Y}_{t-1}, \underline{X}_t; \theta) \end{cases},$$

isto é, onde os momentos condicionais podem depender simultaneamente dos valores passados do processo (dinâmica endógena) e dos valores presente e passados de certas características exógenas ($X_t, t \in \mathbb{Z}$), bem como de um parâmetro vectorial θ desconhecido.

Neste capítulo detalharemos a determinação dos estimadores da máxima verosimilhança dos parâmetros envolvidos num modelo de regressão com erros ARCH.

Relativamente à outra classe de modelos estudada neste texto, a dos modelos bilineares, optamos por apresentar aqui uma metodologia não clássica de inferência estatística baseada na separação assintótica de leis de probabilidade ⁽¹⁾. Esta metodologia, proposta inicialmente em Geffroy (1980), é aqui aplicada à construção de um teste convergente para modelos bilineares diagonais simples segundo um estudo apresentado em Gonçalves, Jacob e Mendes-Lopes (2000). Tal teste permitirá distinguir entre um modelo bilinear simples e um processo de erro sem o recurso aos momentos. Tem assim, entre outras, a vantagem de poder ser usado em modelos bilineares sem momentos, enquanto que uma tal discriminação que os envolva obriga, como já observamos, à existência de momentos de ordem superior a 2.

6.1 Estimação num modelo com erros ARCH

Consideremos o modelo de regressão com erros ARCH, ($Y_t, t \in \mathbb{Z}$), definido por

$$Y_t = X_t b + \varepsilon_t,$$

onde ($X_t, t \in \mathbb{Z}$) é um processo estocástico real de variáveis exógenas e ($\varepsilon_t, t \in \mathbb{Z}$) segue um modelo ARCH(q) condicionalmente

¹Dado um modelo estatístico $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T), P_\theta^T)_{\theta \in \{\theta_0, \theta_1\}}$, $T \in \mathbb{N}$, dizemos, de acordo com Geffroy(1980), que as duas famílias de leis, $P_{\theta_0}^T$ and $P_{\theta_1}^T$ se *separam assintoticamente* se existe uma sucessão de borelianos, A_T , tais que

$$\lim_{T \rightarrow +\infty} P_{\theta_0}^T(A_T) = 1 \text{ e } \lim_{T \rightarrow +\infty} P_{\theta_1}^T(A_T) = 0.$$

Gaussiano tal que

$$\begin{cases} E(\varepsilon_t | \underline{Y}_{t-1}, \underline{X}_t) = 0 \\ V(\varepsilon_t | \underline{Y}_{t-1}, \underline{X}_t) = h_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \end{cases},$$

com $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, q$, desconhecidos.

Este modelo é um caso particular do modelo geral anteriormente considerado com

$$\begin{cases} m_t(\theta) = X_t b \\ h_t^2(\theta) = \alpha_0 + \sum_{i=1}^q \alpha_i (Y_{t-i} - X_{t-i} b)^2 \end{cases}$$

e onde $\theta' = (b, \alpha_0, \alpha_1, \dots, \alpha_q)$; o processo de inovação segue um modelo ARCH(q).

Estimemos θ pelo método da máxima verosimilhança condicional (Gouriéroux, Monfort, 1989).

Consideremos T observações de Y a partir de um certo instante inicial t_0 . Sem perda de generalidade tomemos $t_0 = 0$ e designemos por y_1, \dots, y_T tais observações.

Como ε é condicionalmente Gaussiano, a lei condicional de Y_t dado $(\underline{Y}_{t-1}, \underline{X}_t)$ é normal de valor médio $X_t b$ e variância $h_t^2(\theta)$. Assim a função de verosimilhança associada à observação (y_1, \dots, y_T) e condicional ao passado é

$$\begin{aligned} L_T(y_1, \dots, y_T; \theta) &= f_{(Y_1, \dots, Y_T)}^{Y_0, X_T}(y_1, \dots, y_T) \\ &= f_{Y_1}^{Y_0, X_T}(y_1) f_{Y_2}^{Y_1, X_T}(y_2) \dots f_{Y_T}^{Y_{T-1}, X_T}(y_T) \\ &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi} h_t(\theta)} \exp\left(-\frac{1}{2} \left(\frac{y_t - x_t b}{h_t(\theta)}\right)^2\right) \\ &= \left(\prod_{t=1}^T \frac{1}{\sqrt{2\pi} h_t(\theta)}\right) \exp\left(-\frac{1}{2} \sum_{t=1}^T \left(\frac{y_t - x_t b}{h_t(\theta)}\right)^2\right), \end{aligned}$$

designando por $f_Z^{\mathcal{B}}$ a densidade condicional à tribo \mathcal{B} de um vector aleatório Z e por x_t o valor particular de X_t .

A log-verossimilhança condicional ao passado é então

$$\begin{aligned} \log L_T(y_1, \dots, y_T; \theta) &= -\frac{T}{2} \log 2\pi - \sum_{t=1}^T \log h_t(\theta) - \frac{1}{2} \sum_{t=1}^T \left(\frac{y_t - x_t b}{h_t(\theta)} \right)^2 \\ &= -\frac{T}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right) - \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2}{\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2}. \end{aligned}$$

O estimador, $\hat{\theta}_T = (\hat{b}, \hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_q)'$ é então definido como solução do problema de maximização $\max_{\theta} (\log L_T(y_1, \dots, y_T; \theta))$; assim, tal estimador é solução do sistema

$$\begin{cases} \frac{\partial}{\partial \alpha_0} \log L_T(y_1, \dots, y_T; \theta) = 0 \\ \forall i \in \{1, \dots, q\}, \frac{\partial}{\partial \alpha_i} \log L_T(y_1, \dots, y_T; \theta) = 0 \\ \frac{\partial}{\partial b} \log L_T(y_1, \dots, y_T; \theta) = 0 \end{cases} .$$

Mas, tendo em conta que

$$\forall i \in \{0, 1, \dots, q\}, \frac{\partial}{\partial \alpha_i} \varepsilon_t = 0 \text{ e } \frac{\partial}{\partial b} \varepsilon_t = -x_t,$$

aquele sistema pode explicitar-se na forma

$$\begin{cases} -\frac{1}{2} \sum_{t=1}^T \frac{1}{\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2} + \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2}{\left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right)^2} = 0 \\ \forall i \in \{1, \dots, q\}, -\frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_{t-i}^2}{\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2} + \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2 \varepsilon_{t-i}^2}{\left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right)^2} = 0 \\ \sum_{t=1}^T \frac{\sum_{i=1}^q \alpha_i x_{t-i} \varepsilon_{t-i}}{\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2} - \sum_{t=1}^T \frac{\varepsilon_t (-x_t) \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right) + \varepsilon_t^2 \sum_{i=1}^q \alpha_i x_{t-i} \varepsilon_{t-i}}{\left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 \right)^2} = 0 \end{cases}$$

ou de modo equivalente

$$\left\{ \begin{array}{l} \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2 - h_t^2}{h_t^4} = 0 \\ \forall i \in \{1, \dots, q\}, \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_{t-i}^2 (\varepsilon_t^2 - h_t^2)}{h_t^4} = 0 \\ \sum_{t=1}^T \frac{\varepsilon_t x_t h_t^2 - (\varepsilon_t^2 - h_t^2) \sum_{i=1}^q \alpha_i x_{t-i} \varepsilon_{t-i}}{h_t^4} = 0 \end{array} \right. .$$

Das propriedades do estimador da máxima verosimilhança (Gouriéroux, Monfort, 1989) decorrem, sob certas condições de regularidade, a convergência e a normalidade assintótica do estimador obtido. Tem-se, com efeito, que a lei limite de

$$\sqrt{T} (\hat{\theta}_T - \theta) = \sqrt{T} (\hat{b}_T - b, \hat{\alpha}_0 - \alpha_0, \hat{\alpha}_1 - \alpha_1, \dots, \hat{\alpha}_q - \alpha_q)'$$

é uma lei normal centrada cuja matriz de covariância é a inversa da matriz de informação

$$I = E \left(\frac{\partial \log f_{Y_t}^{Y_{t-1}, X_t}(y_t)}{\partial \theta} \frac{\partial \log f_{Y_t}^{Y_{t-1}, X_t}(y_t)}{\partial \theta'} \right).$$

Prova-se que (Gouriéroux, 1992, cap. IV)

$$I^{-1} = \begin{bmatrix} I_{bb}^{-1} & 0 \\ 0 & I_{\alpha\alpha}^{-1} \end{bmatrix},$$

onde $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_q)'$, $I_{bb} = E \left(\frac{\partial^2 \log f_{Y_t}^{Y_{t-1}, X_t}(y_t)}{\partial b^2} \right)$ e

$$I_{\alpha\alpha} = E \left(\frac{\partial \log f_{Y_t}^{Y_{t-1}, X_t}(y_t)}{\partial \alpha} \frac{\partial \log f_{Y_t}^{Y_{t-1}, X_t}(y_t)}{\partial \alpha'} \right).$$

Este facto permite concluir que os estimadores de b e de α , respectivamente \hat{b}_T e $\hat{\alpha}_T = (\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_q)'$, são assintoticamente não correlacionados. Assim, a estimação de b e α pode ser considerada separadamente sem perda de eficiência. Além disso, qualquer um dos parâmetros b ou α pode ser estimado sem perda de eficiência usando um estimador convergente do outro.

Este aspecto levou Engle a considerar, logo no seu trabalho pioneiro, a estimação de (b, α') em duas etapas a partir do estimador dos mínimos quadrados de b . Tal estimador permitiu-lhe obter, ainda pelo mesmo método, um estimador eficiente de α ; o estimador de b é então melhorado, numa segunda etapa, usando a referida estimação eficiente de α .

De facto, para o modelo linear inicialmente considerado, a estimação de b pelo método dos mínimos quadrados conduz ao estimador

$$\tilde{b}_T = \frac{\sum_{t=1}^T Y_t X_t}{\sum_{t=1}^T X_t^2}.$$

Desta estimação deduzem-se os resíduos estimados

$$\tilde{\varepsilon}_t = Y_t - \tilde{b}_T X_t.$$

Podem então deduzir-se estimadores convergentes de $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_q)$ fazendo uma regressão pelo método dos mínimos quadrados $\tilde{\varepsilon}_t^2$ sobre $(1, \tilde{\varepsilon}_{t-1}^2, \dots, \tilde{\varepsilon}_{t-q}^2)$ tendo em conta que

$$\tilde{\varepsilon}_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \tilde{\varepsilon}_{t-i}^2 + u_t.$$

Tais estimadores $\tilde{\alpha}_{0,T}, \tilde{\alpha}_{1,T}, \dots, \tilde{\alpha}_{q,T}$ são usados numa segunda etapa para, tendo em conta o fenómeno de heteroscedasticidade, melhorar a estimação de b . Com efeito, a variância condicional pode agora ser aproximada por

$$\tilde{h}_t^2 = \tilde{\alpha}_{0,T} + \sum_{i=1}^q \tilde{\alpha}_{i,T} \tilde{\varepsilon}_{t-i}^2.$$

Assim, o estimador de 2ª etapa de b , $\tilde{\tilde{b}}_T$, é então obtido fazendo uma regressão de Y_t sobre X_t mas tomando não a variância do

erro mas a respectiva variância condicional, designadamente, \tilde{h}_t^2 . Então

$$\tilde{b}_T = \frac{\sum_{t=1}^T \left(\frac{Y_t X_t}{\tilde{h}_t^2} \right)}{\sum_{t=1}^T \left(\frac{X_t^2}{\tilde{h}_t^2} \right)}.$$

O estudo destes novos estimadores permitiu concluir que são, tais como os estimadores da máxima verosimilhança, assintoticamente não correlacionados mas, relativamente a estes, menos eficientes. Prova-se que estes dois estimadores estão próximos na vizinhança da homoscedasticidade (*i.e.*, quando $\alpha_1 = 0, \dots, \alpha_q = 0$), mas que para certos valores dos parâmetros $\alpha_1, \dots, \alpha_q$ o ganho em eficiência pode ser bastante grande usando o estimador da máxima verosimilhança. A melhor eficiência deste estimador à medida que nos afastamos da homoscedasticidade tem naturalmente a ver com o facto de que a sua determinação envolve a lei condicional e consequentemente de modo mais directo a própria variância condicional.

Um estudo mais completo das questões aqui abordadas pode ser encontrado em Gouriéroux, 1992a, cap. IV.

6.2 Um teste para modelos bilineares

Para um modelo bilinear simples diagonal de ordem 1, vamos testar a significância do parâmetro real presente estabelecendo a separação assintótica da sucessão de distribuições obtidas sob a hipótese nula, correspondente a um processo de ruído branco, e sob a hipótese alternativa, relativa ao referido modelo bilinear. A definição de separação assintótica permite-nos observar que esta metodologia atribuirá um papel mais simétrico às duas hipóteses em teste do que a metodologia clássica de Neyman-Pearson, onde a hipótese nula é sempre privilegiada.

O estudo que vamos desenvolver está bastante dependente das propriedades de estacionaridade forte e de ergodicidade de um modelo bilinear. Uma breve referência à ergodicidade de processos será então considerada no primeiro parágrafo desta secção.

6.2.1 Ergodicidade de uma série temporal

A noção de ergodicidade de uma série temporal $X = (X_t, t \in \mathbb{Z})$, definida sobre um espaço de probabilidade (Ω, \mathcal{A}, P) , está essencialmente ligada à análise do comportamento, em média, do conjunto das trajectórias de X e ao facto de apenas acontecimentos de probabilidade zero ou um serem invariantes para um tal processo.

Começemos por precisar tal noção de invariância.

Seja P_X a lei do processo X sobre $(\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R}^{\mathbb{Z}}))$ ⁽²⁾.

Consideremos a seguinte função translação τ de $\mathbb{R}^{\mathbb{Z}}$ em $\mathbb{R}^{\mathbb{Z}}$,

$$\tau[(x_t, t \in \mathbb{Z})] = (y_t, t \in \mathbb{Z}), \quad \text{com } y_t = x_{t+1}.$$

Diz-se que um acontecimento A de $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$ é *invariante* se $\tau(A) = A$. A classe dos acontecimentos invariantes forma uma sub- σ -álgebra de $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$, a que se chama *σ -álgebra invariante* do processo X .

Além disso, P_X é *invariante* por τ se $P_X(\tau^{-1}(A)) = P_X(A)$, para todo o $A \in \mathcal{B}(\mathbb{R}^{\mathbb{Z}})$.

Definição. Um processo fortemente estacionário $X = (X_t, t \in \mathbb{Z})$ diz-se *ergódico* se, para quaisquer $A, B \in \mathcal{B}(\mathbb{R}^{\mathbb{Z}})$,

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=1}^n P_X(A \cap \tau^k(B)) = P_X(A) P_X(B),$$

com $\tau^{k+1}(B) = \tau(\tau^k(B))$, $k \in \mathbb{N}$, $\tau^1(B) = \tau(B)$.

Da definição anterior conclui-se que se X é um processo ergódico e A um acontecimento invariante então $P_X(A) = 0$ ou $P_X(A) = 1$. De facto, fazendo $B = A$ na igualdade presente na definição vem $P_X(A) = (P_X(A))^2$ pois $\tau^k(A) = A$.

Aliás, nota-se que esta propriedade é uma caracterização da ergodicidade de X (cf. Rosenblatt (1978)). Assim, a σ -álgebra invariante de um processo ergódico é trivial.

² $\mathcal{B}(\mathbb{R}^{\mathbb{Z}})$ é a menor σ -álgebra sobre $\mathbb{R}^{\mathbb{Z}}$ que torna mensuráveis todas as aplicações coordenadas.

Intuitivamente isto significa que qualquer trajectória de X cobre completamente, por translações sucessivas, o espaço $\mathbb{R}^{\mathbb{Z}}$, o que está naturalmente associado a uma grande irregularidade das trajectórias. Esta imparcialidade (independência em média) das trajectórias permite obter uma propriedade média do tipo lei dos grandes números, devida a Birkoff e von Neumann, a que se chama *teorema ergódico*.

Teorema. Seja $X = (X_t, t \in \mathbb{Z})$ uma série temporal fortemente estacionária e ergódica. Então, se g é uma função de $\mathbb{R}^{\mathbb{Z}}$ em \mathbb{R} , P_X -integrável, tem-se

$$\lim_{T \rightarrow +\infty} \frac{1}{T} [g(x) + g \circ \tau(x) + \dots + g \circ \tau^T(x)] \stackrel{P_X - q.c.}{=} \int_{\mathbb{R}^{\mathbb{Z}}} g(x) P_X(dx).$$

Suponhamos que g é da forma

$$\forall x = (x_n)_{n \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}}, \quad g(x) = f(x_1, \dots, x_k),$$

com f de \mathbb{R}^k em \mathbb{R} tal que $E[f(X_1, \dots, X_k)]$ existe.

O teorema ergódico permite escrever

$$\lim_{T \rightarrow +\infty} \frac{1}{T} [f(X_1, \dots, X_k) + f(X_2, \dots, X_{k+1}) + \dots + f(X_{T+1}, \dots, X_{T+k})] \stackrel{P - q.c.}{=} E[f(X_1, \dots, X_k)].$$

Em particular, se $f = \mathbb{I}_A$, com $A \subset \mathbb{R}^k$, vem

$$\lim_{T \rightarrow +\infty} \frac{1}{T} [\mathbb{I}_A(X_1, \dots, X_k) + \dots + \mathbb{I}_A(X_{T+1}, \dots, X_{T+k})] \stackrel{P - q.c.}{=} P[(X_1, \dots, X_k) \in A],$$

o que permite concluir que a lei conjunta de X_1, \dots, X_k (ou, mais geralmente, a família das leis de dimensão finita do processo) é determinada pela sucessão X_1, \dots, X_T , $T \rightarrow +\infty$. Isto é, os processos ergódicos são identificáveis a partir de uma trajectória infinita.

Observe-se finalmente que a validade das convergências anteriores para quaisquer f e k (resp. A e k) é equivalente à ergodicidade do processo fortemente estacionário X (Azencott, Dacunha-Castelle, 1984, p. 32).

Dois casos importantes de processos ergódicos dizem respeito às sucessões de variáveis aleatórias independentes e identicamente distribuídas e às funções mensuráveis de processos ergódicos. Com efeito, qualquer sucessão $(X_t, t \in \mathbb{Z})$ de v.a.i.i.d. constitui um processo ergódico (Azencott, Dacunha-Castelle, 1984, p. 32); além disso, se $(X_t, t \in \mathbb{Z})$ é um processo ergódico o mesmo acontece com o processo $(Z_t, t \in \mathbb{Z})$, tal que

$$Z_t = f(\dots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \dots),$$

com f uma função mensurável definida sobre $\mathbb{R}^{\mathbb{Z}}$ (Rosenblatt, 1978, p.33).

6.2.2 O modelo: hipóteses e propriedades gerais

Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico real verificando o modelo bilinear simples diagonal de ordem 1, isto é,

$$X_t = \varphi X_{t-1} \varepsilon_{t-1} + \varepsilon_t, \forall t \in \mathbb{Z},$$

onde

(C1) $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ é uma sucessão de v.a.i.i.d., centradas podendo ou não ter variância;

(C2) φ é um número real tal que $E(|\log |\varepsilon_t||) < +\infty$ e
 $E(\log |\varepsilon_t|) + \log |\varphi| < 0$.

Nestas condições, sabemos do capítulo anterior que o processo X tal que

$$X_t \stackrel{q.c.}{=} \sum_{n=1}^{+\infty} \varphi^n \varepsilon_{t-n}^2 \prod_{i=1}^{n-1} \varepsilon_{t-i} + \varepsilon_t, \quad \forall t \in \mathbb{Z},$$

onde, por convenção, $\prod_{i=1}^{n-1} \varepsilon_{t-i} = 1$ se $n = 1$, é $P - q.c.$ a única solução fortemente estacionária do modelo anterior.

Além disso, sendo uma função mensurável do processo ergódico ε , é também um processo ergódico.

Por outro lado, sob as condições anteriores, o modelo é invertível desde que

$$(C3) \ E(|\log |X_t||) < +\infty \text{ e } E(\log |X_t|) + \log |\varphi| < 0.$$

Tem-se então

$$\varepsilon_t \stackrel{q.c.}{=} \sum_{n=1}^{+\infty} (-\varphi)^n X_{t-n}^2 \prod_{i=1}^{n-1} X_{t-i} + X_t, \quad \forall t \in \mathbb{Z},$$

convencionando-se $\prod_{i=1}^{n-1} X_{t-i} = 1$ se $n = 1$.

A avaliação de $E(|\log |X_t||)$ é em geral difícil. É no entanto possível encontrar boas condições suficientes de inversibilidade. Se, por exemplo, ε é um ruído branco Gaussiano centrado e reduzido, temos que a condição $|\varphi| < 0.88$ é suficiente para a verificação de $E(\log |X_t|) + \log |\varphi| < 0$ (Gonçalves, Martins, M. Lopes, 2002). Assim, nestas condições o modelo bilinear diagonal simples correspondente é fortemente estacionário, ergódico e invertível.

Vamos então assumir que o modelo considerado verifica as condições gerais (C1), (C2) e (C3) relativas à estacionaridade forte, ergodicidade e inversibilidade do mesmo. Sob estas condições, vamos também considerar o processo $Y = (Y_t, t \in \mathbb{Z})$ tal que

$$Y_t = X_t \varepsilon_t \stackrel{(q.c.)}{=} X_t \left(\sum_{n=1}^{+\infty} (-\varphi)^n X_{t-n}^2 \prod_{i=1}^{n-1} X_{t-i} + X_t \right).$$

Note-se que este processo é fortemente estacionário e ergódico pois $Y_t = f(X_t, X_{t-1}, X_{t-2}, \dots)$.

6.2.3 Um procedimento de decisão

Para o modelo anteriormente definido, consideremos os dois conjuntos de hipóteses em ensaio estatístico

$$H_0 : \varphi = 0 \text{ contra } H_1 : \varphi = \beta \ (\beta > 0 \text{ fixo}),$$

$$H_0 : \varphi = 0 \text{ contra } H'_1 : \varphi = \beta \ (\beta < 0 \text{ fixo}).$$

Para cada um dos ensaios, vamos construir um procedimento de decisão que permite distinguir os modelos relacionados com estas

hipóteses estabelecendo a separação assintótica das duas famílias de leis de probabilidade associadas aos correspondentes processos estocásticos.

As regiões de separação são naturalmente diferentes para cada um dos conjuntos de hipóteses, pelo que começamos por definir as regiões de separação associadas a cada um dos casos. No entanto, como a análise dos procedimentos é muito semelhante, limitar-nos-emos a detalhar o caso correspondente à distinção entre H_0 e H_1 .

Denotemos por P_φ e E_φ , respectivamente, a lei do modelo e a sua função esperança matemática quando o parâmetro do modelo é φ .

Na construção do procedimento de decisão usamos T observações do processo X , denotadas por x_1, x_2, \dots, x_T .

Seja $g : \mathbb{R} \rightarrow \mathbb{R}^+$ uma função simétrica, estritamente positiva, não decrescente sobre \mathbb{R}^+ e P_φ -integrável, que designamos por *função peso*. Definam-se os subconjuntos de \mathbb{R}^2 :

$$D = \left\{ (u, v) : u > 0; v < \frac{\beta}{2}u \right\} \cup \left\{ (u, v) : u < 0; v > \frac{\beta}{2}u \right\}, \quad \beta > 0,$$

$$D^* = \left\{ (u, v) : u < 0; v < \frac{\beta}{2}u \right\} \cup \left\{ (u, v) : u > 0; v > \frac{\beta}{2}u \right\}, \quad \beta < 0,$$

e considerem-se as seguintes regiões do espaço $\prod_{-\infty}^T \mathbb{R}$,

$$A_T = \left\{ x^{(T)} : \sum_{t=2}^T g(\beta y_{t-1}) [2 \mathbf{1}_D(y_{t-1}, x_t) - 1] \geq 0 \right\}, \quad \beta > 0,$$

$$A_T^* = \left\{ x^{(T)} : \sum_{t=2}^T g(\beta y_{t-1}) [2 \mathbf{1}_{D^*}(y_{t-1}, x_t) - 1] \geq 0 \right\}, \quad \beta < 0,$$

onde $T \in \mathbb{N}$, $x^{(t)} = (\dots, x_{t-1}, x_t) \in \prod_{-\infty}^t \mathbb{R}$ e y_t designa o valor particular de Y_t associado.

Supondo que a lei de ε_t tem mediana única e nula, vamos mostrar que a sucessão de borelianos $(A_T, T \in \mathbb{N})$ (resp., A_T^* , $T \in \mathbb{N}$) separa assintoticamente as leis P_0 e P_β .

A definição das regiões de separação tem uma justificação heurística simples.

Com efeito, consideremos $\beta > 0$ e $y_{t-1} > 0$ e suponhamos que a observação particular do processo nos conduz a

$$\sum_{t=2}^T [2 \mathbf{1}_D(y_{t-1}, x_t) - 1] > 0.$$

Neste caso, as parcelas

$$2 \mathbf{1}_D(y_{t-1}, x_t) - 1 = 2 \mathbf{1}_{\{x_t < \frac{\beta}{2} y_{t-1}\}} - 1, \quad t = 2, \dots, T,$$

tomam mais vezes o valor 1 do que o valor -1 . Então, atendendo a que a mediana condicional da lei de X_t dado ε_{t-1} é φy_{t-1} , parece mais verosímil que seja $\varphi = 0$. Por outro lado, se

$$\sum_{t=2}^T [2 \mathbf{1}_D(y_{t-1}, x_t) - 1] < 0$$

a hipótese mais verosímil parece ser $\varphi = \beta$.

Mais, quando $\frac{\beta}{2} y_{t-1}$ é um valor muito próximo de zero, os acontecimentos $\{X_t < \frac{\beta}{2} y_{t-1}\}$ e $\{X_t > \frac{\beta}{2} y_{t-1}\}$ têm probabilidades muito próximas. Assim, é natural introduzir uma função peso que dê maior relevância a maiores valores de βy_{t-1} .

O teorema seguinte assegura a convergência do procedimento de decisão acima definido.

Teorema. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico real verificando o modelo bilinear simples anteriormente definido, sujeito às condições (C1), (C2) e (C3). Se a lei de ε_t tem mediana única e nula então $(A_T, T \in \mathbb{N})$ (resp. $(A_T^*, T \in \mathbb{N})$) é uma sucessão de

regiões de aceitação de um procedimento de decisão que permite discriminar

$$H_0 : \varphi = 0 \text{ contra } H_1 : \varphi = \beta, \beta > 0$$

(resp., H_0 contra $H'_1 : \varphi = \beta, \beta < 0$).

Prova. Definam-se as funções

$$\Psi(y_{t-1}, x_t) = g(\beta y_{t-1}) [2 \mathbf{1}_D(y_{t-1}, x_t) - 1], \quad t = 2, \dots, T$$

e seja $\bar{\Psi}_T$ a v.a.r.

$$\bar{\Psi}_T = \frac{1}{T} \sum_{t=2}^T \Psi_t(Y_{t-1}, X_t) = \frac{1}{T} \sum_{t=2}^T \Psi_t.$$

Do teorema ergódico decorre

$$\lim_{T \rightarrow +\infty} \bar{\Psi}_T = E_\varphi(\Psi(Y_1, X_2)) \quad (\text{q.c.}).$$

Estudemos o sinal deste limite, sob cada uma das hipóteses, quando $\beta > 0$.

Temos que

$$\begin{aligned} E_\varphi(\bar{\Psi}_T) &= E_\varphi(E_\varphi(\bar{\Psi}_T / \underline{X}_1)) \\ &= E_\varphi(g(\beta Y_1) [2E_\varphi(\mathbf{1}_D(Y_1, X_2) / \underline{X}_1) - 1]) \end{aligned}$$

e

$$\begin{aligned} E_\varphi(\mathbf{1}_D(Y_1, X_2) / \underline{X}_1) &= E_\varphi(\mathbf{1}_{\mathbb{R}^+}(Y_1) \mathbf{1}_{]-\infty, \frac{\beta}{2} Y_1[}(X_2) \\ &\quad + \mathbf{1}_{\mathbb{R}^-}(Y_1) \mathbf{1}_{] \frac{\beta}{2} Y_1, +\infty[}(X_2) / \underline{X}_1) \\ &= \mathbf{1}_{\mathbb{R}^+}(Y_1) P_\varphi \left[X_2 < \frac{\beta}{2} Y_1 / \underline{X}_1 \right] \\ &\quad + \mathbf{1}_{\mathbb{R}^-}(Y_1) P_\varphi \left[X_2 > \frac{\beta}{2} Y_1 / \underline{X}_1 \right]. \end{aligned}$$

Sob H_0 , obtém-se

$$E_0(\mathbf{1}_D(Y_1, X_2) / \underline{X}_1) = \begin{cases} F(\frac{\beta}{2} Y_1^-) & \text{se } Y_1 > 0, \\ 1 - F(\frac{\beta}{2} Y_1) & \text{se } Y_1 < 0, \\ 0 & \text{se } Y_1 = 0, \end{cases}$$

onde F é a função de distribuição da lei comum às v.a.r. ε_t .

Assim, como $Y_1 = X_1\varepsilon_1 \neq 0$, q.c., pois $E(|\log |\varepsilon_t||) < +\infty$ e $E(|\log |X_t||) < +\infty$, obtém-se

$$E_0(\mathbf{1}_D(Y_1, X_2)/\underline{X}_1) > \frac{1}{2} \text{ (q.c.)}$$

tendo em conta a nulidade e unicidade da mediana da lei comum às v.a.r. ε_t .

Então

$$g(\beta Y_1) [2E_0(\mathbf{1}_D(Y_1, X_2)/\underline{X}_1) - 1] > 0 \text{ (q.c.)}$$

e

$$\lim_{T \rightarrow \infty} (\text{q.c.}) \overline{\Psi}_T > 0.$$

Assim, existe um número natural T_0 tal que, para $T \geq T_0$, se tem $\overline{\Psi}_T \geq 0$, ou, de modo equivalente, $\mathbf{1}_{\{\overline{\Psi}_T \geq 0\}} = 1$. Consequentemente

$$\lim_{T \rightarrow \infty} (\text{q.c.}) \mathbf{1}_{\{\overline{\Psi}_T \geq 0\}} = 1.$$

Pelo teorema da convergência dominada chegamos então a

$$\lim_{T \rightarrow \infty} P_0 [\{\overline{\Psi}_T \geq 0\}] = \lim_{T \rightarrow \infty} P_0 (A_T) = 1.$$

Por outro lado, sob H_1 , tem-se

$$\begin{aligned} E_\beta(\mathbf{1}_D(Y_1, X_2)/\underline{X}_1) &= \mathbf{1}_{\mathbb{R}^+}(Y_1) P_\beta \left[\beta Y_1 + \varepsilon_2 < \frac{\beta}{2} Y_1 / \underline{X}_1 \right] + \\ &\quad + \mathbf{1}_{\mathbb{R}^-}(Y_1) P_\beta \left[\beta Y_1 + \varepsilon_2 > \frac{\beta}{2} Y_1 / \underline{X}_1 \right] \\ &= \begin{cases} F(-\frac{\beta}{2} Y_1^-) & \text{se } Y_1 > 0, \\ 1 - F(-\frac{\beta}{2} Y_1) & \text{se } Y_1 < 0, \\ 0 & \text{se } Y_1 = 0; \end{cases} \end{aligned}$$

assim, a unicidade e nulidade da mediana da lei comum às v.a.r. ε_t implica

$$E_\beta(\mathbf{1}_D(Y_1, X_2)/\underline{X}_1) < \frac{1}{2},$$

pelo que

$$g(\beta Y_1) [2E_\beta(\mathbf{1}_D(Y_1, X_2)/\underline{X}_1) - 1] < 0, \quad (\text{q.c.})$$

e então, tal como vimos anteriormente,

$$\lim_{T \rightarrow \infty} P_\beta [\{\bar{\Psi}_T \geq 0\}] = P_\beta(A_T) = 0. \quad \blacksquare$$

Em Gonçalves, Jacob e Mendes-Lopes (2000) foi ainda estabelecida a velocidade de convergência de $(P_0(A_T))_{T \in \mathbb{N}}$, que se concluiu ser exponencial.

De facto, pode enunciar-se o seguinte resultado, no caso correspondente a $\beta > 0$ (se $\beta < 0$ tem-se um resultado análogo para $P_0(A_T^*)$):

Teorema. Seja $X = (X_t, t \in \mathbb{Z})$ um processo estocástico real verificando um modelo bilinear simples diagonal de ordem 1 sujeito às condições (C1), (C2) e (C3). Se a lei marginal de $(\varepsilon_t, t \in \mathbb{Z})$, de função de distribuição F , é absolutamente contínua de densidade simétrica e decrescente em \mathbb{R}^+ e se a função de peso g é definida por $g(x) = 2F(|x|) - 1$ então o procedimento de decisão proposto verifica

$$P_0(A_T) \geq 1 - \left\{ E_0 \left[\exp \left(-\frac{1}{2} g^2 \left(\frac{\beta}{2} \varepsilon_1^2 \right) \right) \right] \right\}^{T-1}, \quad \forall T \in \mathbb{N}.$$

6.2.4 Implementação do procedimento de decisão como um teste

O resultado anterior dá-nos essencialmente um resultado teórico, uma vez que as regiões de decisão A_T (resp. A_T^*) dependem, através de Y_t , do processo de erro não observável. De facto, $Y_t = Y_t(\varphi)$ é função de X_t e de ε_t que, por sua vez, é função de X_t, X_{t-1}, \dots e de φ .

Na prática, para usar este procedimento como um teste das hipóteses

$$H_0 : \varphi = 0 \quad \text{contra} \quad H_1 : \varphi = \beta, \quad \beta > 0$$

(resp. H_0 contra $H'_1 : \varphi = \beta, \beta < 0$),

podemos, por exemplo, considerar

$$Y_t = Y_t(\beta) = X_t \left(\sum_{n=1}^{+\infty} (-\beta)^n X_{t-n}^2 \prod_{i=1}^{n-1} X_{t-i} + X_t \right)$$

e assim Y_t escrever-se-á como função do valor conhecido β e do processo observável X .

Vejamus que, nestas condições, se $\beta > 0$, P_0 e P_β se separaram ainda assintoticamente e, conseqüentemente, definimos deste modo um teste convergente para aquelas hipóteses.

Com efeito, sob H_1 , estamos nas mesmas condições do teorema anterior. Assim, a prova desta separação assintótica resume-se a estabelecer que, sob H_0 ,

$$\lim_{T \rightarrow \infty} P_0 [\{\bar{\Psi}_T \geq 0\}] = P_0 (\hat{A}_T) = 1,$$

$$\text{com } \hat{A}_T = \left\{ x^{(T)} : \sum_{t=2}^T g(\beta y_{t-1}(\beta)) [2 \mathbf{1}_D(y_{t-1}(\beta), x_t) - 1] \geq 0 \right\}.$$

Esta prova segue exactamente os mesmos passos que a anterior substituindo, para todo o t , Y_t por $Y_t(\beta)$.

Em conclusão, se $\beta > 0$ e sendo $y_t(\beta)$ um valor particular de $Y_t(\beta)$, este teste conduz-nos a aceitar H_0 se (x_1, \dots, x_T) é tal que

$$\sum_{t=2}^T g(\beta y_{t-1}(\beta)) [2 \mathbf{1}_D(y_{t-1}(\beta), x_t) - 1] \geq 0,$$

e a rejeitá-la no caso contrário.

Nota-se que a região crítica do teste tem directamente em conta as trajectórias do processo.

É óbvio que, se $\beta < 0$, tudo se passa, com alterações evidentes, de modo semelhante.

6.2.5 Estudo de simulação

Ilustração do comportamento do teste

Ilustremos a aplicabilidade deste teste com um estudo experimental de simulação usando um processo estocástico real $X = (X_t, t \in \mathbb{Z})$ seguindo o modelo bilinear

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

onde $(\varepsilon_t, t \in \mathbb{Z})$ é um ruído branco Gaussiano com variância unitária e $\varphi \in [0, 0.88]$ de modo a que o modelo seja forte e fracamente estacionário, ergódico e invertível.

A importância da função de peso presente na estatística de teste é estudada supondo em primeiro lugar as observações igualmente pesadas ($g = 1$) e, depois, introduzindo ponderações correspondentes à função $g(x) = x^2$. Ambas as funções são simétricas, não decrescentes em \mathbb{R}^+ e P_φ -integráveis; em particular, a integrabilidade de x^2 é equivalente à existência de $E(X_t^2)$, que nas condições deste estudo está assegurada.

O comportamento do teste é analisado com base em amostras de tamanhos 50 e 70, respectivamente.

Os valores de Y_t podem ser estimados por

$$\begin{cases} \hat{y}_t(\beta) = x_t \left(\sum_{n=1}^{t-1} (-\beta)^n x_{t-n}^2 \prod_{i=1}^{n-1} x_{t-i} + x_t \right), & t = 2, \dots, T \\ \hat{y}_1 = x_1^2 \end{cases},$$

o que corresponde a considerar na definição de $Y_t(\beta)$ as observações antes do instante 1 iguais a zero.

Para avaliar o comportamento do nosso teste quando H_0 é verdadeira, efectuou-se um estudo de simulação considerando $\varphi = 0$ e testando o modelo correspondente contra 5 alternativas ($\beta = 0.2, \beta = 0.25, \beta = 0.3, \beta = 0.4$ e $\beta = 0.5$). Para cada uma delas calculamos a proporção de rejeições de H_0 em 60 réplicas do modelo e determinamos os intervalos de confiança a 95% para esta proporção a partir de amostras de tamanho 30 da mesma.

Os resultados obtidos apresentam-se na tabela 1.

T		50	50	70	70
β	teste	$g(x)=1$	$g(x)=x^2$	$g(x)=1$	$g(x)=x^2$
0.2]0.26, 0.3[]0.17, 0.21[]0.23, 0.26[]0.13, 0.16[
0.25]0.23, 0.27[]0.12, 0.16[]0.18, 0.22[]0.09, 0.11[
0.3]0.19, 0.23[]0.09, 0.12[]0.15, 0.19[]0.05, 0.07[
0.4]0.13, 0.16[]0.04, 0.06[]0.08, 0.11[]0.02, 0.03[
0.5]0.08, 0.11[]0.007, 0.02[]0.05, 0.06[]0.001, 0.008[

Tabela 1. Estimativa intervalar para o erro de 1^a espécie

Observamos que este teste tem em geral um bom desempenho; de facto, se considerarmos, por exemplo, $T = 50$, vemos que o nível decresce à medida que o parâmetro sob a alternativa cresce. Além disso, observamos que este decréscimo é mais significativo quando as observações estão sujeitas a ponderação. Salientamos ainda a diminuição da proporção de rejeições de H_0 quando aumenta o número de observações.

Com o objectivo de estimar o erro de 2^a espécie, consideramos vários valores de β ($\beta = 0.2, 0.25, 0.3, 0.4, 0.5$). Registamos, em cada caso, a proporção de aceitação de H_0 (sendo H_0 falsa) em 60 réplicas do modelo. Os intervalos de confiança a 95% para esta proporção, correspondentes a amostras de dimensão 30, são apresentados na tabela 2.

T		50	50	70	70
β	teste	$g(x)=1$	$g(x)=x^2$	$g(x)=1$	$g(x)=x^2$
0.2]0.26, 0.3[]0.19, 0.23[]0.25, 0.29[]0.14, 0.16[
0.25]0.22, 0.26[]0.13, 0.16[]0.19, 0.22[]0.09, 0.12[
0.3]0.18, 0.21[]0.1, 0.12[]0.14, 0.17[]0.06, 0.08[
0.4]0.13, 0.17[]0.04, 0.06[]0.09, 0.11[]0.02, 0.03[
0.5]0.07, 0.1[]0.01, 0.02[]0.04, 0.07[]0.003, 0.01[

Tabela 2. Estimativa intervalar para o erro de 2^a espécie

Sob a veracidade da hipótese alternativa, o teste apresenta também um bom comportamento, particularmente evidente quando as observações são efectivamente ponderadas; tanto para $T = 50$

como para $T = 70$, a estimativa para o erro de 2ª espécie diminuiu fortemente, neste caso, à medida que β aumenta. A precisão da estimativa aumenta, em geral, com T .

Finalmente, podemos observar que estas simulações ilustram bem o comportamento simétrico deste teste, no sentido de que não privilegia nenhuma das hipóteses em análise. Além disso, levam-nos ainda a conjecturar que a velocidade de convergência da potência é elevada (possivelmente exponencial, como para o nível) e fortemente dependente da ponderação g das observações.

Análise comparativa do comportamento do teste

Outros estudos experimentais podem ser empreendidos. Designadamente, podem efectuar-se estudos de comparação com testes do “tipo Neyman-Pearson”. A este propósito, e seguindo o trabalho desenvolvido em Gonçalves, Martins, Mendes-Lopes (2002), detalhemos tais estudos comparando este procedimento, baseado na separação assintótica (aqui designado por AS) dos modelos em teste, com um teste não paramétrico (aqui designado por NP) do “tipo Neyman-Pearson” proposto em Wandji (1998).

Consideramos, neste estudo de comparação, a função de peso presente no teorema relativo à velocidade de convergência e que se prova conduzir ao majorante mínimo da sucessão de níveis.

Apresentemos, de forma breve, o referido teste NP .

Trata-se de um teste para modelos $(X_t, t \in \mathbb{Z})$ da forma

$$X_{t+1} = \sum_{k=0}^p a_k X_{t-k} + V(Y_t) + \tilde{\sigma}(U_t)\varepsilon,$$

onde $Y_t = (X_t, \dots, X_{t-q}, \varepsilon_t, \dots, \varepsilon_{t-Q})$, $U_t = (X_t, \dots, X_{t-L})$, p , q , Q e L são inteiros não negativos conhecidos e a função $V(\cdot)$ tem uma forma paramétrica conhecida. Além disso, os parâmetros a_k , $k = 0, \dots, p$, a distribuição de ε_t e a função $\tilde{\sigma}(\cdot)$ não são necessariamente conhecidos e $(\varepsilon_t, t \in \mathbb{Z})$ é uma sucessão de v.a.i.i.d. centradas e reduzidas tais que ε_{t+1} é independente da σ -álgebra \underline{X}_t .

As hipóteses em teste são

$$H_0 : V = V_0 \text{ contra } H_1 : V = V_0 + V_*.$$

Note-se que, sendo $(\varepsilon_t, t \in \mathbb{Z})$ uma sucessão de v.a.i.i.d. com a lei $N(0, 1)$, o modelo

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

é um caso particular do modelo anterior, com $p = q = Q = 0$, $a_0 = 0$, $V((X_t, \varepsilon_t)) = \varphi X_t \varepsilon_t$ e $\tilde{\sigma}(\cdot) = 1$, reduzindo-se as hipóteses em teste a

$$H_0 : \varphi = 0 \text{ contra } H_1 : \varphi = \beta.$$

O teste NP impõe a X as condições que a seguir detalhamos.

Existe $\gamma > 0$ tal que, sob $H_0 \cup H_1$,

i) o processo $((X_t, \varepsilon_t), t \in \mathbb{Z})$ é fortemente estacionário e α -misturador ⁽³⁾ com coeficientes de mistura, $(\alpha_k, k \geq 0)$, tais que

$$\sum_{k=0}^{+\infty} (k+1) \alpha_k^{\gamma/(2+\gamma)} < +\infty;$$

ii) $E(|X_1|^{2+\gamma}) < +\infty$.

Por outro lado, no caso do modelo bilinear em estudo, a estatística do teste NP toma a forma

$$W_T = \frac{1}{T} \left(\sum_{t=1}^T X_t \right)^2,$$

tendo-se que, sob H_0 , W_T segue a lei $\chi^2(1)$.

³Seja $X = (X_t, t \in \mathbb{Z})$ um p.e.r. definido sobre um espaço de probabilidade (Ω, \mathcal{A}, P) . Para j e l tais que $-\infty < j < l < +\infty$, seja \mathcal{A}_j^l a σ -álgebra gerada por $(X_j, X_{j+1}, \dots, X_{l-1}, X_l)$ e seja

$$\alpha_n = \sup_{j \in \mathbb{Z}} \{ |P(A \cap B) - P(A)P(B)|, A \in \mathcal{A}_{-\infty}^j, B \in \mathcal{A}_{j+n}^{+\infty} \}.$$

O processo X diz-se α -misturador se $\lim_{n \rightarrow +\infty} \alpha_n = 0$.

Assim, fixado um nível $\alpha \in]0, 1[$, a hipótese H_0 é rejeitada se

$$W_T > q_{1-\alpha},$$

onde $q_{1-\alpha}$ representa o quantil de probabilidade $(1 - \alpha)$ da distribuição $\chi^2(1)$.

Relativamente à potência do teste NP contra alternativas fixas, os resultados de Wandji permitem concluir que

$$\lim_{T \rightarrow +\infty} P_\beta(W_T > \chi) = 1, \text{ para cada } \chi > 0,$$

onde P_β designa a lei do processo X , sob a alternativa.

A comparação entre os procedimentos AS e NP vai ser efectuada para o modelo anterior mas supondo $|\varphi| < 0.75$, para garantir a aplicabilidade do teste NP .

Note-se que o teste AS é aplicável pois o processo ε é fortemente estacionário e ergódico e, como $|\varphi| < 0.88$, o processo X é fortemente estacionário, ergódico e invertível.

Por outro lado, o teste NP é também aplicável.

De facto,

i) como $\underline{X}_t = \underline{\varepsilon}_t$ a v.a. ε_{t+1} é independente da σ -álgebra \underline{X}_t ;

ii) a partir de uma representação markoviana de (X_t) , prova-se que o processo $((X_t, \varepsilon_t), t \in \mathbb{Z})$ é fortemente estacionário e geometricamente α -misturador, i.e. os coeficientes de mistura, $(\alpha_k, k \geq 0)$, são tais que $\alpha_k \leq ar^k$, para algum $a > 0$ e algum $r, 0 < r < 1$, o que implica que, para qualquer $\gamma > 0$, a série de termo geral $(k+1)(r^{\gamma/(2+\gamma)})^k, k \in \mathbb{N}$, é convergente;

iii) além disso, $E(X_t^4) < +\infty$ para valores de φ tais que $\varphi^4 E(\varepsilon_t^4) < 1$. Assim, se $|\varphi| < 0.75$, aquela condição é verificada com $\gamma = 2$.

Neste estudo comparativo considerou-se no teste NP , como já referimos, a função de peso associada ao teorema de velocidade de convergência, i.e.,

$$g(x) = 2F(|x|) - 1,$$

com F a f.d. da lei $N(0, 1)$.

Para o cálculo de F usou-se a seguinte aproximação, cujo erro para $x > 0$ não excede 10^{-7} ,

$$F(x) \cong 1 - f(x) (b_1\tau + b_2\tau^2 + b_3\tau^3 + b_4\tau^4 + b_5\tau^5), \quad x > 0,$$

onde $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$, $\tau = (1 + 0.2316419x)^{-1}$;

$b_1 = 0.319381530$; $b_2 = -0.356563782$; $b_3 = 1.7814777937$;

$b_4 = -1.821255978$; $b_5 = 1.330274429$.

Consideremos ainda as seguintes notações:

- $P_{AS}^{(0)}$: proporção de rejeições de H_0 quando $\varphi = 0$ (erro de 1ª espécie) usando o procedimento AS ,
- $P_{AS}^{(1)}$: proporção de rejeições de H_1 quando $\varphi = \beta$ (erro de 2ª espécie) usando o procedimento AS ,
- $P_{NP}^{(1)}$: proporção de rejeições de H_1 quando $\varphi = \beta$ (erro de 2ª espécie) usando o procedimento NP com nível 0.05.

Para comparar os erros de 1ª espécie considera-se o teste NP ao nível 0.05; para o teste AS apresentam-se, para cada T e β , intervalos de confiança a 95% para a proporção de rejeições de H_0 quando H_0 é verdadeira. A este propósito convém referir que o nível do teste apresentado poderia também ter sido fixado, em função de T , usando o majorante da sucessão de níveis dado pelo teorema de velocidade de convergência.

Por outro lado, na comparação dos erros de 2ª espécie determinam-se, para cada teste e para os vários valores de β , intervalos de confiança a 95% para a proporção de rejeições de H_1 quando H_1 é verdadeira.

Os quadros 1, 2, 3 e 4 resumem a informação obtida.

β	$P_{AS}^{(0)}$	$P_{AS}^{(1)}$	$P_{NP}^{(1)}$
0.05]0.40, 0.44[]0.39, 0.43[]0.93, 0.95[
0.1]0.29, 0.34[]0.30, 0.34[]0.88, 0.91[
0.2]0.15, 0.19[]0.15, 0.18[]0.70, 0.74[
0.35]0.04, 0.06[]0.05, 0.06[]0.34, 0.38[
0.5]0.007, 0.02[]0.01, 0.02[]0.13, 0.17[

Quadro 1. Erros associados aos testes AS e NP ($T=50$)

β	$P_{AS}^{(0)}$	$P_{AS}^{(1)}$	$P_{NP}^{(1)}$
0.05]0.36, 0.39[]0.35, 0.39[]0.92, 0.94[
0.1]0.23, 0.27[]0.23, 0.28[]0.83, 0.86[
0.2]0.07, 0.11[]0.08, 0.11[]0.47, 0.51[
0.25]0.04, 0.06[]0.04, 0.05[]0.31, 0.34[
0.3]0.02, 0.03[]0.02, 0.03[]0.18, 0.23[

Quadro 2. Erros associados aos testes AS e NP ($T=100$)

β	$P_{AS}^{(0)}$	$P_{AS}^{(1)}$	$P_{NP}^{(1)}$
0.05]0.25, 0.28[]0.26, 0.30[]0.85, 0.87[
0.1]0.11, 0.13[]0.11, 0.14[]0.59, 0.63[
0.14]0.04, 0.06[]0.04, 0.06[]0.32, 0.35[
0.25]0.0, 0.005[]0.001, 0.006[]0.01, 0.02[

Quadro 3. Erros associados aos testes AS e NP ($T=300$)

β	$P_{AS}^{(0)}$	$P_{AS}^{(1)}$	$P_{NP}^{(1)}$
0.05]0.19, 0.23[]0.20, 0.22[]0.75, 0.79[
0.1]0.04, 0.06[]0.03, 0.05[]0.30, 0.33[
0.25]0.00, 0.002[0.0]0.0, 0.002[

Quadro 4. Erros associados aos testes AS e NP ($T=600$)

Comecemos por observar que o teste NP , ao nível $\alpha = 0.05$, parece não ter um bom comportamento relativamente ao erro de

segunda espécie quando as amostras são pequenas, facto que se acentua se β está muito próximo de zero. Mais, a proporção $P_{NP}^{(1)}$ é sempre superior à proporção $P_{AS}^{(1)}$, embora a diferença entre ambas diminua com o aumento do tamanho da amostra, T .

Por outro lado, à medida que T aumenta, decresce o valor de β para o qual o intervalo de confiança para $P_{AS}^{(0)}$ contém o valor do nível considerado para o teste NP , $\alpha = 0.05$, notando-se que o procedimento AS se comporta globalmente melhor do que o teste NP para os valores de β tais que $P_{AS}^{(0)} \leq 0.05$. É de referir, finalmente, que quando a estimativa do erro de primeira espécie do procedimento AS é superior a 0.05, se tem em contrapartida que a estimativa do erro de 2ª espécie do procedimento NP é muito superior à do correspondente erro do procedimento AS .

Aplicação do teste a outros modelos bilineares

O teste aqui apresentado pode ser aplicado a uma classe muito vasta de modelos incluindo, em particular, os modelos lineares.

Em Gonçalves, Martins e Mendes-Lopes (2001) é desenvolvido este procedimento de decisão para um qualquer modelo bilinear simples.

De facto, seja X um processo seguindo o modelo

$$X_t = \varphi X_{t-k} \varepsilon_{t-l} + \varepsilon_t, \quad t \in \mathbb{Z},$$

com ε ruído branco independente verificando as hipóteses anteriormente referidas, sob as quais X é fortemente estacionário, ergódico e invertível.

O teste de região de aceitação

$$A_T = \left\{ x^{(T)}: \sum_{t=M+1}^T g(\beta(x_{t-k}\varepsilon_{t-l})) [2\mathbf{1}_D((x_{t-k}\varepsilon_{t-l}), x_t) - 1] \geq 0 \right\}$$

é um teste convergente de H_0 contra H_1 ($T > M$, $M = \max(k, l)$).

A velocidade de convergência exponencial para a sucessão de níveis é também obtida com a mesma escolha da função de peso, isto é, com $g(x) = 2F(|x|) - 1$, sendo F a função de distribuição de ε_t .

Vamos ilustrar a aplicação deste teste efectuando um estudo de simulação para o modelo superdiagonal $X = (X_t, t \in \mathbb{Z})$ tal que

$$X_t = \varphi X_{t-3} \varepsilon_{t-1} + \varepsilon_t,$$

onde ε é um ruído branco independente de Cauchy standard e $\varphi \in [0, 0.53]$.

Observa-se que o processo X considerado é fortemente estacionário, ergódico e invertível.

Neste estudo foram utilizadas as funções de peso

$$g = g_1 = 1 \quad \text{e} \quad g(x) = g_2(x) = \frac{2}{\pi} \arctg\left(\frac{|x|}{2}\right),$$

isto é, comparar-se-á o caso em que não há ponderação das observações com aquele em que é usada, nessa ponderação, a função de peso que permite estabelecer a velocidade de convergência. Além disso serão geradas amostras de dimensões 20 e 50.

A estimativa de $Y_t(\beta) = (X_{t-2}\varepsilon_t)(\beta)$ usada na construção da região de aceitação foi

$$\begin{cases} \hat{y}_t(\beta) = x_{t-2} \left(x_t \sum_{n=1}^{t-3} (-\beta)^n x_{t-n} \prod_{i=1}^n x_{t-i-2} \right), & t = 4, \dots, T \\ \hat{y}_3 = x_1 x_3 \end{cases},$$

tendo-se considerado na definição de $Y_t(\beta)$ as observações antes do instante 3 iguais a zero.

Avaliação do nível do teste

Para avaliar o nível de significância do teste foi gerado o modelo referido com $\varphi = 0$ (H_0 verdadeira) e testado contra 5 alternativas ($\beta = 0.01$, $\beta = 0.05$, $\beta = 0.1$, $\beta = 0.25$ e $\beta = 0.5$). Na tabela abaixo estão registados os intervalos de confiança a 95% para a proporção de rejeições de H_0 em 60 réplicas do modelo. Estes intervalos de confiança foram obtidos repetindo 30 vezes aquele procedimento.

T		20	20	50	50
β	teste	g_1	g_2	g_1	g_2
0.01]0.41, 0.46[]0.34, 0.39[]0.41, 0.44[]0.24, 0.27[
0.05]0.33, 0.38[]0.21, 0.24[]0.25, 0.30[]0.08, 0.1[
0.1]0.27, 0.31[]0.14, 0.18[]0.18, 0.22[]0.03, 0.05[
0.25]0.18, 0.21[]0.05, 0.07[]0.07, 0.09[]0.002, 0.008[
0.5]0.09, 0.12[]0.02, 0.03[]0.01, 0.02[]0.0, 0.002[

Tabela 3. Estimativa intervalar para o erro de 1^a espécie

Constata-se de novo um bom desempenho do teste. O nível decresce à medida que o parâmetro sob a alternativa cresce, sendo o decréscimo muito significativo quando as observações estão sujeitas a ponderação. A importância da função de peso torna-se assim evidente. De modo natural, diminui a proporção de rejeições de H_0 quando aumenta o número de observações.

Avaliação do erro de 2^a espécie do teste

A avaliação do erro de 2^a espécie é feita gerando vários modelos sob a alternativa (H_1 verdadeira), isto é, considerando vários valores de β ($\beta = 0.01$, $\beta = 0.05$, $\beta = 0.1$, $\beta = 0.25$ e $\beta = 0.5$). Em cada caso, regista-se a proporção de aceitações de H_0 em 60 réplicas do modelo. Os intervalos de confiança a 95% para esta proporção, obtidos replicando o procedimento 30 vezes, apresentam-se na tabela seguinte.

T		20	20	50	50
β	teste	g_1	g_2	g_1	g_2
0.01]0.44, 0.49[]0.37, 0.41[]0.39, 0.44[]0.25, 0.29[
0.05]0.34, 0.39[]0.23, 0.26[]0.25, 0.30[]0.07, 0.1[
0.1]0.24, 0.29[]0.14, 0.18[]0.15, 0.19[]0.03, 0.04[
0.25]0.12, 0.16[]0.05, 0.07[]0.02, 0.04[]0.004, 0.01[
0.5]0.05, 0.06[]0.01, 0.02[]0.002, 0.008[]0.0, 0.002[

Tabela 4. Estimativa intervalar para o erro de 2^a espécie

Observa-se que a estimativa para o erro de 2ª espécie diminui fortemente à medida que β aumenta. Além disso, a precisão da estimativa aumenta, em geral, com T . O bom comportamento é particularmente evidente quando as observações são efectivamente ponderadas. Realça-se ainda o comportamento simétrico do teste.

Como observação final salienta-se a grande potencialidade de aplicação desta metodologia de teste a modelos de séries temporais sem momentos.

6.3 Modelação de uma série real

Vamos, mais uma vez, ilustrar a aplicação dos resultados probabilistas e estatísticos associados a modelos de séries temporais, lineares e não lineares, através da procura de um modelo teórico que se ajuste bem a uma série de dados reais. O procedimento seguido baseia-se na metodologia proposta por Box e Jenkins para modelos lineares, que descrevemos no final do quarto capítulo. Neste estudo utilizamos de novo o *software* estatístico EViews que, como já foi referido, é particularmente adequado para a modelação de séries temporais.

Consideremos a série temporal relativa às cotações diárias de fecho das acções da Pararede no mercado Euronext, durante o período de 17 de Dezembro de 2003 a 28 de Maio de 2007.

A trajectória desta série, presente na figura 6.1, mostra alguma volatilidade instantânea.

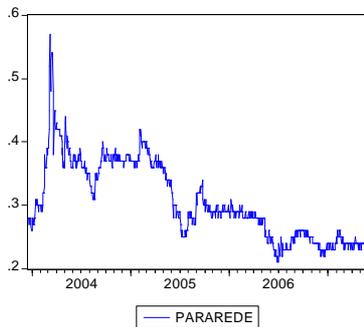


Fig. 6.1 Cotações diárias de fecho da Pararede

Façamos a sua descrição estatística (ver figura 6.2).

Esta análise empírica permite-nos concluir que estamos perante uma distribuição estatística não compatível com a lei normal, não centrada e leptocúrtica.

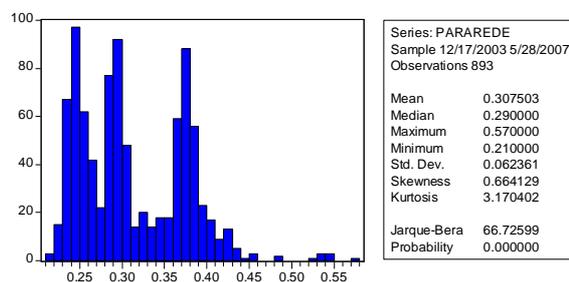


Fig. 6.2 Análise descritiva da série Pararede

Observemos o gráfico das autocorrelações e das autocorrelações parciais (figura 6.3).

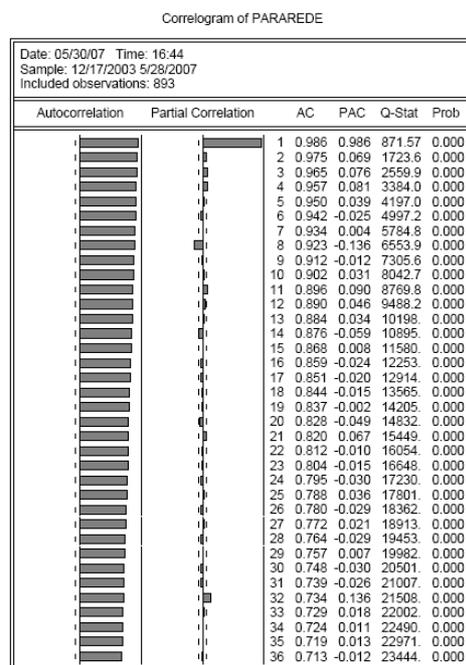


Fig. 6.3 Correlograma da série Pararede

Este gráfico indicia a necessidade de considerar a série diferenciada. Analisemos, pois, o gráfico das autocorrelações e das autocorrelações parciais da série das diferenças (figura 6.4).

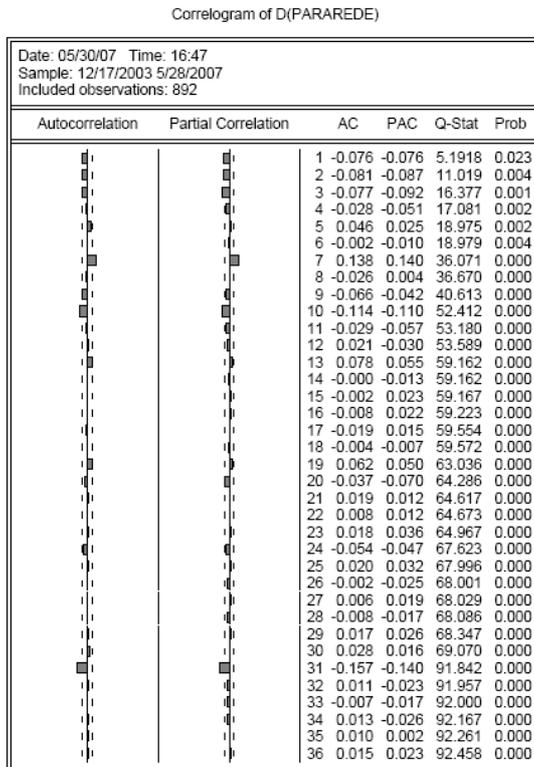


Fig. 6.4 Correlações da série diferenciada

Da figura 6.4 parece surgir a necessidade de introduzir um modelo auto-regressivo de ordem 7, AR(7). Fazendo a estimação de um tal modelo, obtêm-se os valores presentes na figura 6.5.

Note-se que só os coeficientes φ_1 (AR(1)), φ_2 (AR(2)), φ_3 (AR(3)) e φ_7 (AR(7)) são significativamente diferentes de zero (p -valor correspondente menor ou igual a 0.01).

Dependent Variable: D(PARAREDE)				
Method: Least Squares				
Date: 05/29/07 Time: 18:27				
Sample (adjusted): 12/29/2003 5/18/2007				
Included observations: 885 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.57E-05	0.000291	-0.156889	0.8754
AR(1)	-0.091492	0.033434	-2.736523	0.0063
AR(2)	-0.100872	0.033576	-3.004294	0.0027
AR(3)	-0.086975	0.033707	-2.580330	0.0100
AR(4)	-0.037007	0.033800	-1.094883	0.2739
AR(5)	0.037271	0.033693	1.106187	0.2689
AR(6)	0.002902	0.033544	0.086514	0.9311
AR(7)	0.140079	0.033381	4.196349	0.0000
R-squared	0.043848	Mean dependent var	-4.52E-05	
Adjusted R-squared	0.036216	S.D. dependent var	0.010034	
S.E. of regression	0.009850	Akaike info criterion	-6.393609	
Sum squared resid	0.085096	Schwarz criterion	-6.350350	
Log likelihood	2837.172	F-statistic	5.745450	
Durbin-Watson stat	1.999558	Prob(F-statistic)	0.000002	
Inverted AR Roots	.71	.44+ .64i	.44- .64i	-.16-.72i
	-.16+.72i	-.69+.35i	-.69-.35i	

Fig. 6.5 Modelo AR estimado

O modelo estimado é estacionário e o ruído associado é a inovação, pois as raízes do polinômio AR estimado são em módulo superiores a 1 (figura 6.6).

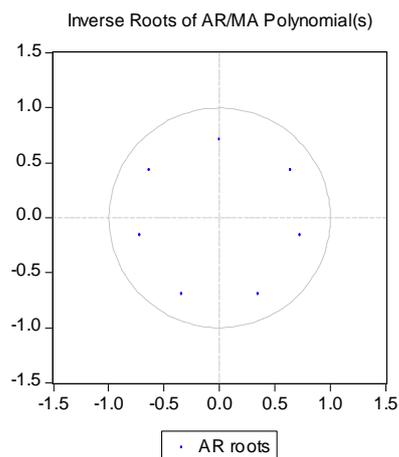


Fig. 6.6 Raízes do polinômio AR

O correlograma dos resíduos (figura 6.7) permite-nos inferir que estamos perante um processo de erro.

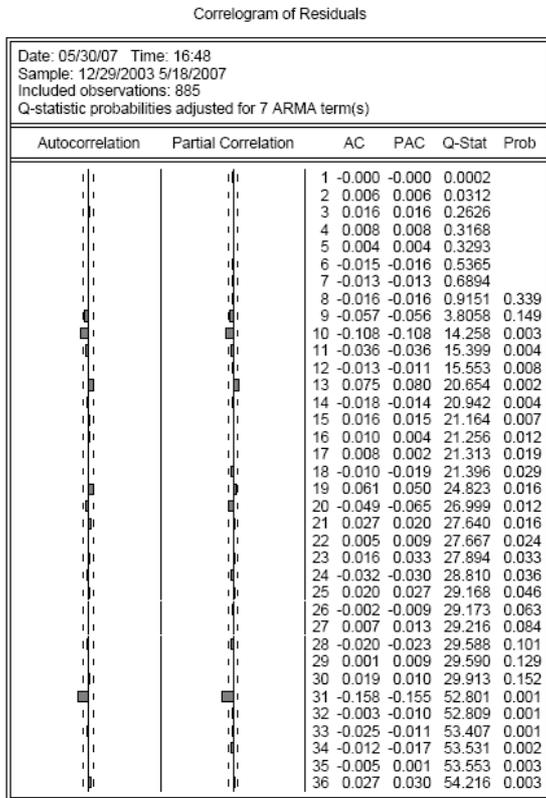


Fig. 6.7 Correlações dos resíduos após modelação AR

Testemos a heteroscedasticidade dos resíduos usando o teste ARCH LM que testa a hipótese H_0 : ausência de heteroscedasticidade.

O p -valor obtido, como pode observar-se na figura 6.8, leva à rejeição da hipótese de ausência de heteroscedasticidade, isto é, o ruído é heteroscedástico.

ARCH Test:				
F-statistic	117.1038	Prob. F(1,882)	0.000000	
Obs*R-squared	103.6126	Prob. Chi-Square(1)	0.000000	
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Date: 05/30/07 Time: 16:56				
Sample (adjusted): 12/30/2003 5/18/2007				
Included observations: 884 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.32E-05	1.43E-05	4.413309	0.0000
RESID^2(-1)	0.342366	0.031638	10.82145	0.0000
R-squared	0.117209	Mean dependent var	9.61E-05	
Adjusted R-squared	0.116208	S.D. dependent var	0.000442	
S.E. of regression	0.000416	Akaike info criterion	-12.73028	
Sum squared resid	0.000153	Schwarz criterion	-12.71945	
Log likelihood	5628.782	F-statistic	117.1038	
Durbin-Watson stat	1.988997	Prob(F-statistic)	0.000000	

Fig. 6.8 Teste à heteroscedasticidade dos resíduos

Façamos então um novo ajustamento por um modelo AR(7) com processo de erro ARCH(9).

Correlogram of Residuals Squared

Date: 05/30/07 Time: 16:54				
Sample: 12/29/2003 5/18/2007				
Included observations: 885				
Q-statistic probabilities adjusted for 7 ARMA term(s)				
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat Prob
■	■	1 0.342	0.342	104.08
■	■	2 0.103	-0.016	113.51
■	■	3 0.113	0.004	124.93
■	■	4 0.067	-0.000	128.87
■	■	5 0.042	0.017	130.46
■	■	6 0.148	0.138	149.91
■	■	7 0.135	0.042	166.18
■	■	8 0.105	0.046	176.12 0.000
■	■	9 0.293	0.260	253.10 0.000
■	■	10 0.313	0.152	341.14 0.000
■	■	11 0.049	-0.129	343.32 0.000
■	■	12 0.021	-0.009	343.72 0.000
■	■	13 0.024	-0.035	344.24 0.000
■	■	14 0.034	0.029	345.25 0.000
■	■	15 0.027	-0.054	345.92 0.000
■	■	16 0.011	-0.092	346.04 0.000
■	■	17 0.023	0.011	346.52 0.000
■	■	18 0.032	-0.037	347.47 0.000
■	■	19 -0.007	-0.140	347.51 0.000
■	■	20 -0.018	-0.022	347.81 0.000
■	■	21 -0.005	0.053	347.83 0.000
■	■	22 0.015	0.042	348.05 0.000
■	■	23 0.058	0.063	351.07 0.000
■	■	24 0.055	0.018	353.84 0.000
■	■	25 0.043	0.088	355.52 0.000
■	■	26 -0.001	0.011	355.52 0.000
■	■	27 -0.002	-0.014	355.52 0.000
■	■	28 0.013	0.054	355.68 0.000
■	■	29 -0.007	0.027	355.72 0.000
■	■	30 0.054	0.061	358.44 0.000
■	■	31 0.157	0.119	381.00 0.000
■	■	32 0.043	-0.111	382.70 0.000
■	■	33 0.020	-0.020	383.09 0.000
■	■	34 0.013	-0.047	383.24 0.000
■	■	35 0.027	-0.005	383.92 0.000
■	■	36 0.006	-0.011	383.95 0.000

Fig. 6.9 Correlações dos resíduos ao quadrado, após modelação AR

Tivemos em conta neste ajustamento o correlograma do quadra-

do do processo, conforme figura 6.9, e as propriedades dos modelos ARCH no que diz respeito à representação auto-regressiva desse processo.

As estimativas para os parâmetros do modelo AR(7)-ARCH(9) encontram-se na figura 6.10.

Dependent Variable: D(PARAREDE)				
Method: ML - ARCH				
Date: 05/29/07 Time: 18:16				
Sample (adjusted): 12/29/2003 5/18/2007				
Included observations: 885 after adjustments				
Failure to improve Likelihood after 22 iterations				
Variance backcast: ON				
GARCH = C(9) + C(10)*RESID(-1)^2 + C(11)*RESID(-2)^2 + C(12)*RESID(-3)^2 + C(13)*RESID(-4)^2 + C(14)*RESID(-5)^2 + C(15)*RESID(-6)^2 + C(16)*RESID(-7)^2 + C(17)*RESID(-8)^2 + C(18)*RESID(-9)^2				
	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.000170	0.000136	-1.254275	0.2097
AR(1)	-0.377928	0.048930	-7.723886	0.0000
AR(2)	-0.165142	0.042108	-3.921834	0.0001
AR(3)	-0.127221	0.032907	-3.866067	0.0001
AR(4)	-0.115950	0.022566	-5.138276	0.0000
AR(5)	-0.012099	0.031152	-0.388396	0.6977
AR(6)	-0.024961	0.032373	-0.771040	0.4407
AR(7)	0.040557	0.026664	1.521057	0.1282
Variance Equation				
C	3.45E-05	3.04E-06	11.35750	0.0000
RESID(-1)^2	0.317268	0.039962	7.939299	0.0000
RESID(-2)^2	0.103821	0.034189	3.036718	0.0024
RESID(-3)^2	0.006713	0.029572	0.227018	0.8204
RESID(-4)^2	-0.021309	0.014302	-1.489927	0.1362
RESID(-5)^2	0.077828	0.025643	3.035076	0.0024
RESID(-6)^2	0.002864	0.018773	0.152581	0.8787
RESID(-7)^2	-0.009015	0.019907	-0.452872	0.6506
RESID(-8)^2	0.043536	0.024156	1.802250	0.0715
RESID(-9)^2	0.017732	0.017245	1.028213	0.3038
R-squared	-0.049524	Mean dependent var	-4.52E-05	
Adjusted R-squared	-0.070103	S.D. dependent var	0.010034	
S.E. of regression	0.010380	Akaike info criterion	-6.792478	
Sum squared resid	0.093406	Schwarz criterion	-6.695145	
Log likelihood	3023.672	Durbin-Watson stat	1.441817	
Inverted AR Roots	.45	.38+ .55i	.38-.55i	-.16+ .60i
	-.16-.60i	-.64+ .33i	-.64-.33i	

Fig. 6.10 Estimação do modelo AR ARCH

Apresentemos, de modo resumido, a análise dos resíduos do novo ajustamento.

1/ O correlograma dos resíduos (figura 6.11) é compatível com o de um ruído branco.

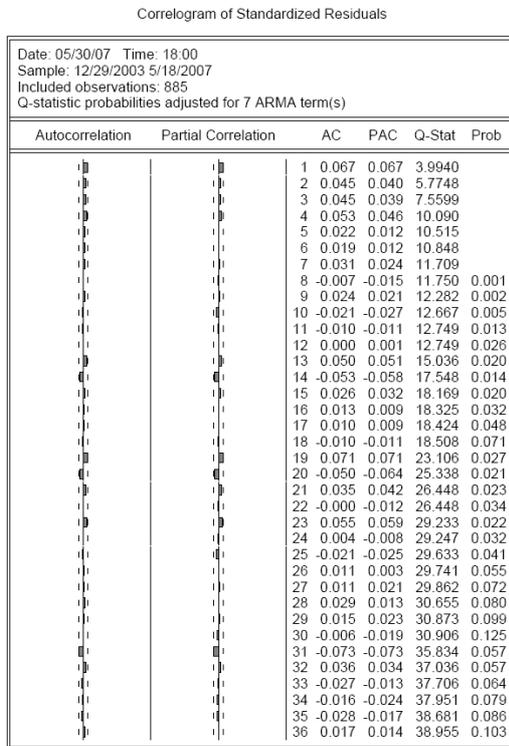


Fig. 6.11 Correlações dos resíduos após modelo AR ARCH

A análise descritiva do resíduo está presente na figura 6.12. É clara a não compatibilidade com uma lei Gaussiana.

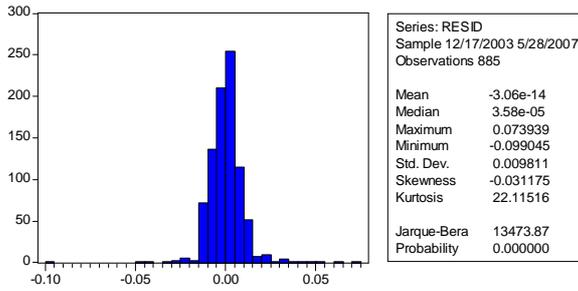


Fig. 6.12 Análise descritiva do resíduo após ajustamento AR-ARCH

2/ O teste ARCH LM à heteroscedasticidade dos resíduos está resumido na figura 6.13.

Com um p -valor desta ordem de grandeza somos obviamente conduzidos à aceitação da hipótese de ausência de heteroscedasticidade, isto é, o novo resíduo é homoscedástico.

ARCH Test:				
F-statistic	0.262897	Prob. F(1,882)	0.608265	
Obs*R-squared	0.263415	Prob. Chi-Square(1)	0.607784	
Test Equation:				
Dependent Variable: WGT_RESID^2				
Method: Least Squares				
Date: 05/30/07 Time: 18:03				
Sample (adjusted): 12/30/2003 5/18/2007				
Included observations: 884 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.992305	0.078992	12.56213	0.0000
WGT_RESID^2(-1)	0.017264	0.033671	0.512735	0.6083
R-squared	0.000298	Mean dependent var	1.009761	
Adjusted R-squared	-0.000835	S.D. dependent var	2.118407	
S.E. of regression	2.119292	Akaike info criterion	4.342301	
Sum squared resid	3961.412	Schwarz criterion	4.353125	
Log likelihood	-1917.297	F-statistic	0.262897	
Durbin-Watson stat	1.999089	Prob(F-statistic)	0.608265	

Fig. 6.13 Teste à heteroscedasticidade do resíduo após modelação AR ARCH

O modelo retido é, então, sendo X_t a cotação de fecho da Pararede no dia t ,

$$Y_t = X_t - X_{t-1}$$

com

$$\begin{aligned} Y_t &= \varphi_0 + \varphi_1 Y_{t-1} + \dots + \varphi_7 Y_{t-7} + \varepsilon_t \\ \varepsilon_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_9 \varepsilon_{t-9}^2 + u_t \end{aligned}$$

com as estimativas para os coeficientes $\varphi_0, \varphi_1, \dots, \varphi_7, \alpha_0, \dots, \alpha_9$ presentes na figura 6.14 e $u = (u_t, t \in \mathbb{Z})$ um ruído branco.

```
Substituted Coefficients:
=====
D(PARAREDE) = -0.0001702503234 + [AR(1)=-0.3779282994,AR(2)=-0.1651419395,AR(3)=-0.1272214459,AR(4)=-
0.1159504824,AR(5)=-0.01209931804,AR(6)=-0.02496107739,AR(7)=0.04055675344]
GARCH = 3.452306302e-005 + 0.3172681558*RESID(-1)^2 + 0.1038212183*RESID(-2)^2 + 0.00671329506*RESID(
-3)^2 - 0.02130917541*RESID(-4)^2 + 0.07782787867*RESID(-5)^2 + 0.00286440962*RESID(-6)^2 - 0.0090151858
69*RESID(-7)^2 + 0.04353596263*RESID(-8)^2 + 0.01773161464*RESID(-9)^2
```

Fig. 6.14 Modelo estimado

O software utilizado permite ainda comparar, visualmente, as trajetórias da série diferenciada e da correspondente série estimada bem como a trajetória do resíduo associado, como se ilustra na figura 6.15.

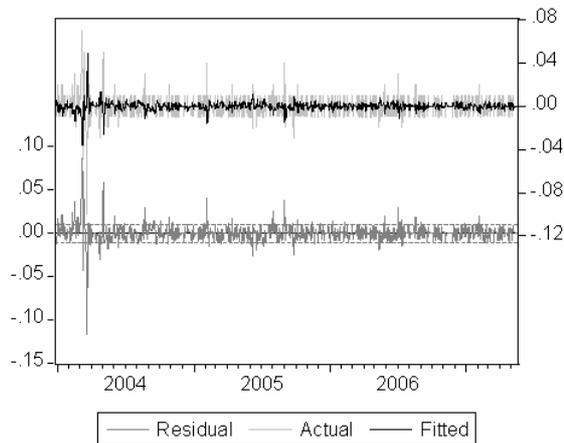


Fig. 6.15 Gráficos das séries Pararede diferenciada, ajustada e residual

Façamos a previsão dos valores da série Pararede no período de 21 a 28 de Maio de 2007 (figura 6.16) e do desvio padrão condicional (figura 6.17).

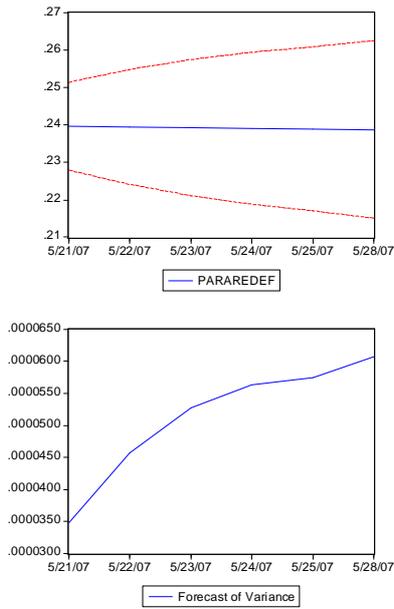


Fig. 6.16 Previsão da série Pararede e da variância

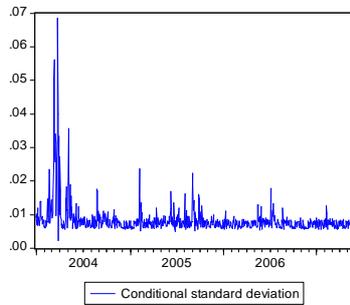


Fig. 6.17 Desvio padrão condicional

Como curiosidade refere-se que os valores das cotações de fecho das acções da Pararede observados durante aquele período foram os apresentados no quadro 6.18.

21/5/2007	22/5/2007	23/5/2007	24/5/2007	25/5/2007	28/5/2007
0.23	0.24	0.24	0.24	0.23	0.24

Quadro 6.18 Valores da cotação de fecho da Pararede

Bibliografia

Anderson, T.W. (1971) *The statistical analysis of time series*, Wiley, New York.

Ash, R. B. (1972) *Real Analysis and Probability*, Academic Press, New York.

Ash, R.B., M.F. Gardner (1975) *Topics in Stochastic Processes*, Academic Press, New York.

Azencott, R., D. Dacunha-Castelle (1984) *Séries d'observations irrégulières, modélisation et prévision*, Masson, Paris.

Bachelier, T. (1900) *Théorie de la speculation*, Gauthier-Villars, Paris, Reprinted in D. Cootner ed.(1964) - The random character of stock market prices, *MIT Press, Cambridge*, 17-78.

Beguín J.M., Ch. Gouriéroux, A. Monfort (1980) Identification of an ARMA process: the corner method, *in Time series*, Editor T. Anderson, North-Holland.

Billingsley, P. (1968) *Convergence of Probability Measures*, Wiley.

Bollerslev, T.(1986) Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 37, 307-327.

Bollerslev, T.(1987) A conditionally heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics*, 69, 542-547.

Bollerslev, T. (1988) On the correlation structure for the generalized autoregressive conditional heteroscedastic process, *Journal of time series analysis*, 9, 121-131.

Box, G.E.P., G.M. Jenkins (1976) *Time series analysis, forecasting and control*, 2^a edição, Holden-Day, San Francisco.

Brockett, R.W. (1976) Volterra series and geometric control

theory, *Automatica*, 12, 167-176.

Brockwell, P.J., R.A. Davies (1991) *Time series: theory and methods*, 2^a edição, Springer-Verlag.

Dacunha-Castelle, D., M. Duflo (1983), *Probabilités et statistiques*, Vol. 1 - *Problèmes à temps fixe*, 2 - *Problèmes à temps mobile*, Masson.

Davis, P., J. (1965), The gamma function and related functions, in M. Abramowitz and N. Stegun eds., *Handbook of mathematical functions*, Dover Publications Inc., New York.

Droesbeke, J.J., B. Fichet, P. Tassi (eds) (1989) *Séries chronologiques: théorie et pratique des modèles ARIMA*, Economica, Paris.

Elie, L. (1992) Les processus ARCH comme approximations de processus en temps continue, *Modèles ARCH et Applications à la Finance*, Journées d'Étude en Statistique, Pub. de l'A.S.U., 193-223.

Engle, R. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of the U.K. inflation, *Econometrica*, 50, 987-1008.

Engle, R. (1983) Estimates of the variance of the U.S. inflation based on the ARCH model, *Journal of Money Credit and Banking*, 15, 286-301.

Engle, R., T. Bollerslev (1986) Modelling the persistence of conditional variances, *Econometric Reviews*, 5, 1-50.

Engle, R., D. Lilien, R. Robins (1987) Estimation of time varying risk premiums in the term structure, *Econometrica*, 55, 391-408.

Fama, E.F. (1965) The behaviour of stock market prices, *Journal of Business*, 38, 34-105.

Foata, D., Fuchs, A. (1998) *Calcul des probabilités*, 2e édition, Dunod.

Gabr, M.M. (1988) On the third order moment structure and bispectral analysis of some bilinear time series, *Journal of Time Series Analysis*, 9, 1, 11-20.

Gallant, A.R. (1987) *Non linear statistical models*, Wiley, New York.

Geffroy, J. (1980) Asymptotic separation of distributions and convergence properties of tests and estimators, in: *Asymptotic*

theory of statistical tests and estimation, I.M. Chakravarti, ed., Academic Press, 159-177.

Glasbey, C.A. (1982) A generalization of partial autocorrelation useful in identification of ARMA models, *Technometrics*, 24, 223-228.

Gonçalves, E., N. Mendes-Lopes (1994) The Generalized Threshold ARCH model: wide sense stationarity and asymptotic normality of the temporal aggregate, *Pub.Inst.Stat.Univ.Paris*, XXXVIII, fasc.2, 19-35.

Gonçalves, E., N. Mendes-Lopes (1996) Stationarity of GTARCH processes, *Statistics*, 28, 171-178.

Gonçalves, E., N. Mendes-Lopes (1996a), Séries temporais com erros condicionalmente heteroscedásticos, *Bom senso e sensibilidade: Traves mestras da Estatística*, Actas do III Congresso Anual da SPE, Ed. Salamandra, 517-526.

Gonçalves, E., N. Mendes-Lopes (1998) Some statistical results on autoregressive conditionally heteroscedastic models, *Journal of Statistical Planning and Inference*, 68, 193-202.

Gonçalves, E., N. Mendes-Lopes (2000) *Probabilidades - Princípios Teóricos*, Escolar Editora, Lisboa.

Gonçalves, E., P. Jacob, N. Mendes-Lopes (2000) A decision procedure for bilinear time series based on the asymptotic separation, *Statistics*, 33, 333-348.

Gonçalves, E., C. Martins e N. Mendes Lopes (2001), Asymptotic separation-based tests for noise processes against first order diagonal dependence, *Annales ISUP*, vol. 45, fasc. 2-3, p. 9-27.

Gonçalves, E., C. Martins, N. Mendes-Lopes (2002), Estudo comparativo de ensaios estatísticos entre processos de erro e modelos bilineares ergódicos, *Actas do IX Congresso Anual da SPE, Novos Rumos em Estatística*, p. 259-271.

Gouriéroux, Ch. (1987) Une approche géométrique des processus ARMA, *Annales d'Economie et de Statistique*, 8, 135-159.

Gouriéroux, Ch. (1992) Les modèles heteroscedastiques univariées, *Modèles ARCH et Applications à la Finance*, Journées d'Étude en Statistique, Pub. de l'A.S.U., 57-76.

Gouriéroux, Ch. (1992a) *Modèles ARCH et Applications Financières*, Economica.

Gouriéroux, Ch., A. Monfort (1989) *Statistique et modèles économétriques*, vols. I, II, Economica, Paris.

Gouriéroux, Ch., A. Monfort (1990) *Séries temporelles et modèles dynamiques*, Economica, Paris.

Gouriéroux, Ch., A. Monfort (1991) Modèles hétéroscedastiques et applications à la Finance, *Document de travail*, CREST-ENSAE, Paris.

Granger, C.W.J., A. Andersen (1978) *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.

Granger, W.J., D. Orr (1972) Infinite variance and research strategy in time series analysis, *JASA*, 64, 275-285.

Guégan, D. (1994) *Séries chronologiques non linéaires à temps discret*, Economica, Paris.

Hannan, E.J. (1970) *Multiple time series*, Wiley.

Higgins, M.L., A.K. Bera (1990) A class of nonlinear ARCH models, *Int. Econ. Rev.*, 33, 137-158.

Hsieg, D.A. (1989) Modelling heteroskedasticity in daily foreign exchange rate changes, *Journal of Business*, 62, 339-368.

Kendall, G. (1973) *Time Series*, Griffin.

Koopmans, L. (1974) *The spectral analysis of time series*, Academic Press.

Li, W.K. (1984) On the autocorrelation structure and identification of some bilinear time series, *Journal of Time Series Analysis*, 5, 173-181.

Loève, M. (1977) *Probability theory I*, 4th edition, Springer-Verlag.

Mandelbrot, B. (1963) The variation of certain speculative prices, *Journal of Business*, 36, 394-419.

Martins, C.M. (1997) On the autocorrelations related to a bilinear model with non-independent shocks, *Stat. & Prob. Letters*, 36, 245-250.

Martins, C.M. (1999) A note on the third order moments structure of a bilinear model with non-independent shocks, *Portugaliae Mathematica*, Vol. 56, Fasc. 1, 115-125.

Martins, C.M. (1999a) Higher order moments of bilinear sub-diagonal models with non-independent shocks, *Annales de l'I.S.U.P.*, 43, 1, 29-42.

Martins, C.M. (2000) *Modelos Bilineares em Séries Temporais: Propriedades Probabilísticas e Decisão Estatística*, Tese de Doutorado, Departamento de Matemática, Univ. Coimbra.

Métivier, M (1972) *Notions fondamentales de théorie des probabilités*, Dunod.

Milhoj, A.(1985) The moment structure of ARCH processes, *Scand. Journal of Statistics*, 12, 281-292.

Mohler, R.R.(1973) *Bilinear control processes*, Academic Press, New York.

Murteira, B.J.F., D.A. Muller, K.F. Turkman (1993) *Análise de séries cronológicas*, McGraw-Hill de Portugal.

Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model, *Econometric Theory*, 6, 318-334.

Nelson, D.B. (1990a) ARCH models as diffusion approximations, *Journal of Econometrics*, 45, 7-38.

Neveu, J. (1970) *Bases mathématiques du calcul des probabilités*, 2e. ed., Masson.

Pham Dinh, T. (1984) A note on some statistics useful in identifying the order of autoregressive moving average model, *Journal of Time Series Analysis*, 5, 273-279.

Quinn, B.G. (1982) Stationarity and invertibility of simple bilinear models, *Stoch. Processes and Appl.*, 12, 225-230.

Rabemananjara, R., Zakoian, J.M. (1993) Threshold ARCH models and asymmetries in volatility, *Journal of Applied Econometrics*, 8, 31-49.

Rosenblatt, M. (1978) Dependence and asymptotic independence for random processes, *Studies in Probability Theory, Studies in Mathematics*, 18, 24-44, The Mathematical Association of America.

Rudin, W. (1987) *Real and complex analysis*, 3 ed., McGraw-Hill Book Company.

Samuelson, P.(1973) Proof that properly discounted present values of assets vibrate randomly, *Bell Journal*, 4.

Tong, H. (1990) *Non-linear time series: a dynamical system approach*, Oxford University Press, Oxford.

Wandji, J.N. (1998) A nonparametric goodness-of-fit test for a class of parametric autoregressive models, *J. Statist. Plann. Inf.*,

71, 57-74.

Weiss, A.A. (1984) ARMA models with ARCH errors, *Journal of Time Series Analysis*, 5, 129-143.

Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing, *Econometric Theory*, 2, 107-131.

Zakoian, J.M. (1990) Modèles autoregressifs à seuil, *Thèse de Docteur de l'Université en Math. Appliquées*, Univ. Paris IX.

Zakoian, J.M. (1994) Threshold heteroskedastic models, *Journal of Economic Dynamics and Control*, 931-955.

Yule, G.U. (1927) On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, *Philos. Trans. Royal Society*, London A 226, 267-298.

Esmeralda Gonçalves e **Nazaré Mendes Lopes** são Professoras no Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra, onde têm desenvolvido as suas carreiras científica e pedagógica no domínio das Probabilidades e Estatística, área de especialidade em que realizaram os respectivos doutoramentos. Os seus trabalhos de investigação têm sido dedicados à Análise Estocástica de Séries Temporais e à Estimação Funcional, com resultados publicados em revistas nacionais e internacionais da especialidade. São ainda autoras e co-autoras de monografias em temas gerais do foro das Probabilidades, Estatística e Processos Estocásticos.

ISBN: 978-972-8890-16-2
Impressão e Distribuição:
Instituto Nacional de Estatística

