

Rodrigo De Losso da Silveira Bueno

ECONOMETRIA DE SÉRIES TEMPORAIS

Econometria de séries temporais

Rodrigo De Losso da Silveira Bueno
Fundação Getúlio Vargas - CFC

CENGAGE
Learning

Austrália • Brasil • Japão • Coréia • México • Cingapura • Espanha • Reino Unido • Estados Unidos

Econometria de Séries Temporais**Rodrigo De Losso da Silveira Bueno**

Gerente Editorial: Patricia La Rosa

Editora de Desenvolvimento: Lígia Cantarelli

Supervisor de Produção Editorial: Fabiana
AlbuquerqueProdutora Editorial: Gisele Gonçalves Bueno
Quirino de Souza

Copidesque: Norma Gusukuma

Revisão de provas: Adriane Peçanha
Cristiane Mayumi Morinaga

Composição: Roberto Maluhy Jr & Mika Mitsui

Capa: Eduardo Bertolini

© 2008 Cengage Learning Edições Ltda.

Todos os direitos reservados. Nenhuma parte deste livro
poderá ser reproduzida, sejam quais forem os meios
empregados, sem a permissão, por escrito, da Editora.
Aos infratores aplicam-se as sanções previstas nos artigos
102, 104, 106 e 107 da Lei nº 9.610, de 19 de fevereiro de 19

Para informações sobre nossos produtos, entre em
contato pelo telefone **0800 11 19 39**

Para permissão de uso de material desta obra, envie
seu pedido para **direitosautorais@cengage.com**

© 2008 Cengage Learning. Todos os direitos reservados.

ISBN13: 978-85-221-0642-4

ISBN10: 85-221-0642-8

Cengage Learning

Condomínio E-Business Park

Rua Werner Siemens, 111 – Prédio 20 – Espaço 3

Lapa de Baixo – 05069-900 – São Paulo – SP

Tel.: (11) 3665-9900 – Fax: (11) 3665-9901

sac@thomsonlearning.com.br

www.thomsonlearning.com.br

Para suas soluções de curso e aprendizado, visite
www.cengage.com.br

Impresso no Brasil.

Printed in Brazil.

1 2 3 4 5 6 7 12 11 10 09 08

Dedico este livro a meu filho,
Luís Felipe, que nunca me
permitiu desistir.

Apresentação

A econometria das séries temporais está ligada a um grande número de problemas econômicos e financeiros. Aqueles que não acessam esse instrumental têm sérias dificuldades para realizar análise econométrica em uma variedade de campos da economia. Além disso, não conseguem acompanhar grande parte da literatura aplicada, em que as ferramentas de séries de tempo são utilizadas. Quando as pessoas não possuem o conhecimento técnico, fica-lhes difícil avaliar a verdadeira contribuição da análise empírica para conhecer a realidade.

O livro *Econometria de séries temporais* do professor Rodrigo De Losso da Silveira Bueno veio preencher uma lacuna importante nas publicações sobre econometria de séries temporais no Brasil. Faltava uma publicação em português, na qual o assunto fosse tratado com rigor, mas sem perda do senso prático. Nesta obra, o rigor e a simplicidade procuram caminhar juntos.

Após a apresentação das definições básicas, uma discussão do conceito de estacionariedade é levada a efeito. Os modelos ARMA são apresentados, e métodos de estimação são discutidos. A atualidade do livro pode ser vista no capítulo destinado ao método de estimação conhecido como método generalizado dos momentos, GMM. A relação entre variáveis estacionárias é estabelecida, até mesmo sob expectativas racionais. Em seguida, discutem-se suas propriedades e as circunstâncias para seu uso são elaboradas.

Já dentro da perspectiva de variáveis não estacionárias, inicialmente o modelo VAR é introduzido, com uma discussão sobre as características do modelo como proposto originalmente por Chistopher Sims. A versão do modelo com variáveis estacionárias completa a discussão do VAR.

O conceito de cointegração é apresentado com propriedade, de forma que aqueles que seguiram o livro e aprenderam não terão problemas em entender esse importante conceito da econometria de séries de tempo. O mesmo ocorre para os modelos GARCH.

Os apêndices apresentam o instrumental matemático necessário para a compreensão dos conceitos discutidos.

O título contém exemplos e exercícios que contribuirão para consolidar o aprendizado.

É uma leitura que recomendo tanto para alunos de graduação quanto para alunos de pós-graduação e profissionais interessados em macroeconomia e finanças. Obviamente os professores também terão uma boa oportunidade para se reciclar, já que o livro é moderno no seu conteúdo, assim como na sua apresentação.

Denisard C. O. Alves

Professor Titular do Departamento de Economia da Faculdade de Economia,
Administração e Contabilidade da Universidade de São Paulo



Sumário

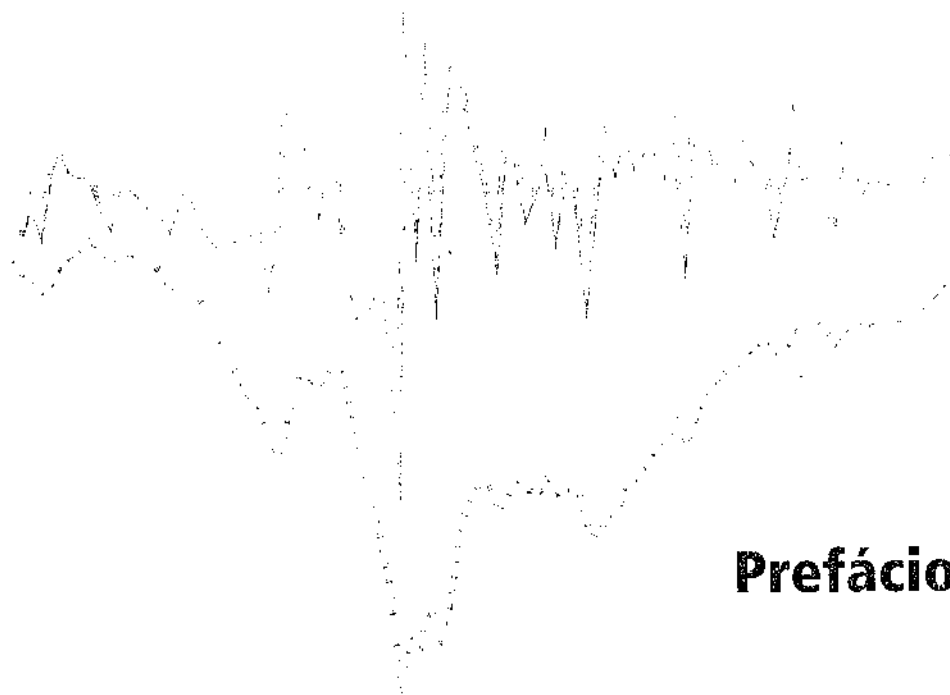
Apresentação	vii
Prefácio	xv
1 Introdução	1
2 Fundamentos Estatísticos	5
2.1 Esperanças Condicional e Incondicional	5
2.2 Processos Auto-Regressivos	10
2.3 Processos Estocásticos	11
2.4 Autocovariância e Autocorrelação	14
2.5 Estacionaridade	15
2.6 Ergodicidade	16
2.7 Ruído Branco	19

2.8 Médias Móveis	20
2.8.1 Médias Móveis de Ordem 1 – MA (1), 20	
2.8.2 Médias Móveis de Ordem q – MA (q), 22	
2.9 Processos Auto-Regressivos	24
2.9.1 Processo Auto-Regressivo de Ordem 1 – AR (1), 24	
2.9.2 Processo Auto-Regressivo de Ordem 2 – AR (2), 27	
2.9.3 Processo Auto-Regressivo de Ordem p – AR (p), 29	
2.10 Processo Auto-Regressivo de Médias Móveis – ARMA (p,q)	30
2.11 Função Geradora de Autocovariâncias	32
2.12 Filtros	33
2.13 Invertibilidade	34
 3 Processos Estacionários	37
3.1 Função de Autocorrelação – FAC	40
3.2 Função de Autocorrelação Parcial – FACP	42
3.3 Identificação	43
3.3.1 FAC, FACP e Ljung-Box, 43	
3.3.2 Critério de Informação, 46	
3.3.3 Identificação de Modelos AR, MA e ARMA, 48	
3.4 Estimação Condicional	52
3.4.1 Função de Verossimilhança para um AR (p), 53	
3.4.2 Função de Verossimilhança para um MA (q), 54	
3.4.3 Função de Verossimilhança para um ARMA (p,q), 56	
3.5 Estimação Exata*	58
3.5.1 Função de Verossimilhança para um AR (p), 58	
3.5.2 Função de Verossimilhança para um MA (q), 59	
3.5.3 Função de Verossimilhança para um ARMA (p,q), 68	
3.6 Diagnóstico de Resíduos	68
3.6.1 Teste de Normalidade, 70	
3.6.2 Teste Jarque-Bera, 71	
3.6.3 Teste LM, 72	
3.6.4 Teste ARCH-LM, 73	
3.6.5 Teste Reset, 74	
3.7 Exemplos Simulados	76
3.8 Previsão	80
3.9 Sazonalidade – ARMA (p,q)(P,Q)_s	83

4 Processos Não Estacionários	89
4.1 Tendência Estacionária e Estocástica	90
4.2 Passeios Aleatórios	92
4.3 Removendo a Tendência	94
4.4 Regressão Espúria	95
4.5 Testes de Raiz Unitária	97
4.5.1 <i>Dickey-Fuller</i> , 97	
4.5.2 <i>Dickey-Fuller Aumentado</i> , 100	
4.5.3 <i>Demais Testes de Dickey e Fuller</i> , 105	
4.5.4 <i>Phillips-Perron</i> , 105	
4.5.5 <i>KPSS</i> , 111	
4.5.6 <i>ERS</i> , 113	
4.5.7 <i>NG e Perron</i> , 118	
4.5.8 <i>CrITÉrio de Informação e Janela Ótima</i> , 120	
4.5.9 <i>Raízes Unitárias Sazonais</i> , 122	
4.6 Decomposição de Beveridge-Nelson	124
5 GMM	133
5.1 Introdução	133
5.2 Especificação	139
5.3 Estimação	140
5.4 Propriedades do Estimador	142
5.5 Estimando a Autocovariância	144
5.6 Casos Especiais do GMM	149
5.7 Testes Usando GMM	153
5.8 Apreçamento de Ativos	155
6 Vetor Auto-Regressivo – VAR	161
6.1 Especificação de Modelo	167
6.2 Testando Hipóteses	170
6.3 Inferência	172
6.4 Verificação	174
6.4.1 <i>Teste de Ljung-Box</i> , 174	
6.4.2 <i>Teste de Breusch-Godfrey</i> , 176	
6.4.3 <i>Teste de Normalidade</i> , 177	

6.5	Previsão	179
6.6	Função Resposta ao Impulso	181
6.6.1	<i>Intervalo de Confiança</i> , 185	
6.7	Decomposição da Variância	186
6.8	Teste de Granger – Causalidade	189
6.9	VAR Estrutural	192
6.10	Decomposição de Blanchard e Quah	196
6.11	Estimação do Modelo Estrutural*	200
6.11.1	<i>Teste LR para Sobreidentificação</i> , 202	
7	Vetor de Correção de Erros – VECM	203
7.1	Teste de Cointegração de Engle-Granger	208
7.1.1	<i>Teste de Engle-Granger com Variáveis I(2)</i> , 211	
7.2	Modelo de Correção de Erros	212
7.3	Teste de Cointegração de Johansen	216
7.3.1	<i>Teste de Hipóteses</i> , 226	
7.3.2	<i>Estimação de Φ por Máxima Verossimilhança*</i> , 227	
8	Heterocedasticidade Condicional	233
8.1	Modelos GARCH	234
8.1.1	<i>ARCH (q)</i> , 234	
8.1.2	<i>GARCH (p,q)</i> , 235	
8.1.3	<i>GARCH-M</i> , 236	
8.1.4	<i>TGARCH (p,q)</i> , 237	
8.1.5	<i>PGARCH</i> , 237	
8.1.6	<i>EGARCH (p,q)</i> , 238	
8.1.7	<i>EWMA</i> , 239	
8.2	Testes para Detecção de Modelos GARCH	239
8.2.1	<i>Teste FAC e FACP</i> , 240	
8.2.2	<i>Teste de Ljung-Box</i> , 241	
8.2.3	<i>Teste Multiplicador de Lagrange ou de Engle</i> , 241	
8.3	Identificação de Modelos GARCH	242
8.4	Estimação de Modelos GARCH	242
8.5	Inferência em Modelos Univariados	244
8.6	Previsão de Modelos GARCH	245

8.7	GARCH Multivariado	248
8.7.1	<i>VECH</i> , 249	
8.7.2	<i>Restrições e Inferência</i> , 251	
8.7.3	<i>BEKK</i> , 252	
8.7.4	<i>CCC e DCC</i> , 253	
8.7.5	<i>Alisamento Exponencial e EWMA</i> , 256	
8.8	Teste de Resíduos ARCH-LM	259
Exercícios		261
	Revisão de Econometria	261
	Fundamentos Estatísticos	263
	Processos Estacionários	265
	Processos Não Estacionários	266
	GMM	266
	Vetor Auto-Regressivo – VAR	268
	Vetor de Correção de Erros – VECM	270
	Heterocedasticidade Condicional	271
Apêndices		273
	Equações a Diferenças Estocásticas	273
	Substituição Recursiva	275
	Solução para Equações a Diferenças de Ordem 2	278
	Símbolos Usados	291
Referências		293



Prefácio

A econometria é usada para explicar fatos passados, testar teorias e prever resultados de políticas ou eventos futuros. Esses objetivos se misturam no livro, mas enfatiza-se a necessidade de os modelos serem sempre economicamente fundamentados, a fim de evitar endogeneidade, regressão espúria, omissão de variáveis, entre outros problemas. O fundamento econômico também serve para definir restrições sobre as equações do modelo e, ainda, para definir o papel do erro na estimação, haja vista que a informação do econometrista nem sempre coincide com a dos agentes econômicos.

Não obstante, é comum ignorar esses fatos, dando ensejo a críticas sobre os resultados e ao papel das expectativas racionais nas dinâmicas observadas. O ponto central deste livro – o estudo de técnicas de séries temporais para estimação de modelos, inferência e avaliação de políticas – não prescinde desse pano de fundo. A acuidade do modelo econométrico, contudo, nem sempre é perfeita, seja pela limitação dos dados, do modelo ou do próprio econometrista. Por exemplo, podem-se prever os resultados da adoção de uma determinada política econômica, desde que os parâmetros do modelo estrutural possam ser identificados. Essa talvez seja a melhor maneira de se utilizar a econometria, pois os resultados serão mais precisos se as hipóteses do modelo forem verdadeiras e se for possível, assim, identificar os parâmetros do modelo. Todavia, na prática,

sempre existe um grau de arbitrariedade quanto às hipóteses utilizadas, e elas podem não ser verdadeiras, de modo que as previsões econométricas têm de ser observadas com cuidado.

A modelagem estatística será tanto melhor quanto mais inspirada na teoria econômica. Os resultados econométricos não serão válidos se não estiverem solidamente fundamentados na estatística. A interpretação e análise dos resultados, por sua vez, dependem muito menos de estatística do que de economia. Entretanto, é mais comum do que se pensa esquecer a economia em um trabalho empírico. Não fosse assim, dificilmente haveria o que se chama “regressão espúria”.

Do ponto de vista acadêmico e mesmo prático, a simples emergência de relações estatísticas tem importância limitada. Sempre se podem encontrar correlações sem significado econômico ou fazer previsões a partir de relações estatísticas ocorridas por acaso.

A econometria de séries temporais é particularmente interessante para as previsões futuras porque é formulada de forma dinâmica, geralmente em equações a diferenças. Assim, conhecendo o passado (ou explicando o passado), vislumbra-se o futuro pela formulação recursiva do problema. A questão fundamental é como tratar o termo aleatório inerente a qualquer formulação econométrica. Em particular, a inferência estatística depende das propriedades desse termo aleatório. Porém, os resultados de uma política econômica requerem a identificação de um modelo estrutural previamente formulado.

O objetivo desta obra é prover economistas, administradores, engenheiros e profissionais de mercado com um tratamento de séries temporais pela ótica da economia. Muitos exemplos usados são de economia e finanças. Tentou-se privilegiar a intuição sem prejuízo fundamental à formalidade.

Enders (2004) e Hamilton (1994) são as referências básicas sobre as quais este texto foi construído, e o leitor reconhecerá muito desses títulos aqui, seja pela notação, seja mesmo pela ordenação de assuntos em cada capítulo. Não pretendo “inventar a roda”, por isso sigo o que há de melhor neles e em outras referências. Entretanto, embora sejam excelentes referências bibliográficas (a primeira é mais básica e a segunda, mais formal), os alunos têm tido dificuldade para adaptar-se a qualquer um deles. O problema não era só o idioma, mas também a organização dos capítulos e a inexistência de exemplos mais próximos da realidade local. Assim, neste livro, uso exemplos com dados brasileiros para ilustrar os conceitos. Os dados dos exemplos, das simulações e para os exercícios estarão disponíveis no site www.cengage.com.br.

Didaticamente, procuro destacar os pontos importantes ou contra-intuitivos ao longo do texto. Em geral, resumo em passos os procedimentos estudados para estimar ou testar um modelo. Esses resumos tendem a ser aplicados nos exemplos

que seguem e procuram indicar mais diretamente o encadeamento lógico do que foi estudado até então naquela seção específica.

Refiro-me constantemente a um *software* para ilustrar modelos, conceitos e simulações. É o Eviews, mas o leitor pode utilizar qualquer outro com funções similares, como os muitos softwares livres na internet: R, Ox, JMulti, EasyReg, Gretl. Não creio ser possível estudar este livro sem o auxílio de algum pacote econométrico, em que o leitor coloca a “mão na massa”, estimando ou simulando modelos. É por esse motivo que julguei desnecessário agregar ao texto as tabelas com os testes estatísticos usados. Em geral, os pacotes econométricos desenvolveram ferramentas específicas para os testes, e essas ferramentas contêm as tabelas estatísticas.

Esta obra pressupõe um leitor com conhecimentos de equações a diferenças e de trigonometria. Entretanto, o apêndice contém tais conceitos fundamentais que serão necessários ao estudo. Além disso, requerem-se conhecimentos de econometria básica no nível do livro *Introductory econometrics*, de Wooldridge (2006), e de cálculo no nível do primeiro ano de graduação. Algum conhecimento de álgebra matricial também é requerido, particularmente na parte multivariada. Há seções técnicas, indicadas por um asterisco (*), que não precisam ser lidas inicialmente. São seções, em geral, que discutem questões de estimação. Não pretendem ser profundas, mas apenas mostrar nuances da estimação dos modelos discutidos.

É indicado para alunos avançados em cursos de graduação em economia, administração, contabilidade e engenharia e alunos de pós-graduação. O texto tem sido usado com muito sucesso entre meus alunos dos mestrados acadêmico e profissional na pós-graduação na EESP-FGV e cursos de especialização como o MBA de Pricing e Risco da BM&F. Isso é resultado de um esforço grande, porém ainda inesgotado, de escrever em uma linguagem mais simples um assunto verdadeiramente difícil, como a econometria de séries temporais. Isso o torna apropriado a várias audiências, até mesmo profissionais de mercado, que queiram alguma referência em português sobre o tema. Por outro lado, a estratégia aumenta o risco de cometer equívocos formais em busca da simplicidade.

Este livro é produto das aulas que tive com a professora Vera L. Fava, da USP, e o professor Pierre Perron, da Boston University. A eles, agradeço profundamente por terem forjado em mim a curiosidade por aprender e o gosto por séries temporais.

Não tenho com quem compartilhar a responsabilidade pelos erros, porém tenho de agradecer a meus alunos que apontaram vários equívocos enquanto escrevia este livro. Agradeço a meus assistentes, particularmente Wagner Oliveira Monteiro, que leram várias vezes os rascunhos que antecederam esta versão final,

compuseram exemplos, sugeriram alterações e simplificaram meu trabalho. Agradeço especialmente ao Ricardo Suganuma, que fez comentários pertinentes sobre esta obra, lendo-a detalhadamente, e a enriqueceu sobremaneira. Finalmente, convém mencionar e agradecer aos órgãos que financiaram parcialmente este título, como o GVPesquisa e, sobretudo, a Escola de Economia de São Paulo – Fundação Getúlio Vargas.

Consciente da possibilidade de enganos e omissões nesta obra, humildemente peço a indulgência do leitor e solicito-lhe que envie suas críticas, comentários e sugestões ao e-mail rodrigo.bueno@fgv.br.

São Paulo, janeiro de 2008.

O autor



As séries temporais podem ser estacionárias ou não estacionárias. Além disso, podem ser estocásticas ou determinísticas. A série temporal estacionária determinística mais simples é uma constante, μ , isto é:

$$y_t = \mu.$$

Essa série será estacionária e estocástica se a ela for acrescentado um componente aleatório idêntica e independentemente extraído de uma distribuição normal, $\epsilon_t \sim N(0, \sigma^2)$:

$$y_t = \mu + \epsilon_t.$$

Há inúmeras séries estacionárias determinísticas. Por exemplo:

$$y_t = \mu + 0,5y_{t-1},$$

dado y_0 . Essa série converge para 2μ , à medida que t aumenta. Ela se tornará estacionária e estocástica se lhe for acrescentado ϵ_t proveniente de uma distribuição normal¹. Em resumo, uma série estacionária é aquela que flutua em torno de uma mesma média.

¹ A estacionaridade da série depende da distribuição do termo aleatório. Por isso, é importante a hipótese sobre a distribuição.

A série não estacionária tem uma tendência, que pode ter uma natureza determinística ou estocástica. A série não estacionária determinística, acrescida de um componente aleatório extraído de uma distribuição normal, flutua em torno de uma tendência temporal. Por exemplo, a seguinte série é não estacionária com tendência determinística:

$$y_t = \mu + \delta t + \epsilon_t.$$

A série não estacionária com tendência estocástica move-se em torno de médias flutuantes. Uma série não estacionária puramente estocástica tem a seguinte configuração, visualizada no gráfico inferior esquerdo da Figura 1.1:

$$y_t = y_{t-1} + \epsilon_t.$$

Caso se acrescente uma constante, μ , a esse modelo, comumente denominada *drift*, trata-se de um modelo não estacionário com tendência estocástica e *drift*, conforme gráfico inferior direito da Figura 1.1.

A forma de estimação econométrica depende de como classificar essas séries, se estacionárias ou não. Portanto, a primeira preocupação é definir essa condição. As séries estacionárias são análogas a séries convergentes da matemática. As séries não estacionárias podem ser divergentes. A matemática está particularmente

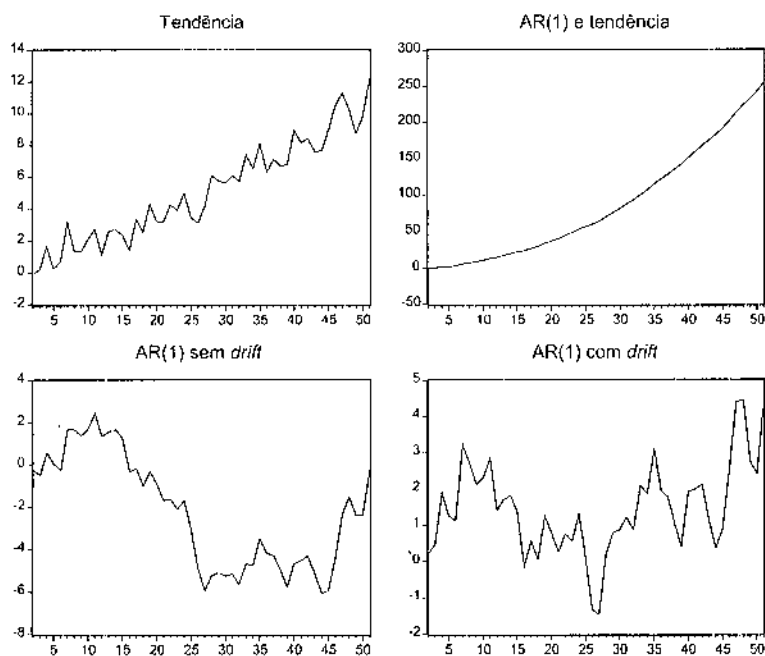


Figura 1.1 *Processos estocásticos.*

interessada em séries convergentes, porque, em geral, permitem análises de equilíbrio. O mesmo ocorre em econometria, porque as inferências estatísticas só terão validade se os resíduos da série temporal estimada forem estacionários. É preciso ter em mente a idéia de que sempre se buscam resíduos estacionários, pois só nesse caso é possível confiar nos testes estatísticos de coeficientes e da regressão. As razões mais técnicas desses argumentos serão dadas mais adiante. Mas a idéia central é a seguinte: as séries não estacionárias não têm média e variância constantes ao longo do tempo, contrariamente às séries estacionárias; logo, não há dados suficientes para estimar média e variância de uma série não estacionária.

Em economia, existem séries estacionárias e não estacionárias. Em geral, retorno de ações e crescimento do PIB, entre outros exemplos, são séries estacionárias. Índices de preços e nível de produto são exemplos comuns de séries não estacionárias. Veja o exemplo de uma série real estacionária na Figura 1.2.

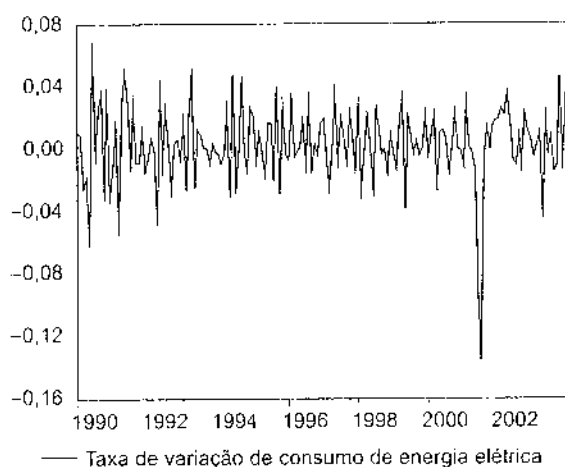


Figura 1.2 *Série estacionária – Brasil.*

O capítulo seguinte retoma alguns fundamentos estatísticos que embasam a econometria de séries temporais. Eles definem formalmente o que é uma série estacionária e ajudam, nos capítulos seguintes, a entender como diferenciar uma série estacionária de uma não estacionária e como estimar séries estacionárias e não estacionárias.

2

Fundamentos Estatísticos

Antes de estudar em profundidade a modelagem de séries temporais, alguns conceitos estatísticos devem estar consolidados. Assim, o objetivo do capítulo é estudar as propriedades estatísticas dos processos estocásticos e reconhecer quais restrições devem-se impor sobre uma série temporal, para que seja factível calcular os momentos da série. O capítulo segue Hamilton (1994) em algumas partes. A estratégia é estudar o processo auto-regressivo, que é intuitivo para uma infinidade de casos, de forma a motivar as restrições necessárias a uma série temporal, para ser usada em estimações.

2.1

Esperanças Condicional e Incondicional

Em econometria, é muito importante entender o conceito de esperança condicional e esperança incondicional (ou não condicional). Em séries temporais, isso toma forma bastante própria que merece uma breve discussão, já que se trata de um conceito fundamental sobre o qual se baseia grande parte do apelo dessa metodologia econométrica.

Para iniciar, convém lembrar a definição de esperança não condicional ou *incondicional*. Assim, considere o espaço amostral Ω . Nesse espaço, pode-se calcular a esperança não condicional de uma variável aleatória Y :

$$E(y|\Omega) = E(y).$$

Quando não se determina claramente a que conjunto se refere a esperança, trata-se do espaço amostral, Ω . Assim, define-se a *lei das expectativas totais*:

$$E[E(y|\Omega)] = E(y|\Omega) = E(y).$$

Um exemplo particularmente interessante dessa lei é o seguinte. Considere X a matriz de variáveis explicativas de um modelo econométrico e ϵ_i uma perturbação aleatória. Assuma que $E(\epsilon_i|X) = 0$. Então, da lei das expectativas totais resulta que:

$$E\left[\underbrace{E(\epsilon_i|X)}_{=0}\right] = E(\epsilon_i) = 0.$$

Isto é, a *média incondicional* de ϵ_i é nula para $i = 1, 2, \dots, n$.

Seja, agora, \mathfrak{F} todos os subconjuntos do espaço amostral Ω , sobre o qual y está definido. Pode-se, então, definir a lei das expectativas iteradas. Considere os conjuntos $A, B \subseteq \mathfrak{F}$, então, por expectativas iteradas:

$$E[E(y|A, B)|A] = E(y|A).$$

Ou seja, o menor conjunto de informação é que determina a média condicional. Há uma consequência importante da lei das expectativas iteradas. Considere a variável $x \in X$; então, qual a esperança incondicional do produto $x\epsilon_i$? Pode-se usar a lei das expectativas iteradas para resolver o problema:

$$E(x\epsilon_i) = E[E(x\epsilon_i|X)] = E[xE(\epsilon_i|X)] = 0.$$

Dada essa pequena introdução, aplicam-se esses conceitos ao caso de séries temporais. Em séries temporais, a cronologia das observações é fundamental e não pode ser quebrada. Assim, é imediatamente intuitivo que o conjunto de informação de um agente no período t , I_t é maior que no período $t-1$, I_{t-1} . Isso pode ser caracterizado formalmente:

$$\mathfrak{F} \supset I_t \supset I_{t-1} \supset \dots \supset I_0 \supseteq \Omega.$$

Ignorando a formalidade, intuitivamente o conjunto \mathfrak{F} é muito maior que I_t . Ele contém toda a informação gerada até t , naquela ordem cronológica, porém poderia conter toda a informação até t gerada de forma distinta daquela observada. O essencial é entender que, contido em I_t , está toda a informação passada, incluindo a cronologia ordenada das informações. O conjunto Ω representa, intuitivamente, o espaço amostral de onde se sorteia o resultado observado em cada instante de tempo.

Considere um modelo AR (1), sobre o qual a próxima seção debruça-se em profundidade:

$$y_t = \mu + \phi y_{t-1} + \epsilon_t,$$

em que se assume por simplicidade que $E(\epsilon_t) = E(\epsilon_t | I_{t-1}) = 0$.

A esperança de y_{t+2} condicional à informação I_{t+1} é:

$$E(y_{t+2} | I_{t+1}) = \mu + \phi E(y_{t+1} | I_{t+1}) = \mu + \phi y_{t+1}. \quad (1)$$

A esperança não condicional é:

$$E(y_{t+2}) = \mu + \phi E(y_{t+1}).$$

Nesse ponto, é conveniente observar que nada garante que $E(y_{t+2}) = E(y_{t+1})$. Uma razão para explicar esse fato é que a distribuição de y pode ter mudado entre $t+1$ e o período seguinte. Essa mudança pode ser explicada economicamente, a partir de um evento relevante como a mudança de política econômica.

Exemplo 2.1

Considere a Figura 2.1, em que se desenha a evolução de y_t ao longo de três datas.

No ponto $t = 0$, as probabilidades condicional e incondicional de y_1 e y_2 são idênticas, conforme mostram os cálculos a seguir:

$$E(y_1 | I_0) = E(y_1) = 3 \times 0,7 + 2 \times 0,3 = 2,7;$$

$$E(y_2 | I_0) = E(y_2) = 2 \times (0,7 \times 0,6) + 4 \times (0,7 \times 0,4) +$$

$$3 \times (0,3 \times 0,5) + 5 \times (0,3 \times 0,5) = 3,16.$$

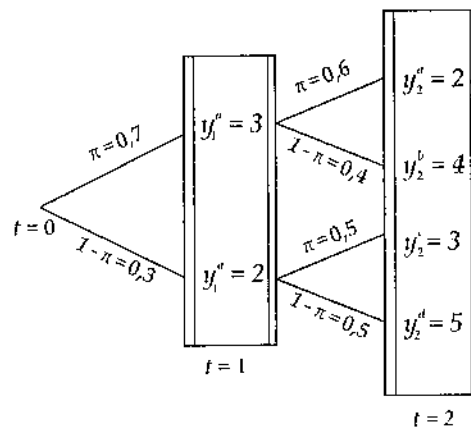


Figura 2.1 Probabilidade condicional e incondicional.

Observe aqui que $E(y_1) \neq E(y_2)$.

No ponto $t = 1$, a esperança condicional se diferencia claramente da incondicional, pois vai depender de onde se encontra y , ou em y_1^a ou em y_1^b , para que se determine $E(y_2|I_1)$:

$$E(y_2|I_1 = y_1^a) = 2 \times 0,6 + 4 \times 0,4 = 2,8;$$

$$E(y_2|I_1 = y_1^b) = 3 \times 0,5 + 5 \times 0,5 = 4;$$

$$E[E(y_2|I_1)] = 2,8 \times 0,7 + 4 \times 0,3 = 3,16 \approx E(y_2).$$

A última linha mostra o seguinte fato interessante: a esperança da esperança condicional é igual à esperança incondicional. Portanto, sabendo-se que prevalece a esperança condicional tomada sob o conjunto com informação mais limitado, é forçoso concluir que o conjunto de informação da média incondicional é mais limitado que o da média condicional.

Prosseguindo com a análise desse modelo simples, e se a esperança de y_{t+2} fosse condicionada à I_t ? Teríamos, então:

$$E(y_{t+2}|I_t) = \mu + \phi E(y_{t+1}|I_t).$$

Agora, veja o que ocorre quando se toma a esperança de $E(y_{t+2}|I_{t+1})$, condicional a um conjunto de informação contido em I_t :

$$E[E(y_{t+2}|I_{t+1})|I_t] = \mu + \phi E[E(y_{t+1}|I_{t+1})|I_t] = \mu + \phi E(y_{t+1}|I_t).$$

Da última igualdade, conclui-se que:

$$E[E(y_{t+2}|I_{t+1})|I_t] = E(y_{t+2}|I_t).$$

Esse resultado é esperado porque prevalece a esperança sob o conjunto de informação mais limitado – nesse caso, I_t em relação a I_{t+1} . Similarmente, tomando a esperança incondicional da equação (1), tem-se que:

$$E[E(y_{t+2}|I_{t+1})] = \mu + \phi E[E(y_{t+1}|I_{t+1})] = \mu + \phi E(y_{t+1}).$$

Logo, não é difícil concluir que $E[E(y_{t+2}|I_{t+1})] = E(y_{t+2})$. Assim, pela lógica da lei das expectativas iteradas, o conjunto de informação sobre o qual toma-se a média incondicional está contido no conjunto de informação da média condicional.

A lógica estatística do resultado é a seguinte: a variável y_t assume valores em $\Omega \subset I_s, \forall s$. Em outras palavras, os valores que y_t assume não podem estar dissociados da informação contida em I_s . A lógica econômica é a seguinte: a informação condicional permite obter a esperança condicional de y_{t+h} . Essa esperança

condicional possui um erro quadrático médio de previsão menor do que se fosse usada a média incondicional de y . Portanto, o conjunto condicionante contém mais informação que o conjunto incondicional e, por isso, o erro quadrático médio é menor.

Outra forma de ver isso é imaginando duas variáveis z e x , de modo que a média incondicional de z seja dada por:

$$E(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f(x, z) dx dz,$$

em que $f(x, z)$ é a função densidade de probabilidade.

A média condicional de z , dado x , é:

$$E(z|x) = \int_{-\infty}^{\infty} z f(z|x) dz,$$

em que

$$f(z|x) = \frac{f(x, z)}{f(x)}; \quad f(x) = \int_{-\infty}^{\infty} f(x, z) dz.$$

Logo, tem-se que:

$$E(z|x) = \int_{-\infty}^{\infty} z \frac{f(x, z)}{f(x)} dz.$$

Tomando a esperança incondicional, encontra-se:

$$\begin{aligned} E\{E(z|x)\} &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} z \frac{f(x, z)}{f(x)} dz \right] f(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z f(x, z) dz dx \\ &= E(z). \end{aligned}$$

Uma discussão mais profunda sobre esses fundamentos encontra-se em Gallant (1997), em que interessa particularmente o teorema 2.5, pelo qual o autor demonstra a lei das expectativas iteradas. O teorema contém a demonstração da lei das expectativas totais, que é um caso particular da lei das expectativas iteradas.

Exemplo 2.2

A lei das expectativas iteradas está associada às expectativas racionais. É interessante entender um exemplo em que há falha nas expectativas iteradas e, portanto, na racionalidade¹. Dois sujeitos A e B discutem previsões sobre um jogo de futebol.

¹ Exemplo criado por Ricardo Suganuma

A: Qual é sua previsão para o jogo no fim de semana?

B: Meu time vai vencer por 2×0 .

A: E qual acha que será a sua previsão amanhã para o mesmo jogo?

B: Meu time vai empatar.

O sujeito B não entendeu a lei das expectativas iteradas, pois não pode ter ocorrido mudança no conjunto de informação dele (a menos que, durante a conversa, tenha se lembrado de que A é seu chefe e torce pelo outro time; e nesse caso, não usou toda a informação disponível no momento da decisão, ou seja, o sujeito não é racional à luz da teoria).

2.2

Processos Auto-Regressivos

Um processo auto-regressivo imediato é o seguinte:

$$y_t = \phi y_{t-1} + \epsilon_t,$$

$$\epsilon_t \sim i.i.d.(0, \sigma^2),$$

em que *i.i.d.* implica idêntica e independentemente distribuído.

Esse processo era muito popular na época megainflacionária que o Brasil viveu até meados dos anos 1990. Para visualizar a razão, suponha que y_t represente a inflação. Dizia-se que a inflação no Brasil era inercial, o que se traduzia formalmente por $\phi = 1$. Nesse contexto, é imediato tentar estimar a série por mínimos quadrados ordinários e testar se $H_0 : \phi = 1 \times H_1 : \phi < 1$. Entretanto, quando se trata de séries temporais, esse tipo de estimação pode gerar sérios problemas. Considere o principal deles, calculando os momentos incondicionais desse processo.

$$E(y_t) = \phi E(y_{t-1}) + E(\epsilon_t).$$

Por simplicidade, assuma que os momentos incondicionais são iguais, isto é, $E(y_t) = E(y_{t-1})$. Se isso não fosse assim, $E(y_t) \neq E(y_{t-1})$ e, de imediato, não se poderiam estimar os momentos da série por falta de dados. Não seria possível estimar t médias e variância dessa série, por exemplo, com apenas t observações. Portanto, é preciso que $E(y_t) = E(y_{t-1})$. Nessas condições, conclui-se que:

$$E(y_t) = 0.$$

Uma forma econômica de entender essa relação incondicional é lembrar a condição de equilíbrio de longo prazo para uma variável determinística $y_t = y_{t-1}$

para todo t . O “equivalente” dessa condição em uma economia estocástica é ter as esperanças incondicionais iguais para todo t .

Outro problema fundamental surge no cálculo da variância:

$$\text{var}(y_t) = \phi^2 \text{var}(y_{t-1}) + \text{var}(e_t) + 2\phi \text{cov}(y_{t-1}, e_t).$$

Ora, como $e_t \sim i.i.d.$, então $\text{cov}(y_{t-1}, e_t) = 0$. Admitindo que a série é gerada pelo mesmo processo, $\text{var}(y_t) = \text{var}(y_{t-1})$, conclui-se que:

$$\text{var}(y_t) = \frac{\sigma^2}{1 - \phi^2}.$$

Dessa forma, se $|\phi| > 1$, a variância de y_t seria negativa, o que é um absurdo. Se $|\phi| = 1$, a variância de y_t é infinita, o que impossibilita, em princípio, a inferência estatística.

Desse pequeno exemplo, então, é possível concluir que é necessário estabelecer algumas restrições sobre a série temporal para que se possa estimá-la. Em particular, uma condição necessária para estimar a série é que $|\phi| < 1$.

De fato, os processos auto-regressivos são muito importantes porque definem se uma série temporal estocástica é “estável” ou estacionária. Um processo auto-regressivo estacionário possui coeficientes que fazem y_t flutuar ao redor de uma dada média, isto é, y_t não explode. Uma discussão mais detalhada sobre esse assunto encontra-se no apêndice.

A seguir, serão desenvolvidos alguns conceitos estatísticos primários, aplicáveis a séries temporais, para serem usados em previsões e testes de modelos.

2.3

Processos Estocásticos

Observa-se uma série temporal $\{y_1, y_2, \dots, y_T\}$, decorrente de uma variável aleatória Y . A série temporal observada é *uma possível* realização do processo estocástico gerador dos dados. Por que *uma possível*? Para entender essa questão, imagine que amanhã ou chove ou faz sol. Suponha que amanhã acabe fazendo sol. Entretanto, poderia ter chovido. Uma possível realização foi o tempo de sol; a outra possível seria chover.

Figurativamente, pode-se pensar em uma dimensão do mundo fazendo sol e uma outra, chovendo. O leitor encontra-se na dimensão que faz sol e não vê a “outra dimensão”, como se vivesse em uma região seca, mas houvesse outra região chuvosa ao mesmo tempo.

Para entender o que significa “outra realização” mais concretamente, considere o exemplo a seguir. O leitor está em um local em que há sol, entretanto

possui um amigo que lhe relata chuva no lugar em que se encontra. A realização observada para o leitor foi o sol, embora pudesse ter havido chuva, como ocorreu a seu amigo. Do ponto de vista estatístico, imagina-se que o mesmo processo gerador de dados que determinou o sol determina a chuva em outra localidade.

Essa noção é importante para entender o conjunto de restrições que deverá ser feito para estudar uma série temporal. Na prática, infelizmente, o econometrista só consegue visualizar uma possível realização, mas precisará fazer inferências que seriam fáceis se observasse todas as possíveis realizações.

Veja na Figura 2.2 duas possibilidades de seqüências, Y , em que uma tem a ver com essa realidade concreta observada e a outra, com a realidade virtual experimentada pelo amigo. Essa variável pode ter qualquer distribuição de probabilidades, mas, em particular, usa-se uma normal para simulá-la: $Y \sim N(0, \sigma^2)$. Convença-se de que, na prática, só se observa uma delas.

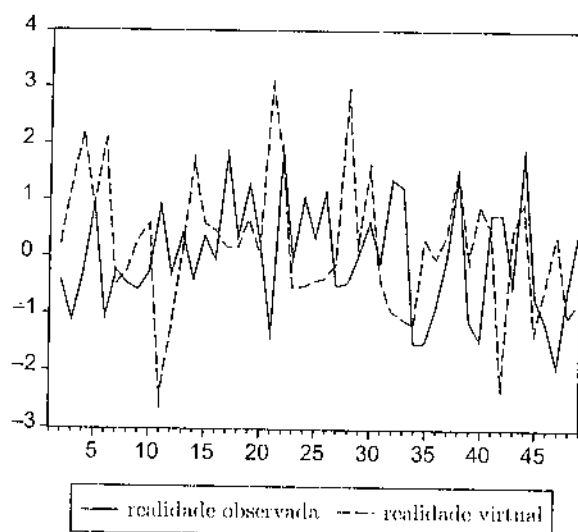


Figura 2.2 Realizações de um processo estocástico.

O que se observa, portanto, é uma possível realização de um processo estocástico. Na mesma linha de raciocínio, é possível gerar artificialmente as várias seqüências de realizações de um processo estocástico por meios computacionais. Suponha que se queira gerar S seqüências, cada uma com infinitas observações, como se houvesse S diferentes regiões geográficas. Então, haveria um conjunto com as seguintes seqüências:

$$\left(\left\{ y_t^{(1)} \right\}_{t=-\infty}^{\infty}, \left\{ y_t^{(2)} \right\}_{t=-\infty}^{\infty}, \dots, \left\{ y_t^{(S)} \right\}_{t=-\infty}^{\infty} \right).$$

Isso significa que, em cada instante de tempo t , há S estados da natureza. Se fosse possível selecionar as S observações associadas à data t , teríamos o conjunto de dados $\{y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(S)}\}$, cuja distribuição, por construção, é normal. Nessas circunstâncias, os vários momentos da série poderiam ser interferidos. Com essa noção intuitiva, pode-se mencionar a seguinte definição baseada em Spanos (1986), sabendo que Ω representa o conjunto amostral e \mathfrak{S} os subconjuntos de Ω , cuja probabilidade associada a cada um deles é P :

Definição 2.1 *Sejam $(\Omega, \mathfrak{S}, P)$ um espaço de probabilidades e \mathbb{Z} um conjunto de índices de números reais. Defina a função $y(\cdot, \cdot)$ por $y(\cdot, \cdot) : S \times \mathbb{Z} \rightarrow \mathbb{R}$. A sequência ordenada de variáveis aleatórias $\{y(\cdot, t), t \in \mathbb{Z}\}$ é chamada de processo estocástico (aleatório).*

A definição anterior significa que, para cada $t \in \mathbb{Z}$, $y(\cdot, t)$ representa uma variável aleatória sobre S . Além disso, para cada $s \in S$, a sequência $y(s, \cdot)$ representa uma realização de um processo estocástico. Desse modo, $y(s, t)$ é simplesmente um número real.

Definição 2.2 *As funções $\{y(s, \cdot), s \in \Omega\}$ em \mathbb{Z} são realizações do processo $\{y(\cdot, t), t \in \mathbb{Z}\}$.*

Formalmente, a esperança não condicional da variável aleatória y_t é:

$$E(y_t) \equiv \int_{-\infty}^{\infty} y_t f_{y_t}(y_t) dy_t = p \lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S y_t^{(s)},$$

em que $p \lim$ é o limite de probabilidade.

Por que é necessário qualificar como esperança não condicional? Mais tarde será vista a diferença entre esperança condicional e não condicional e será possível compreender plenamente o significado. Por ora, basta entender a expressão como a esperança que não depende de informações passadas; depende apenas da distribuição da variável aleatória Y . Por isso, a integral é tomada no conjunto de medida de Y , razão por que o integrando é dy_t .

Nota 2.1 *O conjunto de medida está associado a y_t , não ao tempo.*

Exemplo 2.3

Calcule a esperança não condicional de Y . Sejam $\epsilon_t \sim N(0, \sigma^2)$ e $y_t = \mu + \epsilon_t$; então:

$$E(y_t) = \mu + E(\epsilon_t) = \mu.$$

Se, por outro lado, fosse especificado que $y_t = \delta t + \epsilon_t$, então:

$$E(y_t) = \delta t.$$

Essa última esperança também é não condicional. Para fundir os dois exemplos em um só, pode-se escrever:

$$E(y_t) = \mu_t = \begin{cases} \mu, & \text{se } y_t = \mu + \epsilon_t; \\ \delta t, & \text{se } y_t = \delta t + \epsilon_t. \end{cases}$$

A definição dada é muito geral e torna difícil a modelagem de fenômenos econômicos sem maiores restrições. Observe que se pode ter a esperança como dependente do tempo. Nesse caso, com T observações, teriam de ser estimados T esperanças e os momentos de ordem superior. Porém, isso é impossível, dado que o número de observações é menor do que o número de parâmetros. Por essa razão, é preciso estudar as propriedades estatísticas dos processos estocásticos, de forma a reconhecer quais as restrições adicionais a impor sobre a série de tempo, a fim de calcular os momentos da série.

2.4

Autocovariância e Autocorrelação

Dada uma particular realização, s , de um processo estocástico, a função de autocovariância é definida como:

$$\gamma_{jt} \equiv E(y_t - \mu_t)(y_{t-j} - \mu_{t-j}) = p \lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S (y_t^{(s)} - \mu_t)(y_{t-j}^{(s)} - \mu_{t-j}).$$

Exemplo 2.4

Sejam $\epsilon_t \sim N(0, \sigma^2)$ e $y_t = \mu + \epsilon_t$, então:

$$\gamma_{jt} = E(y_t - \mu_t)(y_{t-j} - \mu_{t-j}) = E(\epsilon_t \epsilon_{t-j}) = \begin{cases} \sigma^2, & j = 0 \\ 0, & j \neq 0. \end{cases}$$

Nota 2.2 A variância é dada por γ_{0t} . Observe que as variâncias não condicionais de $y_t = \mu + \epsilon_t$ e $y_t = \delta t + \epsilon_t$ são idênticas.

Dada a autocovariância, pode-se facilmente calcular a autocorrelação definida da seguinte forma:

$$\rho_{jt} \equiv \frac{\gamma_{jt}}{\gamma_{0t}}.$$

2.5

Estacionaridade

O conceito de estacionaridade é a principal idéia que se deve ter para estimar uma série temporal. É fundamentalmente a constatação de estacionaridade que permitirá proceder a inferências estatísticas sobre os parâmetros estimados com base na realização de um processo estocástico. Se nem a esperança nem a autocovariância dependem do tempo, então y_t é fracamente estacionário².

Definição 2.3 O processo estocástico, ou a série temporal, $\{y_t, t \in \mathbb{Z}\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ é fracamente estacionário se:

- a) $E|y_t|^2 < \infty$;
- b) $E(y_t) = \mu$, para todo $t \in \mathbb{Z}$, e
- c) $E(y_t - \mu)(y_{t-j} - \mu) = \gamma_j$.

A primeira condição afirma apenas que o segundo momento não centrado deve ser *finito*, ainda que desigual em diferentes períodos. A segunda condição afirma que a média é *igual* para todo período, mesmo que a distribuição da variável aleatória vá se alterando ao longo do tempo. A terceira condição estabelece que a *variância é sempre igual* para todo período de tempo e que a *autocovariância não depende do tempo*, mas apenas da distância temporal entre as observações.

Outro conceito importante de estacionaridade, desta vez a estrita, usado em certas demonstrações estatísticas é:

Definição 2.4 O processo estocástico, ou a série temporal, $\{y_t, t \in \mathbb{Z}\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ é estritamente estacionário se a função de distribuição de $\{y_t\}_{t=1}^k$ for igual à função de distribuição de $\{y_{t+h}\}_{t=1}^k$, $h \in \mathbb{Z}$, isto é:

$$F(y_{t_1}, y_{t_2}, \dots, y_{t_k}) = F(y_{t_1+h}, y_{t_2+h}, \dots, y_{t_k+h}).$$

Intuitivamente, estacionaridade estrita significa que os gráficos da função de distribuição da série em quaisquer dois intervalos de tempo de igual tamanho exibirão propriedades estatísticas similares (Brockwell e Davis, 1991). Significa, na prática, que os momentos populacionais, quando existem, são independentes de t .

² O conceito de estacionaridade fraca também é conhecido como *estacionaridade, covariância estacionária, estacionaridade lato sensu* ou *estacionaridade de segunda ordem*.

A estacionaridade estrita não implica nem é implicada pela estacionaridade fraca. Por exemplo, uma distribuição de Cauchy pode definir uma série estritamente estacionária, mas não será uma série estacionária, já que seu momento de primeira ordem não é bem definido. Por outro lado, pode existir uma distribuição diferente para dois tempos distintos, mas que satisfaça a definição de estacionaridade, ou seja, tenha o primeiro e o segundo momentos iguais.

Nota 2.3 *Em resumo, estacionaridade se associa vagamente à estacionaridade estrita e vice-versa:*

1. *Estacionaridade forte \Rightarrow estacionaridade. Nada garante que existam momentos finitos e, se existirem, se são iguais ou independentes de t ;*
2. *Estacionaridade \Rightarrow estacionaridade forte. Nada garante que não haja alteração de distribuição por translação temporal;*
3. *Estacionaridade + normalidade \Rightarrow estacionaridade forte. A distribuição não se alterará por translação temporal.*

Nota 2.4 *Estacionaridade estrita é uma propriedade mais forte que distribuição idêntica, contudo é mais fraca que distribuições idêntica e independentemente distribuídas. Sequências i.i.d. são estritamente estacionárias, mas sequências estritamente estacionárias não precisam ser necessariamente independentes.*

Nota 2.5 *A condição necessária para estacionaridade fraca é que as raízes da equação característica devem estar fora do círculo unitário.*

Nota 2.6 *Visualmente, observa-se estacionaridade se uma série flutua em torno de uma média fixa e se a variância da série é constante ao longo do tempo. Não obstante, são necessários testes estatísticos para verificar ou não a estacionaridade da série.*

2.6 Ergodicidade

Com a propriedade de estacionaridade apenas não é possível estimar uma série temporal. Esta seção deixa claro por que isso ocorre, por meio de um exemplo retirado de Hamilton (1994). Essencialmente, para modelagem de séries temporais e sua estimação é necessário satisfazer a propriedade de ergodicidade.

Essa propriedade permite usar uma série temporal para calcular as médias em cada instante de tempo. Como as médias são todas iguais, basta uma única realização da série para viabilizar o cálculo.

Suponha uma particular realização, s , de um processo estocástico, justamente a única série que se observa. A média temporal dessa realização é dada por:

$$\bar{y}^{(s)} = \frac{1}{T} \sum_{t=1}^T y_t^{(s)}.$$

Se \bar{y} convergir para $E(y_t)$, existe ergodicidade. Ou seja, se a média temporal convergir para a média não condicional, há ergodicidade. Tendo isso, a série temporal pode ser estimada normalmente, mesmo com uma realização apenas do processo estocástico.

Definição 2.5 Um processo fracamente estacionário é ergódico para o primeiro momento se:

$$\begin{aligned} E(\bar{y}^{(s)}) &\equiv p \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T y_t^{(s)} \\ &= p \lim_{S \rightarrow \infty} \frac{1}{S} \sum_{s=1}^S y_t^{(s)} \\ &\equiv E(y_t). \end{aligned}$$

No fundo, o que se tenta impor é o seguinte: primeiro, a esperança de cada uma das observações é igual (restrição de estacionaridade fraca). Segundo, pode-se estimar essa esperança tomando a média temporal das observações (restrição de ergodicidade).

Prova-se que $\{y_t\}$ é ergódico para a média se a soma das covariâncias for finita (veja prova em Brockwell e Davis, 1991):

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty. \quad (2)$$

Algumas observações são importantes agora.

Nota 2.7 Para haver ergodicidade, a série deve ser fracamente estacionária. Portanto, ergodicidade é mais restrita que estacionaridade.

Nota 2.8 Se a autocovariância vai a zero suficientemente rápido quando j aumenta, pode-se provar que a série é ergódica.

Definição 2.6 Um processo fracamente estacionário é ergódico para o segundo momento se:

$$\frac{1}{T-j} \sum_{t=1}^T (y_t - \mu) (y_{t-j} - \mu) \xrightarrow{P} \gamma_j, \text{ para todo } j,$$

em que \xrightarrow{P} significa convergência em probabilidade.

Na prática, requerer estacionaridade acaba sendo o mesmo que requerer ergodicidade. Entretanto, o exemplo a seguir mostra um caso de estacionaridade sem existir ergodicidade.

Exemplo 2.5

(Hamilton, 1994) Seja a realização de um processo estocástico $\{y_t^i\}_{t=-\infty}^{\infty}$, cuja média é dada por $\mu^i \sim N(0, \lambda^2)$. Considere:

$$y_t^s = \mu^s + \epsilon_t, \mu^s \perp \epsilon_t \sim i.i.N(0, \sigma^2).$$

Perceba que:

$$\begin{aligned} \mu_t &= E(\mu^s + \epsilon_t) = 0; \\ \gamma_{0t} &= E(\mu^s + \epsilon_t)^2 = \lambda^2 + \sigma^2; \\ \gamma_{jt} &= E(\mu^s + \epsilon_t)(\mu^s + \epsilon_{t-j}) = \lambda^2, \quad j \neq 0. \end{aligned}$$

O processo $\{y_t^s\}_{t=-\infty}^{\infty}$ satisfaz a definição de estacionaridade fraca. Contudo, a média temporal dessa variável não converge para zero, conforme mostra-se a seguir:

$$\frac{1}{T} \sum_{t=1}^T (\mu^s + \epsilon_t) = \mu^s + \frac{1}{T} \sum_{t=1}^T \epsilon_t \xrightarrow{P} \mu^s \neq 0.$$

Veja que o processo não satisfaz a condição suficiente $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, a menos que $\lambda^2 = 0$.

As condições de estacionaridade podem ser verificadas empírica e teoricamente, ao contrário das de ergodicidade. Por outro lado, muitas séries econômicas parecem nem ser estacionárias. Nesses casos, algumas condições adicionais têm de ser impostas, como será visto mais adiante.

2.7

Ruído Branco

Um processo fundamental das séries temporais estocásticas discretas é o ruído branco. Uma sequência $\{\epsilon_t\}$ é um ruído branco se cada valor nela tiver média zero, variância constante e não for correlacionado com qualquer realização da própria série (autocorrelação igual a zero). A média zero é conveniência, pois seria possível especificar um ruído branco cuja média fosse diferente de zero. Entretanto, pode-se centrar em zero tal série, sem prejuízo de suas demais propriedades.

Definição 2.7 *Seja uma sequência $\{\epsilon_t\}_{t=-\infty}^{\infty}$ de variáveis aleatórias. Se:*

$$\begin{aligned} E(\epsilon_t) &= 0; \forall t \\ E(\epsilon_t^2) &= \sigma^2; \forall t \\ E(\epsilon_t \epsilon_{t-j}) &= 0, \text{ todo } j \neq 0, \end{aligned}$$

diz-se que o processo é um ruído branco, cuja representação é $RB(0, \sigma^2)$.

Tal processo é chamado ruído branco porque sua função densidade espectral é horizontal como a luz branca e o processo provoca alterações na série assim como as ondas eletromagnéticas produzem ruídos na sintonização de um rádio. Portanto, um ruído branco é, ao mesmo tempo, temporalmente homogêneo, estacionário e sem memória³.

Nota 2.9 *Séries independentes como ϵ_t implicam que $E(\epsilon_t \epsilon_{t-j}) = 0$, mas não são implicadas por essa condição. Independência é uma propriedade mais forte do que autocovariância nula.*

Exemplo 2.6

A Figura 2.3 mostra um ruído branco ($\sigma^2 = 1$).

³ O conceito de sem memória aqui precisa ser qualificado. O sentido é o de ser um processo em que a dependência temporal, se existe, está implícita na série. Uma série $\epsilon_t \sim i.i.d.(0, \sigma^2)$ certamente é um ruído branco, entretanto nem todo ruído branco é *i.i.d.*, não obstante satisfaça as condições da definição.

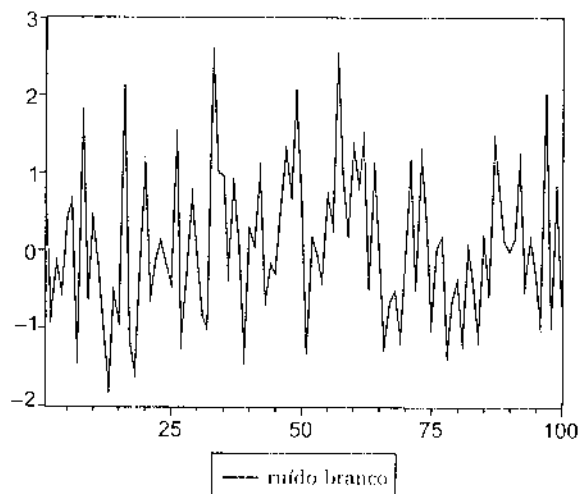


Figura 2.3 *Possível realização de um ruído branco.*

2.8

Médias Móveis

Aqui inicia-se a forja das técnicas de estimação de séries temporais. A estratégia é partir das formas mais simples e, vagarosamente, incrementar o modelo. A importância de entender bem esses processos é que, por meio deles, será identificado o modelo que melhor se ajusta à série de interesse.

2.8.1 Médias Móveis de Ordem 1 – *MA (1)*

Considere o seguinte processo estocástico:

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1},$$

em que ϵ_t é um ruído branco, conforme definido anteriormente.

Uma vez que y_t depende do erro contemporâneo, ϵ_t , e do erro imediatamente passado, então o processo é chamado médias móveis de ordem 1 e denotado como *MA (1)*. Se o processo dependesse de ϵ_{t-2} , então seria chamado de *MA (2)*, e assim por diante.

É possível que o leitor já tenha ouvido falar de médias móveis em outros contextos e considere essa definição estranha. Por exemplo, é comum calcular a média aritmética das vendas dos últimos 12 meses em cada período de tempo e chamar esse processo de médias móveis, em que móvel aqui se aplica porque o cálculo desliza um período à chegada de nova informação. Esse procedimento

é feito para identificar algum tipo de tendência, expurgando-se a influência de sazonalidade. O contexto da discussão de média móvel neste texto é mais geral. Em primeiro lugar, o processo de média móvel estará sempre associado aos erros do modelo. Além disso, os pesos poderão ser diferentes conforme a importância das observações passadas, em contraposição aos pesos idênticos que costumemente se associam aos valores usados para calcular a média dos últimos 12 meses.

Será que o processo y_t satisfaz a definição de estacionaridade? Para responder a isso, é preciso calcular a esperança, a variância e as autocovariâncias do processo. Os cálculos a seguir fazem isso.

$$E(y_t) = \mu + E(\epsilon_t) + \theta E(\epsilon_{t-1}) = \mu;$$

$$\begin{aligned} \text{Var}(y_t) &= E(y_t - \mu)^2 = E(\epsilon_t + \theta\epsilon_{t-1})^2 \\ &= E(\epsilon_t^2 + 2\theta\epsilon_t\epsilon_{t-1} + \theta^2\epsilon_{t-1}^2) \\ &= \sigma^2 + 0 + \theta^2\sigma^2 = (1 + \theta^2)\sigma^2; \end{aligned}$$

$$\begin{aligned} E[(y_t - \mu)(y_{t-1} - \mu)] &= E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-1} + \theta\epsilon_{t-2})] \\ &= E(\epsilon_t\epsilon_{t-1} + \theta\epsilon_t\epsilon_{t-2} + \theta\epsilon_{t-1}^2 + \theta^2\epsilon_{t-1}\epsilon_{t-2}) \\ &= \sigma^2\theta. \end{aligned}$$

A esperança é constante e finita para cada t . A variância é finita. A autocovariância não depende do t . Além disso, as outras autocovariâncias são nulas:

$$\begin{aligned} 0 &= E[(y_t - \mu)(y_{t-j} - \mu)] \\ &= E[(\epsilon_t + \theta\epsilon_{t-1})(\epsilon_{t-j} + \theta\epsilon_{t-j-1})] \\ &= E(\epsilon_t\epsilon_{t-j} + \theta\epsilon_t\epsilon_{t-j-1} + \theta\epsilon_{t-1}\epsilon_{t-j} + \theta^2\epsilon_{t-1}\epsilon_{t-j-1}), \quad j > 1. \end{aligned}$$

Como esperança e autocovariâncias não são funções do tempo, o processo é fracamente estacionário, independentemente do valor de θ .

Observe que a condição dada pela equação (2) também é satisfeita, gerando um processo ergódico:

$$\sum_{j=0}^{\infty} |\gamma_j| = [(1 + \theta^2) + |\theta|]\sigma^2 < \infty.$$

A autocorrelação só existe para a primeira defasagem e é dada por:

$$\rho_1 = \frac{\theta\sigma^2}{(1 + \theta^2)\sigma^2} = \frac{\theta}{1 + \theta^2}.$$

A seguir, três processos de médias móveis de ordem 1 são simulados na Figura 2.4.

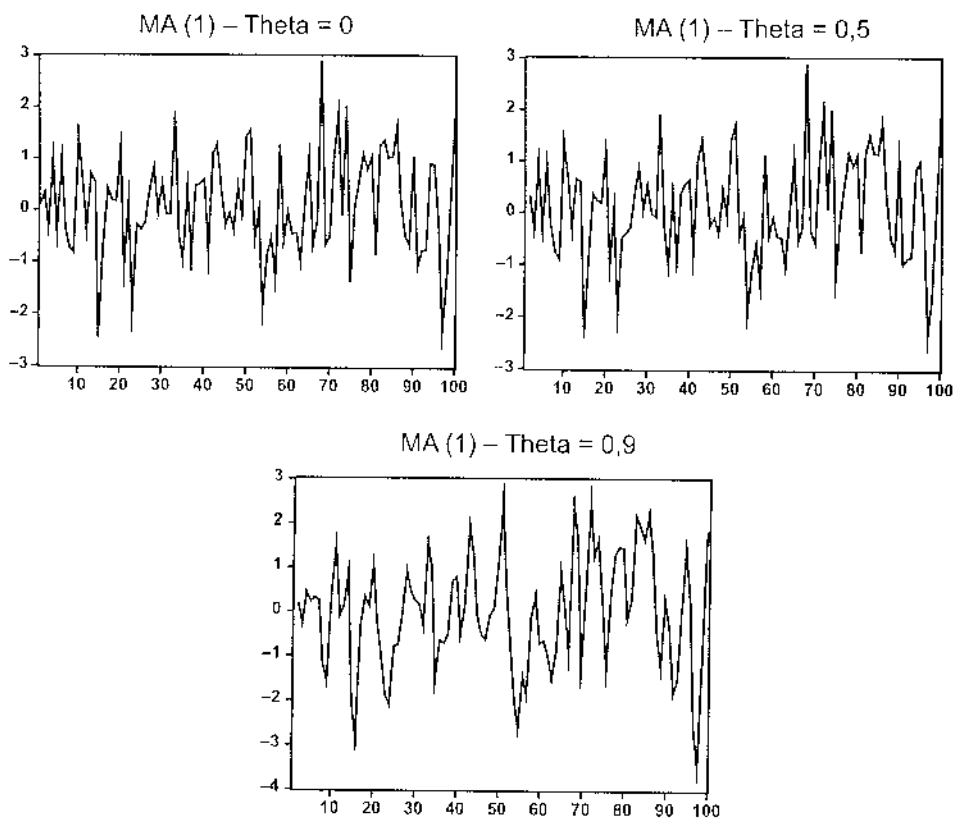


Figura 2.4 Processos $MA(1)$ com vários valores para ϕ .

2.8.2 Médias Móveis de Ordem q – $MA(q)$

A seguir, generaliza-se o processo de médias móveis para q defasagens:

$$y_t = \mu + \sum_{j=0}^q \theta_j \epsilon_{t-j}, \theta_0 = 1.$$

Observe se esse processo satisfaz a primeira condição de estacionaridade.

$$E(y_t) = \mu + \sum_{j=0}^q \theta_j E(\epsilon_{t-j}) = \mu.$$

Como as correlações cruzadas são nulas, tem-se que:

$$\text{Var}(y_t) = E(y_t - \mu)^2 = E\left(\sum_{j=0}^q \theta_j \epsilon_{t-j}\right)^2 = \sum_{j=0}^q \theta_j^2 E(\epsilon_{t-j}^2) = \sigma^2 \sum_{j=0}^q \theta_j^2.$$

Agora, para $j = 1, 2, \dots, q$, tem-se:

$$\begin{aligned} \gamma_j &= E\left(\sum_{i=0}^q \theta_i \epsilon_{t-i} \sum_{i=0}^q \theta_i \epsilon_{t-i-j}\right) \\ &= E\left[\theta_j \epsilon_{t-j}^2 + \theta_{j+1} \theta_1 \epsilon_{t-j-1}^2 + \theta_{j+2} \theta_2 \epsilon_{t-j-2}^2 + \dots + \theta_q \theta_{q-j} \epsilon_{t-q}^2\right]. \end{aligned}$$

Para $j > q$, não haverá ϵ 's em datas comuns. Logo a esperança será nula. Portanto:

$$\gamma_j = \begin{cases} [\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \dots + \theta_q\theta_{q-j}] \sigma^2, & \text{para } j = 1, 2, \dots, q; \\ 0, & \text{para } j > q. \end{cases}$$

Exemplo 2.7

Seja um MA (2), qual é a autocovariância?

$$\gamma_j = \begin{cases} \sigma^2 (1 + \theta_1^2 + \theta_2^2), & j = 0 \\ \sigma^2 (\theta_1 + \theta_1\theta_2), & j = 1 \\ \sigma^2 \theta_2, & j = 2 \\ 0, & j > 2. \end{cases}$$

Conseqüentemente, a função de autocorrelação é dada por:

$$\rho_j = \begin{cases} 1, & j = 0 \\ \frac{(\theta_1 + \theta_1\theta_2)}{(1 + \theta_1^2 + \theta_2^2)}, & j = 1 \\ \frac{\theta_2}{(1 + \theta_1^2 + \theta_2^2)}, & j = 2 \\ 0, & j > 2. \end{cases}$$

Como será visto mais tarde, a autocorrelação pode ser colocada em um gráfico como função da defasagem. Nesse gráfico verifica-se que, a partir da defasagem q , a autocorrelação desaparece. Essa propriedade permite identificar a ordem do processo de médias móveis. Logo, um processo que é truncado na autocorrelação q é identificado como um MA (q), desde que a autocorrelação parcial, a ser definida no próximo capítulo, seja decrescente assintoticamente.

Nota 2.10 Quando $q = \infty$, reescreve-se o processo $MA(\infty)$ da seguinte forma: $y_t = \mu + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \mu + \psi(L) \epsilon_t$, em que $\psi_0 = 1$ e $\psi(L)$ representa a polinomial de ordem infinita dada por $\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots$. O termo L representa o operador defasagem em que $L^j y_t = y_{t-j}$.⁴

Nota 2.11 Se um processo é fracamente estacionário, então $\sum_{j=0}^{\infty} \psi_j^2 < \infty$.

Nota 2.12 Se $\sum_{j=0}^{\infty} |\psi_j| < \infty$, então:

- $\sum_{j=0}^{\infty} \psi_j^2 < \infty$ e
- $\sum_{j=0}^{\infty} |\gamma_j| < \infty$, o que implica ergodicidade para a média.

2.9

Processos Auto-Regressivos

Esta seção retoma o exemplo inicial, $AR(1)$, em detalhes e o utiliza para ilustrar os conceitos fundamentais dos processos auto-regressivos.

2.9.1 Processo Auto-Regressivo de Ordem 1 – $AR(1)$

Considere o seguinte processo estocástico:

$$y_t = c + \phi y_{t-1} + \epsilon_t,$$

em que ϵ_t é um ruído branco.

O processo assim definido é chamado auto-regressivo de ordem 1 e denotado como $AR(1)$. Será que esse processo é estável e tem variância finita, admitindo que $|\phi| < 1$? Uma forma de responder a essa pergunta é usando o operador defasagem, de modo a obter um processo de médias móveis infinito e, em seguida, encontrar a esperança e as autocovariâncias do processo. Usa-se esse procedimento aqui para diferenciá-lo daquele do início do capítulo, no qual encontrou-se o mesmo resultado a partir

⁴ No apêndice, encontra-se uma discussão sobre as propriedades desse termo.

do uso da esperança não condicional. Assim, reescrevendo o processo auto-regressivo de ordem 1, pode-se encontrar um MA (∞):

$$y_t = c + \phi y_{t-1} + \epsilon_t \implies (1 - \phi L) y_t = c + \epsilon_t \implies \\ y_t = \frac{c}{1 - \phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} = \mu + \psi(L) \epsilon_t,$$

em que

$$\mu = \frac{c}{1 - \phi} \quad \text{e} \quad \psi(L) = (1 - \phi L)^{-1} = 1 + \phi L + \phi^2 L^2 + \dots^5$$

Pode-se, então, calcular:

$$E(y_t) = \mu + \sum_{j=0}^{\infty} \phi^j E(\epsilon_{t-j}) = \mu; \\ \text{Var}(y_t) = E(y_t - \mu)^2 = E\left(\sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}\right)^2 = \sum_{j=0}^{\infty} \phi^{2j} E(\epsilon_{t-j}^2) = \frac{\sigma^2}{1 - \phi^2}.$$

A autocovariância de defasagem j é:

$$E[(y_t - \mu)(y_{t-j} - \mu)] = E\left[\left(\sum_{s=0}^{\infty} \phi^s \epsilon_{t-s}\right)\left(\sum_{s=0}^{\infty} \phi^s \epsilon_{t-s-j}\right)\right] \\ = \sigma^2 (\phi^j + \phi^{j+2} + \phi^{j+4} + \dots) \\ = \left(\frac{\phi^j}{1 - \phi^2}\right) \sigma^2.$$

Como média e autocovariâncias não são funções do tempo, o processo é fracamente estacionário, independentemente do valor de ϕ . A autocorrelação de ordem j é dada por:

$$\rho_j = \frac{\left(\frac{\phi^j}{1 - \phi^2}\right) \sigma^2}{\frac{\sigma^2}{1 - \phi^2}} = \phi^j, \quad j = 1, 2, \dots$$

A Figura 2.5 mostra quatro processos auto-regressivos. É possível observar que sua trajetória é bem diferente conforme o valor do coeficiente ϕ , em todos admitindo que $\mu = 0$.

⁵ Para verificar essa igualdade, observe que $(1 - \phi L)(1 + \phi L + \phi^2 L^2 + \dots) = 1$.

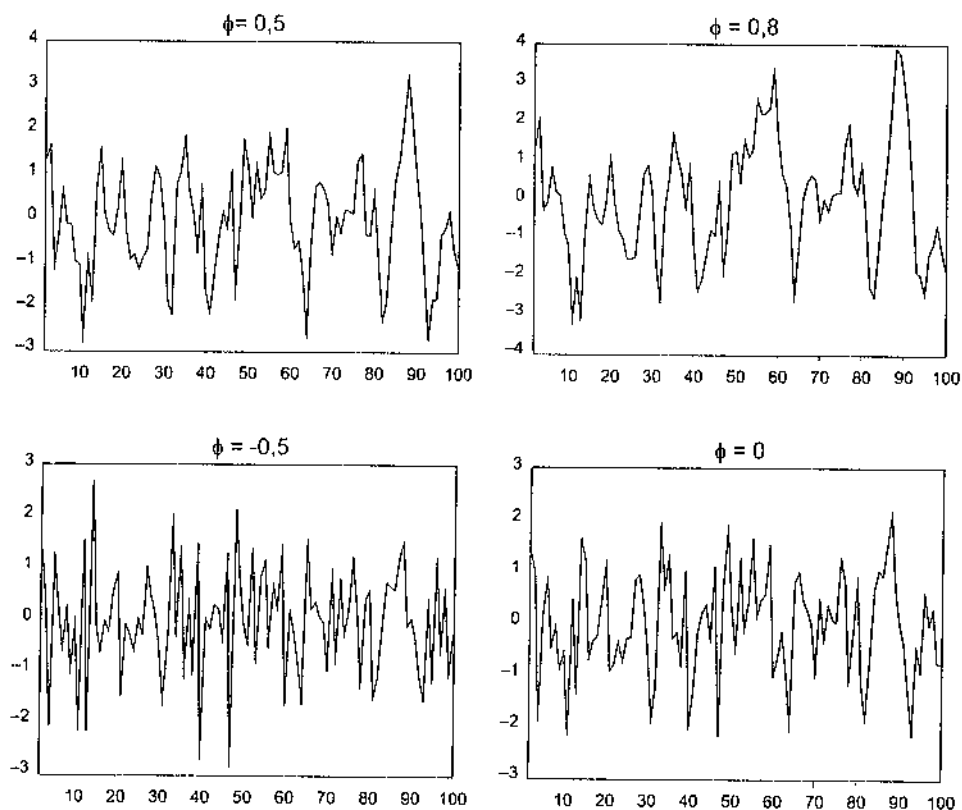


Figura 2.5 Processos AR (1) com vários valores para ϕ .

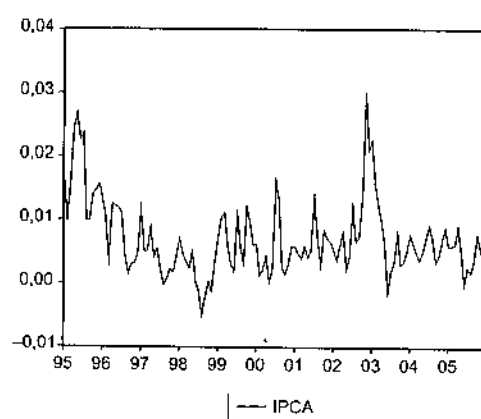


Figura 2.6 Inflação mensal: IPCA.

Comparando as duas figuras superiores, o processo em que $\phi = 0,8$ parece mais resistente a mudanças do que o processo em que $\phi = 0,5$. Comparando as duas figuras à esquerda, o processo inferior dá a impressão de ser mais volátil, embora ambos tenham a mesma variância. Essa volatilidade se reflete nos valores da autocovariância. Quando $\phi = 0$, é difícil definir um padrão para os dados. Trata-se, de fato, de um ruído branco.

A Figura 2.6 mostra a inflação mensal dada pelo IPCA entre janeiro de 1995 e dezembro de 2005. Como será visto adiante, essa variável segue um processo $AR(1)$ em que $\phi \simeq 0,71$.

2.9.2 Processo Auto-Regressivo de Ordem 2 – $AR(2)$

O estudo dos processos auto-regressivos de ordem 2 é importante porque eles geram intuição para processos de ordem maior. Eles complementam os estudos do apêndice a respeito de equações a diferenças de ordem 2.

Seja o seguinte processo auto-regressivo:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t.$$

O problema agora é encontrar a função de autocorrelação para ver as condições necessárias para a estacionaridade do processo. Isso é importante para adquirir intuição para casos de ordem superior. Sob estacionaridade, é verdade que $E(y_t) = E(y_{t-1})$. Logo, pode-se calcular a esperança não condicional de y_t :

$$\begin{aligned} E(y_t) &= c + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + E(\epsilon_t) \implies \\ E(y_t) &\equiv \mu = \frac{c}{1 - \phi_1 - \phi_2}. \end{aligned}$$

Dada a esperança não condicional do processo, é conveniente reescrevê-lo de outra forma, a fim de tornar alguns cálculos mais fáceis:

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \epsilon_t.$$

Multiplicando ambos os lados dessa equação por $(y_{t-j} - \mu)$ e tomando a esperança, e como $(y_{t-j} - \mu)$ não contém qualquer elemento correlacionado com ϵ_t , se $j > 0$, tem-se que:

$$\begin{aligned} E(y_t - \mu)(y_{t-j} - \mu) &= \phi_1 E(y_{t-1} - \mu)(y_{t-j} - \mu) \\ &\quad + \phi_2 E(y_{t-2} - \mu)(y_{t-j} - \mu) + E[\epsilon_t(y_{t-j} - \mu)]. \end{aligned}$$

Logo, por definição, encontra-se:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}, \quad j = 1, 2, \dots$$

Ou seja, a autocovariância segue um processo auto-regressivo de ordem 2. Para calcular a função de autocorrelação, é preciso apenas dividir a equação anterior por γ_0 :

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}, \quad j = 1, 2, \dots$$

Nota 2.13 Esse conjunto de equações está contido na família mais geral, conhecida como equações de Yule-Walker.

Exemplo 2.8

Pode-se usar a equação anterior para calcular a função de autocorrelação desse processo, de forma a resolver um sistema de equações simultâneas:

$$j = 1: \quad \rho_1 = \phi_1 + \phi_2 \rho_1 \implies \rho_1 = \frac{\phi_1}{1 - \phi_2};$$

$$j = 2: \quad \rho_2 = \phi_1 \rho_1 + \phi_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2;$$

$$j = s: \quad \rho_s = \phi_1 \rho_{s-1} + \phi_2 \rho_{s-2}.$$

Para calcular a variância, basta encontrar:

$$\begin{aligned} E(y_t - \mu)^2 &= \phi_1 E(y_{t-1} - \mu)(y_t - \mu) \\ &\quad + \phi_2 E(y_{t-2} - \mu)(y_t - \mu) + E[\epsilon_t(y_t - \mu)] \implies \\ \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2 = \phi_1 \gamma_0 \rho_1 + \phi_2 \gamma_0 \rho_2 + \sigma^2 \implies \\ \gamma_0 &= \gamma_0 \left[\frac{\phi_1^2}{1 - \phi_2} + \frac{\phi_1^2 \phi_2}{1 - \phi_2} + \phi_2^2 \right] + \sigma^2 \implies \\ \gamma_0 &= \frac{(1 - \phi_2) \sigma^2}{1 - \phi_2 - \phi_1^2 - \phi_1^2 \phi_2 + \phi_2^3 - \phi_2^2} \\ &= \frac{(1 - \phi_2) \sigma^2}{-\phi_2^2(1 - \phi_2) - \phi_1^2(\phi_2 + 1) + 1 - \phi_2} \\ &= \frac{(1 - \phi_2) \sigma^2}{(1 - \phi_2)^2(\phi_2 + 1) - \phi_1^2(\phi_2 + 1)} \\ &= \frac{(1 - \phi_2) \sigma^2}{(\phi_2 + 1) \left[(1 - \phi_2)^2 - \phi_1^2 \right]} \\ \gamma_0 &= \frac{(1 - \phi_2) \sigma^2}{(\phi_2 + 1)(1 - \phi_2 - \phi_1)(1 - \phi_2 + \phi_1)}. \end{aligned}$$

2.9.3 Processo Auto-Regressivo de Ordem p – AR (p)

O processo auto-regressivo de ordem p é definido como:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t = c + \sum_{j=1}^p \phi_j y_{t-j} + \epsilon_t.$$

Se as raízes da polinomial $(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p)$ estiverem fora do círculo unitário (Hamilton, 1994) ou, equivalentemente, se as raízes da polinomial $(\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \dots - \phi_p)$ estiverem dentro do círculo unitário (veja Enders, 2004), o processo será fracamente estacionário e poderá ser representado como um MA (∞):

$$y_t = \mu + \psi(L) \epsilon_t,$$

em que

$$\mu = \frac{c}{1 - (\phi_1 + \phi_2 + \dots + \phi_p)},$$

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)^{-1}.$$

Similarmente à seção anterior, pode-se reescrever o processo AR (p) da seguinte forma:

$$y_t - \mu = \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \dots + \phi_p (y_{t-p} - \mu) + \epsilon_t.$$

Multiplicando por $(y_{t-j} - \mu)$ e tomando a esperança:

$$\begin{aligned} E(y_t - \mu)(y_{t-j} - \mu) &= \phi_1 E(y_{t-1} - \mu)(y_{t-j} - \mu) \\ &+ \phi_2 E(y_{t-2} - \mu)(y_{t-j} - \mu) + \dots \\ &+ \phi_p E(y_{t-p} - \mu)(y_{t-j} - \mu) + E[\epsilon_t(y_{t-j} - \mu)], \end{aligned}$$

tem-se:

$$\gamma_j = \begin{cases} \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}, & j = 1, 2, \dots \\ \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma^2, & j = 0. \end{cases}$$

Dado que $\gamma_j = \gamma_{-j}$, há um sistema de $p + 1$ equações simultâneas que pode ser resolvido para $\gamma_0, \gamma_1, \dots, \gamma_p$. Dividindo tudo por γ_0 , encontra-se o sistema de equações de Yule-Walker:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \quad j = 1, 2, \dots, p.$$

2.10

Processo Auto-Regressivo de Médias Móveis – ARMA (p,q)

O processo auto-regressivo de médias móveis é simplesmente a combinação dos processos vistos anteriormente. Assim, um ARMA (p, q) é escrito como:

$$y_t = c + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=0}^q \theta_j \epsilon_{t-j}. \quad (3)$$

Tomando a esperança incondicional da equação anterior, tem-se:

$$\begin{aligned} E(y_t) &= c + \sum_{i=1}^p \phi_i E(y_{t-i}) \implies \\ E(y_t) \equiv \mu &= \frac{c}{1 - \sum_{i=1}^p \phi_i}. \end{aligned} \quad (4)$$

Subtraindo a equação (4) da equação (3) e aplicando a definição de $E(y_t)$, resulta a seguinte equação:

$$y_t = \mu + \sum_{i=1}^p \phi_i (y_{t-i} - \mu) + \sum_{j=0}^q \theta_j \epsilon_{t-j} = \mu + \psi(L) \epsilon_t,$$

em que

$$\psi(L) = \frac{1 + \sum_{j=1}^q \theta_j L^j}{1 - \sum_{i=1}^p \phi_i L^i}.$$

Reescrevendo o processo ARMA em termos dos desvios em relação à média não condicional e multiplicando por $(y_{t-h} - \mu)$, tem-se:

$$(y_t - \mu)(y_{t-h} - \mu) = \sum_{i=1}^p \phi_i (y_{t-i} - \mu)(y_{t-h} - \mu) + \sum_{j=0}^q \theta_j \epsilon_{t-j} (y_{t-h} - \mu).$$

Tomando a esperança para $h > q$, pode-se encontrar um processo auto-regressivo de ordem p , pois $E[\epsilon_{t-j}(y_{t-h} - \mu)] = 0$:

$$\gamma_h = \phi_1 \gamma_{h-1} + \phi_2 \gamma_{h-2} + \dots + \phi_p \gamma_{h-p}, \quad h = q+1, q+2, \dots$$

Se $h \leq q$, a função de autocovariância torna-se bem complicada, pois há correlação entre ϵ_{t-j} e $(y_{t-h} - \mu)$.

Exemplo 2.9

Seja um processo ARMA (1, 1):

$$y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

O problema é calcular a autocovariância desse processo.

$$\gamma_0 = E(\phi_1 y_{t-1} y_t + \epsilon_t y_t + \theta_1 \epsilon_{t-1} y_t) = \phi_1 \gamma_1 + \sigma^2 + \theta_1 (\phi_1 + \theta_1) \sigma^2;$$

$$\gamma_1 = E(\phi_1 y_{t-1} y_{t-1} + \epsilon_t y_{t-1} + \theta_1 \epsilon_{t-1} y_{t-1}) = \phi_1 \gamma_0 + \theta_1 \sigma^2;$$

$$\gamma_2 = E(\phi_1 y_{t-1} y_{t-2} + \epsilon_t y_{t-2} + \theta_1 \epsilon_{t-1} y_{t-2}) = \phi_1 \gamma_1;$$

$$\vdots$$

$$\gamma_h = E(\phi_1 y_{t-1} y_{t-h} + \epsilon_t y_{t-h} + \theta_1 \epsilon_{t-1} y_{t-h}) = \phi_1 \gamma_{h-1}.$$

Resolvendo as duas primeiras equações simultaneamente, resulta:

$$\gamma_0 = \frac{1 + \theta_1^2 + 2\phi_1\theta_1}{1 - \phi_1^2} \sigma^2;$$

$$\gamma_1 = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 - \phi_1^2} \sigma^2;$$

$$\gamma_2 = \phi_1 \gamma_1;$$

$$\vdots$$

$$\gamma_h = \phi_1^{h-1} \gamma_1.$$

Para estimar os modelos ARMA, é preciso que sejam “estáveis”, significando que as raízes das equações características dos processos AR e MA estejam simultaneamente fora do círculo unitário. Outra maneira de formalizar isso é dada a seguir.

Definição 2.8 Considere o modelo $\alpha(L)y_t = \epsilon_t$, em que $\alpha(L) = (1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p)$. O processo é dito estável se

$$\alpha(z) \neq 0$$

para todo número complexo, z , satisfazendo $|z| \leq 1$.

A definição é bastante difícil, porém extremamente precisa. Para descomplicar, considere os seguintes exemplo e contra-exemplo.

Exemplo 2.10

Seja um AR (1): $y_t = \phi y_{t-1} + \epsilon_t$. Se ocorrer $|\phi| < 1$, então também sucede que

$$1 - \phi z \neq 0$$

para todo complexo z , tal que $|z| \leq 1$. Ou seja, tentativamente considere $\phi = 0,8$ e $z = 0,8 < 1$, então $1 - 0,64 \neq 0$. De fato, para qualquer valor de z , tal que $|z| \leq 1$, o resultado é diferente de zero.

Agora, suponha que $\phi = 1,25$ e $z = 0,8 < 1$. Nesse caso, $1 - 1,25 \times 0,8 = 0$, contrariando a definição. Portanto, a definição implica que as raízes da equação característica estão fora do círculo unitário.

No caso do AR (1), é bem fácil ver que a definição será sempre satisfeita, desde que $|\phi| < 1$. O problema fica mais complicado quando o processo autorregressivo for de ordem maior; entretanto, mesmo assim, a definição ainda se mantém.

2.11**Função Geradora de Autocovariâncias**

Existe uma maneira fácil de encontrar as autocovariâncias, $\{\gamma_j\}_{j=-\infty}^{\infty}$, quando elas são absolutamente somáveis – o caso que geralmente se deseja trabalhar. O truque é usar a função geradora de autocovariâncias definida por:

$$g_y(z) = \sigma^2 \psi(z) \psi(z^{-1}),$$

em que z é um escalar complexo.

As autocovariâncias são representadas pelos coeficientes de z^j , em que j é o expoente de z e indica a ordem de defasagem da autocovariância. Os exemplos a seguir devem esclarecer o ponto.

Exemplo 2.11

Seja um MA (1), tal que $\psi(L) = 1 + \theta L$. Logo:

$$\begin{aligned} g_y(z) &= \sigma^2 (1 + \theta z) (1 + \theta z^{-1}) \\ &= \sigma^2 \left[\theta z^{-1} + (1 + \theta^2) z^0 + \theta z \right]. \end{aligned}$$

Observe que em um MA (1):

$$\begin{aligned} \gamma_0 &= (1 + \theta^2) \sigma^2; \quad \gamma_1 = \theta \sigma^2; \\ \gamma_j &= 0, \quad j > 1. \end{aligned}$$

Exemplo 2.12

Seja um AR (1), tal que $\psi(L) = \frac{1}{1-\phi L}$, então:

$$\begin{aligned} g_y(z) &= \frac{\sigma^2}{(1-\phi z)(1-\phi z^{-1})} \\ &= \sigma^2 \left(1 + \phi z + \phi^2 z^2 + \dots\right) \left(1 + \phi z^{-1} + \phi^2 z^{-2} + \dots\right). \end{aligned}$$

O coeficiente de z^j é:

$$\text{para } j = 0: \quad \sigma^2 \left(1 + \phi^2 + \phi^4 + \dots\right) = \frac{\sigma^2}{1-\phi^2};$$

$$\text{para } j = 1: \quad \sigma^2 \left(\phi + \phi^3 + \phi^5 + \dots\right) = \frac{\phi\sigma^2}{1-\phi^2};$$

$$\vdots$$

$$\text{para } j = h: \quad \sigma^2 \left(\phi^h + \phi^{h+2} + \phi^{h+4} + \dots\right) = \frac{\sigma^2 \phi^h}{1-\phi^2}.$$

2.12

Filtros

É comum proceder a algumas transformações em séries econômicas. Isso é chamado filtragem. Ou seja, é a idéia de executar algum ajuste prévio na série antes de trabalhar com ela, significando passar um filtro na série. Por exemplo, deseja-se diferenciá-la de modo a estacionarizá-la. A função geradora de autocovariância poderá ser útil para encontrar as autocovariâncias da série filtrada. Considere o modelo:

$$y_t = (1 + \theta L) \epsilon_t.$$

Suponha que se transforme y da seguinte forma:

$$x_t = (1 - L) y_t = \Delta y_t.$$

Substituindo a primeira equação nesta última, obtém-se o seguinte MA (2):

$$x_t = (1 - L)(1 + \theta L) \epsilon_t.$$

A função geradora de autocovariância de x pode ser caracterizada assim:

$$\begin{aligned} g_x(z) &= \sigma^2 (1-z)(1+\theta z) \left(1+z^{-1}\right) \left(1+\theta z^{-1}\right) \\ &= (1-z) \left(1+z^{-1}\right) g_y(z). \end{aligned}$$

Generalizando, é cabível passar um filtro $h(L)$ na série y , tal que:

$$x_t = h(L) y_t,$$

em que

$$h(L) = \sum_{j=-\infty}^{\infty} h_j L^j;$$

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

Então, é fácil ver que:

$$g_x(z) = h(z) h(z^{-1}) g_y(z).$$

2.13

Invertibilidade

Similar à idéia de convergência das equações a diferenças, os processos de médias móveis devem ser invertíveis. Isso significa escrever um $MA(q)$ como um $AR(\infty)$ se certas condições forem satisfeitas. Seja um processo $MA(1)$:

$$y_t = \mu + \epsilon_t + \theta \epsilon_{t-1} = (1 + \theta L) \epsilon_t.$$

Se $|\theta| < 1$, pode-se reescrever esse processo da seguinte forma:

$$\begin{aligned} (y_t - \mu) (1 + \theta L)^{-1} &= \epsilon_t \Rightarrow \\ (y_t - \mu) \left(1 - \theta L + \theta^2 L^2 - \dots \right) &= \epsilon_t \Rightarrow \\ y_t - \mu &= \sum_{j=1}^{\infty} (-\theta)^j (y_{t-j} - \mu) + \epsilon_t. \end{aligned}$$

No caso do $MA(q)$, haverá invertibilidade se as raízes características da polinomial $(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)$ estiverem fora do círculo unitário. A razão de impor as condições de invertibilidade em um processo de médias móveis ficará mais clara no Capítulo 3. Resumidamente, são condições necessárias a dois propósitos principais: primeiro, sem invertibilidade, a série não poderia ser estimada recursivamente, usando observações passadas; segundo, para haver unicidade de resultados.⁶

⁶ Um terceiro propósito seria para gerar a função de autocorrelação parcial, cuja definição será dada mais tarde.

Para o primeiro ponto, considere o modelo anterior, admitindo que $|\theta| > 1$. Nesse caso, pode-se inverter a equação usando alguns truques.

$$(y_t - \mu)(1 + \theta L)^{-1} = \epsilon_t \implies \left(\frac{y_t - \mu}{1 + \theta^{-1}L^{-1}} \right) (\theta^{-1}L^{-1}) = \epsilon_t.$$

Agora, o denominador da fração é do tipo que permite a inversão para uma progressão geométrica, pois $|\theta^{-1}| < 1$. Logo, tem-se que:

$$\begin{aligned} \theta \epsilon_{t+1} &= (y_t - \mu) \left(1 - \theta^{-1}L^{-1} + \theta^{-2}L^{-2} - \theta^{-3}L^{-3} + \dots \right) \implies \\ y_t - \mu &= \sum_{j=1}^{\infty} (-\theta)^{-j} (y_{t+j} - \mu) + \theta \epsilon_{t+1}. \end{aligned}$$

Como se vê, y_t depende das observações futuras. Logo, o modelo não pode ser estimado se as raízes do processo MA estiverem dentro do círculo unitário.

Para entender o segundo ponto, considere inicialmente um processo $y_t^a = (1 + \theta L) \epsilon_t$. Esse processo possui:

$$\begin{aligned} \text{var}(y_t^a) &= (1 + \theta^2) \sigma^2; \\ \rho(1) &= \frac{\theta}{1 + \theta^2} = \frac{1}{\frac{1}{\theta} + \theta}. \end{aligned}$$

Considere agora o processo $y_t^b = (1 + \theta^{-1}L) \epsilon_t$, $\epsilon_t \sim RB(0, \theta^2 \sigma^2)$. Esse processo possui:

$$\begin{aligned} \text{var}(y_t^b) &= \left(1 + \frac{1}{\theta^2} \right) \theta^2 \sigma^2 = (1 + \theta^2) \sigma^2 = \text{var}(y_t^a); \\ \rho(1) &= \frac{\frac{1}{\theta}}{1 + \frac{1}{\theta^2}} = \frac{1}{\frac{1}{\theta} + \theta}. \end{aligned}$$

A conclusão é que as séries y^a e y^b são indistinguíveis.

Equivalentemente, em termos da função geradora de autocovariância, tem-se

$$\begin{aligned} g_y(z) &= \tilde{\sigma}^2 \left(1 + \frac{1}{\theta} z^{-1}\right) \left(1 + \frac{1}{\theta} z\right) \\ &= \sigma^2 (1 + \theta z^{-1}) (1 + \theta z), \end{aligned}$$

desde que se defina

$$\sigma^2 = \frac{\tilde{\sigma}^2}{\theta^2}.$$

Ou seja, as funções de autocovariância são idênticas para um processo $MA(1)$ em que $y_t = \epsilon_t + \theta\epsilon_{t-1}$ e $y_t = \epsilon_t + \frac{1}{\theta}\epsilon_{t-1}$. Isso pode gerar confusão e significa que um processo $MA(1)$ com raízes fora do círculo unitário teria os mesmos momentos que outro com raízes dentro do círculo unitário. Para evitar esse problema, impõem-se as raízes sempre fora do círculo unitário⁷.

⁷ Em princípio, poder-se-ia imaginar um processo $MA(q)$ em que algumas de suas raízes estariam fora do círculo unitário e outras, dentro. Tal processo poderia ser tornado totalmente invertível, desde que se ajustasse a variância do processo adequadamente. Para detalhes, veja Hamilton (1994, Capítulo 3) e Hansen e Sargent (1981, p. 102).

3

Processos Estacionários

Este capítulo aborda a modelagem de Box, Jenkins e Reinsel (1994) para séries temporais univariadas estacionárias. Se uma série for considerada não estacionária, deve ser diferenciada. Caso contrário, pode-se usar a série original, estacionária, e aplicar a metodologia deste capítulo para modelá-la e proceder às inferências e previsões. A modelagem de séries temporais univariadas consiste nos seguintes passos:

1. Identificar as ordens p e q do modelo.
2. Estimar o modelo.
3. Verificar se os resíduos estimados não rejeitam a hipótese nula de que são um ruído branco. Se não rejeitam, passa-se ao próximo passo. Se rejeitam, retorna-se ao primeiro passo.
4. Prever.

Como se verá, nem sempre é fácil identificar as ordens do modelo univariado, razão pela qual podem-se cometer equívocos e encontrar resíduos indesejados. É preciso cuidado na análise e, freqüentemente, paciência.

Nesta introdução à análise de processos estacionários, é importante observar que qualquer processo estacionário, mesmo não sendo linear, tem uma representação linear. Isso é muito poderoso, pois implica que se pode decompor um

processo estacionário qualquer em dois componentes lineares, um determinístico e um estocástico. Trata-se do teorema de Wold, formalmente enunciado a seguir, de acordo com Priestley (1981) e Perron (1990).

Teorema 3.1 *Considere um processo estacionário qualquer, y_t . Tal processo pode ser representado por dois processos mutuamente não correlacionados, um puramente determinístico, outro puramente estocástico e que pode ser escrito como um MA (∞):*

$$y_t = u_t + d_t,$$

em que

1. u_t e d_t não são correlacionados;
2. u_t é regular, sendo representado por

$$u_t = \sum_{s=0}^{\infty} \psi_s \epsilon_{t-s},$$

$$\psi_0 = 1, \quad \sum_{s=0}^{\infty} \psi_s^2 < \infty, \quad \epsilon_t \sim RB(0, \sigma^2),$$

sendo $E(\epsilon_s d_t) = 0, \forall s, t$. Além disso, a sequência $\{\psi_s\}_{s=0}^{\infty}$ e o processo $\{\epsilon_t\}$ são unicamente determinados.

3. d_t é singular no sentido de que pode ser previsto a partir do seu próprio passado com variância de predição nula.

Prova

Por se tratar de um importante teorema, segue uma prova¹ simplificada nesta seção². Primeiro defina I_t como o espaço linear gerado pela combinação de y_t, y_{t-1}, \dots , de tal modo que I_t é uma sequência informacional crescente, tal que $I_t \supset I_{t-1}$. Além disso, seja $\text{proj}[y_t | I_{t-1}]$ a projeção ortogonal de y_t sobre o espaço gerado por I_{t-1} e defina o erro de previsão um passo à frente como:

$$c_t = y_t - \text{proj}[y_t | I_{t-1}].$$

¹ Esta prova pode ser pulada em uma primeira leitura.

² É conveniente ver também a prova de Sargent (1987, p. 286-288).

Dessa equação, claramente:

$$\begin{aligned} E(\epsilon_t | I_{t-1}) &= E(y_t - \text{proj}[y_t | I_{t-1}] | I_{t-1}) = 0 \\ &= \text{proj}[y_t | I_{t-1}] - \text{proj}[y_t | I_{t-1}] = 0. \end{aligned}$$

Em consequência: ϵ_t é independente de I_{t-1} , isto é, ϵ_t é ortogonal a I_{t-1} , $\epsilon_t \perp I_{t-1}$. Similarmente, isso é válido para todo $s < t-1$. De fato, considere qualquer par $\{\epsilon_s, \epsilon_t\}$, $s < t$. Nesse caso, $\epsilon_t \perp I_s \subseteq I_{t-1}$ e $\epsilon_s \in I_s$, resultando $\epsilon_t \perp \epsilon_s$, de modo que $E(\epsilon_t \epsilon_s) = 0$, e, portanto, ϵ_t é um ruído branco. Fica claro, a partir disso, que I_t é um subespaço gerado pelos ϵ^t s:

$$I_t = I_{t-1} \oplus \{\epsilon_t\} \implies I_t = \sum_{s=0}^m \{\epsilon_{t-s}\} \oplus I_{t-m-1} \implies I_t = \sum_{s=0}^{\infty} \{\epsilon_{t-s}\} \oplus I_{-\infty},$$

em que

$$\begin{aligned} \{\epsilon_t\} &\text{ é um espaço 1-dimensional gerado por } \epsilon_t \text{ e} \\ I_{-\infty} &\equiv \lim_{m \rightarrow \infty} I_{t-m-1}. \end{aligned}$$

Seja $I_t(\epsilon)$ o subespaço gerado por $\epsilon_t, \epsilon_{t-1}, \dots$ e considere u_t a projeção de y_t sobre $I_t(\epsilon)$, então u_t deve ser a combinação linear de $\epsilon_t, \epsilon_{t-1}, \dots$ e, portanto, $u_t = \sum_{s=0}^{\infty} \psi_s \epsilon_{t-s}$. Assim:

$$E(u_t \epsilon_{t-s}) = \psi_s E(\epsilon_{t-s}^2).$$

Como $y_t = u_t + y_t - u_t$, então:

$$E(u_t \epsilon_{t-s}) = E(y_t \epsilon_{t-s}) - E[(y_t - u_t) \epsilon_{t-s}] = E(y_t \epsilon_{t-s}),$$

pois o termo $(y_t - u_t)$ é ortogonal ao espaço que contém ϵ_{t-s} .

Logo:

$$\psi_s = \frac{E(y_t \epsilon_{t-s})}{E(\epsilon_{t-s}^2)}.$$

De fato, para $s = 0$ e como $y_t = \text{proj}[y_t | I_{t-1}] + \epsilon_t$, $\psi_0 = 1$.

Escrevendo $d_t = y_t - u_t$, verifica-se que:

1. Para $s \leq t$, $E(d_t \epsilon_s) = 0$.
2. Para $s > t$, $v_t \in I_t$. Assim, como $\epsilon_s \perp I_t$, então $\epsilon_s \perp d_t$, de modo que $E(\epsilon_s d_t) = 0$.

Disso conclui-se que $u_s \perp d_t$ para todo s, t . Se u_t e d_t são ortogonais, então:

$$\text{Var}(u_t) + \text{Var}(d_t) = \text{Var}(y_t) < \infty, \text{ porque } y_t \text{ é estacionário.}$$

Então:

$$\sum_{s=0}^{\infty} \psi_s^2 E(\epsilon_{t-s})^2 \leq \text{var}(y_t) < \infty.$$

Resta provar que d_t é singular. Primeiro note que $d_t \in I_t \equiv I_{t-1} \oplus \{\epsilon_t\}$. Como $\epsilon_t \perp d_t$, é forçoso concluir que $d_t \in I_{t-1}$. Igualmente, $d_t \perp \epsilon_{t-1}$ e assim sucessivamente, já que $d_t \perp \epsilon_s, \forall s, t$:

$$\begin{aligned} I_{t-1} &\equiv I_{t-2} \oplus \{\epsilon_{t-1}\} \implies d_t \in I_{t-2} \\ &\vdots \\ d_t &\in I_{-\infty} = \cap_{-\infty}^{\infty} I_t. \end{aligned}$$

Portanto, d_t é determinado pelo passado remoto de y_t .

O passo final é provar que a projeção de d_t sobre I_{t-1} é a mesma que a projeção de d_t sobre $I_t(d)$, em que $I_t(d)$ é o espaço gerado por d_t, d_{t-1}, \dots . Para isso, observe que:

$$I_{t-1} = I_{t-1}(d) \oplus I_{t-1}(\epsilon).$$

De novo, como $d_t \perp \epsilon_s, \forall s, t$, $d_t \in I_{t-1}(d)$ e, portanto, $I_t(d) = I_{t-1}(d)$, para todo t e $d_t \in I_{t-1}(d), s \geq 0$. Logo,

$$d_t \in I_{-\infty}(d) = \cap_{-\infty}^{\infty} I_t(d).$$

Conclui-se que d_t é determinado por seu próprio passado remoto e pode ser previsto com variância nula.

Nota 3.1 Embora a decomposição seja linear mesmo para processos não-lineares, o que significa que se pode aproximar um processo não-linear por um linear, é claro que o processo não-linear tende a aproximar melhor o processo estacionário.

3.1

Função de Autocorrelação – FAC

A função de autocorrelação é o gráfico da autocorrelação contra a defasagem. A função de autocorrelação permitirá identificar a ordem q de um processo MA. Veja o porquê, lembrando as funções de autocorrelação de alguns processos já conhecidos:

Modelo	FAC
$MA(q)$	$\rho_j = \frac{\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \dots + \theta_q\theta_{q-j}}{\sum_{j=0}^q \theta_j^2}, \quad j = 1, 2, \dots, q$
$AR(1)$	$\rho_j = \phi^j, \quad j = 1, 2, \dots$
$AR(p)$	$\rho_j = \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \dots + \phi_p\rho_{j-p}, \quad j = 1, 2, \dots$
$ARMA(1,1)$	$\begin{cases} \rho_1 = \frac{(1 + \phi_1\theta_1)(\phi_1 + \theta_1)}{1 + \theta_1^2 + 2\phi_1\theta_1} \\ \rho_j = \phi_1\rho_{j-1} = \phi_1^{j-1}\rho_1, \quad j > 1. \end{cases}$

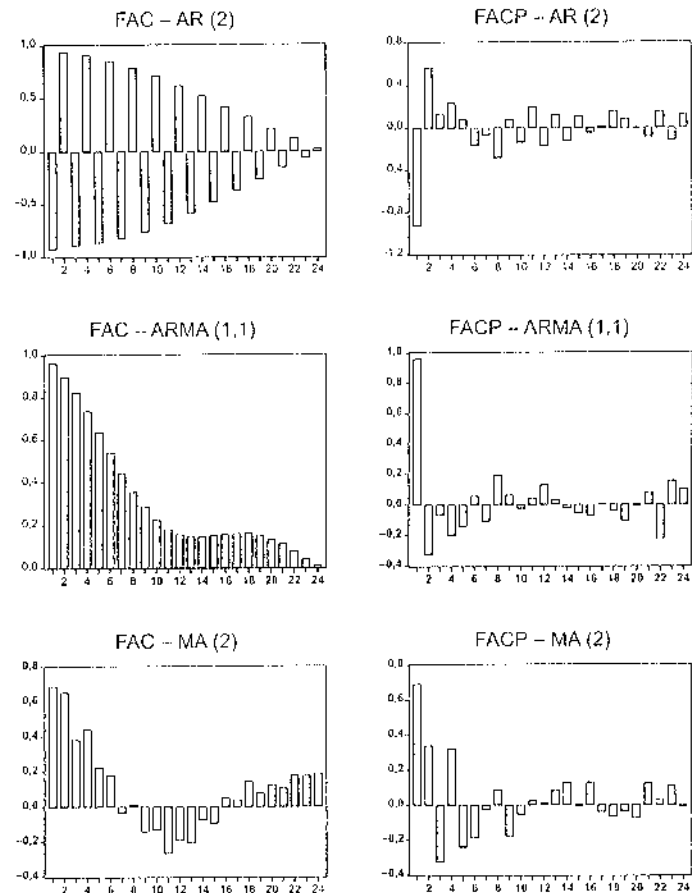


Figura 3.1 $MA(2)$: $\theta_1 = 0,8$, $\theta_2 = 0,9$; $AR(2)$: $\phi_1 = -0,5$, $\phi_2 = 0,4$; $ARMA(1,1)$: $\phi = \theta = 0,9$.

A tabela anterior mostra que a autocorrelação do *MA* torna-se zero a partir da defasagem q , enquanto a dos outros modelos decai exponencialmente. A Figura 3.1 confirma esse padrão. O modelo *ARMA*, por sua vez, passa a decair geometricamente a partir da defasagem q . Essas características permitem determinar padrões sobre as funções de autocorrelação e autocorrelação parcial que ajudam a definir as ordens p e q de um modelo univariado.

Esses padrões, todavia, nem sempre são muito cristalinos, por serem estocásticos. Uma mesma série pode sugerir padrões alternativos de modelagem, em razão da dificuldade de inferir qual é o padrão gerador daquela função. Além disso, dependendo da magnitude dos coeficientes de $\phi(L)$ e $\theta(L)$, o padrão que emerge na FAC é diferente do padrão do verdadeiro processo gerador de dados.

O que importa, na prática, é extrair o máximo de informações possível da série, a ponto de gerar um resíduo que seja estatisticamente um ruído branco. Essa seria a melhor modelagem possível a ser obtida a partir dos dados observados. Supostamente, essa modelagem levaria às melhores previsões estatísticas.

Convém pontuar, entretanto, que a informação do econometrista normalmente é diferente da informação disponível ao agente na sua tomada de decisão. Como resultado, pode ser impossível modelar corretamente uma série, ou a modelagem ideal é, na verdade, um resultado estatístico artificial.

3.2

Função de Autocorrelação Parcial – FACP

Em um modelo *AR* (1), existe uma correlação implícita entre y_t e y_{t-2} . Isso está presente na FAC, por meio do decaimento exponencial. Entretanto, podem-se filtrar as correlações, de forma a manter apenas a correlação pura entre duas observações. Esse processo de filtragem implica gerar a função de autocorrelação parcial – FACP –, pela qual eliminam-se as correlações implícitas entre duas observações. Dessa forma, em um *AR* (1), a autocorrelação parcial entre y_t e y_{t-2} desaparece.

Formalmente, a função de autocorrelação parcial é o gráfico de $\hat{\phi}_{j,j}$ contra j estimado a partir das seguintes regressões em que a série original tem sua média subtraída:

$$y_t = \phi_{j,1}y_{t-1} + \phi_{j,2}y_{t-2} + \dots + \phi_{j,j}y_{t-j} + e_t, \quad j = 1, 2, \dots,$$

em que e_t é um erro.

Em outras palavras, o procedimento consiste em regredir y_t contra y_{t-1} e obter $\hat{\phi}_{1,1}$. Em seguida, deve-se regredir y_t contra y_{t-1} e y_{t-2} . Serão obtidos os coeficientes $\hat{\phi}_{2,1}$ e $\hat{\phi}_{2,2}$, dos quais interessa apenas este último; e assim por diante.

A partir de agora, é importante imaginar teoricamente o que deve acontecer. Se a análise recair, por exemplo, sobre um modelo $AR(2)$, os coeficientes obtidos a partir de $j > 2$ deverão ser iguais a zero. Genericamente, em um $AR(p)$, serão encontrados coeficientes diferentes de zero, até $\hat{\phi}_{p,p}$, e estatisticamente iguais a zero, a partir de então.

Em um $MA(q)$, dada a condição de invertibilidade que torna esse processo um $AR(\infty)$, pode-se mostrar que os coeficientes $\hat{\phi}_{j,j}$ decaem exponencialmente. Quando se tem um modelo $ARMA(p,q)$, há decaimento exponencial a partir da defasagem p .

Nota 3.2 Na prática, Enders (2004) sugere calcular a função de autocorrelação parcial até $j = T/4$, em que T é o tamanho da amostra.

3.3 Identificação

3.3.1 FAC, FACP e LJUNG-BOX

A partir das idéias anteriores, podem-se identificar mais facilmente os modelos $AR(p)$ e $MA(q)$ e, com dificuldades, o modelo $ARMA(p,q)$. A FAC define a defasagem do MA . A FACP define a defasagem do AR . No primeiro caso, sabe-se que a função de autocorrelação decai exponencialmente com o aumento de defasagens, e a função de autocorrelação parcial é truncada a partir da defasagem p . No segundo caso, ocorre o inverso. A função de autocorrelação é truncada na defasagem q , e a função de autocorrelação parcial decai exponencialmente. No caso de uma $ARMA(p,q)$, ambas as funções decaem exponencialmente a partir da defasagem de truncagem. Essa defasagem é, freqüentemente, difícil de reconhecer visualmente. A FAC começará a decair a partir da defasagem q , e a FACP, a partir da defasagem p . O quadro a seguir resume as condições apresentadas:

Modelo	FAC	FACP
$AR(p)$	Decai exponencialmente	Truncada na defasagem p
$MA(q)$	Truncada na defasagem q	Decai exponencialmente
$ARMA(p,q)$	Decai exponencialmente se $j > q$	Decai exponencialmente se $j > p$

Nota 3.3 Visualmente, é difícil identificar quando se inicia o decaimento exponencial.

Nota 3.4 Tsay (2005) propõe uma metodologia para identificar modelos ARMA (p, q) , entretanto parece ser uma metodologia muito trabalhosa vis-à-vis o processo de verificação de resíduos a ser discutido adiante.

Outro problema surge para estimar os valores da variância da autocorrelação e da autocorrelação parcial, porque dependem dos verdadeiros valores dos coeficientes, os quais são desconhecidos.

Do ponto de vista prático, as estimativas dos coeficientes são obtidas conforme o procedimento a seguir:

1. Obtenha a média amostral da série y_t :

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

2. Em seguida, calcule a autocorrelação amostral,

$$\hat{\rho}_j : \hat{\rho}_j = \frac{\frac{\sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y})}{T}}{\frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T}}, \quad j = 1, 2, \dots,$$

em que o acento circunflexo indica parâmetro estimado³.

3. Depois trace o gráfico de $\hat{\rho}_j$ contra j .
4. Em grandes amostras, a variância das estimativas para um ruído branco pode ser *aproximada*⁴ por T^{-1} . Esse é o referencial básico para traçar o intervalo de confiança, já que o objetivo é obter resíduos que são estatisticamente um ruído branco.
5. Se a série for um MA (q) , então a variância das estimativas pode ser aproximada por $T^{-1}(1 + 2 \sum_{s=1}^{j-1} \rho_s^2)$, para $j > q$. Se os valores de $\hat{\rho}_j$ forem estatisticamente diferentes de zero, ou seja, se a hipótese nula $H_0 : \rho_j = 0$ contra $H_1 : \rho_j \neq 0$ é rejeitada, considera-se $\hat{\rho}_j$ diferente de zero. Há *softwares*

³ A razão de se dividir a autocovariância por T e não por $T - j$, como se esperaria, foge ao escopo deste texto. Pode-se mostrar que o erro quadrático médio é menor dividindo-se por T .

⁴ Também conhecida como aproximação de Bartlett.

que calculam esse intervalo de confiança para ajudar na identificação da FACP, aproximando ρ_s por $\hat{\rho}_s$. Outros pacotes simplificam e usam o intervalo do ruído branco mesmo, em razão da indisponibilidade de ρ_s . No caso, é importante entender que a rejeição da nula sob o intervalo de um ruído branco poderia ser falsa se o intervalo correto fosse calculado.

6. Para a FACP, a variância de $\hat{\phi}_{i,j}$ também é dada por T^{-1} para um $AR(p)$, quando $j > p$. Assim, se a estimativa estiver entre dois desvios-padrão, $\pm 2T^{-\frac{1}{2}}$, não se rejeita a hipótese nula de coeficiente igual a zero.

Nota 3.5 *Necessita-se dos verdadeiros valores de ρ_j para calcular o intervalo de confiança. Por isso, o procedimento descrito para a identificação de modelos ARMA deve ser usado apenas como uma regra de bolso. Na verdade, as variâncias das estimativas de autocorrelação dependem umas das outras, de modo que há uma contaminação no intervalo de confiança de cada uma dessas estimativas [veja Priestley (1981) para uma discussão sobre isso, particularmente o Capítulo 5].*

Em outras palavras, a aproximação de Bartlett para amostras limitadas não é muito boa. O mesmo ocorre com a função de autocorrelação parcial.

Uma maneira melhor de verificar se há ordens maiores no modelo ARMA é usando a estatística de Ljung-Box, cujas hipóteses nula e alternativa são mais bem inferidas dessa forma, porque as autocorrelações são conjuntamente identificadas.

Sejam a hipótese nula dada por $H_0 : \sum_{j=1}^n \rho_j = 0$ e a alternativa dada por $H_1 : \sum_{j=1}^n \rho_j \neq 0$, então pode-se testar essa hipótese usando a seguinte estatística:

$$Q = T(T+2) \sum_{j=1}^n \frac{\hat{\rho}_j^2}{T-j} \xrightarrow{d} \chi_n^2,$$

em que $\xrightarrow{d} \chi_n^2$ indica convergência em distribuição para uma distribuição qui-quadrada com n graus de liberdade⁵.

Ora, como o teste indica, se uma das autocorrelações for diferente de zero, há evidência de existência de um modelo ARMA (p, q) . Se estimado o modelo e os resíduos obtidos não apresentarem mais evidência de autocorrelação usando essa estatística, o modelo estará bem estimado.

⁵ As hipóteses nula e alternativa desse teste poderiam ser formuladas de outra forma, sem alterar a estatística. Nesse caso, a hipótese nula é: $H_0 : \rho_j = 0$, para todo j , contra $H_1 : \rho_j \neq 0$, para algum j .

Nota 3.6 Na prática, identifica-se o modelo por meio da FAC e FACP. Em seguida, usa-se a estatística de Ljung-Box sobre os resíduos estimados para confirmar e reforçar os resultados. Caso os resultados não se confirmem por essa estatística, adicionam-se novas defasagens e repete-se o processo de verificação de resíduos.

3.3.2 Critério de Informação

O critério de informação é uma forma de encontrar o número ideal de parâmetros de um modelo. Para entendê-lo, tenha em mente que, a cada regressor adicional, a soma dos resíduos não vai aumentar e, freqüentemente, irá diminuir. A redução se dá à custa de mais regressores. Para balancear a redução dos erros e o aumento do número de regressores, o critério de informação associa uma penalidade a esse aumento. Se a penalidade for menor que a diminuição da soma de resíduos, o regressor adicional deve ser incorporado ao modelo. Se a penalidade for maior que a diminuição da soma, o regressor adicional traz mais custos que benefícios. A idéia do critério de informação é minimizar uma função baseada nos resíduos, penalizada pelo número de regressores.

É difícil identificar o modelo $ARMA(p, q)$ visualmente. Frequentemente, dois ou mais modelos possíveis geram resíduos cujos testes indicam ser um ruído branco. O melhor modelo será o mais parcimonioso, satisfeito que os resíduos sejam os menores possíveis. Isto é, o modelo mais parcimonioso, portanto com menor número de parâmetros, deverá gerar menos imprecisão de estimativas justamente porque tem menos parâmetros que um modelo com mais parâmetros.

As estatísticas descritas a seguir tratam desse problema e são usadas para escolher o melhor modelo. O critério de especificação tem, em geral, a seguinte forma:

$$C = \ln \hat{\sigma}^2(T) + c_T \varphi(T),$$

em que

$\hat{\sigma}^2(T) = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2$ é a variância estimada dos resíduos;

c_T representa o número de parâmetros estimados;

$\varphi(T)$ é a ordem do processo, que penaliza a falta de parcimônia.

O primeiro termo da equação mede a adequação do processo. É claro que, quanto menor a variância dos resíduos, melhor. Entretanto, tal redução de variância foi obtida à custa da imposição de mais parâmetros ao processo, cuja penalização ocorre na segunda parcela.

Nota 3.7 Note que o número de observações, T , é invariante ao número de parâmetros estimados. Isso significa ser preciso comparar séries com o mesmo número de observações.

Há três principais critérios de informação. A estatística de Schwarz é dada pela seguinte expressão, denotada por BIC (Bayesian Information Criterion) ou SBC (Schwarz Bayesian Criterion):

$$BIC(p, q) = \ln \hat{\sigma}^2 + n \frac{\ln T}{T},$$

em que

$n = p + q$, se o modelo não tem constante, e

$n = p + q + 1$, se há constante no modelo,

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{\epsilon}_t^2}{T}.$$

A estatística de Akaike, denotada por AIC (Akaike Information Criterion), é dada por:

$$AIC(p, q) = \ln \hat{\sigma}^2 + n \frac{2}{T}.$$

Finalmente, a estatística de Hannan-Quinn, HQ , é dada por:

$$HQ(p, q) = \ln \hat{\sigma}^2 + n \frac{2}{T} \ln \ln T.$$

Quanto mais parâmetros são estimados no mesmo período de amostra, menor será o erro estimado, mas isso será penalizado na segunda parcela da estatística. Por isso, deseja-se o menor AIC , HQ ou BIC possível.

Nota 3.8 Modelos com diferentes amostras não podem ser comparados por esses critérios. As amostras devem ser iguais para comparação.

Nota 3.9 Enquanto o critério BIC é consistente assintoticamente, tendendo a escolher um modelo mais parcimonioso que o AIC , AIC funciona melhor em pequenas amostras, não obstante seja viesado para escolher modelos sobreparametrizados. O critério HQ também é assintoticamente consistente, porém é menos forte que o critério BIC .

Nota 3.10 Os resultados valem tanto para processos estacionários quanto para processos integrados (Lütkepohl e Krätzig, 2004).

De forma geral, se $T \geq 16$, então⁶:

$$BIC \leq HQ \leq AIC.$$

Em resumo, os critérios de informação são:

Critério de informação	Definição
<i>AIC</i>	$\ln \hat{\sigma}^2 + n \frac{2}{T}$
<i>BIC</i>	$\ln \hat{\sigma}^2 + n \frac{\ln T}{T}$
<i>HQ</i>	$\ln \hat{\sigma}^2 + n \frac{2}{T} \ln \ln T$

A seguir, são mostrados exemplos das várias possibilidades visuais que se podem encontrar com séries de tempo reais. São exemplos das configurações mais comuns.

3.3.3 Identificação de Modelos AR, MA e ARMA

Exemplo 3.1

A Figura 3.2 apresenta os gráficos de autocorrelação e autocorrelação parcial de um processo MA (1): $y_t = \epsilon_t + 0,4\epsilon_{t-1}$. Os intervalos de confiança são baseados na variância de um ruído branco, portanto trata-se da versão simplificada de cálculo do intervalo.

Tratando-se de um MA (q), é de se notar que a função de autocorrelação é de fato truncada na primeira defasagem. A FACP deveria decair exponencialmente. No caso dessa simulação, e de inúmeras séries temporais reais também, a FACP aparece truncada na primeira defasagem, o que pode gerar dúvida se se trata

⁶ Embora usados indistintamente para identificar modelos ARMA(p, q), na verdade, os critérios sugeridos são indicados para modelos auto-regressivos exclusivamente. A razão é que os modelos ARMA(p, q) podem ter as ordens p e q simultaneamente superespecificadas, e estas se cancelariam de um lado e outro. Veja Lütkepohl e Krätzig (2004), que propõem um procedimento alternativo, conhecido como procedimento de Hannan-Rissanen, para identificar mais corretamente o modelo ARMA por meio do critério de informação.

de um modelo ARMA (1,1). A solução da dúvida é simples: estime os dois modelos e tome aquele cujos resíduos são um ruído branco. Se ambos tiverem resíduos que são um ruído branco, escolha o modelo cujo critério de informação é o menor possível.

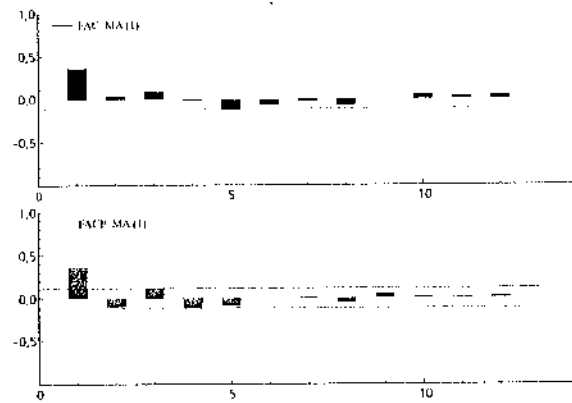


Figura 3.2 Funções de autocorrelação e autocorrelação parcial de um MA (1).

Exemplo 3.2

A Figura 3.3 apresenta os gráficos de autocorrelação e autocorrelação parcial de um processo AR (1): $y_t = 0,6y_{t-1} + \epsilon_t$.

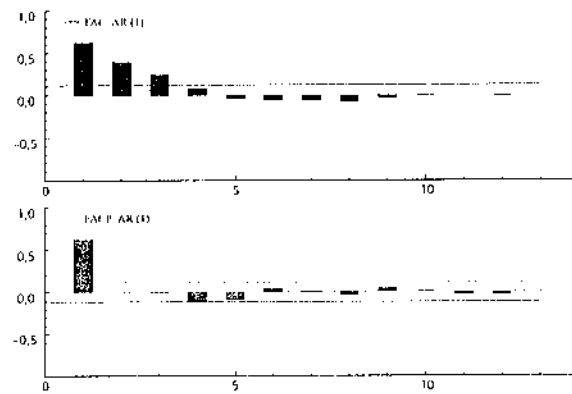


Figura 3.3 Funções de autocorrelação e autocorrelação parcial de um AR (1).

Claramente, verifica-se que a função de autocorrelação parcial é truncada na primeira defasagem e parece haver um decaimento exponencial da autocorrelação. Trata-se de um modelo AR (1).

Exemplo 3.3

A Figura 3.4 apresenta os gráficos de autocorrelação e autocorrelação parcial de um processo MA (2): $y_t = \epsilon_t + 0,4\epsilon_{t-1} + 0,3\epsilon_{t-2}$.

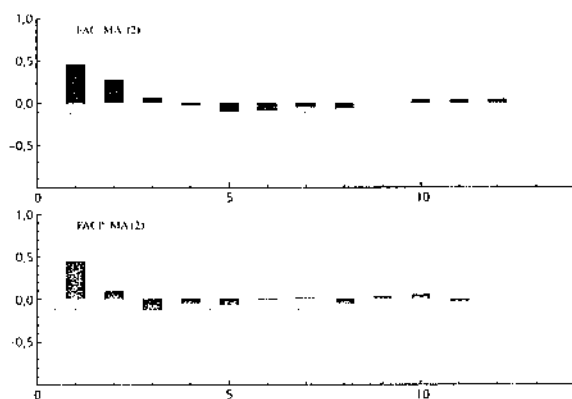


Figura 3.4 Funções de autocorrelação e autocorrelação parcial de um MA(2).

De novo, essa simulação representa um MA (2), mas de difícil identificação, pois pode ser confundido com um ARMA (1, 2) ou mesmo um AR (1), quando se tomam em conjunto a FAC e a FACP. De novo, o procedimento usual é gerar os resultados de todas as combinações possíveis e tomar aquela cujos resíduos são um ruído branco e com o menor valor do critério de informação.

Exemplo 3.4

A Figura 3.5 representa a seguinte especificação de um modelo AR (2): $y_t = 0,6y_{t-1} + 0,4y_{t-2} + \epsilon_t$.

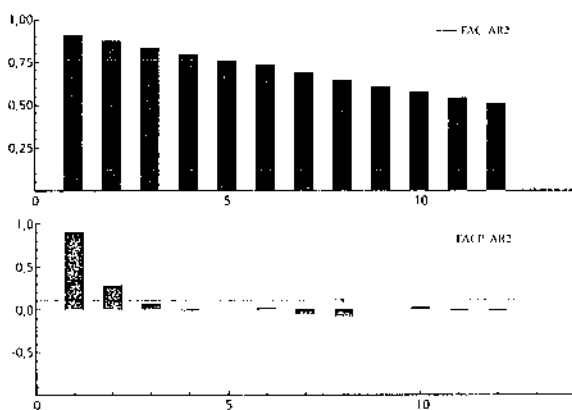


Figura 3.5 Funções de autocorrelação e autocorrelação parcial de um AR (2).

O decaimento da FAC é lento, sugerindo que os coeficientes do processo somam 1. Não obstante, as raízes estão fora do círculo unitário, sugerindo um processo estacionário.

Exemplo 3.5

A Figura 3.6 representa a FAC e a FACP de um ARMA (1,1), com a seguinte especificação: $y_t = 0,6y_{t-1} + \epsilon_t + 0,3\epsilon_{t-1}$.

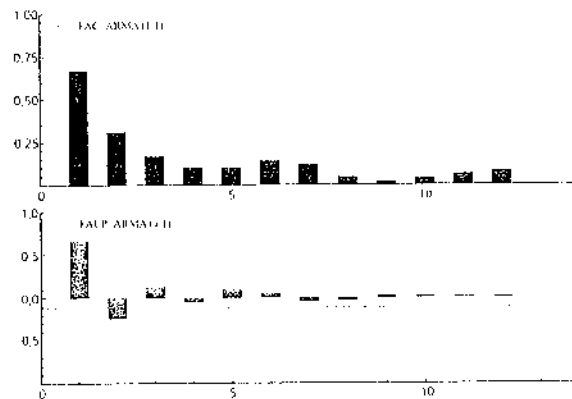


Figura 3.6 Funções de autocorrelação e autocorrelação parcial de um ARMA (1,1).

Em ambas as funções, há decaimento exponencial. É muito difícil identificar onde eles se iniciam. A regra prática é a seguinte: avalie o critério de informação de todas as combinações possíveis de modelos a partir do ponto em que os coeficientes da FAC e FACP tornam-se estatisticamente insignificantes. Tome o modelo com o menor valor. No exemplo, define-se $\text{ARMA}(p_{\max}, q_{\max}) = \text{ARMA}(5,7)$, cuja especificação é justificável se houver interpretação econômica. Se o pesquisador for um pouco mais liberal, poderá ajustar $\text{ARMA}(p_{\max}, q_{\max}) = \text{ARMA}(3,3)$. Nesse último caso, haveria 16 combinações a considerar, pois $p = 0, 1, 2, 3$ e $q = 0, 1, 2, 3$.

Exemplo 3.6

Considere a taxa mensal de inflação medida pelo IPCA. A Figura 3.7 apresenta o gráfico de autocorrelação e autocorrelação parcial dessa série.

Como pode ser notado, a autocorrelação decai exponencialmente, indicando que a série possui componente de média móvel igual a zero. A autocorrelação parcial é truncada na primeira defasagem, indicando um processo AR (1). A interpretação econômica para esse resultado é que apenas a inflação do mês passado possui efeito significativo sobre a inflação corrente. A taxa de inflação

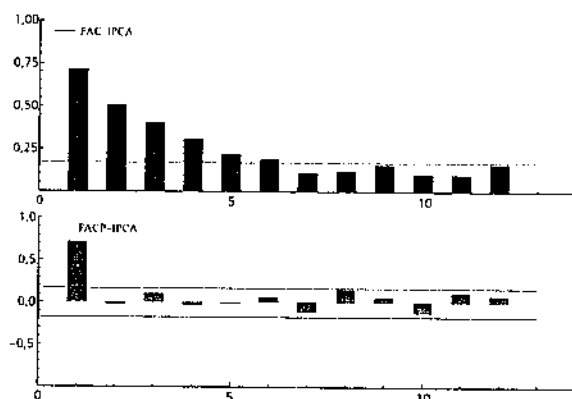


Figura 3.7 Funções de autocorrelação e autocorrelação parcial do IPCA.

em meses anteriores ($t - 2, t - 3, \dots$) não possui efeitos sobre a inflação corrente. Trata-se de um resultado esperado, pois a literatura sobre o assunto sempre supõe algum tipo de inercialidade sobre a inflação. O IPCA será usado na próxima seção para entender o processo de estimação.

Em resumo, define-se a máxima ordem p e q do modelo na defasagem referente à última hipótese nula rejeitada nos gráficos das funções de autocorrelação e autocorrelação parcial. Depois, estimam-se todas as combinações de modelos possíveis, de forma a identificar o modelo *minimizando* a estatística de informação ou critério de especificação. Por trás dessa idéia, está a de buscar o modelo mais parcimonioso possível, que, em geral, gera melhores previsões, segundo Box, Jenkins e Reinsel (1994).

3.4

Estimação Condicional

Esta seção mostra como estimar os modelos univariados estacionários usando o método de máxima verossimilhança. Inicia-se com o método de máxima verossimilhança condicional. Trata-se de um método que é assintoticamente equivalente ao da máxima verossimilhança exata, a ser visto depois. A vantagem do primeiro método é a facilidade de especificar e, conseqüentemente, de estimar a função de verossimilhança. O problema é não ser tão eficiente quanto o método exato, sobretudo para pequenas amostras. Em geral, o pesquisador assume distribuição normal ou t -student e procura estimar o vetor de parâmetros $\Psi = (c, \phi_1, \phi_2, \dots, \phi_p; \theta_1, \theta_2, \dots, \theta_q)$ que maximiza a probabilidade de a amostra pertencer à distribuição especificada. A seguir, são estudados os casos principais, assumindo distribuição normal, segundo Hamilton (1994).

3.4.1 Função de Verossimilhança para um AR (p)

A idéia de estimação condicional para um processo AR (p) é usar as p primeiras observações como valores iniciais para maximizar a função de verossimilhança condicional:

$$\ln f(y_t | y_{t-1}, y_{t-2}, \dots, y_1; \Psi) = -\frac{T-p}{2} \ln(2\pi\sigma^2) - \sum_{t=p+1}^T \frac{(y_t - c - \sum_{i=1}^p \phi_i y_{t-i})^2}{2\sigma^2}.$$

Observe que o vetor de parâmetros dessa função, Ψ , é o mesmo do de mínimos quadrados ordinários. A vantagem do método é a facilidade de estimação para o caso auto-regressivo. O que muda, nesse caso condicional, entre máxima verossimilhança e mínimos quadrados, é a estimativa de σ^2 .

Quando T é suficientemente grande, as p primeiras observações contribuem marginalmente para o total da verossimilhança. Desse modo, a verossimilhança condicional e a exata acabam tendo as mesmas propriedades assintóticas, desde que as raízes de $\phi(L)$ estejam fora do círculo unitário.

Nota 3.11 Máximas verossimilhanças condicional e exata têm as mesmas propriedades assintóticas, assim como o método de mínimos quadrados condicionais, quando as raízes de $\phi(L)$ estão fora do círculo unitário. Do ponto de vista econométrico, essas estimativas são todas consistentes, embora o método de máxima verossimilhança exato seja preferido por ser o mais eficiente.

Na estimação, é importante ter estacionaridade estrita, pois isso garante que certas condições necessárias para inferência sejam satisfeitas. Além disso, embora geralmente assumamos distribuição normal, se a distribuição do processo gerador de dados não for normal, as estimativas ainda assim serão consistentes. A intuição para esse resultado é que, se a estimação por mínimos quadrados, cujas estimativas são as mesmas da maximização da função de verossimilhança, é consistente, o mesmo deve ocorrer com a máxima verossimilhança. No entanto, será preciso corrigir a matriz de covariância estimada. Com a correção, as inferências estatísticas continuam válidas. Trata-se do método de estimação de *quasi-máxima verossimilhança*. White (1982) propõe uma forma de correção da variância dessas estimativas.

O grande problema da estimação, contudo, é quando existe um modelo de médias móveis, já que não se observam os erros, necessários para calcular os parâmetros da média móvel. A próxima seção sugere uma forma de resolver esse problema.

Exemplo 3.7

Usa-se o IPCA para exemplificar o uso da técnica de estimação condicional. O modelo estimado foi o seguinte:

$$\begin{aligned}y_t &= \mu + u_t; \\ u_t &= \phi u_{t-1} + \epsilon_t.\end{aligned}$$

Note que o modelo pode ser sintetizado em um AR (1) da seguinte forma:

$$y_t = \mu + \phi y_{t-1} + \epsilon_t.$$

Como $u_{t-1} = y_{t-1} - \mu$, tem-se:

$$\begin{aligned}y_t &= \mu + \phi (y_{t-1} - \mu) + \epsilon_t \\ &= \mu (1 - \phi) + \phi y_{t-1} + \epsilon_t \\ &= c + \phi y_{t-1} + \epsilon_t.\end{aligned}$$

Como $\mu = c / (1 - \phi)$, é claro que $\mu (1 - \phi) = c$

Alguns softwares, como o Eviews, usam essa especificação para estimar o modelo, porque o resultado gera diretamente a média incondicional, μ , que é um coeficiente de maior interesse que o c .

Por razões que serão aclaradas na seção de diagnóstico de resíduos, a série de IPCA a estimar foi dessazonalizada, usando o método $X - 11$. A análise da série dessazonalizada quanto às funções FAC e FACP não se alterou. O resultado obtido é descrito na equação a seguir:

$$\begin{aligned}\text{IPCA}_t &= 0,612 + u_t; \\ &\quad (0,129) \\ u_t &= 0,801 u_{t-1} + \epsilon_t. \\ &\quad (0,048)\end{aligned}$$

A série original tem 149 observações. Na regressão, foram utilizadas 148 em razão do condicionamento da amostra à primeira observação. A inflação de longo prazo está em torno de 0,61% ao mês. O coeficiente do AR (1) representa o ϕ e indica uma considerável inercialidade, a despeito do Plano Real. O desvio-padrão dos coeficientes, indicado entre parênteses embaixo dos coeficientes estimados, foi ajustado para heterocedasticidade e autocorrelação e indica alta significância.

3.4.2 Função de Verossimilhança para um MA (q)

Considere inicialmente um MA (1), como forma de ganhar alguma intuição com relação a um modelo de ordem $q > 1$:

$$y_t = \epsilon_t + \theta \epsilon_{t-1}.$$

Primeiro, fixam-se os valores iniciais dos erros à sua esperança incondicional. No caso do $MA(1)$, fixa-se $\epsilon_0 = 0$. Trata-se de uma hipótese inicial bem razoável, porque é desejável que os resíduos sejam nulos. Com isso, podem-se obter os erros em função da variável observada:

$$\begin{aligned}\epsilon_1 &= y_1 - \theta\epsilon_0 = y_1; \\ \epsilon_2 &= y_2 - \theta\epsilon_1 = y_2 - \theta y_1; \\ \epsilon_3 &= y_3 - \theta\epsilon_2 = y_3 - \theta(y_2 - \theta y_1); \\ &\vdots \\ \epsilon_t &= \sum_{i=0}^{t-1} (-\theta)^i y_{t-i}.\end{aligned}$$

É claro que essa derivação vale se $|\theta| < 1$, razão pela qual se requereu invertibilidade anteriormente. Se $|\theta|$ não é próximo de 1, esse condicionante terá poucos efeitos práticos, de modo que a função condicional vai ser uma boa aproximação para a incondicional⁷. Assim, a função de verossimilhança condicional a maximizar, supondo distribuição normal, fica:

$$\ln f(y_t, y_{t-1}, \dots, y_1 | \epsilon_0 = 0; \Psi) = -\frac{T}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \left(\sum_{i=0}^{t-1} (-\theta)^i y_{t-i} \right)^2.$$

As derivadas não são exatas e, portanto, ela é maximizada, usando-se métodos numéricos. O método de mínimos quadrados não-lineares também poderia ser utilizado, observada a diferença de estimativa sobre o desvio-padrão, σ .

Generalizando, pode-se estimar um $MA(q)$ fixando os valores iniciais por meio da esperança incondicional:

$$\epsilon_0 = \epsilon_{-1} = \dots = \epsilon_{-q+1} = 0.$$

Definindo $\epsilon_0 = (\epsilon_0, \epsilon_{-1}, \dots, \epsilon_{-q+1})$ e assumindo invertibilidade, a função de verossimilhança condicional é dada por:

$$\ln f(y_t, y_{t-1}, \dots, y_1 | \epsilon_0 = 0; \Psi) = -\frac{T}{2} \ln 2\pi\sigma^2 - \sum_{t=1}^T \frac{\epsilon_t^2}{2\sigma^2}.$$

⁷ Entretanto, quando $|\theta|$ estiver próximo de 1, esse tipo de condicionante não será muito bom, devendo-se preferir a máxima verossimilhança exata. Infelizmente, não é possível saber qual opção utilizar *a priori*.

Novamente o problema do método é quando as raízes da polinomial estão próximas do círculo unitário.

Exemplo 3.8

Continuando o exemplo de IPCA, estima-se um modelo MA (3), suficiente para branquear os resíduos. A seguir, é mostrado o resultado da estimação:

$$IPCA_t = 0,640 + 0,855\epsilon_{t-1} + 0,622\epsilon_{t-2} + 0,256\epsilon_{t-3} + \epsilon_t.$$

(0,074)
(0,077)
(0,091)
(0,078)

3.4.3 Função de Verossimilhança para um ARMA (p,q)

Para o caso de um ARMA (p,q) condicional, assume-se que os erros iniciais sejam nulos, de forma que $\epsilon_0 = \epsilon_{-1} = \dots = \epsilon_{-q+1} = 0$.

O componente auto-regressivo, por sua vez, pode ser condicionado de duas formas. Na primeira, fixa-se $y_0 = y_{-1} = \dots = y_{-p+1} = \bar{y}$, a média temporal da variável que supostamente converge para sua esperança não condicional. Nesse caso, inicia-se a estimação usando a amostra toda, ou seja, as observações de 1 a T. Formalmente, dados os valores iniciais $y_0 \equiv (y_0, y_{-1}, \dots, y_{-p+1})$ e $\epsilon_0 = (\epsilon_0, \epsilon_{-1}, \dots, \epsilon_{-q+1})$, calculam-se os resíduos:

$$\epsilon_t = y_t - c - \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j}, \quad t = 1, 2, \dots, T,$$

e maximiza-se:

$$\ln f(y_T, y_{T-1}, \dots, y_1 | y_0, \epsilon_0; \Psi) = -\frac{T}{2} \ln 2\pi\sigma^2 - \sum_{t=1}^T \frac{\epsilon_t^2}{2\sigma^2}.$$

Outra possibilidade é assumir os valores iniciais de y como os valores observados. Nesse caso, inicia-se a estimação em $t = p + 1$. Essa última alternativa é a recomendada por Box, Jenkins e Reinsel (1994). Dessa forma, fixa-se:

$$\epsilon_p = \epsilon_{p-1} = \dots = \epsilon_{p-q+1} = 0$$

e maximiza-se:

$$\begin{aligned} \ln f(y_T, y_{T-1}, \dots, y_1 | y_p, y_{p-1}, \dots, y_1, \epsilon_p = \epsilon_{p-1} = \dots = \epsilon_{p-q+1} = 0; \Psi) \\ = -\frac{T-p}{2} \ln 2\pi\sigma^2 - \sum_{t=p+1}^T \frac{\epsilon_t^2}{2\sigma^2}. \end{aligned}$$

Numericamente, procede-se à maximização condicionada da seguinte forma.

1. Defina um vetor de parâmetros iniciais Ψ^0 , dados os valores iniciais, (y_0, e_0) .
2. Use esse vetor de parâmetros para construir a série inicial de erros:

$$\tilde{e}_{0,t} = y_t - c^0 - \sum_{i=1}^p \phi_i^0 y_{t-i} - \sum_{j=1}^q \theta_j^0 \tilde{e}_{0,t-j}, \text{ para todo } t.$$

3. Encontre o vetor de parâmetros Ψ^1 que maximiza a verossimilhança que contém os seguintes erros:

$$\epsilon_{1,t} = y_t - c^1 - \sum_{i=1}^p \phi_i^1 y_{t-i} - \sum_{j=1}^q \theta_j^1 \tilde{e}_{0,t-j}.$$

4. Com Ψ^1 , recalcule os erros, dados os valores iniciais, (y_0, e_0) :

$$\tilde{e}_{1,t} = y_t - c^1 - \sum_{i=1}^p \phi_i^1 y_{t-i} - \sum_{j=1}^q \theta_j^1 \tilde{e}_{1,t-j}.$$

5. Em seguida, encontre o vetor de parâmetros Ψ^2 que maximiza a verossimilhança que contém os seguintes erros:

$$\epsilon_{2,t} = y_t - c^2 - \sum_{i=1}^p \phi_i^2 y_{t-i} - \sum_{j=1}^q \theta_j^2 \tilde{e}_{1,t-j}.$$

6. De modo geral, para $n \geq 2$, calcule:

$$\tilde{e}_{n,t} = y_t - c^n - \sum_{i=1}^p \phi_i^{n-1} y_{t-i} - \sum_{j=1}^q \theta_j^n \tilde{e}_{n,t-j}.$$

7. Em seguida, encontre Ψ^{n+1} que maximiza

$$\epsilon_{n+1,t} = y_t - c^{n+1} - \sum_{i=1}^p \phi_i^{n+1} y_{t-i} - \sum_{j=1}^q \theta_j^{n+1} \tilde{e}_{n,t-j}.$$

8. Maximize a verossimilhança que contém os erros $\epsilon_{n,t}$, para encontrar um vetor de parâmetros tal que $\ln f(\cdot; \Psi^{n+1}) > \ln f(\cdot; \Psi^n)$.
9. Volte ao passo 6.
10. Pare quando $|\Psi^{n+1} - \Psi^n| < \epsilon$, em que $|\cdot|$ representa a norma. Geralmente, fixa-se $\epsilon < 10^{-5}$.

3.5

Estimação Exata*

3.5.1 Função de Verossimilhança para um AR (p)

Primeiro, é preciso montar as funções de verossimilhança exata para um processo $AR(p)$, de forma a estimar o vetor de parâmetros. Inicialmente, considere as primeiras p observações do processo $AR(p)$, definido pelo vetor $Y_p = (y_1, y_2, \dots, y_p)'$. É preciso encontrar a função de verossimilhança não condicional para as p primeiras observações. Daí em diante, condicionam-se as observações restantes às anteriores. Procedendo-se assim, obtém-se a função de verossimilhança exata. Essa metodologia é chamada *decomposição do erro de previsão* da função de verossimilhança.

Considere a matriz de covariância, V_p , dessas primeiras p observações:

$$\sigma^2 V_p = \begin{bmatrix} E(y_1 - \mu)^2 & E(y_1 - \mu)(y_2 - \mu) & \cdots & E(y_1 - \mu)(y_p - \mu) \\ E(y_2 - \mu)(y_1 - \mu) & E(y_2 - \mu)^2 & \cdots & E(y_2 - \mu)(y_p - \mu) \\ \vdots & \vdots & \ddots & \vdots \\ E(y_p - \mu)(y_1 - \mu) & E(y_p - \mu)(y_2 - \mu) & \cdots & E(y_p - \mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 \end{bmatrix},$$

em que

$$E(y_t) = \mu = \frac{c}{\phi(1)}; \quad \phi(1) = 1 - \sum_{i=1}^p \phi_i.$$

Dessa forma, a função densidade de probabilidade das primeiras p observações é:

$$\begin{aligned} f(y_p, y_{p-1}, \dots, y_1; \Psi) &= (2\pi)^{-\frac{p}{2}} |\sigma^2 V_p|^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma^2} (Y_p - \mu)' V_p^{-1} (Y_p - \mu) \right] = \\ &= (2\pi\sigma^2)^{-\frac{p}{2}} |V_p|^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma^2} (Y_p - \mu)' V_p^{-1} (Y_p - \mu) \right], \end{aligned}$$

em que μ é um vetor $p \times 1$ com elementos todos iguais a μ .

Para $t > p$, pode-se condicionar a função de verossimilhança aos valores anteriores, conforme representado a seguir:

$$\begin{aligned} f(y_t | y_{t-1}, y_{t-2}, \dots, y_1; \Psi) &= f(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}; \Psi) \\ &= (2\pi\sigma^2)^{-\frac{1}{2}} \exp \left[-\frac{(y_t - c - \sum_{i=1}^p \phi_i y_{t-i})^2}{2\sigma^2} \right]. \end{aligned}$$

A função de verossimilhança com amostra completa é:

$$\begin{aligned} f(y_t | y_{t-1}, y_{t-2}, \dots, y_1; \Psi) &= f(y_p, y_{p-1}, \dots, y_1; \Psi) \\ &\quad \times \prod_{t=p+1}^T f(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}; \Psi). \end{aligned}$$

Dessa forma, a função log-verossimilhança é dada por:

$$\begin{aligned} \ln f(y_t | y_{t-1}, y_{t-2}, \dots, y_1; \Psi) &= \frac{1}{2} \ln |V_p^{-1}| - \frac{1}{2\sigma^2} (Y_p - \mu)' V_p^{-1} (Y_p - \mu) \\ &\quad - \frac{T}{2} \ln (2\pi\sigma^2) - \sum_{t=p+1}^T \frac{(y_t - c - \sum_{i=1}^p \phi_i y_{t-i})^2}{2\sigma^2}. \end{aligned}$$

Essa função tem de ser maximizada numericamente. O poderio computacional atual permite isso, assim como a inversão de V_p , sem maiores problemas.

Nota 3.12 Em vez de considerar apenas as p primeiras observações para construir a matriz de covariância, pode-se proceder usando todas as T observações. Isso seria possível se fosse construída uma matriz de covariância maior. Pode-se mostrar que, assintoticamente, os resultados seriam idênticos.

3.5.2 Função de Verossimilhança para um MA (q)

Para um modelo MA (q), pode-se derivar a função de verossimilhança exata, condicional aos dados e à estrutura do processo. Preliminarmente, é necessário estudar a estrutura do processo. Seguindo Box, Jenkins e Reinsel (1994), convém visualizar as equações do modelo, assumindo, por simplicidade, que $\mu = 0$.

$$\begin{cases} y_1 = \epsilon_1 + \sum_{j=1}^q \theta_j \epsilon_{1-j} \\ y_2 = \epsilon_2 + \sum_{j=1}^q \theta_j \epsilon_{2-j} \\ \vdots \\ y_T = \epsilon_T + \sum_{j=1}^q \theta_j \epsilon_{T-j} \end{cases}$$

Agora defina os vetores $Y \equiv (y_1, y_2, \dots, y_T)'$; $\epsilon_* \equiv (\epsilon_{-(q-1)}, \epsilon_{-(q-2)}, \dots, \epsilon_0)'$ e $a \equiv (\epsilon_1, \epsilon_2, \dots, \epsilon_T)'$. Em seguida, defina a matriz triangular inferior L_θ de dimensão $T \times T$ que contém 1's na diagonal principal, θ_1 na primeira diagonal secundária, θ_2 na segunda diagonal secundária, e assim sucessivamente:

$$L_\theta = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ \theta_1 & 1 & 0 & 0 & 0 & \dots & 0 \\ \theta_2 & \theta_1 & 1 & 0 & 0 & \dots & 0 \\ \theta_3 & \theta_2 & \theta_1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \theta_1 & 1 & 0 \\ 0 & 0 & 0 & \dots & \theta_2 & \theta_1 & 1 \end{bmatrix}$$

Defina a matriz $F_{T \times q} = (B'_{q \times q}, 0'_{q \times (T-q)})'$, de dimensão $(T \times q)$, em que 0 é uma matriz de zeros de dimensão $(T - q) \times q$ e B_q é uma matriz triangular superior de dimensão $q \times q$, configurada da seguinte forma:

$$B_\theta = \begin{bmatrix} \theta_q & \theta_{q-1} & \dots & \theta_1 \\ 0 & \theta_q & \dots & \theta_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_q \end{bmatrix}.$$

Dessa forma, pode-se escrever:

$$Y = L_\theta a + F \epsilon_*.$$

Invertendo L_θ para isolar a , obtém-se:

$$a = L_\theta^{-1} Y - L_\theta^{-1} F \epsilon_*.$$

Agora defina $\epsilon \equiv (\epsilon'_*, a')'$ e as matrizes $L_{(T+q) \times T} = (0'_{T \times q}, L_\theta^{-1})'$, em que 0 é uma matriz de dimensão $(q \times T)$, e $M_{(T+q) \times q} = (I_q, F' L_\theta^{-1})'$, em que I_q é uma matriz identidade de dimensão q . Dessa maneira, escreve-se

$$\epsilon = LY - M \epsilon_*.$$

Segundo Perron (1990), pode-se dizer que a idéia principal é derivar a distribuição conjunta de $\{y_t\}_{t=1}^T$ a partir da distribuição conjunta dos erros $\{\epsilon_t\}_{t=-(q-1)}^T$. Ora, o problema evidente dessa especificação é não observar os erros. Logo, a seção segue buscando formas de estimar consistentemente esses

erros inicialmente, mantendo a estrutura exata da função de verossimilhança. Assim, a função densidade de probabilidade conjunta de ϵ é:

$$f(\epsilon) = (2\pi\sigma^2)^{-\frac{T+q}{2}} \exp \left[-\frac{\epsilon' \epsilon}{2\sigma^2} \right].$$

O jacobiano da transformação de $\{\epsilon_t\}_{t=1}^T$ em $\{y_t\}_{t=1}^T$ é unitário, como se pode observar pela matriz L , de modo que a função densidade de probabilidade conjunta de (Y, ϵ_*) é:

$$f(Y, \epsilon_*; \Psi) = (2\pi\sigma^2)^{-\frac{T+q}{2}} \exp \left[-\frac{\tilde{Q}(\theta_1, \theta_2, \dots, \theta_q, \epsilon_*)}{2\sigma^2} \right],$$

em que

$$\tilde{Q}(\theta_1, \theta_2, \dots, \theta_q, \epsilon_*) = (LY - M\epsilon_*)' (LY - M\epsilon_*).$$

Os erros a partir de $t = 1$ dependem de ϵ_* , razão pela qual é necessário encontrar uma forma de estimá-lo corretamente. Para isso, imagine preliminarmente que, de alguma forma, os parâmetros Ψ sejam conhecidos. Nesse caso, seria simples obter ϵ_* a partir da projeção de LY sobre M , ou seja, a partir da regressão linear:

$$LY = M\epsilon_* + u,$$

de modo que a projeção LY sobre M , para um Y fixado, resulta no coeficiente $\hat{\epsilon}_*$, ou seja:

$$\hat{\epsilon}_* = (M'M)^{-1} M'LY.$$

Dessa configuração, está claro que, conhecidos os parâmetros Ψ , recuperam-se $\hat{\epsilon}_*$ e, assim, todos os demais erros. A função de verossimilhança, entretanto, exige que se conheça o parâmetro ϵ . Por isso, ainda é necessário trabalhar um pouco mais para separar a função de verossimilhança em duas partes. A primeira parte servirá para obter Ψ a partir de Y unicamente; a segunda parte servirá para obter a função concentrada em ϵ_* , de forma a definir esse parâmetro condicional a Y e Ψ . Somando e subtraindo $-2(LY - M\hat{\epsilon}_*)' M\hat{\epsilon}_* + \hat{\epsilon}_* M' M \epsilon + \epsilon M' M \hat{\epsilon}_*$, mostra-se via manipulações algébricas maçantes que:

$$\begin{aligned} \tilde{Q}(\theta_1, \theta_2, \dots, \theta_q, \epsilon_*) &= (LY - M\epsilon_*)' (LY - M\epsilon_*) \\ &= (LY - M\hat{\epsilon}_*)' (LY - M\hat{\epsilon}_*) + (\epsilon_* - \hat{\epsilon}_*)' M' M (\epsilon_* - \hat{\epsilon}_*) \\ &\quad - 2(LY - M\hat{\epsilon}_*)' M (\epsilon_* - \hat{\epsilon}_*). \end{aligned}$$

Note que se parte da função de verossimilhança exata a partir do parâmetro ϵ_* e chega-se a expressões que dependem desse parâmetro e mais uma parcela

totalmente independente dele. O último termo é nulo porque $\hat{u} = (LY - M\hat{\epsilon}_*)$ é ortogonal a M . Logo, fazendo $Q(\theta_1, \theta_2, \dots, \theta_q) \equiv (LY - M\hat{\epsilon}_*)'(LY - M\hat{\epsilon}_*)$, tem-se:

$$\tilde{Q}(\theta_1, \theta_2, \dots, \theta_q, \epsilon_*) = Q(\theta_1, \theta_2, \dots, \theta_q) + (\epsilon_* - \hat{\epsilon}_*)' M' M (\epsilon_* - \hat{\epsilon}_*).$$

Por conseguinte, a função densidade de probabilidade de (Y, ϵ_*) é:

$$f(Y, \epsilon_*; \Psi) = (2\pi\sigma^2)^{-\frac{T+q}{2}} \exp \left[-\frac{Q(\theta_1, \theta_2, \dots, \theta_q) + (\epsilon_* - \hat{\epsilon}_*)' M' M (\epsilon_* - \hat{\epsilon}_*)}{2\sigma^2} \right].$$

Essa configuração permite decompor $f(Y, \epsilon_*; \Psi)$ em $f(Y; \Psi)$ e $f(\epsilon_*; Y, \Psi)$, em que a primeira função é a densidade marginal dos dados e a segunda é a densidade marginal dos erros iniciais, condicional aos dados. Logo, tem-se:

$$f(Y; \Psi) = (2\pi\sigma^2)^{-\frac{T}{2}} |M' M|^{-\frac{1}{2}} \exp \left[-\frac{Q(\theta_1, \theta_2, \dots, \theta_q)}{2\sigma^2} \right]$$

e

$$f(\epsilon_*; Y, \Psi) = (2\pi\sigma^2)^{-\frac{q}{2}} |M' M|^{\frac{1}{2}} \exp \left[-\frac{(\epsilon_* - \hat{\epsilon}_*)' M' M (\epsilon_* - \hat{\epsilon}_*)}{2\sigma^2} \right].$$

O estimador de máxima verossimilhança de Ψ pode ser obtido unicamente a partir da maximização de $f(Y; \Psi)$. A probabilidade incondicional de Ψ , dado Y , é igual a $f(Y; \Psi)$.

Perron (1990) nota que, para T grande, $|M' M|^{-\frac{1}{2}}$ circunda 1, exceto quando as raízes da polinomial estão próximas do círculo unitário. De fato, para um $MA(1)$, $|M' M|^{-\frac{1}{2}} \simeq (1 - \theta^2)^{\frac{1}{2}}$. Porém, se $|\theta| \rightarrow 1$, a aproximação é ruim. Os exemplos a seguir ilustram esse ponto.

Exemplo 3.9

Considere um $MA(1)$. Então,

$$M' = \begin{bmatrix} 1 & \theta & -\theta^2 & \theta^3 & \dots & (-1)^{T+1} \theta^T \end{bmatrix}.$$

Assim:

$$\begin{aligned} |M' M|^{-\frac{1}{2}} &= (1 + \theta^2 + \theta^4 + \dots + \theta^{2T})^{-\frac{1}{2}} \\ &= \left(\frac{1 - \theta^2}{1 - \theta^{2T}} \right)^{\frac{1}{2}} \simeq (1 - \theta^2)^{\frac{1}{2}}. \end{aligned}$$

A igualdade da segunda linha decorre da fórmula da soma de uma progressão geométrica, e a aproximação vale quando T é grande. Colocando alguns números,

$$(\theta; T) = (0, 4; 10) : \left(\frac{1 - \theta^2}{1 - \theta^{2T}} \right)^{\frac{1}{2}} = 0,917 \simeq 0,917 = (1 - \theta^2)^{\frac{1}{2}};$$

$$(\theta; T) = (0, 8; 10) : \left(\frac{1 - \theta^2}{1 - \theta^{2T}} \right)^{\frac{1}{2}} = 0,603 \simeq 0,600 = (1 - \theta^2)^{\frac{1}{2}};$$

$$(\theta; T) = (0, 8; 100) : \left(\frac{1 - \theta^2}{1 - \theta^{2T}} \right)^{\frac{1}{2}} = 0,600 \simeq 0,600 = (1 - \theta^2)^{\frac{1}{2}}.$$

Exemplo 3.10

No caso de um MA (2), a matriz M , em que $T = 5$, é

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \theta_2 & \theta_1 \\ -\theta_1\theta_2 & \theta_1 - \theta_1^2 \\ -\theta_2(\theta_2 - \theta_1^2) & -\theta_1^2 - \theta_1(\theta_2 - \theta_1^2) \\ -\theta_2(\theta_1^3 - 2\theta_1\theta_2) & -\theta_1(\theta_1^3 - 2\theta_1\theta_2) - \theta_1(\theta_2 - \theta_1^2) \\ \theta_2(\theta_1^4 - 3\theta_1^2\theta_2 + \theta_2^2) & \theta_1(\theta_1^4 - 3\theta_1^2\theta_2 + \theta_2^2) - \theta_1(\theta_1^3 - 2\theta_1\theta_2) \end{bmatrix}$$

Assim, expandindo um pouco mais para $T = 7$, obtêm-se os seguintes valores:

$$(\theta_1; \theta_2) = (0, 2; 0, 4) : |M'M|^{-\frac{1}{2}} = 0,882;$$

$$(\theta_1; \theta_2) = (0, 2; 0, 6) : |M'M|^{-\frac{1}{2}} = 0,767;$$

$$(\theta_1; \theta_2) = (1, 2; -0, 6) : |M'M|^{-\frac{1}{2}} = 0,069.$$

Portanto, se se pode ignorar $|M'M|^{-\frac{1}{2}}$, a função de verossimilhança reduz-se à minimização de

$$Q(\theta_1, \theta_2, \dots, \theta_q) = (LY - M\hat{\epsilon}_*)'(LY - M\hat{\epsilon}_*).$$

Sabe-se que $\hat{\epsilon}_*$ é a esperança de ϵ_* condicional a Y e aos parâmetros do modelo. Usando $\epsilon = LY - M\epsilon_*$, tem-se:

$$E(\epsilon|Y, \Psi) = LY - ME(\epsilon_*|Y, \Psi) = LY - M\hat{\epsilon}_*.$$

Por conseguinte, a função de verossimilhança aproximada equivale a minimizar

$$Q(\theta_1, \theta_2, \dots, \theta_q) = \sum_{t=(q-1)}^T \hat{\epsilon}_t^2,$$

em que $\hat{\epsilon}_t$ é a esperança condicional de ϵ_t , obtida a partir da equação $LY - M\hat{\epsilon}_*$.

Neste ponto, convém destacar que a decomposição da função de verossimilhança em duas partes também foi interessante para verificar a necessidade de só maximizar a primeira parcela. Isso ocorre porque, afinal, o interesse é estimar os parâmetros do modelo.

Apesar do esforço, ainda não se atingiu o objetivo final de encontrar explicitamente os erros como função dos valores observados de Y , de forma a obter $\hat{\epsilon}_t$. Há duas maneiras de resolver isso. A primeira é a seguinte:

1. Escolha os valores iniciais de Ψ ;
2. Obtenha $\hat{\epsilon}_* = (M'M)^{-1} M'LY$;
3. Calcule $\hat{\epsilon} = (LY - M\hat{\epsilon}_*)$;
4. Em seguida, minimize $Q(\theta_1, \theta_2, \dots, \theta_q) = \hat{\epsilon}'\hat{\epsilon}$ e obtenha um novo conjunto de parâmetros;
5. Repita 2 até a convergência dos parâmetros.

Na segunda maneira, prevê-se para trás, segundo o procedimento de Box, Jenkins e Reinsel (1994) denominado *backforecasting*. A idéia é usar as últimas observações para fazer a previsão para trás, baseado no fato de que um modelo estacionário pode ser equivalentemente representado por um modelo com erros do futuro. O uso das últimas observações para prever as anteriores é equivalente a utilizar um modelo cuja variável atual é função das variáveis do futuro. Sendo assim, é preciso usar o operador avanço para resolver a previsão.

Para visualizar o procedimento, considere um $MA(q)$ convencional:

$$y_t = \theta(L)\epsilon_t.$$

Seja o operador avanço, F , tal que $Fe_t = \epsilon_{t+1}$, em que $F = L^{-1}$. Considere o modelo anterior a partir do operador avanço:

$$\begin{aligned} y_t &= \theta(F)\epsilon_t, \\ \epsilon_t &\sim RB(0, \sigma_\epsilon^2). \end{aligned}$$

Tome a esperança condicional a Y e Ψ em ambos os lados dessa equação, assumindo como hipótese inicial que os erros antes e depois da amostra são nulos, isto é, $E(e_{-j}|Y, \Psi) = 0$, para $j = 1, 2, \dots$ e $j = -(T+i)$, $i > 0$:

$$E(y_t|Y, \Psi) = \theta(F)E(e_t|Y, \Psi), \quad (5)$$

para todo $t = 1, 2, \dots, T$. Ora, $E(y_t|Y, \Psi) = y_t$, $t = 1, 2, \dots, T$. Portanto, o procedimento é o seguinte:

1. Encontre $E(e_{T-j}|Y, \Psi)$ para $j = 0, 1, 2, \dots, T-1$, usando a equação (5);
2. Em seguida, aplique a equação (5) para encontrar $E(y_t|Y, \Psi)$ para $t = 0, -1, -2, \dots, -(q-1)$;
3. Utilize o resultado encontrado no item anterior para obter uma estimativa para $E(e_t|Y, \Psi)$:

$$E(e_t|Y, \Psi) = \theta^{-1}(L)E(y_t|Y, \Psi) \text{ para } t = -(q-1), -(q-2), \dots, T.$$

Exemplo 3.11

O modelo com operador avanço em um MA(1):

$$y_t = (1 + \theta_1 F)e_t = e_t + \theta_1 e_{t+1};$$

$$E(y_t|Y, \Psi) = E(e_t|Y, \Psi) + \theta_1 E(e_{t+1}|Y, \Psi), \quad t = 1, 2, \dots, T.$$

Usando $E(e_{T+1}|Y, \Psi) = 0$, então:

$$y_T = E(e_T|Y, \Psi);$$

$$y_t = E(e_t|Y, \Psi) + \theta_1 E(e_{t+1}|Y, \Psi), \quad t = 1, 2, \dots, T-1 \Rightarrow$$

$$E(e_t|Y, \Psi) = y_t - \theta_1 E(e_{t+1}|Y, \Psi).$$

Fazendo $E(e_0|Y, \Psi) = 0$, resulta que $E(y_0|Y, \Psi) = \theta_1 E(e_1|Y, \Psi)$. Com isso, retorna-se ao modelo inicial:

$$E(y_t|Y, \Psi) = E(e_t|Y, \Psi) + \theta_1 E(e_{t-1}|Y, \Psi), \quad t = 1, 2, \dots, T.$$

Para isso, inicie com $E(y_0|Y, \Psi) = E(e_0|Y, \Psi)$, sabendo que $E(y_0|Y, \Psi) = \theta_1 E(e_1|Y, \Psi)$. Dessa forma:

$$E(y_t|Y, \Psi) = E(e_t|Y, \Psi) + \theta_1 E(e_{t-1}|Y, \Psi) \Rightarrow$$

$$E(e_t|Y, \Psi) = E(y_t|Y, \Psi) - \theta_1 E(e_{t-1}|Y, \Psi)$$

$$= y_t - \theta_1 E(e_{t-1}|Y, \Psi), \quad t = 1, 2, \dots, T.$$

Exemplo 3.12

O modelo com operador avanço em um MA (2):

$$y_t = (1 + \theta_1 F + \theta_2 F^2) e_t = e_t + \theta_1 e_{t+1} + \theta_2 e_{t+2};$$

$$E(y_t|Y, \Psi) = E(e_t|Y, \Psi) + \theta_1 E(e_{t+1}|Y, \Psi) + \theta_2 E(e_{t+2}|Y, \Psi), \quad t = 1, 2, \dots, T.$$

Usando $E(e_{T+i}|Y, \Psi) = 0$, $i = 1, 2$, então:

$$y_T = E(e_T|Y, \Psi);$$

$$y_{T-1} = E(e_{T-1}|Y, \Psi) + \theta_1 E(e_T|Y, \Psi) \implies$$

$$E(e_{T-1}|Y, \Psi) = y_{T-1} - \theta_1 E(e_T|Y, \Psi)$$

$$E(e_t|Y, \Psi) = y_t - \theta_1 E(e_{t+1}|Y, \Psi) - \theta_2 E(e_{t+2}|Y, \Psi), \quad t = 1, 2, \dots, T-2.$$

Fazendo $E(e_{-j}|Y, \Psi) = 0$, $j = 0, 1$, resulta que:

$$E(y_0|Y, \Psi) = \theta_1 E(e_1|Y, \Psi) + \theta_2 E(e_2|Y, \Psi);$$

$$E(y_{-1}|Y, \Psi) = \theta_2 E(e_1|Y, \Psi).$$

Finalmente, retorna-se ao modelo inicial:

$$E(y_t|Y, \Psi) = E(e_t|Y, \Psi) + \theta_1 E(e_{t+1}|Y, \Psi) + \theta_2 E(e_{t+2}|Y, \Psi), \quad t = 1, 2, \dots, T.$$

Para isso, inicie com $E(y_{-j}|Y, \Psi) = E(e_{-j}|Y, \Psi)$, $j = 0, 1$. Dessa forma:

$$E(y_t|Y, \Psi) = E(e_t|Y, \Psi) + \theta_1 E(e_{t+1}|Y, \Psi) + \theta_2 E(e_{t+2}|Y, \Psi) \implies$$

$$E(e_t|Y, \Psi) = y_t - \theta_1 E(e_{t+1}|Y, \Psi) - \theta_2 E(e_{t+2}|Y, \Psi), \quad t = 1, 2, \dots, T.$$

Segundo Box, Jenkins e Reinsel (1994), o sistema especificado dessa forma recursiva resulta em estimativas muito mais eficientes, particularmente quando as raízes da polinomial do MA (q) estão próximas do círculo unitário.

Hamilton (1994) sugere ainda outra forma de estimar o modelo. Para obter a função de verossimilhança exata, é preciso definir a matriz de covariância com atenção a todas as observações, ou seja, aplicando a decomposição do erro de previsão. Assim, defina $Y = (y_1, y_2, \dots, y_T)'$ e obtenha a matriz de covariância, Ω :

$$\Omega = \begin{bmatrix} E(y_1 - \mu)^2 & E(y_1 - \mu)(y_2 - \mu) & \cdots & E(y_1 - \mu)(y_T - \mu) \\ E(y_2 - \mu)(y_1 - \mu) & E(y_2 - \mu)^2 & \cdots & E(y_2 - \mu)(y_T - \mu) \\ \vdots & \vdots & \ddots & \vdots \\ E(y_T - \mu)(y_1 - \mu) & E(y_T - \mu)(y_2 - \mu) & \cdots & E(y_T - \mu)^2 \end{bmatrix}.$$

A partir da defasagem q , a autocovariância do MA é nula. Logo, a matriz de covariância fica:

$$\Omega = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_q & \cdots & \ddots & \cdots & 0 \\ \gamma_1 & \gamma_0 & \gamma_1 & \ddots & \gamma_{q-1} & \ddots & \cdots & 0 & \vdots \\ \gamma_2 & \gamma_1 & \ddots & \ddots & \vdots & \ddots & \gamma_q & \vdots & \ddots \\ \ddots & \ddots & \ddots & \ddots & \gamma_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \gamma_0 & \ddots & \cdots & \ddots & \gamma_q \\ \ddots & \ddots & \ddots & \ddots & \gamma_1 & \ddots & \ddots & \ddots & \gamma_{q-1} \\ \ddots & \ddots & \gamma_q & \ddots & \gamma_2 & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \gamma_1 \\ 0 & \cdots & \ddots & \ddots & \cdots & \ddots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}.$$

Essa matriz é usada na função de verossimilhança:

$$f_Y(Y; \Psi) = (2\pi)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (Y - \mu)' \Omega^{-1} (Y - \mu) \right].$$

Pode-se fatorar triangularmente Ω , tal que:

$$\Omega = A\Lambda A',$$

em que

A é uma matriz triangular inferior;

Λ é uma matriz diagonal.

Essa decomposição é necessária para especificar a série recursivamente. Como há muitos zeros em Ω distribuídos de uma forma regular, é relativamente fácil encontrar A e Λ .

Exemplo 3.13

Supondo um MA (1), pode-se escrever a função de verossimilhança

$$f_Y(Y; \Psi) = (2\pi)^{-\frac{T}{2}} |\Lambda|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \tilde{Y}' \Lambda^{-1} \tilde{Y} \right],$$

em que

$|A| = 1$, por ser uma matriz triangular cuja diagonal é formada por 1s;

$\tilde{Y} = A^{-1}(Y - \mu)$, sendo que dessa equação pode-se mostrar que:

$$\tilde{y}_t = y_t - \mu - \frac{\theta [1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-2)}]}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}} \tilde{y}_{t-1}, \quad t > 1,$$

$$\tilde{y}_1 = y_1 - \mu,$$

$$\lambda_{tt} = E(\tilde{y}_t^2) = \sigma^2 \frac{\theta [1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-2)}]}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}}.$$

Interpreta-se \tilde{y}_t como sendo o resíduo da projeção linear de y_t sobre uma constante e seus valores passados, enquanto λ_{tt} é interpretado como o erro quadrático médio dessa projeção linear.

3.5.3 Função de Verossimilhança para um ARMA (p,q)

A função de verossimilhança exata do modelo ARMA depende da matriz Ω , que se torna bem complicada na presença de componentes auto-regressivos. O procedimento de estimação, no entanto, é semelhante ao do processo MA, embora os componentes auto-regressivos tornem a expressão bem mais complexa.

3.6

Diagnóstico de Resíduos

Estimado o modelo, deve-se verificar como os resíduos ficaram. Na formulação inicial do modelo temporal, assumiu-se que os erros eram um ruído branco. Ora, o mesmo deve acontecer com os resíduos estimados. Isso significa verificar se a FAC e FACP dos resíduos estimados mostram-se sem qualquer memória. Se a hipótese nula é rejeitada, isso implica dizer que há informação ainda não captada pelo economista, o que pode gerar previsões pobres.

Ora, se se verifica que o modelo estimado não produziu ruídos brancos, então descarta-se esse modelo e testam-se outras possibilidades, até que se encontre um cujos resíduos comportem-se como um ruído branco. A recomendação usual é utilizar os testes já apresentados sobre os resíduos também.

Se não se rejeita a hipótese nula de não-autocorrelação dos resíduos via FAC, FACP e Ljung-Box, os resíduos comportam-se como um ruído branco. Isso é correto, por uma razão muito sutil. Embora tais testes tenham sido indicados para séries *observadas*, a utilização de séries estimadas implicaria um intervalo de confiança maior do que realmente é calculado. Logo, se a nula não é rejeitada sob a hipótese de séries observadas, com maior razão não se rejeita a nula com as séries estimadas.

Maior problema ocorre quando se rejeita a nula, pois isso poderia não acontecer se fossem usados os intervalos de confiança corretos. Isso é cometer o erro

do tipo I. Para minimizar esse problema, a recomendação é olhar apenas para o Ljung-Box nos resíduos, pois trata-se de um teste menos sujeito a equívocos.

Nota 3.13 Quando ocorrer de rejeitar a nula pelo Ljung-Box, um conselho prático é olhar para a FAC e a FACP para tentar descobrir que autocorrelação sobressai e que deveria ser mais bem modelada. Não obstante, como já foi dito, um procedimento melhor é identificar a série desde o começo, usando os critérios de informação e descartando os modelos cujos resíduos falham no teste de Ljung-Box.

Exemplo 3.14

O exemplo refere-se ao modelo AR (1) estimado anteriormente para a série do IPCA. Mostram-se os testes FAC, FACP e Ljung-Box dos resíduos estimados. Sem o ajustamento sazonal $X - 11$, a nula seria rejeitada porque as autocorrelações múltiplas de 12 seriam significantes. Isso indica a presença de sazonalidade. Uma vez feito esse ajustamento e estimado o modelo, encontra-se o resultado mostrado na Figura 3.8.

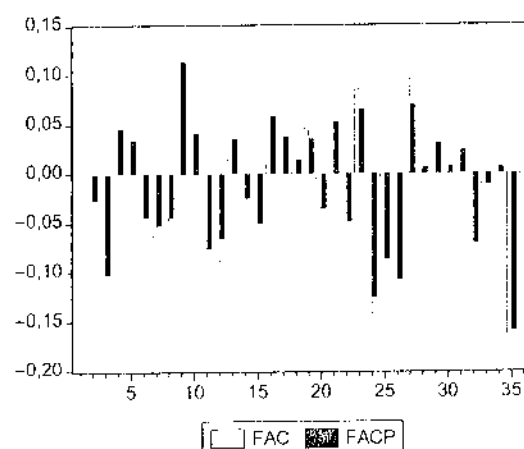


Figura 3.8 FAC e FACP dos resíduos.

Testes de Autocorrelação								
Defasagem	1	2	3	4	5	10	15	20
FAC	0	-0,02	-0,10	0,04	0,04	0,05	-0,04	-0,05
FACP	0	-0,02	-0,10	0,04	0,03	0,04	-0,05	0,04
Q	2×10^{-5}	0,11	1,72	2,05	2,30	6,39	9,29	10,7
Prob		0,73	0,42	0,56	0,68	0,70	0,81	0,93

3.6.1 Teste de Normalidade

Uma maneira de testar a série é verificando se, além de ruído branco, os resíduos são normalmente distribuídos. Pode-se visualizar isso ou testar a normalidade. A visualização consiste em estimar a distribuição da série usando um ponderador.

No que se refere à normalidade, a idéia é fazer o gráfico da densidade estimada usando o *kernel* como ponderador. Usa-se o histograma para representar os resultados, no qual o eixo horizontal apresenta o intervalo dos valores e o vertical, a frequência. Em geral, o *kernel* do estimador dos resíduos é dado por:

$$\hat{f}_h(\epsilon) = \frac{\sum_{t=1}^T K\left(\frac{\epsilon - \hat{\epsilon}_t}{h}\right)}{Th},$$

em que

h é a largura da janela ou parâmetro de suavização;

$K(\cdot)$ é a função *kernel*, tipicamente uma função densidade de probabilidade simétrica ao redor de zero;

$\hat{\epsilon}_t^s = (\hat{\epsilon}_t - \bar{\hat{\epsilon}})/\hat{\sigma}$ é o resíduo padronizado.

A ponderação serve para suavizar os pontos observados. Isso é feito colocando-se menos peso nos erros mais distantes do ponto sob avaliação. Alguns tipos de função K estão listados a seguir, sendo $I(\cdot)$ a função indicador que é 1, quando seu argumento é verdadeiro, e 0, em caso contrário:

Kernel	$K(u)$
Gaussiana	$\frac{\exp\left(-\frac{u^2}{2}\right)}{\sqrt{2\pi}}$
Bi-ponderada	$\frac{15}{16} (1 - u^2)^2 I(u < 1)$
Triangular	$(1 - u) I(u < 1)$
Epanechnikov	$\frac{3}{4} \frac{\left(1 - \frac{u^2}{5}\right)}{\sqrt{5}} I(u < \sqrt{5})$

A escolha da função *kernel* geralmente não altera significativamente a função densidade. Entretanto, quanto maior h , mais suave será a função. De fato, a escolha de h é crucial, razão pela qual vários métodos para defini-lo têm sido propostos. Silverman (1986, p. 48) recomenda o seguinte:

$$h = \frac{0,9}{\sqrt[5]{T}} \min\left(\hat{\sigma}, \frac{\text{distância entre quartis}}{1,34}\right).$$

Exemplo 3.15

Apresenta-se aqui a Figura 3.9 dos resíduos, seguindo o kernel triangular:

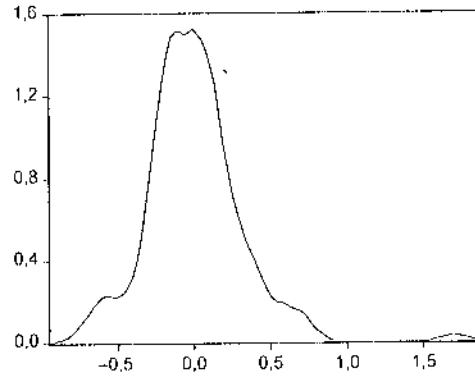


Figura 3.9 Kernel triangular.

A figura mostra que os resíduos fogem da distribuição normal padrão em razão de uma observação outlier ocorrida em novembro de 2002. A data exata só pode ser determinada por inspeção visual da série. Isso sugere a observação de que a inspeção visual da série é fundamental para definir se há outliers e a razão de sua existência. A exclusão desse outlier provavelmente tornaria a série normal.

Entretanto, não se podem excluir outliers de séries temporais, pois isso quebraria a dependência temporal das observações, que é justamente o que o econometrista está tentando extrair. Um tratamento aos outliers deve ser dado. Para isso, veja Balke e Fomby (1994) e Fox (1972).

3.6.2 Teste Jarque-Bera

Trata-se de um teste para verificar se os momentos da série estimada são iguais aos da normal. Sob essa hipótese, a assimetria é igual a zero e a curtose é igual a 3. Portanto, deve-se testar a hipótese conjunta:

$$H_0 : E(\hat{\epsilon}_t^s)^3 = 0 \wedge E(\hat{\epsilon}_t^s)^4 = 3$$

×

$$H_1 : E(\hat{\epsilon}_t^s)^3 \neq 0 \vee E(\hat{\epsilon}_t^s)^4 \neq 3.$$

Para implementá-lo, usa-se a estatística:

$$JB = \frac{T}{6} \left[\frac{\sum_{t=1}^T (\hat{\epsilon}_t^s)^3}{T} \right]^2 + \frac{T}{24} \left[\frac{\sum_{t=1}^T (\hat{\epsilon}_t^s)^4}{T} - 3 \right]^2 \xrightarrow{d} \chi^2_2.$$

É importante notar que a rejeição da hipótese nula indica não-normalidade, porém a não-rejeição não indica normalidade. De fato, a não-rejeição indica apenas que o terceiro e o quarto momentos da distribuição empírica coincidem com os da normal.

Exemplo 3.16

Usando o IPCA sazonalmente ajustado, rejeita-se a nula do teste de Jarque-Bera, em função daquela observação extrema, conforme mostra a tabela a seguir. Perceba que a série é assimétrica (o valor da estatística deveria ser 0 para haver assimetria) e tem excesso de curtose ou tem caudas mais espessas do que as da distribuição normal (para ter caudas da mesma espessura que a distribuição gaussiana, o valor da estatística deveria ser 3). O valor positivo da assimetria também indica que a série é assimétrica para a direita. Essas observações podem ser corroboradas por meio de inspeção visual da Figura 3.10.

Teste Jarque-Bera			
Média	0,644	Assimetria	1,487
Mediana	0,472	Curtose	5,830
Máximo	2,884	Jarque-Bera	104,683
Mínimo	-0,457	Probabilidade	0,000
Desvio-padrão	0,527		

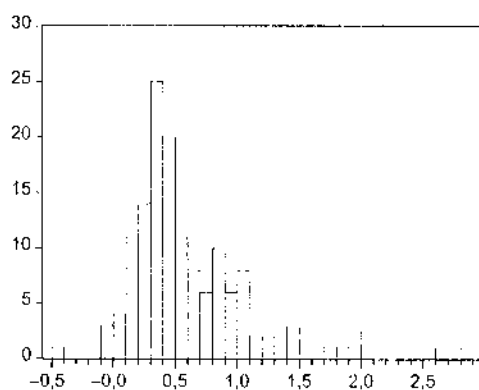


Figura 3.10 Histograma IPCA.

3.6.3 Teste LM

O teste LM para autocorrelação dos resíduos é também conhecido como teste de Breusch-Godfrey. Para implementá-lo, regrida:

$$\hat{\epsilon}_t = \beta_1 \hat{\epsilon}_{t-1} + \beta_2 \hat{\epsilon}_{t-2} + \cdots + \beta_h \hat{\epsilon}_{t-h} + u_t.$$

Teste:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_h = 0 \times H_1 : \beta_1 \neq 0, \text{ ou } \beta_2 \neq 0, \text{ ou } \cdots \text{ ou } \beta_h \neq 0.$$

Pode-se mostrar que:

$$LM_h = T \times R^2 \xrightarrow{d} \chi_h^2$$

sob a hipótese nula. Rejeita-se a nula se o valor calculado exceder o valor tabelado.

Exemplo 3.17

Usando ainda o exemplo do IPCA ajustado sazonalmente, não se rejeita a nula de inexistência de autocorrelação quando se aplica o teste LM ou Breusch-Godfrey sobre os resíduos estimados, segundo a estatística qui-quadrado ao nível de significância de 39%. Perceba também pelos resultados da tabela a seguir que os valores das estatísticas t dos resíduos defasados não são significativos, indicando que não ajudam a explicar o resíduo vigente. Isso reforça a ausência de autocorrelação dos resíduos. Ainda se observa o valor baixo do R^2 e da estatística F , indicando a insignificância da equação estimada no teste.

Teste de LM ou Breusch-Godfrey				
Estatística F	0,988	Probabilidade	0,400	
$T \times R^2$	3,006	Prob. χ^2	0,390	
Variáveis	Coeficiente	Desvio-Padrão	Estatística t	P-Valor
C	-0,017	0,130	-0,131	0,896
AR (1)	0,106	0,091	1,164	0,246
Resíduos(-1)	-0,111	0,124	-0,894	0,373
Resíduos(-2)	-0,113	0,110	-1,022	0,308
Resíduos(-3)	-0,169	0,101	-1,671	0,096
R^2	0,020	Durbin Watson	2,002	

3.6.4 Teste ARCH-LM

O teste ARCH-LM serve para identificar sinais de heterocedasticidade condicional. Para implementá-lo, regrida:

$$\hat{\epsilon}_t^2 = \beta_1 \hat{\epsilon}_{t-1}^2 + \beta_2 \hat{\epsilon}_{t-2}^2 + \cdots + \beta_h \hat{\epsilon}_{t-h}^2 + u_t.$$

Teste:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_h = 0 \times H_1 : \beta_1 \neq 0, \text{ ou } \beta_2 \neq 0, \text{ ou } \dots \text{ ou } \beta_h \neq 0.$$

Pode-se mostrar que:

$$ARCH-LM_h = T \times R^2 \xrightarrow{d} \chi_h^2$$

sob a hipótese nula. Rejeita-se a nula se o valor calculado exceder o valor tabelado.

Exemplo 3.18

Mais uma vez usando o IPCA, não se rejeita a nula de inexistência de heterocedasticidade condicional na inflação modelada ao nível de significância de 30%. Nesse teste é utilizado o resíduo defasado ao quadrado como variável dependente. Perceba que novamente as estatísticas t não são significantes, o R^2 tem um valor inexpressivo e a probabilidade da estatística F tem um valor alto.

Teste ARCH-LM				
Estatística F	1,227	Probabilidade	0,302	
$T \times R^2$	3,689	Prob. χ^2	0,296	
Variáveis	Coeficiente	Desvio-Padrão	Estatística t	P-Valor
C	0,080	0,024	3,217	0,001
Resíduos ² (-1)	0,132	0,083	1,577	0,116
Resíduos ² (-2)	0,059	0,083	0,715	0,475
Resíduos ² (-3)	-0,062	0,083	-0,753	0,452
R^2	0,025	Durbin Watson	2,007	

3.6.5 Teste Reset

O teste de erro de especificação de regressão, RESET, testa a presença de não-linearidades na série. A hipótese nula é de linearidade contra a alternativa de não-linearidade. Se os resíduos do modelo estimado forem independentes, não deverão ser correlacionados com qualquer outra variável, particularmente com os regressores da regressão original, bem como com seus valores estimados e suas potências.

Para implementar o teste, estime o modelo:

$$y_t = x_t' \beta + \epsilon_t.$$

Considere $\hat{\beta}$ o parâmetro estimado por mínimos quadrados ordinários dessa regressão. Agora, proceda à seguinte regressão adicional:

$$\hat{\epsilon}_t = x_t' \beta + \sum_{j=2}^h \varphi_j \hat{y}_t^j + v_t.$$

Sob a hipótese de que a regressão original está correta:

$$H_0 : \varphi_2 = \varphi_3 = \dots = \varphi_h = 0.$$

Assim, não haverá erro de especificação se a seguinte estatística:

$$RESET_h = \frac{\left(\sum_{t=1}^T \hat{\epsilon}_t^2 - \sum_{t=1}^T \hat{v}_t^2 \right)}{\frac{h-1}{\sum_{t=1}^T \hat{v}_t^2}} \xrightarrow{d} F(h-1, T-K-h+1),$$

em que K é a dimensão de x_t .

Na prática, $h = 2$ ou 3 é suficiente para detectar possíveis problemas de especificação.

Exemplo 3.19

Primeiro, estima-se o modelo auto-regressivo com inflação e toma-se a soma dos quadrados dos resíduos:

$$\sum_{t=1}^{131} \hat{\epsilon}_t^2 = 0,001437.$$

Em seguida, estime o modelo:

$$\hat{\epsilon}_t = c + \beta y_{t-1} + \sum_{j=2}^h \varphi_j \hat{y}_t^j + v_t, h=3.$$

A soma dos quadrados dos resíduos dessa regressão é:

$$\sum_{t=1}^{131} \hat{v}_t^2 = 0,001382.$$

Calcule a estatística $RESET_3$:

$$RESET_3 = \frac{\left(\sum_{t=1}^{131} \hat{\epsilon}_t^2 \dots \sum_{t=1}^{131} \hat{\theta}_t^2 \right)}{\frac{3-1}{\sum_{t=1}^{131} \hat{\theta}_t^2}} \xrightarrow{d} F(2, 127);$$

$$RESET_3 = \frac{0,001437 - 0,00138}{\frac{2}{0,00138 \cdot 127}} = 2,5154.$$

A esse valor não se rejeita a nula a 5%, mas rejeita-se a 10%, pois nesse caso o nível de significância pode ser calculado em 8,49%. A seguir, é apresentada uma tabela com o nível de significância para vários h s:

h	2	3	4	5
Prob.	98,00%	8,49%	6,08%	11,68%

Outros testes comuns para diagnosticar resíduos são: teste de Chow para estabilidade, análise recursiva, teste CUSUM etc.

3.7

Exemplos Simulados

Exemplo 3.20

A seguir, desenvolve-se um exemplo completo de uma série simulada. O intuito é mostrar os resultados se for adotado um modelo equivocado e se for adotado um modelo corretamente especificado. Para esse propósito, é mais conveniente usar uma série simulada. Assim, considere a seguinte série simulada com:

$$y_t = 1 + \epsilon_t + 0,3\epsilon_{t-1} - 0,4\epsilon_{t-2}, \epsilon_t \sim N(0, 1).$$

A série foi gerada com 502 observações, das quais serão utilizadas apenas as observações entre 101 e 300, cujo gráfico está na Figura 3.11.

Uma forma de identificar a série é observando a FAC, FACP e Ljung-Box para autocorrelações, conforme mostra a Figura 3.12.

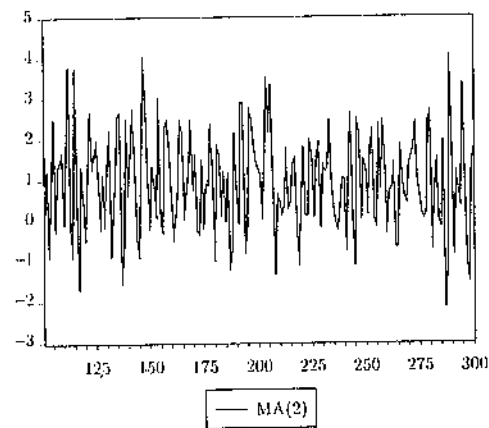


Figura 3.11 MA (2) – Simulado.

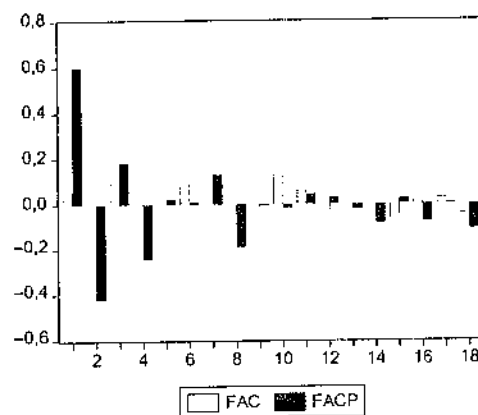


Figura 3.12 MA (2) – Autocorrelações.

Testes de Autocorrelação								
Defasagem	1	2	3	4	5	10	15	20
FAC	0,060	-0,415	0,090	0,014	-0,135	0,133	-0,062	-0,045
FACP	0,060	-0,420	0,183	-0,243	0,024	-0,016	0,026	-0,111
Q	0,7321	35,895*	37,539*	37,579*	41,885*	63,749*	66,678*	67,325*

* Indica significância a 1%.

A FACP indica decaimento exponencial a partir da defasagem 2 na correlação parcial, como algum problema na defasagem 8. A FAC indica truncagem

degenerada na defasagem 2. Ignorando totalmente o processo gerador de dados, o econometrista é levado a crer que se trata de um MA (2) degenerado, significando um modelo do tipo: $y_t = \epsilon_t + \theta\epsilon_{t-2}$.

Inicia-se, pois, estimando-se um MA (2) degenerado, embora o leitor saiba que o verdadeiro modelo não é esse. O objetivo dessa regressão é ver o que acontece com os resíduos quando o modelo estimado não é exatamente o verdadeiro. Os resultados são apresentados a seguir:

$$MA = 0,967 - 0,492\epsilon_{t-2} + \epsilon_t$$

(0,039) (0,061)

Feita a estimação, é o caso de verificar os resíduos da regressão. Os resíduos precisam mostrar-se um ruído branco, pelo critério de Ljung-Box. Caso contrário, é preciso reestimar um novo modelo. A Figura 3.13 mostra que os resíduos ainda não são um ruído branco. A figura permite supor, pela observação das correlações, que há algo na primeira e/ou terceira autocorrelação e correlação parcial.

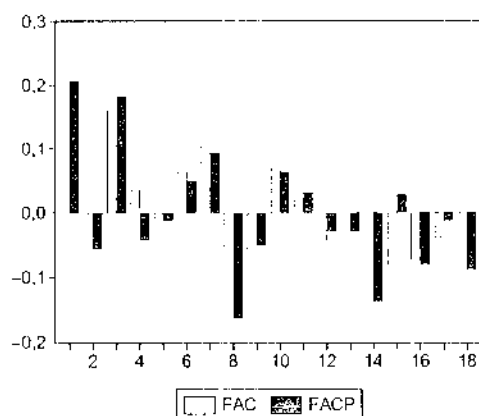


Figura 3.13 MA (2) degenerado – Autocorrelações.

Testes de Autocorrelação								
Defasagem	1	2	3	4	5	10	15	20
FAC	0,206	-0,011	0,160	0,035	-0,030	0,070	-0,083	-0,092
FACP	0,206	-0,056	0,182	0,043	0,012	0,061	0,028	-0,088
Q	8,625	8,649*	13,901*	14,153*	14,339*	24,231*	29,295*	32,667*

* Indica significância a 1%.

Seria o caso de testar as várias combinações possíveis. O teste nos levaria a concluir, pelos critérios de informação, que o melhor modelo é um MA (2) usual. É possível ver que os coeficientes estimados como μ, θ_1 e θ_2 , respectivamente 0,964, 0,250 e -0,531, são próximos dos coeficientes usados na simulação, 1, 0,3 e -0,4.

$$MA = 0,964 + 0,250\epsilon_{t-1} - 0,531\epsilon_{t-2} + \epsilon_t$$

(0,054) (0,057) (0,060)

O critério de informação (não apresentado) é menor que no modelo degenerado. Além disso, as raízes determinam que o modelo é invertível, como é desejado. Nessas condições, o correlograma dos resíduos estimados deve refletir a exatidão do modelo. De fato, a Figura 3.14 mostra isso. Tanto a FAC e a FACP estão dentro do intervalo de confiança, como o teste de Ljung-Box não rejeita a hipótese nula de ausência de autocorrelação residual.

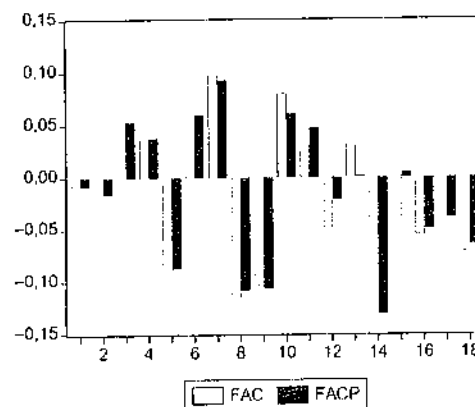


Figura 3.14 MA (2) – Autocorrelação.

Testes de Autocorrelação								
Defasagem	1	2	3	4	5	10	15	20
FAC	0,08	-0,016	0,054	0,037	-0,089	0,080	-0,040	-0,072
FACP	0,08	-0,016	0,052	0,038	-0,087	0,061	-0,005	-0,064
Q	0,013	0,065	0,652	0,941	3,658	11,828	15,552	17,369

A prática é difícil e exige um pouco de experiência na presença de série com configurações mais complicadas na FAC e FACP.

Outras séries, como as financeiras com dados diários, dificilmente são passíveis de “branquear” os resíduos. O problema é que séries de alta frequência

vêm acompanhadas de um ruído considerável, sazonalidades temporárias etc. Uma forma razoável de modelar a série é assumir uma constante na média e modelar apenas a volatilidade, como será visto.

3.8

Previsão

A previsão é relativamente fácil e direta. Inicia-se mostrando a mecânica de previsão do modelo mais simples possível: um $AR(1)$. Em seguida, passa-se ao modelo $ARMA(1,1)$, suficiente para entender os modelos mais complexos. Por isso, considere o seguinte modelo:

$$y_{t+1} = c + \phi y_t + \epsilon_{t+1}.$$

Logo:

$$\begin{aligned} E_t(y_{t+1}) &= c + \phi y_t = y_{t+1} - \epsilon_{t+1}; \\ E_t(y_{t+2}) &= c + \phi E_t(y_{t+1}) = c + \phi(c + \phi y_t); \\ &\vdots \\ E_t(y_{t+h}) &= c \sum_{i=1}^{h-1} \phi^{i-1} + \phi^h y_t. \end{aligned}$$

O erro de previsão no período h , $e_t(h)$, é dado por:

$$\begin{aligned} e_t(1) &= y_{t+1} - E_t(y_{t+1}) = \epsilon_{t+1}; \\ e_t(2) &= y_{t+2} - E_t(y_{t+2}) = c + \phi y_{t+1} + \epsilon_{t+2} - c - \phi E_t(y_{t+1}) \\ &= \phi \epsilon_{t+1} + \epsilon_{t+2}; \\ e_t(3) &= y_{t+3} - E_t(y_{t+3}) = c + \phi y_{t+2} + \epsilon_{t+3} - c - \phi E_t(y_{t+1}) \\ &= \epsilon_{t+3} + \phi \epsilon_{t+2} + \phi^2 \epsilon_{t+1}; \\ &\vdots \\ e_t(h) &= y_{t+h} - E_t(y_{t+h}) = \epsilon_{t+h} + \phi \epsilon_{t+h-1} + \phi^2 \epsilon_{t+h-2} + \cdots + \phi^{h-1} \epsilon_{t+1}. \end{aligned}$$

Tomando a esperança dos erros de previsão, verifica-se que são iguais a zero. Portanto, as previsões não são viesadas. É importante encontrar a variância do erro de previsão. Ela pode ser calculada por:

$$\begin{aligned} \text{Var}(e_t(h)) &= \text{Var}\left(\epsilon_{t+h} + \phi \epsilon_{t+h-1} + \phi^2 \epsilon_{t+h-2} + \cdots + \phi^{h-1} \epsilon_{t+1}\right) \\ &= \sigma^2 \left(1 + \phi^2 + \phi^4 + \cdots + \phi^{2(h-1)}\right). \end{aligned}$$

Nota 3.14 Note que a variância aumenta com o horizonte de previsão, não obstante seja a taxas decrescentes. No limite, quando $h \rightarrow \infty$, a variância de previsão converge à variância não condicional $\sigma^2 / (1 - \phi^2)$.

O intervalo de confiança para resíduos normais é dado da seguinte forma:

$$c \sum_{i=1}^{h-1} \phi^{i-1} + \phi^h y \pm 2\sigma \left(1 + \phi^2 + \phi^4 + \dots + \phi^{2(h-1)} \right)^{\frac{1}{2}}.$$

Nos modelos ARMA de ordem maior, a variância de previsão torna-se bastante complicada. Para dar uma idéia do que ocorre com esses modelos, considere um ARMA (2, 1):

$$y_t = \mu + \sum_{i=1}^2 \phi_i y_{t-i} + \epsilon_t + \theta_1 \epsilon_{t-1}.$$

Calcula-se, a seguir, o erro de previsão para alguns passos à frente:

$$\begin{aligned} e_t(1) &= y_{t+1} - E_t(y_{t+1}) \\ &= \mu + \sum_{i=1}^2 \phi_i y_{t+1-i} + \epsilon_{t+1} + \theta_1 \epsilon_t - \mu - \sum_{i=1}^2 \phi_i y_{t+1-i} - \theta_1 \epsilon_t \\ &= \epsilon_{t+1}; \end{aligned}$$

$$\begin{aligned} e_t(2) &= y_{t+2} - E_t(y_{t+2}) \\ &= \mu + \sum_{i=1}^2 \phi_i y_{t+2-i} + \epsilon_{t+2} + \theta_1 \epsilon_{t+1} - \mu - \sum_{i=1}^2 \phi_i E_t(y_{t+2-i}) \\ &= \phi_1 (y_{t+1} - E_t(y_{t+1})) + \epsilon_{t+2} + \theta_1 \epsilon_{t+1} \\ &= \phi_1 e_t(1) + \epsilon_{t+2} + \theta_1 \epsilon_{t+1}; \end{aligned}$$

$$\begin{aligned} e_t(3) &= y_{t+3} - E_t(y_{t+3}) \\ &= \mu + \sum_{i=1}^2 \phi_i y_{t+3-i} + \epsilon_{t+3} + \theta_1 \epsilon_{t+2} - \mu - \sum_{i=1}^2 \phi_i E_t(y_{t+3-i}) \\ &= \phi_1 [y_{t+2} - E_t(y_{t+2})] + \phi_2 [y_{t+1} - E_t(y_{t+1})] + \epsilon_{t+3} + \theta_1 \epsilon_{t+2} \\ &= \phi_1 e_t(2) + \phi_2 e_t(1) + \epsilon_{t+3} + \theta_1 \epsilon_{t+2}; \end{aligned}$$

\vdots

$$e_t(h) = \phi_1 e_t(h-1) + \phi_2 e_t(h-2) + \epsilon_{t+h} + \theta_1 \epsilon_{t+h-1}, h > 2.$$

O cálculo da variância pode ser feito a partir da fórmula anterior. Observe que as covariâncias cruzadas entre $e_t(h-1)$, $e_t(h-2)$ e os erros não são nulas, complicando sobremaneira a fórmula.

Na prática, não se sabem os verdadeiros valores dos coeficientes, uma vez que são estimados. Logo, é preciso considerar também a variância das estimativas dos coeficientes. Quando isso é feito, a fórmula da variância fica bem difícil de ser calculada, pois é preciso encontrá-la a partir de:

$$e_t(h) = \hat{\phi}_1 e_t(h-1) + \hat{\phi}_2 e_t(h-2) + \epsilon_{t+h} + \hat{\theta}_1 \epsilon_{t+h-1}.$$

Uma discussão detalhada sobre esse assunto pode ser encontrada em Hamilton (1994), embora em uma linguagem um pouco mais carregada e formal.

Feitas as previsões e definidos seus intervalos de confiança, algumas medidas de desempenho são utilizadas para avaliar as previsões. Três delas são básicas: o erro quadrático médio, MS(P)E – *mean square (prediction) error*; o erro absoluto médio, MA(P)E – *mean absolute (prediction) error*, e o erro absoluto percentual médio, MAPE – *mean absolute percentual error*.

Para calculá-los, deve-se deixar uma porção de observações fora da amostra. Por exemplo, em uma amostra com $T+H$ observações, deixam-se as últimas H observações fora da amostra e estima-se o modelo com as T observações restantes. Assim, essas estatísticas são calculadas da seguinte forma:

$$MSE_{t,H} = \sqrt{\frac{\sum_{h=1}^H e_t(h)^2}{H}};$$

$$MAE_{t,H} = \frac{\sum_{h=1}^H |e_t(h)|}{H};$$

$$MAPE_{t,H} = \sum_{h=1}^H \left| \frac{e_t(h)}{Hy_{t+h}} \right|.$$

Se a amostra for suficientemente grande, pode-se deixar de 1/4 a 1/3 fora da amostra por razões de previsão. Na metodologia até aqui proposta, não se usa a informação adicional para realimentar as previsões. Isto é, para prever, por exemplo, y_{T-h+2} , usa-se $E_t(y_{T-h+1})$. A isso, denomina-se previsão dinâmica, cuja variância acaba sendo maior. Quando se usa na previsão de y_{T-h+s} a observação $y_{T-h+s-1}$, fala-se da previsão estática. Em qualquer caso, os coeficientes estimados foram calculados com as observações até T .

Nota 3.15 Na prática, a cada nova observação reestimam-se os coeficientes e refazem-se todas as previsões.

No exemplo a seguir, usa-se o modelo estimado para calcular essas estatísticas, supondo previsão estática e dinâmica. Ficam fora da amostra as últimas 50 observações. Comparam-se os resultados entre o modelo degenerado, equivocado e denotado por $MA((2))$, e o $MA(2)$ convencional.

	$MA((2))$		$MA(2)$	
	Dinâmico	Estático	Dinâmico	Estático
MSE	1,316	1,131	1,330	1,094
MAE	1,084	0,964	1,091	0,908
MAPE	193,35%	119,96%	193,15%	129,11%

Primeiro observe que o modelo estático gera sempre os melhores resultados, por se tratar de um modelo que prevê um passo à frente usando a última observação do período anterior. Observe que o modelo degenerado, ainda que errado, às vezes é superior ao modelo correto. Isso ocorre porque o modelo degenerado é mais parcimonioso. Assim, há uma compensação entre viés e variância de previsão. O modelo degenerado pode ser mais viesado, porém a variância de previsão tende a ser menor em termos de erro quadrático.

Nota 3.16 O modelo mais parcimonioso tem menos parâmetros estimados. Portanto, a incerteza resultante da estimação tende a ser menor.

3.9

Sazonalidade – $ARMA(p,q)(P,Q)_s$

Séries econômicas podem ser sazonais, em decorrência de safras agrícolas, férias, clima e datas especiais, como Natal. Há casos em que se deve estimar o modelo dessazonalizando cada uma das variáveis utilizadas. Em outros casos, deve-se estimar o modelo e corrigir simultaneamente a sazonalidade. Em geral, dessazonalizam-se as variáveis quando se misturam variáveis sazonais com outras não sazonais. Corrige-se a sazonalidade simultaneamente com estimação, por exemplo, no caso de séries univariadas.

É preciso cuidado para não dessazonalizar séries sem sazonalidade, pois o procedimento pode distorcer artificialmente a série. Por exemplo: taxas de juros

não devem ser dessazonalizadas. Por outro lado, a inflação possui sazonalidade. Assim, se for relacionar taxa de juros à inflação, esta deve estar depurada de fatores sazonais.

Há várias formas de tratar a sazonalidade. Tradicionalmente, usam-se *dummies* para os períodos sazonais. Em séries temporais, usam-se métodos um pouco mais elaborados. O problema, na prática, é que as séries sazonais tornam as funções FAC e FACP bastante complicadas.

Para adquirir intuição sobre o que ocorre, considere o seguinte modelo sazonal bem simples:

$$y_t = \phi_4 y_{t-4} + \epsilon_t.$$

A FAC desse modelo será tal que: $\rho_i = (\phi_4)^{\frac{i}{4}}$, quando $\frac{i}{4}$ for um número inteiro, exibindo assim um padrão de decaimento exponencial em defasagens múltiplas de 4. A FACP vai apresentar correlação parcial diferente de zero somente na defasagem 4.

Quando o modelo sazonal é dado por:

$$y_t = \epsilon_t + \theta_4 \epsilon_{t-4},$$

ocorre o inverso. A FAC é truncada na defasagem 4, e a FACP tem decaimento sazonal nas defasagens múltiplas de 4. A Figura 3.15 dá uma idéia do que se procura.

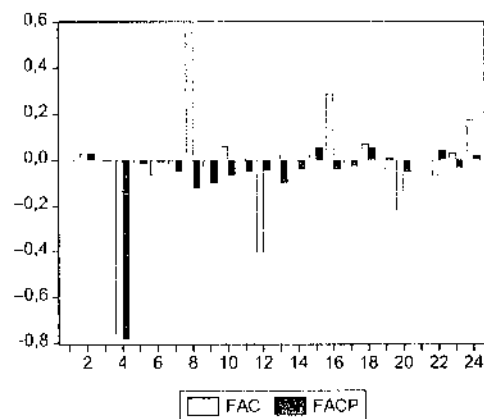


Figura 3.15 *MA (4) degenerado – Autocorrelação.*

Podem-se verificar dois tipos de sazonalidade em séries temporais que interagem com padrões não sazonais. A primeira é a *sazonalidade aditiva*. Por exemplo, um $ARMA(1,1)$ poderá ser sazonal na defasagem 4 via coeficiente auto-regressivo ou de médias móveis:

$$y_t = \phi_1 y_{t-1} + \phi_4 y_{t-4} + \epsilon_t + \theta_1 \epsilon_{t-1} \text{ ou}$$

$$y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_4 \epsilon_{t-4}.$$

Testes de Autocorrelação						
Defasagem	1	6	12	18	19	24
FAC	0,001	-0,066	-0,406	0,070	-0,040	0,175
FACP	0,001	-0,007	0,047	0,052	0,009	0,021
Q	0,0002*	308,44*	553,52*	601,13*	601,97*	646,26*

* Indica significância a 1%.

O padrão de identificação dessa sazonalidade é o mesmo que o anterior. A complicação decorre da interação com componentes não sazonais. Veja, na Figura 3.16, como é difícil identificar a sazonalidade de um modelo do tipo $ARMA(1, (1, 4))$, em que o segundo parêntese interno define as ordens do MA degenerado pelo fato de ser sazonal.

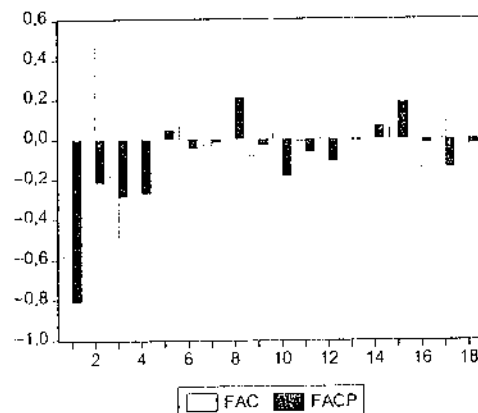


Figura 3.16 $ARMA(1, (1, 4))$ – Autocorrelações.

Testes de Autocorrelação					
Defasagem	1	6	12	18	24
FAC	-0,80	0,08	0,02	-0,07	-0,03
FACP	-0,80	-0,04	-0,11	-0,03	0,04
Q	100,35*	218,21*	220,30*	229,49*	242,73*

* Indica significância a 1%.

O outro tipo é a *sazonalidade multiplicativa*. Ainda supondo sazonalidade de ordem 4, no caso multiplicativo de um AR e de um MA, resultaria, respectivamente:

$$(1 - \phi_1 L) (1 - \phi_4 L^4) y_t = (1 + \theta_1 L) \epsilon_t \Rightarrow$$

$$y_t = \phi_1 y_{t-1} + \phi_4 y_{t-4} + \phi_1 \phi_4 y_{t-5} + \epsilon_t + \theta_1 \epsilon_t;$$

$$(1 - \phi_1 L) y_t = (1 + \theta_4 L) (1 + \theta_1 L^4) \epsilon_t \Rightarrow$$

$$y_t = \phi_1 y_{t-1} + \epsilon_t + \theta_1 \epsilon_t + \theta_4 y_{t-4} + \theta_1 \theta_4 \epsilon_{t-5}.$$

A Figura 3.17 mostra um $ARMA(p, q)(P, Q)_s$, em que p e q dão as ordens dos coeficientes não sazonais, P é o número de coeficientes sazonais auto-regressivos multiplicativos, Q é o número de coeficientes sazonais de médias móveis multiplicativos e s é o período sazonal. Por exemplo:

$$\begin{aligned} ARMA(2, 1)(1, 2)_{12} : (1 - \phi_1 L - \phi_2 L^2) (1 - \phi_{12} L^{12}) y_t \\ = (1 - \theta_1 L) (1 - \theta_{12} L^{12} - \theta_{24} L^{24}) \epsilon_t. \end{aligned}$$

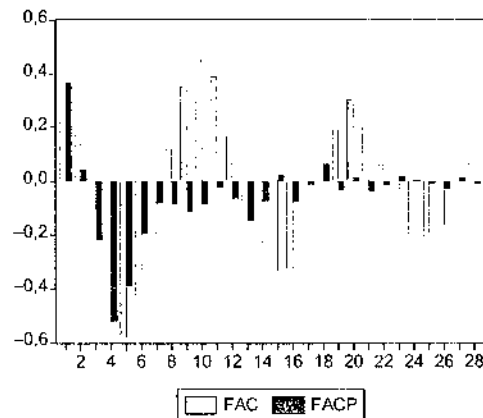


Figura 3.17 $ARMA(1, 1)(1, 0)_4$ – Autocorrelação.

Para o caso de um $ARMA(1, 1)(1, 0)_4$, ou seja, a sazonalidade é de ordem 1, definida a cada 4 períodos, veja a Figura 3.17. É possível desconfiar que há sazonalidade multiplicativa na defasagem 4 proveniente de um AR, porque ao redor de 4 as autocorrelações 5 e 3 parecem ser significativas. Quanto a saber que se trata de um $ARMA(1, 1)$, a análise fica toda prejudicada visualmente, e o

econometrista vai ter de usar tentativa e erro. Com a experiência, o econometrista rapidamente acertaria o modelo.

<i>Testes de Autocorrelação</i>					
<i>Defasagem</i>	1	6	12	18	24
<i>FAC</i>	0,367	-0,436	0,164	0,111	-0,196
<i>FACP</i>	0,367	0,195	-0,065	0,064	0,005
<i>Q</i>	67,171*	493,73*	776,51*	944,44*	1057,6*

* Indica significância a 1%.

4

Processos Não Estacionários

Quando a série temporal não é estacionária, não se pode estimá-la trivialmente. O problema essencial já foi exposto: é impossível estimar todos os momentos da série. Em particular, é impossível fazer inferências estatísticas. Para visualizar melhor a idéia, lembre-se de que a variância não condicional de um AR (1) é:

$$\text{var}(y_t) = \frac{1}{1 - \phi^2}.$$

Ora, se $\phi = 1$, o que caracteriza uma série não estacionária de raiz unitária, a variância explode. Todavia, pode-se resolver esse problema. Em séries univariadas, o procedimento básico é diferenciar a série tantas vezes quantas sejam necessárias para estacionarizá-la. O caso mais comum é encontrar uma série com raiz unitária, de forma que basta a primeira diferença dessa série para estacionarizá-la, entretanto mais diferenças poderão ser necessárias.

De modo geral, uma série econômica tem três partes distintas:

$$y_t = \text{tendência} + \text{componente estacionário} + \text{ruído}.$$

A seguir, definem-se algumas particularidades de não-estacionariedade, distinguindo os tipos de tendências que se podem encontrar.

4.1

Tendência Estacionária e Estocástica

Suponha o seguinte modelo:

$$y_t = y_0 + \delta t + \psi(L) \epsilon_t.$$

Esse modelo é chamado *tendência estacionária*, porque flutua em torno de uma tendência determinística, sem jamais se distanciar de tal tendência. Outra forma comum de ver isso é observando que a subtração da tendência determinística gera um componente estacionário.

Por outro lado, a série também poderia ser estacionarizada pela primeira diferença, isto é:

$$\Delta y_t \equiv (1 - L) y_t = y_t - y_{t-1} = \delta + (1 - L) \psi(L) \epsilon_t.$$

Essa diferenciação estacionariza a série, entretanto introduz ruído por tornar o erro não invertível. Logo, se uma série é tendência estacionária, é melhor estimá-la usando a variável explicativa t .

Considere agora o seguinte modelo, em que se supõe que as variações esperadas de y são dadas por:

$$\Delta y_t = \delta + \epsilon_t \implies y_t = y_{t-1} + \delta + \epsilon_t.$$

Compondo recursivamente y , obtém-se:

$$y_t = y_0 + \delta t + \sum_{i=1}^t \epsilon_i.$$

Veja que a variável aleatória y é dada pela composição de todos os choques havidos, $\sum_{i=1}^t \epsilon_i$. Define-se tal série como *tendência estocástica* ou *diferença estacionária*, pois a diferença da série é estacionária. Nesse caso, os choques produzem mudanças permanentes na série y , ainda que aleatórias.

Séries cuja tendência é estocástica são chamadas *séries integradas*. A ordem, d , de integração, depende do número de somatórios dos erros. Séries integradas são denotadas por $I(d)$. Séries integradas com erros estacionários são chamadas séries *ARIMA* (p, d, q) . Diferenciando d vezes a série, obtém-se uma série estacionária.

No caso anterior, diz-se que a série é integrada de ordem 1, $I(1)$, porque os erros são todos somados uma única vez, como indica a presença do somatório. Para entender melhor a idéia, é conveniente ver como isso funciona em uma série $I(2)$.

Uma série $I(2)$ precisa ser diferenciada duas vezes para ser estacionarizada, isto é:

$$\begin{aligned}\Delta^2 y_t &= \Delta \Delta y_t \\ &= \Delta (y_t - y_{t-1}) \\ &= y_t - y_{t-1} - (y_{t-1} - y_{t-2}) \\ \Delta^2 y_t &= y_t - 2y_{t-1} + y_{t-2}.\end{aligned}$$

É conveniente notar outra forma de derivar a mesma relação. Uma série integrada de ordem 2 tem duas raízes unitárias, significando que:

$$(1 - L)^2 y_t = (1 - 2L + L^2) y_t = y_t - 2y_{t-1} + y_{t-2}.$$

Ora, suponha que a série $I(2)$ seja um ruído branco, de forma que:

$$\begin{aligned}(1 - L)^2 y_t &= \epsilon_t \Rightarrow \\ y_t &= 2y_{t-1} - y_{t-2} + \epsilon_t.\end{aligned}$$

Pode-se agora ver como a série fica expressa usando substituição recursiva. Para isso, inicialmente, defina $(1 - L) y_t = x_t$. Então, tem-se:

$$\begin{aligned}x_t &= x_{t-1} + \epsilon_t \Rightarrow \\ x_t &= x_1 + \sum_{i=2}^t \epsilon_i.\end{aligned}$$

Substituindo x_t pela definição, encontra-se:

$$\begin{aligned}y_t &= y_{t-1} + y_{t-1} - y_0 + \sum_{i=2}^t \epsilon_i \\ &= y_{t-2} + 2y_{t-1} - 2y_0 + \sum_{i=2}^t \epsilon_i + \sum_{i=2}^{t-1} \epsilon_i \\ &= \dots \\ y_t &= ty_1 - (t-1)y_0 + \sum_{j=2}^t \sum_{i=2}^j \epsilon_i, \quad t \geq 2.\end{aligned}$$

Nota 4.1 O duplo somatório sobre ϵ significa $I(2)$.

A seguir, definem-se algumas variações dos processos apresentados nesta seção.

4.2

Passeios Aleatórios

Esta seção mostra as propriedades de alguns processos não estacionários. É interessante comparar essas propriedades com as dos processos estacionários e verificar que as condições de estacionaridade não se verificam.

No caso $I(1)$ com $\delta = 0$, define-se o *passeio aleatório* ou *tendência estocástica pura* pela equação:

$$y_t = y_{t-1} + \epsilon_t.$$

De forma interessante, a previsão condicional j passos à frente é dada pela observação atual, isto é:

$$E_t(y_{t+j}) = y_t + \sum_{s=1}^j E_t(\epsilon_{t+s}) = y_t.$$

A covariância é dependente do tempo:

$$\text{Var}(y_t) = \text{Var}\left(\sum_{j=1}^t \epsilon_j\right) = t\sigma^2;$$

$$\text{Cov}(y_t, y_{t-j}) = E\left(\sum_{j=1}^t \epsilon_j\right)\left(\sum_{s=1}^{t-j} \epsilon_s\right) = (t-j)\sigma^2.$$

A correlação não é direta. Divide-se a covariância pelo produto do desvio-padrão em t e $t-j$:

$$\rho_j = \frac{(t-j)\sigma^2}{\sqrt{t\sigma}\sqrt{(t-j)\sigma}} = \sqrt{\frac{t-j}{t}} = \sqrt{1 - \frac{j}{t}}.$$

Nota 4.2 Em um processo não estacionário, a autocorrelação demora a cair, pois j/t se reduz lentamente.

Ao adicionar um *drift* ao modelo, encontra-se o *passeio aleatório com drift*:

$$\begin{aligned} y_t &= y_{t-1} + \delta + \epsilon_t \\ &= y_0 + \delta t + \sum_{i=1}^t \epsilon_i. \end{aligned}$$

Veja que, nesse caso, o comportamento de y_t depende de um componente determinístico e de outro estocástico. Além disso, é fácil calcular a previsão j passos à frente:

$$E_t(y_{t+j}) = y_t + \delta j + \sum_{s=1}^j E_t(\epsilon_{t+s}) = y_t + \delta j.$$

É possível generalizar o modelo de passeio aleatório adicionando um ruído a ele. É o *passeio aleatório com ruído*:

$$y_t = y_0 + \sum_{j=1}^t \epsilon_j + \eta_t,$$

em que

$\{\eta_t\}$ é um ruído branco;

$\epsilon_t \perp \eta_{t-j}$.

Pode-se, com isso, encontrar que:

$$\Delta y_t = \epsilon_t + \Delta \eta_t.$$

A importância desse modelo é ser $I(1)$ com uma correlação menor que naquele passeio aleatório puro, em razão da presença de σ_η^2 . Isso é caracterizado no cálculo da covariância feito a seguir.

$$\text{Var}(y_t) = \text{Var}\left(\sum_{s=1}^t \epsilon_s + \eta_t\right) = t\sigma^2 + \sigma_\eta^2;$$

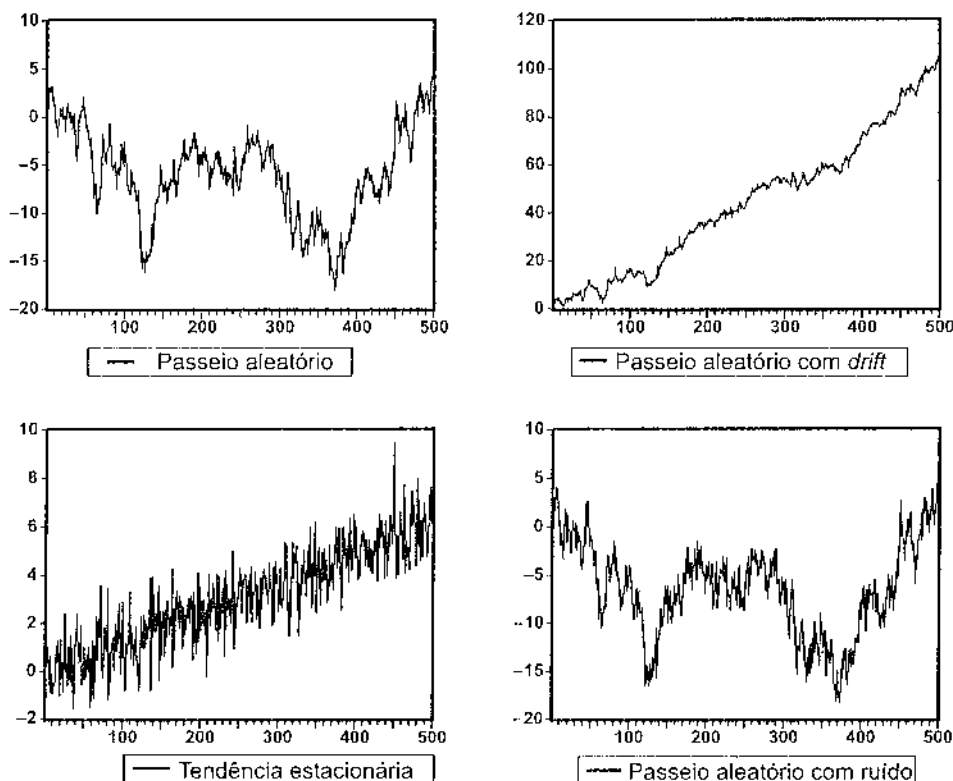
$$\text{Cov}(y_t, y_{t-j}) = E\left(\sum_{s=1}^t \epsilon_s + \eta_t\right)\left(\sum_{s=1}^j \epsilon_s + \eta_{t-j}\right) = (t-j)\sigma^2;$$

$$\rho_s = \frac{(t-j)\sigma^2}{\sqrt{(t\sigma^2 + \sigma_\eta^2)[(t-j)\sigma^2 + \sigma_\eta^2]}} < \sqrt{1 - \frac{j}{t}}.$$

A Figura 4.1 mostra os vários modelos discutidos até o momento.

O modelo mais geral possível inclui tendência determinística e estocástica e resíduos que seguem um processo $ARMA(p, q)$. O modelo é chamado *tendência geral mais componente irregular*:

$$y_t = \delta t + y_0 + \sum_{i=1}^t \epsilon_i + \psi(L)\eta_t. \quad (6)$$

Figura 4.1 *Séries temporais com tendência.*

4.3

Removendo a Tendência

No modelo com tendência estocástica, basta diferenciá-lo para estacionarizá-lo, inclusive se houver tendência determinística. De fato, tome a diferença da equação (6), para obter:

$$\Delta y_t = \delta + \epsilon_t + \psi(L) \Delta \eta_t.$$

É claro que, se y_t for integrado de ordem d , torna-se a d -ésima diferença para estacionarizá-lo. Porém, mais importante é saber como estimar uma série que é tendência determinística. O procedimento é simples e descrito a seguir.

No caso em que y_t é uma função do tempo, constituindo uma série de tendência determinística, o procedimento é estimar y_t contra o tempo e armazenar os resíduos. Os resíduos armazenados constituem nova série que deverá ser modelada de modo separado. Formalmente:

1. Estime por mínimos quadrados ordinários:

$$y_t = \delta_0 + \delta_1 t + \delta_2 t^2 + \cdots + \delta_n t^n + e_t,$$

em que $e_t = \psi(L)\epsilon_t$, e obtenha os resíduos estimados: \hat{e}_t .

2. Estime o modelo $ARMA(p, q)$ a partir dos resíduos estimados, conforme o capítulo anterior.

Para determinar n , use os testes estatísticos convencionais como t , F ou AIC/BIC .¹ Em geral, estima-se o modelo com um n máximo, predeterminado, n_{\max} . Se o teste de t sobre $\delta_{n_{\max}}$ não é significativo, retira-se t^n e estima-se o modelo até t^{n-1} , repetindo o teste t -student. Procedo-se assim até rejeitar que $\delta_{n-1} = 0$. Em seguida, confirme se os coeficientes conjuntamente são TODOS iguais a zero, usando o teste de F . Reconfirme com AIC e BIC .

Nota 4.3 É proibido diferenciar uma série que é tendência estacionária, porque isso adiciona ruído à série original. Da mesma forma, é proibido estimar uma série que é tendência estocástica usando tendência determinística, porque isso não elimina a tendência estocástica.

4.4 Regressão Espúria

A necessidade de teoria econômica para definir variável explicada e explicativa torna-se muito importante na presença de raiz unitária. Podem-se encontrar relações econométricas entre duas ou mais variáveis econômicas sem qualquer relação de causalidade entre uma e outra por puro acaso. Isso é particularmente comum quando se trata de séries não estacionárias. Por exemplo, a regressão de uma variável $I(1)$ com outra $I(1)$ obtida independentemente gera alto R^2 e significativo t -estatístico. Contudo, o resultado é sem significado econômico.

Para ver isso, considere a seguinte experiência. Gere duas séries $I(1)$ independentemente uma da outra e regreda uma contra a outra. Qual resultado você

¹ Na presença de componentes unicamente determinísticos e estacionários, os testes convencionais são válidos. Na presença de tendência estocástica, muitos problemas surgem. Veja uma excelente discussão sobre isso em Hamilton (1994, capítulo 16).

obtem? Em 75% das vezes, parecerá que elas são correlacionadas (veja Patterson, 2000, para exemplos). Considere um exemplo:

$$y_t = y_{t-1} + \epsilon_{y,t};$$

$$z_t = z_{t-1} + \epsilon_{z,t}.$$

A Figura 4.2 mostra duas séries geradas independentemente uma da outra, como no modelo anterior.

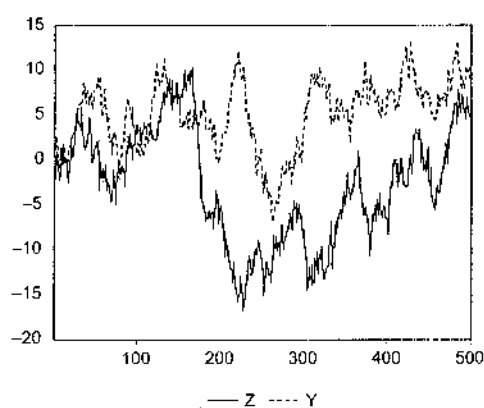


Figura 4.2 Séries com relação espúria.

Agora, regride-se y_t contra z_t :

$$y_t = \alpha + \beta z_t + e_t.$$

Nas simulações com 300 observações, com a amostra entre 201 e 300, obteve-se o seguinte resultado:

$$y_t = -6,37 - 0,770z_t + e_t, \quad R^2 = 0,30,$$

(1,544) (0,142)

em que o desvio-padrão está entre parênteses.

O R^2 não é muito baixo, e os coeficientes são altamente significantes. Isso é tipicamente uma regressão espúria. Assim, em resumo, tem-se o seguinte:

- a) Se $\{y_t\}$ e $\{z_t\}$ são ambos estacionários, a regressão convencional se aplica sem problemas;
- b) Se $\{y_t\}$ e $\{z_t\}$ são integrados de diferentes ordens, regredir um contra o outro é espúrio e sem significado;

- c) Se $\{y_t\}$ e $\{z_t\}$ são integrados de mesma ordem, e o resíduo ainda é integrado, a regressão é espúria;
- d) Se $\{y_t\}$ e $\{z_t\}$ são integrados de mesma ordem, e os resíduos são estacionários, há cointegração que será objeto de discussão mais adiante.

A conclusão é ter cuidado com a regressão que se faz. A especificação de um modelo econômico freqüentemente ajuda a não cometer absurdos. Os melhores modelos, em geral, geram regressões que fazem sentido, com resíduos estacionários.

Estatisticamente, pode-se mostrar que, quando se faz uma regressão espúria conforme exposto:

- a) α e β não convergem em probabilidade para uma constante;
- b) as estatísticas $t_{\hat{\alpha}}$ e $t_{\hat{\beta}}$ também não convergem em probabilidade para nenhuma distribuição, inexistindo, pois, valores críticos;
- c) o teste de Durbin-Watson converge para zero;
- d) o R^2 tem distribuição assintótica degenerada e
- e) os coeficientes da autocorrelação dos resíduos convergem para 1.

Uma discussão mais detalhada sobre esse ponto pode ser vista em Patterson (2000).

4.5

Testes de Raiz Unitária

4.5.1 Dickey-Fuller

A inspeção visual de uma série raramente permite distingui-la como de tendência estocástica ou tendência determinística. E, freqüentemente, quando se age assim, incorre-se em equívocos. Então, foram desenvolvidos testes para a verificação de existência de raízes unitárias. O primeiro foi feito por Dickey e Fuller (1979, 1981).

Para entender o teste, considere o seguinte modelo:

$$y_t = \phi y_{t-1} + \epsilon_t.$$

A tendência inicial é estimar esse modelo e usar um teste convencional de t sobre ϕ , tendo como hipótese nula $H_0 : \phi = 1$. Em geral, os pacotes econométricos

reportam os testes nos coeficientes contra a nula de serem iguais a zero. Então, o teste poderia ser alterado, subtraindo y_{t-1} de ambos os lados:

$$\Delta y_t = (\phi - 1) y_{t-1} + \epsilon_t = \alpha y_{t-1} + \epsilon_t, \quad (7)$$

em que se define $\alpha \equiv \phi - 1$.

Assim, $H_0 : \phi = 1$ é equivalente a $H_0 : \alpha = 0$. O problema é que, sob a nula, a distribuição do teste, infelizmente, não é convencional, ou seja, não é igual à distribuição t estatística, pois y_t não é estacionário. A distribuição do teste pode ser encontrada analiticamente usando o teorema central do limite funcional e pode ser visualizada por meio de experimentos de Monte Carlo. A forma mais básica de visualização consiste em montar o seguinte experimento:

1. Gere uma sequência de erros normais com esperança nula e variância σ^2 , $\{\epsilon_t\}$, com $T + n$ observações;
2. Gere a sequência $\{y_t\}$ sob a hipótese nula de raiz unitária, dado y_0 ;
3. Estime a equação (7) usando as T últimas realizações² e armazene o valor da estatística t ;
4. Retorne a 1 S vezes (em geral, $S \geq 10.000$);
5. Faça o gráfico da distribuição da estatística t .

A Figura 4.3 mostra a distribuição empírica dessa estatística em que $T = 100$, $n = 50$, $S = 10.000$ e $y_0 = 0$. Note que a distribuição é assimétrica, concentrada em valores negativos.

Procedendo dessa maneira, deve-se concluir que a média da estatística t não é zero, como poderia ser esperado de uma distribuição convencional. De fato, em 10% das vezes, a estatística $t < -1,60$; em 5%, $t < -1,95$; e em 1%, $t < -2,60$. Ou seja, o uso da estatística t olhando para a tabela convencional implicaria cometer o erro do tipo I³ com muito mais frequência.

Dessa forma, Dickey e Fuller (1979) recalcularam o valor da estatística t . O valor dessa estatística se altera, conforme se define a equação de regressão e segundo o tamanho da amostra. Dickey e Fuller usaram as seguintes equações

² Assim, atenua-se o efeito de y_0 até a n -ésima primeira realização.

³ Erro do tipo I é rejeitar a nula quando ela é verdadeira.

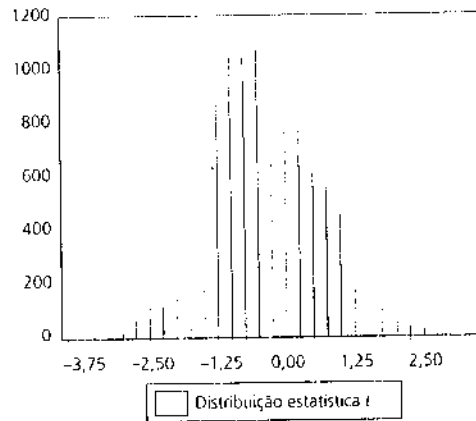


Figura 4.3 Distribuição da estatística t-student sob $H_0 : \phi = 1$.

de estimação e suas respectivas estatísticas, dependendo se consideravam a existência de *drift* e tendência determinística:

$$\Delta y_t = \alpha y_{t-1} + \epsilon_t \rightarrow \tau;$$

$$\Delta y_t = \mu + \alpha y_{t-1} + \epsilon_t \rightarrow \tau_\mu;$$

$$\Delta y_t = \mu + \delta t + \alpha y_{t-1} + \epsilon_t \rightarrow \tau_\tau.$$

Sob $H_0 : \alpha = 0$, as três estatísticas associadas foram obtidas por meio de simulações de Monte Carlo. A primeira estatística desconsidera a existência de *drift* e tendência determinística; a segunda desconsidera a existência de tendência determinística. Na terceira, há um intercepto e tendência determinística⁴, $\mu + \delta t$. Repetiram o experimento para $T = 25, 50, 100, 250, 300, \infty$, e, do ponto de vista prático, ∞ significa $T \geq 1.000$.

Como calcular a estatística do teste de Dickey e Fuller (1979, 1981)? Supondo o modelo sem *drift* ou tendência determinística, o cálculo de $\hat{\tau}$ segue, essencialmente, os passos de um teste convencional de estatística.

Primeiro, supondo uma série com $T + 1$ observações, $y_0, y_1, y_2, \dots, y_T$, calcule a regressão por mínimos quadrados ordinários e subtraia 1 do parâmetro ϕ , para

⁴ A terceira equação, $\Delta y_t = \mu + \delta t + \alpha y_{t-1} + \epsilon_t$, sob a hipótese nula de $\alpha = 0$, pode ser escrita como:

$$y_t = \mu + \delta t + y_{t-1} + \epsilon_t \Rightarrow$$

$$y_t = y_0 + \left(\mu + \frac{\delta}{2} \right) t + \frac{\delta}{2} t^2 + \sum_{i=1}^t \epsilon_i.$$

Ou seja, y_t é tendência quadrática determinística e tendência estocástica.

proceder ao teste sob $H_0 : \alpha = 0$:

$$\hat{\alpha} = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} - 1.$$

Em seguida, calcule a variância amostral:

$$S^2 = \frac{1}{T-1} \sum_{t=1}^T (\Delta y_t - \hat{\alpha} y_{t-1})^2.$$

Calcule, a seguir, o desvio-padrão do coeficiente $\hat{\alpha}$, $s(\hat{\alpha})$:

$$s(\hat{\alpha}) = \frac{S}{\sqrt{\sum_{t=1}^T y_{t-1}^2}}.$$

Finalmente, obtenha o valor calculado da estatística t :

$$\hat{t} = \frac{\hat{\alpha}}{s(\hat{\alpha})}.$$

O critério de rejeição é se $\hat{t} < \tau$, em que τ são os valores críticos obtidos por meio de experimentos de Monte Carlo; então rejeita-se a nula e conclui-se que a série não possui raiz unitária. Por outro lado, não rejeitar a nula significa admitir que a série possua uma raiz unitária.

4.5.2 Dickey-Fuller Aumentado

O problema do teste anterior é que Dickey e Fuller (1979) consideraram o erro como um ruído branco. Mas, frequentemente, o erro é um processo estacionário qualquer. Esse problema pode causar distorções no poder do teste. Felizmente, é fácil corrigir esse problema.

De fato, suponha que y_t seja um processo auto-regressivo de ordem p , com raiz unitária:

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_{p-1} y_{t-p+1} + \phi_p y_{t-p} + \epsilon_t.$$

Como testar esse modelo para raiz unitária? Usa-se o que se convencionou chamar teste de Dickey-Fuller aumentado. A idéia é estimar o modelo com as variáveis auto-regressivas. Essa é uma forma de corrigir o desvio do valor correto da estatística, ou seja, intuitivamente trata-se de encontrar os desvios de y_t em relação à sua “média”, para deslocar a distribuição de α em direção a zero, caso a hipótese nula seja verdadeira.

Do ponto de vista prático, é preciso introduzir tantas variáveis auto-regressivas quantas forem necessárias para que o teste de resíduos não rejeite a hipótese de que se trata de um ruído branco. Para entender o teste, adicione e subtraia $\phi_p y_{t-p+1}$ à equação anterior:

$$\begin{aligned} y_t &= \mu + \phi_1 y_{t-1} + \cdots + \phi_{p-1} y_{t-p+1} + \phi_p y_{t-p} + \phi_p y_{t-p+1} - \phi_p y_{t-p+1} + \epsilon_t \\ &= \mu + \phi_1 y_{t-1} + \cdots + (\phi_{p-1} + \phi_p) y_{t-p+1} - \phi_p \Delta y_{t-p+1} + \epsilon_t. \end{aligned}$$

Utilizando o mesmo procedimento, desta vez com $(\phi_{p-1} + \phi_p) y_{t-p+2}$:

$$\begin{aligned} y_t &= \mu + \phi_1 y_{t-1} + \cdots + (\phi_{p-1} + \phi_p) y_{t-p+2} - (\phi_{p-1} + \phi_p) y_{t-p+2} \\ &\quad + (\phi_{p-1} + \phi_p) y_{t-p+1} - \phi_p \Delta y_{t-p+1} + \epsilon_t \\ &= \mu + \phi_1 y_{t-1} + \cdots + (\phi_{p-2} + \phi_{p-1} + \phi_p) y_{t-p+2} \\ &\quad - (\phi_{p-1} + \phi_p) \Delta y_{t-p-2} - \phi_p \Delta y_{t-p+1} + \epsilon_t. \end{aligned}$$

Repetindo isso p vezes, obtém-se ao final:

$$\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i} + \epsilon_t,$$

em que

$$\alpha = - \left(1 - \sum_{i=1}^p \phi_i \right); \quad \lambda_i = - \sum_{j=i}^p \phi_{j+1}.$$

O teste então pode ser feito, usando os mesmos valores críticos encontrados por Dickey e Fuller. Ou seja, desde que se faça a correção do modelo, de forma a considerar as demais variáveis defasadas, em nada se alteram os valores críticos do teste ou a interpretação do modelo.

E se o modelo for ainda mais complexo, com termos de médias móveis, o que fazer? O procedimento anterior pode ser adotado mesmo no caso de um modelo $ARIMA(p, 1, q)$, já que sempre se pode transformar um $MA(q)$ em um $AR(\infty)$. Surge, daí, o problema de como estimar um modelo de infinitas defasagens. Sucede que se provou que um modelo $ARIMA(p, 1, q)$ pode ser bem aproximado por um $ARIMA(n, 1, 0)$, em que $n \leq T^{1/4}$ (veja Said e Dickey, 1984).

Nota 4.4 Não obstante a adição de componentes auto-regressivos ao modelo originalmente proposto por Dickey e Fuller, experimentos de Monte Carlo mostraram que o valor da estatística t permanece inalterado.

Mais tarde, MacKinnon (1991, 1996) desenvolveu uma metodologia bastante interessante para obter a estatística t com qualquer número de observações. O Eviews reporta os resultados de 1996, que são baseados em simulações cujo erro é *i.i.d.*, pois o programa admite que a correção paramétrica por meio de inclusão de componentes auto-regressivos não deve alterar os valores críticos.

O leitor poderá perguntar-se como escolher a ordem p para executar o teste de raiz unitária. O número de defasagens a incluir deve ser tal que os resíduos estimados sejam um ruído branco, de acordo com testes discutidos no Capítulo 3. Poucas defasagens não “branqueiam” os resíduos, afetando a distribuição de α . Excesso de defasagens diminui sobremaneira o poder do teste em direção da rejeição da hipótese nula por reduzir o número de graus de liberdade. Assim, para definir p , existem duas possibilidades equivalentes:

1. A primeira é acrescentar o número de defasagens suficientes para encontrar resíduos que sejam isentos de autocorrelação. Para saber se o resíduo é isento de autocorrelação, deve-se proceder ao teste de Ljung-Box.
2. A segunda possibilidade é fixar um p_{\max} relativamente alto. Em seguida, estimar o modelo por mínimos quadrados ordinários para $p_{\max}, p_{\max} - 1, \dots, 0$ e coletar os valores de algum dos critérios de informação como Hannan-Quinn, Schwarz ou Akaike ou utilizando testes estatísticos convencionais até que se rejeite a hipótese nula, usando como nível de significância 20%.

No primeiro caso, deve-se escolher o valor da defasagem que resulta no menor Schwarz ou Akaike e confirmar o resultado verificando se os resíduos comportam-se como um ruído branco.⁵

A segunda forma é válida porque Sims, Stock e Watson (1990) mostraram que os coeficientes das variáveis estacionárias convergem para uma distribuição normal. Contudo, adverte-se para não executar o procedimento de modo inverso, pois experimentos de Monte Carlo resultam na escolha de uma defasagem menor que a ideal, quando se inicia o processo de escolha de um modelo mais parcimonioso para um mais extenso.

⁵ O Eviews 5.x executa o procedimento de escolha do p ideal automaticamente. Porém, não verifica se os resíduos gerados por essa defasagem são ruídos brancos.

Como definir o p_{\max} ? Um critério de escolha do máximo p foi proposto por Schwert (1989), conforme a seguinte fórmula:

$$p_{\max} = 12 \operatorname{int} \left[\left(\frac{T}{100} \right)^{\frac{1}{3}} \right],$$

em que $\operatorname{int}(x)$ é a parte inteira de x .

Assim, uma série com 100 observações teria um p_{\max} de 12 defasagens. Outra série com 200 observações teria 14 defasagens, no máximo.

Exemplo 4.1

Para exemplificar os testes de raiz unitária discutidos, serão utilizados, neste exemplo e nos seguintes, dados da série da taxa de câmbio mensal de julho de 1997 até junho de 2007 e da série do IPCA abrangendo o período de janeiro de 1995 até maio de 2007. Em cada uma dessas séries não foi necessário nenhum tipo de transformação para a aplicação dos testes. A importância desses testes emerge da restrição existente em alguns modelos econométricos que só podem ser aplicados, por exemplo, em séries que sejam estacionárias. Considere a série de dólar mensal representada na Figura 4.4.

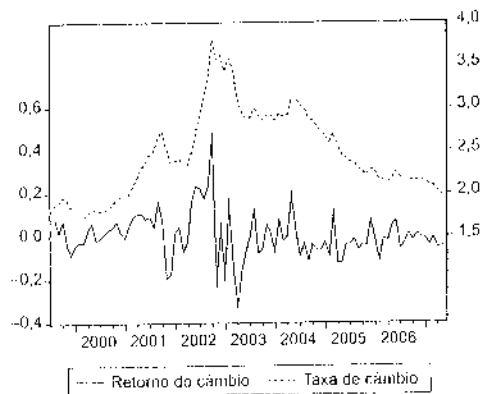


Figura 4.4 Taxa de câmbio.

É bem possível que essa série contenha uma raiz unitária. Por quê? Porque verifica-se que a série demora para reverter a tendência. Isso caracteriza um elevado componente auto-regressivo, possivelmente uma série integrada. Na parte inferior, mostra-se a série diferenciada em que se verifica que ela flutua ao redor de zero.

Mesmo assim, é mais seguro proceder ao teste de raiz unitária. O Eviews fixa o máximo $p_{\max} = 11$ e, pelo critério de Schwarz, encontra $p = 1$. Feito isso, encontra o seguinte valor da estatística, em que se especifica o modelo com constante $\hat{\tau}_\mu = -1,413$. O próprio Eviews reporta que $P(\tau_\mu < -1,676) = 57,26\%$. Logo, não é possível rejeitar a hipótese nula, e conclui-se que a série possui uma raiz unitária. A título informativo, os valores de rejeição da hipótese nula são a 1%, 5% e 10%, respectivamente iguais a $-3,50$, $-2,89$ e $-2,58$.

O segundo exemplo é o da inflação mensal, versão IPCA. É interessante verificar se há raiz unitária depois do período megainflacionário vivido pelo Brasil. A Figura 4.5 mostra o IPCA ao longo do tempo.

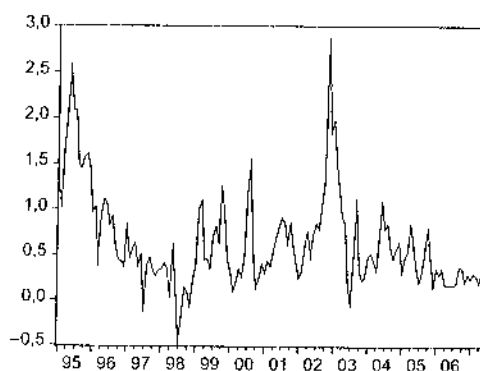


Figura 4.5 IPCA depois do Plano Real.

Não é possível ter certeza de que não existe raiz unitária a partir da figura. Procedendo ao teste de raiz unitária, como no exemplo anterior, rejeita-se a existência de raiz unitária, pois o valor da estatística $\hat{\tau}_\mu = -4,91$ é menor que o valor crítico a 1% de $-3,47$, conforme:

Teste de Dickey-Fuller		
Estatística t		$-4,917$
	1%	$-3,4749$
Valores Críticos	5%	$-2,8810$
	10%	$-2,5772$

Quando se procede ao teste usando tendência, o coeficiente dessa variável é estatisticamente insignificante, logo é desnecessário incluir a tendência no teste. Entretanto, o coeficiente da constante é sempre significativo, segundo o teste t .

4.5.3 Demais Testes de Dickey e Fuller

O teste de Dickey e Fuller pode ser feito conjuntamente para dois ou três coeficientes. Considere as seguintes especificações:

$$\Delta y_t = \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i+1} + \epsilon_t;$$

$$\Delta y_t = \mu + \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i+1} + \epsilon_t;$$

$$\Delta y_t = \mu + \delta t + \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i+1} + \epsilon_t.$$

Dickey e Fuller (1981) calcularam estatísticas F para testes conjuntos, chamando-as $\Phi_i, i = 1, 2, 3$. Essas estatísticas não têm uma distribuição convencional, daí a necessidade de renomeá-las. Hipóteses nulas e respectivas estatísticas são fornecidas a seguir:

$$H_0 : \alpha = \mu = 0 \rightarrow \Phi_1;$$

$$H_0 : \alpha = \delta = \mu = 0 \rightarrow \Phi_2;$$

$$H_0 : \alpha = \delta = 0 \rightarrow \Phi_3.$$

Essas estatísticas são construídas da mesma forma que os testes convencionais:

$$\Phi_i = \frac{(\hat{\epsilon}' \hat{\epsilon}_{\text{restrita}} - \hat{\epsilon}' \hat{\epsilon}_{\text{não restrita}}) / r}{\hat{\epsilon}' \hat{\epsilon}_{\text{não restrita}} / (T - k)},$$

em que

r é o número de restrições, igual a 2 ou 3;

T é o número de observações;

k é o número de parâmetros estimados no modelo não restrito.

Como de praxe, grandes valores de Φ_i , quando comparados aos valores simulados, sugerem a rejeição da hipótese nula.

4.5.4 Phillips-Perron

Embora o procedimento de Said e Dickey (1984) permita proceder ao teste de raiz unitária usando um modelo auto-regressivo para representar um modelo $ARIMA(p, 1, q)$, seria interessante um teste de raiz unitária que fosse especificado independentemente das ordens p e q do modelo. Afinal, a idéia é essencialmente testar o coeficiente α do processo.

É nesse contexto que surge o teste de Phillips e Perron (1988), o qual faz uma correção não paramétrica ao teste de Dickey e Fuller, permitindo que seja

consistente mesmo que haja variáveis defasadas dependentes e correlação serial nos erros.⁶ O teste de Phillips e Perron torna desnecessária, assim, a especificação de um modelo com ordem auto-regressivo suficiente para expurgar a correlação serial dos resíduos. Os detalhes de sua derivação fogem muito ao escopo desta obra e não serão apresentados. As equações estimadas e os testes designados são idênticos aos de Dickey e Fuller. Assim, a interpretação também é análoga. A maioria dos *softwares* já incorporou o teste de Phillips e Perron. Assim, considere as seguintes possibilidades com as respectivas estatísticas associadas:

$$\begin{aligned}\Delta y_t &= \alpha y_{t-1} + u_t \rightarrow z_t, \\ \Delta y_t &= \mu + \alpha y_{t-1} + u_t \rightarrow z_{t,\mu}, \\ \Delta y_t &= \mu + \delta t + \alpha y_{t-1} + u_t \rightarrow z_{t,\tau},\end{aligned}$$

em que u_t é um processo estacionário.

Phillips e Perron (1988) também definem testes diretamente sobre os coeficientes do modelo, em vez de usar a estatística t , como anteriormente. Eles chamaram tais testes de z_α , para enfatizar que são testes sobre a distribuição dos coeficientes, e comparam os resultados com os testes baseados na distribuição da estatística t , ambos sob a hipótese nula de raiz unitária. Entretanto, do ponto de vista prático, não há diferença entre eles.

A seqüência de estimação é semelhante à do teste de Dickey e Fuller mencionado anteriormente. Esquemáticamente, a correção, $z_{t,\mu}$, empregada por Phillips e Perron para τ_μ , é seqüencialmente estimada da seguinte forma, dado y_0 (o procedimento para z_t e $z_{t,\tau}$ é análogo, por isso omitido):

1. Estime as seguintes médias:

$$\bar{y} = \frac{\sum_{t=1}^T y_t}{T}, \bar{y}_{-1} = \frac{\sum_{t=1}^T y_{t-1}}{T};$$

2. Estime o parâmetro de maior interesse:

$$\hat{\alpha} = \frac{\sum_{t=1}^T (y_{t-1} - \bar{y}_{-1})(y_t - \bar{y})}{\sum_{t=1}^T (y_{t-1} - \bar{y}_{-1})^2} \dots 1;$$

⁶ Ou seja, o teste de Dickey e Fuller é um caso especial do teste de Phillips e Perron se o desvio-padrão calculado dos resíduos for numericamente igual ao desvio-padrão do resíduo de longo prazo, calculado por meio da densidade espectral da variável dependente na frequência zero (veja Phillips e Perron, 1988).

3. Estime a constante ou *drift*:

$$\hat{\mu} := \bar{y} - (\hat{\kappa} + 1) \bar{y}_{-1};$$

4. Estime a variância populacional da regressão:

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{u}_t^2}{T} = \frac{\sum_{t=1}^T (\Delta y_t - \hat{\mu} - \hat{\kappa} y_{t-1})^2}{T};$$

5. Calcule o desvio-padrão do parâmetro de interesse:

$$s(\hat{\kappa}) := \frac{\hat{\sigma}}{\sqrt{\sum_{t=1}^T y_{t-1}^2}};$$

6. Calcule a estatística de Dickey e Fuller:

$$\hat{\tau}_{\mu} = \frac{\hat{\kappa}}{s(\hat{\kappa})};$$

7. Estime a variância de longo prazo, HAC:

$$\hat{\nu}^2 = \hat{\sigma}^2 + \frac{2}{T} \sum_{j=1}^M \omega\left(\frac{j}{M+1}\right) \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j};$$

8. Calcule a estatística de Phillips e Perron:

$$\hat{z}_{t,\mu} = \hat{\tau}_{\mu} \left(\frac{\hat{\sigma}}{\hat{\nu}} \right) - \frac{1}{2} \left(\frac{\hat{\nu}^2 - \hat{\sigma}^2}{\hat{\sigma} \sqrt{T^{-2} \sum_{t=1}^T y_{t-1}^2}} \right).$$

O termo $\hat{\nu}^2$ representa a variância de longo prazo, em que estão incluídas todas as autocovariâncias do processo u_t . Seria ideal calcular $\sum_{j=-\infty}^{\infty} \gamma_j$. Ocorre que não existem infinitas observações, logo é preciso truncar j em algum ponto. É claro que se podia ir de $-T$ a T autocovariâncias, porém quanto mais distante a autocovariância, menos informação ela produz em troca de muito mais ruído, por isso é necessário calcular $\sum_{j=-M}^M \gamma_j$, em que $\lim_{M,T \rightarrow \infty} \frac{M}{T} \rightarrow 0$, isto é, a banda de truncagem cresce a uma taxa menor que o número de observações. Como $\gamma_j = \gamma_{-j}$, pode-se escrever:

$$\sum_{j=-M}^M \gamma_j = \gamma_0 + 2 \sum_{j=1}^M \gamma_j.$$

Para o cálculo amostral, então, dever-se-ia estimar:

$$\sum_{j=-M}^M \hat{\gamma}_j = \frac{\sum_{t=1}^T \hat{u}_t^2}{T} + \frac{2}{T} \sum_{j=1}^M \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j}.$$

Porém, por razões de amostras finitas, é preciso ponderar as observações mais distantes das observações mais recentes. Essa ponderação é dada pela função $\omega\left(\frac{j}{M+1}\right)$, também conhecida como função janela. Essa função pode ser especificada de várias maneiras, e as mais comuns são as seguintes:

$$\text{Bartlett: } \omega(z) = \begin{cases} 1 - |z|, & \text{se } |z| \leq 1. \\ 0, & \text{se } |z| > 1. \end{cases}$$

$$\text{Parzen: } \omega(z) = \begin{cases} 1 - 6z^2 + 6z^3, & \text{se } 0 \leq z \leq \frac{1}{2}. \\ 2(1 - z)^2, & \text{se } \frac{1}{2} \leq z \leq 1. \\ 0, & \text{do contrário.} \end{cases}$$

$$\text{Quadrática: } \omega(z) = \frac{3}{\left(\frac{6\pi}{5}z\right)^2} \left[\frac{\sin\left(\frac{6\pi}{5}z\right)}{\frac{6\pi}{5}z} - \cos\left(\frac{6\pi}{5}z\right) \right].$$

Andrews (1991) estuda as propriedades de cada uma dessas funções e suas aplicações. Perron (1990) recomenda o uso da janela de Parzen, embora grande parte dos trabalhos empíricos ainda use a janela de Bartlett.

Outro componente importante do teste é fixar M . Não é trivial definir que valor M deveria ser. Os pacotes econométricos, em geral, definem esse valor a partir da amostra em termos do critério de Newey-West (1994) ou Andrews (1991). Para os propósitos deste texto é melhor seguir o primeiro.

Definida a janela, procede-se à correção não paramétrica definida pela estatística $\hat{z}_{t,\mu}$. Ela segue princípios derivados assintoticamente e é bastante complexa, porque multiplica $\hat{\tau}_\mu$ por $\hat{\sigma}/\hat{\sigma}$ e subtrai

$$\frac{1}{2} \left(\frac{\hat{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma} \sqrt{T^{-2} \sum_{t=1}^T y_{t-1}^2}} \right).$$

A justificativa formal para essa correção foge ao escopo desta obra. Intuitivamente, é preciso corrigir um viés de estimação, decorrente do uso de um erro que segue um processo estacionário desconhecido. O termo

$$\frac{1}{2} \left(\frac{\hat{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma} \sqrt{T^{-2} \sum_{t=1}^T y_{t-1}^2}} \right)$$

é subtraído para centrar a distribuição de z_t em zero. O termo $\hat{\sigma}/\hat{\sigma}$ é multiplicado para corrigir a amplitude de distribuição do teste.

É interessante observar que, na ausência de autocovariância serial, isto é, quando $\sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j} = 0$, então o teste de Phillips e Perron é idêntico ao teste de Dickey e Fuller, pois $\hat{\sigma}^2 = \hat{\sigma}^2 \Rightarrow \hat{z}_{t,\mu} = \hat{t}_\mu$.

A seguir, exemplifica-se o teste usando as mesmas séries dos exemplos da seção anterior.

Exemplo 4.2

Considere a série da taxa de câmbio, usando o teste de Phillips e Perron. A tabela seguinte é obtida a partir da saída do Eviews. A idéia é explicar cada um dos seus termos. Para obtê-la, considerou-se o modelo com constante, a função peso, $\omega[j/(M+1)]$, segue a especificação de Parzen, o valor de M é obtido pelo critério de Newey-West.

Com isso, o Eviews gera uma saída em que se podem ler, entre outras coisas:

Valores Críticos Teóricos		Valores Estimados	
	$z_{t,\mu}$	$\hat{z}_{t,\mu}$	-1,541
1%	-3,501	Prob ($z_{t,\mu} < \hat{z}_{t,\mu}$)	50,83%
5%	-2,892	M	4
10%	-2,583	Variância Residual: $\hat{\sigma}^2$	0,0116
		HAC: $\hat{\sigma}^2$	0,2155

Como a estatística calculada é maior que os valores teóricos, não se rejeita a nula. O teste conclui que a série possui raiz unitária.

Exemplo 4.3

Refazendo o exemplo da inflação, mas usando o teste de Phillips, mesma função, peso e critério para a escolha de M , encontram-se:

Valores Críticos Teóricos		Valores Estimados	
	$z_{t,\mu}$	$\hat{z}_{t,\mu}$	-4,789
1%	-3,474	Prob ($z_{t,\mu} < \hat{z}_{t,\mu}$)	0,00%
5%	-2,881	M	5
10%	-2,577	Variância Residual: $\hat{\sigma}^2$	0,1509
		HAC: $\hat{\sigma}^2$	0,1393

Aqui, rejeita-se a nula como anteriormente, pois o valor estimado é menor que o valor crítico.

Perron (1990) aduz sérias objeções ao teste de Dickey e Fuller sob a hipótese de tendência temporal e intercepto. Assim, sugere fortemente o uso do teste de Phillips e Perron em seu lugar, porque corrige os problemas do teste.

O teste de raiz unitária é muito problemático quando α é próximo de 0, mas não é zero. Isso significa que o poder do teste (probabilidade de rejeitar a hipótese nula quando ela é falsa) é muito baixo. Em outras palavras, há uma tendência de se aceitar a hipótese nula equivocadamente. Para reduzir isso, Perron sugere alguns procedimentos:

1. Inicie o teste com a regressão mais geral possível, isto é, estimando $\Delta y_t = \mu + \delta t + \alpha y_{t-1} + u_t$ e teste $H_0 : \alpha = 0 \times H_1 : \alpha < 0$. Se rejeitar a hipótese nula, não é necessário avançar.
2. Se não rejeitar a hipótese nula, pode ser em virtude do baixo poder do teste. Este poderá ser melhorado se não tiver tendência temporal e se não houver *drift*, isto é, se $\mu = \delta = 0$. Então, primeiro teste:

$$H_0 : \begin{bmatrix} \alpha \\ \delta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \times H_1 : \begin{bmatrix} \alpha \\ \delta \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

usando a estatística Φ_3 , tabulada por Phillips e Perron (1988). Se se rejeita H_0 , conclua pela existência de raiz unitária, via estatística $z_{t,\tau}$ de Phillips e Perron.

3. Se não se rejeita H_0 , teste:

$$\begin{bmatrix} \alpha \\ \delta \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \times H_1 : \begin{bmatrix} \alpha \\ \delta \\ \mu \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

usando a estatística Φ_2 de Phillips e Perron. Se não se rejeita H_0 , teste para raiz unitária, sem tendência e *drift*, usando a estatística z_t de Phillips e Perron.

4. Se se rejeita H_0 usando Φ_2 , teste sem tendência o modelo

$$\begin{bmatrix} \alpha \\ \mu \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \times H_1 : \begin{bmatrix} \alpha \\ \mu \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

usando a estatística Φ_1 de Phillips e Perron. Se se rejeita H_0 , então teste o modelo só com intercepto, usando a estatística $z_{t,\mu}$.

5. Se não se rejeita H_0 usando Φ_1 , então teste para raiz unitária usando a estatística z_t .

4.5.5 KPSS

Um dos problemas do teste de raiz unitária desenvolvido por Dickey e Fuller (1979, 1981) é seu baixo poder. Isso significa que o teste de raiz unitária não consegue rejeitar a nula para uma infinidade de séries econômicas. Por essa razão, outros testes vêm sendo desenvolvidos. Um deles é o teste KPSS (mnemônico de seus autores Kwiatkowski, Phillips, Schmidt e Shin) idealizado em 1992. A hipótese nula é de estacionaridade da série, isto é, $H_0 : y_t \sim I(0)$ contra $H_1 : y_t \sim I(1)$. A idéia dos autores é usar o teste complementando os testes de raiz unitária. O argumento é que se pode, assim, distinguir a raiz unitária de séries cujos dados não são suficientemente conclusivos.

Para entender o teste, assuma que o processo gerador de dados é:

$$y_t = x_t + u_t,$$

em que $x_t = x_{t-1} + v_t$, $v_t \sim i.i.d. (0, \sigma^2)$ e u_t é um processo estacionário.

A idéia é testar a variância de passeio aleatório x_t . Se essa variância for nula, então o processo é estacionário. Nessa estrutura, a hipótese a ser testada é a seguinte:

$$H_0 : \sigma^2 = 0 \times H_1 : \sigma^2 > 0.$$

Pode-se acrescentar uma tendência determinística ao modelo da seguinte forma:

$$y_t = \mu + \delta t + x_t + u_t.$$

O procedimento usual, nesse caso, é:

1. Estimar a série em análise contra as variáveis determinísticas:

$$y_t = \mu + \delta t + e_t.$$

2. Calcular os resíduos dessa regressão, em todo t :

$$\hat{e}_t = y_t - \hat{\mu} - \hat{\delta}t.$$

3. Definir a soma parcial dos resíduos como

$$S_t = \sum_{j=1}^t \hat{e}_j.$$

4. Usar teste de multiplicador de Lagrange, KPSS, assim estabelecido:

$$KPSS = \sum_{t=1}^T \frac{S_t^2}{T^2 \hat{\sigma}^2},$$

em que $\hat{\sigma}^2$ é a variância de longo prazo, definida como no teste de Phillips e Perron:

$$\hat{\sigma}^2 = \hat{\sigma}^2 + \frac{2}{T} \sum_{j=1}^M \omega \left(\frac{j}{M+1} \right) \sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j};$$

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{e}_t^2}{T}.$$

Se y_t é um processo estacionário, então S_t será $I(1)$ e o numerador do KPSS é um estimador da variância de S_t que, por sua vez, tem um limite assintótico. O termo no denominador assegura que a distribuição é livre de ruídos. Se, por outro lado, y_t é $I(1)$, o numerador vai crescer sem limites, o que faz a estatística se tornar bastante grande.

Para visualizar melhor, observe que⁷:

$$e_t = x_t + u_t.$$

Se a variância de x_t é positiva, então $x_t = x_0 + \sum_{j=1}^t v_j$, isto é:

$$e_t = x_0 + \sum_{j=1}^t v_j + u_t.$$

Substituindo essa equação em S_t , encontra-se:

$$S_t = \sum_{j=1}^t \left(x_0 + \sum_{h=1}^j v_h + u_j \right)$$

$$= tx_0 + \sum_{j=1}^t \sum_{h=1}^j v_h + \sum_{j=1}^t u_j.$$

É possível perceber aqui que S_t é, pelo menos, $I(1)$ em razão do termo $\sum_{j=1}^t u_j$. Tomando $\sum_{t=1}^T S_t^2$, obtém-se uma variável $I(2)$, que, ajustada por T^2 , resulta em uma distribuição limitada, caso a variância de x_t seja nula. Se não for, o termo $\sum_{j=1}^t \sum_{h=1}^j v_h$ fará com que a estatística KPSS exploda, e o teste resultará em um valor elevado, tornando impossível não rejeitar a nula.

⁷ Por amor à simplicidade, ignora-se o circunflexo da variável estimada.

Os testes KPSS são definidos em η , η_μ e η_τ , conforme se especifica o modelo sem constante ou tendência, com constante, ou com constante e tendência, respectivamente.

Exemplo 4.4

Seja o exemplo com a série da taxa de câmbio e de inflação. Os parâmetros são os mesmos, mas mudam a hipótese nula e os valores críticos. Nesse teste, a hipótese nula consiste de a série ser estacionária. Com isso, o EViews gera uma saída em que se podem ler, entre outras coisas:

Câmbio	Valores Críticos Teóricos	Valores Estimados	
	η_μ	$\hat{\eta}_\mu$	0,324
1%	0,739		
5%	0,463	M	7
10%	0,347	Variância Residual: $\hat{\sigma}^2$	0,250905
		HAC: $\hat{\sigma}^2$	1,748361

Aqui a hipótese nula é de estacionaridade da série. Nesse caso, como $\hat{\eta}_\mu < \eta_\mu$, não se rejeita a nula. No caso do KPSS, isso significa que a série de câmbio é estacionária.

IPCA	Valores Críticos Teóricos	Valores Estimados	
	η_μ	$\hat{\eta}_\mu$	0,316
1%	0,739		
5%	0,463	M	8
10%	0,347	Variância Residual: $\hat{\sigma}^2$	0,331847
		HAC: $\hat{\sigma}^2$	1,47527

Aqui, não se rejeita a nula, consistentemente com os testes de Dickey-Fuller e Phillips-Perron.

4.5.6 ERS

A seção anterior mencionou que o teste KPSS foi desenvolvido porque os testes de Dickey e Fuller têm baixo poder. Para visualizar o problema, considere o seguinte modelo:

$$y_t = d_t + u_t; \quad u_t = \alpha u_{t-1} + e_t; \quad e_t = \psi(L) \epsilon_t;$$

$$d_t = \sum_{h=0}^H \delta_h t^h \equiv \delta' x_t.$$

A parcela d_t representa os termos determinísticos de *drift* e tendência, isto é, $\max H = 1$, porém poderia conter mais regressores. Quando $\alpha = 1$, y_t é integrado de ordem 1. Admite-se, ainda, que $E(u_0^2) < \infty$, $\sum_{j=1}^{\infty} j |\psi_j| < \infty$. Note que essa última hipótese é mais forte do que foi vista assumindo absoluta somabilidade $\sum_{j=1}^{\infty} |\psi_j| < \infty$.

O poder baixo do teste emerge quando $\alpha \rightarrow 1$, porém $\alpha < 1$. Essencialmente, baixo poder significa que o teste comete muito frequentemente o erro do tipo II, isto é, não rejeita a nula quando ela é falsa. Por exemplo, Perron e NG (1996, tabela 3) relatam que o teste ADF tem um poder de 25,8% quando $\delta_h = 0$ e $\alpha = 0,95$, para uma amostra de $T = 200$ observações. Isto é, em 74,2% das simulações do modelo, o teste ADF não rejeitou a nula quando ela era falsa. A mesma tabela mostra que o poder aumenta para 92,5% quando $\alpha = 0,85$.

O poder do teste se reduz consideravelmente na presença de termos determinísticos. Elliot, Rothemberg e Stock (1996, tabela II) relatam que o poder do teste ADF para $H = 1$, $\psi(L) = 1/(1 - \phi_1 L)$, $\alpha = 0,95$ e $\phi_1 = 0,5$ é de 10%. Ou seja, o teste só rejeita a nula de raiz unitária em 10% dos casos, embora seja de fato falsa em todos os casos. Por outro lado, o poder aumenta para 67% quando $\alpha = 0,7$ e $\phi_1 = 0,5$.

Elliot, Rothemberg e Stock argumentam que o poder do teste pode ser aumentado se, de alguma forma, os termos determinísticos forem expurgados da regressão do teste. Eles sugerem, então, uma modificação no teste ADF, rebatizando o teste de DF-GLS. O procedimento operacional do teste é relativamente simples e segue os seguintes passos:

- a) Dada a seqüência qualquer observada $\{z_t\}_{t=0}^T$, defina a seqüência

$$\begin{pmatrix} z_0^{\bar{\alpha}} \\ z_t^{\bar{\alpha}} \end{pmatrix} \equiv (z_0, (1 - \bar{\alpha}L)z_t), t = 1, 2, \dots, T,$$

para algum $\bar{\alpha} \equiv 1 + \frac{c}{T}$;

- b) Encontre $\hat{\delta}(\bar{\alpha})$ que minimiza a seguinte função:

$$L(\bar{\alpha}) = \min_{\{\delta_h(\bar{\alpha})\}_{h=0}^H} \left[y_t^{\bar{\alpha}} - \hat{\delta}(\bar{\alpha})' x_t^{\bar{\alpha}} \right]' \left[y_t^{\bar{\alpha}} - \hat{\delta}(\bar{\alpha})' x_t^{\bar{\alpha}} \right];$$

- c) Em seguida, obtenha a série com os termos determinísticos expurgados, em que o sobrescrito d representa *detrended*:

$$y_t^d \equiv y_t - \hat{\delta}(\bar{\alpha})' x_t;$$

d) Proceda ao teste de Dickey-Fuller usando a nova sequência:

$$\Delta y_t^d = \alpha y_{t-1}^d + \sum_{i=1}^p \lambda_i \Delta y_{t-i}^d + \epsilon_t. \quad (8)$$

Como y_t^d é livre de termos determinísticos, é desnecessário incluir constante ou tendência. Entretanto, é preciso saber se a série original possui tendência determinística ou constante para definir o valor de c :

$$c = \begin{cases} -7, & \text{se } H = 0; \\ -13,5, & \text{se } H = 1. \end{cases}$$

O valor de c decorre de experimentos de Monte Carlo, de forma a maximizar o poder do teste $\alpha = 1$ contra $\bar{\alpha} = \bar{\alpha}$, quando se fixa o poder em 50%. A intuição do teste é que o poder vai aumentar conforme $\bar{\alpha}$ se distancie de α .

Do ponto de vista prático, na verdade, o valor de c fixado para um poder de 50% funciona bem para faixas de poder que variam de 25% a 95%. O resultado fundamental de Elliot, Rothemberg e Stock é que o poder do teste ADF aumenta consideravelmente, passando de 10% para 26%, quando $\alpha = 0,95$ e $\phi_1 = 0,5$, e para 95%, quando α passa a 0,70.

Elliot, Rothemberg e Stock propõem ainda outro teste, identificado como ERS – Point Optimal, baseado na estatística LR – taxa de verossimilhança. Para executá-lo, devem-se seguir estes passos:

1. Obtenha os resíduos, indexados pelo número de defasagens, da regressão:

$$\Delta y_t = d_t + \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i} + \epsilon_{t,p};$$

2. Calcule a variância amostral desses resíduos:

$$\hat{\sigma}_p^2 = \sum_{t=p+1}^T \frac{\hat{\epsilon}_{t,p}^2}{T-p};$$

3. Calcule a variância de longo prazo (ou densidade auto-regressiva espectral na frequência zero):

$$v_{AR}^2 = \frac{\hat{\sigma}_p^2}{[1 - \hat{\lambda}(1)]^2},$$

$$\text{em que } \hat{\lambda}(1) = \sum_{i=1}^p \hat{\lambda}_i;$$

4. Finalmente, calcule a estatística P_T , ajustada pela possível correlação serial dos resíduos:

$$P_T = \frac{|L(\bar{\alpha}) - \bar{\alpha}L(1)|}{\sigma_{AR}^2}.$$

A intuição do teste é a seguinte. Se a série for integrada, a diferenciação gerará uma série de variância pequena se $\bar{\alpha} = 1$, porém o valor de $L(\bar{\alpha} \neq 1)$ será grande. Logo, P_T será grande e não se rejeita a hipótese nula. Se a série for estacionária, a diferenciação da série em $L(\bar{\alpha} = 1)$ será estacionária, e o mesmo acontecerá com $L(\bar{\alpha} \neq 1)$. Os valores serão baixos e, conseqüentemente, P_T terá um valor baixo. Disso, conclui-se que a regra de rejeição da nula é a seguinte: se $P_T^{\text{calculado}} < P_T^{\text{crítico}}$, rejeita-se a nula de raiz unitária.

Uma variante do teste é usar como variância de longo prazo o estimador:

$$\hat{\sigma}^2 = \hat{\sigma}^2 + \frac{2}{T} \sum_{j=1}^M \omega\left(\frac{j}{M+1}\right) \sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j};$$

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^T \hat{e}_t^2}{T},$$

em que \hat{e}_t é calculado como resíduo da regressão de y_t contra d_t :

$$\hat{e}_t = y_t - \hat{\mu} - \hat{\delta}t.$$

Simulações desses autores apontam para uma certa superioridade do teste DF-GLS.

Exemplo 4.5

Seja o exemplo com a série da taxa de câmbio. Primeiro, considere o teste DF-GLS, em que o *Eviews* fixa o máximo $p_{\max} = 11$ e seleciona o p ideal pelo critério de Schwarz, admitindo-se constante apenas.

Câmbio	Valores Críticos Teóricos	Valores Estimados
	τ_{μ}^{GLS}	$\hat{\tau}_{\mu}^{\text{GLS}} \quad -1,057$
1%	-2,590	$p \quad 1$
5%	-1,944	
10%	-1,614	

Pelo critério de ERS-Point Optimal, com janela AR Spectral-mínimos quadrados ordinários e M escolhido pelo critério de Schwarz, obtêm-se os seguintes resultados:

Câmbio	Valores Críticos Teóricos	Valores Estimados	
	P_T	\hat{P}_T	12,655
1%	1,944	M	1
5%	3,099	HAC: $\hat{\sigma}^2$	0,021
10%	4,149		

Em ambos os casos, não é possível rejeitar a hipótese de raiz unitária, e conclui-se que a série a possui.

Exemplo 4.6

Considere agora o exemplo da série de inflação. Primeiro, considere o teste DF-GLS, em que o Evietos fixa o máximo $p_{\max} = 12$ e seleciona o p ideal pelo critério de Akaike, admitindo-se constante apenas.

Câmbio	Valores Críticos Teóricos	Valores Estimados	
	τ_{μ}^{GLS}	$\hat{\tau}_{\mu}^{\text{GLS}}$	2,653
1%	-2,581	p	13
5%	-1,942		
10%	-1,615		

Nesse caso, rejeita-se a nula e conclui-se que a inflação no Brasil é estacionária, resultado que corrobora os testes anteriores.

Usando o teste ERS-Point Optimal, serão encontrados:

IPCA	Valores Críticos Teóricos	Valores Estimados	
	P_T	\hat{P}_T	2,889
1%	1,930	M	13
5%	3,139	HAC: $\hat{\sigma}^2$	0,151
10%	4,248		

Nesse caso, rejeita-se a nula a 5%, porém não se rejeita a 1%. Comparado com os testes anteriores, o resultado é relativamente inconclusivo. A análise não se altera se se utilizam outras formas de cálculo da variância de longo prazo, como pela densidade espectral na frequência zero usando o modelo auto-regressivo.

4.5.7 NG e Perron

Além da questão de poder, o teste de raiz unitária sofre do problema de tamanho do teste, quando a raiz do processo de médias móveis é muito alta, isto é, quando $\theta \rightarrow 1$. A razão desse problema é fácil de visualizar. Considere o seguinte modelo $ARMA(1,1)$:

$$y_t = \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}.$$

Fatorando, pode-se obter:

$$(1 - \phi L) y_t = (1 + \theta L) \epsilon_t.$$

Ora, se existe raiz unitária, $\phi = 1$. Mas, se $\theta \rightarrow -1$, então os lados direito e esquerdo da equação cancelam-se e y_t , no limite, vai se comportar como um ruído branco. Do ponto de vista prático, se θ estiver próximo de $-0,9$, a rejeição da hipótese nula é muito mais freqüente do que se desejaria, em razão das distorções de tamanho. Para se ter uma idéia, NG e Perron (2001) relatam que o tamanho do teste DF-GLS quando $\theta = -0,8$, $T = 100$ é de 62,4%, enquanto o ideal seria de 5% ou 10%. A Figura 4.6 mostra duas séries simuladas com os mesmos erros. Porém, a série que flutua ao redor de zero foi calculada com $\theta = -0,8$. Embora as duas séries sejam integradas, é difícil reconhecer isso visualmente na série com médias móveis.

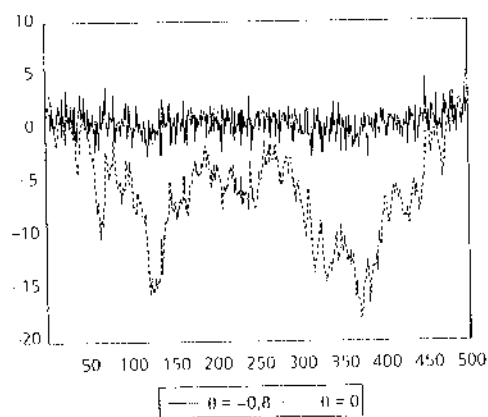


Figura 4.6 Passeios aleatórios com diferentes médias móveis.

O teste de raiz unitária sobre a série em que $\theta = -0,8$ rejeita fortemente a hipótese, não obstante tenha simulado um passeio aleatório.

Cientes desse problema de tamanho, Perron e NG (1996) propõem modificações aos testes de Phillips e Perron (1988) que simulações mostraram melhorar

sensivelmente o tamanho do teste. São os M testes, para lembrar “modificados”. São três testes modificados, admitindo-se que já tenha sido expurgada alguma eventual tendência:

$$Mz_{\alpha} = \frac{\left(\frac{y_T^2}{T} - v_{AR}^2\right)}{2 \frac{\sum_{t=1}^T y_{t-1}^2}{T^2}} = z_{\alpha} + \frac{T}{2} (\hat{\alpha} - 1)^2;$$

$$Mz_t = Mz_{\alpha} \times MSB = z_t + \frac{1}{2} \sqrt{\frac{\sum_{t=1}^T y_{t-1}^2}{v_{AR}^2}} (\hat{\alpha} - 1)^2;$$

$$MSB = \sqrt{\frac{\sum_{t=1}^T y_{t-1}^2}{T^2 v_{AR}^2}}.$$

Simulações de Monte Carlo mostram consideráveis ganhos de tamanho a partir desse teste, particularmente quando se emprega o estimador da variância de longo prazo v_{AR}^2 no lugar de v^2 . Por exemplo, em uma amostra em que $T = 200$ e $\theta = -0,8$ o tamanho do teste Mz_t diminui de 96,1% com v^2 para 6,3% com v_{AR}^2 (veja tabelas 1 e 2 de Perron e NG, 1996).

Entretanto, os autores ainda não usavam o procedimento de expurgo de tendência de Elliot, Rothemberg e Stock (1996). Tal expurgo, em tese, poderia melhorar ainda mais o tamanho do teste, particularmente porque o teste é muito sensível às variáveis determinísticas. É nesse contexto que, de fato, NG e Perron (2001) mostram que um considerável ganho de tamanho poderia ser obtido usando os testes Mzs juntamente com o procedimento de expurgo proposto, denominando esses testes de Mz^{GLS} . Além disso, NG e Perron propõem ainda o teste modificado de “Point Optimal”:

$$MP_T^{GLS} = \begin{cases} \frac{\frac{c^2}{T^2} \sum_{t=1}^T y_{t-1}^d - \frac{c}{T} (y_T^d)^2}{v_{AR}^2}, & \text{quando } H = 0; \\ \frac{\frac{c^2}{T^2} \sum_{t=1}^T y_{t-1}^d - \frac{1-c}{T} (y_T^d)^2}{v_{AR}^2}, & \text{quando } H = 1. \end{cases}$$

Assim como as demais regras de decisão para raiz unitária, se o valor calculado dessa estatística for menor que o valor crítico, rejeita-se a hipótese de raiz unitária.

Permanecia outro problema, contudo. Os testes são sensíveis ao tamanho da defasagem auto-regressiva p . Por exemplo, NG e Perron mostram por simulações de Monte Carlo que o tamanho do teste DF-GLS com uma amostra de

250 observações, $H = 0$ e $\theta = -0,8$, reduz-se de 98,5%, quando $p = 0$, para 9,9%, quando $p = 10$. Os testes Mz^{GLS} resultam em tamanhos menores que os testes DF-GLS, porém sofrem do mesmo problema. Por exemplo, o teste Mz_a^{GLS} , $T = 250$ e $\theta = -0,8$, reduz de um tamanho de 98,7%, quando $p = 0$, para um tamanho de 3,5%, quando $p = 10$.

Nessas circunstâncias, é preciso desenvolver uma técnica dependente da amostra para selecionar a defasagem ótima. É o que NG e Perron (2001) fazem nesse artigo, propondo modificações nos tradicionais critérios de informação de Akaike (AIC), Schwarz (BIC) e Hannan-Quinn (HQ).

4.5.8 Critério de Informação e Janela Ótima

Para o teste de NG e Perron (2001), eles sugerem modificações aos critérios de informação convencionais. Essas modificações decorrem de simulações de Monte Carlo, que mostraram gerar melhores resultados em termos de poder e tamanho do teste. A tabela a seguir mostra a natureza dessas modificações, essencialmente obtidas pela inclusão do coeficiente τ aos critérios originais:

Critério de Informação	Definição
Modified AIC – MAIC	$\ln \hat{\sigma}^2 + (n + \tau) \frac{2}{T}$
Modified BIC – MBIC	$\ln \hat{\sigma}^2 + (n + \tau) \frac{\ln T}{T}$
Modified HQ – MHQ	$\ln \hat{\sigma}^2 + (n + \tau) \frac{2}{T} \ln \ln T$

em que

n é o número de parâmetros estimados na regressão 8;

$$\tau = \hat{\alpha}^2 \sum_{t=p_{\max}+1}^T \frac{(y_{t-1}^d)^2}{\hat{\sigma}_p^2}, \quad \hat{\sigma}_p^2 = \sum_{t=p_{\max}+1}^T \frac{\hat{\epsilon}_{t,p}^2}{T - p_{\max}}.$$

Observe aqui que $\hat{\epsilon}_{t,p}^2$ é calculado a partir da regressão

$$\Delta y_t = d_t + \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i} + \epsilon_{t,p},$$

em que p é fixado otimamente. Entretanto, alguns resíduos estimados, aqueles entre p e p_{\max} , não devem ser contabilizados. Isso é feito para que se assegure que o número de observações utilizadas para calcular a estatística seja idêntico em todos os modelos com diferentes defasagens. Se o número de observações utilizadas fosse diferentes, os testes seriam incomparáveis.

Em termos práticos, NG e Perron recomendam que se use o método MAIC.

No caso de usar o método não paramétrico de Bartlett, Parzen ou janela quadrática, para definir a função $\lambda(\cdot)$, é preciso saber o tamanho da banda, isto é, o valor M . A definição da janela é importante para a melhor estimativa. Há duas formas de definir o valor M . Uma delas é utilizar o procedimento descrito em Andrews (1991); a outra é usar o procedimento de Newey e West (1994).

Exemplo 4.7

Seja o exemplo com a série da taxa de câmbio, usando como opção de janela espectral o GLS-detrended AR com constante, com $M = 7$, definido pelo critério AIC modificado. Então, tem-se:

Câmbio		$Mz_{\alpha,\mu}^{GLS}$	$Mz_{t,\mu}^{GLS}$	MSB^{GLS}	MP_T^{GLS}
Valor Calculado		-1,570	-0,883	0,562	15,540
Valores Críticos	1%	-13,800	-2,580	0,174	1,780
	5%	-8,100	-1,980	0,233	3,170
	10%	-5,700	-1,620	0,275	4,450

Os testes demonstram a não-rejeição da hipótese nula, de forma que o câmbio parece ter uma raiz unitária. O resultado do teste é invariante a outras especificações de janela, ou cálculo paramétrico da variância de longo prazo.

Exemplo 4.8

Considere agora o exemplo da série de inflação, usando a janela espectral GLS-detrended AR com constante, com $M = 12$, definido pelo critério AIC modificado. Então, tem-se:

IPCA		$Mz_{\alpha,\mu}^{GLS}$	$Mz_{t,\mu}^{GLS}$	MSB^{GLS}	MP_T^{GLS}
Valor Calculado		-1,13786	-0,59311	0,52125	15,9673
Valores Críticos	1%	-13,800	-2,580	0,174	1,780
	5%	-8,100	-1,980	0,233	3,170
	10%	-5,700	-1,620	0,275	4,450

Nesse caso, rejeita-se a nula a 5%, mas não a 1%. O resultado do teste é invariante a outras especificações de janela, ou cálculo paramétrico da variância de longo prazo.

4.5.9 Raízes Unitárias Sazonais

A maneira mais direta de tratar raízes sazonais é usando variáveis *dummies* para captá-las. Imaginando dados trimestrais, pode-se estimar a seguinte regressão:

$$\Delta y_t = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \alpha y_{t-1} + \sum_{i=1}^p \lambda_i \Delta y_{t-i+1} + \epsilon_t.$$

Experimentos de Monte Carlo demonstram que a distribuição do teste sobre α não se altera na presença de sazonalidade determinística, mesmo na presença de tendência temporal, t .

Se não é possível usar *dummies* e há raiz unitária sazonal, o teste muda consideravelmente. Seguindo a exposição de Enders (2004), suponha dados trimestrais, de modo que:

$$(1 - \phi_1 L)(1 + \phi_2 L)(1 - i\phi_3 L)(1 + i\phi_4 L)y_t = \epsilon_t.$$

Se houver raiz unitária sazonal, então $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 1$, gerando $(1 - L^4)$. Mas há outras possibilidades, listadas a seguir:

- a) Se $\phi_1 = 1$, y_t é o típico caso de um passeio aleatório, testado como já sabido;
- b) Se $\phi_2 = 1$, a sequência tende a se replicar a cada seis meses, portanto há uma raiz unitária semi-anual, já que a solução homogênea é: $y_t + y_{t-1} = 0$. Por exemplo, se $y_t = 1, y_{t+1} = -1, y_{t+2} = 1, \dots$
- c) Se $\phi_3 = 1$ ou $\phi_4 = 1$, a sequência tem uma raiz unitária de ciclo anual. Para ver isso, suponha $y_t = 1$, então $y_{t+1} = i, y_{t+2} = i^2 = -1, y_{t+3} = -i, y_{t+4} = 1$.

Para entender o teste, expanda $(1 - \phi_1 L)(1 + \phi_2 L)(1 - i\phi_3 L)(1 + i\phi_4 L)$ por Taylor em torno de $\phi_1 = \phi_2 = \phi_3 = \phi_4 = 1$.

$$\begin{aligned} & (1 - \phi_1 L)(1 + \phi_2 L)(1 - i\phi_3 L)(1 + i\phi_4 L) \\ &= (1 - L^4) - L(1 + L + L^2 + L^3)(\phi_1 - 1) \\ &+ L(1 - L + L^2 - L^3)(\phi_2 - 1) - iL(1 - L^2)(1 + iL)(\phi_3 - 1) \\ &+ iL(1 - L^2)(1 - iL)(\phi_4 - 1). \end{aligned}$$

Definindo $\alpha_j \equiv \phi_j - 1$, para todo $j = 1, 2, 3, 4$, e notando que $i(1 + iL) = i - L$ e $i(1 - iL) = i + L$, pode-se escrever:

$$\begin{aligned} (1 - L^4) y_t = & \alpha_1 (1 + L + L^2 + L^3) y_{t-1} - \alpha_2 (1 - L + L^2 - L^3) y_{t-1} \\ & + (1 - L^2) [\alpha_3 (i - L) - \alpha_4 (i + L)] y_{t-1} + \epsilon_t. \end{aligned}$$

Ou seja:

$$\begin{aligned} (1 - L^4) y_t = & \alpha_1 (1 + L + L^2 + L^3) y_{t-1} - \alpha_2 (1 - L + L^2 - L^3) y_{t-1} \\ & + (1 - L^2) [i(\alpha_3 - \alpha_4) - (\alpha_3 + \alpha_4)L] y_{t-1} + \epsilon_t. \end{aligned}$$

Definindo $2\alpha_3 = \alpha_6 - i\alpha_5$ e $2\alpha_4 = \alpha_6 + i\alpha_5$, tem-se que $(\alpha_3 - \alpha_4)i = \alpha_5$ e $(\alpha_3 + \alpha_4) = \alpha_6$. Disso resulta que:

$$\begin{aligned} (1 - L^4) y_t = & \alpha_1 (1 + L + L^2 + L^3) y_{t-1} - \alpha_2 (1 - L + L^2 - L^3) y_{t-1} \\ & + (1 - L^2) (\alpha_5 - \alpha_6 L) y_{t-1} + \epsilon_t. \end{aligned}$$

Podem-se adicionar à regressão, se conveniente, *dummies* sazonais, tendência temporal e intercepto. Todos os valores já foram tabelados, bastando consultar as tabelas apropriadas.

1. Com base nessas derivações, monte séries auxiliares:

$$x_{t-1} = y_{t-1} + y_{t-2} + y_{t-3} + y_{t-4};$$

$$z_{t-1} = y_{t-1} - y_{t-2} + y_{t-3} - y_{t-4};$$

$$m_{t-1} = y_{t-1} - y_{t-3}.$$

2. Estime a regressão aumentada de Dickey e Fuller:

$$\begin{aligned} (1 - L^4) y_t = & \alpha_1 x_{t-1} - \alpha_2 z_{t-1} + \alpha_5 m_{t-1} - \alpha_6 m_{t-2} \\ & + \sum_{i=1}^p \lambda_i (1 - L^4) \Delta y_{t-i} + \epsilon_t. \end{aligned}$$

3. Se não se rejeita $\alpha_1 = 0$ pelo teste τ , então existe raiz unitária não sazonal. Se não se rejeita que $\alpha_2 = 0$, existe uma raiz unitária semestral. Se não se rejeita o teste de F conjunto que $\alpha_5 = \alpha_6 = 0$, de modo que o valor calculado seja menor do que o valor crítico, então há sazonalidade anual.⁸

⁸ As hipóteses anteriores não são conjuntamente excludentes. Isso significa que a série pode ter raiz unitária semestral e anual simultaneamente, por exemplo.

4.6

Decomposição de Beveridge-Nelson

Agora que já se sabe estimar uma série temporal e o significado de séries estacionárias, pode-se apresentar um importante instrumento para obter o componente permanente e transitório de uma série. Trata-se da decomposição de Beveridge e Nelson (1981). Esse resultado é importante para entender o efeito de curto prazo dos ciclos econômicos e inferir a tendência de longo prazo da série analisada. Por exemplo, suponha que haja uma queda do PIB, então, seria importante saber se essa queda é primordialmente devida a fatores cíclicos ou trata-se de uma tendência de longo prazo.

O segredo matemático da decomposição é perceber a seguinte identidade:

$$\psi(L) = \psi(1) + (1 - L)\psi^*(L),$$

em que

$$\psi(1) = \sum_{j=0}^{\infty} \psi_j;$$

$$\psi^*(L) = \sum_{k=0}^{\infty} \psi_k^* L^k \text{ e}$$

$$\psi_k^* = \sum_{j=k+1}^{\infty} \psi_j, \text{ sendo } \psi_0 = 1.$$

Para ver que vale essa identidade, considere:

$$\begin{aligned} \psi^*(L) &= \sum_{k=0}^{\infty} \psi_k^* L^k \\ &= - \sum_{k=0}^{\infty} L^k \sum_{j=k+1}^{\infty} \psi_j \\ &= - \sum_{j=1}^{\infty} \psi_j - L \sum_{j=2}^{\infty} \psi_j - L^2 \sum_{j=3}^{\infty} \psi_j - \dots \end{aligned}$$

Portanto:

$$\begin{aligned} \psi(1) + (1 - L)\psi^*(L) &= \psi(1) - \sum_{j=1}^{\infty} \psi_j - L \sum_{j=2}^{\infty} \psi_j - L^2 \sum_{j=3}^{\infty} \psi_j - \dots \\ &\quad + L \sum_{j=1}^{\infty} \psi_j + L^2 \sum_{j=2}^{\infty} \psi_j + L^3 \sum_{j=3}^{\infty} \psi_j + \dots \\ &= 1 + L\psi_1 + L^2\psi_2 + L^3\psi_3 + \dots \\ &= \sum_{j=0}^{\infty} L^j \psi_j = \psi(L). \end{aligned}$$

Estabelecida a identidade, considere um modelo *ARIMA* $(p, 1, q)$ com *drift*:

$$y_t = y_{t-1} + \delta + \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = y_{t-1} + \delta + \psi(L) \epsilon_t.$$

Esse modelo pode ser reescrito da seguinte forma, por substituição recursiva:

$$y_t = y_0 + \delta t + \psi(L) \sum_{j=1}^t \epsilon_j. \quad (9)$$

Aplicando a identidade à equação (9), obtém-se:

$$\begin{aligned} y_t &= y_0 + \delta t + [\psi(1) + (1-L)\psi^*(L)] \sum_{j=1}^t \epsilon_j \\ &= y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j + (1-L)\psi^*(L) \sum_{j=1}^t \epsilon_j \\ &= y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j + \psi^*(L) (\epsilon_t - \epsilon_0). \end{aligned} \quad (10)$$

Beveridge e Nelson decompõem a última expressão em duas parcelas, admitindo-se por simplicidade que $\epsilon_0 = 0$:

$$\begin{aligned} y_t &= p_t + c_t; \\ p_t &= y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j; \\ c_t &= \psi^*(L) \epsilon_t. \end{aligned}$$

Da segunda equação, pode-se subtrair p_{t-1} para obter:

$$p_t = p_{t-1} + \delta + \psi(1) \epsilon_t, \quad p_0 = y_0.$$

A variável p_t representa o componente permanente do modelo e se trata de passeio aleatório com *drift* puro, cuja inovação tem variância $\psi(1)^2 \sigma^2$. O fator $\psi(1)$ pode ser maior ou menor que 1, de modo que a variância do componente permanente pode ser maior ou menor que a variância de y_t . Se for maior, o componente cíclico tende a suavizar os choques de longo prazo na série. Se for menor, o componente cíclico tende a acentuar os choques havidos sobre p_t .

O termo c_t representa o componente cíclico ou transitório do modelo, que será estacionário se for verdade a seguinte hipótese técnica que se assume valer:

$\sum_{j=0}^{\infty} |\psi_j| j < \infty$. Essa hipótese é um pouco mais forte que $\sum_{j=0}^{\infty} |\psi_j| < \infty$, porém é necessária para demonstrar que $\sum_{k=0}^{\infty} |\psi_k^*| < \infty$.

A seguir, procura-se introduzir mais intuição econômica ao modelo, buscando-se explicar a razão econômica dessa decomposição. Para ver isso, tome y_{t+h} :

$$y_{t+h} = y_t + \delta h + \psi(L) \sum_{s=1}^h \epsilon_{t+s} = y_t + \delta h + \sum_{s=1}^h \sum_{k=0}^{\infty} \psi_k \epsilon_{t+s-k}.$$

Como $E_t(\epsilon_{t+s-k}) = 0$, para $k < s$, segue-se da equação anterior que a previsão de y_{t+h} , dada a informação em t , é:

$$E_t(y_{t+h}) = y_t + h\delta + \sum_{s=1}^h \sum_{k=s}^{\infty} \psi_k \epsilon_{t+s-k}.$$

Definindo o termo $s - k = -j$ e substituindo na equação anterior, tem-se que:

$$\begin{aligned} E_t(y_{t+h}) &= y_t + h\delta + \sum_{k-j=1}^h \sum_{k=k-j}^{\infty} \psi_k \epsilon_{t-j} \\ &= y_t + h\delta + \sum_{k=j+1}^h \sum_{j=0}^{\infty} \psi_k \epsilon_{t-j}. \end{aligned}$$

Invertendo os somatórios, obtém-se que:

$$E_t(y_{t+h}) = y_t + h\delta + \sum_{j=0}^{\infty} \left(\sum_{k=j+1}^h \psi_k \right) \epsilon_{t-j}.$$

Como o termo determinístico é muito simples de prever, o problema fundamental recai sobre aquilo que resta além dele, sendo o interesse maior o de encontrar a previsão de longo prazo. Assim, a tendência estocástica é obtida quando se toma $h \rightarrow \infty$, do termo $E_t(y_{t+h}) - h\delta$:

$$\begin{aligned} \lim_{h \rightarrow \infty} [E_t(y_{t+h}) - h\delta] &= y_t + \sum_{j=0}^{\infty} \left(\sum_{k=j+1}^{\infty} \psi_k \right) \epsilon_{t-j} \\ &= y_t - \sum_{j=0}^{\infty} \psi_j^* L^j \epsilon_t = y_t - \psi^*(L) \epsilon_t. \end{aligned}$$

Ora, como $y_t = y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j + \psi^*(L) \epsilon_t$, é forçoso concluir que:

$$\lim_{h \rightarrow \infty} [E_t(y_{t+h}) - h\delta] = y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j = p_t.$$

Conseqüentemente, o componente permanente é um passeio aleatório, essencialmente um resultado da previsão de longo prazo da série ajustada em relação à tendência determinística.

Nota 4.5 Há, ainda, outra forma de derivar o resultado anterior. Tome a equação (10), porém h períodos à frente:

$$y_{t+h} = y_0 + \delta(t+h) + \psi(1) \sum_{j=1}^{t+h} \epsilon_j + c_{t+h}.$$

A previsão de y_{t+h} , dada a informação em t , é:

$$E_t(y_{t+h}) = y_0 + \delta(t+h) + \psi(1) \sum_{j=1}^t \epsilon_j + E_t(c_{t+h}) \Rightarrow$$

$$E_t(y_{t+h}) - \delta h = y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j + E_t(c_{t+h}).$$

Agora, pode-se encontrar a previsão de longo prazo:

$$\lim_{h \rightarrow \infty} [E_t(y_{t+h}) - \delta h] = y_0 + \delta t + \psi(1) \sum_{j=1}^t \epsilon_j = p_t,$$

pois $\lim_{h \rightarrow \infty} E_t(c_{t+h}) = 0$, haja vista que c_{t+h} é um processo estacionário com média nula.

Conclui-se, assim, que o componente permanente é um passeio aleatório, essencialmente resultado da previsão de longo prazo da série ajustada em relação à tendência determinística.

Exemplo 4.9

Considere um modelo $ARIMA(0, 1, 2)$:

$$y_t = y_{t-1} + \delta + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}.$$

É claro que $\psi_1 = \theta_1$, $\psi_2 = \theta_2$ e $\psi_j = 0$, $j > 2$. Então:

$$\psi(1) = 1 + \theta_1 + \theta_2;$$

$$\psi_0^* = - \sum_{j=1}^{\infty} \psi_j = -(\theta_1 + \theta_2);$$

$$\psi_1^* = - \sum_{j=2}^{\infty} \psi_j = -\theta_2;$$

$$\psi_k^* = - \sum_{j=1}^{\infty} \psi_j = 0, \quad k > 1.$$

Logo, tem-se que:

$$p_t = p_{t-1} + \delta + (1 + \theta_1 + \theta_2) \epsilon_t;$$

$$c_t = -(\theta_1 + \theta_2) \epsilon_t - \theta_2 \epsilon_{t-1}.$$

Para operacionalizar o modelo da forma mais simples⁹, seguem-se estes passos:

1. Estime o modelo *ARIMA* $(p, 1, q)$ e obtenha $\hat{\psi}(L) \equiv \hat{\theta}(L) / \hat{\phi}(L)$, $\hat{\epsilon}_t$ e $\hat{\delta}$;
2. Obtenha $\hat{\psi}(1)$;
3. Calcule $\hat{p}_t = \hat{\delta} + \hat{p}_{t-1} + \hat{\psi}(1) \hat{\epsilon}_t$, dado $p_0 = y_0$;
4. Calcule $\hat{c}_t = y_t - \hat{p}_t$.

Exemplo 4.10

Neste exemplo, demonstra-se o procedimento de Beveridge-Nelson com uma série simulada com 199 observações, cujos valores iniciais de y_t são dados iguais a zero, de acordo com o seguinte *ARIMA* $(3, 1, 0)$:

$$y_t = y_{t-1} + 0,2 + 0,3\Delta y_{t-1} - 0,8\Delta y_{t-2} + \epsilon_t.$$

Essa série foi estimada e resultou em (desvio-padrão entre parênteses):

$$y_t = y_{t-1} + \underset{(0,051)}{0,193} + \underset{(0,047)}{0,341\Delta y_{t-1}} - \underset{(0,047)}{0,771\Delta y_{t-2}} + \hat{\epsilon}_t.$$

Os componentes p_t e c_t foram obtidos, respectivamente, calculando-se:

$$\hat{p}_t = \hat{p}_{t-1} + 0,193 + \frac{1}{1 - 0,342 + 0,771} \hat{\epsilon}_t;$$

$$\hat{c}_t = y_t - \hat{p}_t.$$

A Figura 4.7 mostra a série decomposta.

⁹ Rigorosamente, deve-se estimar o modelo e proceder às previsões, conforme explicado a seguir.

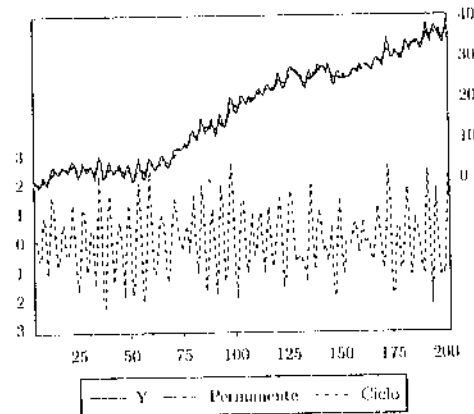


Figura 4.7 Decomposição de Beveridge-Nelson – Simulação de um ARIMA (3, 1, 0).

Exemplo 4.11

Há outro exemplo interessante a seguir. Do site do Banco Central do Brasil, pode-se baixar a série de reservas internacionais em dólar, no conceito liquidez internacional, de dezembro/1970 a abril/2004. Extraído o log da série e constatada, pelo teste de NG-Perron, a existência de raiz unitária, analisaram sua FAC e FACP. Tal análise sugeriu a existência de sazonalidade trimestral e anual, de modo que se estimou o seguinte modelo:

$$(1 - L)(1 - \phi_1 L)(1 - \phi_4 L^4)(1 - \phi_{12} L^{12})y_t = c + \epsilon_t.$$

Como resultado, o modelo estimado com 423 observações foi (desvio-padrão entre parênteses, corrigido por Newey-West):

$$(1 - L) \begin{pmatrix} 1 - 0,211L \\ (0,069) \end{pmatrix} \begin{pmatrix} 1 - 0,168L^4 \\ (0,066) \end{pmatrix} \begin{pmatrix} 1 - 0,136L^{12} \\ (0,073) \end{pmatrix} y_t = 0,010 + \hat{\epsilon}_t, \quad (0,007)$$

Os componentes p_t e c_t foram obtidos, respectivamente, calculando-se:

$$\hat{p}_t = \hat{p}_{t-1} + 0,010 + \frac{1}{(1 - 0,211)(1 - 0,168)(1 - 0,136)} \hat{\epsilon}_t;$$

$$\hat{c}_t = y_t - \hat{p}_t.$$

A Figura 4.8 mostra as séries decompostas.

Observa-se da figura que, desde o início de 1992, as reservas internacionais vêm aumentando permanentemente. Dividindo-se o período de 1992 em diante em três subperíodos, observa-se, entre maio/1998 e abril/2000, uma reversão dessa tendência. Entretanto, desde maio/2000 as reservas não param de subir, podendo-se observar que mais recentemente essa tendência se acentuou.

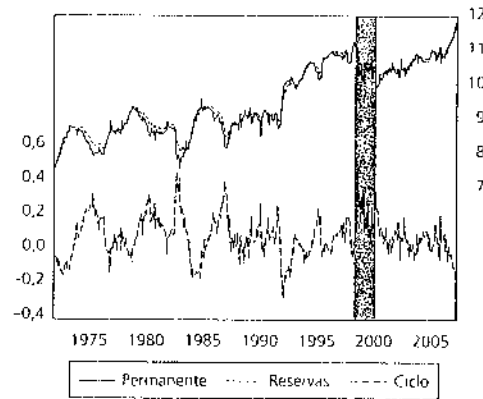


Figura 4.8 Decomposição de Beveridge-Nelson: reservas internacionais do Brasil (log).

Beveridge e Nelson (1981) sugerem outro procedimento para decompor a série. Inicialmente, observe que:

$$y_{t+h} = \Lambda y_{t+h} + \Delta y_{t+h-1} + \cdots + \Delta y_{t+1} + y_t.$$

Isso significa que a tendência sempre pode ser prevista a partir do valor corrente de y_t mais a soma das previsões das variações de y_t . Por sua vez, as variações de y_t são estacionárias e podem ser modeladas. Assim, depois de subtrair δh de ambos os lados da equação anterior para que a previsão não seja explosiva, a previsão do termo estocástico é:

$$\lim_{h \rightarrow \infty} [E_t(y_{t+h}) - \delta h] = \lim_{h \rightarrow \infty} [E_t(\Delta y_{t+h-1} + \Delta y_{t+h-2} + \cdots + \Delta y_{t+1}) - \delta h] + y_t.$$

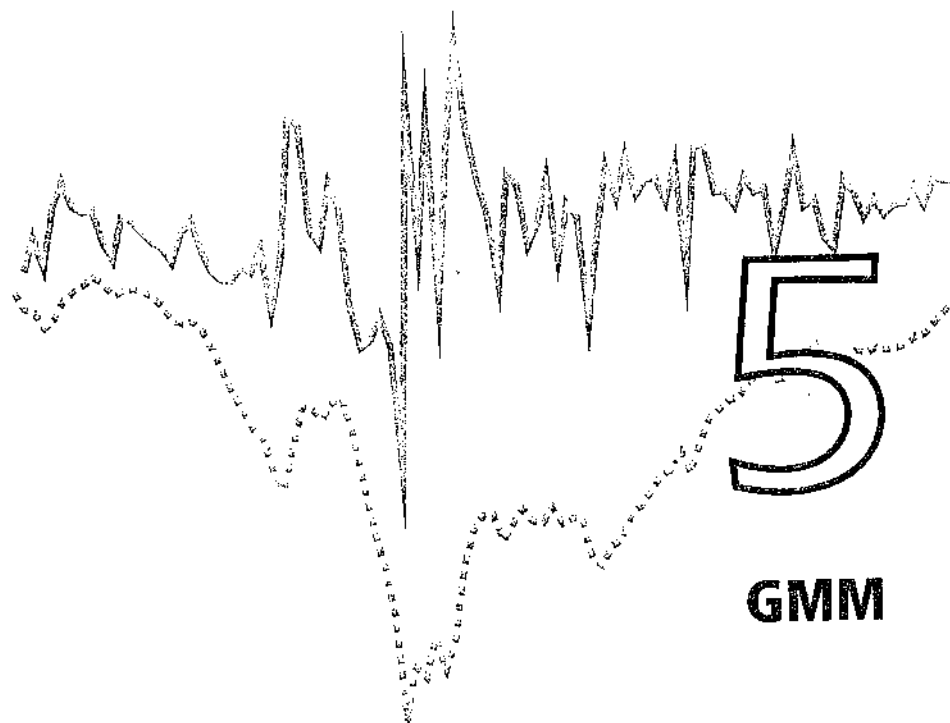
Assim, o problema é encontrar a previsão de variação de cada y_t para cada t e todo h . Uma vez feito isso, o componente irregular será simplesmente a subtração de y_t de sua tendência estocástica e determinística:

$$c_t = y_t - \lim_{h \rightarrow \infty} E_t(y_{t+h} - h\delta).$$

A forma de operacionalizar esse procedimento é a seguinte:

1. Estime o modelo na primeira diferença usando o procedimento de Box, Jenkins e Reinsel (1994);

2. Para cada $t = 1, 2, \dots, T$ use os parâmetros do modelo estimado para prever $E_t(\Delta y_{t+h})$ para $h = 1, 2, \dots, H$. Beveridge e Nelson (1981) usaram até $H = 100$;
3. Construa $p_t = \sum_{h=1}^H E_t(\Delta y_{t+h}) + y_t$;
4. Para encontrar o ciclo, calcule $c_t = y_t - p_t$.



5.1

Introdução

O método generalizado dos momentos é uma maneira de estimação bastante geral, ressuscitada por Hansen (1982). O autor demonstrou as propriedades assintóticas desse estimador no caso não-linear. Sua contribuição é importante, porque o método permite a estimação de modelos de expectativas racionais de modo muito direto.

Mais do que uma forma geral de estimação, o método dos momentos é mais direto de entender, pois trata-se apenas de igualar um momento a um determinado valor¹. O autor contribuiu no sentido de estender o significado de momento, embora em sua origem o método se associasse diretamente aos momentos de uma distribuição. Não se trata mais de encontrar os parâmetros relativos ao primeiro ou ao terceiro momento de uma distribuição, porém os parâmetros de momentos decorrentes de condições econômicas. Assim, igualar um momento a um determinado valor não é mais do que satisfazer uma média amostral.

¹ A grande dificuldade em entender GMM vem do costume de usar outros métodos.

Para entender primitivamente o método dos momentos, recorre-se a um exemplo bastante simples. Considere o seguinte modelo:

$$y_t = \beta_1 x_t + \epsilon_t.$$

O termo β é um escalar a ser estimado por algum método. O termo x_t pode representar alguma defasagem de y_t ou algum termo exógeno que explica y_t . O termo ϵ_t é um ruído branco idêntica e independentemente distribuído.

Dadas essas hipóteses, é natural concluir que x_t e ϵ_t sejam independentes entre si. Essa independência é formalizada pela seguinte condição de momento populacional:

$$E(x_t \epsilon_t) = 0.$$

A equação anterior é exatamente uma condição de momento. A forma de obtê-la usando uma amostra finita de observações é calculando a seguinte média:

$$\frac{\sum_{t=1}^T x_t \epsilon_t}{T} = 0.$$

Essa é a condição de momento que permite estimar β_1 . Além de propor o método para estimar esse parâmetro, Hansen (1982) derivou a distribuição de β_1 , de tal modo que se pudessem proceder a inferências estatísticas. Prosseguindo no exemplo, pode-se estimar β_1 usando essa média, substituindo $\epsilon_t = y_t - \beta_1 x_t$:

$$\begin{aligned} \sum_{t=1}^T x_t (y_t - \beta_1 x_t) = 0 &\implies \\ \hat{\beta}_1 = \frac{\sum_{t=1}^T x_t y_t}{\sum_{t=1}^T x_t^2}. \end{aligned}$$

Claramente, esse é o resultado da estimação de β_1 pelo método de mínimos quadrados ordinários. E, de fato, pode-se demonstrar facilmente que mínimos quadrados ordinários são um caso especial do GMM.

O trabalho de Hansen (1982) a respeito desse método de estimação é geral em duas dimensões explicadas a seguir. A primeira dimensão vale-se do relaxamento da hipótese de que x_t é independente de ϵ_t . De fato, a estimação de β_1 prescinde dessa hipótese se for possível encontrar outra variável, z_t , que seja correlacionada com x_t , todavia não o seja com ϵ_t . Ora, se existe z_t , tal que $E(z_t x_t) \neq 0$ e $E(z_t \epsilon_t) = 0$, pode-se estimar β_1 usando a seguinte média amostral:

$$\begin{aligned} \frac{\sum_{t=1}^T z_t \epsilon_t}{T} = 0 &\implies \sum_{t=1}^T z_t (y_t - \beta_1 x_t) = 0 \implies \\ \hat{\beta}_1 = \frac{\sum_{t=1}^T z_t y_t}{\sum_{t=1}^T z_t x_t}. \end{aligned}$$

De fato, esse é o estimador de variáveis instrumentais, IV, também conhecido como método de mínimos quadrados a dois ou três estágios. Naturalmente, essa forma de estimação também é um caso especial do método generalizado dos momentos. Convém introduzir um exemplo em que a variável explicativa seja endógena, para mostrar o poder do método de estimação GMM.

Exemplo 5.1

Suponha que a teoria econômica estabeleça a seguinte relação:

$$y_t = \beta x_t + u_t,$$

porém, infelizmente, $\text{cov}(x_t, u_t) \neq 0$. Esse modelo não pode ser estimado por mínimos quadrados ordinários, sob pena de se obter um coeficiente estimado viesado. Logo, é recomendável que se usem variáveis instrumentais – no caso, GMM – para corrigir o problema.

Para visualizar o tamanho do problema, as variáveis estocásticas serão simuladas. Com isso, é possível estabelecer qual o valor esperado do coeficiente β , caso ele seja estimado corretamente por GMM. Imagine, portanto, que o modelo seja simulado da seguinte forma:

$$y_t = \phi y_{t-1} + \epsilon_t;$$

$$x_t = \delta y_t + v_t,$$

em que ϵ_t e v_t são ambos ruídos brancos independentes entre si.

A utilização do modelo econômico por mínimos quadrados ordinários gera coeficientes viesados, haja vista que:

$$\begin{aligned} \text{cov}(x_t, u_t) &= \text{cov}(x_t, y_t - \beta x_t) = \text{cov}(x_t, y_t) - \beta \text{var}(x_t) \\ &= \delta \text{var}(y_t) - \beta \left[\delta^2 \text{var}(y_t) + \text{var}(v_t) \right] \\ &= \delta \text{var}(y_t) [1 - \beta \delta \text{var}(y_t)] + \beta \text{var}(v_t) \neq 0. \end{aligned}$$

A seguir, quantifica-se essa relação com 300 observações, fazendo $\phi = 0,8$, $\delta = 0,1$ e ϵ_t e v_t ruídos brancos com variância unitária e média nula. Ante essa configuração, o modelo é:

$$y_t = \beta x_t + u_t = \beta \delta y_t + \beta v_t + u_t.$$

Ora, para que o lado direito iguale-se ao lado esquerdo da equação anterior em termos expectationais – isto é, para que $E(y_t) = \beta \delta E(y_t)$ –, é necessário que $E(v_t) = E(u_t) = 0$ e que $\beta = 1/\delta$. Portanto, se $\delta = 0,1$, é forçoso esperar que $\beta = 10$. Estatisticamente, espera-se que o valor estimado, corrigindo

para endogeneidade, seja ao redor de 10 para o coeficiente β . A seguir, veja os resultados. Primeiro, estimando o modelo simplesmente por mínimos quadrados ordinários, obtém-se:

$$y_t = 0,205x_t + \hat{u}_t.$$

(0,103)

Observe que a magnitude do viés é bastante elevada. Porém, pode-se corrigir o problema usando variável instrumental. Uma variável instrumental natural para esse exemplo simulado é y_{t-1} , pois

$$\begin{aligned}\text{cov}(x_t, y_{t-1}) &= \text{cov}(\delta y_t + v_t, y_{t-1}) \\ &= \delta \text{cov}(\phi y_{t-1} + \epsilon_t, y_{t-1}) \\ &= \delta \phi \text{var}(y_{t-1}) \neq 0.\end{aligned}$$

Por outro lado:

$$\begin{aligned}\text{cov}(u_t, y_{t-1}) &= \text{cov}(y_t - \beta x_t, y_{t-1}) \\ &= \text{cov}\left(\overbrace{y_t - \beta \delta y_t}^{=0} - \beta v_t, y_{t-1}\right) \\ &= -\beta \text{cov}(v_t, y_{t-1}) = 0.\end{aligned}$$

Estimando por GMM e usando como variável instrumental y_{t-1} :

$$y_t = 10,154x_t + \hat{u}_t.$$

(4,275)

Há outro candidato a instrumento também, que é x_{t-1} , pois:

$$\begin{aligned}\text{cov}(x_t, x_{t-1}) &= \text{cov}(\delta y_t + v_t, \delta y_{t-1} + v_{t-1}) \\ &= \delta^2 \text{cov}(\phi y_{t-1} + \epsilon_t, y_{t-1}) \\ &= \delta^2 \phi \text{var}(y_{t-1}) \neq 0; e\end{aligned}$$

$$\begin{aligned}\text{cov}(u_t, x_{t-1}) &= \text{cov}(y_t - \beta x_t, x_{t-1}) \\ &= \text{cov}\left(\overbrace{y_t - \beta \delta y_t}^{=0} - \beta v_t, x_{t-1}\right) \\ &= -\beta \text{cov}(v_t, x_{t-1}) = 0.\end{aligned}$$

Usando como instrumento x_{t-1} , o resultado não é satisfatório, pois o coeficiente está longe do que deveria ser e não é significativo:

$$y_t = 5,408x_t + \hat{u}_t.$$

(5,230)

Não obstante, o exemplo motiva o que segue, pois nada impede de usar os dois instrumentos para estimar β .

Outra forma natural de extensão do modelo é a possibilidade de existir mais de uma variável correlacionada com x_t , porém independente de ϵ_t . Para fins de exposição, suponha existirem as variáveis z_t e m_t . Nesse caso, as condições de momento seriam:

$$E \begin{pmatrix} m_t \epsilon_t \\ z_t \epsilon_t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

De novo, a idéia é obter as médias amostrais, usando essas variáveis instrumentais:

$$\begin{aligned} \sum_{t=1}^T z_t (y_t - \beta_1 x_t) &= 0; \\ \sum_{t=1}^T m_t (y_t - \beta_1 x_t) &= 0. \end{aligned}$$

O problema evidente aqui é a existência de duas equações e apenas uma incógnita, ou seja, há duas condições de momento para estimar apenas um parâmetro. Esse exemplo simples mostra o problema geral do método de momentos, que é a possibilidade de existir mais momentos que os parâmetros a serem estimados. Nesse caso, mais uma vez, Hansen (1982) trata dessa possibilidade da seguinte forma: ele propõe ponderar cada um dos momentos de modo a obter os parâmetros que minimizam uma função quadrática. Em outras palavras, ante a existência de mais momentos que parâmetros a serem estimados, o sistema de equações fica sobreidentificado. Nessas condições, a idéia é ponderar cada um dos momentos de uma determinada maneira e encontrar um função quadrática que possa ser minimizada. É *mais ou menos a idéia da estimação de mínimos quadrados*, porém no caso dos momentos é a matriz de *covariância dos momentos* que serve como ponderador, em vez de ser a matriz de variância dos parâmetros. A ponderação tem uma intuição simples: os momentos de maior variância deverão ser aqueles a receber menor peso. Isso ocorrerá naturalmente se tomar o inverso da matriz de covariância dos momentos a ser definida oportunamente.

Formalmente, a idéia é encontrar uma matriz W , simétrica e positiva semidefinida de forma a minimizar:

$$E \left[\begin{pmatrix} m_t \epsilon_t & z_t \epsilon_t \end{pmatrix} W \begin{pmatrix} m_t \epsilon_t \\ z_t \epsilon_t \end{pmatrix} \right].$$

Em termos de médias amostrais, significa:

$$\min_{\beta_1} \left(\sum_{t=1}^T m_t (y_t - \beta_1 x_t) \quad \sum_{t=1}^T z_t (y_t - \beta_1 x_t) \right) W \begin{pmatrix} \sum_{t=1}^T m_t (y_t - \beta_1 x_t) \\ \sum_{t=1}^T z_t (y_t - \beta_1 x_t) \end{pmatrix}.$$

Supondo que W seja a matriz identidade, significando que os momentos têm igual peso, o problema é encontrar β de forma a minimizar:

$$\begin{aligned} \min_{\beta_1} & \left(\sum_{t=1}^T m_t (y_t - \beta_1 x_t) \quad \sum_{t=1}^T z_t (y_t - \beta_1 x_t) \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sum_{t=1}^T m_t (y_t - \beta_1 x_t) \\ \sum_{t=1}^T z_t (y_t - \beta_1 x_t) \end{pmatrix} \\ & \Leftrightarrow \min_{\beta_1} \left(\sum_{t=1}^T m_t (y_t - \beta_1 x_t) \quad \sum_{t=1}^T z_t (y_t - \beta_1 x_t) \right) \begin{pmatrix} \sum_{t=1}^T m_t (y_t - \beta_1 x_t) \\ \sum_{t=1}^T z_t (y_t - \beta_1 x_t) \end{pmatrix} \\ & \Leftrightarrow \min_{\beta_1} \left[\sum_{t=1}^T m_t (y_t - \beta_1 x_t) \right]^2 + \left[\sum_{t=1}^T z_t (y_t - \beta_1 x_t) \right]^2. \end{aligned}$$

Resolvendo essa minimização, ela resulta em:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T \sum_{j=1}^T y_t x_j (m_t m_j + z_t z_j)}{\sum_{t=1}^T \sum_{j=1}^T x_t x_j (m_t m_j + z_t z_j)}.$$

Hansen (1982) explicita qual é a matriz ideal para W . Nas seções posteriores, discute-se como encontrar essa matriz. Além disso, a contribuição do autor nesse caso foi desenvolver um teste para identificar se um determinado momento adicional ao modelo contribui para a estimação do parâmetro β . Se o momento adicional não contribuir, rejeita-se a hipótese nula, significando que a variável instrumental adicional ao modelo está correlacionada com o erro.

Exemplo 5.2

Continuando o exemplo anterior, mas usando x_{t-1} e y_{t-1} como instrumentos, obtém-se:

$$y_t = 10,049x_t + \hat{u}_t.$$

(4,2 (8))

Note a melhora de precisão e de variância do coeficiente estimado em relação ao exemplo anterior.

As próximas seções aprofundam o estudo do GMM, usando como referências Hall (2005), Hamilton (1994), Cochrane (2005) e as notas de aula de Sheppard (2006). Pretende-se mostrar, em particular, que o método GMM permite eliminar hipóteses bastante fortes a respeito da distribuição dos erros, ao mesmo tempo que pode ser aplicado em uma família enorme de problemas econômicos. De fato, um dos grandes atrativos do GMM é a total inexistência de hipóteses sobre a distribuição dos erros. Em contrapartida, trata-se de um método cujo desempenho em pequenas amostras é bastante deficiente.

5.2 Especificação

Esta seção formaliza a especificação do GMM em termos de seus momentos populacionais e sua contrapartida amostral. Para isso, considere inicialmente a seguinte definição:

Definição 5.1 (*Condição de Momento Populacional*) Sejam w_t um vetor de variáveis aleatórias, θ_0 um vetor de parâmetros $k \times 1$ e $g(\cdot)$ um vetor $m \times 1$ de funções reais. A condição de momento populacional é definida como:

$$E[g(w_t, \theta_0)] = 0.$$

A contrapartida amostral dessa condição de momento é imediata e corresponde a:

$$g_T(w, \theta) = \frac{\sum_{t=1}^T g(w_t, \theta)}{T}.$$

Note que $g_T(w, \theta)$ representa a condição de momento e deve ser comparada a $E[g(w_t, \theta_0)]$. Observe ainda que o subscrito t inexistente na condição de momento amostral.

O estimador GMM é definido como o valor de θ que resolve o seguinte problema:

$$\begin{aligned} \min_{\theta} J_T(\theta) &= g_T(w, \theta)' W_T g_T(w, \theta) \\ \hat{\theta}^{GMM} &= \arg \min_{\theta} J_T(\theta), \end{aligned}$$

em que W_T é uma matriz $m \times m$ positiva semidefinida. Essa matriz deve convergir em probabilidade para uma matriz definida e poderá depender dos dados.

Naturalmente o número de momentos, m , deve ser maior que o número de parâmetros, k , sob pena de não ser possível estimar o modelo. De fato, a especificação do GMM pode ser:

- subidentificada, isto é, $m < k$;

- identificada, isto é, $m = k$;
- sobreidentificada, isto é, $m > k$.

5.3 Estimação

Em princípio, pode-se diferenciar J_T com relação a θ , para encontrar a solução do problema. Assim, as condições de primeira ordem são:

$$2G_T(w, \theta)' W_T g_T(w, \theta) = 0,$$

em que a matriz $G_T(w, \theta)$, de dimensão $m \times k$, é o jacobiano das condições de momento tomadas com respeito a θ :

$$G_T(w, \theta) = T^{-1} \sum_{t=1}^T \frac{\partial g(w_t, \theta)}{\partial \theta'}.$$

As condições de primeira ordem são obtidas com relação a funções que podem ser muito não-lineares. Logo, o problema pode não ter solução analítica e precisar de métodos numéricos para ser resolvido. Além disso, é preciso ter cuidado para obter uma solução que represente o mínimo global, o que pode ser computacionalmente difícil.

A matriz W_T é uma parte importante do problema de otimização, sobretudo quando $m > k$. A questão fundamental é: qual a matriz ideal de pesos? Pode-se mostrar que a matriz ideal de pesos é $W_T = S^{-1}$, em que:

$$S = \lim_{T \rightarrow \infty} \text{var} \left(T^{1/2} g_T(w, \theta_0) \right).$$

A matriz S representa a covariância de longo prazo dos momentos. A intuição para o resultado já foi mencionada e é repetida por conveniência. Momentos de variância maior devem ter um peso menor. Isso é obtido invertendo-se a matriz de covariância dos momentos.

Para obter essa covariância, é preciso usar o parâmetro θ_0 , que é normalmente desconhecido. De modo que, para resolver o problema, será preciso, pois, usar algum método de estimação de múltiplos estágios. No primeiro estágio, impõe-se algum W_T . O parâmetro então resultante, $\hat{\theta}_1$ pode ser usado para calcular a matriz de covariância, \hat{S}_1 , e assim sucessivamente. É comum impor inicialmente a matriz identidade, I_m .

Admitindo-se que se usa um método de estimação consistente para S , há três métodos de estimação principais para o GMM. O primeiro é o método

a dois estágios, em que se impõe a matriz identidade no primeiro passo e, no passo seguinte, encontram-se os parâmetros do modelo. O segundo é o método seqüencial, que é uma repetição do método a dois estágios, em que cada novo vetor de parâmetros é usado para estimar a nova matriz de covariância. Finalmente, o último método importante é aquele de atualização contínua dos parâmetros, em que a matriz de covariância é simultaneamente estimada com o vetor de parâmetros. A seguir, discutem-se alguns detalhes de cada método.

No método a dois estágios, impõe-se no primeiro passo a matriz identidade para W_T e encontra-se o vetor de parâmetros $\hat{\theta}_1$, de forma a resolver:

$$G_T(w, \theta)' g_T(w, \theta) = 0.$$

Desse resultado e na ausência de momentos autocorrelacionados, estima-se

$$\hat{S}(w, \hat{\theta}_1) = \frac{\sum_{t=1}^T g(w_t, \hat{\theta}_1) g(w_t, \hat{\theta}_1)'}{T},$$

a ser utilizado no segundo estágio.

Exemplo 5.3

No exemplo da seção anterior, encontrou-se:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T \sum_{j=1}^T y_t x_j (m_t m_j + z_t z_j)}{\sum_{t=1}^T \sum_{j=1}^T x_t x_j (m_t m_j + z_t z_j)}.$$

Supondo que os momentos não são autocorrelacionados, a matriz de covariância pode ser obtida calculando-se:

$$\begin{aligned} \hat{S}(w, \hat{\beta}_1) &= T^{-1} \sum_{t=1}^T \begin{pmatrix} m_t (y_t - \hat{\beta}_1 x_t) \\ z_t (y_t - \hat{\beta}_1 x_t) \end{pmatrix} \begin{pmatrix} m_t (y_t - \hat{\beta}_1 x_t) & z_t (y_t - \hat{\beta}_1 x_t) \end{pmatrix} \\ &= T^{-1} \sum_{t=1}^T \begin{pmatrix} [m_t (y_t - \hat{\beta}_1 x_t)]^2 & (y_t - \hat{\beta}_1 x_t)^2 z_t m_t \\ (y_t - \hat{\beta}_1 x_t)^2 z_t m_t & [z_t (y_t - \hat{\beta}_1 x_t)]^2 \end{pmatrix}. \end{aligned}$$

O método de estimação seqüencial é uma repetição do método de estimação a dois estágios, objetivando encontrar os parâmetros do modelo. A iteração encerra-se quando:

$$\max |\hat{\theta}_l - \hat{\theta}_{l-1}| < \epsilon,$$

em que

l representa o número de iterações;

ϵ representa o critério de convergência, por exemplo, $\epsilon < 10^{-6}$.

É importante mencionar que o procedimento seqüencial tende a ter melhores propriedades de pequenas amostras do que o modelo a dois estágios, não obstante ambos sejam assintoticamente equivalentes.

Nota 5.1 *É possível fixar o número máximo de iterações, haja vista que, se o estimador a dois estágios é consistente, o método seqüencial também o será. Nesse sentido, o critério de encerramento fica sendo $\max |\hat{\theta}_I - \hat{\theta}_{I-1}| < \epsilon$, se $I < I_{\max}$, ou $\hat{\theta}_{I_{\max}}$, se $|\hat{\theta}_{I_{\max}} - \hat{\theta}_{I_{\max}-1}| > \epsilon$.*

No critério de atualização contínua, o modelo encontra θ , que satisfaz

$$\min_{\theta} J_T(\theta) = g_T(w, \theta)' S^{-1}(w, \theta) g_T(w, \theta).$$

Logo, trata-se de uma forma muito mais complicada de estimação, até porque as condições de primeira ordem são bem diferentes das condições de primeira ordem do problema anterior. Se o estimador converge, trata-se do método cujas propriedades de pequenas amostras são bem melhores que as anteriores. Sheppard (2006) sugere usar o estimador seqüencial para encontrar os valores iniciais da função a ser minimizada e depois utilizar o estimador de atualização contínua. Do contrário, não será incomum encontrar W_T bem grande, o que fará com que a J_T seja pequena, por meio da matriz de ponderação e não pela redução de g_T a zero. Além disso, os parâmetros obtidos quando isso ocorre são inconsistentes.

5.4

Propriedades do Estimador

Fundamentalmente, o GMM é um estimador consistente e assintoticamente normal se $g(w_t, \theta)$ são ergódicos e possuem poucos momentos. Se isso ocorre, então são necessárias algumas condições técnicas adicionais para que as estimativas tenham sentido.

Outras condições são as seguintes. Como já mencionado, a matriz W_T deve ser positiva definida para que J_T tenha mínimo único. É preciso também que o sistema seja identificado. Ocorre que a identificação é problemática, pois há modelos razoáveis que admitem vários resultados (veja capítulo 3 de Hall, 2005, para mais detalhes). Para obter identificação, será necessário que o posto de G_T , quando avaliado em θ_0 , seja igual ao número de parâmetros do modelo.

O estimador GMM será consistente para qualquer escolha da matriz de ponderação W_T . Isto é, dada qualquer matriz positiva definida W_T , então:

$$\hat{\theta} - \theta_0 \xrightarrow{p} 0.$$

O estimador é consistente para qualquer matriz W_T porque os momentos populacionais são nulos em um único ponto, justamente aquele em que $J(\theta)$ é minimizada.

Quanto à normalidade do estimador, a covariância vai depender da forma de estimação dos parâmetros, que, por sua vez, dependem de como se determina W_T . Se W_T é eficiente, isto é, se $W_T = S^{-1}$, então o estimador é assintoticamente normal com matriz de covariância expressa de forma matematicamente simples²:

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G'S^{-1}G)^{-1}\right).$$

Se a matriz W_T não for eficiente, a expressão da covariância do estimador torna-se mais complicada:

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N\left(0, (G'WG)^{-1}G'WSWG(G'WG)^{-1}\right).$$

Uma substituição simples mostra que, se $W = S^{-1}$, a fórmula mais complicada reduz-se à anterior.

O que se pode dizer a respeito das condições de momento do modelo? Se W_T é eficiente, os momentos serão também assintoticamente normais, conforme a seguinte representação:

$$\sqrt{T}W_T^{\frac{1}{2}}\hat{g}_T(w, \hat{\theta}) \xrightarrow{d} N\left(0, I_m - W^{\frac{1}{2}}G(G'WG)^{-1}G'W^{\frac{1}{2}}\right).$$

De onde vem essa matriz de covariância complicada? Se o vetor de parâmetros populacionais fosse conhecido, a expressão $W_T^{\frac{1}{2}}\hat{g}_T(w, \hat{\theta})$ seria assintoticamente normal com variância unitária. Essa variância é representada pela matriz I_m . O termo seguinte da matriz de covariância resulta da necessidade de se estimar um vetor de parâmetros desconhecidos. Consequentemente, cada parâmetro estimado resulta na perda de um grau de liberdade; logo, se se estimam k parâmetros, o número de graus de liberdade da matriz de covariância (ou posto) será de $m - k$. Em particular, suponha um modelo em que $m = k$, isto é, o modelo é exatamente identificado, tal que $\hat{g}_T(w, \hat{\theta}) = 0$. Nesse caso, fazendo $W_T = I$ e $G = X$, a covariância será dada por:

$$I_{k=m} - X(X'X)^{-1}X',$$

² Observe que G é uma matriz $m \times k$. Logo, a dimensão de $G'S^{-1}G$ será $k \times k$, em que k é a dimensão do vetor de parâmetros.

que é uma matriz idempotente conhecida da regressão linear que projeta os erros de estimação (dos momentos).

Se, por outro lado, a escolha de W_T é ótima, então $W = S^{-1}$, de modo que, substituindo essa matriz na covariância, resulta que a covariância de g_T é:

$$\text{var}(\hat{g}_T) = S - G \left(G' S^{-1} G \right)^{-1} G'.$$

Para ver isso, note que:

$$\begin{aligned} \text{var} \left(W_T^{\frac{1}{2}} \hat{g}_T(w, \hat{\theta}) \right) &= W_T^{\frac{1}{2}} E \left[\hat{g}_T(w, \hat{\theta}) \hat{g}_T(w, \hat{\theta})' \right] W_T^{\frac{1}{2}} \\ &= I_m - W^{\frac{1}{2}} G \left(G' W G \right)^{-1} G' W^{\frac{1}{2}}. \end{aligned}$$

Substituindo W por S^{-1} :

$$S^{-\frac{1}{2}} E \left[\hat{g}_T(w, \hat{\theta}) \hat{g}_T(w, \hat{\theta})' \right] S^{-\frac{1}{2}} = I_m - S^{-\frac{1}{2}} G \left(G' S^{-1} G \right)^{-1} G' S^{-\frac{1}{2}}.$$

Pré e pós-multiplicando a expressão em ambos os lados por $S^{\frac{1}{2}}$, o resultado se segue.

Se W_T não é escolhido de forma eficiente, então a distribuição das condições de momento fica:

$$\sqrt{T} W_T^{\frac{1}{2}} \hat{g}_T(w, \hat{\theta}) \xrightarrow{d} N \left(0, N W^{\frac{1}{2}} S W^{\frac{1}{2}} N' \right),$$

em que $N = I_m - W^{\frac{1}{2}} G \left(G' W G \right)^{-1} G' W^{\frac{1}{2}}$.

5.5

Estimando a Autocovariância

Para inferência, é necessário calcular a variância de longo prazo. Em um modelo univariado, isso significa somar os termos de autocovariância de uma variável aleatória, isto é, calcular a expressão:

$$\hat{S} = \frac{\sum_{t=1}^T \sum_{j=0}^{T-1} y_t y_{t-j}}{T}.$$

Entretanto, no caso multivariado, em que há, por exemplo, seqüências de vários momentos distintos, a autocovariância não é simétrica. Para ver esse fato, defina a função de autocovariância cruzada entre duas variáveis aleatórias, y e x , como:

$$\gamma_{xy}(j) = E \left[(x_t - \mu_x) (y_{t-j} - \mu_y) \right], j = 0, 1, 2, \dots$$

$$\gamma_{yx}(j) = E \left[(y_t - \mu_y) (x_{t-j} - \mu_x) \right], j = 0, 1, 2, \dots$$

Pode-se mostrar que:

$$\gamma_{xy}(j) = \gamma_{yx}(-j).$$

A função de correlação cruzada pode ser definida como:

$$\rho_{xy}(j) = \frac{\gamma_{xy}(j)}{\sigma_x \sigma_y}, j = 0, \pm 1, \pm 2, \dots$$

Nota 5.2 Em geral, $\rho_{xy}(j) \neq \rho_{xy}(-j)$; logo, a função de correlação cruzada não é necessariamente simétrica ao redor de $j = 0$. Para ver isso, considere a forma de estimação de $\gamma_{xy}(j)$ e $\gamma_{yx}(j)$, sendo, por simplicidade, $\mu_x = \mu_y = 0$:

$$\hat{\gamma}_{xy}(j) = \frac{\sum_{t=j+1}^T x_t y_{t-j}}{T} \neq \frac{\sum_{t=j+1}^T x_{t-j} y_t}{T} = \hat{\gamma}_{yx}(j).$$

É preciso convencer-se de que os dados para calcular a expressão da direita diferem daqueles usados para calcular a expressão da esquerda. Logo, apenas por sorte a autocovariância será simétrica.

Com essa noção em mente, pode-se definir a função de autocovariância multivariada como:

$$\Gamma_j = E(g_t g'_{t-j}),$$

em que se denota

$$g_t = g(w_t, \hat{\theta}) = \frac{\sum_{t=1}^T g(w_t, \hat{\theta})}{T}.$$

Ou seja, tomam-se os momentos centrados em sua média para evitar a inconsistência do estimador, caso a condição de momento populacional não seja válida, isto é, caso $E[g(w_t, \theta)] \neq 0$. O problema de adotar esse procedimento ocorrerá quando a condição de momento for válida, pois, nessas circunstâncias, expurgar a média reduz a variação dos momentos e, assim, a matriz de covariância.

A contraparte amostral dessa autocovariância pode ser estimada por:

$$\hat{\Gamma}_j = \frac{\sum_{t=j+1}^T g_t g'_{t-j}}{T}.$$

Para calcular S , segue-se Hall (2005). Por conveniência, repete-se a definição de S :

$$\begin{aligned}
 S &:= \lim_{T \rightarrow \infty} \text{var} \left(T^{\frac{1}{2}} g_T(w, \theta_0) \right) \\
 &= \lim_{T \rightarrow \infty} E \left[T^{-\frac{1}{2}} \sum_{t=1}^T g_t - E \left(T^{-\frac{1}{2}} \sum_{t=1}^T g_t \right) \right] \times \left[T^{-\frac{1}{2}} \sum_{t=1}^T g_t - E \left(T^{-\frac{1}{2}} \sum_{t=1}^T g_t \right) \right]' \\
 &= \lim_{T \rightarrow \infty} E \left\{ T^{-1} \left(\sum_{t=1}^T [g_t - E(g_t)] \right) \times \left(\sum_{t=1}^T [g_t - E(g_t)]' \right) \right\} \\
 &= \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \sum_{s=1}^T E \{ [g_t - E(g_t)] [g_s - E(g_s)]' \}.
 \end{aligned}$$

A última expressão pode ser mais bem visualizada imaginando-se uma matriz $T \times T$, em que é necessário somar todos os seus termos. Uma forma de fazer isso é somando seus termos diagonais. Na diagonal principal, somam-se:

$$\sum_{t=1}^T E \{ [g_t - E(g_t)] [g_t - E(g_t)]' \} = T\Gamma_0.$$

Nas duas diagonais adjacentes à principal, somam-se:

$$\begin{aligned}
 &\sum_{t=2}^T E \{ [g_t - E(g_t)] [g_{t-1} - E(g_{t-1})]' \} \\
 &\quad + \sum_{t=2}^T E \{ [g_{t-1} - E(g_{t-1})] [g_t - E(g_t)]' \} = (T-1) (\Gamma_1 + \Gamma_1').
 \end{aligned}$$

Seguindo a mesma lógica para as diagonais seguintes, tem-se, para a diagonal que dista j linhas e colunas da principal:

$$\begin{aligned}
 &\sum_{t=j+1}^T E \{ [g_t - E(g_t)] [g_{t-j} - E(g_{t-j})]' \} \\
 &\quad + \sum_{t=j+1}^T E \{ [g_{t-j} - E(g_{t-j})] [g_t - E(g_t)]' \} = (T-j) (\Gamma_j + \Gamma_j').
 \end{aligned}$$

Conseqüentemente,

$$S = \Gamma_0 + \lim_{T \rightarrow \infty} \left[\sum_{j=1}^{T-1} \left(\frac{T-j}{T} \right) (\Gamma_j + \Gamma_j') \right] = \Gamma_0 + \sum_{j=1}^{\infty} (\Gamma_j + \Gamma_j').$$

A contraparte amostral da autocovariância multivariada é dada por:

$$\hat{S} = \Gamma_0 + \sum_{j=1}^T (\Gamma_j + \Gamma_j').$$

Esse estimador dificilmente será simétrico e, portanto, positivo definido. Além disso, é impossível estimar todas as autocovariâncias, particularmente quando j estiver próximo de T . Se fossem estimadas, a variância da estimativa seria elevada em razão do número de observações baixo e da perda dos graus de liberdade. Logo, é preciso introduzir estrutura sobre ele.

Se não houver autocorrelação dos momentos, pode-se simplesmente calcular a seguinte estimativa consistente:

$$\hat{S} := \frac{\sum_{t=1}^T g(w_t, \theta) g(w_t, \theta)'}{T}.$$

Se houver autocorrelação dos momentos, pode-se usar o estimador de covariância de Newey-West, o qual assegura que a matriz de covariância de longo prazo seja positiva definida. O estimador de Newey-West pondera a covariância amostral de tal modo que covariâncias mais distantes do presente têm peso nulo. Para calcular a covariância amostral, é necessário definir a função peso e janela (ou defasagem) máxima de autocovariância, M . A definição dessa janela depende do problema em tela, porém os pacotes econométricos sugerem automaticamente o valor ótimo dessa janela. Por exemplo, o número de defasagens proporcional a $T^{1/5}$ minimiza o erro quadrático médio assintótico do estimador da covariância, cuja fórmula é, usando a função peso de Bartlett:

$$\hat{S}^{NW} = \Gamma_0 + \sum_{j=1}^M \frac{M+1-j}{M+1} (\Gamma_j + \Gamma_j').$$

Outra possibilidade para calcular a matriz de covariância de longo prazo é usando a densidade auto-regressiva espectral. Isto é, se as condições de momento seguem um $AR(p)$, pode-se estimar:

$$g_t = A_1 g_{t-1} + A_2 g_{t-2} + \dots + A_p g_{t-p} + \epsilon_t,$$

em que

ϵ_t é um vetor $m \times 1$;

A_i são matrizes $m \times m$.

Uma forma de escolher a defasagem p é usando o critério de informação *BIC*. Se o modelo estiver correto, a covariância amostral de ϵ_t será dada por:

$$\hat{S}_\epsilon = \frac{\sum_{t=p+1}^T \hat{\epsilon}_t \hat{\epsilon}_t'}{T-p}.$$

Dessa forma, a covariância de longo prazo pode ser obtida por:

$$\hat{S}^{AR} = (I - \hat{A}_1 - \hat{A}_2 - \dots - \hat{A}_p)^{-1} \hat{S}_\epsilon (I - \hat{A}_1 - \hat{A}_2 - \dots - \hat{A}_p)^{-1'}.$$

O estimador espectral da variância necessita de menos parâmetros para ser estimado, comparado ao estimador de Newey-West.

É possível combinar o estimador de covariância de Newey-West com o estimador espectral. Tal procedimento foi batizado como “pré-branqueamento” seguido de “coloração”. O procedimento supõe que os momentos tenham um alto grau de persistência, de forma que algumas defasagens auto-regressivas possam representar essa persistência. Como permanece alguma persistência nos dados, o estimador de Newey-West pode ser acionado para fazer esse ajuste fino e estimar consistentemente a covariância dos resíduos. O procedimento é interessante porque a quantidade de parâmetros a ser estimada permanece baixa.

Para visualizar o procedimento, suponha que os momentos sigam um processo auto-regressivo de ordem p :

$$g_t = \sum_{i=1}^p A_i g_{t-i} + \epsilon_t.$$

Inicialmente, obtenha a covariância dos resíduos usando Newey-West:

$$\hat{S}_\epsilon^{NW} = \Xi_0 + \sum_{j=1}^M \frac{M+1-j}{M+1} (\Xi_j + \Xi_j'),$$

em que

$$\Xi_j = \frac{\sum_{t=j+1}^T \epsilon_t \epsilon_{t-j}'}{T}.$$

A matriz de covariância de longo prazo combina os dois resultados:

$$\hat{S}^{PWRC} = \left(I - \sum_{i=1}^p \hat{A}_i \right)^{-1} \hat{S}_\epsilon^{NW} \left(I - \sum_{i=1}^p \hat{A}_i \right)^{-1'}.$$

O Eviews faz o “pré-branqueamento” sob a suposição de $p = 1$. Mesmo que $p > 1$, em princípio essa imprecisão não teria problema, pois a matriz de

covariância ajustada pelos resíduos é consistentemente estimada. Para selecionar o número de defasagens, o Eviews usa três métodos. No primeiro, a janela é fixada pelo usuário ou pelo critério de Newey-West:

$$M = \text{int} \left[4 \left(\frac{T}{100} \right)^{\frac{2}{5}} \right].$$

Os outros dois métodos de escolha dependem de a função de ponderação ser de Bartlett ou quadrática:

$$M = \begin{cases} \text{int} \left[1,1447 (\alpha(1) T)^{\frac{1}{3}} \right], & \text{para Bartlett;} \\ \text{int} \left[1,3221 (\alpha(2) T)^{\frac{1}{5}} \right], & \text{para quadrática.} \end{cases}$$

Para selecionar M , pode-se agora usar o critério de Andrews e a variável de Newey-West, que variam a respeito da forma como calculam $\alpha(1)$ e $\alpha(2)$. O método de Andrews é paramétrico e faz a hipótese de que os momentos amostrais seguem um $AR(1)$. O método de Newey-West é não paramétrico e baseado no cálculo da matriz de covariância ponderada pela função de Bartlett ou quadrática.

Neste ponto, é importante lembrar que o estimador com atualização contínua usa a teoria desenvolvida nesta seção para especificar a matriz $S(w, \theta)$:

$$S(w, \theta) = \Gamma_{0,T}(\theta) + \sum_{j=1}^M \omega_{j,T} \left(\Gamma_{j,T}(\theta) + \Gamma'_{j,T}(\theta) \right),$$

$$\Gamma_{j,T}(\theta) = \frac{\sum_{t=j+1}^T g_t g'_{t-j}}{T}.$$

O estimador de atualização contínua, além das propriedades desejáveis de pequenas amostras, é importante porque torna o GMM invariante a alterações na curvatura da função momento. Isso ocorre porque tal alteração afeta simultaneamente a função momento e a matriz de covariância, S , justamente porque é dependente dos parâmetros do modelo.

5.6

Casos Especiais do GMM

Esta seção segue em detalhes Sheppard (2006). O objetivo é mostrar alguns casos especiais do GMM e a conexão entre GMM e os métodos de estimação mais conhecidos.

O autor analisa as diversas classes de estimadores em econometria, definindo três classes distintas: os M-estimadores (pelo máximo), os R-estimadores (pelo

posto) e os L-estimadores (por combinações lineares). Os estimadores de máxima verossimilhança, mínimos quadrados ordinários e GMM, são da primeira classe.

O GMM engloba todos os M-estimadores exceto aqueles baseados na distância mínima clássica – CMD. A idéia é minimizar a distância entre o vetor final de parâmetros em relação ao vetor inicial de parâmetros. A particularidade interessante desse método é que o vetor final de parâmetros em geral possui menos componentes que o vetor inicial. Esse tipo de estimação era interessante quando era difícil estimar modelos não-lineares. Entretanto, o desenvolvimento computacional mais recente tornou desnecessário partir para essa linha de ataque.

Ocorre que o GMM reina absoluto entre as classes dos M-estimadores, com exceção ao CMD, como apontado. De fato, nos primórdios, o GMM era usado para estimar os momentos de uma distribuição. Por exemplo, uma distribuição normal de dados enseja a estimação da média e variância:

$$\mu = \frac{\sum_{t=1}^T y_t}{T}; \quad \sigma^2 = \frac{\sum_{t=1}^T (y_t - \mu)^2}{T}.$$

A condição de momento é dada por:

$$g_T(w, \theta) = \begin{bmatrix} \frac{\sum_{t=1}^T y_t}{T} - \mu \\ \frac{\sum_{t=1}^T (y_t - \mu)^2}{T} - \sigma^2 \end{bmatrix} = T^{-1} \begin{bmatrix} \sum_{t=1}^T \hat{\epsilon}_t \\ \sum_{t=1}^T (\hat{\epsilon}_t^2 - \sigma^2) \end{bmatrix}.$$

Se os dados são idêntica e independentemente distribuídos, a matriz S pode ser estimada consistentemente por:

$$\begin{aligned} \hat{S} - E(\hat{S}) &= T^{-1} \sum_{t=1}^T \begin{bmatrix} \hat{\epsilon}_t \\ \hat{\epsilon}_t^2 - \sigma^2 \end{bmatrix} \begin{bmatrix} \hat{\epsilon}_t \\ \hat{\epsilon}_t^2 - \sigma^2 \end{bmatrix}' \\ &= T^{-1} \begin{bmatrix} \sum_{t=1}^T \hat{\epsilon}_t^2 & \sum_{t=1}^T \hat{\epsilon}_t^3 \\ \sum_{t=1}^T \hat{\epsilon}_t^3 & \sum_{t=1}^T (\hat{\epsilon}_t^2 - \sigma^2)^2 \end{bmatrix} \Rightarrow \\ E(\hat{S} - E(\hat{S})) &= \begin{bmatrix} \sigma^2 & 0 \\ 0 & 3\sigma^4 \end{bmatrix}. \end{aligned}$$

Para estimar $G_T(w, \hat{\theta})$ consistentemente, obtenha:

$$\begin{aligned} G_T(w, \hat{\theta}) &= T^{-1} \sum_{t=1}^T \begin{bmatrix} \frac{\partial (y_t - \mu)}{\partial \mu} & \frac{\partial (y_t - \mu)}{\partial \mu} \\ \frac{\partial (y_t - \mu)}{\partial \sigma^2} & \frac{\partial [(y_t - \mu)^2 - \sigma^2]}{\partial \sigma^2} \end{bmatrix} \\ &= T^{-1} \begin{bmatrix} -T & 0 \\ 0 & -T \end{bmatrix} = -I_2. \end{aligned}$$

Conseqüentemente:

$$\hat{G}_T \hat{S} \hat{G}_T = \hat{S}.$$

Outro caso clássico para o uso do GMM é o de mínimos quadrados ordinários, MQO. A idéia é usar a condição de ortogonalidade entre erros e variáveis explicativas para determinar as condições de momento:

$$g_T(w, \theta) = \frac{X'(y - X\beta)}{T},$$

em que

X é a matriz de variáveis explicativas $T \times k$;

β é vetor de parâmetros $k \times 1$;

y é um vetor de variáveis a serem explicadas $T \times 1$.

A solução é conhecida:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Se os dados provêm de uma seqüência de diferença martingal estacionária, então pode-se estimar a matriz S consistentemente usando:

$$\hat{S} = T^{-1} \sum_{t=1}^T x'_t \hat{\varepsilon}_t \hat{\varepsilon}_t' x_t = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2 x'_t x_t.$$

A matriz G_T pode ser obtida por:

$$G_T(w, \hat{\theta}) = T^{-1} \frac{\partial X'(y - X\beta)}{\partial \beta} = -T^{-1} X'X.$$

A combinação dessas duas expressões resulta:

$$\begin{aligned} (G'S^{-1}G)^{-1} &= \left[(-T^{-1}X'X)' \left(T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2 x'_t x_t \right)^{-1} (-T^{-1}X'X) \right]^{-1} \\ &= \hat{\Sigma}_{XX}^{-1} \hat{S} \Sigma_{XX}^{-1}. \end{aligned}$$

Essa matriz é conhecida como estimador robusto de covariância heterocedástica de White.

O GMM também aninha os estimadores de máxima verossimilhança, MLE, e quasi-máxima verossimilhança, QMLE. O QMLE é usado quando se desconhece a distribuição verdadeira dos dados. A conseqüência de usar QMLE é a violação da igualdade da matriz de informação, de modo a ser necessário calcular os

erros-padrão robustos. As condições de momento do estimador de máxima verossimilhança são os escores da função de verossimilhança, isto é, a primeira derivada:

$$g_T(w, \theta) = \sum_{t=1}^T \nabla_{\theta'} l(w_t, \theta),$$

em que ∇_{θ} representa o vetor gradiente da função de verossimilhança $l(\cdot, \cdot)$.

Se os escores são seqüências de diferenças martingais, S pode ser estimada naturalmente por:

$$\hat{S} = T^{-1} \sum_{t=1}^T \nabla_{\theta'} l(w_t, \theta) (\nabla_{\theta'} l(w_t, \theta))'.$$

A matriz G_T é então obtida por meio da segunda derivada da função de verossimilhança:

$$\hat{G}_T(w, \hat{\theta}) = T^{-1} \sum_{t=1}^T \nabla_{\theta\theta'} l(w_t, \theta).$$

Não é difícil verificar que essas expressões são equivalentes às matrizes de informação da estimação por máxima verossimilhança:

$$\begin{aligned} E(\hat{S}) &= T^{-1} \sum_{t=1}^T E[\nabla_{\theta'} l(w_t, \theta) (\nabla_{\theta'} l(w_t, \theta))'] \\ &= E[\nabla_{\theta'} l(w_t, \theta) (\nabla_{\theta'} l(w_t, \theta))'] = \mathcal{I}. \end{aligned}$$

Além disso,

$$\begin{aligned} E[\hat{G}_T(w, \hat{\theta})] &= T^{-1} \sum_{t=1}^T E[\nabla_{\theta\theta'} l(w_t, \theta)] \\ &= E[\nabla_{\theta\theta'} l(w_t, \theta)] = -\mathcal{J}. \end{aligned}$$

A covariância do GMM é dada por:

$$\hat{G}_T(w, \hat{\theta})'^{-1} \hat{S} \hat{G}_T(w, \hat{\theta}) = \mathcal{J}^{-1} \mathcal{I} \mathcal{J}^{-1}.$$

Se a igualdade da matriz de informação é válida, então $\mathcal{J}^{-1} = \mathcal{I}$, de modo que a expressão anterior simplifica para \mathcal{J}^{-1} , como no caso de máxima verossimilhança. Se, no entanto, trata-se de um caso de QMLE, então é preciso usar a matriz de covariância robusta para evitar problemas de tamanho dos testes.

5.7

Testes Usando GMM

O GMM minimiza uma função representando as condições de momentos devidamente ponderadas. Se essas condições de momentos estiverem corretas, vão ter média 0. Isso conduz naturalmente a um teste de superidentificação usando o valor minimizado da função. Trata-se do teste J , definido como:

$$J = T g_T(w, \theta)' S^{-1} g_T(w, \theta) \sim \chi_{m-k}^2.$$

Ou seja, o número de graus de liberdade dessa função é o excesso de momentos em relação ao número de parâmetros estimados.

Equivalentemente, o teste J pode ser implementado usando:

$$J = T Q_T(\theta).$$

O que significa a rejeição do teste de superidentificação? A rejeição do teste representa a existência de momentos que não são estatisticamente iguais a zero, ou seja, está-se rejeitando o modelo, pois a condição de momento não é válida. Outra interpretação é a seguinte: se a inclusão de mais um momento não implica a rejeição da nula, esse momento é válido e contribui para estimar os parâmetros do modelo.³

Se não é possível um estimador consistente para W_T , deve-se computar o teste usando a seguinte expressão:

$$J^{WT} = T g_T(w, \theta)' \left\{ \left[I_m - G (G'WG)^{-1} G'W \right] S \left[I_m - G (G'WG)^{-1} G'W \right] \right\}^{-1} g_T(w, \theta),$$

que decorre da normalidade assintótica das condições de momento estimadas.

O GMM permite testes não-lineares sobre os parâmetros. Para isso, defina o vetor $q \times 1$ de funções reais $R(\theta)$. A hipótese nula do teste é: $R(\theta) = 0$. Pode-se usar o método delta para determinar a distribuição dessa função, desde que a função $R(\cdot)$ seja continuamente diferenciável em θ e $T^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ seja assintoticamente normal. Para fins de esclarecimento, define-se o método delta a seguir.

³ Há casos de modelos com inúmeros momentos possíveis. Quantos momentos usar? A resposta não é fácil. Mas, é preciso entender que momentos redundantes podem comprometer as propriedades do estimador. Assim, para pequenas amostras, com 50 a 100 observações, Hall (2005) recomenda de 2 a 5 momentos a mais que o número de parâmetros.

Definição 5.2 Seja $T^{\frac{1}{2}} (\hat{\theta} - \theta_0) \xrightarrow{A} N(0, \Sigma)$, em que Σ é a matriz de covariância positiva definida e \xrightarrow{A} representa convergência assintótica em distribuição. Se $R(\theta) : \mathbb{R}^k \rightarrow \mathbb{R}^q$ for continuamente diferenciável em θ , com $q \leq k$, então:

$$T^{\frac{1}{2}} [R(\hat{\theta}) - R(\theta_0)] \xrightarrow{A} N\left(0, \frac{\partial R(\theta_0)}{\partial \theta'} \Sigma \left(\frac{\partial R(\theta_0)}{\partial \theta'}\right)'\right).$$

O teste de Wald pode ser calculado usando a fórmula:

$$W = T \times \left[R(\hat{\theta})' \left[\frac{\partial R(\theta_0)}{\partial \theta'} \Sigma \left(\frac{\partial R(\theta_0)}{\partial \theta'}\right)'\right]^{-1} R(\hat{\theta}) \right].$$

Nota 5.3 O teste de Wald não é padrão quando o posto do jacobiano assumir múltiplos valores, conforme o teste que se queira fazer. Por exemplo: suponha $H_0 : \theta_1 \theta_2 = 0$. Nesse caso, podem ocorrer três possibilidades: $\theta_1 = 0$ e $\theta_2 \neq 0$, $\theta_1 \neq 0$ e $\theta_2 = 0$ ou $\theta_1 = \theta_2 = 0$. Nos dois primeiros casos, o posto do jacobiano

$$\frac{\partial R(\theta)}{\partial \theta} = \begin{bmatrix} \theta_2 \\ \theta_1 \end{bmatrix} \text{ é 1. Mas, no último, não é.}$$

Dadas essas preliminares, considere que os parâmetros estimados têm a seguinte distribuição:

$$T^{\frac{1}{2}} (\hat{\theta} - \theta_0) \xrightarrow{A} N\left(0, (G'S^{-1}G)^{-1}\right).$$

No caso de uma restrição linear $H_0 : R\theta - q = 0$, o teste é idêntico ao caso de mínimos quadrados ordinários:

$$Wald = T (R\hat{\theta} - q)' \left[R (G'S^{-1}G)^{-1} R' \right]^{-1} (R\hat{\theta} - q)' \xrightarrow{A} \chi_q^2,$$

em que o posto de R é q .

No caso de restrições não-lineares, pode-se usar o método delta. Na hipótese de se estimar W de forma ineficiente, sabe-se que:

$$T^{\frac{1}{2}} (\hat{\theta} - \theta_0) \xrightarrow{A} N\left(0, (G'WG)^{-1} G'WSWG (G'WG)^{-1}\right),$$

de modo que

$$Wald = T (R\hat{\theta} - q)' V^{-1} (R\hat{\theta} - q)' \xrightarrow{A} \chi_q^2,$$

em que

$$V = R (G'WG)^{-1} G'WSWG (G'WG)^{-1} R'.$$

O teste de razão de verossimilhança, LR, também é válido. Considere $\hat{\theta}$ o parâmetro não restrito e $\tilde{\theta}$ a solução de

$$\begin{aligned}\tilde{\theta}^{GMM} &= \arg \min_{\theta} Q_T(\theta) \\ \text{sujeito a } R\theta - q &= 0,\end{aligned}$$

em que $Q_T = g_T(w, \theta)' S_{\hat{\theta}}^{-1} g_T(w, \theta)$, sendo $S_{\hat{\theta}}$ uma estimativa da covariância das condições de momento sob a hipótese não restrita. A estatística, então, é obtida por:

$$LR = T \left(g_T(w, \tilde{\theta})' S_{\hat{\theta}}^{-1} g_T(w, \tilde{\theta}) - g_T(w, \hat{\theta})' S_{\hat{\theta}}^{-1} g_T(w, \hat{\theta}) \right) \xrightarrow{A} \chi_q^2.$$

Note que a matriz de covariância deve ser estimada sob a hipótese de parâmetros não restritos e deve ser a mesma para ambos os modelos.

O teste de multiplicador de Lagrange, LM, é obtido resolvendo:

$$\tilde{\theta}^{GMM} = \arg \min_{\theta} Q_T(\theta) + \lambda' (R\theta - q).$$

A estatística é difícil de derivar, mas resulta em:

$$LM = T g_T(w, \tilde{\theta})' S_{\hat{\theta}}^{-1} G \left(G' S_{\hat{\theta}}^{-1} G \right)^{-1} G' S_{\hat{\theta}}^{-1} g_T(w, \tilde{\theta}) \xrightarrow{A} \chi_q^2.$$

O LM é vantajoso quando a nula é mais fácil de estimar do que a alternativa, já que se utiliza apenas dos parâmetros estimados sob a hipótese nula.

5.8

Apreçamento de Ativos

Esta seção discute um problema prático importante em economia. Trata-se da estimação de modelos de apreçamento intertemporais. Como o GMM é diretamente aplicável a esses problemas e foi inicialmente apresentado usando um modelo de equilíbrio geral estocástico de apreçamento de ativos, trata-se de um caso interessante para aprofundar as idéias do GMM e mostrar um modelo imediato decorrente de expectativas racionais.

Considere um agente representativo que maximiza sua utilidade ao longo do tempo. Em particular, e sem perda de generalidade, considere um agente que vive dois períodos e maximiza a seguinte utilidade⁴:

⁴ Veja Cochrane (2005) para detalhes.

$$\max_{\{c_t\}} u(c_t) + \beta E_t [u(c_{t+1})],$$

em que

c_t é o consumo do indivíduo no período t ;

$0 < \beta < 1$ é a taxa de desconto subjetiva por adiar o consumo, representando o valor do futuro para o indivíduo. O indivíduo será mais impaciente à medida que essa taxa for menor;

E_t é o operador esperança condicional às informações até t .

A maximização do agente está sujeita à sua restrição orçamentária, que é dada pelas seguintes equações:

$$c_t = w_t - \sum_{j=1}^n p_{j,t} q_{j,t};$$

$$c_{t+1} = w_{t+1} + \sum_{j=1}^n x_{j,t+1} q_{j,t},$$

em que

w_t é a dotação do indivíduo no período t . Pode-se pensar nisso como o consumo original sem a compra de ativo;

$p_{j,t}$ é o preço do ativo j no período t ;

$q_{j,t}$ é quantidade do ativo j adquirida no período t ;

$x_{j,t}$ é o preço do ativo j comprado em $t - 1$. Por exemplo, poderá ser o preço da ação no período seguinte.

O modelo gera o seguinte resultado de interesse mais diretamente:

$$p_{j,t} = E_t (m_{t+1} x_{j,t+1}),$$

em que $m_{t+1} = \beta(u'(c_{t+1})/u'(c_t))$ é o fator estocástico de desconto ou *price kernel*.

Trata-se, de fato, de uma taxa de desconto, que muda conforme os estados da natureza se revelem. Se não se sabe o que acontecerá amanhã, podem-se adotar cenários. Para cada cenário, há uma taxa de desconto dos fluxos futuros. Essa taxa de desconto é representada por m_t .

Podem-se dividir ambos os lados da equação anterior por p_t para obter a equação fundamental de apreçamento:

$$1 = E_t (m_{t+1} R_{j,t+1}),$$

em que $R_{j,t+1} = x_{j,t+1}/p_{j,t}$ é o retorno bruto (discreto) futuro obtido pela compra do ativo j .

A equação anterior está dizendo, em palavras: se o fator estocástico de desconto exerce bem a sua função de descontar os retornos brutos corretamente, então, em média, os retornos descontados dos diversos ativos vão ser iguais a 1.

Exemplo 5.4

Apreçamento de ativos com dados brasileiros

Issler e Piqueira (2002) usam o método generalizado dos momentos com dados brasileiros e estimam os parâmetros estruturais do modelo CCAPM (consumption capital asset pricing model) por meio de três classes de funções utilidade distintas: função utilidade potência (CRRA), utilidade com hábito externo e aversão ao desapontamento (Kreps-Porteus). Esses parâmetros estruturais estão associados à aversão ao risco, à elasticidade de substituição intertemporal no consumo e à taxa de desconto intertemporal da utilidade futura.

Neste exemplo são efetuadas as estimações do trabalho citado utilizando uma amostra com observações compreendendo o período de 1968 a 2005. Os dados utilizados são o consumo das famílias (c_t), o Ibovespa (ib_t) e a taxa de juros dada pelo CDB (r_t). Todos os dados foram extraídos do site do Ipea. As estimações foram efetuadas utilizando janela variável de Newey e West (1994), pré-branqueamento e kernel quadrático.

CRRA: para esta especificação, a equação de Euler⁵ toma a seguinte forma:

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{t+1}) \right] = 1$$

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + ib_{t+1}) \right] = 1$$

em que o valor de β é interpretado como a taxa de desconto intertemporal, γ é o coeficiente de aversão relativa ao risco. Por meio desses parâmetros, pode-se calcular o valor da elasticidade intertemporal de substituição do consumo dada por $\psi = 1/\gamma$.

A estimação do sistema de equações de Euler foi realizada utilizando janela variável de Newey-West. Os resultados são apresentados a seguir, em que * indica significância a 5%.

1. As estimativas de β são próximas de 0,9.

⁵ A equação de Euler é a equação resultante das condições de primeira ordem do problema de otimização.

CRRA				
β	γ	$T \times J$	Prob.	Instrumentos
0,874* (0.012)	0,177 (0.149)	6,588	36,063%	constante, $(c_{t-1}), (ib_{t-1}), (r_{t-1})$
0,843* (0.0007)	-0,065* (0.001)	23,94	2,073%	constante, $(c_{t-1}), (c_{t-2}), (ib_{t-1}),$ $(ib_{t-2}), (r_{t-1}), (r_{t-2})$
0,862* (0.0001)	-0,080* (0.0002)	1240,49	0,000%	constante, $(c_{t-1}), (c_{t-2}), (ib_{t-1}), (ib_{t-2}),$ $(ib_{t-3}), (r_{t-1}), (r_{t-2}), (r_{t-3})$

2. As estimativas de γ apresentam resultados menos robustos. Tem-se uma estimativa positiva, entre zero e um, mas não significativa. Por outro lado, quando os resultados são significativos, o parâmetro apresenta valor negativo.
3. Foram utilizadas 35 observações para a estimação do modelo, dado que tínhamos 36 originalmente e usamos 1 defasagem de cada variável como instrumento. Então, para o cálculo do teste J, foi utilizado um valor de T igual a 35. Os graus de liberdade são determinados por meio da subtração do número de momentos (igual ao número de instrumentos multiplicado pelo número de equações estimadas) e o número de parâmetros estimados. Assim, para o primeiro sistema, temos 6 graus de liberdade (8 momentos e 2 parâmetros). Com essas informações, infere-se que o primeiro sistema estimado com apenas 1 defasagem de cada variável como instrumento não é rejeitado pelo teste J, ou seja, nenhum dos instrumentos é correlacionado com os erros. Para o segundo sistema, não se rejeita a 5% e rejeita-se a 1% a hipótese de que nenhum dos instrumentos é correlacionado com os erros. Para o terceiro sistema, rejeita-se a hipótese de sobreidentificação.

Hábito Externo: o sistema de equação de Euler para essa modelagem é representado a seguir:

$$E_t \left[\beta \left(\frac{c_t}{c_{t-1}} \right)^{-\kappa(\gamma-1)} \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{t+1}) \right] = 1$$

$$E_t \left[\beta \left(\frac{c_t}{c_{t-1}} \right)^{-\kappa(\gamma-1)} \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + ib_{t+1}) \right] = 1.$$

Os parâmetros β e γ possuem o mesmo significado contido no caso da função utilidade CRRA. A diferença aqui é o parâmetro κ . Ele mede a separabilidade do consumo em relação ao consumo passado na função utilidade. Ou seja, permite a existência de um efeito positivo do consumo passado – entendido como “hábito” do consumidor – na função utilidade instantânea do agente. Assim, impede-se a

Hábito Externo					
β	γ	κ	$T \times J$	Prob.	Instrumentos
0,889* (0,054)	0,047 (0,183087)	-0,899 (0,672)	6,408	26,852%	constante, (c_{t-1}) , (ib_{t-1}) , (r_{t-1})
0,831* (0,007)	-0,264* (0,041)	-0,925* (0,062)	7,595	74,906%	constante, (c_{t-1}) , (c_{t-2}) , (ib_{t-1}) , (ib_{t-2}) , (r_{t-1}) , (r_{t-2})
0,859* (0,015)	0,159* (0,052)	-0,687* (0,086)	11,356	72,697%	constante, (c_{t-1}) , (c_{t-2}) , (ib_{t-1}) , (ib_{t-2}) , (ib_{t-3}) , (r_{t-1}) , (r_{t-2}) , (r_{t-3})

separabilidade da função utilidade e tempo. Para que esse efeito seja corroborado, é necessário que $\kappa > 0$. Os resultados das estimações são apresentados a seguir e, como antes, * indica significância a 5%.

Nenhum dos sistemas de equações de Euler foi rejeitado pelo teste J das restrições de sobreidentificação – lembrando que, desta vez, têm-se, para o primeiro modelo, 35 observações com 5 graus de liberdade. Os valores de β novamente apresentaram valor próximo a 0,9. Em relação a γ , tem-se a mesma situação do caso anterior: se o valor é positivo, não é significativo, e se for significativo, é negativo. Para o parâmetro κ , ele sempre apresenta valor negativo, sendo não significante em apenas uma das estimações. Em resumo, a situação é parecida com a encontrada por Issler e Piqueira (2002), e os resultados não são muito intuitivos.

Kreps-Porteus: essa modelagem permite a separação entre o comportamento observável por causa da aversão ao risco e do atribuível à substituição intertemporal – uma das críticas ao modelo com utilidade CRRA. Sua estrutura é apresentada a seguir:

$$E_t \left[\beta^\eta \left(\frac{c_{t+1}}{c_t} \right)^{\eta(p-1)} ib_{t+1}^{\eta-1} (1 + r_{t+1}) \right] = 1$$

$$E_t \left[\frac{1}{\eta} \left\{ \left[\beta \left(\frac{c_t}{c_{t-1}} \right)^{(p-1)} ib_{t+1} \right]^\eta - 1 \right\} \right] = 0.$$

1. Para essa especificação, o coeficiente β possui a mesma interpretação dada. Para calcular γ , é preciso utilizar a fórmula $\gamma = 1 - \kappa$, em que $\eta = \kappa/\rho$. E calcula-se ψ por $\psi = 1/(1 - \rho)$. O desvio-padrão de γ e ψ foi obtido via método delta.

Os resultados mostram que o valor estimado de β diminui para 0,60. As estimativas de ψ foram significativamente diferentes de zero, evidenciando alguma disposição dos agentes em alterar seu padrão de consumo diante das variações na taxa de juros, ou seja, alguma substituição intertemporal no consumo. Novamente foram estimados valores negativos para a taxa de

aversão ao risco. A estimação de nenhum sistema de equação de Euler foi rejeitada pelo teste de sobreidentificação, isto é, em nenhum dos casos os instrumentos utilizados estavam correlacionados aos erros.

<i>Kreps-Porteus</i>						
β	ρ	ψ	γ	$T \times J$	Prob.	Instrumentos
0,618* (0,027)	0,562* (0,107)	2,288* (0,315)	-0,346* (0,105)	6,048	19,559%	constante, (c_{t-1}), (ib_{t-1}), (r_{t-1})
0,604* (0,015)	0,754* (0,082)	4,077* (1,862)	-0,179* (0,022)	7,735	65,470%	constante, (c_{t-1}), (c_{t-2}), (ib_{t-1}), (ib_{t-2}), (r_{t-1}), (r_{t-2})
0,593* (0,012)	0,736* (0,074)	3,802* (1,156)	-0,025* (0,012)	6,766	97,751%	constante, (c_{t-1}), (c_{t-2}), (ib_{t-1}), (ib_{t-2}), (ib_{t-3}), (r_{t-1}), (r_{t-2}), (r_{t-3})



Os modelos econômicos em geral são expressos por meio de diversas variáveis. Portanto, o uso de modelos univariados como visto até o momento é limitado para expressar modelos econômicos. O vetor auto-regressivo permite que se expressem modelos econômicos completos e se estimem os parâmetros desse modelo. É interessante observar que os modelos em *VAR* definem restrições entre as equações do modelo. Estudar essas restrições e usá-las para identificar os parâmetros estruturais do *VAR* constitui um objetivo fundamental da metodologia.

De modo geral, pode-se expressar um modelo auto-regressivo de ordem p por um vetor com n variáveis endógenas, X_t , que estão conectadas entre si por meio de uma matriz A , conforme segue:

$$AX_t = B_0 + \sum_{i=1}^p B_i X_{t-i} + B\epsilon_t, \quad (11)$$

em que

A é uma matriz $n \times n$ que define as restrições contemporâneas entre as variáveis que constituem o vetor $n \times 1$, X_t ;

B_0 é um vetor de constantes $n \times 1$;

B_i são matrizes $n \times n$;

B é uma matriz diagonal $n \times n$ de desvios-padrão;

ϵ_t é um vetor $n \times 1$ de perturbações aleatórias não correlacionadas entre si contemporânea ou temporalmente, isto é:

$$\epsilon_t \sim i.i.d. (0; I_n).$$

A equação (11) expressa as relações entre as variáveis endógenas, frequentemente decorrentes de um modelo econômico teoricamente estruturado, e por isso chama-se *forma estrutural*. Os choques ϵ_t são denominados choques estruturais porque afetam individualmente cada uma das variáveis endógenas. Os choques estruturais são considerados independentes entre si porque as inter-relações entre um choque e outro são captadas indiretamente pela matriz A . Logo, a independência dos choques dá-se sem perda de generalidade.

Por causa da endogeneidade das variáveis, esse modelo é normalmente estimado em sua *forma reduzida*, isto é, estima-se o seguinte modelo:

$$\begin{aligned} X_t &= A^{-1}B_0 + \sum_{i=1}^p A^{-1}B_i X_{t-i} + A^{-1}B\epsilon_t \\ &= \Phi_0 + \sum_{i=1}^p \Phi_i X_{t-i} + e_t, \end{aligned}$$

em que

$$\Phi_i \equiv A^{-1}B_i, \quad i = 0, 1, \dots, p \quad B\epsilon_t \equiv Ae_t.$$

É conveniente visualizar melhor esse modelo por meio de um exemplo bivariado de ordem 1, inspirado em Enders (2004). A partir dessa simplificação, uma série de resultados intuitivos que valem para modelos de ordem maior pode ser desenvolvida, facilitando o entendimento da metodologia. Considere, então, o seguinte modelo bivariado:

$$\begin{aligned} y_t &= b_{10} - a_{12}z_t + b_{11}y_{t-1} + b_{12}z_{t-1} + \sigma_y \epsilon_{yt}; \\ z_t &= b_{20} - a_{21}y_t + b_{21}y_{t-1} + b_{22}z_{t-1} + \sigma_z \epsilon_{zt}. \end{aligned}$$

Trata-se de uma especificação inicial bem razoável, pela qual as variáveis são mutuamente influenciadas uma pela outra, tanto contemporaneamente como pelos seus valores defasados. Esse modelo não pode ser estimado diretamente, já que ambas as variáveis contemporâneas z_t e y_t são individualmente correlacionadas com os erros ϵ_{yt} ou ϵ_{zt} , respectivamente. Isso ocorre porque cada uma dessas variáveis depende contemporaneamente da outra (efeito *feedback*). O objetivo do VAR é desenvolver técnicas para evitar esse problema, visando-se encontrar a trajetória da variável de interesse ante um choque nesses erros, ou seja, um choque estrutural.

As hipóteses assumidas para esse modelo são:

- a) y_t e z_t são ambos estacionários;
- b) $\epsilon_{yt} \sim RB(0, 1)$ e $\epsilon_{zt} \sim RB(0, 1)$;
- c) $\epsilon_{yt} \perp \epsilon_{zt} \implies \text{Cov}(\epsilon_{yt}, \epsilon_{zt}) = 0$.

Nota 6.1 O efeito feedback se sucede porque y_t e z_t afetam um ao outro.

O modelo apresentado está na *forma estrutural* e pode ser escrito em matrizes:

$$\underbrace{\begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix}}_{\equiv A} \underbrace{\begin{bmatrix} y_t \\ z_t \end{bmatrix}}_{\equiv X_t} = \underbrace{\begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix}}_{\equiv B_0} + \underbrace{\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}}_{\equiv B_1} \underbrace{\begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix}}_{\equiv X_{t-1}} + \underbrace{\begin{bmatrix} \sigma_y & 0 \\ 0 & \sigma_z \end{bmatrix}}_{\equiv B} \underbrace{\begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix}}_{\equiv \epsilon_t}$$

$$\implies AX_t = B_0 + B_1X_{t-1} + B\epsilon_t.$$

A forma reduzida desse modelo simplificado é:

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + e_t; \quad (12)$$

$$\Phi_0 \equiv A^{-1}B_0; \quad \Phi_1 \equiv A^{-1}B_1;$$

$$Ae_t \equiv B\epsilon_t.$$

A condição de estabilidade é ter os autovalores de $(I - \Phi_1 L)$ fora do círculo unitário. Observe que nessa especificação não existem problemas em se estimar o VAR, já que os erros transformados, e_t , não estão correlacionados com os regressores, não são autocorrelacionados, mas são contemporaneamente correlacionados entre si. De fato, é muito importante entender os erros compostos, e_t , para extrair maneiras de identificar os parâmetros adicionais que faltam. Observe que:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \equiv A^{-1}B\epsilon_t = \begin{bmatrix} \frac{\sigma_y \epsilon_{yt} - a_{12} \sigma_z \epsilon_{zt}}{1 - a_{12} a_{21}} \\ \frac{\sigma_z \epsilon_{zt} - a_{21} \sigma_y \epsilon_{yt}}{1 - a_{12} a_{21}} \end{bmatrix}.$$

Dessa forma, é fácil perceber que:

$$E(e_t) = 0;$$

$$\text{Cov}(e_t) \equiv \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \frac{\sigma_y^2 + a_{12}^2 \sigma_z^2}{(1 - a_{12} a_{21})^2} & -\frac{a_{21} \sigma_y^2 + a_{12} \sigma_z^2}{(1 - a_{12} a_{21})^2} \\ -\frac{a_{21} \sigma_y^2 + a_{12} \sigma_z^2}{(1 - a_{12} a_{21})^2} & \frac{\sigma_z^2 + a_{21}^2 \sigma_y^2}{(1 - a_{12} a_{21})^2} \end{bmatrix}.$$

Os erros não são autocorrelacionados, pois:

$$\text{Cov} \left(e_{it}, e_{i(t-j)} \right) = E \left[\left(\frac{\sigma_i e_{it} - a \sigma_{\sim i} e_{\sim i t}}{1 - a_{12} a_{21}} \right) \left(\frac{\sigma_i e_{i(t-j)} - a \sigma_{\sim i} e_{\sim i(t-j)}}{1 - a_{12} a_{21}} \right) \right] = 0,$$

$i = y, z; j \neq 0; a = a_{12}, a_{21}; \sim$ representa negação.

Assim, existe correlação contemporânea nos erros da regressão da forma reduzida. Mais tarde, esse fato será usado para identificar os parâmetros do modelo estrutural. De fato, ao estimar o modelo na forma reduzida, perdem-se informações a tal ponto que, sem algumas hipóteses, normalmente decorrentes da teoria econômica, é impossível identificar os parâmetros estruturais, ou seja, é impossível recuperar os parâmetros do modelo original.

Nota 6.2 *A grande questão dos modelos VAR é se, a partir da forma reduzida, consegue-se recuperar as informações contidas na forma estrutural.*

Neste ponto, é preciso deixar claro por que é necessário recuperar os parâmetros estruturais. Somente por meio de modelos estruturais, corretamente identificados, é possível designar políticas econômicas cujos resultados estarão em conformidade com o esperado. Do contrário, apenas usando a forma reduzida, a imposição de uma determinada política poderá alterar a trajetória das variáveis em razão das expectativas racionais. Isto é, a crítica de Lucas se aplica completamente ao modelo reduzido, exceto se forem impostas hipóteses muito particulares.

Para ter uma idéia de como funciona um sistema de variáveis aleatórias, simulou-se um VAR¹ estacionário na Figura 6.1. As séries superiores são positivamente correlacionadas porque especificaram-se $\phi_{11} = \phi_{22} = 0,6$ e $\phi_{12} = \phi_{21} = 0,2$. As séries inferiores são negativamente correlacionadas, pois alterou-se o modelo anterior, fazendo $\phi_{12} = \phi_{21} = -0,2$.

Se ambas as séries são estacionárias, pode-se estimar o modelo normalmente. Porém, se as séries são não estacionárias, alguns cuidados devem ser tomados e são objeto de cuidadoso estudo nas próximas seções.

É importante ter alguma intuição sobre como se comportam as séries multivariadas não estacionárias. Nesse caso, as raízes da polinomial devem estar sobre o círculo unitário, de forma a mostrar um passeio aleatório das séries. As séries poderão ser individualmente independentes ou poderão estar conectadas entre

¹ Assumindo $a_{12} = 0,8$ e $a_{21} = 0,2$.

si, como em um VAR. Assim, a Figura 6.2 mostra a simulação de um passeio aleatório com e sem *drift*. O caso com *drift* é evidenciado no conjunto superior. O caso sem *drift* é evidenciado no conjunto inferior de séries. Para as séries não estacionárias e sem *drift*, impuseram-se $\Phi = 0$; $\phi_{11} = \phi_{22} = 0,7$ e $\phi_{12} = \phi_{21} = 0,3$. Para as séries com *drift*, definiu-se o modelo anterior, porém alterou-se ϕ_{10} para 0,5. Note que basta uma das constantes ser diferente de zero para que ambas as séries tenham um *drift*.

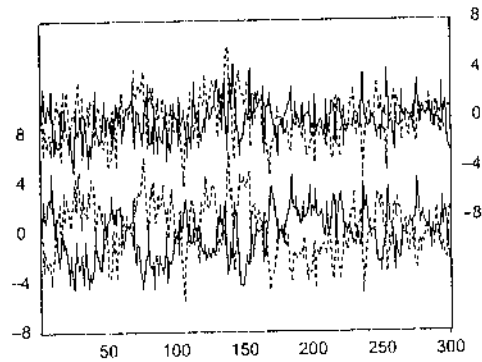


Figura 6.1 VAR (1): séries superiores - eixo à direita: $\Phi_0 = 0$; $\phi_{11} = \phi_{22} = 0,6$ e $\phi_{12} = \phi_{21} = 0,2$; séries inferiores: $\phi_{12} = \phi_{21} = -0,2$. A série pontilhada: z_t .

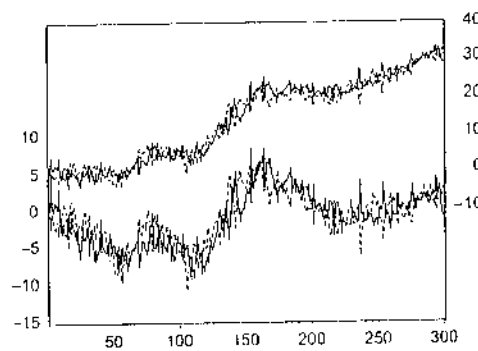


Figura 6.2 VAR (1): séries inferiores - eixo à esquerda: $\Phi = 0$; $\phi_{11} = \phi_{22} = 0,7$ e $\phi_{12} = \phi_{21} = 0,3$; séries superiores: $\phi_{10} = 0,5$. A série pontilhada: z_t .

O que se nota em ambas as figuras é que as séries y_t e z_t movimentam-se conjuntamente, mesmo na presença de raiz unitária. Quando isso acontece, estuda-se o caso que se convencionou chamar cointegração, a ser visto mais tarde.

Por enquanto, o estudo se concentra na discussão sobre o caso de variáveis estacionárias, porém deve-se mencionar que Sims (1980) e Sims, Stock e Watson (1990) admitem a mistura de variáveis estacionárias e não estacionárias em um

modelo VAR. Argumentam que o VAR é uma metodologia fundamentalmente interessada nas inter-relações entre as variáveis. Logo, a forma correta de dimensionar essas inter-relações é por meio de um VAR completo, cujas variáveis são tomadas no nível, em vez de estacionarizá-las por diferenças como foi feito no caso univariado. Trata-se de um ponto controverso e sujeito às necessidades da pesquisa para se definir como proceder.

A generalização do modelo para um VAR de ordem p com variáveis exógenas é direta e feita por meio da seguinte representação econométrica:

$$X_t = \Phi_0 + \sum_{i=1}^p \Phi_i X_{t-i} + GZ_t + e_t,$$

em que

X_t é um vetor $n \times 1$ de variáveis endógenas, como anteriormente;

G é uma matriz de coeficientes $n \times g$;

Z_t é um vetor $g \times 1$ de variáveis exógenas que pode incluir variáveis determinísticas.

Às vezes, convém transformar um VAR (p) em um VAR (1). Isso geralmente é necessário quando se deseja obter a função resposta ao impulso decorrente de um choque estrutural ou para saber se uma determinada estrutura multivariada é estacionária. (A função resposta ao impulso será abordada mais adiante.) Por outro lado, a transformação em um VAR (1) costuma facilitar as manipulações algébricas, razão pela qual se considera o seguinte exemplo.

Exemplo 6.1

Seja um VAR (2):

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + e_t.$$

Como transformá-lo em um VAR (1)? Para isso, reescreva o modelo da seguinte forma:

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_0 \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 \\ I & \mathbf{0} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} e_t \\ \mathbf{0} \end{bmatrix}.$$

Diferentemente dos modelos univariados que se preocupam fundamentalmente com aspectos de previsão, o VAR busca responder qual a trajetória da série, dado um *choque estrutural*. Por trajetória, o pesquisador deseja conhecer o tempo que um choque afeta uma série, se ela muda de patamar ou não, para que patamar vai, entre outras informações. Enquanto os detalhes desse estudo serão

dados mais adiante, é preciso observar que o VAR resulta na estimação de uma infinidade de coeficientes. Um VAR (p), por exemplo, com n variáveis endógenas teria $n + n^2 p$ coeficientes a estimar, já que as matrizes Φ_j têm dimensão $n \times n$ e as n primeiras variáveis referem-se à constante, sem contar ainda os coeficientes de possíveis variáveis exógenas.

Muitas vezes os coeficientes estimados serão estatisticamente insignificantes, até porque algumas variáveis são normalmente colineares, entretanto deve-se evitar impor restrições sobre os coeficientes, sob pena de perder informações relevantes, a menos que sejam restrições econômicas bem fundamentadas.

Nota 6.3 *Discutivelmente, o VAR busca fundamentalmente a trajetória das variáveis endógenas ante um choque estrutural. Nada impede, contudo, de ser usado para previsão também.*

6.1

Especificação de Modelo

Esta seção procura responder à seguinte questão: como selecionar a ordem p de um modelo VAR? Que critérios podem ser utilizados nessa tarefa?

Trata-se de uma tarefa difícil, porque certamente a defasagem necessária para obter “resíduos brancos” com relação à primeira variável endógena não é a mesma para obter na segunda. A regra é usar tantas defasagens quantas forem necessárias para obter “resíduos brancos” em todas as variáveis endógenas. Na prática, é um evento muito difícil, pois há uma probabilidade bastante alta de que a autocorrelação de uma determinada variável seja diferente de zero mesmo se p for alto. Isso acontece porque probabilisticamente 5% dos valores estimados estariam fora do intervalo de confiança. Em contraste, usar um p muito alto, em um modelo altamente complexo em que se estimam muitos coeficientes cruzados, torna o poder do teste estatístico bastante deficiente. Em suma, bom senso e parcimônia devem ser utilizados nesse processo, embora haja alguns procedimentos objetivos a seguir, conforme discute-se agora.

Imediatamente, como no caso univariado, pode-se usar o critério de informação para definir a ordem de defasagem do modelo VAR. Considere um VAR (m), em que $m = 0, 1, 2, \dots, p_{\max}$. O problema é escolher a ordem p que minimiza a seguinte fórmula geral do critério de informação:

$$Cr(m) = \ln |\hat{\Sigma}| + c_T \varphi(m),$$

modelo VAR. Argumentam que o VAR é uma metodologia fundamentalmente interessada nas inter-relações entre as variáveis. Logo, a forma correta de dimensionar essas inter-relações é por meio de um VAR completo, cujas variáveis são tomadas no nível, em vez de estacionarizá-las por diferenças como foi feito no caso univariado. Trata-se de um ponto controverso e sujeito às necessidades da pesquisa para se definir como proceder.

A generalização do modelo para um VAR de ordem p com variáveis exógenas é direta e feita por meio da seguinte representação econométrica:

$$X_t = \Phi_0 + \sum_{i=1}^p \Phi_i X_{t-i} + GZ_t + e_t,$$

em que

X_t é um vetor $n \times 1$ de variáveis endógenas, como anteriormente;

G é uma matriz de coeficientes $n \times g$;

Z_t é um vetor $g \times 1$ de variáveis exógenas que pode incluir variáveis determinísticas.

Às vezes, convém transformar um VAR (p) em um VAR (1). Isso geralmente é necessário quando se deseja obter a função resposta ao impulso decorrente de um choque estrutural ou para saber se uma determinada estrutura multivariada é estacionária. (A função resposta ao impulso será abordada mais adiante.) Por outro lado, a transformação em um VAR (1) costuma facilitar as manipulações algébricas, razão pela qual se considera o seguinte exemplo.

Exemplo 6.1

Seja um VAR (2):

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + e_t.$$

Como transformá-lo em um VAR (1)? Para isso, reescreva o modelo da seguinte forma:

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix}.$$

Diferentemente dos modelos univariados que se preocupam fundamentalmente com aspectos de previsão, o VAR busca responder qual a trajetória da série, dado um *choque estrutural*. Por trajetória, o pesquisador deseja conhecer o tempo que um choque afeta uma série, se ela muda de patamar ou não, para que patamar vai, entre outras informações. Enquanto os detalhes desse estudo serão

si, como em um VAR. Assim, a Figura 6.2 mostra a simulação de um passeio aleatório com e sem *drift*. O caso com *drift* é evidenciado no conjunto superior. O caso sem *drift* é evidenciado no conjunto inferior de séries. Para as séries não estacionárias e sem *drift*, impuseram-se $\Phi = 0$; $\phi_{11} = \phi_{22} = 0,7$ e $\phi_{12} = \phi_{21} = 0,3$. Para as séries com *drift*, definiu-se o modelo anterior, porém alterou-se ϕ_{10} para 0,5. Note que basta uma das constantes ser diferente de zero para que ambas as séries tenham um *drift*.

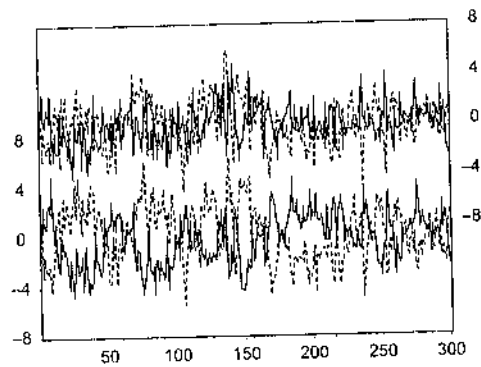


Figura 6.1 VAR (1): séries superiores – eixo à direita: $\Phi_0 = 0$; $\phi_{11} = \phi_{22} = 0,6$ e $\phi_{12} = \phi_{21} = 0,2$; séries inferiores: $\phi_{12} = \phi_{21} = -0,2$. A série pontilhada: z_t .

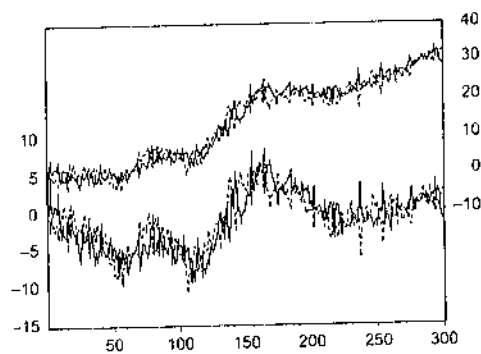


Figura 6.2 VAR (1): séries inferiores – eixo à esquerda: $\Phi = 0$; $\phi_{11} = \phi_{22} = 0,7$ e $\phi_{12} = \phi_{21} = 0,3$; séries superiores: $\phi_{10} = 0,5$. A série pontilhada: z_t .

O que se nota em ambas as figuras é que as séries y_t e z_t movimentam-se conjuntamente, mesmo na presença de raiz unitária. Quando isso acontece, estuda-se o caso que se convencionou chamar cointegração, a ser visto mais tarde.

Por enquanto, o estudo se concentra na discussão sobre o caso de variáveis estacionárias, porém deve-se mencionar que Sims (1980) e Sims, Stock e Watson (1990) admitem a mistura de variáveis estacionárias e não estacionárias em um

modelo VAR. Argumentam que o VAR é uma metodologia fundamentalmente interessada nas inter-relações entre as variáveis. Logo, a forma correta de dimensionar essas inter-relações é por meio de um VAR completo, cujas variáveis são tomadas no nível, em vez de estacionarizá-las por diferenças como foi feito no caso univariado. Trata-se de um ponto controverso e sujeito às necessidades da pesquisa para se definir como proceder.

A generalização do modelo para um VAR de ordem p com variáveis exógenas é direta e feita por meio da seguinte representação econométrica:

$$X_t = \Phi_0 + \sum_{i=1}^p \Phi_i X_{t-i} + GZ_t + e_t,$$

em que

X_t é um vetor $n \times 1$ de variáveis endógenas, como anteriormente;

G é uma matriz de coeficientes $n \times g$;

Z_t é um vetor $g \times 1$ de variáveis exógenas que pode incluir variáveis determinísticas.

Às vezes, convém transformar um VAR (p) em um VAR (1). Isso geralmente é necessário quando se deseja obter a função resposta ao impulso decorrente de um choque estrutural ou para saber se uma determinada estrutura multivariada é estacionária. (A função resposta ao impulso será abordada mais adiante.) Por outro lado, a transformação em um VAR (1) costuma facilitar as manipulações algébricas, razão pela qual se considera o seguinte exemplo.

Exemplo 6.1

Seja um VAR (2):

$$X_t = \Phi_0 + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + e_t.$$

Como transformá-lo em um VAR (1)? Para isso, reescreva o modelo da seguinte forma:

$$\begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} = \begin{bmatrix} \Phi_0 \\ 0 \end{bmatrix} + \begin{bmatrix} \Phi_1 & \Phi_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ X_{t-2} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix}.$$

Diferentemente dos modelos univariados que se preocupam fundamentalmente com aspectos de previsão, o VAR busca responder qual a trajetória da série, dado um *choque estrutural*. Por trajetória, o pesquisador deseja conhecer o tempo que um choque afeta uma série, se ela muda de patamar ou não, para que patamar vai, entre outras informações. Enquanto os detalhes desse estudo serão

dados mais adiante, é preciso observar que o VAR resulta na estimação de uma infinidade de coeficientes. Um VAR (p), por exemplo, com n variáveis endógenas teria $n + n^2p$ coeficientes a estimar, já que as matrizes Φ_i têm dimensão $n \times n$ e as n primeiras variáveis referem-se à constante, sem contar ainda os coeficientes de possíveis variáveis exógenas.

Muitas vezes os coeficientes estimados serão estatisticamente insignificantes, até porque algumas variáveis são normalmente colineares, entretanto deve-se evitar impor restrições sobre os coeficientes, sob pena de perder informações relevantes, a menos que sejam restrições econômicas bem fundamentadas.

Nota 6.3 *Discutivelmente, o VAR busca fundamentalmente a trajetória das variáveis endógenas ante um choque estrutural. Nada impede, contudo, de ser usado para previsão também.*

6.1

Especificação de Modelo

Esta seção procura responder à seguinte questão: como selecionar a ordem p de um modelo VAR? Que critérios podem ser utilizados nessa tarefa?

Trata-se de uma tarefa difícil, porque certamente a defasagem necessária para obter “resíduos brancos” com relação à primeira variável endógena não é a mesma para obter na segunda. A regra é usar tantas defasagens quantas forem necessárias para obter “resíduos brancos” em todas as variáveis endógenas. Na prática, é um evento muito difícil, pois há uma probabilidade bastante alta de que a autocorrelação de uma determinada variável seja diferente de zero mesmo se p for alto. Isso acontece porque probabilisticamente 5% dos valores estimados estariam fora do intervalo de confiança. Em contraste, usar um p muito alto, em um modelo altamente complexo em que se estimam muitos coeficientes cruzados, torna o poder do teste estatístico bastante deficiente. Em suma, bom senso e parcimônia devem ser utilizados nesse processo, embora haja alguns procedimentos objetivos a seguir, conforme discute-se agora.

Imediatamente, como no caso univariado, pode-se usar o critério de informação para definir a ordem de defasagem do modelo VAR. Considere um VAR (m), em que $m = 0, 1, 2, \dots, p_{\max}$. O problema é escolher a ordem p que minimiza a seguinte fórmula geral do critério de informação:

$$Cr(m) = \ln |\hat{\Sigma}| + c_T q(m),$$

em que

$$\hat{\Sigma} = \frac{\sum_{t=1}^T \hat{e}_t \hat{e}_t'}{T};$$

c_T é uma seqüência que depende do tamanho da amostra;

$\varphi(m)$ é uma função que penaliza VAR de grandes ordens. Por exemplo, $\varphi(m)$ pode representar o número de parâmetros estimados no modelo VAR.

Nota 6.4 Como no caso univariado, o tamanho amostral tem de ser mantido constante para tornar o critério de informação comparável. Logo, o tamanho da amostra, comum a todas as ordens, será $T - p_{\max}$.

A versão multivariada dos critérios AIC, BIC e HQ é uma generalização da versão univariada da seguinte forma:

$$\begin{aligned} AIC(m) &= \ln |\hat{\Sigma}(m)| + \frac{2}{T} mn^2; \\ BIC(m) &= \ln |\hat{\Sigma}(m)| + \frac{\ln T}{T} mn^2; \\ HQ(m) &= \ln |\hat{\Sigma}(m)| + \frac{\ln \ln T}{T} 2mn^2, \end{aligned}$$

em que mn^2 é o número total de parâmetros estimados em todas as equações.

De novo, o critério AIC superestima assintoticamente a ordem do VAR com probabilidade positiva, ao passo que os dois últimos critérios estimam a ordem consistentemente sob hipóteses bem gerais, se o processo gerador de dados tiver uma ordem finita no VAR e se $p_{\max} > p$, em que p é a verdadeira ordem do modelo.

Lütkepohl e Krätzig (2004) comentam que, para $T \geq 16$, a seguinte relação ocorre, mesmo para pequenas amostras²:

$$\hat{p}(BIC) \leq \hat{p}(HQ) \leq \hat{p}(AIC).$$

Há outra maneira de escolher a ordem de defasagem, seguindo a idéia dos modelos univariados, em que se aplicam testes seqüenciais para definir a ordem p do modelo VAR. Para começar, estabeleça o p_{\max} e considere $H_0 : \Phi_{p_{\max}} = 0 \times H_1 : \Phi_{p_{\max}} \neq 0$. Se o teste não for rejeitado, repete-se o procedimento

² O programa Eviews possui uma opção automática de seleção da defasagem, exatamente conforme descrito aqui. Inclui também testes pela mesma razão, como descrito na próxima seção.

considerando $p_{\max} = 1$. Quando a nula for rejeitada, terá sido encontrada a ordem de defasagem do modelo.

O problema desse procedimento é estabelecer p_{\max} . Se p_{\max} é muito pequeno, os resíduos estimados não serão um ruído branco, quando se fizerem os testes sobre eles, pós-estimação. Contudo, se p_{\max} é muito grande, o impacto sobre a probabilidade de erros como um todo poderá ser severamente afetado, de modo que é difícil confiar nos intervalos de confiança gerados. Pode-se confiar apenas parcialmente nesses testes, porém, como já foi mencionado, nada substituirá o bom senso do pesquisador para definir quais parâmetros serão considerados insignificantes. Procurando esclarecer melhor esse assunto, quando há muitos parâmetros estimados, resultante de um excesso de defasagens, então a matriz de covariância dos parâmetros tende a gerar intervalos de confiança bastante largos, conduzindo a erros do tipo II.

Exemplo 6.2

Considere o seguinte vetor de variáveis $Z_t = [y_t, e_t, \pi_t, i_t]$, em que y_t representa o hiato do produto, e_t a variação cambial, π_t a taxa de inflação e i_t a taxa de juros (Selic), todas as variáveis em base mensal. Modelos VAR com essas variáveis são comuns na literatura macroeconômica. No Brasil, Minella et al (2003) estimam um modelo semelhante para diversos períodos anteriores ao deste exemplo. Os critérios de Akaike e Schwartz para este modelo são apresentados a seguir.

Defasagens	Akaike	Schwartz
1	-22,088	-21,451
2	-22,498	-21,342
3	-22,026	-20,343
4	-21,366	-19,146
5	-20,943	-18,179
6	-20,788	-17,470

Nessas condições, o modelo indica duas defasagens, segundo o critério Akaike, e uma defasagem segundo o critério BIC.

Exemplo 6.3

Considere o exemplo de endogeneidade estudado no capítulo sobre GMM. Por conveniência, lembre-se de que o modelo foi simulado da seguinte forma:

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t; \\ x_t &= \delta y_t + v_t, \end{aligned}$$

em que ϵ_t e ν_t são ambos ruídos brancos independentes.

Pode-se reescrever esse modelo na forma matricial:

$$\begin{bmatrix} 1 & 0 \\ -\delta & 1 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ \nu_t \end{bmatrix}.$$

Reescrevendo o modelo na forma reduzida, tem-se:

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ \delta\phi & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} e_{yt} \\ e_{xt} \end{bmatrix}.$$

No momento de estimar o modelo, a matriz de coeficientes Φ_1 conterá quatro coeficientes estimados. Espera-se que os coeficientes da coluna 2 sejam insignificantes e que os coeficientes da coluna 1 satisfaçam aproximadamente a seguinte restrição:

$$\begin{aligned} \phi_{11,1} &= \phi; \\ \phi_{21,1} &= \delta\phi, \end{aligned}$$

em que δ e ϕ são valores usados na simulação. No caso da simulação, definiram-se $\delta = 0,1$ e $\phi = 0,8$.

Estimando-se o VAR (1), obtêm-se os seguintes coeficientes e desvios-padrão (entre parênteses):

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0,794 & 0,097 \\ (0,035) & (0,063) \\ 0,077 & -0,002 \\ (2,374) & (0,058) \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \hat{e}_{yt} \\ \hat{e}_{xt} \end{bmatrix}.$$

Note que os coeficientes da coluna 2 são não significantes e que os da coluna 1 são significantes e têm o valor aproximado que era esperado:

$$\begin{aligned} \phi_{11,1} &\simeq 0,8; \\ \phi_{21,1} &\simeq 0,1 \times 0,8 = 0,08. \end{aligned}$$

6.2

Testando Hipóteses

A idéia de testar hipóteses em modelos multivariados é muito semelhante à do caso univariado. A diferença é que, em vez de calcular a soma dos quadrados dos resíduos, calcula-se o determinante da matriz de covariância dos resíduos do modelo restrito e do não restrito, como foi aprendido no caso univariado.

Para iniciar o teste, convém explicitar o modelo geral, com p defasagens, uma constante e g variáveis exógenas, conforme a equação a seguir:

$$X_t = \Phi_0 + \sum_{i=1}^p \Phi_i X_{t-i} + GZ_t + e_t.$$

Definição 6.1 O modelo anterior é estacionário se os autovalores da polinomial $\sum_{i=1}^p \Phi_i L^i$ estiverem dentro do círculo unitário³.

Uma vez especificado o modelo não restrito, estabelecem-se quais as restrições a serem impostas e seguem-se os procedimentos de teste:

1. Estima-se o modelo sem restrição e calcula-se a matriz de covariância dos resíduos, denotada por Σ_u . Essa matriz tem a mesma dimensão do número de variáveis endógenas ou que compõem o vetor X_t . Ou seja, a matriz Σ_u terá dimensão $n \times n$;
2. Em seguida, estima-se o modelo com restrição, excluindo $k \leq g$ variáveis exógenas e/ou m defasagens, e calcula-se a matriz de covariância dos resíduos $n \times n$; Σ_r ;
3. Calcula-se a razão de verossimilhança da seguinte forma:

$$LR = (T - c) (\log |\Sigma_r| - \log |\Sigma_u|) \rightarrow \chi_r^2,$$

em que

T é o número de observações utilizadas na regressão;

$c = 1 + g + np$ é o número de parâmetros estimados em cada equação do sistema não restrito, incluindo a constante e as variáveis exógenas⁴;

$r = mn^2 + kn$ é o número de restrições no sistema.

Exemplo 6.4

Suponha um sistema de n equações com número máximo de defasagens $p = 6$. Deseja-se testar se $p = 4$. Imaginando um sistema com constante e sem variáveis exógenas, então $c = 1 + pn = 1 + 6n$. O número de restrições é dado pela quantidade de parâmetros que se deixa de estimar no modelo restrito que é de $r = mn^2 = 2n^2$.

³ O Eviews mostra esses autovalores graficamente também.

⁴ Do contrário, $c = pn$.

Convencionalmente, se o valor calculado da estatística LR for menor que o valor tabelado, não se pode rejeitar a hipótese nula.

6.3

Inferência

A inferência sobre os coeficientes estimados em modelos multivariados tem algumas particularidades bastante diferentes dos modelos univariados. A mensagem importante é que se pode proceder a inferências estatísticas sobre os coeficientes individual ou coletivamente, mesmo ante a existência de variáveis endógenas não estacionárias de ordem 1, sempre que o modelo puder ser reescrito de maneira que os coeficientes sob inspeção passem a multiplicar variáveis estacionárias (veja Sims, Stock e Watson, 1990). Essa conclusão não é muito convencional e, na verdade, é bastante surpreendente. Antes de discutir um pouco mais formalmente esse problema, considere o seguinte exemplo que procura esclarecer o ponto.

Exemplo 6.5

Sejam $y_t \sim I(1)$ e $z_t \sim I(1)$ na seguinte equação extraída da primeira linha de um VAR (2) bivariado:

$$y_t = \phi_{11,1}y_{t-1} + \phi_{12,1}z_{t-1} + \phi_{11,2}y_{t-2} + \phi_{12,2}z_{t-2} + \epsilon_{yt}.$$

O subíndice depois da vírgula designa a defasagem a que pertence a matriz Φ_i . Os números antes da vírgula indicam a coordenada a que pertence aquele coeficiente.

É possível fazer inferências sobre os coeficientes $\phi_{11,2}$ e $\phi_{12,2}$ desse modelo. Para isso, adicione e subtraia $\phi_{11,2}y_{t-1}$ e $\phi_{12,2}z_{t-1}$:

$$\begin{aligned} y_t &= \phi_{11,1}y_{t-1} + \phi_{11,2}y_{t-1} + \phi_{12,1}z_{t-1} + \phi_{12,2}z_{t-1} \\ &\quad + \phi_{11,2}y_{t-2} - \phi_{11,2}y_{t-1} + \phi_{12,2}z_{t-2} - \phi_{12,2}z_{t-1} + \epsilon_{yt} \\ &= (\phi_{11,1} + \phi_{11,2})y_{t-1} + (\phi_{12,1} + \phi_{12,2})z_{t-1} + \\ &\quad - \phi_{11,2}\Delta y_{t-1} - \phi_{12,2}\Delta z_{t-1} + \epsilon_{yt}. \end{aligned}$$

Como Δy_{t-1} e Δz_{t-1} são ambas variáveis estacionárias, pode-se testar conjuntamente por $F \sim$ Snedecor, por exemplo, a hipótese $\phi_{11,2} = \phi_{12,2} = 0$, isto é, se o modelo é, de fato, um VAR (2).

Esse tipo de transformação é usado para indicar que tipo de inferência pode ser feito. No modelo, pouco se pode dizer a respeito dos coeficientes que multiplicam as variáveis não estacionárias y_{t-1} e z_{t-1} depois da transformação.

Pode-se estimar cada equação por mínimos quadrados ordinários, se os erros não forem serialmente correlacionados. Estimções por mínimos quadrados ordinários são consistentes e assintoticamente eficientes. Como os regressores à direita são todos iguais, a estimação pelo método SUR (*seemingly unrelated regression*) não contribuiria adicionalmente à eficiência do método de estimação.

Sob condições convencionais⁵, pode-se mostrar que a matriz de coeficientes estimados $\hat{\Phi} = [\hat{\Phi}_1 : \hat{\Phi}_2 : \dots : \hat{\Phi}_p]$ satisfaz:

$$\sqrt{T} \text{vech}(\hat{\Phi} - \Phi) \xrightarrow{d} N(0, \Sigma_{\Phi}),$$

em que *vech* é um operador que empilha as colunas de uma matriz uma sobre as outras.

O estimador por mínimos quadrados ordinários de variáveis $I(1)$ tem a mesma distribuição assintótica do estimador de variáveis $I(0)$, com uma diferença: a matriz de covariância dos coeficientes de variáveis não estacionárias é singular. Nessas circunstâncias, é possível testar os coeficientes dos modelos individualmente usando o teste *t*. Entretanto, hipóteses conjuntas não podem, em geral, ser estimadas porque podem ter uma distribuição degenerada, conforme já mostrado.

Nota 6.5 *Seja um modelo multivariado com variáveis não estacionárias, então inferências sobre hipóteses conjuntas só poderão ser feitas se o modelo puder ser representado de tal forma que os coeficientes a serem testados passem a multiplicar variáveis estacionárias. Como consequência, algumas hipóteses podem ser testadas convencionalmente, outras não, tanto pelo teste *F* como pelo teste χ^2 .*

Quando o modelo contém variáveis determinísticas, a distribuição dos parâmetros dessas variáveis também não é convencional e, portanto, certas hipóteses não podem ser testadas.

Cabe agora alguma intuição sobre a razão de se poder testar normalmente os coeficientes individuais de um modelo multivariado não estacionário, porém não é possível fazer o mesmo quando se trata de um modelo univariado não estacionário. A questão fundamental é perceber que, em um modelo multivariado, a não-estacionaridade emerge ou se verifica quando um dos autovalores da matriz

⁵ Essas condições, bem como as demais propriedades discutidas nesta seção, podem ser vistas em Lütkepohl (2005, seções 7.2 e 7.6), Enders (2004, p. 285-287), Lütkepohl e Krätzig (2004, seção 3.3) e Sims, Stock e Watson (1990) e nas referências indicadas por esses autores.

de coeficientes está sobre ou fora do círculo unitário. No modelo univariado, os autovalores se confundem com as raízes da equação característica, enquanto no modelo multivariado isso não chega a ocorrer estritamente. O exemplo a seguir procura esclarecer essa afirmação.

Exemplo 6.6

Considere o seguinte modelo VAR (1):

$$X_t = \begin{bmatrix} 0,7 & 0,3 \\ 0,3 & 0,7 \end{bmatrix} X_{t-1} + e_t.$$

Os autovalores da matriz Φ_1 são 1 e 0,4. Trata-se, portanto, de um modelo multivariado não estacionário. Esses autovalores estão associados aos coeficientes da matriz Φ_1 , porém indiretamente. Logo, a estatística t permanece válida para a inferência individual dos coeficientes.

6.4

Verificação

Os testes de diagnóstico de resíduos dos modelos univariados podem ser generalizados para o caso multivariado. Esta seção discute um subconjunto desses testes.

Cabe observar que a generalização é praticamente direta e, muitas vezes, evidente. A dificuldade maior talvez seja enxergar as matrizes de interesse para cada teste em particular. A seguir, é apresentado um resumo de cada um dos testes.

6.4.1 Teste de Ljung-Box

O teste de autocorrelação de Ljung-Box tem a mesma lógica do modelo univariado, porém preocupa-se em ver se as autocorrelações multivariadas são nulas. Assim, a hipótese nula deste teste é:

$$H_0 : E(e_t e'_{t-j}) = 0, \text{ para todo } j = 1, 2, \dots, J > p.$$

versus

$$H_1 : E(e_t e'_{t-j}) \neq 0, \text{ para algum } j.$$

A estatística do teste é parecida com a de Ljung-Box:

$$Q = T \sum_{j=1}^J \text{tr}(\hat{C}_j' \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}) \stackrel{d}{\rightarrow} \chi^2_{n^2(J-p)},$$

em que $\hat{C}_j = \sum_{t=j+1}^T \hat{e}_t \hat{e}'_{t-j} / T$ é a matriz de autocovariância na defasagem j .

Alternativamente, usa-se a estatística de Ljung-Box ajustada, a qual tem potencialmente melhores propriedades para pequenas amostras:

$$Q = T^2 \sum_{j=1}^J \frac{1}{T-j} \text{tr} \left(\hat{C}_j' \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1} \right) \xrightarrow{d} \chi_{n^2(j-p)}^2.$$

Ambas as estatísticas existem no Eviews, e o exemplo a seguir mostra isso.

Exemplo 6.7

Considere o modelo não estacionário sem drift simulado no início deste capítulo. Estimado o modelo, verificam-se os resíduos usando esses testes. O Eviews permite calcular esses valores automaticamente, e o resultado é dado na tabela a seguir:

Defasagens	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	0,196	NA	0,197	NA	NA
3	7,057	0,531	7,121	0,524	8
9	22,362	0,897	22,751	0,886	32
12	32,360	0,903	33,139	0,884	44

O símbolo df representa o número de graus de liberdade para a distribuição χ^2 . É importante notar nesse exemplo que, embora o modelo seja não estacionário por construção, as autocorrelações caem rapidamente, de modo que o teste de Ljung-Box rejeita a nula de existência de autocorrelação serial. Esse é um resultado que não ocorreria em um modelo univariado.

Nota 6.6 Os parâmetros do vetor de cointegração de um modelo VECM (a ser visto mais tarde) não contam para a definição do número de graus de liberdade.

Nota 6.7 Se J é muito pequeno, a aproximação à distribuição será bastante pobre, ao passo que um valor alto para J pode resultar em perda de poder.

Nota 6.8 Notando que a diferença fundamental entre as estatísticas de Ljung-Box pura e a ajustada é dada pelo fator

$$\frac{T}{T-j} = \left(1 - \frac{j}{T}\right)^{-1},$$

convém observar que a estatística tende a ser muito alta para valores de j mais elevados, contribuindo para a rejeição da hipótese nula. Por isso, é preciso ser parcimonioso quanto ao valor que j toma.

6.4.2 Teste de Breusch-Godfrey

O objetivo é testar se existe autocorrelação de resíduos no modelo:

$$\hat{\varepsilon}_t = \Theta_1 \hat{\varepsilon}_{t-1} + \Theta_2 \hat{\varepsilon}_{t-2} + \cdots + \Theta_h \hat{\varepsilon}_{t-h} + u_t$$

e verificar se

$$H_0 : \Theta_1 = \Theta_2 = \cdots = \Theta_h = 0 \times H_1 : \Theta_1 \neq 0 \vee \Theta_2 \neq 0 \vee \cdots \vee \Theta_h \neq 0.$$

Para isso, utiliza-se a regressão auxiliar:

$$\begin{aligned} \hat{\varepsilon}_t = & \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} \\ & + \Theta_1 \hat{\varepsilon}_{t-1} + \Theta_2 \hat{\varepsilon}_{t-2} + \cdots + \Theta_h \hat{\varepsilon}_{t-h} + u_t. \end{aligned}$$

Para entender o processo de obtenção da regressão auxiliar, veja Johnston e Dinardo (1997, p. 182-187). Basicamente, a idéia é a seguinte. Trata-se de um teste de multiplicador de Lagrange, LM. Para executá-lo, é sempre necessário estimar $\hat{\varepsilon}_t$ contra o gradiente da função de verossimilhança. No presente caso, esse gradiente é formado pela matriz $[X_{t-1} \ X_{t-2} \ \cdots \ X_{t-p} \ \hat{\varepsilon}_{t-1} \ \hat{\varepsilon}_{t-2} \ \cdots \ \hat{\varepsilon}_{t-h}]$. Johnston e Dinardo comentam que uma importante característica desse teste é funcionar simultaneamente contra a hipótese alternativa de um processo $MA(q)$ para os resíduos.

O teste é executado em dois estágios. No primeiro, é estimado por mínimos quadrados ordinários, de forma que os $\hat{\varepsilon}_t$ s em que $t < 0$ são substituídos por zero. Desse modelo, calcula-se:

$$\hat{\Sigma}_u = \frac{\sum_{t=1}^T \hat{u}_t \hat{u}_t'}{T}.$$

Em seguida, estima-se o modelo impondo a hipótese nula para obter os resíduos restritos, \hat{u}_t^r , e calcula-se:

$$\hat{\Sigma}_r = \frac{\sum_{t=1}^T \hat{u}_t^r \hat{u}_t^{r'}}{T}.$$

O teste LM é definido como:

$$LM_h = T \left[n - \text{tr} \left(\hat{\Sigma}_u \hat{\Sigma}_r^{-1} \right) \right] \xrightarrow{d} \chi_{hm}^2.$$

O exemplo a seguir esclarece esse teste.

Exemplo 6.8

Considere o modelo não estacionário sem drift simulado no início deste capítulo. Estimado o modelo, verificam-se os resíduos usando o teste LM. O Eviews permite calcular esses valores automaticamente, e o resultado é dado na tabela a seguir:

Defasagens	Q-Stat	Prob.	df
1	1,678	0,795	4
3	6,523	0,887	12
9	4,208	1,000	36
12	4,282	1,000	48

O Eviews reporta o resultado das probabilidades com n^2 graus de liberdade. Por isso, é preciso olhar a tabela correta. A tabela anterior contém as probabilidades corretas, mas se fossem usados n^2 a hipótese nula não seria rejeitada também. No caso do exemplo, como não se rejeita a nula com menos graus de liberdade do que seria correto, com mais razão não se rejeitaria se o número de graus de liberdade fosse ainda maior.

6.4.3 Teste de Normalidade

Para entender o teste, é preciso observar que a assimetria e a curtose, respectivamente os momentos 3 e 4, têm uma distribuição normal. Ou seja, defina os momentos 3 e 4 conforme segue:

$$m_3 = (m_{31}, m_{32}, \dots, m_{3n})', \text{ com } m_{3i} = \frac{\sum_{t=1}^T \epsilon_{it}^3}{T};$$

$$m_4 = (m_{41}, m_{42}, \dots, m_{4n})', \text{ com } m_{4i} = \frac{\sum_{t=1}^T \epsilon_{it}^4}{T}.$$

Pode-se demonstrar, então, que essas medidas têm a seguinte distribuição:

$$\sqrt{T} \begin{pmatrix} m_3 \\ m_4 - 3_n \end{pmatrix} \sim N \left[0; \begin{pmatrix} 6I_n & 0 \\ 0 & 24I_n \end{pmatrix} \right],$$

em que $3_n = (3, 3, \dots, 3)'$ é um vetor $n \times 1$ de 3s.

Trata-se, mais uma vez, da versão multivariada do teste de Jarque-Bera.

Na prática, primeiro deve-se obter a matriz de covariância dos resíduos estimados:

$$\hat{\Sigma}_e = \frac{\sum_{t=1}^T (\hat{e}_t - \bar{\hat{e}}) (\hat{e}_t - \bar{\hat{e}})'}{T}.$$

Em seguida, deve-se calcular sua matriz raiz quadrada: $\hat{\Sigma}_e^{\frac{1}{2}}$. No caso da matriz de covariância, isso é imediato por ser uma forma quadrática. Tecnicamente, trata-se de usar procedimentos matemáticos de ortogonalização de matrizes para obter a matriz raiz quadrada, a partir dos autovetores da matriz de covariância. Deve-se observar que o termo \bar{e}_t é desnecessário na presença de interceptos no modelo VAR. Essa matriz deve ser utilizada para padronizar os resíduos, conforme se segue:

$$\hat{e}_t^s = \hat{\Sigma}_e^{-\frac{1}{2}} (\hat{e}_t - \bar{e}_t).$$

Em seguida, devem-se calcular a assimetria, m_3 , e a curtose, m_4 , dos n resíduos:

$$\begin{aligned} \hat{m}_3 &= (\hat{m}_{31}, \hat{m}_{32}, \dots, \hat{m}_{3n})', \text{ com } \hat{m}_{3i} = \frac{\sum_{t=1}^T (\hat{e}_{it}^s)^3}{T}; \\ \hat{m}_4 &= (\hat{m}_{41}, \hat{m}_{42}, \dots, \hat{m}_{4n})', \text{ com } \hat{m}_{4i} = \frac{\sum_{t=1}^T (\hat{e}_{it}^s)^4}{T}. \end{aligned}$$

As estatísticas são dadas por:

$$s_3^2 = T \frac{\hat{m}_3' \hat{m}_3}{6}; \quad s_4^2 = T \frac{(\hat{m}_4 - 3n)' (\hat{m}_4 - 3n)}{24}.$$

Ambas as estatísticas têm distribuição χ_n^2 sob a nula: $s_3^2 = s_4^2 = 0$. Portanto, podem-se testar individualmente as estatísticas usando a distribuição χ_n^2 .

Alternativamente, pode-se usar a distribuição conjunta de ambos os testes, $JB_{2n} = s_3^2 + s_4^2$:

$$JB_{2n} = s_3^2 + s_4^2 \xrightarrow{d} \chi_{2n}^2.$$

Esse procedimento foi sugerido por Doornik e Hansen (1994). E, embora Lütkepohl (2005) proponha a normalização dos resíduos usando a fatorização de Choleski, os primeiros autores criticam o segundo com o argumento de que tal fatorização depende da ordem das variáveis, o que, por sua vez, pode alterar a estatística do teste.

Exemplo 6.9

Considere o modelo não estacionário sem drift simulado no início deste capítulo. Estimado o modelo, verificam-se os resíduos usando o teste LM. O Eviews permite calcular esses valores automaticamente, e o resultado é dado a seguir:

$$JB_4 = 3,620 \sim \chi_4^2$$

$$\text{Prob}(3,620) = 0,460.$$

Logo, não se pode rejeitar a nula de normalidade dos resíduos.

6.5

Previsão

A previsão é análoga aos processos univariados, razão que ajudará a evitar detalhes desnecessários sobre o procedimento. Há um aspecto importante, contudo. A complexidade dos modelos multivariados é muito maior, por isso certos cuidados devem ser tomados, e algumas matrizes utilizadas podem-se tornar difíceis de visualizar. Uma forma de superar o obstáculo é abrindo as matrizes do modelo, o que não poderá ser feito aqui por economia de espaço, mas será deixado a cargo do leitor.

Quando se conhece o processo gerador de dados, a previsão h passos à frente é dada por:

$$E(X_{t+h}|I_t) = X_{t+h|t} = \Phi_1 X_{t+h-1|t} + \Phi_2 X_{t+h-2|t} + \cdots + \Phi_p X_{t+h-p|t},$$

em que $X_{t+j|t} = X_{t+j}$ para $j \leq 0$.

Transformando X_t em um modelo de médias móveis infinito, pelo fato de X_t ser estacionário, obtém-se:

$$X_{t+h} = \left(I - \sum_{j=1}^p \Phi_j L^j \right)^{-1} e_{t+h} = e_{t+h} + \Psi_1 e_{t+h-1} + \Psi_2 e_{t+h-2} + \cdots$$

Conseqüentemente, a previsão correspondente é dada por:

$$X_{t+h|t} = \sum_{j=h}^{\infty} \Psi_j e_{t+h-j}.$$

Desse modo, o erro de previsão será obtido extraíndo-se de X_{t+h} o termo $X_{t+h|t}$:

$$X_{t+h} - X_{t+h|t} = \sum_{j=0}^{h-1} \Psi_j e_{t+h-j},$$

em que $\Psi_0 = I_n$.

A expectativa de previsão dos erros é zero. E o erro quadrático médio de previsão é dado por:

$$\Sigma_X(h) = E(X_{t+h} - X_{t+h|t})(X_{t+h} - X_{t+h|t})' = \sum_{j=0}^{h-1} \Psi_j \Sigma \Psi_j'.$$

O resultado ocorre porque a matriz de erros não é autocorrelacionada.

Se o processo é estacionário, a incerteza da previsão é limitada, mesmo para previsões de longo prazo. Filosoficamente, isso é interessante porque sabe-se,

assim, que o erro de previsão mais distante tem um limite bem definido. Por outro lado, processos integrados têm erro de previsão indeterminado quando o horizonte se eleva, mas isso não exclui a possibilidade de que a previsão de alguns componentes ou combinações lineares de variáveis integradas tenha o erro de previsão limitado também.

Na presença de variáveis exógenas, entre as quais algumas determinísticas, pode-se estender a fórmula anterior facilmente. No caso das determinísticas, por definição conhecem-se suas trajetórias futuras. No caso de variáveis exógenas estocásticas, é preciso definir um modelo para elas, juntamente com o modelo multivariado. Se tais variáveis estiverem sob o controle pleno do tomador de decisões, então pode ser desejável prever a trajetória das endógenas condicionalmente a uma trajetória futura específica das exógenas.

Exemplo 6.10

Considere o seguinte modelo VAR (1), estimado a partir do modelo simulado sem drift:

$$X_t = \begin{bmatrix} 0,694 & 0,332 \\ 0,298 & 0,676 \end{bmatrix} X_{t-1}.$$

Usando esse modelo para prever 20 passos à frente, tem-se a configuração apresentada na Figura 6.3.

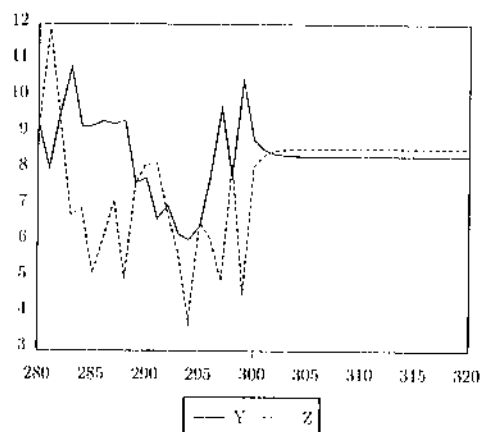


Figura 6.3 Previsão de y e z 20 passos à frente.

É importante notar que as variáveis são não estacionárias e convergem para a previsão multivariada do último valor. Em contraposição, considere o seguinte

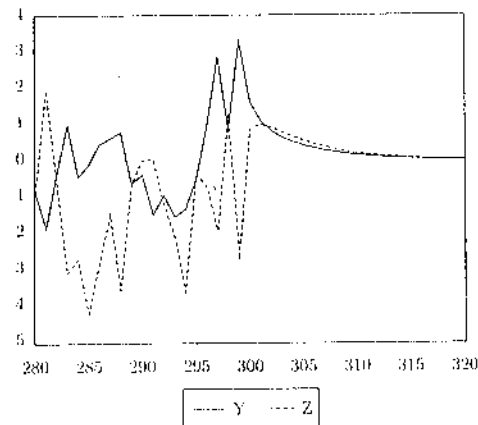


Figura 6.4 Previsão de y e z 20 passos à frente.

modelo VAR (1), estimado a partir do modelo simulado estacionário:

$$X_t = \begin{bmatrix} 0,578 & 0,186 \\ 0,284 & 0,611 \end{bmatrix} X_{t-1}.$$

Usando o modelo para prever 20 passos à frente, a Figura 6.4 mostra que as variáveis convergem para sua média de longo prazo, no redor de zero.

6.6

Função Resposta ao Impulso

De modo geral, o modelo VAR não permite identificar todos os parâmetros da forma estrutural, a não ser que se imponham restrições adicionais. Para ver isso, observe que no sistema restrito dado pela equação (12) consegue-se estimar seis parâmetros na equação da média, mais $\text{Var}(e_1)$, $\text{Var}(e_2)$ e $\text{Cov}(e_1, e_2)$, ou seja, há nove parâmetros estimados. No sistema primitivo, contudo, há dez parâmetros a calcular: além dos oito coeficientes estruturais, há ainda as variâncias de cada um dos choques.

Assim, Sims (1980) sugere um sistema recursivo para identificar o modelo. Trata-se, em outras palavras, de impor que alguns coeficientes sejam iguais a zero, de uma certa forma. Geralmente, usam-se argumentos econômicos para definir quais deles são iguais a zero. A sugestão de Sims impõe que o efeito *feedback* seja limitado. No caso mais simples, de um modelo bivariado, seja imposto, por

exemplo, que $a_{12} = 0$. Como resultado, o modelo fica:

$$\begin{aligned}y_t &= a_{10} + b_{11}y_{t-1} + b_{12}z_{t-1} + \epsilon_{yt}; \\z_t &= a_{20} - a_{21}y_t + b_{21}y_{t-1} + b_{22}z_{t-1} + \epsilon_{zt}.\end{aligned}$$

Essa restrição é importante porque torna os parâmetros estruturais restantes identificáveis, conforme se observa no exemplo bivariado:

$$\begin{aligned}A^{-1} &= \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} y_t \\ z_t \end{bmatrix} &= \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix} \begin{bmatrix} a_{10} \\ a_{20} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} \\ &\quad + \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix} \begin{bmatrix} \sigma_y & 0 \\ 0 & \sigma_z \end{bmatrix} \begin{bmatrix} \epsilon_{yt} \\ \epsilon_{zt} \end{bmatrix}.\end{aligned}$$

Sabendo-se ainda que $a_{12} = 0$, então os erros reduzidos ficam:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} \sigma_y \epsilon_{yt} \\ \sigma_z \epsilon_{zt} - a_{21} \sigma_y \epsilon_{yt} \end{bmatrix},$$

de modo que

$$\begin{aligned}\text{Var}(e_1) &= \sigma_y^2; \\ \text{Var}(e_2) &= \sigma_z^2 + a_{21}^2 \sigma_y^2; \\ \text{Cov}(e_1, e_2) &= -a_{21} \sigma_y^2.\end{aligned}$$

Essas três equações combinam-se com as demais estimativas para identificar o modelo. Ou seja, usando essas três equações mais as equações a seguir, identificam-se os parâmetros estruturais do modelo.

$$\begin{aligned}\phi_{10} &= a_{10}; & \phi_{20} &= a_{20} - a_{10}a_{21}; \\ \phi_{11} &= b_{11}; & \phi_{12} &= b_{12}; \\ \phi_{21} &= -a_{21}b_{11} + b_{21}; & \phi_{22} &= -a_{21}b_{12} + b_{22}.\end{aligned}$$

A metodologia proposta por Sims pode ser generalizada para um vetor com n variáveis endógenas. Trata-se de uma maneira triangular de decompor os resíduos, chamada *decomposição de Choleski*.

No caso de n variáveis endógenas, a matriz de covariância é de dimensão $n \times n$. As condições de identificação requerem a imposição de $(n^2 - n)/2$ restrições.

Como a decomposição de Choleski é triangular, forçando que a porção superior da diagonal tenha zeros, isso equivale a impor as restrições requeridas. O problema dessa imposição é definir a ordenação das variáveis, que é arbitrária, ainda que atribuída a razões econômicas. A ordenação das variáveis define a forma das restrições, de modo que diferentes ordenações geram diferentes restrições.

Se os autovalores da polinomial $(I - \sum_{i=1}^p \Phi_i L^i)$ estiverem fora do círculo unitário, pode-se escrever um VAR (p) como um vetor de médias móveis infinito VMA (∞). Assim, considerando o exemplo de VAR (1) bivariado, tem-se:

$$X_t = \bar{X} + \sum_{i=0}^{\infty} \Phi_1^i e_{t-i} = \bar{X} + \sum_{i=0}^{\infty} \frac{\Phi_1^i}{1 - a_{12}a_{21}} \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix} \begin{bmatrix} \sigma_y e_{y,t-i} \\ \sigma_z e_{z,t-i} \end{bmatrix},$$

em que $\bar{X} \equiv (I - \Phi_1)^{-1} \Phi_0$ é a média de longo prazo.

Defina a matriz:

$$\Psi_i = \frac{\Phi_1^i}{1 - a_{12}a_{21}} \begin{bmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{bmatrix}.$$

Desse modo:

$$\begin{aligned} X_t &= \bar{X} + \sum_{i=0}^{\infty} \Psi_i e_{t-i} \\ &= \bar{X} + \sum_{i=0}^{\infty} \begin{bmatrix} \psi_{i,11} & \psi_{i,12} \\ \psi_{i,21} & \psi_{i,22} \end{bmatrix} \begin{bmatrix} \sigma_y e_{y,t-i} \\ \sigma_z e_{z,t-i} \end{bmatrix}. \end{aligned}$$

Os elementos da matriz Ψ_i são os multiplicadores de impacto de um choque sobre as variáveis endógenas. Assim, o impacto total de um choque de e_{yt} sobre y_{t+h} é dado pela soma dos coeficientes $\psi_{i,11}$, $i = 0, 1, 2, \dots, h$. E sobre z_{t+h} , devem-se somar os coeficientes $\psi_{i,21}$. Os coeficientes, quando desenhados em um gráfico contra i , geram a função resposta ao impulso. A soma dos coeficientes, quando desenhada em um gráfico contra i , gera a função resposta ao impulso acumulada.

O exemplo estacionário anterior com correlação positiva será usado para gerar os gráficos da função resposta ao impulso. Inicialmente, impõe-se a simulação que $a_{12} = 0$ e $a_{21} = 0,8$. Em seguida, estima-se o modelo simulado e gera-se a função resposta ao impulso usando a decomposição de Choleski que assume $a_{12} = 0$. Isso está caracterizado na Figura 6.5.

Deve-se observar que o choque em z_t não tem efeito contemporâneo em y_t , por construção da hipótese de identificação, justamente refletindo que $a_{12} = 0$, conforme se vê a partir da especificação estrutural. É também interessante observar que o efeito se dissipa ao longo do tempo muito rapidamente.

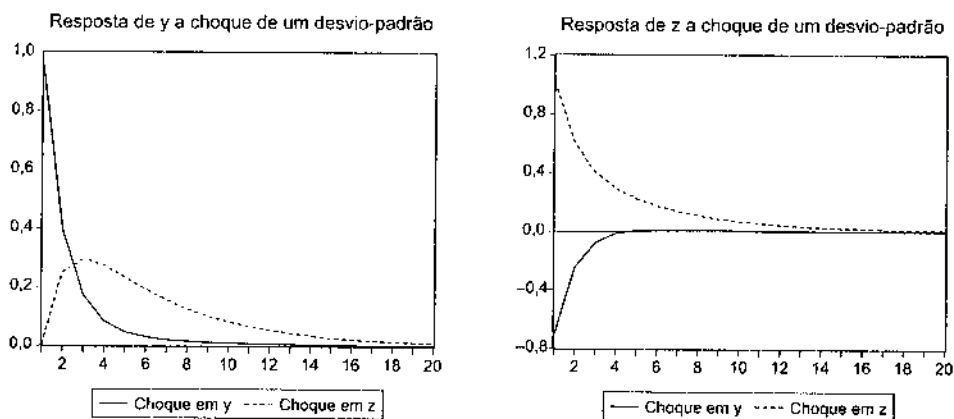


Figura 6.5 Função resposta ao impulso -- decomposição de Choleski.

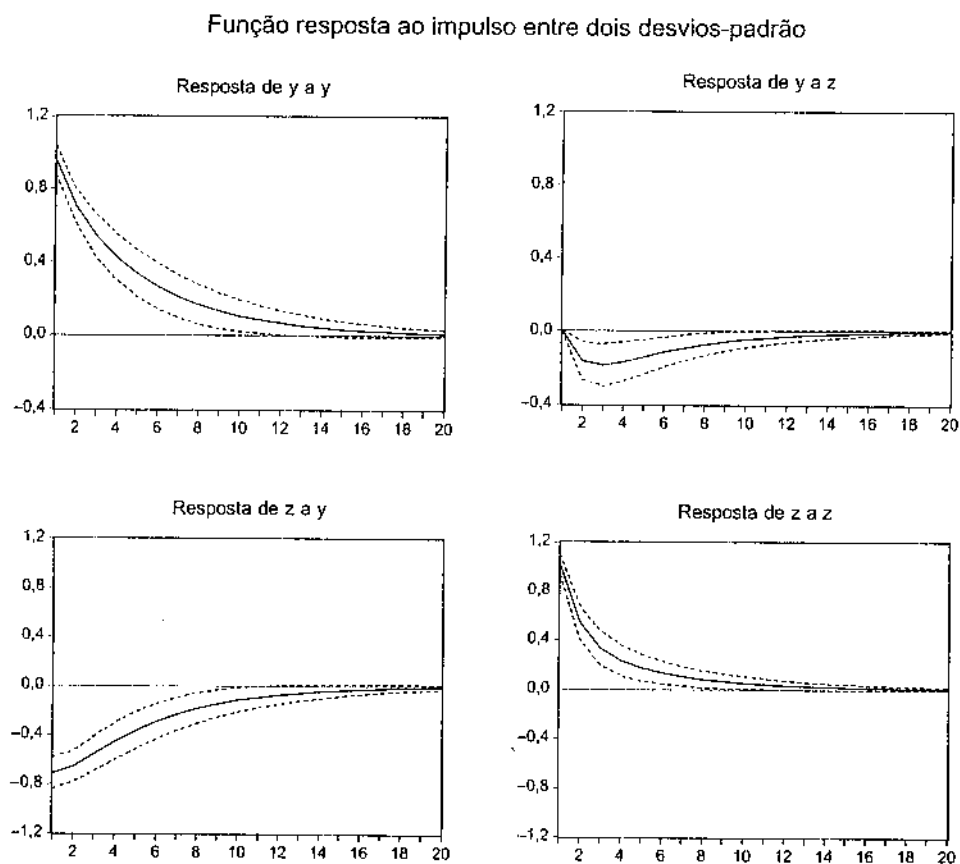


Figura 6.6 Função resposta ao impulso e intervalo de confiança.

6.6.1 Intervalo de Confiança

A função resposta ao impulso é calculada mediante coeficientes estimados. Logo, é claro que há um intervalo de confiança a ser considerado nessas estimativas. Esse intervalo pode ser calculado de forma analítica ou por métodos de experimentos de Monte Carlo.

O método analítico torna-se bem complicado quando se imagina um problema multivariado, em razão das covariâncias cruzadas. Não se procede à derivação analítica do intervalo de confiança neste espaço, mas o leitor interessado poderá consultar Lütkepohl (2005) e Hamilton (1994). Entretanto, muitos softwares já calculam o intervalo de confiança automaticamente. Na Figura 6.6, mostram-se as séries simuladas com correlação negativa e os respectivos intervalos de confiança, calculados analiticamente, com base nos coeficientes estimados e suas variâncias.

Função resposta ao impulso entre dois desvios-padrão

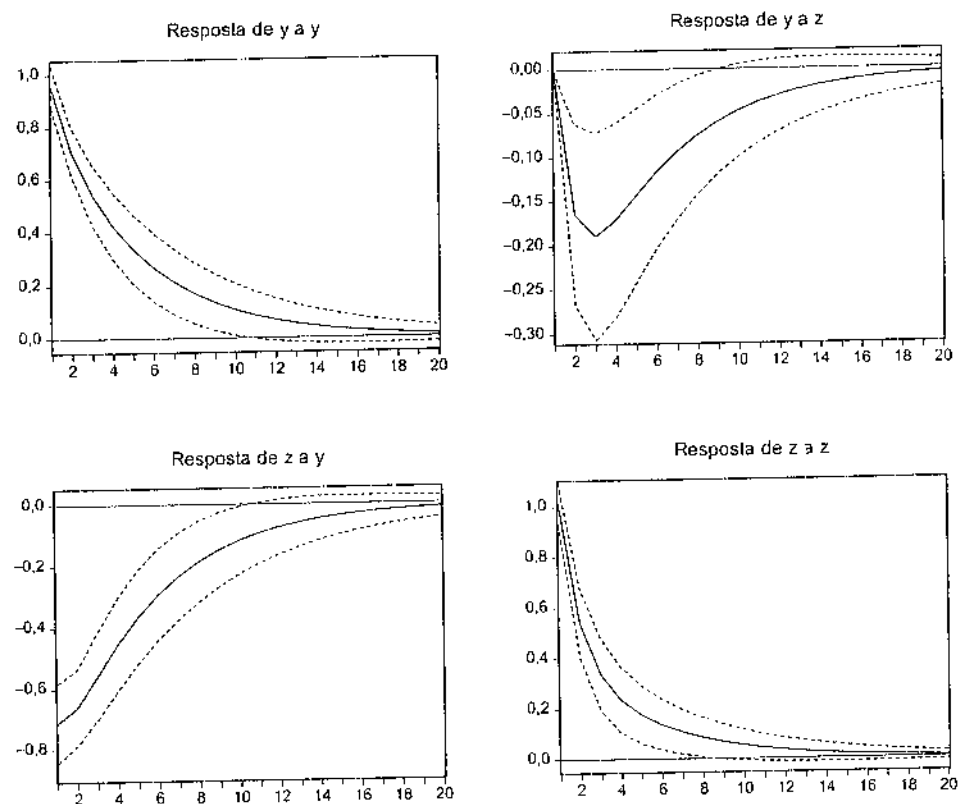


Figura 6.7 Desvio-padrão Monte Carlo – decomposição de Choleski. Séries simuladas assumindo-se $a_{12} = 0$ e $\phi_{12} = \phi_{21} = -0,2$, 1.000 repetições.

Outra maneira de calcular o intervalo de confiança é usando experimentos de Monte Carlo. Para isso, aja da seguinte maneira:

1. Estime o modelo multivariado e armazene os resíduos estimados, $\{\hat{\epsilon}_t\}$;
2. Sorteie os resíduos armazenados com reposição e simule uma nova série usando as matrizes Φ estimadas no passo anterior. Por exemplo, no caso de um VAR (1):

$$\hat{X}_t = \hat{\Phi}_0 + \hat{\Phi}_1 \hat{X}_{t-1} + \hat{\epsilon}_t;$$

3. Reestime o modelo e a nova função resposta ao impulso;
4. Repita o processo milhares de vezes;
5. Para construir um intervalo com 95% de confiança, exclua 2,5% das menores e maiores respostas.

O procedimento foi adotado com relação ao exemplo anterior. A Figura 6.7 apresenta esse método. Para esse caso, os intervalos analítico e por Monte Carlo mostraram-se muito semelhantes, razão pela qual amplificam-se os resultados a seguir.

6.7

Decomposição da Variância

Há ainda outra forma de analisar os resultados do modelo, por meio da decomposição da variância. Trata-se de uma forma de dizer que porcentagem da variância do erro de previsão decorre de cada variável endógena ao longo do horizonte de previsão.

Para entender a idéia, considere o exemplo das seções anteriores, VAR (1) com duas variáveis endógenas, y e z :

$$X_{t+h} = \bar{X} + \sum_{i=0}^{\infty} \Psi_i \epsilon_{t+h-i}.$$

Calcule o erro de previsão:

$$X_{t+h} - E_t(X_{t+h}) = \sum_{i=0}^{h-1} \Psi_i \epsilon_{t+h-i}.$$

Esmiuçando apenas y_{t+h} :

$$\begin{aligned} y_{t+h} - E_t(y_{t+h}) = & \psi_{0,11}\epsilon_{yt+h} + \psi_{1,11}\epsilon_{yt+h-1} + \dots + \psi_{h-1,11}\epsilon_{yt+1} \\ & + \psi_{0,12}\epsilon_{zt+h} + \psi_{1,12}\epsilon_{zt+h-1} + \dots + \psi_{h-1,12}\epsilon_{zt+1}. \end{aligned}$$

Logo:

$$\sigma_y^2(h) = \sigma_y^2(\psi_{0,11}^2 + \psi_{1,11}^2 + \dots + \psi_{h-1,11}^2) + \sigma_z^2(\psi_{0,12}^2 + \psi_{1,12}^2 + \dots + \psi_{h-1,12}^2).$$

Agora, pode-se decompor a variância do erro de previsão em seus diversos elementos. No caso do modelo bivariado, a variância é decomposta em dois pedaços para cada período de tempo, dividindo-se ambos os lados por $\sigma_y^2(h)$:

$$1 = \frac{\sigma_y^2(\psi_{0,11}^2 + \psi_{1,11}^2 + \dots + \psi_{h-1,11}^2)}{\sigma_y^2(h)} + \frac{\sigma_z^2(\psi_{0,12}^2 + \psi_{1,12}^2 + \dots + \psi_{h-1,12}^2)}{\sigma_y^2(h)}.$$

A tabela a seguir decompõe a variância de y_{t+h} , quando $h = 1, 2, \dots, 5$.

Decomposição da variância de $y_{t+h} - E_t(y_{t+h})$			
h	$\sigma_y(h)$	$\frac{\sigma_y^2(\psi_{0,11}^2 + \psi_{1,11}^2 + \dots + \psi_{h-1,11}^2)}{\sigma_y^2(h)}$	$\frac{\sigma_z^2(\psi_{0,12}^2 + \psi_{1,12}^2 + \dots + \psi_{h-1,12}^2)}{\sigma_y^2(h)}$
1	0,956	100,00	0,00
2	1,063	94,53	5,47
3	1,116	88,09	11,91
4	1,152	83,24	16,76
5	1,176	80,03	19,97

A tabela anterior mostra que o erro de previsão aumenta com o horizonte de previsão, porém a importância do erro atribuída a cada variável se altera. No quinto passo, 20% do erro se deve à variável z .

O mesmo é feito com a variável z_{t+h} :

Decomposição da variância de $z_{t+h} - E_t(z_{t+h})$			
h	$\sigma_z(h)$	$\frac{\sigma_y^2(\psi_{0,21}^2 + \psi_{1,21}^2 + \dots + \psi_{h-1,21}^2)}{\sigma_z^2(h)}$	$\frac{\sigma_z^2(\psi_{0,22}^2 + \psi_{1,22}^2 + \dots + \psi_{h-1,22}^2)}{\sigma_z^2(h)}$
1	1,244	32,96	67,04
2	1,404	25,51	74,49
3	1,462	29,04	70,96
4	1,490	27,07	72,93
5	1,506	26,06	73,94

Exemplo 6.11

O Banco Central do Brasil (BC) utiliza em suas projeções para inflação futura quatro modelos VAR (veja o Relatório de Inflação 2T04). Aqui, vamos-nos

restringir a um deles, o chamado modelo VAR 1, para verificar o impacto que variáveis como a variação da taxa de câmbio nominal e a taxa de juros Selic real têm sobre os preços livres.

O modelo VAR1 inclui as seguintes variáveis endógenas: variação da taxa de câmbio nominal, variação da taxa Selic real, inflação dos preços livres e inflação dos preços administrados. Além disso, são incluídas 11 dummies sazonais referentes a cada mês do ano e uma dummy de tendência para o período de desinflação (1995 : 1 a 1998 : 6). Para o exemplo apresentado aqui, foram utilizados dados que abrangem o período de setembro de 1994 a maio de 2007. Os dados de inflação utilizados não necessitam de nenhuma transformação. Para os dados da taxa de câmbio nominal, basta tirar a primeira diferença da série para se determinar a variação da taxa de câmbio nominal. No caso da variação da taxa Selic real, é necessária a utilização de alguns cálculos que podem ser verificados na planilha correspondente no exemplo.

A estimação realizada segue o modelo do Banco Central utilizando duas defasagens. Outro ponto a ser lembrado é que não são recuperados os valores dos parâmetros estruturais dos modelos. A justificativa é que o modelo estimado tem como função prever a inflação futura e determinar a relação entre as variáveis. Dado o enfoque sobre os preços, analisam-se a função resposta ao impulso e a decomposição da variância sobre os preços livres. A Figura 6.8 apresenta o resultado da função resposta ao impulso dos preços livres para um choque de um desvio-padrão na variação da taxa de juros real e na variação da taxa de câmbio nominal. Pode-se perceber que o câmbio tem uma maior influência sobre os preços livres do que a taxa de juros. Esse fato é corroborado também pela decomposição da variância mostrada na tabela a seguir.

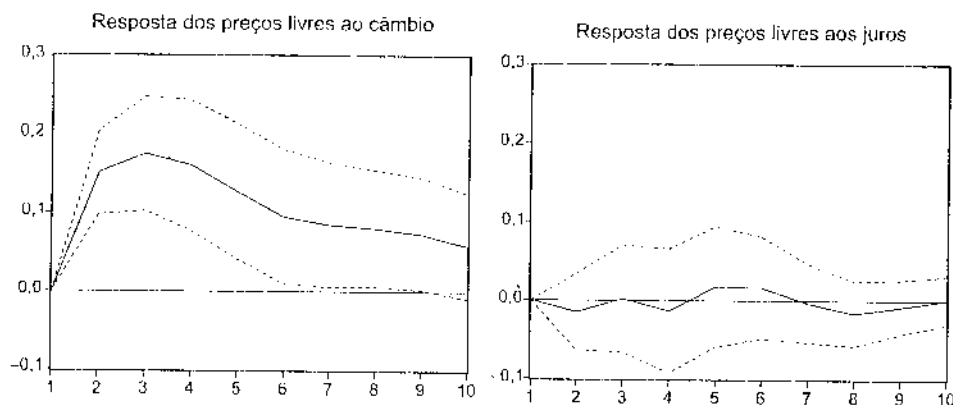


Figura 6.8 Resposta ao impulso sobre os preços.

Decomposição da variância dos preços livres				
Período	Livres	Administrados	Câmbio	Juros
1	100,00	0,00	0,00	0,00
2	82,93	1,69	15,24	0,14
3	72,00	1,65	26,24	0,11
4	64,43	1,82	33,57	0,17
5	60,20	2,62	36,90	0,27
6	58,43	2,58	38,62	0,37
7	57,56	2,53	39,51	0,35
8	56,62	2,53	40,41	0,43
9	55,65	2,78	41,13	0,44
10	55,02	2,90	41,64	0,44

Note que nos primeiros meses a variância da inflação dos preços livres é quase totalmente explicada pela própria inflação dos preços livres. Ao longo do tempo, ganham importância a desvalorização cambial (chamada pass-through) e o efeito dos preços administrados. Observe ainda que o efeito dos juros é pequeno.

6.8

Teste de Granger – Causalidade

Continuando com o sistema bivariado que tem sido usado, pergunta-se se uma variável é capaz de prever outra e em que condições. Em outras palavras, a questão fundamental é saber se o escalar y ajuda a prever o escalar z . Se isso não acontece, então diz-se que y não-Granger-causa z .

A forma de responder a essa pergunta é usando um teste F convencional, válido quando os coeficientes de interesse puderem ser escritos de modo a multiplicar variáveis estacionárias. Assim, o teste é feito da seguinte maneira:

1. Estime $z_t = \phi_{20} + \sum_{i=1}^p \phi_{i,21} y_{t-i} + \sum_{i=1}^p \phi_{i,22} z_{t-i} + e_{2t}$;
2. Teste se y não-Granger-causa z usando o teste de F , sob

$$H_0 : \phi_{1,21} = \phi_{2,21} = \dots = \phi_{p,21} = 0 \times H_1 : \phi_{i,21} \neq 0, i = 1, 2, \dots, p,$$

em que a estatística do teste é dada por:

$$S_1 = \frac{(e_v^2 - e_u^2) / p}{e_u^2 / (T - 2p - 1)} \xrightarrow{d} F(p, T - 2p - 1),$$

em que r representa restrito e u , não restrito. Se $S_1 > F^{5\%}$, rejeita-se a hipótese nula de que y não-Granger-causa z .

3. Um teste equivalente é:

$$S_2 = \frac{T(e_r^2 - e_u^2)}{e_u^2} \xrightarrow{d} \chi_p^2.$$

Rejeita-se a nula se $S_2 > \chi_p^2(5\%)$.

Nota 6.9 Enders (2004) refere-se a esse teste assumindo que todas as variáveis são estacionárias, ou que os coeficientes de interesse multiplicam variáveis estacionárias.

Nota 6.10 Teste de causalidade de Granger não é a mesma coisa que teste de exogeneidade. Para que z_t seja exógeno a y_t , é preciso que z_t não seja afetado contemporaneamente por y_t . A forma reduzida do VAR não permite que se faça esse tipo de teste. O teste de causalidade de Granger inclui, pois, valores correntes e passados de y_t sobre z_t .

Pode-se fazer o mesmo teste em contextos de mais variáveis, cujo nome é teste de bloco-exogeneidade, não obstante Enders (2004) sugira o nome de bloco-causalidade, talvez mais apropriado para o contexto do teste. Estima-se o modelo com e sem restrição e utiliza-se o teste F , como visto anteriormente.

É importante observar que, em sistemas com $n > 2$, o teste de causalidade é mais complicado de ser feito e a interpretação necessita de maiores cuidados. O problema que pode ocorrer ao não rejeitar a nula é não perceber a dinâmica mais complicada do modelo em que uma variável, apesar de não causar diretamente outra, por exemplo, y_{2t} não-Granger-causa y_{1t} , pode causá-la indiretamente. Isso ocorrerá quando y_{2t} causar y_{3t} , que, por sua vez, causa y_{1t} . O teste de Granger-causalidade não foi desenvolvido para esse tipo de caso.

Para entender melhor o argumento anterior, considere, como exemplo, um sistema trivariado (Lütkepohl e Krätzig, 2004):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \sum_{i=1}^p \begin{bmatrix} \phi_{i,11} & \phi_{i,12} & \phi_{i,13} \\ \phi_{i,21} & \phi_{i,22} & \phi_{i,23} \\ \phi_{i,31} & \phi_{i,32} & \phi_{i,33} \end{bmatrix} X_{t-i} + e_t.$$

O teste de bloco-exogeneidade (causalidade) sugere empreender o seguinte teste para verificar a causalidade de y_{2t} a y_{1t} :

$$H_0 : \phi_{i,12} = 0, \forall i.$$

Entretanto, pode ser que y_{2t} cause y_{1t} indiretamente, por intermédio de y_{3t} . Portanto, para sistemas de equações de maior dimensão, as restrições para testar Granger-causalidade são complicadas restrições não-lineares. Mais detalhes sobre esse assunto podem ser encontrados em Lütkepohl e Krätzig (2004), McCrorie e Chambers (2006) e Dufour e Renault (1998).

Exemplo 6.12

Um exemplo bastante simples de aplicação do teste de causalidade de Granger é encontrado em Taylor (2001). O autor testa a eficácia do comitê norte-americano de política monetária – FOMC – em definir a taxa de juros, denominada Fed Funds, a partir de um teste de Granger envolvendo a meta para os juros e sua taxa efetiva.

No caso brasileiro, podem-se utilizar a meta para a taxa Selic definida pelo Copom e a taxa Selic efetiva observada no mercado financeiro (séries encontradas no site do Banco Central do Brasil). Se o Banco Central, de fato, determina a taxa Selic, então alterações na taxa efetiva praticada no mercado financeiro deveriam responder às alterações na meta definida pelo Banco Central. Por outro lado, não há razões para acreditar que alterações nas taxas de mercado afetem a meta definida pelo Banco Central (as alterações diárias na taxa Selic são causadas principalmente pela necessidade de caixa dos bancos naquele dia).

O teste de causalidade de Granger deve ser realizado utilizando variáveis que sejam estacionárias. Por meio da tabela a seguir, pode-se verificar que as séries de taxa de juros efetiva e de meta são estacionárias somente em primeira diferença. Devido a esse fato, utilizam-se as séries em primeira diferença para realizar o teste de causalidade de Granger.

Teste Ng-Perron				
	MZ _a	MZ _t	MSB	MPT
Meta – Nível	-5,706	-1,440	0,252	5,030
Meta – Primeira Diferença	-943,032	-21,714	0,023	0,025
Efetiva – Nível	-5,153	-1,345	0,261	5,420
Efetiva – Primeira Diferença	-14,426	-2,685	0,186	1,698
1%	-13,800	-2,580	0,174	1,780
5%	8,100	-1,980	0,233	3,170
10%	-5,700	-1,620	0,275	4,450

O resultado do teste de Granger, incluindo as variáveis “meta”, que é o objetivo para a taxa Selic definida pelo Copom, e “efetiva”, que é a taxa Selic

efetiva diária, encontra-se a seguir. Como os dados são diários e apenas para os dias úteis, utilizam-se 20 defasagens de cada variável. Lembrando que as variáveis estão em primeira diferença, então os resultados do teste informam se a variação de uma variável ajuda a explicar a variação da outra defasagem. Como pode ser observado na tabela, rejeita-se a hipótese de que a "meta" não causa (no sentido de Granger) a taxa "efetiva", ao passo que se pode rejeitar a hipótese de que a taxa "efetiva" causa a taxa "meta", o que é exatamente o resultado esperado.

<i>Teste de Causalidade de Granger</i>			
Hipótese Nula	Obs	Estatística F	Probabilidade
META não-Granger-causa EFETIVA	1858	4,503	0,000
EFETIVA não-Granger-causa META		0,395	0,992

6.9

VAR Estrutural

Existem outras formas de definir restrições sobre a matriz A , de modo a identificar os parâmetros estruturais. Nessas outras, seguem-se mais estritamente argumentos econômicos a ponto de impor restrições que chegam a sobreidentificar o modelo, se o número de restrições for superior ao número de coeficientes estimados na forma reduzida. Ou seja, enquanto a metodologia de Sims (1980) usava a economia para definir a ordenação das variáveis, é possível buscar restrições econômicas de forma mais abrangente. Isto é, usa-se a teoria econômica para definir as restrições da matriz A completamente.

Esta seção trata do VAR estrutural, com o qual se podem recuperar os parâmetros estruturais do modelo. Não se constitui, portanto, da simples imposição de uma estrutura ordenada de resíduos, como é o caso da decomposição de Choleski, mas da busca de restrições plausíveis sobre os parâmetros.

Nota 6.11 *Poderiam ser tentadas várias ordenações no VAR para a escolha da melhor, quando se usa a decomposição de Choleski. Contudo, quando os resíduos são correlacionados, ordens alternativas das variáveis não são uma solução prática.*

Para entender o procedimento, considere um VAR (1) com n variáveis:

$$AX_t = B_0 + B_1 X_{t-1} + B\epsilon_t.$$

Quando se estima a forma reduzida, obtêm-se os resíduos \hat{e}_t , tal que:

$$B\hat{e}_t = A\hat{e}_t.$$

O problema é restringir o sistema, de forma a recuperar ϵ_t conforme a hipótese de que cada resíduo do sistema estrutural é independente um do outro. Para começar, usa-se a matriz de covariância dos erros da forma reduzida:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix},$$

em que cada elemento é estimado por⁶:

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T \hat{e}_{it} \hat{e}_{jt}.$$

Como Σ é simétrica, existem $n(n+1)/2$ elementos diferentes. Como os elementos da diagonal de A são unitários, a matriz contém $(n^2 - n)$ elementos desconhecidos. Além disso, desconhecem-se n valores relativos à variância dos elementos do vetor ϵ . Portanto, há n^2 elementos desconhecidos e $n(n+1)/2$ conhecidos. Logo, é preciso impor $n^2 - (n(n+1)/2) = n(n-1)/2$ restrições sobre o modelo estrutural.

Nota 6.12 No caso da decomposição de Choleski, é exatamente isso o que acontece, pois os elementos acima da diagonal são iguais a zero e perfazem um total de $n(n-1)/2$ elementos.

Nota 6.13 A imposição de $n(n-1)/2$ restrições não é suficiente para a identificação exata, pois existem não-linearidades no sistema de equações, as quais podem gerar várias soluções como mostra um exemplo adiante.

Então, de modo geral, é verdade que:

$$BB' = \text{Var}(A\epsilon_t) = A \text{Var}(\epsilon_t) A'.$$

⁶ O acento circunflexo representa resíduo estimado, como vem sendo a praxe notacional deste texto.

Usando essa estrutura, em um caso bivariado, discutem-se as várias possibilidades de restrições que podem ser impostas ao modelo.

Exemplo 6.13

Suponha o seguinte sistema de equações, em que se admite conhecer a matriz de covariância de e_t :

$$\begin{aligned} \begin{bmatrix} \sigma_{\epsilon_1}^2 & 0 \\ 0 & \sigma_{\epsilon_2}^2 \end{bmatrix} &= \begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix} \begin{bmatrix} 0,5 & 0,4 \\ 0,4 & 0,6 \end{bmatrix} \begin{bmatrix} 1 & a_{21} \\ a_{12} & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0,5 + 0,4a_{12} & 0,4 + 0,6a_{12} \\ 0,5a_{21} + 0,4 & 0,4a_{21} + 0,6 \end{bmatrix} \begin{bmatrix} 1 & a_{21} \\ a_{12} & 1 \end{bmatrix} \\ \therefore \begin{bmatrix} 0,5 + 0,8a_{12} + 0,6a_{12}^2 & 0,5a_{21} + 0,4a_{12}a_{21} + 0,4 + 0,6a_{12} \\ 0,5a_{21} + 0,4 + 0,4a_{12}a_{21} + 0,6a_{12} & 0,5a_{21}^2 + 0,8a_{21} + 0,6 \end{bmatrix}. \end{aligned}$$

Há três equações independentes com quatro incógnitas ($a_{12}, a_{21}, \sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2$):

$$\begin{aligned} \sigma_{\epsilon_1}^2 &= 0,5 + 0,8a_{12} + 0,6a_{12}^2; \\ 0 &= 0,5a_{21} + 0,4 + 0,4a_{12}a_{21} + 0,6a_{12}; \\ \sigma_{\epsilon_2}^2 &= 0,5a_{21}^2 + 0,8a_{21} + 0,6. \end{aligned}$$

Impondo a restrição de Choleski, $a_{12} = 0$, calculam-se:

$$\begin{aligned} \sigma_{\epsilon_1}^2 &= 0,5; \\ 0 &= 0,5a_{21} + 0,4 \implies a_{21} = -0,8 \\ \sigma_{\epsilon_2}^2 &= 0,5a_{21}^2 + 0,8a_{21} + 0,6 \implies \sigma_{\epsilon_2}^2 = 0,28. \end{aligned}$$

Portanto, a recuperação dos erros estruturais é imediata, já que os parâmetros de A multiplicando e_t geram ϵ_t . A partir deles, gera-se a função resposta ao impulso. No caso de um VAR (1) bivariado, essa função é:

$$\Psi_i = \Phi_1^i \begin{bmatrix} 1 & 0 \\ -a_{21} & 1 \end{bmatrix}.$$

Exemplo 6.14

É possível impor outras restrições, normalmente decorrentes da teoria econômica. Por exemplo, uma inovação em ϵ_{2t} pode ter um efeito unitário em y_{1t} . Se isso é

verdade, poderia ser imposto que $a_{12} = 1$ e se resolver o sistema dado no exemplo anterior. Nesse caso, é imediato encontrar:

$$\begin{aligned}\sigma_{\epsilon_1}^2 &= 1,9; \\ a_{21} &= -\frac{10}{9}; \\ \sigma_{\epsilon_2}^2 &= 0,5 \frac{100}{81} - 0,8 \frac{10}{9} + 0,6 = 0,328.\end{aligned}$$

Exemplo 6.15

Há casos em que as restrições podem não gerar resultados únicos. Por exemplo, ante a imposição de que a correlação cruzada no modelo bivariado estrutural é única, $a_{12} = a_{21} = a$, não há unicidade de resultados:

$$\begin{aligned}0 &= 1,1a + 0,4 + 0,4a^2 \implies a = \begin{cases} -0,173 \\ -0,928 \end{cases}; \\ \sigma_{\epsilon_1}^2 &= 0,5 + 0,8a + 0,6a^2 \implies \sigma_{\epsilon_1}^2 = \begin{cases} 0,414 \\ 0,460 \end{cases}; \\ \sigma_{\epsilon_2}^2 &= 0,5a^2 + 0,8a + 0,6 \implies \sigma_{\epsilon_2}^2 = \begin{cases} 0,477 \\ 0,288 \end{cases}.\end{aligned}$$

Na Figura 6.9, impõe-se que $a_{12} = 1$ para a função resposta ao impulso da simulação sob $a_{12} = 0, b_{12} = b_{21} = 0,2$. A imposição é falsa, porque se sabe o verdadeiro processo gerador de dados ($a_{12} = 0$), porém serve para

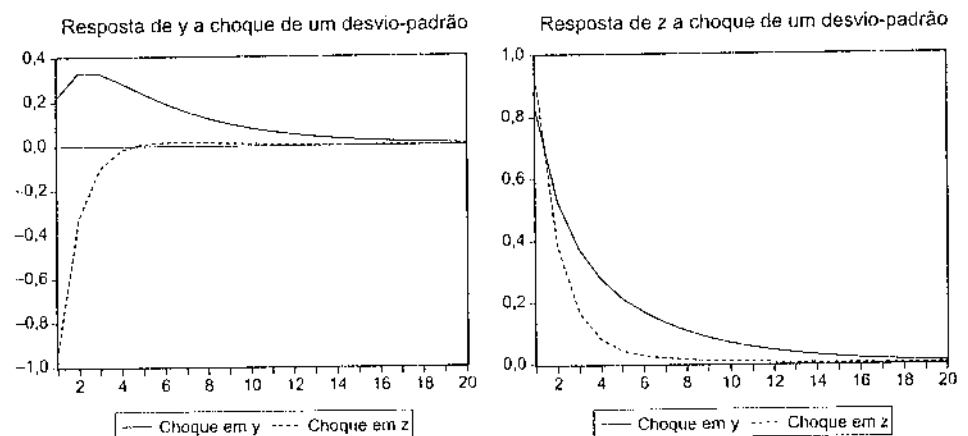


Figura 6.9 Função resposta ao impulso – VAR estrutural imposto, erroneamente, $a_{12} = 1$ a séries simuladas com $a_{12} = 0$ e $b_{12} = b_{21} = 0,2$.

ilustrar a técnica e verificar os efeitos de que a imposição de uma restrição falsa gera resultados discrepantes. Por isso, a Figura 6.9 deve ser comparada com a Figura 6.5.

A comparação entre ambas as figuras mostra que a imposição falsa de restrições gera funções resposta ao impulso bastante diferentes entre si. Nos exemplos usados aqui, é fácil ver que a imposição é falsa. Entretanto, na prática, não se sabe qual o verdadeiro processo gerador de dados, por isso é difícil saber se a restrição imposta é falsa ou não. A recomendação é preferir a teoria econômica a restrições *ad hoc*, como é o caso do uso da decomposição de Choleski. A teoria econômica ajuda bastante a definir as restrições e deve ser usada para testes de robustez, quando existem várias possibilidades diferentes.

Nota 6.14 *A simulação do modelo foi feita de forma que a decomposição de Choleski funcionasse perfeitamente. Com dados reais, é muito mais provável que as imposições econômicas façam mais sentido do que imaginar que o processo gerador resulta de um modelo que pode ser representado segundo a decomposição de Choleski.*

6.10

Decomposição de Blanchard e Quah

Blanchard e Quah (1989) sugerem uma forma de identificação com base em restrições determinadas pela teoria econômica. A idéia é impor restrições a respeito do comportamento de longo prazo de uma variável a partir do choque estrutural em outra variável. A seguir, discute-se com alguns detalhes essa metodologia, seguindo a exposição dos autores mencionados.

O exemplo de Blanchard e Quah é motivador. Eles supõem um vetor $X_t = (\Delta y_t, u_t)'$, em que y_t representa o logaritmo do produto, e u_t , a taxa de desemprego. As variáveis respondem a um vetor de choques dado por $\epsilon_t = (\epsilon_{t,d}, \epsilon_{t,s})$. O produto é afetado por choques de demanda, e o desemprego, por choques de oferta. Os choques entre oferta e demanda não são correlacionados, ainda que possam existir argumentos contrários a essa hipótese. Eles interpretam os choques da seguinte forma: os choques de demanda têm efeitos temporários; os choques de oferta têm efeitos permanentes. Finalmente, eles usam essa interpretação para impor restrições apropriadas e identificar o modelo.

Para formalizar essas restrições no modelo econométrico, convém reescrevê-lo na forma estrutural dada a seguir, ignorando a existência de constante e normalizando $B = I$:

$$AX_t = \sum_{j=1}^p B_j X_{t-j} + \epsilon_t,$$

em que $\text{var}(\epsilon_t) = I$.

O modelo na forma reduzida fica:

$$X_t = \sum_{j=1}^p A^{-1} B_j X_{t-j} + A^{-1} \epsilon_t = \sum_{j=1}^p \Phi_j X_{t-j} + e_t,$$

em que

$$A^{-1} B_j \equiv \Phi_j; \quad A e_t = \epsilon_t.$$

Sendo X_t estacionário, o modelo pode ser escrito segundo a representação de Wold. Haverá duas representações equivalentes de X_t , dependendo de o modelo partir da forma reduzida ou estrutural. Partindo da forma reduzida, pode-se escrevê-lo como:

$$\begin{aligned} \left(I - \sum_{j=1}^p \Phi_j L^j \right) X_t = e_t &\implies X_t = \left(I - \sum_{j=1}^p \Phi_j L^j \right)^{-1} e_t \implies \\ X_t &= \left(\sum_{j=0}^{\infty} C_j L^j \right) e_t, \end{aligned} \quad (13)$$

em que

$$C_0 \equiv I; \quad \left(I - \sum_{j=1}^p \Phi_j L^j \right)^{-1} \equiv \left(\sum_{j=0}^{\infty} C_j L^j \right); \quad \text{var}(e_t) = \Sigma.$$

Partindo da forma estrutural, obtém-se:

$$\begin{aligned} \left(A - \sum_{j=1}^p B_j L^j \right) X_t = \epsilon_t &\implies X_t = \left(A - \sum_{j=1}^p B_j L^j \right)^{-1} \epsilon_t \implies \\ X_t &= \left(\sum_{j=0}^{\infty} A_j L^j \right) \epsilon_t, \end{aligned} \quad (14)$$

em que

$$\left(A - \sum_{j=1}^p B_j L^j \right)^{-1} \equiv \left(\sum_{j=0}^{\infty} A_j L^j \right).$$

Dado que $C_0 = I$, das equações (14) e (13) resultam as seguintes relações:

$$\begin{aligned} A_0 \epsilon_t = e_t &\implies A_0 \epsilon_{t-j} = e_{t-j}; \\ A_j \epsilon_{t-j} = C_j e_{t-j} &\implies A_j \epsilon_{t-j} = C_j A_0 \epsilon_{t-j} \therefore \\ A_j &= C_j A_0, \quad j > 0. \end{aligned}$$

Finalmente, convém traçar a relação entre a matriz A_0 e a matriz A . Para isso, considere a decomposição do modelo VAR estrutural com a decomposição estrutural de Wold.

$$\begin{aligned} A \left(I - \sum_{j=1}^p A^{-1} B_j L^j \right) X_t = \epsilon_t &\implies X_t = \left(I - \sum_{j=1}^p A^{-1} B_j L^j \right)^{-1} A^{-1} \epsilon_t \\ &= \left(\sum_{j=0}^{\infty} C_j L^j \right) A^{-1} \epsilon_t = \sum_{j=0}^{\infty} C_j A^{-1} \epsilon_{t-j}. \end{aligned}$$

Comparando essa equação com a equação (14), conclui-se que:

$$\begin{aligned} A_j &= C_j A^{-1} \implies \\ C_j A_0 &= C_j A^{-1} \therefore \\ A_0 &= A^{-1}. \end{aligned}$$

Essa relação é importante porque serão impostas restrições à matriz A_0 . Ora, não deve tomar muito tempo do leitor entender que as imposições sobre A_0 são equivalentes a impor restrições sobre A , como feito anteriormente.

Com essas relações, pode-se verificar o primeiro conjunto de restrições existentes até o momento e que são análogas ao que já foi visto nas seções anteriores. Do termo $A_0 \epsilon_t$, verifica-se que:

$$\begin{aligned} A_0 \epsilon_t &= e_t \implies \\ A_0 A_0' &= \Sigma. \end{aligned}$$

Esse sistema de equações resulta, como já foi visto, em $\frac{n(n+1)}{2}$ equações. No caso bivariado, são explicitamente dadas por:

$$\begin{aligned} a_{11,0}^2 + a_{12,0}^2 &= \text{var}(e_1); \\ a_{11,0}a_{21,0} + a_{12,0}a_{22,0} &= \text{cov}(e_1, e_2); \\ a_{21,0}^2 + a_{22,0}^2 &= \text{var}(e_2). \end{aligned}$$

A partir desse ponto, é necessário introduzir as restrições de longo prazo, como sugerido por Blanchard e Quah (1989). A imposição dos autores pode ser visualizada partindo-se da equação (14) no modelo bivariado. Tal equação pode ser reescrita como:

$$X_t = \sum_{j=0}^{\infty} \begin{bmatrix} a_{11,j} & a_{12,j} \\ a_{21,j} & a_{22,j} \end{bmatrix} \epsilon_{t-j}.$$

Portanto, para que o choque de demanda não tenha efeitos de longo prazo sobre o produto, a seguinte relação deve ser imposta:

$$\sum_{j=0}^{\infty} a_{11,j} = 0.$$

Ora, tendo-se A_0 e C_j , este obtido da estimação do modelo reduzido, e valendo a relação $A_j = C_j A_0$, pode-se obter A_j . Cabe aqui um pouco mais de detalhes sobre o modelo. O vetor X é estacionário, de modo que nenhum choque tem efeitos de longo prazo sobre as variáveis endógenas, variação de produto e desemprego, no caso. A restrição imposta, $\sum_{j=0}^{\infty} a_{11,j} = 0$, implica que $\epsilon_{t,dt}$ não afeta o nível do produto. Por que é assim? Usando o argumento de Blanchard e Quah, $a_{11,j}$ é o efeito do choque da demanda sobre Δy depois de j períodos; logo, $\sum_{j=0}^{\infty} a_{11,j}$ representa o efeito dos choques sobre o nível do produto. Se esse efeito é nulo, explica-se a necessidade de impor a restrição.

Para operacionalizar o procedimento:

1. Estime o modelo VAR (p) e obtenhas as matrizes $\hat{\Phi}_j$;
2. Obtenha as matrizes C_j resolvendo:

$$\left(\sum_{j=0}^{\infty} \hat{C}_j L^j \right) = \left(I - \sum_{j=1}^p \hat{\Phi}_j L^j \right)^{-1};$$

3. Sabendo que $A_j = C_j A_0$, imponha a restrição de longo prazo;
4. Resolva o sistema de equações e encontre os coeficientes de A_0 .

No caso bivariado, o passo três é fácil de visualizar, ignorando a situação de se tratar de coeficientes estimados, pois:

$$\begin{bmatrix} a_{11,j} & a_{12,j} \\ a_{21,j} & a_{22,j} \end{bmatrix} = \begin{bmatrix} c_{11,j} & c_{12,j} \\ c_{21,j} & c_{22,j} \end{bmatrix} \begin{bmatrix} a_{11,0} & a_{12,0} \\ a_{21,0} & a_{22,0} \end{bmatrix}.$$

Portanto, pode-se escrever:

$$\begin{bmatrix} a_{11,j} & a_{12,j} \\ a_{21,j} & a_{22,j} \end{bmatrix} = \begin{bmatrix} a_{11,0}c_{11,j} + a_{21,0}c_{12,j} & a_{12,0}c_{11,j} + a_{22,0}c_{12,j} \\ a_{11,0}c_{21,j} + a_{21,0}c_{22,j} & a_{12,0}c_{21,j} + a_{22,0}c_{22,j} \end{bmatrix}.$$

Impondo a restrição de longo prazo, encontra-se a quarta equação do sistema a ser resolvida juntamente com as demais três equações:

$$\sum_{j=0}^{\infty} a_{11,j} = 0 = a_{11,0} \sum_{j=0}^{\infty} c_{11,j} + a_{21,0} \sum_{j=0}^{\infty} c_{12,j}.$$

Por conveniência, repetem-se aqui as quatro equações a serem resolvidas para encontrar os coeficientes da matriz A_0 :

$$\begin{aligned} a_{11,0}^2 + a_{12,0}^2 &= \text{var}(e_1); \\ a_{11,0}a_{21,0} + a_{12,0}a_{22,0} &= \text{cov}(e_1, e_2); \\ a_{21,0}^2 + a_{22,0}^2 &= \text{var}(e_2); \\ a_{11,0} \sum_{j=0}^{\infty} c_{11,j} + a_{21,0} \sum_{j=0}^{\infty} c_{12,j} &= 0. \end{aligned}$$

É claro que são necessários métodos computacionais para resolver o problema, haja vista a complicação que é inverter $(I - \sum_{j=1}^p \Phi_j L^j)$ ⁷. Felizmente, os pacotes econométricos fazem isso facilmente.

Uma vez obtida a matriz A_0 , podem-se recuperar os erros estruturais da seguinte relação:

$$c_t = A_0^{-1} e_t,$$

e, assim, encontrar a função resposta ao impulso e a decomposição da variância.

A partir dos erros estruturais, pode-se construir a resposta do produto a choques de demanda e de oferta. Blanchard e Quah (1989) advertem para não associar choques de demanda ao componente transitório do ciclo de negócios, e choques de oferta, ao componente permanente de tendência. Eles argumentam que o componente transitório pode ocorrer em virtude de uma combinação de choques na oferta e demanda, assim como choques de oferta afetarão simultaneamente os componentes cíclico e permanente da economia.

6.11

Estimação do Modelo Estrutural*

É possível estimar o modelo estrutural por máxima verossimilhança, depois de impor as restrições estabelecidas. Para isso, considere mais uma vez o modelo estrutural:

$$AX_t = \sum_{i=1}^p B_i X_{t-i} + B\epsilon_t, \quad c_t \sim i.i.d.N(0, I_n).$$

⁷ Sempre é possível visualizar a inversão de $(I - \sum_{j=1}^p \Phi_j L^j)$ usando substituição recursiva. No caso de um VAR(1), é possível derivar $X_t = \sum_{j=0}^{\infty} \Phi_1^j \epsilon_{t-j}$, o que implica que $C_j = \Phi_1^j$. Surgem complicações em modelos de ordem maior, porém fáceis de resolver usando um computador.

Empilham-se as matrizes X_t s, de modo que $Y_t = [X'_{t-1} \ X'_{t-2} \ \dots \ X'_{t-p}]'$, $A = A^{-1}[B_1, B_2, \dots, B_p]$. Logo, pode-se reescrever o modelo anterior como:

$$X_t = AY_t + A^{-1}B\epsilon_t,$$

em que B é a matriz de desvio-padrão de dimensão $n \times n$.

A função de verossimilhança será dada por:

$$\begin{aligned} \ln(A, A, B) = & -\frac{n(T-p)}{2} \ln 2\pi - \frac{(T-p)}{2} \ln |A^{-1}BB'A^{-1}| \\ & - \frac{1}{2} \sum_{t=p+1}^T (X_t - AY_t)' (A^{-1}BB'A^{-1})^{-1} (X_t - AY_t). \end{aligned}$$

Usando algumas propriedades para manipulação de matrizes que podem ser estudadas em detalhes em Lütkepohl (2005), pode-se reescrever o termo

$$\sum_{t=p+1}^T (X_t - AY_t)' (A^{-1}BB'A^{-1})^{-1} (X_t - AY_t)$$

da seguinte forma:

$$(T-p) \text{tr} \left[(X - AZ)' (A^{-1}BB'A^{-1})^{-1} (X - AZ) \right],$$

em que $X = [X'_{p+1}, X'_{p+2}, \dots, X'_T]'$ e $Z = [Y'_{p+1}, Y'_{p+2}, \dots, Y'_T]'$.

Assim, obtém-se:

$$\begin{aligned} \ln(A, A, B) = & \text{constante} + \frac{(T-p)}{2} (\ln |A|^2 - \ln |B|^2) \\ & - \frac{(T-p)}{2} \text{tr} [A'B'^{-1}B^{-1}A(X - AZ)'(X - AZ)]. \end{aligned}$$

Se não houver restrições sobre as matrizes A e B , pode-se estimar

$$\hat{\Sigma} = \frac{(X - \hat{A}Z)(X - \hat{A}Z)'}{T},$$

em que $(X - \hat{A}Z)$ são os resíduos do VAR reduzido, $\hat{A} = XZ'(ZZ')^{-1}$. Daí, é possível concentrar a função anterior para maximizar:

$$\begin{aligned} \ln_c(A, B) = & \text{constante} + \frac{(T-p)}{2} (\ln |A|^2 - \ln |B|^2) \\ & - \frac{(T-p)}{2} \text{tr} [A'B'^{-1}B^{-1}A\hat{\Sigma}]. \end{aligned}$$

6.11.1 Teste LR para Sobreidentificação

Ante a possibilidade de a economia prover mais restrições do que as necessárias para identificar o sistema, um teste de verossimilhança pode ser empreendido. Para isso, estimam-se os modelos sem e com restrição, obtendo-se o valor da verossimilhança em ambos os casos. Em seguida, procede-se ao teste de razão de verossimilhança.

1. Calcule o valor da matriz de covariância não restrita, dada por $\hat{\Sigma}$.
2. Como se sabe que $BB' = A\hat{\Sigma}A'$, pode-se reescrever $\hat{\Sigma}$ como:

$$\hat{\Sigma} = A^{-1}BB'A'^{-1},$$

impor as restrições sobre A e B para maximizar a função de verossimilhança e encontrar os demais coeficientes livres:

$$-\frac{T}{2} \ln |A^{-1}BB'A'^{-1}| - \frac{1}{2} \sum_{t=1}^T \hat{e}_t' (A^{-1}BB'A'^{-1})^{-1} \hat{e}_t.$$

3. Calcule a nova matriz de covariância restrita:

$$\hat{\Sigma}_r = \hat{A}^{-1}\hat{B}\hat{B}'\hat{A}'^{-1}.$$

4. Proceda ao teste estatístico de razão de verossimilhança dado por:

$$LR = T (\ln |\hat{\Sigma}_r| - \ln |\hat{\Sigma}|) \xrightarrow{d} \chi_{q'}^2,$$

em que q é o número de restrições acima de $n(n-1)/2$.

5. Se há dois conjuntos de restrições tais que $q_1 > q_2$, pode-se testar a hipótese de as restrições serem equivalentes pelo teste:

$$LR = T (\ln |\hat{\Sigma}_{r_1}| - \ln |\hat{\Sigma}_{r_2}|) \xrightarrow{d} \chi_{q_1 - q_2}^2.$$



Vetor de Correção de Erros – VECM

Uma das grandes críticas dos modelos *ARIMA* é o seu caráter eminentemente estatístico, muitas vezes sem qualquer significado econômico. De fato, quando se diferencia uma série para estacionarizá-la, perdem-se muitas de suas características (a constante, por exemplo). Além disso, séries de tempo com defasagens degeneradas – por exemplo, uma série explicada apenas pela primeira e quinta defasagens – são difíceis de justificar economicamente. No entanto, se séries de tempo não estacionárias têm uma dinâmica em comum, no sentido que será dado neste capítulo, pode-se especificar um modelo *VAR* mais completo denominado modelo vetor de correção de erros – *VECM*. Esse modelo possui significado econômico, pois pode-se dizer que suas variáveis, em virtude da dinâmica comum, têm um componente de longo prazo e um de curto prazo.

Para entender melhor essa terminologia, inicialmente define-se o que é cointegração, segundo Engle e Granger (1987):

Definição 7.1 (Engle e Granger) Os elementos do vetor X_t , $n \times 1$, são ditos cointegrados de ordem (d, b) , denotados por $X_t \sim CI(d, b)$, se:

- i. Todos os elementos de X_t são integrados de ordem d , ou seja, são $I(d)$;
- ii. Existe um vetor não nulo, β , tal que:

$$u_t = X_t' \beta \sim I(d - b), b > 0.$$

A primeira condição diz que todas as variáveis constantes de X_t devem ter a mesma ordem para que possam ser cointegradas. Na verdade, trata-se de uma condição bastante restrita, pois há modelos em economia que relacionam variáveis estacionárias com variáveis integradas. Essa condição será relaxada mais tarde, porém no momento retenha-a na mente para compreender o conceito.

As variáveis contidas em X_t guardam uma relação de equilíbrio de longo prazo. O termo que se utiliza é longo prazo porque as variáveis não são estacionárias e, portanto, têm uma tendência estocástica. Essa tendência estocástica justifica a designação de relação de longo prazo. Se a tendência estocástica for comum a todas as variáveis, diz-se que existe um equilíbrio de longo prazo.

Formalmente, diz-se que existe equilíbrio de longo prazo quando $X_t' \beta = 0$, isto é, o vetor β , chamado de vetor de cointegração, define uma combinação linear entre os elementos de X_t perfeita no sentido de seguir uma tendência em comum, sem desvio. Entretanto, no curto prazo há desvios dessa tendência comum, de modo que o termo u_t é o erro de equilíbrio, porque expressa os desvios temporários do equilíbrio de longo prazo. Neste momento, convém introduzir um exemplo para esclarecer as idéias apresentadas. Considere um vetor $\beta = [\tilde{\beta}_1 \quad \tilde{\beta}_2]'$ que define o equilíbrio de longo prazo entre as variáveis integradas de ordem 1, x_{1t} e x_{2t} , então:

$$\begin{bmatrix} x_{1t} & x_{2t} \end{bmatrix} \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \tilde{\beta}_1 x_{1t} + \tilde{\beta}_2 x_{2t} = 0.$$

Note que multiplicar ambos os lados da equação por um número $k < \infty$ não altera o equilíbrio de longo prazo. Isso significa que, na verdade, o vetor de cointegração não é único e, por isso, pode haver ambigüidades para defini-lo corretamente. Resolve-se esse problema multiplicando ambos os lados dessa equação por $1/\tilde{\beta}_1$, de forma a "normalizar" o vetor de cointegração:

$$\begin{bmatrix} x_{1t} & x_{2t} \end{bmatrix} \begin{bmatrix} 1 \\ \beta_2 \end{bmatrix} = x_{1t} + \beta_2 x_{2t} = 0, \quad \beta_2 \equiv \frac{\tilde{\beta}_2}{\tilde{\beta}_1}.$$

Portanto, o coeficiente da primeira variável endógena é sempre fixado em 1 e, daqui em diante, as referências a β serão feitas à sua versão normalizada. Convém dizer também que, mesmo quando há desequilíbrio de curto prazo, essa normalização não afetará as propriedades de u_t , por se tratar apenas de uma simples mudança de volatilidade nesse termo de desequilíbrio.

Normalmente, séries econômicas têm ordem de integração menor que 2. Séries estacionárias têm ordem de integração 0, a exemplo dos retornos de ações. Em outras palavras, como já sabido, a ordem de integração é o número de

diferenças que deve ser aplicado à série para torná-la estacionária. A diferença $(d - b)$ é a ordem de integração da série resultante da aplicação do vetor β a X_t .

Há uma forma interessante de interpretar u_t a partir disso. Dada a normalização que se faz, o termo u_t pode ser considerado o resíduo de uma coordenada do vetor X_t contra as demais variáveis. Assim, se as variáveis são cointegradas, a aplicação de β sobre essas variáveis vai gerar um resíduo, cuja ordem de integração será menor que a ordem das variáveis originais. Para ganhar intuição sobre o que isso significa, convém imaginar um modelo bivariado, cujas variáveis são integradas de ordem 1. Se essas variáveis cointegram, significa que o resíduo resultante de uma contra outra é estacionário, ou seja, de ordem 0. Economicamente, isso significa que há uma relação de longo prazo entre as coordenadas desse vetor, pois a ordem de integração da série diminui. Essa relação de longo prazo é perturbada por choques de curto prazo representados por u_t . Porém, sendo estacionário, trata-se de uma perturbação que tende a dissipar-se ao longo do tempo, de modo que as variáveis originais voltam a seu equilíbrio de longo prazo. O tempo necessário para a dissipação do choque vai depender do ajustamento do modelo, evidenciado por um parâmetro chamado ajustamento, o qual será discutido mais tarde.

A teoria de cointegração preocupa-se, então, com dois pontos fundamentais. O primeiro é testar os resíduos u_t , para constatar tratar-se de uma variável estacionária. O segundo ponto é, dado que u_t é estacionário, usar essa informação para ajustar melhor o modelo VAR, denominado por causa desse ajuste de quase-VAR. Quando se utiliza a informação decorrente desse fato, tem-se o modelo VECM, em que se incorpora o erro de equilíbrio, por isso a denominação modelo vetor de correção de erros.

Considerando a necessidade de obter uma perturbação estacionária, a questão que emerge é a de como testar se os resíduos são, de fato, estacionários. Parece ser natural, nesse caso, pensar no teste de raiz unitária sobre os resíduos, e é no que constitui o teste de Engle e Granger, objeto de estudo na próxima seção. Essencialmente, os autores desenvolvem o teste de raiz unitária para os resíduos de variáveis cointegradas.

Para visualizar como as séries se comportam, a Figura 7.1 reproduz uma simulação em que há duas séries que cointegram: $y_t = \mu_t + \epsilon_{yt}$ e $z_t = \mu_t + \epsilon_{zt}$, em que μ_t é um passeio aleatório. Os resíduos são estacionários, e a inspeção visual sugere, verdadeiramente, que eles não divergem.

A simulação procura esclarecer por que séries não estacionárias teriam uma mesma trajetória de longo prazo. Acaba mostrando que a razão é ter um elemento de longo prazo comum entre elas. Na prática, é nas razões econômicas que se procurará a justificativa para o equilíbrio de longo prazo das variáveis.

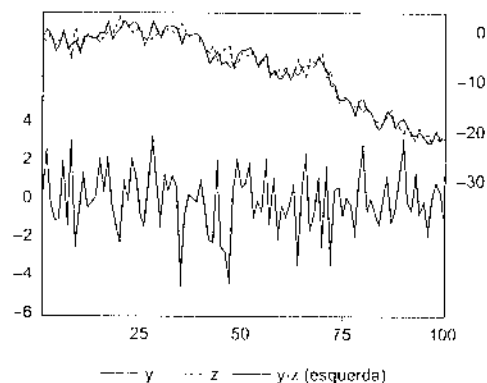


Figura 7.1 Variáveis cointegradas: $y_t = \mu_t + c_{yt}$ e $z_t = \mu_t + c_{zt}$, em que μ_t é um passeio aleatório.

Nem sempre, contudo, é imediato inferir se as séries cointegram. Esse problema é particularmente mais evidente em modelos com mais de duas variáveis. De fato, a Figura 7.2 contém séries simuladas a partir do mesmo valor inicial e que parecem divergir ao longo do tempo. Entretanto, as séries foram construídas de modo a que cointegassem. Na verdade, as séries não cointegram quando tomadas duas a duas, entretanto, quando simultaneamente combinadas, elas geram resíduos estacionários.

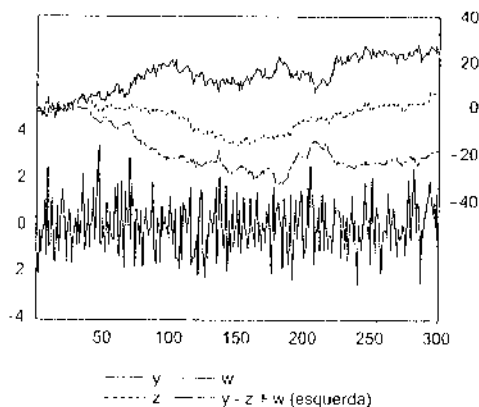


Figura 7.2 $y_t = y_{t-1} + c_{yt}$, $z_t = z_{t-1} + c_{zt}$, $w_t = w_{t-1} + c_{wt}$.

Se duas variáveis têm ordens de integração diferentes, qualquer combinação linear entre elas resultará em uma variável cuja ordem de integração será a de maior ordem. Em outras palavras, a ordem de integração da variável de maior ordem domina a da variável de menor ordem. Esse fato sugere a necessidade de as variáveis serem de mesma ordem para haver cointegração, o que é uma

conclusão verdadeira. Contudo, em um modelo em que o número de variáveis endógenas é maior que 2, nem todas as variáveis precisam ter a mesma ordem de integração para existir cointegração. Supondo o caso de três variáveis, $X_t = [y_t, x_t, z_t]'$, sendo duas integradas de ordem 2, y_t e x_t , e uma integrada de ordem 1, z_t , é possível imaginar que as duas variáveis de maior ordem sejam $C(2, 1)$ e a resultante, quando combinada com z_t , seja $C(1, 1)$. Isto é, pode-se imaginar um vetor β entre y_t e x_t que gere uma variável, $w_t \sim I(1)$, e, a seguir, um vetor π que combine w_t e z_t , resultando em um modelo estacionário. Formalmente, as contas são:

$$\frac{1}{\pi_1} y_t + \frac{\beta_0}{\pi_1} x_t = w_t \sim I(1);$$

$$\pi_1 w_t + \pi_2 z_t = u_t \sim I(0).$$

Outra forma mais direta de escrever é usando um único vetor $\beta = [1 \ \beta_0 \ \pi_2]'$:

$$\begin{bmatrix} y_t & x_t & z_t \end{bmatrix} \begin{bmatrix} 1 \\ \beta_0 \\ \pi_2 \end{bmatrix} = y_t + \beta_0 x_t + \pi_2 z_t = u_t.$$

A conclusão dessa abstração é a necessidade de se ter pelo menos duas variáveis integradas de mesma ordem na ordem máxima de integração entre todas as variáveis, para existir cointegração. A combinação linear das coordenadas do vetor X_t será estacionária, isto é, u_t será $I(0)$ se, aplicando o teste de raiz unitária nos resíduos estimados, rejeita-se a hipótese nula de raiz unitária, conforme será esclarecido mais adiante. É claro que u_t pode ser expresso como um $ARMA(p, q)$, da seguinte forma: $u_t = (\theta(L) / \phi(L)) \epsilon_t$.

O tipo de combinação descrito determina uma definição mais abrangente do que a de Engle e Granger para cointegração, enunciada por Campbell e Perron (1991):

Definição 7.2 (Campbell e Perron) Os elementos do vetor X_t , $n \times 1$ são ditos cointegrados de ordem (d, b) , denotados por $X_t \sim CI(d, b)$, se existe pelo menos um vetor β não nulo tal que:

$$u_t = X_t' \beta \sim I(d - b), b > 0.$$

A vantagem mais prática dessa formulação é poder testar hipóteses econômicas não abrangidas pela definição de Engle e Granger. Se fosse mantida aquela definição e se houvesse um modelo econômico que relacionasse variáveis não estacionárias com estacionárias, tal modelo não seria possível de ser testado adequadamente.

Segundo essa definição, não é preciso que os elementos de X_t tenham a mesma ordem de integração, o que é desejável pela abrangência de casos que interessam, mas engloba também possibilidades triviais e de interesse limitado. Um exemplo trivial ocorre ante a existência de uma variável $I(0)$ e uma variável $I(1)$, de maneira que o exemplo trivial ocorre ante a existência de uma variável $I(0)$ e uma variável $I(1)$, de maneira que o vetor de cointegração terá coeficiente zero para a variável $I(1)$. De qualquer forma, a importância dessa definição está na abrangência de todos os casos de interesse. Em geral, estudam-se apenas os casos em que as variáveis endógenas são integradas de ordem 1 ou há várias integradas de ordem 1 e algumas de ordem zero.

Convém dizer finalmente que o modelo pode ser estendido para incluir constante e tendência determinística:

$$u_t = X_t' \beta - \mu_1 t - \mu_0.$$

7.1

Teste de Cointegração de Engle-Granger

O teste de cointegração de Engle e Granger (1987) é indicado para ser feito sobre uma única equação. Trata-se de um teste imediato e fácil de implementar. Em um modelo de várias variáveis, é possível existir mais de um vetor de cointegração. Especificar quais equações devem ser testadas constitui um problema a ser resolvido antes do teste. Não obstante, a solução estará na especificação das relações econômicas entre essas variáveis.

Para iniciar, suponha um sistema bivariado nas variáveis y_t e x_t , integradas de ordem 1. Engle e Granger propõem uma metodologia a três passos para determinar se essas variáveis são $CI(1, 1)$. A metodologia consiste em estimar a relação de longo prazo e armazenar os resíduos. Se as variáveis forem cointegradas, os resíduos serão estacionários. Portanto, o objetivo é fazer o teste de raiz unitária nos resíduos. Se a hipótese nula de raiz unitária dos resíduos não for rejeitada, as variáveis não serão cointegradas. Em contraste, a rejeição da hipótese nula implica que as variáveis são cointegradas, pois os resíduos serão estacionários. Procede-se ao teste da seguinte forma:

1. Execute o teste de raiz unitária nas variáveis de interesse e certifique-se de que são $I(1)$;

¹ O caso de haver variáveis $I(2)$ em conjunto com $I(1)$ será discutido a seguir.

2. Estime a relação de longo prazo e obtenha \hat{u}_t , em que o circunflexo representa parâmetro estimado;
3. Faça o teste de raiz unitária nos resíduos estimados, usando o procedimento ADF:

$$\Delta \hat{u}_t = \alpha \hat{u}_{t-1} + \sum_{i=1}^{p-1} \lambda_{i+1} \Delta \hat{u}_{t-i} + v_t.$$

A não-rejeição de $H_0 : \alpha = 0$ implica que os resíduos têm raiz unitária, de modo que as variáveis não cointegram.

Como os resíduos são estimados, não é possível usar os valores tabulados convencionalmente. Trata-se, agora, de usar uma tabela apropriada a variáveis estimadas, como aquela encontrada no próprio trabalho de Engle e Granger (1987). Outra alternativa é utilizar o procedimento de MacKinnon (1991), o qual inova porque abarca todos os tamanhos de amostra. A tabela depende do número de observações, número de variáveis endógenas e existência ou não de constante e tendência linear.

O coeficiente β é estimado por mínimos quadrados ordinários. Na presença de cointegração, ele será superconsistente. O conceito de superconsistência significa que o coeficiente se aproxima mais rapidamente do seu verdadeiro valor do que se fosse estimado com variáveis estacionárias diferenciadas² quando a amostra é aumentada. O resultado é em virtude do fato de as variáveis serem integradas, pois o efeito de não-estacionaridade se superpõe ao de estacionaridade, mesmo que haja significativa autocorrelação dos resíduos.

Nota 7.1 *Tendência determinística domina a não-estacionaridade de primeira ordem, que domina estacionaridade, no sentido de convergência assintótica.*

Em um sistema de n variáveis endógenas, haverá no máximo $n - 1$ vetores de cointegração. Para entender por que não se pode ter mais do que $n - 1$ vetores de cointegração, suponha o caso mais simples em que $n = 2$, isto é, um sistema bivariado (y_t, x_t) em que ambas as variáveis são $I(1)$ e possuem um vetor de

² Isto é, se a relação de cointegração é dada por $y_t = \beta x_t + \varepsilon_t$, o mesmo β pode ser estimado quando se subtrai y_{t-1} no lado esquerdo da equação anterior e $\beta x_{t-1} + \varepsilon_{t-1}$ no lado direito, resultando no modelo $\Delta y_t = \beta \Delta x_t + \Delta \varepsilon_t$. Entretanto, a taxa de convergência do estimador nesse caso é bem mais lenta que no caso de estimar o coeficiente no nível.

cointegração $(1, -b)$, tal que $y_t - bz_t \sim I(0)$. Agora, suponha que exista outro vetor de cointegração, $(1, -c)$, que torna as séries $I(0)$: $y_t - cz_t \sim I(0)$, então é forçoso concluir a seguinte relação:

$$y_t - cz_t = \underbrace{y_t - bz_t}_{\sim I(0)} + \underbrace{(b - c)z_t}_{I(1) \text{ ou } b=c}.$$

Em palavras, se $b = c$, então existe somente um vetor de cointegração. Se $b \neq c$, então a segunda parcela do lado direito da equação passa a ser importante. Nesse caso, para que haja estacionaridade no lado esquerdo, é preciso que $z_t \sim I(0)$, o que é um absurdo, pois foi suposto que $z_t \sim I(1)$.

Uma vez concluído o teste de raiz unitária sobre os resíduos e constatado que são estacionários, o modelo original pode ser escrito na forma de correção de erros. Esse modelo é como um VAR convencional, porém se acrescenta o erro de equilíbrio e, por isso, chama-se VECM. Para visualizar um modelo VECM simples, considere que a relação de cointegração entre y_t e z_t seja:

$$y_t = \mu + \beta z_t + u_t.$$

Então, existem formas de manipular o modelo VAR de modo que, se existe cointegração, pode-se reescrever o modelo original de tal maneira que os resíduos entram explicitamente no VAR resultante. Os detalhes do procedimento serão explicados mais detalhadamente quando for discutido o modelo de correção de erros. Assim, introduzir os erros no modelo VAR implica estimar o seguinte modelo, para o caso bivariado $X_t = [y_t \quad z_t]'$:

$$\begin{aligned}\Delta y_t &= \alpha_1 \hat{u}_{t-1} + \sum_{j=1}^{p-1} \lambda_{11,j+1} \Delta y_{t-j} + \sum_{j=1}^{p-1} \lambda_{12,j+1} \Delta z_{t-j} + e_{yt}, \\ \Delta z_t &= \alpha_2 \hat{u}_{t-1} + \sum_{j=1}^{p-1} \lambda_{21,j+1} \Delta y_{t-j} + \sum_{j=1}^{p-1} \lambda_{22,j+1} \Delta z_{t-j} + e_{zt}.\end{aligned}$$

Pede-se ao leitor que aguarde mais alguns instantes e deriva-se essa relação a partir do modelo original. Por enquanto, apenas verifique como se usa o resíduo da cointegração para estimar o VAR. É claro que, em relação ao modelo VAR original, simplesmente nas diferenças, há alteração da magnitude dos coeficientes. O importante é perceber, contudo, que não incluir os resíduos de cointegração no modelo implica incorrer no erro de omissão de variáveis explicativas, que resulta em estimativas de parâmetros viesadas.

No caso mais geral em que $n > 2$, haverá $r \leq n - 1$ vetores de cointegração (se $r = 0$, não existe cointegração entre as variáveis). Para cada vetor de cointegração,

haverá um resíduo estacionário. Mais claramente, na situação em que $r < n$, deve-se considerar a regressão de cada equação diferenciada com relação aos r resíduos e às n variáveis endógenas, de forma que \hat{u}_{t-1} passaria a ser um vetor de resíduos e α_i , $i = 1, 2, \dots, n$ seria um vetor $r \times 1$ a multiplicar o vetor \hat{u}_{t-1} .

7.1.1 Teste de Engle-Granger com Variáveis $I(2)$

Multicointegração está associada a situações em que combinações lineares de variáveis $I(2)$ com variáveis $I(1)$ são estacionárias. Um exemplo é quando se consideram as variáveis x_{1t} e x_{2t} integradas de ordem 2 e o vetor $y_t \sim I(1)$. Então, é possível ter:

$$x_{1t} = b_2 x_{2t} + y_t' \omega + u_t.$$

Note que $b_2 = 0$ impediria a obtenção de estacionaridade, porque, sob essa hipótese, teríamos $x_{1t} \sim I(2)$ combinado com variáveis $I(1)$, tornando impossível a existência de estacionaridade.

A especificação anterior pode ser desdobrada em outras relações com as mesmas variáveis. Um exemplo seria usar as variáveis de maior integração em sua diferença, conforme a seguinte especificação:

$$x_{1t} = b_2 x_{2t} + y_t' \omega + \gamma_2 \Delta x_{2t} + u_t.$$

Sejam x_{1t} , x_{2t} e x_{3t} variáveis $I(2)$ e y_t , um vetor de variáveis $I(1)$. Então, o teste é feito quase como na seção anterior, da seguinte forma:

1. Estima-se o modelo completo:

$$x_{1t} = \mu + \delta t + b_2 x_{2t} + b_3 x_{3t} + \gamma_2 \Delta x_{2t} + \gamma_3 \Delta x_{3t} + y_t' \omega + u_t.$$

2. Procede-se ao teste de raiz unitária nos resíduos estimados:

$$\Delta \hat{z}_t = \alpha \hat{z}_{t-1} + \sum_{i=1}^{p-1} \lambda_{i+1} \Delta \hat{z}_{t-i} + u_t.$$

A não-rejeição de $H_0 : \alpha = 0$ implica que os resíduos têm raiz unitária e que as variáveis não cointegram. Os valores críticos nesse caso dependem do número de regressores que são $I(2)$, $m_2 = 1, 2$, do número de regressores que são $I(1)$, $m_1 = 0, 1, \dots, 4$, do número de observações e da forma dos regressores determinísticos. Veja Enders (2004) para os detalhes das tabelas.

7.2

Modelo de Correção de Erros

Esta seção elabora mais a versão do modelo de correção de erros, estendendo-se para possibilitar a existência de muitas variáveis endógenas na presença de diversos vetores de cointegração, culminando com a descrição completa do modelo vetor de correção de erros, VECM.

O modelo vetor de correção de erros é uma versão mais completa do VAR. A idéia básica é que o VAR com variáveis não estacionárias, mas diferenciadas, omite variáveis relevantes. O VECM corrige esse problema. Para entendê-lo, considere o modelo multivariado VAR no nível, ignorando a possível existência de constante:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + e_t.$$

Esse é o modelo multivariado em que cada X_t é um vetor $n \times 1$ de variáveis endógenas. A ordem p de defasagem pode ser escolhida de tal maneira que os resíduos estimados sejam não autocorrelacionados entre si ou de acordo com o critério de Schwarz.

Pode-se reescrever a equação anterior mais convenientemente da seguinte maneira:

$$\begin{aligned} \left[I - (\Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_p L^p) \right] X_t &= e_t \implies \\ \Phi(L) X_t &= e_t. \end{aligned}$$

Observe, em particular, que, quando $L = 1$, então:

$$\Phi(1) = [I - (\Phi_1 + \Phi_2 + \cdots + \Phi_p)] \equiv -\Phi.$$

Essa formulação será adequada daqui a pouco, quando serão apresentadas algumas definições e teoremas que ajudarão a determinar o VECM. O polinômio característico de $\Phi(L)$ é dado por:

$$\Phi(Z) = I - \sum_{i=1}^p \Phi_i Z^i,$$

em que Z é uma matriz diagonal com n elementos³.

O termo $\Phi(Z)$ vai ajudar a determinar a existência ou não de raiz unitária e a quantidade de vetores de cointegração a partir do seu posto. O polinômio

³ Observe que, sendo Z uma matriz diagonal com elementos idênticos na diagonal, então: $AZ = Az$, em que z é o elemento da diagonal de Z .

característico tem o mesmo significado do caso univariado, em que, para que haja estacionaridade, todos os autovalores da matriz Φ devem estar fora do círculo unitário.

No caso multivariado, há alterações na forma de obter determinados resultados análogos ao caso univariado, porém as idéias fundamentais são idênticas e usadas para explicar os resultados de forma mais intuitiva. Assim, sabe-se da álgebra linear que, se o determinante de uma matriz é nulo, seu posto não é pleno. Isto é, $|\Phi(I)| = 0 \iff \text{posto}(\Phi) < n$, ou seja, o processo tem uma raiz unitária, de forma que $\Phi(Z)$ pode ser fatorada de tal maneira que:

$$\Phi(Z) = (I - Z)(I - \lambda_1 Z)(I - \lambda_2 Z) \cdots (I - \lambda_{p-1} Z).$$

O posto de uma matriz é o número de linhas ou colunas independentes dessa matriz. O posto será sempre menor ou igual ao menor número entre colunas e linhas. Ou seja, para uma matriz qualquer A com dimensão $r \times n$, $\text{posto}(A) \leq \min[r, n]$. Uma propriedade da matriz quadrada é que, se seu posto é pleno, ou seja, para uma matriz $r \times r$, com posto igual a r , o determinante dessa matriz é não nulo. Com isso em mente, é possível enunciar o teorema de Granger. O teorema de Granger serve para separar a matriz de cointegração da matriz de ajustamento, cujas idéias fundamentais já foram discutidas anteriormente.

Teorema 7.1 (Granger) Se $|\Phi(Z)|_1 = 0$ implica que $Z \geq 1$ e $0 < \text{posto}(\Phi) = r < n$, então existem as matrizes α e β de dimensão $n \times r$ tal que: $\Phi = \alpha\beta'$.

Como se verá, β é chamada matriz de cointegração e α , matriz de ajustamento. O teorema de Granger expressa a idéia de que se pode decompor Φ em duas matrizes multiplicativas quando a matriz não tem posto pleno. Essa será uma propriedade bastante adequada para dar uma interpretação econômica a séries temporais e é a base para desenvolver o teste de cointegração multivariado de Johansen, a ser visto adiante.

A partir desses conceitos, pode-se formular o modelo vetor de correção de erros. Esse modelo é obtido por manipulações algébricas⁴ da equação $\Phi(L)X_t = e_t$. Inicialmente o modelo será desenvolvido, admitindo-se tratar de um VAR(3), depois será enunciado o resultado para o caso de um VAR(p), deixando ao leitor a tarefa de derivar o modelo mais geral. Assim, seja:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \Phi_3 X_{t-3} + e_t.$$

⁴ De forma semelhante à técnica utilizada para obter a equação do teste de Dickey-Fuller aumentado

A essa equação, some e subtraia $\Phi_3 X_{t-2}$:

$$\begin{aligned} X_t &= \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \Phi_3 X_{t-2} - \Phi_3 X_{t-2} + \Phi_3 X_{t-3} + e_t \\ &= \Phi_1 X_{t-1} + (\Phi_2 + \Phi_3) X_{t-2} - \Phi_3 \Delta X_{t-2} + e_t. \end{aligned}$$

Ao resultado anterior, some e subtraia $(\Phi_2 + \Phi_3) X_{t-1}$:

$$\begin{aligned} X_t &= \Phi_1 X_{t-1} + (\Phi_2 + \Phi_3) X_{t-1} - (\Phi_2 + \Phi_3) X_{t-1} \\ &\quad + (\Phi_2 + \Phi_3) X_{t-2} - \Phi_3 \Delta X_{t-2} + e_t \\ &= (\Phi_1 + \Phi_2 + \Phi_3) X_{t-1} - (\Phi_2 + \Phi_3) \Delta X_{t-1} - \Phi_3 \Delta X_{t-2} + e_t. \end{aligned}$$

Finalmente, subtraia X_{t-1} de ambos os lados para obter a versão do VAR na forma VECM:

$$\begin{aligned} X_t - X_{t-1} &= -X_{t-1} + (\Phi_1 + \Phi_2 + \Phi_3) X_{t-1} - (\Phi_2 + \Phi_3) \Delta X_{t-1} - \Phi_3 \Delta X_{t-2} + e_t \\ \Delta X_t &= -[I - (\Phi_1 + \Phi_2 + \Phi_3)] X_{t-1} - (\Phi_2 + \Phi_3) \Delta X_{t-1} - \Phi_3 \Delta X_{t-2} + e_t \\ &= \Phi X_{t-1} + \sum_{i=1}^2 \Lambda_i \Delta X_{t-i} + e_t, \end{aligned}$$

em que $\Lambda_i = -\sum_{j=1+i}^3 \Phi_j$, $i = 1, 2$.

O caso geral é derivado de forma semelhante e direta a partir do que foi feito anteriormente:

$$\Delta X_t = \Phi X_{t-1} + \sum_{i=1}^{p-1} \Lambda_i \Delta X_{t-i} + e_t,$$

em que $\Lambda_i = -\sum_{j=1+i}^p \Phi_j$, $i = 1, 2, \dots, p-1$.

Se há raiz unitária, é claro que $\Phi(I) = 0$, de modo que $\Phi = \alpha\beta'$. Então, pode-se dizer que β é a matriz que tem r vetores de cointegração e α é a matriz de ajustamento, com r vetores de ajustamento.

O modelo de correção de erros é assim chamado porque explica ΔX_t por dois componentes: os fatores de curto prazo, $\sum_{i=1}^{p-1} \Lambda_i \Delta X_{t-i}$, e a relação de longo prazo dada entre as coordenadas do vetor de variáveis endógenas, ΦX_{t-1} , considerando que haja cointegração.

Sempre é possível associar ao VAR a correção de erros na presença de cointegração? Haveria algum caso em que isso não seria possível? O teorema da representação de Granger mostra que a resposta à primeira pergunta é sim, e à segunda, não.

Teorema 7.2 (teorema da representação de Granger) Se $X_t \sim CI(1, 1)$, X_t tem representação em forma de VECM.

Convém dizer que Campbell e Perron (1991) não mostram qualquer objeção com relação à sua definição de cointegração com o teorema da representação. Com efeito, em seu artigo, o VECM é colocado sem qualquer restrição.

Há um caso interessante de cointegração não abrangido pelo desenvolvimento feito até aqui: é o caso do VAR (0). Trata-se de um modelo univariado em que há cointegração entre variável explicada e explicativas, porém sem a necessidade de variáveis defasadas. Esse caso é desenvolvido a seguir, iniciando por meio de um exemplo.

Exemplo 7.1

Considere o modelo normalizado $y_t = \pi x_t + e_t$. Subtraia y_{t-1} de ambos os lados da equação e some e subtraia πx_{t-1} no lado direito:

$$y_t - y_{t-1} = \pi x_t - \pi x_{t-1} - y_{t-1} + \pi x_{t-1} + e_t \implies$$

$$\Delta y_t = \pi \Delta x_t - (y_{t-1} - \pi x_{t-1}) + e_t \implies$$

$$\Delta y_t = \pi \Delta x_t - \begin{bmatrix} 1 & -\pi \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + e_t.$$

Em vez de estimar o modelo como se apresenta, poderia ser estimado da seguinte forma. Primeiro, estima-se a relação entre y_t e x_t para obter os resíduos estimados da relação de cointegração. Em seguida, estima-se o modelo na primeira diferença, incluindo os resíduos estimados da relação de cointegração para obter o coeficiente de ajustamento α e o coeficiente π . Isto é:

$$1. \text{ Obtenha } \hat{u}_t = y_t - \beta_1 x_t;$$

$$2. \text{ Use o resíduo estimado para estimar a regressão:}$$

$$\Delta y_t = \pi \Delta x_t + \alpha \hat{u}_{t-1} + e_t.$$

Formalmente, operou-se a decomposição da matriz $\begin{bmatrix} 1 & -\pi \end{bmatrix}$ em $\alpha \beta'$. Esquematicamente, isso está contido nas seguintes equações:

$$\Delta y_t = \pi \Delta x_t + \alpha \hat{\beta}' \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + e_t;$$

$$\Delta y_t = \pi \Delta x_t + \alpha \hat{u}_{t-1} + e_t.$$

O que é importante notar aqui é que o procedimento permite estimar três parâmetros no modelo, em vez de apenas um se não fosse considerada a possibilidade de cointegração: α , β_1 e π . Alves e Bueno (2003) aplicam esse modelo para

estimar a elasticidade no preço da demanda por gasolina, a elasticidade cruzada da demanda de gasolina em relação ao álcool e a elasticidade renda da gasolina, tanto no curto quanto no longo prazo.

Assim, pode-se interpretar α como a velocidade de ajustamento quando o modelo sai de sua trajetória de longo prazo. Para entender melhor, suponha que y_{t-1} descole de x_{t-1} , de modo que $u_{t-1} > 0$. Quanto maior α , mais efeito terá sobre y_t , no sentido de reajustar o modelo em direção a seu equilíbrio de longo prazo. A correção é dinâmica, por meio da diferença entre y_t e y_{t-1} .

7.3

Teste de Cointegração de Johansen

Johansen propõe um teste para definir o posto da matriz Φ e, assim, estimar os vetores de cointegração contidos na matriz β . Sua metodologia é interessante porque é empreendida simultaneamente à estimação do modelo de cointegração. Isto é, a metodologia de Johansen permite a estimação do VECM simultaneamente aos vetores de cointegração.

Mesmo definindo o posto de Φ , pode não ser possível identificar α e β . Por isso, propõe-se uma normalização aos vetores de cointegração, de forma a restringir as possibilidades que essas matrizes podem assumir.

Nota 7.2 As matrizes α e β não são identificáveis, pois qualquer matriz F não singular resulta: $\Phi = \alpha F (F^{-1}) \beta'$, em que F é uma matriz $r \times r$. Para a normalização, a matriz F será tal que representa o inverso do menor principal de β de dimensão $r \times r$. Isto é, a matriz F é utilizada para normalizar os vetores de cointegração.

Exemplo 7.2

Suponha o seguinte sistema de equações não-estacionárias na forma de vetor de cointegração:

$$y_t = bx_t + cz_t + u_{1t};$$

$$x_t = dy_t + ez_t + u_{2t}.$$

Colocando esse sistema na forma vetorial, encontra-se:

$$\begin{bmatrix} 1 & -b & c \\ -d & 1 & -e \end{bmatrix} \begin{bmatrix} y_t \\ x_t \\ z_t \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}.$$

A matriz que multiplica o vetor $X_t = [y_t \ x_t \ z_t]'$ é a matriz β , sem qualquer normalização. É uma matriz de posto 2, pois a última linha é uma combinação das anteriores. Aplicando a inversa de

$$\begin{bmatrix} 1 & -b \\ -d & 1 \end{bmatrix}$$

sobre essa matriz, pode-se normalizá-la:

$$\begin{aligned} \begin{bmatrix} 1 & -b \\ -d & 1 \end{bmatrix}^{-1} &= \frac{1}{1-bd} \begin{bmatrix} 1 & b \\ d & 1 \end{bmatrix} \Rightarrow \\ \frac{1}{1-bd} \begin{bmatrix} 1 & b \\ d & 1 \end{bmatrix} \begin{bmatrix} 1 & -b & -c \\ -d & 1 & -e \end{bmatrix} &= \begin{bmatrix} 1 & 0 & \frac{1}{bd-1}(c+be) \\ 0 & 1 & \frac{1}{bd-1}(e+cd) \end{bmatrix}. \end{aligned}$$

Para identificar o posto, Johansen propõe dois testes, baseados em uma estimação de máxima verossimilhança com restrição. A idéia de Johansen é usar a configuração multivariada e procurar o posto da matriz Φ de uma forma bastante inteligente. A derivação matemática desses testes é complexa e é deixada para outra seção, opcional. O leitor poderá procurar nas referências bibliográficas, ao final do texto, as fontes para entender tecnicamente os procedimentos sugeridos por Johansen.

Entretanto, é possível entender os testes mais intuitivamente. Essencialmente, Φ é uma matriz $n \times n$, cujo posto é $r < n$, se houver cointegração. Se o posto dessa matriz for n , então as variáveis endógenas são todas estacionárias. Se o posto da matriz for nulo, então não existe cointegração e as variáveis são não estacionárias. A idéia de posto nulo é análoga à idéia de raiz unitária no modelo univariado. Naquele caso, o coeficiente que multiplicava y_{t-1} era nulo ante a presença de raiz unitária. No caso multivariado, posto nulo significa que a matriz $\Phi = 0$, portanto trata-se de uma raiz unitária multivariada.

Sabendo que o determinante de uma matriz é o produto de seus autovalores e imaginando que o posto de Φ está entre 0 e n , então é possível afirmar que Φ terá r autovalores diferentes de zero e $n - r$ autovalores iguais a zero. O problema, portanto, é encontrar esses autovalores.

Para estabelecer o teste, convém, neste momento, introduzir variáveis determinísticas ao modelo, uma vez que os valores críticos dependem da configuração dessas variáveis. As variáveis determinísticas farão parte da variável X_t no nível e do vetor de cointegração:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \delta' d_t + e_t,$$

em que

$d_t = [1, t]'$ é um vetor com variáveis determinísticas e que poderia incluir também *dummies* sazonais ou outras variáveis determinísticas;

δ é uma matriz de coeficientes, cuja dimensão é compatível com d_t , nesse caso com dimensão $2 \times n$.

Reescrevendo o modelo anterior na forma VECM, obtém-se:

$$\Delta X_t = \Phi X_{t-1} + \sum_{i=1}^{p-1} \Lambda_i \Delta X_{t-i} + \delta' d_t + e_t.$$

Maximizando essa especificação com restrições sobre a matriz de covariância, conforme procedimento descrito no fim deste capítulo, é possível obter os autovalores da matriz Φ . Esses autovalores podem ser ordenados, sem perda de generalidade, do maior para o menor: $\lambda_1 > \lambda_2 > \cdots > \lambda_n$, cada um deles correspondente a um determinado autovetor que será associado aos vetores de cointegração contidos em β . Do ponto de vista prático, o pacote econométrico a ser utilizado calculará os autovalores e os ordenará, bem como procederá ao teste com o respectivo valor crítico, já que a distribuição do teste, assim como no caso de raiz unitária, não é convencional. A partir desse ponto, há dois testes que podem ser empreendidos, ambos designados por Johansen.

O primeiro teste é o do traço, que assume como hipótese nula a existência de r^* vetores de cointegração contra a hipótese alternativa de $r > r^*$ vetores. Formalmente:

$$H_0 : r = r^* \times H_1 : r > r^*.$$

A estatística do teste é dada por:

$$\lambda_{tr}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i).$$

Nota 7.3 Os autovalores são normalizados de tal maneira que sempre serão menores que 1, portanto o \ln será negativo. Isso é possível por causa da não-identificabilidade da matriz Φ .

A lógica do teste é a seguinte. O posto da matriz Φ é igual ao número de suas raízes características diferentes de zero. Se não existe cointegração, então os autovalores obtidos serão próximos a zero, denotando não-estacionaridade

ou instabilidade da matriz Φ , e $\ln(1 - \lambda_i) \rightarrow 0$. Se isso acontece, a estatística do traço resulta em valores pequenos, de tal modo que não se pode rejeitar a nula. Se, por outro lado, λ_i é significativamente diferente de zero, então $\ln(1 - \lambda_i)$ será negativo. A estatística terá um valor alto, e a nula será rejeitada em favor da alternativa.

O teste é crescente, ou seja, inicia-se com a soma de n “logs” de 1 menos o autovalor, admitindo-se inicialmente que $r^* = 0$. Rejeitando a hipótese nula, significa que há mais de um vetor de cointegração. Então, parte-se para a soma dos $n - 1$ “logs” de 1 menos os autovalores correspondentes, até que não se rejeite mais a hipótese nula de que há r^* vetores de cointegração.

A distribuição desse teste é um processo de difusão, ou seja, não é uma distribuição convencional. A tabela mais recente a respeito dos valores críticos pode ser encontrada em MacKinnon, Haug e Michelis (1999). Normalmente, os pacotes econométricos já possuem essas tabelas embutidas.

O segundo teste é o de razão de verossimilhança, aparentemente com resultados mais robustos que o anterior, mas também com distribuição não convencional. A hipótese nula desse teste é que existem r^* vetores de cointegração; a hipótese alternativa é que existem $r^* + 1$ vetores de cointegração:

$$H_0 : r = r^* \times H_1 : r = r^* + 1.$$

A estatística do teste é dada por:

$$LR(r) = -T \ln(1 - \hat{\lambda}_{r+1}).$$

No fundo, é um teste que verifica qual o máximo autovalor significativo que produz um vetor de cointegração. Esse autovalor máximo correspondente ao vetor de cointegração r^* mostra que há r^* vetores de cointegração. Como o teste anterior, é um teste crescente. Rejeitar H_0 significa que há mais um vetor de cointegração. Não rejeitar H_0 significa que há r^* vetores de cointegração.

É preciso determinar cuidadosamente o modelo de cointegração a ser testado e a ordem de defasagens dos componentes auto-regressivos. Para determinar a ordem de defasagem, é preciso combinar, simultaneamente, o melhor critério de informação com a ausência de autocorrelação dos resíduos. O procedimento padrão pode ser usado aqui: determinar o menor valor de algum critério de informação a partir de uma defasagem máxima e verificar se a autocorrelação dos resíduos é nula para todas as variáveis.

O problema maior é determinar o modelo de cointegração, pois há cinco possibilidades. As possibilidades relacionam-se à existência de constante e tendência no nível do vetor X_t e no vetor de cointegração. Formalmente, os casos

são listados a seguir e, a partir da definição de qual deles cabe usar, definem-se os valores críticos a consultar.

Caso 1 Ausência de intercepto e tendência tanto no vetor de cointegração quanto no nível de X_t . Nesse caso, tem-se:

$$\Phi X_{t-1} + \delta' d_t = \alpha \beta' X_{t-1}.$$

Caso 2 Há intercepto apenas no vetor de cointegração. Nesse caso, tem-se⁵:

$$\Phi X_{t-1} + \delta' d_t = \alpha (\beta' X_{t-1} + \mu_0),$$

em que μ_0 é um vetor $r \times 1$ de constantes.

Caso 3 Há intercepto no vetor de cointegração e tendência linear no nível de X_t . Nesse caso, tem-se:

$$\Phi X_{t-1} + \delta' d_t = \alpha (\beta' X_{t-1} + \mu_0) + \delta_0,$$

em que δ_0 é um vetor $n \times 1$ de constantes.

Caso 4 Há intercepto no vetor de cointegração e tendência linear no vetor de cointegração e no nível. Nesse caso, tem-se:

$$\Phi X_{t-1} + \delta' d_t = \alpha [\beta' X_{t-1} + \mu_0 + \mu_1(t-1)] + \delta_0,$$

em que μ_1 é um vetor $r \times 1$ de constantes.

Caso 5 Há intercepto e tendência linear no vetor de cointegração e tendência quadrática no nível. Então, tem-se:

$$\Phi X_{t-1} + \delta' d_t = \alpha [\beta' X_{t-1} + \mu_0 + \mu_1(t-1)] + \delta_0 + \delta_1 t,$$

em que δ_1 é um vetor $n \times 1$ de constantes.

Especificando o modelo em sua formulação completa, tem-se, portanto:

$$\Delta X_t = \alpha \{ \beta' [X_{t-1} + \mu_0 + \mu_1(t-1)] \} + (\delta_0 + \delta_1 t) + \sum_{i=1}^{p-1} \Lambda_i \Delta X_{t-i} + e_t.$$

⁵ A dimensão de intercepto e tendência é compatível com a dimensão dos vetores de cointegração.

Com essa especificação, os seguintes modelos são possíveis:

1. Sem termos determinísticos: $\mu_0 = \mu_1 = \delta_0 = \delta_1 = 0$;
2. Constante dentro do vetor de cointegração: $\mu_0 \neq 0, \mu_1 = \delta_0 = \delta_1 = 0$;
3. Constante dentro e fora do vetor de cointegração: $\mu_0, \delta_0 \neq 0, \mu_1 = \delta_1 = 0$;
4. Constante dentro e fora do vetor de cointegração e tendência dentro do vetor: $\delta_0, \mu_1, \mu_2 \neq 0, \delta_1 = 0$;
5. Constante dentro e fora do vetor de cointegração e tendência dentro e fora do vetor: $\delta_0, \delta_1, \mu_1, \mu_2 \neq 0$.

O leitor poderá perguntar-se sobre alguns casos que foram deixados de fora desses mencionados. O primeiro é o caso em que não há intercepto no vetor de cointegração, mas tendência linear no nível. Nessa situação, é preciso observar que os elementos de X_t têm todos uma tendência linear, de modo que a combinação entre seus elementos deverá resultar em constante, ou mesmo não eliminar a tendência linear, como ocorre no caso 3. Argumento semelhante poderá ser invocado se o leitor se perguntar se não é possível ter tendência quadrática no nível e apenas constante no vetor de cointegração. Finalmente, é preciso observar que os casos abarcam todas as possibilidades de regressores para que não haja omissão de variável relevante. Assim, estimar um modelo com mais regressores do que, de fato, é necessário não constitui problema.

Como exemplo, suponha três variáveis, x , y e z , em que haja dois vetores de cointegração, quando propriamente normalizados pela matriz F . A representação disso seria:

$$\begin{aligned}\Delta x_t &= \alpha_{x1} (x_{t-1} - bz_{t-1}) + \alpha_{x2} (y_{t-1} - dz_{t-1}) \\ &\quad + \sum_{i=1}^m [\omega_{xi} \Delta x_{t-i} + \psi_{xi} \Delta y_{t-i} + \delta_{xi} \Delta z_{t-i}] + \epsilon_t; \\ \Delta y_t &= \alpha_{y1} (x_{t-1} - bz_{t-1}) + \alpha_{y2} (y_{t-1} - dz_{t-1}) \\ &\quad + \sum_{i=1}^m [\omega_{yi} \Delta x_{t-i} + \psi_{yi} \Delta y_{t-i} + \delta_{yi} \Delta z_{t-i}] + v_t; \\ \Delta z_t &= \alpha_{z1} (x_{t-1} - bz_{t-1}) + \alpha_{z2} (y_{t-1} - dz_{t-1}) \\ &\quad + \sum_{i=1}^m [\omega_{zi} \Delta x_{t-i} + \psi_{zi} \Delta y_{t-i} + \delta_{zi} \Delta z_{t-i}] + \xi_t.\end{aligned}$$

A ordem de defasagem é determinada pela última defasagem, que faz com que as autocorrelações residuais sejam nulas pelo teste de Ljung-Box.

Nota 7.4 Não obstante, é possível calcular o posto da matriz Φ , o que nos permite saber se os elementos do vetor X_t são cointegrados ou não, lembrando que existem, no máximo, $n - 1$ vetores de cointegração.

Nota 7.5

- a) $\text{posto}(\Phi) = n \implies X_t \sim I(0)$;
- b) $\text{posto}(\Phi) = 0$, então as séries não cointegram;
- c) $0 < \text{posto}(\Phi) = r < n$, então há r vetores de cointegração.

Nota 7.6 Como β é superconsistente, podem-se fazer normalmente inferências sobre a matriz α .

Veja que, com posto igual a zero, inexistirá qualquer vetor que torne a combinação linear das coordenadas de X_t estacionária. Por outro lado, quando o posto é pleno, todas as combinações lineares possíveis pela matriz resultam em uma série estacionária. Se isso é sempre possível, é porque todas as séries são estacionárias.

É importante notar o seguinte: se ocorre o caso *a*, pode-se estimar $\Phi(L)X_t = e_t$ sem problemas, por ser o caso VAR convencional; se ocorre o caso *b*, é preciso estimar o sistema nas primeiras diferenças. Se ocorre o caso *c*, não é permitido estimar o modelo na primeira diferença, pois estaria faltando um regressor, o termo $\beta'X_{t-1}$. Estimar o modelo no nível não implica qualquer erro, mas haveria uma perda de eficiência, porque algumas restrições são omitidas. Daí a importância de definir, em um modelo VAR, se as variáveis cointegram ou não.

Nota 7.7 Quando os coeficientes da matriz α são zeros, a variável explicada é dita fracamente exógena. Assim, é preciso reinterpretar a condição de causalidade em sistemas cointegrados. Em um sistema cointegrado, $\{y_t\}$ não Granger causa $\{x_t\}$ se os valores defasados Δy_{t-i} não explicam Δx_t e se x_t não responde aos desvios de equilíbrio de longo prazo. Portanto, x_t é fracamente exógena.

Nota 7.8 O conceito de causalidade de Granger não deve ser usado para testar relações econômicas, mas tão-somente para melhorar a previsão.

É interessante observar que cointegração não significa Granger-causalidade, diferentemente do que afirmam Lütkepohl e Krätzig (2004, p. 150). Conceitualmente, cointegração é relacionada ao equilíbrio de longo prazo entre as

variáveis, ao passo que Granger-causalidade é relacionada à previsão de curto prazo. Ambos os conceitos são considerados na especificação VECM e, por isso, os autores fazem essa afirmação sob o argumento de que X_{t-1} contém variáveis defasadas que explicam ΔX_t . Portanto, haveria causalidade, pois variáveis defasadas ajudam a prever variáveis no presente. O problema é usar esse conceito para dizer que uma variável causa outra em sentido econômico, pois, havendo omissão de variáveis, tal afirmação seria economicamente espúria. A seguir, há um exemplo de causalidade de Granger que é espúria, do ponto de vista econômico. O leitor mais interessado pode consultar Maddala e Kim (1998, Seção 5.9.2) e suas referências, para mais detalhes.

Considere o seguinte modelo, em que z_t é um passeio aleatório:

$$y_t = \phi_y y_{t-1} + z_t + \epsilon_t, \epsilon_t \sim i.i.d. (0, \sigma_\epsilon^2), |\phi_y| < 1;$$

$$x_t = \phi_x x_{t-1} + z_t + v_t, v_t \sim i.i.d. (0, \sigma_v^2), |\phi_x| < 1;$$

$$z_t = z_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d. (0, \sigma_\epsilon^2).$$

Estimando o modelo apenas com as duas primeiras equações, pode-se concluir corretamente que há Granger-causalidade. De fato, como $z_t \sim I(1)$, o mesmo ocorre com x_t e y_t , por construção. Simulando uma série com 300 observações, conforme a seguir, será verificado que as séries x e y são cointegradas e que uma delas causa a outra, no sentido de previsão dado por Granger.

O sentido econômico de causalidade seria espúrio nesse contexto, pois levaria à conclusão falsa de que uma variável causa a outra, quando, na verdade, ambas têm um componente em comum que as causa simultaneamente. O fato de esse

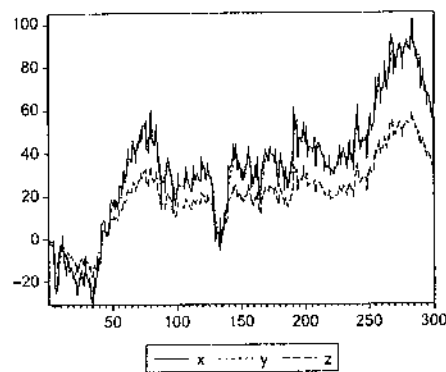


Figura 7.3 Granger-causalidade espúria: $y_t = 0,4y_{t-1} + z_t + \epsilon_t, \epsilon_t \sim i.i.d. (0, 1)$; $x_t = 0,4x_{t-1} + z_t + v_t, v_t \sim i.i.d. (0, 4)$; $z_t = z_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d. (0, 9)$.

componente não ter sido considerado no modelo econométrico leva à falsa conclusão de causalidade econômica⁶.

O teste de cointegração de Johansen pela estatística do traço rejeita a nula a um nível de significância inferior a 1%. O teste de causalidade de Granger com seis defasagens não rejeita a nula de que y_t não Granger causa x_t com significância de 87%. Porém, rejeita a nula de que x_t não Granger causa y_t com significância de 2,1%.

Exemplo 7.3

(O mercado de pão francês) Muito se tem dito sobre a taxa de câmbio e outras variáveis financeiras, porém praticamente nada sobre o nosso pão francês de cada dia. Esse exemplo se propõe a preencher essa lacuna e aproximar a teoria econômica do nosso cotidiano matinal.

Sejam p_1 o preço do trigo percebido pelo produtor, p_2 o preço da farinha de trigo na região metropolitana de São Paulo e p_3 o preço do pão francês nessa mesma região. Propõe-se o seguinte modelo para essas variáveis, impondo uma relação de longo prazo entre as séries, com o objetivo de procurar alguma relação de repasse dos preços na cadeia produtiva do pão francês:

$$\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{3,t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{3,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} [(p_3 - \beta_{10} - \beta_{11}p_1 - \beta_{12}p_2)_{t-1}] + \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} d_{1,t} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}$$

Exemplo 7.4

Veja aqui uma cadeia produtiva de um mercado relativamente competitivo, em que se assume que os custos são repassados ao consumidor.

Pode-se dividir o modelo em duas partes essenciais: a dinâmica de curto prazo e a dinâmica de longo prazo. A dinâmica de longo prazo está representada pela matriz 3×1 com α 's. Nesse modelo, tem-se uma relação de cointegração:

⁶ Essa formulação econométrica geralmente é resultado da falta de modelo econômico que defina as equações do modelo.

a diferença entre o preço do pão francês, da farinha de trigo e do trigo. Os parâmetros α s captam como os desvios de longo prazo são corrigidos. Na dinâmica de curto prazo, existem os parâmetros ϕ s. A interpretação destes é a mesma do caso VAR tradicional, e eles captam o comportamento auto-regressivo das séries. Nesse caso, o modelo inclui apenas uma defasagem das variáveis.

O modelo também inclui um componente determinístico, por meio da *dummy* para o período de câmbio flutuante (a partir de janeiro de 1999). Nesse caso, procura-se ver se há alguma alteração das séries quando a volatilidade da taxa de câmbio aumenta. Note que o trigo é uma commodity e um dos principais produtos da pauta importadora brasileira. Assim, poderia ocorrer que, em momentos de maior incerteza (volatilidade), o preço do trigo ao consumidor se ajustasse com maior velocidade para evitar prejuízos inesperados (aqui também é factível haver um movimento assimétrico de reajuste, que determina elevadas margens de lucro).

O primeiro passo para a estimação do modelo é dado pela realização do teste de Johansen para verificar a existência ou não de cointegração entre as séries. O resultado do teste do traço não rejeita a hipótese nula a 5% de existência de pelo menos dois vetores de cointegração para as séries:

	<i>Autovalor</i>	<i>Traço</i>	<i>Valor crítico a 5%</i>	<i>Prob.</i>
Nenhum*	0,163	47,021	35,192	0,002
1 Vetor*	0,105	21,505	20,261	0,033
2 Vetores	0,038	5,621	9,164	0,222

Os resultados do teste de máximo autovalor indicam a existência de pelo menos um vetor de cointegração a 5%. Então, pode-se afirmar que existe uma relação de longo prazo entre as variáveis.

	<i>Autovalor</i>	<i>Máximo autovalor</i>	<i>Valor crítico a 5%</i>	<i>Prob.</i>
Nenhum*	0,163	25,516	22,299	0,017
1 Vetor	0,105	15,883	15,892	0,050
2 Vetores	0,038	5,621	9,164	0,222

Com os resultados apresentados, pode-se confirmar a existência de pelo menos um vetor de correção de erros e estimar o modelo proposto. Os resultados

são mostrados a seguir (valores significantes a 5% marcados com *):

$$\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{3,t} \end{pmatrix} = \begin{pmatrix} 0,416^* & -0,071 & 0,298 \\ 0,147^* & 0,473^* & 0,117 \\ 0,087^* & 0,262^* & 0,012 \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{3,t-1} \end{pmatrix} \\ + \begin{pmatrix} 0,038^* \\ -0,041 \\ 0,080^* \end{pmatrix} [(p_3 - 1,141^* - 0,926^* p_1 + 2,104^* p_2)_{t-1}] \\ + \begin{pmatrix} 0,001 \\ 0,009^* \\ 0,007^* \end{pmatrix} d_{1,t} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix}$$

Os mercados de trigo e pão francês estão interligados, conforme apontam os coeficientes de ajuste dos desequilíbrios de longo prazo. Choques de longo prazo nos preços desses dois mercados são incorporados à dinâmica (como elevações das cotações internacionais do produto percebidas como duradouras).

Na dinâmica de curto prazo, verifica-se que o preço ao produtor do trigo não é influenciado pelo preço dos outros mercados (apenas 0,416 é significativo, ou seja, um AR (1)). A dinâmica do preço da farinha de trigo responde a oscilações do preço do trigo, porém não às do pão francês (assim como visto anteriormente, isso poderia indicar que as pressões de demanda não afetam os preços no sentido "contrário"). Por fim, o preço do pão responde a oscilações no preço da farinha de trigo (0,23) e não possui um componente auto-regressivo. Todos os sinais são positivos, conforme esperado. É interessante notar que a dummy do período de oscilação cambial não é significativa para o preço do trigo, ao contrário do que ocorreu com a farinha (0,09) e o pão francês (0,07).

7.3.1 Teste de Hipóteses

O aspecto mais interessante do procedimento de Johansen é poder testar formas restritas do vetor de cointegração. Isso é possível porque, se há r vetores de cointegração, apenas essas r combinações lineares de variáveis são estacionárias. Assim, reestimando os modelos impondo as restrições desejadas e se elas são "folgadas" (ou não são "justas"), então o número de vetores de cointegração não deve diminuir.

Assim, estime os modelos restrito e não restrito. Obtenha os autovalores de cada modelo, respectivamente ordenados como $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_n$ e $\hat{\lambda}_1^* > \hat{\lambda}_2^* > \dots > \hat{\lambda}_n^*$. A estatística do teste será dada por:

$$2J = -T \sum_{i=r+1}^n [\ln(1 - \hat{\lambda}_i^*) - \ln(1 - \hat{\lambda}_i)] \xrightarrow{d} \chi_{n-r}^2.$$

A intuição é que autovalores de cada uma das regressões devem ser próximos se a restrição imposta for verdadeira. Dessa forma, baixos valores de J implicam a rejeição da hipótese nula representada pela imposição de restrições sobre as matrizes α e β .

7.3.2 Estimação de Φ por Máxima Verossimilhança*

Esta seção, mais técnica, fornece alguns detalhes sobre como obter o posto da matriz Φ , baseada na exposição de Patterson (2000), Maddala e Kim (1998) e em suas referências. Considere, assim, o modelo VAR completo:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \delta' d_t + e_t,$$

em que $e_t \sim i.i.N(0; \Sigma)$.

Esse modelo pode ser escrito na forma de vetor de correção de erros, e o problema é estimar o modelo por máxima verossimilhança sujeito à restrição de posto incompleto: $\Phi = \alpha\beta'$:

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{p-1} \Lambda_i \Delta X_{t-i} + \delta' d_t + e_t.$$

Patterson sugere renomear algumas variáveis para tornar mais fácil a exposição. Imagine, inicialmente, para facilitar a notação, uma amostra com $T + p$ observações começando em $-p + 1$. Os p primeiros valores são usados como valores iniciais do modelo. Assim, considere o vetor de variável explicada $Y_{0t} \equiv \Delta X_t$, o vetor $Y_{1t} \equiv X_{t-1}$, a matriz de variáveis explicativas $Y_{2t} \equiv [\Delta X_{t-1} \quad \Delta X_{t-2} \quad \cdots \quad \Delta X_{t-p+1} \quad d_t]'$ e a matriz de coeficientes $\Lambda = [\Lambda_1 \quad \Lambda_2 \quad \cdots \quad \Lambda_{p-1} \quad \delta]'$. Portanto, o modelo VECM pode ser reescrito simplifadamente:

$$Y_{0t} = \alpha\beta' Y_{1t} + \Lambda Y_{2t} + e_t. \quad (15)$$

Ou seja, o problema é maximizar a função de verossimilhança, em que há restrições não-lineares de parâmetros, dadas por:

$$\ln L(\alpha, \beta, \Lambda, \Sigma) = -\frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T e_t' \Sigma^{-1} e_t.$$

As condições de primeira ordem desse modelo com relação a Λ são:

$$0 = \sum_{t=1}^T (Y_{0t} - \alpha\beta' Y_{1t} - \hat{\Lambda} Y_{2t}) Y_{2t}' \implies \sum_{t=1}^T Y_{0t} Y_{2t}' = \alpha\beta' \sum_{t=1}^T Y_{1t} Y_{2t}' + \hat{\Lambda} \sum_{t=1}^T Y_{2t} Y_{2t}',$$

em que $\hat{\Lambda}$ é uma estimativa de Λ .

Renomeando as variáveis, para facilitar a exposição, como

$$\Pi_{ij} = \frac{\sum_{t=1}^T Y_{it} Y_{jt}'}{T},$$

a equação anterior pode ser reescrita como:

$$\Pi_{02} = \alpha \beta' \Pi_{12} + \hat{\Lambda} \Pi_{22}.$$

Portanto, pode-se obter $\hat{\Lambda}$ pós-multiplicando ambos os lados da equação anterior por Π_{22}^{-1} e rearranjando os termos adequadamente:

$$\hat{\Lambda} = \Pi_{02} \Pi_{22}^{-1} - \alpha \beta' \Pi_{12} \Pi_{22}^{-1}.$$

Por razões que ficarão mais claras logo adiante, considere ainda as seguintes regressões lineares auxiliares:

1. Y_{0t} contra Y_{2t} , isto é, regrida:

$$Y_{0t} = BY_{2t} + r_{0t} \implies \hat{B} = (Y_{0t} Y_{2t}') (Y_{2t} Y_{2t}')^{-1} = \Pi_{02} \Pi_{22}^{-1}$$

e obtenha os resíduos estimados $\hat{r}_{0t} = Y_{0t} - \Pi_{02} \Pi_{22}^{-1} Y_{2t}$.

2. Y_{1t} contra Y_{2t} e obtenha os resíduos estimados $\hat{r}_{1t} = Y_{1t} - \Pi_{12} \Pi_{22}^{-1} Y_{2t}$.

Observe da equação 15 que:

$$\begin{aligned} \hat{e}_t &= Y_{0t} - \alpha \beta' Y_{1t} - \left(\Pi_{02} \Pi_{22}^{-1} - \alpha \beta' \Pi_{12} \Pi_{22}^{-1} \right) Y_{2t} \\ &= Y_{0t} - \Pi_{02} \Pi_{22}^{-1} Y_{2t} - \alpha \beta' \left(Y_{1t} - \Pi_{12} \Pi_{22}^{-1} Y_{2t} \right) \\ &= \hat{r}_{0t} - \alpha \beta' \hat{r}_{1t}. \end{aligned}$$

Esse resultado é importante para mostrar que não é necessário calcular $\hat{\Lambda}$ para obter \hat{e}_t . Desse modo, as quantidades \hat{r}_{0t} e \hat{r}_{1t} podem ser obtidas pelas regressões auxiliares e, assim, podem-se estimar α e β . A partir desse resultado, o problema se torna maximizar a seguinte função de verossimilhança concentrada:

$$\ln L(\alpha, \beta, \Sigma) = -\frac{T}{2} \ln |\Sigma| - \frac{1}{2} \sum_{t=1}^T (\hat{r}_{0t} - \alpha \beta' \hat{r}_{1t})' \Sigma_t^{-1} (\hat{r}_{0t} - \alpha \beta' \hat{r}_{1t}).$$

Equivalentemente a maximizar essa função, pode-se empreender a regressão por mínimos quadrados ordinários entre \hat{r}_{0t} e \hat{r}_{1t} . Uma forma operacional de fazer isso é estimar α , para um dado β . Isto é, estimar:

$$\begin{aligned}\hat{r}_{0t} &= \alpha (\beta' \hat{r}_{1t}) + \hat{e}_t, \Rightarrow \\ \hat{\alpha}(\beta) &= \left[\left(\sum_{t=1}^T \hat{r}_{0t} \hat{r}_{1t}' \right) \beta \right] \left[\beta' \left(\sum_{t=1}^T \hat{r}_{1t} \hat{r}_{1t}' \right) \beta \right]^{-1} \\ &= S_{01} \beta (\beta' S_{11} \beta)^{-1},\end{aligned}$$

em que $S_{ij} := T^{-1} \sum_{t=1}^T \hat{r}_{it} \hat{r}_{jt}'$, $j = 0, 1$.

Agora, o problema é estimar a matriz de covariância a partir das quantidades obtidas até o momento. Isso pode ser feito com as estimativas anteriores, da seguinte forma, notando-se apenas que $S_{ij} = S_{ji}'$:

$$\begin{aligned}\hat{\Sigma}(\beta) &= \frac{\sum_{t=1}^T \hat{e}_t \hat{e}_t'}{T} - \frac{\sum_{t=1}^T [\hat{r}_{0t} - \alpha(\beta' \hat{r}_{1t})] [\hat{r}_{0t} - \alpha(\beta' \hat{r}_{1t})]'}{T} \\ &= \sum_{t=1}^T \frac{\hat{r}_{0t} \hat{r}_{0t}' - \hat{\alpha}(\beta' \hat{r}_{1t}) \hat{r}_{0t}' - \hat{r}_{0t} \hat{r}_{1t}' \hat{\alpha}' + \hat{\alpha}(\beta' \hat{r}_{1t}) \hat{r}_{1t}' \hat{\alpha}'}{T} \\ &= S_{00} - \hat{\alpha} \beta' S_{10} - S_{01} \beta \hat{\alpha}' + \hat{\alpha} \beta' S_{11} \beta \hat{\alpha}' \\ &= S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{01}' \\ &\quad + S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{11} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{01}' \\ \hat{\Sigma}(\beta) &= S_{00} - S_{01} \beta (\beta' S_{11} \beta)^{-1} \beta' S_{10}.\end{aligned}$$

De novo, o objetivo dessas contas foi o de desaparecer com $\hat{\alpha}(\beta)$, uma vez que, para obtê-lo, é preciso ter β . Entretanto, essa matriz ainda não foi estimada. Porém, partindo da última derivação, pode-se encontrar β a partir da maximização da verossimilhança em β da seguinte função:

$$L(\beta)^{-\frac{1}{2}} = |\hat{\Sigma}(\beta)|.$$

Para prosseguir, convém lembrar um resultado devido a Johansen (1991), pelo qual a matriz de covariância amostral de $[\hat{r}_{0t}' \quad \hat{r}_{1t}' \beta']'$, S , é dada por:

$$S = \begin{bmatrix} S_{00} & S_{01} \beta \\ \beta' S_{10} & \beta' S_{11} \beta \end{bmatrix}.$$

Trata-se de uma matriz quadrada com submatrizes também quadradas e não singulares. Assim, pelas regras 5 e 6 de Lütkepohl (2005, p. 660), se:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

então, se A_{11} e A_{22} são matrizes quadradas não singulares, vale:

$$\begin{aligned} |A| &= |A_{22}| |A_{11} - A_{12} A_{22}^{-1} A_{21}| = |A_{11}| |A_{22} - A_{21} A_{11}^{-1} A_{12}| \implies \\ |A_{11} - A_{12} A_{22}^{-1} A_{21}| &= \frac{|A_{11}| |A_{22} - A_{21} A_{11}^{-1} A_{12}|}{|A_{22}|}. \end{aligned}$$

Conseqüentemente, o problema de estimação é equivalente a maximizar:

$$\frac{|S_{00}| |\beta' S_{11} \beta - \beta' S_{10} S_{00}^{-1} S_{01} \beta|}{|\beta' S_{11} \beta|} = \frac{|S_{00}| |\beta' (S_{11} - S_{10} S_{00}^{-1} S_{01}) \beta|}{|\beta' S_{11} \beta|}.$$

Johansen (1995, lema A.8, p. 224) demonstra que a seguinte maximização

$$\max_X \frac{|X' (A_1 - A_2) X|}{|X' A_1 X|}$$

é equivalente a maximizar a raiz característica da equação $|A_2 - \lambda A_1|$. Assim, fazendo $A_1 = S_{11}$ e $A_2 = S_{10} S_{00}^{-1} S_{01}$, o problema se torna encontrar o autovalor de

$$\begin{aligned} |S_{10} S_{00}^{-1} S_{01} - S_{11} \lambda| &= 0 \quad \text{ou, equivalentemente,} \\ |S_{11}^{-1} S_{10} S_{00}^{-1} S_{01} - I \lambda| &= 0. \end{aligned}$$

Se os autovalores de A são λ_i , então os autovalores de $I - A$ são $1 - \lambda_i$. Logo, se λ_i é o autovalor que resolve a equação anterior, $1 - \lambda_i$, $i = 1, 2, \dots, n$ são os autovalores que resolvem $(I - S_{11}^{-1} S_{10} S_{00}^{-1} S_{01})$, isto é, resolvem o seguinte problema:

$$|(I - S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}) - \lambda I| = 0.$$

O valor do determinante de uma matriz é igual ao produto dos autovalores. Logo, é forçoso concluir que:

$$\begin{aligned} \prod_{i=1}^n (1 - \lambda_i) &= |I - S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}| = |S_{11}^{-1}| |S_{11} - S_{10} S_{00}^{-1} S_{01}| \\ &= \frac{|S_{11} - S_{10} S_{00}^{-1} S_{01}|}{|S_{11}|} = \frac{|S_{00} - S_{10} S_{11}^{-1} S_{01}|}{|S_{00}|}, \end{aligned}$$

em que a última igualdade foi obtida aplicando-se as regras 5 e 6 de Lütkepohl de novo. Portanto, $\max L(\beta)^{-\frac{2}{T}} = |\hat{\Sigma}(\beta)|$ será dado por:

$$L_{\max}^{-\frac{2}{T}} = |S_{00}| \prod_{i=1}^n (1 - \hat{\lambda}_i).$$

Os autovalores são, em seguida, ordenados do maior para o menor: $\lambda_1 > \lambda_2 > \dots > \lambda_n$. A cada autovalor é associado um autovetor correspondente às colunas da matriz β . Porém, se o posto da matriz Φ for r , então só os r primeiros autovetores corresponderão aos vetores de cointegração. Isto é, os autovetores associados aos r autovalores definirão a matriz $\hat{\beta} := [\hat{\beta}_1 \ \hat{\beta}_2 \ \dots \ \hat{\beta}_r]$, que é uma matriz $n \times r$.

O teste de razão de verossimilhança pode ser obtido estimando-se as funções restrita e não restrita:

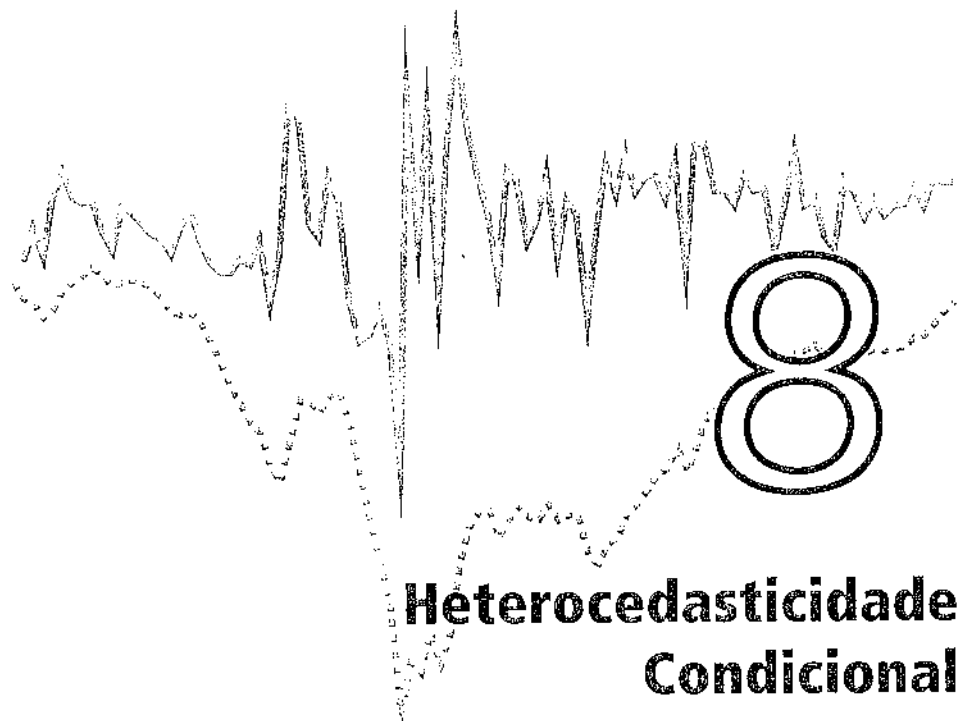
$$L_{\max}^{-\frac{2}{T}} = |S_{00}| \prod_{i=1}^r (1 - \hat{\lambda}_i);$$

$$L_{\max}^{-\frac{2}{T}} = |S_{00}| \prod_{i=1}^n (1 - \hat{\lambda}_i).$$

A estatística do teste é dada, então, por⁷:

$$\begin{aligned} \exp(\lambda_{lr}) &= \frac{|S_{00}| \prod_{i=1}^r (1 - \hat{\lambda}_i)}{|S_{00}| \prod_{i=1}^n (1 - \hat{\lambda}_i)} = \frac{1}{\prod_{i=r+1}^n (1 - \hat{\lambda}_i)} \implies \\ \lambda_{lr} &= -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i). \end{aligned}$$

⁷ Esse teste tende em distribuição para $\text{tr}\{(\int_0^1 dW(r)\tilde{W}(r))'[\int_0^1 \tilde{W}(s)\tilde{W}(s)'ds]^{-1} \int_0^1 \tilde{W}(r)dW(r)'\}$, em que $W(\cdot)$ é um processo de Wiener padrão de dimensão $n - r$ e \tilde{W} é um processo de Wiener sem tendência ou constante, isto é, *demeaned* ou *detrended*.



Heterocedasticidade Condicional

No início da década de 1980, assistiu-se ao surgimento dos modelos autorregressivos de heterocedasticidade condicional, ARCH (veja Engle, 1982). Esses modelos foram, posteriormente, generalizados por Bollerslev (1986), GARCH, originando vasta literatura sobre o assunto, ainda inesgotada, provavelmente.

A razão principal do surgimento desses modelos é que, antes disso, os modelos econométricos de séries de tempo, fossem os financeiros ou os macroeconômicos, enfatizavam apenas o primeiro momento condicional. As dependências temporais de ordem superior eram perturbações aleatórias, consideradas em seus momentos incondicionais. Essas dependências expressam a existência de aglomerações na série e a alternância de períodos de baixa volatilidade com períodos de alta volatilidade. A Figura 8.1 apresenta essas características.

Os modelos de heterocedasticidade condicional surgiram principalmente porque a importância do risco e da incerteza na teoria econômica moderna tornou-se proeminente e pelo fato de modelos como o CAPM de Sharpe (1964) e Lintner (1965) não funcionarem tão bem empiricamente (veja Fama e French, 2004). Talvez fosse necessário incluir momentos de ordem maior no modelo

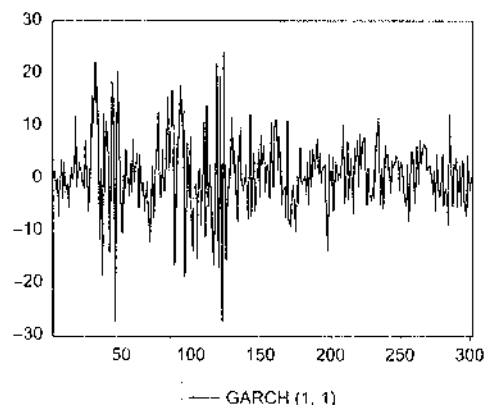


Figura 8.1 Modelo de simulação em que erros seguem GARCH (1, 1).

CAPM para aproximá-lo dos dados empíricos. Assim, desenvolveram-se técnicas que permitem a modelagem temporal de variâncias e covariâncias. De fato, os modelos de heterocedasticidade condicional, GARCH, fundamentam-se na estimação da variância condicional, em vez de considerá-la constante ao longo do tempo.

A contribuição principal desses modelos é a distinção de uso entre momentos de segunda ordem condicionais e não condicionais. Enquanto a matriz de covariância não condicional para as variáveis de interesse pode ser invariante no tempo, a matriz de covariância condicional depende de estados passados da natureza.

Empiricamente, observa-se que as séries financeiras não têm distribuição normal padrão em geral, dada a elevada probabilidade de eventos extremos. Então, os modelos GARCH teriam a capacidade de modelar esse fato estilizado.

A seguir, apresentam-se os principais modelos de heterocedasticidade condicional. Esforça-se para dar a intuição subjacente, e os exemplos explicam como podem ser utilizados empiricamente.

8.1

Modelos GARCH

8.1.1 ARCH (q)

Inicia-se descrevendo o primeiro modelo imaginado e complicando aos poucos; trata-se do modelo de Engle (1982), em que ϵ_t é um processo estocástico real em tempo discreto, condicional à informação em $t - 1$. Suponha que se queira estimar o modelo ARIMA ($p, 0, q$). Se o erro desse processo segue um processo ARCH,

então ele é definido da seguinte forma:

$$\epsilon_t = \sigma_t u_t, u_t \sim i.i.d. (0, 1);$$

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2.$$

É desnecessário que a distribuição do erro, u_t , seja normal, bastando ser idêntica e independentemente distribuída para que os momentos assintóticos da série sejam iguais aos da normal. O fato de $VAR(u_t) = 1$ não implica perda de generalidade¹. Qualquer outro número poderia ser usado, e o ajuste se daria nos coeficientes de σ_t^2 . Com isso, é fácil ver que a variância incondicional é dada por:

$$E(\epsilon_t^2) = E(\sigma_t^2 u_t^2) = E(\sigma_t^2) E(u_t^2) = \sigma_\epsilon^2 = \frac{\omega}{1 - \sum_{i=1}^q \alpha_i}.$$

Algumas restrições ao modelo são necessárias para garantir a estacionaridade e positividade da variância condicional. Em primeiro lugar, uma condição suficiente para que a variância condicional seja sempre positiva é que $\omega > 0$ e $\alpha_i \geq 0, \forall i, i = 1, 2, \dots, q$. Além disso, $\sum_{i=1}^q \alpha_i < 1$ é condição necessária para que a série seja estacionária. Isso significa que as raízes da polinomial $(1 - \sum_{i=1}^q \alpha_i L^i)$ devem estar fora do círculo unitário.

De modo semelhante ao dos modelos *ARIMA*, o modelo condicional no segundo momento pode ser assim expresso:

$$E(\sigma_t^2 | I_{t-1}) = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2.$$

8.1.2 GARCH (p,q)

Infelizmente, é fato que os modelos *ARCH* (q) precisam de muitos parâmetros para serem ajustados corretamente. Para tentar reduzir esse problema, Bollerslev (1986) sugere um modelo potencialmente mais parcimonioso, o modelo de heterocedasticidade condicional auto-regressivo generalizado, *GARCH*, no qual a equação da variância muda para:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2. \quad (16)$$

¹ Note que as características não condicionais de ϵ_t não se alteram, valendo todos os resultados anteriores discutidos ao longo deste trabalho.

Embora pareça que aqui há mais parâmetros que no ARCH, a verdade é que $q_{ARCH} > q_{GARCH} + p_{GARCH}$ em geral. De novo, uma condição suficiente para estacionaridade é que as raízes da polinomial da variância estejam fora do círculo unitário, como no caso dos modelos *ARIMA*.

O modelo GARCH (p, q) pode ser interpretado como um processo autorregressivo em ϵ_t^2 . Para ver isso, defina:

$$v_t = \epsilon_t^2 - \sigma_t^2 = (u_t^2 - 1) \sigma_t^2.$$

Verifica-se, sem muito esforço, que v_t tem média zero e não é autocorrelacionado. Prosseguindo com substituição na equação (16), tem-se que:

$$\epsilon_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j (\epsilon_{t-j}^2 - v_{t-j}) + v_t.$$

Reescrevendo:

$$\epsilon_t^2 = \omega + \sum_{i=1}^{\max[p,q]} (\alpha_i + \beta_i) \epsilon_{t-i}^2 - \sum_{j=1}^p \beta_j v_{t-j} + v_t.$$

A última equação lembra um modelo *ARMA* $[\max(p, q), p]$, sugerindo que a identificação do modelo GARCH pode seguir a mesma metodologia proposta por Box-Jenkins. Essa metodologia vale para a identificação de modelos GARCH apenas se se trata de fato de um modelo GARCH, em vez de um modelo assimétrico como nos exemplos a seguir, e isso nem sempre é fácil de ser verificado.

As restrições impostas sobre os coeficientes dos modelos GARCH, de que eles devem ser todos maiores que zero, podem ser relaxadas. Na verdade, as condições gerais de estacionaridade que devem ser impostas sobre os coeficientes não foram derivadas ainda. Para casos específicos, no entanto, existem alguns resultados importantes (veja Nelson e Cao, 1992).

8.1.3 GARCH-M

As teorias financeiras mostram uma explícita relação entre média e risco, ou variância. Com os modelos GARCH, tornou-se mais fácil esse tipo de especificação, bastando acrescentar à equação da média a variável explicativa ou seu quadrado, da seguinte forma:

$$y_t = \frac{\theta(L)}{\phi(L)} \epsilon_t + \delta \frac{\sigma_t^{1-a}}{1-a},$$

em que:

$$\sigma_t^a = f(\epsilon_{t-j}, \sigma_{t-i}), i, j = 1, 2, \dots, \max[p, q]; \quad a = -1, 0 \quad \text{ou} \quad 1.$$

Se $a = -1$, a variância explica retornos. Se $a = 0$, é o desvio-padrão que explica retornos. Caso $a = 1$, é o logaritmo do desvio-padrão que explica retornos. Se y_t representa os retornos de uma série financeira, intuitivamente espera-se que δ seja positivo.

Antes dos modelos GARCH, a relação entre risco e retorno era obtida usando dados mensais para calcular a média e a variância anual. Em seguida, procedia-se a uma regressão de média contra variância. Com os modelos GARCH-M, deixou de ser necessário adotar esse procedimento, pelo qual necessariamente se perdiam as observações mensais incorporadas às médias anuais. Para mais detalhes, veja Engle, Lilien e Robins (1987).

8.1.4 TGARCH (p, q)

Os modelos anteriores supõem que os impactos dos choques aleatórios são simétricos, mas a evidência empírica contradiz largamente essa idéia. Por exemplo, choques no retorno de ações são assimétricos, pois os negativos são seguidos por maiores aumentos na volatilidade do que os positivos. Além disso, os parâmetros do modelo GARCH têm de ser positivos para que a estacionaridade da série seja mantida. Assim, Zakoïan (1994) propôs o modelo de heterocedasticidade condicional auto-regressivo generalizado com limiar, TGARCH, da seguinte forma:

$$\sigma_t^a = \omega + \sum_{j=1}^p \beta_j \sigma_{t-j}^a + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^a + \sum_{k=1}^r \gamma_k d_{(\varepsilon_{t-k} \leq 0)} |\varepsilon_{t-k}|^a,$$

em que $d_{(\cdot)}$ é a variável *dummy* igual a 1 se o erro satisfaz a condição imposta entre parênteses, e 0, caso contrário.

O coeficiente a representa o expoente da variância. Para Zakoïan, o coeficiente é 1; para Glosten, Jaganathan e Runkle (1993), é 2. Entretanto, o coeficiente pode ser livre, o que vem a ser uma forma melhor de ajustar o modelo, pois o torna extremamente flexível, como é discutido particularmente no modelo PGARCH a seguir.

O modelo implica um aumento de volatilidade quando notícias ruins, representadas por $\varepsilon_{t-k} < 0$, são acompanhadas por um coeficiente positivo, isto é, por $\gamma_k > 0$. É possível observar que o modelo GARCH é um caso especial do TGARCH, bastando, para isso, impor $\gamma_k = 0, \forall k$.

8.1.5 PGARCH

Combinando as idéias dos modelos apresentados, Ding, Granger e Engle (1993) propuseram o modelo assimétrico potência GARCH, cuja assimetria é captada pelo coeficiente γ . O interessante desse modelo também é capturar um processo

de memória longa por meio da liberdade dada ao coeficiente a :²

$$\sigma_t^a = \omega + \sum_{j=1}^p \beta_j \sigma_{t-j}^a + \sum_{i=1}^q \alpha_i (|\epsilon_{t-i}| - \gamma_i \epsilon_{t-i})^a,$$

em que

$$a > 0; \quad |\gamma_i| < 1, i = 1, 2, \dots, r; \quad \gamma_i = 0, r < i \leq p.$$

O expoente da volatilidade condicional, σ_t , pode ser estimado em vez de imposto. O segundo ponto de interesse é a captura da assimetria por meio do parâmetro γ . Um caso particular desse modelo é fixar $a = 1$ e $\gamma_i = 0, \forall i$. Assim, estima-se o desvio-padrão em vez da variância, como fizeram Taylor (1986) e Schwert (1989). Outro caso particular deve-se a Bera e Higgins (1995), que fizeram $\gamma_i = 0, \forall i$, mas deixam o expoente livre.

8.1.6 EGARCH (p, q)

Nelson (1991), também buscando superar as dificuldades do modelo GARCH simétrico, propôs o modelo de heterocedasticidade condicional auto-regressivo generalizado exponencial, EGARCH, que permite a absorção de impactos assimétricos com uma vantagem adicional: a possibilidade de que alguns coeficientes sejam negativos. O modelo é assim especificado:

$$\ln \sigma_t^2 = \omega + \sum_{j=1}^p \beta_j \ln \sigma_{t-j}^2 + \sum_{i=1}^q \alpha_i \left| \frac{\epsilon_{t-i}}{\sigma_{t-i}} - E \left(\frac{\epsilon_{t-i}}{\sigma_{t-i}} \right) \right| + \sum_{k=1}^r \gamma_k \frac{\epsilon_{t-i}}{\sigma_{t-i}}.$$

A especificação logarítmica impede que a variância seja negativa, portanto os parâmetros do modelo podem ser negativos. Os coeficientes γ ajustam a assimetria dos efeitos. Se $\gamma_k = 0$, para todo k , então um choque positivo tem o mesmo efeito na volatilidade que um choque negativo de mesma magnitude. Nesse caso, o impacto é simétrico. Se houver algum $\gamma_k \neq 0$, os impactos serão assimétricos.

Pode-se testar a presença do chamado efeito alavancagem se $\gamma_k < 0$, significando que choques negativos aumentam a volatilidade do modelo mais que choques positivos.

É preciso que as raízes da polinomial $(1 - \sum_{j=1}^p \beta_j L^j)$ estejam fora do círculo unitário, a fim de que a variância seja estacionária. A estacionaridade estrita é dada, segundo Nelson, se $\sum_{i=1}^q \alpha_i^2 < \infty$.

² A apresentação de modelos de memória longa foge ao escopo desta obra.

Para estimar o EGARCH, Nelson sugere usar a função de distribuição erro generalizado, GED:

$$f(u_t) = \frac{\nu \exp\left(-\frac{1}{2} \left|\frac{\epsilon_t}{\lambda}\right|^\nu\right)}{\lambda 2^{\frac{\nu+1}{\nu}} \Gamma\left(\frac{1}{\nu}\right)},$$

em que

$\Gamma(\cdot)$ é a função gama;

$\lambda \equiv \left[\frac{2^{-(2/\nu)} \Gamma(1/\nu)}{\Gamma(3/\nu)}\right]$ é uma constante;

$0 < \nu \leq \infty$ é um parâmetro positivo determinando o grau de curtose da distribuição;

$$-\infty < \epsilon_t < \infty.$$

Em particular, quando $\nu = 2$ e $\lambda = 1$, tem-se a distribuição normal. Se $\nu < 2$, a distribuição tem caudas gordas.

8.1.7 EWMA

O modelo EWMA (*exponential weighted moving average*) é muito utilizado em mercados financeiros e é especificado da seguinte maneira:

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) \epsilon_{t-1}^2.$$

Trata-se do modelo GARCH integrado, IGARCH, caracterizado pelo fato de que a soma dos parâmetros é unitária, isto é, $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1$.

Com dados diários, pode-se escrever que os retornos contínuos, r_t , são:

$$r_t = \mu + \epsilon_t.$$

Se os mercados são eficientes, dificilmente seria encontrada autocorrelação no modelo. Na prática, é possível estimar facilmente esse modelo, via Excel ou restringindo o modelo para ser um IGARCH, em que $\omega = 0$.

8.2

Testes para Detecção de Modelos GARCH

Nesta seção, detalham-se alguns testes para identificação de modelos GARCH. Serão abordados apenas os testes mais comuns, mas há outras alternativas em Bollerslev, Engle e Nelson (1994).

8.2.1 Teste FAC e FACP

Verificou-se que o modelo GARCH (p, q) assemelha-se ao modelo ARMA $(\max(p, q), q)$. Assim, as funções de autocorrelação, FAC, e de autocorrelação parcial, FACP, devem sugerir se a série é heterocedástica, da mesma maneira que dão uma idéia das ordens p e q de um modelo ARMA. Deve-se, pois, proceder obtendo primeiro os quadrados dos erros estimados pela regressão do modelo estabelecido, como se a variância condicional fosse constante:

$$\bar{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2,$$

em que T é o número de resíduos.

Em seguida, calcula-se a FAC amostral para o quadrado dos resíduos, sem esquecer de representá-la graficamente. A função é dada por:

$$\hat{\rho}_j = \frac{\frac{\sum_{t=j+1}^T (\hat{\epsilon}_t^2 - \bar{\sigma}^2) (\hat{\epsilon}_{t-j}^2 - \bar{\sigma}^2)}{T}}{\frac{\sum_{t=1}^T (\hat{\epsilon}_t^2 - \bar{\sigma}^2)}{T}}, \quad j = 1, 2, \dots$$

Pode-se aproximar o desvio-padrão de $\hat{\rho}_j$ em grandes amostras por $T^{-0.5}$. Diz-se que algum modelo ARCH está presente se os valores de $\hat{\rho}_j$ forem estatisticamente diferentes de zero. A função de autocorrelação parcial é obtida pondo em um gráfico o coeficiente estimado, $\phi_{s,s}$, contra a defasagem s , a partir da seguinte equação:

$$\hat{\epsilon}_t^2 = \phi_s + \phi_{s,1} \hat{\epsilon}_{t-1}^2 + \phi_{s,2} \hat{\epsilon}_{t-2}^2 + \dots + \phi_{s,s} \hat{\epsilon}_{t-s}^2 + u_t, \quad s = 1, 2, \dots, q.$$

De modo semelhante, o desvio-padrão pode ser aproximado por $T^{-0.5}$.

Diferentemente dos modelos ARMA, a FAC nos dá a ordem máxima da auto-regressão do GARCH, representado pelo termo $\hat{\epsilon}_{t-j}^2$, e a FACP dá a ordem p das “médias móveis” do GARCH do termo σ_{t-j}^2 . Se $p < \max\{p, q\}$, sabem-se exatamente as ordens p e q do modelo. Do contrário, encontra-se a ordem q por tentativa e erro. As dificuldades de estimação são da mesma natureza que nos modelos ARMA (p, q) convencionais.

Se os modelos forem assimétricos na variância, os testes ajudam a identificar a existência de heterocedasticidade condicional, sem, contudo, servir para definir as ordens p e q .

8.2.2 Teste de Ljung-Box

A estatística de Ljung-Box, Q , pode ser usada para testar a presença de heterocedasticidade condicional. Sua hipótese nula é que a soma das autocorrelações é estatisticamente diferente de zero. Isto é:

$$H_0 : \sum_{j=1}^n \hat{\rho}_j = 0 \times H_A : \hat{\rho}_j \neq 0, \text{ para algum } j,$$

em que

$$Q = T(T+2) \sum_{j=1}^n \frac{\hat{\rho}_j^2}{T-j} \xrightarrow{d} \chi_n^2.$$

8.2.3 Teste Multiplicador de Lagrange ou de Engle

O teste multiplicador de Lagrange, LM, é bastante tradicional (veja White (1980), Bollerslev, Engle e Nelson (1994) e Johnston e Dinardo (1997), para outros detalhes), inicialmente proposto por Engle (1982) para o processo ARCH e, depois, estendido por Lee (1991) para o GARCH. Define-se o teste como:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0 \times H_A : \alpha_i \neq 0, \text{ para algum } i = 1, 2, \dots, q,$$

em que

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{\varepsilon}_{t-1}^2 + \alpha_2 \hat{\varepsilon}_{t-2}^2 + \dots + \alpha_q \hat{\varepsilon}_{t-q}^2 + u_t,$$

cujas estatística é dada por

$$TR^2 \xrightarrow{d} \chi_q^2,$$

em que R^2 representa o coeficiente de determinação da equação estimada.

A regressão terá pouco poder explicativo, se os coeficientes são conjuntamente próximos a zero, de forma que o R^2 será baixo. Portanto, se a hipótese nula não é rejeitada, não há presença de modelos ARCH. Rejeitando-se a nula, os testes, entretanto, não determinam que tipo de modelo ARCH deverá ser usado. Eles apenas indicam se a covariância condicional é, ou não, constante. De fato, mesmo em modelos lineares, a seleção do modelo apropriado não é trivial, particularmente quando se consideram as inúmeras possibilidades que os modelos ARCH oferecem.

Os testes são sensíveis a especificações incorretas na equação da média. Logo, esses testes podem indicar erroneamente a presença de modelos ARCH.

8.3

Identificação de Modelos GARCH

Embora possa ser impossível definir de que tipo de modelo GARCH se trata, existem alguns testes para verificar se o modelo GARCH é simétrico ou assimétrico (Hegerud, 1997). Uma vez que existem inúmeros modelos assimétricos, ainda não se consegue especificar as defasagens do modelo, permanecendo a dificuldade de especificação.

Uma alternativa usual na literatura é escolher o modelo utilizando o critério de Akaike, Schwarz ou Hannan-Quinn, não obstante suas propriedades estatísticas sejam ainda desconhecidas na presença de modelos GARCH. Mas, como o critério de informação (de qualquer tipo) é muito fácil de ser usado, é largamente recomendado. Para outras possibilidades veja Bueno (2002).

8.4

Estimação de Modelos GARCH

Geralmente, usa-se o método de máxima verossimilhança, cuja implementação pressupõe que se faça uma hipótese sobre a verdadeira distribuição. Escolhe-se, em geral, a distribuição normal. Entretanto, Hentschel (1995) argumenta que a presença de leptocurtose sugere o uso da distribuição *t*-student ou distribuição erro generalizado (GED) para a estimação, ainda que esta venha a falhar se a série tiver elevado número de distribuições extremas. Não obstante, mesmo na falta de normalidade dos erros, pode-se utilizar a distribuição normal, entendendo isso como uma estimação de quase-máxima-verossimilhança.

Admitindo que a distribuição dos erros seja normal, a função de verossimilhança é a seguinte:

$$L = L(\Theta; y_t) = \prod_{t=\kappa+1}^T \frac{1}{(2\pi\sigma_t^2)^{\frac{1}{2}}} \exp \left\{ -\frac{\epsilon_t^2}{2\sigma_t^2} \right\},$$

em que

Θ é o vetor de parâmetros a serem estimados da equação da média e da variância;

$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_\kappa^2 = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t^2$ são as variâncias iniciais do processo de otimização, necessárias ao algoritmo;

$\kappa = \max[p, q]$, p e q da equação da variância, é o número de variâncias para iniciar o processo de otimização.

Supondo que a distribuição dos resíduos siga uma t -student, a função de verossimilhança deve ser:

$$L = L(\Theta; y_t) = \prod_{t=k}^T \left[\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{(\nu-2)\pi\sigma_t^2}} \exp\left(1 + \frac{\epsilon_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}} \right],$$

em que

$\Gamma(\cdot)$ é a função gama;

$\nu > 2$.

Quando $\nu \rightarrow \infty$, a distribuição converge para uma normal. Entretanto, para $4 < \nu < \infty$, a curtose, $k = 3\frac{\nu-2}{\nu-4} > 3$, excede a curtose da normal.

A obtenção dos parâmetros iniciais se dá por mínimos quadrados ordinários, ou por ajustamento de um modelo $AR(1)MA$, ou ainda pelos parâmetros obtidos com teste LM.

Para a variância condicional, convém observar que:

- a) As condições de estacionaridade e de não-negatividade são difíceis de serem observadas na prática; mas, há métodos de reparametrização, de maneira a assegurar a hipótese de não-negatividade (Bollerslev, Engle e Nelson, 1994).
- b) Como o procedimento descrito é numérico, há problemas de convergência se o valor inicial não for adequado. Não obstante, a estimação por máxima verossimilhança mantém as propriedades assintóticas usuais, como consistência e distribuição normal assintótica.
- c) Um possível método para assegurar a convergência é estimar os parâmetros por mínimos quadrados ordinários, calcular os erros, calcular σ_t , reestimar os parâmetros por mínimos quadrados generalizados, e assim sucessivamente até a convergência dos coeficientes. Entretanto, esse procedimento não tem tanto sentido hoje, dado o estágio computacional disponível (Bueno, 2002).

Somente é possível proceder como sugerido em c se os coeficientes puderem ser consistentemente estimados pela equação da média. Embora Engle (1982) mostre as condições em que isso pode ser feito para o modelo ARCH(q), Bollerslev, Engle e Nelson (1994) argumentam que a perda de eficiência assintótica pode ser significativa sob heterocedasticidade. Isto é, embora os coeficientes possam não ser viesados, não serão eficientemente estimados como no caso simultâneo. Para esse tipo de procedimento, é preciso que a matriz de covariância condicional seja bloco-diagonal e os erros, gaussianos, ou não se pode estimar a equação da média separadamente da variância (veja Gouriéroux, 1997).

Nota 8.1 Por exemplo, os modelos EGARCH (p, q) e GARCH-M (p, q) não são bloco-diagonais; logo, é preciso que as equações da média e variância sejam estimadas simultaneamente.

Há outra razão para estimar ambas as equações simultaneamente. Bera e Higgins (1995, p. 227) dizem que freqüentemente a estimação do processo GARCH muda a especificação do modelo na equação da média.

8.5

Inferência em Modelos Univariados

A máxima verossimilhança, dadas determinadas condições de regularidade, garante consistência e normalidade assintótica dos estimadores. Mas a verificação de tais condições de regularidade tem-se mostrado extremamente difícil nos modelos GARCH (Bollerslev, Engle e Nelson, 1994). Apenas em modelos específicos conseguem-se provas da validade dessas condições. Por exemplo, Lumsdaine (1996) prova que a seguinte condição sobre o GARCH (1,1) é suficiente para assegurar sua ergodicidade e estacionaridade:

$$E \ln |\alpha_1 u_t^2 + \beta_1| < 0.$$

E, pela desigualdade de Jensen:

$$E \ln |\alpha_1 u_t^2 + \beta_1| < \ln E |\alpha_1 u_t^2 + \beta_1| = \ln (\alpha_1 + \beta_1).$$

Dessa forma, a superfície possível dos parâmetros abarca o caso do IGARCH³ (1,1), no qual $(\alpha_1 + \beta_1) = 1$. Ou seja, não obstante sua semelhança com os modelos integrados de ordem 1, sua interpretação assintótica é completamente diferente.

Nessas condições, as inferências sobre os parâmetros de um modelo GARCH são válidas? A teoria assintótica aplicável a tais especificações é extremamente difícil (Bollerslev, Chou e Kroner, 1992). Mas, garantida aquela condição de regularidade, incluindo no IGARCH (1,1), as inferências convencionais permanecem valendo, desde que um considerável tamanho amostral para o IGARCH esteja disponível. Na prática, os estudos empíricos simplesmente assumem que as condições de regularidade são satisfeitas (Bollerslev, Engle e Nelson, 1994, p. 2.980).

³ Os modelos IGARCH (p, q) são idênticos aos modelos GARCH (p, q), exceto que $\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j = 1$. Assim, no modelo IGARCH (p, q) os coeficientes da equação da variância somam 1, de modo que a variância não condicional passa a tender ao infinito.

8.6

Previsão de Modelos GARCH

Nos modelos econométricos convencionais, com variância não condicional, a incerteza do erro de predição é crescente com o horizonte de predição e não depende da origem da previsão. Na presença de modelos ARCH, a acuidade das previsões dependerá do conjunto de informações disponível de forma não trivial. Assim, a construção do intervalo de confiança do erro e os testes de estabilidade estrutural amostrais exigem a avaliação das variâncias condicionais do erro futuro.

O problema reside no fato de que a distribuição dos erros é leptocúrtica, alterando sensivelmente a maneira de se construir o intervalo de confiança. Na realidade, ainda estão sendo feitas pesquisas no sentido de encontrar as formulações corretas para a determinação do intervalo de confiança e da própria função de predição. Heynen e Kat (1994) apresentam funções para os casos de EGARCH (1, 1) e GARCH (1, 1), que, aliás, são os modelos cujas propriedades assintóticas já são conhecidas.

Não obstante, a maneira das previsões segue os mesmos passos do modelo ARIMA. Prevê-se um passo à frente; com o resultado obtido, faz-se o passo seguinte, e assim sucessivamente.

Heynen e Kat fizeram um estudo de previsão usando índice de ações e câmbio. Seus resultados podem ser sintetizados com o seguinte: em geral, a volatilidade é mais previsível no longo do que no curto prazo. Além disso, o desempenho do modelo de previsão depende sensivelmente do ativo com que se está trabalhando. Ainda segundo eles, índices de ações são mais bem previstos utilizando-se o modelo de volatilidade estocástica, enquanto, para o câmbio, o modelo GARCH (1, 1) apresentou os melhores resultados.

Exemplo 8.1

Em modelos empíricos de finanças, muitas vezes é necessário determinar a volatilidade do retorno de um determinado ativo. Essa informação pode ser utilizada tanto no apuração de uma opção quanto na estimação do risco de um ativo ou de uma carteira. Os dados de retorno em finanças geralmente apresentam determinadas propriedades que podem ser modeladas utilizando modelos com heterocedasticidade condicional ou modelos com estrutura ARCH para os erros.

Neste exemplo será estimada a volatilidade da série de retorno do Índice Ibovespa abrangendo o período de 4 de julho de 1994 até 9 de outubro de 2007. Para isso foi aplicado o logaritmo natural sobre o valor de fechamento do Índice Ibovespa e depois foi calculado o retorno da série subtraindo o valor em t de $t - 1$.

Foram estimados os modelos GARCH, EGARCH, GJR, PARCH e GARCH-M, utilizando tanto a distribuição normal quanto a distribuição *t*-student para os erros. Os resultados são dados a seguir (* indica significância a 5%).

1. **GARCH:** os resultados indicam que todos os coeficientes estimados são significantes a 5%, exceto a constante no modelo com distribuição *t*-student. O modelo capta uma alta persistência da volatilidade em ambas as distribuições (de $0,136 + 0,833 = 0,969$ e $0,106 + 0,869 = 0,975$).

$$\sigma_t^2 = 1,84 \times 10^{-5} + 0,136 \epsilon_t^2 + 0,833 \sigma_{t-1}^2 \quad (\text{normal})$$

(3,17 × 10⁻⁶)* (0,010)* (0,013)*

$$\sigma_t^2 = 1,34 \times 10^{-5} + 0,106 \epsilon_t^2 + 0,869 \sigma_{t-1}^2 \quad (t\text{-student})$$

(3,39 × 10⁻⁶)* (0,013)* (0,016)*

2. **EGARCH:** esse modelo tem uma persistência um pouco menor em relação ao anterior, de 0,951 e 0,961. Além disso, o parâmetro que capta a assimetria do modelo é significativo e possui o sinal esperado (-0,122 e -0,106). Por último, salienta-se a significância de todos os parâmetros estimados.

$$\ln \sigma_t^2 = -0,548 + 0,951 \ln \sigma_{t-1}^2 + 0,216 \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - 0,122 \frac{\epsilon_{t-1}}{\sigma_{t-1}} \quad (\text{normal})$$

(0,054)* (0,006)* (0,016)* (0,010)*

$$\ln \sigma_t^2 = -0,446 + 0,961 \ln \sigma_{t-1}^2 + 0,188 \left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - 0,106 \frac{\epsilon_{t-1}}{\sigma_{t-1}} \quad (t\text{-student})$$

(0,063)* (0,007)* (0,023)* (0,013)*

3. **PARCH:** os resultados corroboram o efeito assimetria detectado no modelo anterior por causa da significância dos parâmetros estimados (0,610 e 0,608); além disso, os valores assumidos são menores do que 1 em módulo. Como esperado, o valor de " α " é positivo.

$$\sigma_t^{1,463} = 0,0002 + 0,847 \sigma_{t-1}^{1,463} + 0,103 \left(|\epsilon_{t-1}| - 0,610 \epsilon_{t-1} \right)^{1,463} \quad (\text{normal})$$

(0,0001)* (0,013)* (0,012)* (0,100)*

$$\sigma_t^{1,520} = 0,0001 + 0,875 \sigma_{t-1}^{1,520} + 0,084 \left(|\epsilon_{t-1}| - 0,608 \epsilon_{t-1} \right)^{1,520} \quad (t\text{-student})$$

(0,0001) (0,016)* (0,018)* (0,147)*

4. **GJR**: novamente um modelo assimétrico e pode-se verificar o efeito assimetria presente, em virtude da significância dos parâmetros (0,195 e 0,154).

$$\sigma_t^2 = 2,62 \times 10^{-5} + 0,834 \sigma_{t-1}^2 + 0,015 \epsilon_{t-1}^2 + 0,195 d_{(\epsilon_{t-1} \leq 0)} |\epsilon_{t-1}|^2 \quad (\text{normal})$$

(3,55 × 10⁻⁶)* (0,015)* (0,011) (0,018)*

$$\sigma_t^2 = 1,83 \times 10^{-5} + 0,870 \sigma_{t-1}^2 + 0,012 \epsilon_{t-1}^2 + 0,154 d_{(\epsilon_{t-1} \leq 0)} |\epsilon_{t-1}|^2 \quad (t\text{-student})$$

(3,48 × 10⁻⁶)* (0,017)* (0,013) (0,020)*

5. **GARCH-M**: para esse modelo, assume-se que $\alpha = -1$, então é a variância que entra como variável na equação da média, ajudando a explicar os retornos. Para a distribuição normal, esse parâmetro foi significativo e positivo, como esperado, pois, quanto maior o risco, maior o retorno. No caso da estimação utilizando a distribuição *t-student*, o parâmetro é positivo, mas não é significativo.

$$y_t = 0,0005 + 0,054 y_{t-1} + 2,508 \frac{\sigma_t^2}{2}$$

(0,0005) (0,021)* (1,136)*

$$\sigma_t^2 = 1,94 \times 10^{-5} + 0,138 \epsilon_{t-1}^2 + 0,829 \sigma_{t-1}^2 \quad (\text{normal})$$

(3,34 × 10⁻⁶)* (0,010)* (0,013)*

$$y_t = 0,001 + 0,047 y_{t-1} + 1,363 \frac{\sigma_t^2}{2}$$

(0,0005) (0,020)* (1,237)

$$\sigma_t^2 = 1,35 \times 10^{-5} + 0,108 \epsilon_{t-1}^2 + 0,868 \sigma_{t-1}^2 \quad (t\text{-student})$$

(3,34 × 10⁻⁶)* (0,010)* (0,016)*

Na Figura 8.2 observa-se os valores da volatilidade (variância) estimada para cada modelo. Entre outras coisas, observa-se que os modelos GARCH e GARCH-M estimaram valores maiores para a volatilidade e que os modelos tendem a perceber situações de maior volatilidade em momentos idênticos.

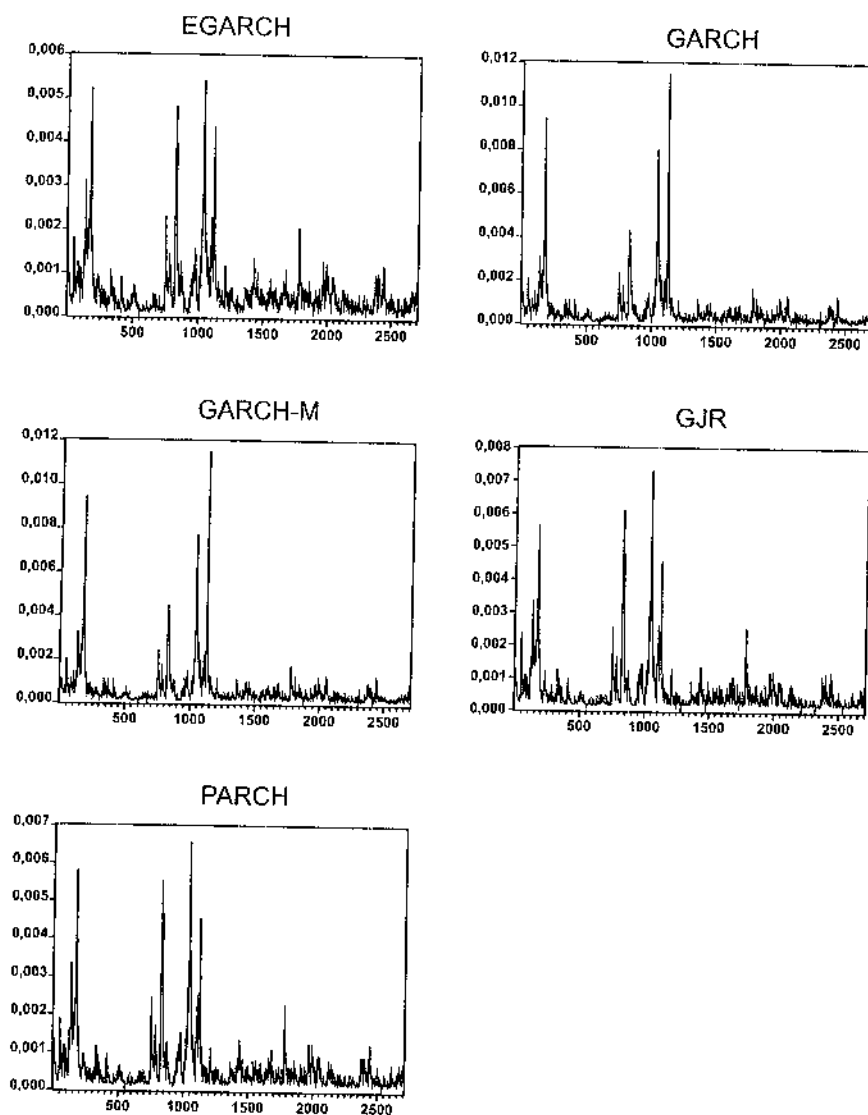


Figura 8.2 Volatilidade estimada utilizando distribuição *t*-student.

8.7

GARCH Multivariado

Nesta seção, estudam-se as três formulações mais comuns. Um dos modelos é o proposto por Bollerslev, Engle e Wooldridge (1988), que estenderam o modelo GARCH-M para o contexto multivariado, usando o operador Vech denotado por MGARCH-M. O outro modelo multivariado é o BEKK, graças a Baba, Engle,

Kraft e Kroner (veja Engle e Kroner, 1995). Finalmente, mostra-se o modelo de correlação condicional dinâmica de Engle e Sheppard (2002).

8.7.1 Vech

Como a matriz de covariância dos retornos é simétrica, os elementos diferentes uns dos outros estão no triângulo inferior dessa matriz. A técnica de estimação consiste, pois, em aplicar um operador nessa matriz, de tal modo que se fique apenas com os elementos do triângulo inferior em forma de vetor. Esse é o chamado operador Vech⁴. Por exemplo, a matriz de covariância de um modelo com n variáveis explicativas no tempo t é assim descrita:

$$\Sigma_t = \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} & \cdots & \sigma_{1n,t} \\ \sigma_{21,t} & \sigma_{22,t} & \cdots & \sigma_{2n,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1,t} & \sigma_{n2,t} & \cdots & \sigma_{nn,t} \end{bmatrix}.$$

Como se supõe que a matriz de covariância é simétrica, então pode-se aplicar o operador Vech para empilhar a porção triangular inferior em um vetor de dimensão $(n(n+1)/2) \times 1$. Ou seja, coloca-se a primeira coluna sobre a segunda; o resultado sobre a terceira; e assim sucessivamente:

$$\text{Vech}(\Sigma_t) = \begin{bmatrix} \sigma_{11,t} & \sigma_{21,t} & \cdots & \sigma_{n1,t} & \sigma_{22,t} & \sigma_{32,t} & \cdots & \sigma_{n2,t} & \cdots & \sigma_{nn,t} \end{bmatrix}'.$$

O modelo MGARCH-M em Vech pode ser formulado assim:

$$Y_t = b + \Sigma_t \delta + \varepsilon_t$$

$$\text{Vech}(\Sigma_t) = C + \sum_{i=1}^q A_i \text{Vech}(\varepsilon_{t-i} \varepsilon_{t-i}') + \sum_{j=1}^p B_j \text{Vech}(\Sigma_{t-j}),$$

em que

Y_t é um vetor $n \times 1$;

b é um vetor de médias (ou constantes) $n \times 1$;

δ é um vetor $n \times 1$;

$\varepsilon_t | I_{t-1} \sim N(0, \Sigma_t)$ é um vetor de perturbações aleatórias, que segue um processo MGARCH;

⁴ O operador Vec procede da mesma maneira, mas transforma todos os elementos da matriz em um único vetor, ou seja, não desconsidera o triângulo superior da matriz de covariância (veja detalhes em Lütkepohl, 2005).

C é um vetor $(n(n+1)/2) \times 1$;

A_i é uma matriz $(n(n+1)/2) \times (n(n+1)/2)$;

B_j é uma matriz $(n(n+1)/2) \times (n(n+1)/2)$;

vech é um operador que empilha as colunas sob a diagonal principal, uma sobre as outras.

Abrindo as equações matriciais anteriores, têm-se:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} & \cdots & \sigma_{1n,t} \\ \sigma_{21,t} & \sigma_{22,t} & \cdots & \sigma_{2n,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1,t} & \sigma_{n2,t} & \cdots & \sigma_{nn,t} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{nt} \end{bmatrix}$$

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{21,t} \\ \vdots \\ \sigma_{n1,t} \\ \sigma_{22,t} \\ \sigma_{32,t} \\ \vdots \\ \sigma_{n2,t} \\ \vdots \\ \sigma_{nn,t} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{n1} \\ c_{22} \\ c_{32} \\ \vdots \\ c_{n2} \\ \vdots \\ c_{nn} \end{bmatrix} + \sum_{i=1}^q \begin{bmatrix} a_{11,i} & a_{12,i} & \cdots & a_{1 \frac{n(n+1)}{2},i} \\ a_{21,i} & a_{22,i} & \cdots & a_{2 \frac{n(n+1)}{2},i} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\frac{n(n+1)}{2}1,i} & a_{\frac{n(n+1)}{2}2,i} & \cdots & a_{\frac{n(n+1)}{2} \frac{n(n+1)}{2},i} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-i}^2 \\ \epsilon_{2,t-i} \epsilon_{1,t-i} \\ \vdots \\ \epsilon_{n,t-i} \epsilon_{1,t-i} \\ \epsilon_{2,t-i}^2 \\ \epsilon_{3,t-i} \epsilon_{2,t-i} \\ \vdots \\ \epsilon_{n,t-i} \epsilon_{2,t-i} \\ \vdots \\ \epsilon_{n,t-i}^2 \end{bmatrix} +$$

$$+ \sum_{j=1}^p \begin{bmatrix} b_{11,j} & b_{12,j} & \cdots & b_{1 \frac{n(n+1)}{2},j} \\ b_{21,j} & b_{22,j} & \cdots & b_{2 \frac{n(n+1)}{2},j} \\ \vdots & \vdots & \ddots & \vdots \\ b_{\frac{n(n+1)}{2}1,j} & b_{\frac{n(n+1)}{2}2,j} & \cdots & b_{\frac{n(n+1)}{2} \frac{n(n+1)}{2},j} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-j} \\ \sigma_{21,t-j} \\ \vdots \\ \sigma_{n1,t-j} \\ \sigma_{22,t-j} \\ \sigma_{32,t-j} \\ \vdots \\ \sigma_{n2,t-j} \\ \vdots \\ \sigma_{nn,t-j} \end{bmatrix}.$$

Na formulação anterior, o número de parâmetros a serem estimados na equação da variância é de $[n(n+1)/2]^2(p+q) + [n(n+1)/2]$. Para evitar que tantos parâmetros sejam estimados, geralmente impõe-se que os elementos das

diagonais secundárias das matrizes A_i e B_i sejam nulos. A economia proporcionada pelo procedimento é de $[n(n+1)/2](p+q)[(n(n+1)/2)-1]$ parâmetros, e o modelo é mais conhecido como Vech-diagonal.

Em um MGARCH (1, 1), as equações da covariância são:

$$\begin{bmatrix} \sigma_{11,t} \\ \sigma_{21,t} \\ \sigma_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} \\ c_{21} \\ c_{22} \end{bmatrix} + \begin{bmatrix} a_{11,1} & 0 & 0 \\ 0 & a_{22,1} & 0 \\ 0 & 0 & a_{33,1} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1}^2 \\ \epsilon_{2,t-1}\epsilon_{1,t-1} \\ \epsilon_{2,t-1}^2 \end{bmatrix} \\ + \begin{bmatrix} b_{11,1} & 0 & 0 \\ 0 & b_{22,1} & 0 \\ 0 & 0 & b_{33,1} \end{bmatrix} \begin{bmatrix} \sigma_{11,t-1} \\ \sigma_{21,t-1} \\ \sigma_{22,t-1} \end{bmatrix}.$$

Assim, a variância de uma variável é explicada apenas pelo quadrado de seus erros passados e de sua variância passada, enquanto a covariância das equações é explicada pelas covariâncias passadas e pelo produto cruzado dos erros passados.

Impõem-se essas restrições pelas seguintes razões: as informações sobre as variâncias são usualmente reveladas nos resíduos quadrados passados e nas variâncias passadas. Se elas se alteram vagarosamente, os resíduos quadrados passados e as variâncias passadas devem ser capazes de prever as variâncias futuras (Engle e Kroner, 1995). O mesmo comportamento se repetiria no caso das covariâncias. Isso significa que as variâncias pouco explicam as covariâncias, e vice-versa.

8.7.2 Restrições e Inferência

Para que se maximize a verossimilhança, é preciso impor restrições para garantir que a matriz de covariância seja positiva definida. Por isso, estima-se a função de verossimilhança sujeita à restrição de que os autovalores da matriz de covariância condicional sejam simultaneamente maiores que zero. Para estacionaridade, é preciso que as raízes da seguinte polinomial estejam fora do círculo unitário:

$$\left| I - \sum_{i=1}^{\max[p,q]} (A_i + B_i) \lambda^i \right| = 0,$$

em que λ é o autovalor associado a cada um dos autovetores.

Como os parâmetros são restritos, surgem sérias dificuldades para a inferência estatística. O método usual de inferência estatística, cálculo da matriz de covariância e construção da estatística t , a partir dos desvios-padrão dos parâmetros, não funciona quando as restrições envolvem desigualdades. Isso se deve ao fato de que as regiões de confiança não serão simétricas em relação

às estimativas. Para superar essa dificuldade, podem-se computar os limites de confiança, por intermédio de simulações (veja Schoenberg, 1997).

8.7.3 BEKK

Para garantir que a matriz de covariância seja positiva semidefinida em cada momento do tempo e reduzir os números de parâmetros estimados, Engle e Kroner (1995) sugerem uma formulação alternativa pouco restritiva:

$$\Sigma_t = C'C + \sum_{i=1}^q A_i' \epsilon_{t-i} \epsilon_{t-i}' A_i + \sum_{j=1}^p B_j' \Sigma_{t-j} B_j,$$

em que C , A_i , B_j são matrizes $n \times n$, mas C é uma matriz triangular superior.

O modelo BEKK é um caso especial do modelo Vech (veja Engle e Kroner, 1995, para uma prova).

A intuição para entender por que a matriz C é triangular superior é lembrar da covariância não condicional em que os elementos diagonais são iguais. O exemplo a seguir mostra claramente por que não se perde generalidade impondo-se triangularidade superior. De qualquer forma, é importante notar que o número de parâmetros a estimar nesse modelo é de $n^2(p+q) + (n(n+1)/2)$, o que proporciona uma economia em relação ao modelo Vech de $(n^2(n+3)(n-1)/4)(p+q)$ parâmetros.

Exemplo 8.2

Exemplificando para $n = 2$, $p = q = 1$, tem-se:

$$\begin{aligned} \begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} &= \begin{bmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{bmatrix}' \begin{bmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{bmatrix} \\ &+ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ &+ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}' \begin{bmatrix} \sigma_{11,t-1} & \sigma_{12,t-1} \\ \sigma_{21,t-1} & \sigma_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}. \end{aligned}$$

Dessa forma, o número de parâmetros a serem estimados reduz-se de 21 para 11, garantindo-se que a matriz de covariância será positiva semidefinida em cada momento do tempo.

Note que:

$$\begin{bmatrix} c_{11} & 0 \\ c_{12} & c_{22} \end{bmatrix}' \begin{bmatrix} c_{11} & c_{12} \\ 0 & c_{22} \end{bmatrix} = \begin{bmatrix} c_{11}^2 & c_{11}c_{12} \\ c_{11}c_{12} & c_{12}^2 + c_{22}^2 \end{bmatrix}.$$

Por inspeção visual, é fácil ver o número de graus de liberdade de que se precisa, por ser necessária uma matriz simétrica. No exemplo, isso implica que existem duas variâncias e uma covariância.

Há duas formas de simplificar o modelo BEKK. No primeiro caso, impõe-se diagonalidade às matrizes A_i e B_j , para obter o chamado DBEKK. Nessa situação, haveria sete parâmetros a estimar no exemplo apresentado. No segundo caso, impõe-se que as matrizes A_i e B_j tornem-se escalares que multiplicam matrizes de uns. Quer dizer, as dinâmicas das variâncias e covariâncias são governadas pelos mesmos parâmetros e só diferem pelos obtidos na matriz C . Esse modelo é chamado de BEKK-escalar.

8.7.4 CCC e DCC

Bollerslev (1990) propôs um modelo em que as covariâncias condicionais são parametrizadas de modo a serem proporcionais ao produto dos desvios-padrão correspondentes (Bollerslev, Engle e Nelson, 1994). Isso diminuiu a intensidade computacional requerida para estimar o modelo e tornou mais fácil impor as condições necessárias para garantir que a matriz de covariância seja positiva semidefinida.

Considere a matriz diagonal D_t^2 , $n \times n$, com covariâncias condicionais ao longo de sua diagonal. Ou seja, $\{D_t^2\}_{ii} = \{\Sigma_t\}_{ii}$ e $\{D_t^2\}_{ij} = 0, \forall j, i, j = 1, 2, \dots, n$. O modelo de correlação condicional constante, CCC, assume que $R_t = R$ é invariante no tempo, de tal modo que a variação temporal em Σ_t é determinada somente pelas variâncias condicionais. Ou seja:

$$\Sigma_t = D_t R D_t.$$

Considerando que as variâncias condicionais ao longo de D_t^2 devem ser todas positivas e assumindo que R seja positiva definida, a sequência de matrizes de covariância condicional, $\{\Sigma_t\}_{t=1,T}$, é positiva definida quase certamente para todo t . Dessa forma, é possível concentrar a função de verossimilhança em R , simplificando bastante a estimação.

Exemplo 8.3

No caso bivariado, tem-se:

$$\Sigma_t = \begin{bmatrix} \sqrt{\sigma_{11,t}} & 0 \\ 0 & \sqrt{\sigma_{22,t}} \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\sigma_{11,t}} & 0 \\ 0 & \sqrt{\sigma_{22,t}} \end{bmatrix}.$$

Mais recentemente, Engle (2003) e Engle e Sheppard (2002) generalizaram o modelo anterior, permitindo que a correlação assumisse uma dinâmica. Assim,

eles introduziram os modelos de correlação condicional dinâmica – DCC. Os modelos de correlação condicional dinâmica permitem que os modelos multivariados sejam estimados a vários passos. O primeiro passo consiste em estimar os modelos individualmente assumindo uma estrutura GARCH. No segundo passo, estima-se a correlação dinâmica a partir dos resíduos padronizados do passo anterior.

Esses artigos mostram duas propriedades importantes desses modelos:

- a) Os parâmetros da estrutura GARCH são independentes da estrutura de correlação sob verossimilhança normal;
- b) A consistência e normalidade dos parâmetros são mantidas, mesmo que os erros não sejam normais, caso em que se pode assumir uma estimação de quasi-verossimilhança.

É claro que a estimação a vários passos requer que o erro padrão dos parâmetros seja modificado, entretanto a matriz robusta de Bollerslev-Wooldridge de cada modelo GARCH univariado permanece consistente, de forma que apenas a covariância assintótica dos parâmetros do estimador da correlação precisa ser ajustada.

A seguir, formaliza-se o modelo DCC. Assume-se que $r_t | I_{t-1} \sim N(0, \Sigma_t)$, em que $\Sigma_t = D_t R_t D_t$ é um vetor de perturbações aleatórias, que segue um processo MGARCH. Aqui r_t pode ser o resultado de uma série filtrada previamente. Segundo Engle e Sheppard (2002, nota de rodapé 2), o erro padrão do modelo não dependerá da escolha do filtro (ARMA, exclusão da média), pois as derivadas parciais cruzadas da função de verossimilhança com respeito aos parâmetros da média e da variância têm esperança nula com função distribuição normal.

A função de verossimilhança é dada por:

$$\begin{aligned}
 L &= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\pi) + \ln |\Sigma_t| + r_t' \Sigma_t^{-1} r_t) \\
 &= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\pi) + \ln |D_t R_t D_t| + r_t' D_t^{-1} R_t^{-1} D_t^{-1} r_t) \\
 &= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\pi) + 2 \ln |D_t| + \ln |R_t| + \epsilon_t' R_t^{-1} \epsilon_t) \\
 L &= -\frac{1}{2} \sum_{t=1}^T (n \ln(2\pi) + 2 \ln |D_t| + r_t' D_t^{-1} D_t^{-1} r_t + \ln |R_t| + \epsilon_t' R_t^{-1} \epsilon_t - \epsilon_t' \epsilon_t),
 \end{aligned}$$

em que $\epsilon_t \equiv r_t' D_t^{-1}$ são os resíduos padronizados, tal que:

$$\epsilon_t \sim N(0, R_t).$$

Um dos objetivos dessa formulação é permitir que o modelo seja estimado mais facilmente, mesmo que a matriz de covariância seja grande. Pode-se dividir a função de verossimilhança em dois conjuntos de parâmetros, Θ_v e Θ_c , da seguinte forma:

$$L(\Theta_v, \Theta_c) = L_v(\Theta_v) + L_c(\Theta_v, \Theta_c),$$

em que o componente de volatilidade, correspondente ao parâmetro Θ_v , é dado por:

$$L_v(\Theta_v) = -\frac{1}{2} \sum_t (n \ln(2\pi) + \ln |D_t|^2 + r_t' D_t^{-2} r_t)$$

e o componente da correlação, correspondente ao vetor Θ_c , é dado por:

$$L_c(\Theta_v, \Theta_c) = -\frac{1}{2} \sum_t (\ln |R_t| + \epsilon_t R_t^{-1} \epsilon_t' - \epsilon_t \epsilon_t').$$

A volatilidade da probabilidade é aparentemente a soma das probabilidades individuais de GARCH:

$$L_v(\Theta_v) = -\frac{1}{2} \sum_t \sum_{i=1}^n \left[\ln(2\pi) + \ln(\sigma_{ii,t}) + \frac{r_{i,t}^2}{\sigma_{ii,t}} \right],$$

que será juntamente maximizada se cada termo for separado. A segunda parte da probabilidade será usada para estimar os parâmetros da correlação.

A forma de estimação a dois passos constitui-se do seguinte, então:

$$\hat{\Theta}_v = \arg \max \{L_v(\Theta_v)\}.$$

Com esse parâmetro, pode-se construir a matriz D_t , da seguinte forma:

$$\sigma_{ii,t} = \omega_i + \sum_{q=1}^{Q_i} \alpha_{iq} r_{i,t-q}^2 + \sum_{p=1}^{P_i} \beta_{ip} \sigma_{ii,t-p}, \quad i = 1, 2, \dots, n.$$

O modelo deve ser estimado com as restrições usuais de não-negatividade e estacionaridade, tal que: $\sum_{q=1}^{Q_i} \alpha_{iq} + \sum_{p=1}^{P_i} \beta_{ip} < 1$. Enfatiza-se que os modelos individuais podem ter diferentes ordens p e q , razão pela qual se colocam os subscritos i nas ordens Q_i e P_i . A especificação do modelo GARCH poderia incluir termos assimétricos e de memória longa, como estudados em seções anteriores.

Dados os parâmetros obtidos nos modelos individuais anteriores, podem-se encontrar os parâmetros da estrutura de correlação, por meio da estimação em segundo estágio:

$$\max_{\Theta_c} \{L_c(\hat{\Theta}_v, \Theta_c)\} = \max_{\Theta_c} -\frac{1}{2} \sum_t (\ln |R_t| + \epsilon_t R_t^{-1} \epsilon_t' - \epsilon_t \epsilon_t').$$

Sob condições regulares bem razoáveis, a consistência da primeira etapa assegurará a consistência da segunda etapa. Como o máximo da segunda etapa será uma função das estimativas do parâmetro da primeira etapa, se a primeira etapa for consistente, o segundo passo também o será, desde que a função seja contínua na vizinhança dos verdadeiros parâmetros (veja Newey e McFadden, 1994, para mais detalhes). A prova dessa proposição pode ser encontrada em Engle e Sheppard (2002).

Nota 8.2 Note que, na última maximização, o termo $\epsilon_t \epsilon_t'$ é desnecessário, pois advém da primeira maximização.

O modelo proposto no segundo estágio supõe encontrar os parâmetros do seguinte modelo dinâmico:

$$Q_t = \left(1 - \sum_{m=1}^M \alpha_m - \sum_{h=1}^H \beta_h\right) \bar{Q} + \sum_{m=1}^M \alpha_m \epsilon_{t-m} \epsilon_{t-m}' + \sum_{h=1}^H \beta_h Q_{t-h},$$

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1},$$

em que \bar{Q} é a covariância não condicional dos resíduos padronizados resultantes da estimação de primeiro estágio e:

$$Q_t^* = \begin{bmatrix} \sqrt{q_{11,t}} & 0 & \cdots & 0 \\ 0 & \sqrt{q_{22,t}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{q_{mm,t}} \end{bmatrix}.$$

A matriz Q_t^* é a matriz dos desvios-padrão padronizados. Assim, um elemento típico de R_t terá a forma:

$$[R_t]_{i,j} \equiv \rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}.$$

Nota 8.3 Os parâmetros α_m e β_h são escalares e iguais para todo q_{ij} , de forma a garantir que a matriz de covariância seja positiva definida para todo t . Essa é a restrição fundamental em relação aos modelos anteriores.

8.7.5 Alisamento Exponencial e EWMA

Alternativas à especificação anterior podem ser tentadas. Por exemplo, pode-se especificar um modelo com um alisamento exponencial do seguinte tipo:

$$\rho_{i,j,t} = \frac{\sum_{s=1}^{t-1} \lambda^s \epsilon_{i,t-s} \epsilon_{j,t-s}}{\sqrt{\left(\sum_{s=1}^{t-1} \lambda^s \epsilon_{i,t-s}^2\right) \left(\sum_{s=1}^{t-1} \lambda^s \epsilon_{j,t-s}^2\right)}}, \lambda \in (0,1).$$

Outra forma é o modelo EWMA multivariado especificado da seguinte forma:

$$q_{ij,t} = (1 - \lambda)(\epsilon_{i,t-1} \epsilon_{j,t-1}) + \lambda(q_{ij,t-1}),$$

$$\rho_{i,j,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}.$$

Normalmente fixa-se $\lambda = 0,94$ para dados diários. Essa recomendação é da RiskMetrics. O problema é considerar isso igual para todos os ativos. A forma multivariada de escrever a equação anterior e comparar com o modelo DCC é:

$$Q_t = (1 - \lambda)(\epsilon_{t-1} \epsilon'_{t-1}) + \lambda Q_{t-1}.$$

Exemplo 8.4

Estimação de hedge

Os agentes participantes do mercado futuro elaboram estratégias de hedge para se proteger de oscilações adversas nos preços das ações. Eles necessitam comprar determinada quantidade de contratos futuros que minimize a volatilidade da riqueza. É necessário o cálculo da taxa ótima de hedge β , determinada pela seguinte fórmula:

$$\beta = \frac{\text{cov}(\Delta \text{Futuro}, \Delta \text{Spot})}{\text{var}(\Delta \text{Futuro})},$$

em que ΔFuturo é a variação absoluta de preços no mercado futuro e ΔSpot é a variação absoluta de preços no mercado à vista.

Dada a volatilidade presente em dados financeiros, é interessante utilizar uma metodologia que permita que a variância e a covariância entre os preços futuros e à vista variem ao longo do tempo e, conseqüentemente, a razão de hedge. Tal possibilidade pode ser alcançada usando modelos GARCH multivariados.

Neste exemplo, é utilizado um BEKK diagonal para a estimação da taxa ótima de hedge usando dados diários de cotação do preço à vista e dos contratos futuros da taxa de câmbio R\$/US\$.

Os dados compreendem o período de 6 de janeiro de 2000 até 16 de fevereiro de 2006. Como em cada instante do tempo haverá mais de um contrato futuro em aberto, foi utilizado o contrato mais próximo da data de vencimento e com maior liquidez. Para os preços do mercado à vista, foi utilizada a série de cotações da

taxa de câmbio R\$/US\$ multiplicada por 1.000. Nos preços obtidos, foi aplicado o logaritmo natural e, finalmente, foi calculado o retorno.

Os resultados da estimação do modelo utilizando uma defasagem para o sistema da variância são dados a seguir:

$$\begin{aligned} \Omega_t = & \begin{bmatrix} 0,0007 & 0 \\ (0,00009) & \\ 0,0004 & 0,0008 \\ (0,0001) & (0,00008) \end{bmatrix}' \begin{bmatrix} 0,0007 & 0 \\ (0,00009) & \\ 0,0004 & 0,0008 \\ (0,0001) & 0,000087 \end{bmatrix} \\ & + \begin{bmatrix} 0,270 & 0 \\ (0,010) & \\ 0 & 0,295 \\ & (0,013) \end{bmatrix}' \epsilon_{t-1} \epsilon_{t-1}' \begin{bmatrix} 0,270 & 0 \\ (0,010) & \\ 0 & 0,295 \\ & (0,013) \end{bmatrix} \\ & + \begin{bmatrix} 0,959 & 0 \\ (0,002) & \\ 0 & 0,952 \\ & (0,003) \end{bmatrix}' \Omega_{t-1} \begin{bmatrix} 0,960 & 0 \\ (0,002) & \\ 0 & 0,952 \\ & (0,003) \end{bmatrix}. \end{aligned}$$

Primeiro, pode-se perceber que todos os parâmetros estimados são estatisticamente significativos. Outro ponto a salientar é a alta persistência (0,960 e 0,952) encontrada na estimação. A Figura 8.3 mostra as estimativas das variâncias dos mercados futuro e à vista e da covariância entre os dois mercados obtidas a partir dos parâmetros estimados. Com esses valores, estimou-se o valor da taxa ótima de hedge para cada instante do tempo, resultado presente na mesma figura (em baixo). Pode-se perceber como o valor da taxa de hedge assume diferentes valores a cada instante do tempo por causa da volatilidade nas séries.

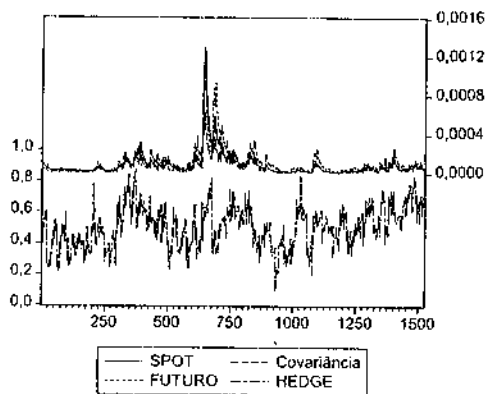


Figura 8.3 Volatilidade e hedge.

8.8

Teste de Resíduos ARCH-LM

Esta seção sugere um procedimento para identificar a existência de GARCH em um contexto multivariado. A idéia é estimar o VAR multivariado e tomar os resíduos daí decorrentes para obter subsídios à existência de covariância condicional. Portanto, seja \hat{e}_t o vetor de resíduos do VAR. O teste prossegue estimando-se:

$$\text{vech}(\hat{e}_t \hat{e}_t') = C_0 + C_1 \text{vech}(\hat{e}_{t-1} \hat{e}_{t-1}') + C_2 \text{vech}(\hat{e}_{t-2} \hat{e}_{t-2}') + \dots + C_q \text{vech}(\hat{e}_{t-q} \hat{e}_{t-q}') + u_t,$$

em que

vech é um operador que empilha as colunas sob a diagonal principal, uma sobre as outras;

C_0 é um vetor $(n(n+1)/2) \times 1$;

C_i é uma matriz $(n(n+1)/2) \times (n(n+1)/2)$.

As hipóteses nula e alternativa são:

$$H_0 : C_1 = C_2 = \dots = C_q = \mathbf{0} \times H_1 : C_1 \neq \mathbf{0} \vee C_2 \neq \mathbf{0} \vee \dots \vee C_q \neq \mathbf{0}.$$

A estatística multivariada é:

$$\text{MAR}CH_{LM}(q) = \frac{Tn(n+1)R_m^2}{2} \xrightarrow{d} \chi_{\frac{q n^2 (n+1)^2}{4}}^2;$$

$$R_m^2 = 1 - \frac{2}{n(n+1)} \text{tr}(\hat{\Sigma}_u \hat{\Sigma}_r^{-1}),$$

em que

$\hat{\Sigma}_u$ é a matriz de covariância irrestrita dos resíduos com dimensão

$$\frac{n(n+1)}{2} \times \frac{n(n+1)}{2};$$

$\hat{\Sigma}_r$ é a matriz de covariância restrita dos resíduos da regressão impondo

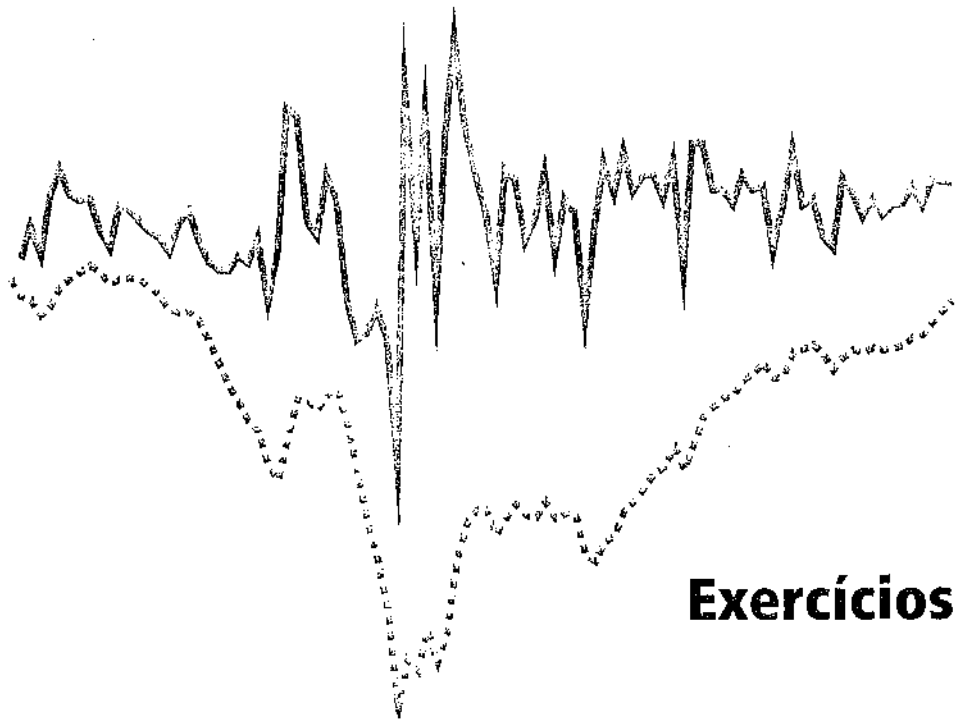
$$C_i = \mathbf{0}.$$

Exemplo 8.5

Considere o modelo não estacionário sem drift simulado no capítulo sobre VAR. Estimado o modelo, verificam-se os resíduos usando esses testes. Estimado o VAR (1), procede-se ao teste LM sobre existência de heterocedasticidade condicional.

<i>Defasagens</i>	<i>Q-Stat</i>	<i>Prob.</i>	<i>df</i>
1	1,678	0,996	9
2	2,924	1,000	18
3	6,523	1,000	27
4	5,032	1,000	36

O *Eviews* reporta o resultado das probabilidades com n^2 graus de liberdade. Por isso, é preciso olhar a tabela usando $qn^2 (n + 1)^2 / 4$ graus de liberdade. A tabela anterior contém as probabilidades corretas; entretanto, se fossem usados n^2 , a hipótese nula não seria rejeitada também. No caso do exemplo, como não se rejeita a nula com menos graus de liberdade do que seria correto, com mais razão não se rejeitaria se o número de graus de liberdade fosse ainda maior.



Exercícios

Revisão de Econometria¹

1. Suponha o seguinte modelo linear: $y = X\beta + \epsilon$, em que y e ϵ são vetores $n \times 1$, X é uma matriz $n \times k$ e β é um vetor $k \times 1$.
 - (a) Determine a(s) hipótese(s) necessária(s) para estimar esse modelo por MQO.
 - (b) Determine a(s) hipótese(s) necessária(s) para que o β estimado, $\hat{\beta}$, exista e seja único.
 - (c) Determine a(s) hipótese(s) necessária(s) para que $\hat{\beta}$ seja não viesado.
 - (d) Determine a(s) hipótese(s) necessária(s) para que $\hat{\beta}$ seja eficiente.
 - (e) Determine a(s) hipótese(s) necessária(s) para que se possa fazer inferência estatística.

¹ Alguns dos exercícios a seguir são adaptações de Bueno (2001).

2. Adão Ismiti queria entender como a produtividade do trabalho se associava à divisão do trabalho. Propôs-se a fazer a seguinte experiência: primeiro, ele regrediu produtividade (p) de n trabalhadores de fábricas de alfinetes contra o número de funções exercidas por esse trabalhador (F), os anos de escolaridade (E), o salário (w) e o número de filhos (N). Formalmente, a regressão ficou assim:

$$p_i = \beta_1 + \beta_2 F_i + \beta_3 E_i + \beta_4 w_i + \beta_5 N_i + u_i.$$

Usando o teste t -student, o eminente pensador não rejeitou a hipótese nula de que $\hat{\beta}_3 = 0$. Em seguida, retirou a variável E da regressão para estimar um modelo mais restrito. Como resultado, $\hat{\beta}_5$ tornou-se, também, estatisticamente não significativo. Por último, retirou a variável N do modelo e o reestimou novamente.

- (a) Explique por que é desnecessário retirar E do modelo para fazer o teste de F em $\hat{\beta}_3$, bastando, pois, apenas o teste de t -student.
 - (b) O procedimento adotado por Ismiti está correto ou equivocado por ter eliminado a variável N do modelo?
3. Considere o modelo linear $y = X\beta + \epsilon$, em que y e ϵ são vetores $n \times 1$, $X < \infty$ é uma matriz $n \times k$, e β é um vetor $k \times 1$, estimado por MQO com constante. Verifique se as afirmações a seguir são (F) falsas ou (V) verdadeiras e justifique sua resposta:
- (a) Ante a existência de perturbações heterocedásticas, as estimativas de β são consistentes.
 - (b) Ante a existência de perturbações heterocedásticas, as estimativas de β são ineficientes.
 - (c) Ante a existência de perturbações heterocedásticas, a matriz de covariância das estimativas é inconsistente.
 - (d) Ante a existência de perturbações heterocedásticas, os testes de hipóteses sobre os coeficientes são inválidos.
4. Considere um modelo de regressão linear múltiplo em que $\hat{\beta}$ exista, seja não viesado e eficiente, haja vista que u é homocedástico. Imponha falsas restrições sobre os parâmetros do modelo.
- (a) Mostre que as estimativas são viesadas.

- (b) Mostre que a variância das estimativas do modelo com restrições é menor que a variância das estimativas do modelo sem restrição.
- (c) Qual a implicação desse resultado em termos de previsão? Qual a intuição desse resultado?

5. Responda:

- (a) Mencione ao menos dois testes para a hipótese de homocedasticidade.
- (b) Mencione ao menos um teste para a hipótese de autocorrelação dos resíduos.
- (c) Se você rejeita a hipótese nula em a., por qual método você estimaria o modelo?
- (d) Se você rejeita a hipótese nula em b., por qual método você estimaria o modelo?

6. Faça os seguintes exercícios:

- (a) Suponha que $\sum_{i=0}^{\infty} |x_i| < \infty$. Mostre que $\sum_{i=0}^{\infty} x_i^2 < \infty$.
- (b) Prove que $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k} = \infty$.
- (c) Prove que $\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{1}{k^2} < \infty$.
- (d) Prove que, se $\sum_{i=0}^{\infty} x_i^2 < \infty$, então $\sum_{i=0}^{\infty} |x_i| < \infty$.

Fundamentos Estatísticos

1. Considere a seguinte afirmação como verdadeira: Seja $\{Z_t\}$ uma sequência de variáveis aleatórias *i.i.d.* $N(0, 1)$, então $\{Z_t\}$ é (estritamente) estacionária.
 - (a) Qual a hipótese básica do resultado anterior? Por quê?
 - (b) É possível afirmar que estacionaridade reforça a hipótese de distribuição idêntica?
 - (c) A hipótese de estacionaridade sobre uma série qualquer é mais fraca que a hipótese *i.i.d.*? Por quê?
2. Defina processo estocástico e ilustre graficamente. Explique o que é a realização de um processo estocástico e por que as séries econômicas podem ser entendidas como geradas por processos estocásticos.

3. Por que são impostas restrições sobre a heterogeneidade temporal e sobre a memória de um processo estocástico?
4. Determine a diferença entre estacionaridade forte (ou estrita) e estacionaridade (fraca). Construa exemplos mostrando quando uma implica a outra, e quando uma não implica a outra.
5. Responda:
 - (a) Mostre algebricamente como um processo $AR(2)$, com raízes fora do círculo unitário é expresso como um $MA(\infty)$.
 - (b) Escreva um $MA(1)$ sob a forma de um $AR(\infty)$.
 - (c) Por que as raízes do processo MA devem estar fora do círculo unitário?
6. Verifique se os modelos a seguir são estacionários e/ou inversíveis, em que L é o operador defasagem:
 - (a) $(1 - L)y_t = (1 - 0,5)\epsilon_t$;
 - (b) $(1 + 0,8L)y_t = (1 - 1,2L)\epsilon_t$;
 - (c) $(1 - 0,7L + 0,4L^2) = (1 - 0,5L)\epsilon_t$;
 - (d) $(1 - 0,7L - 0,4L^2) = (1 - 1,6L + 0,7L^2)\epsilon_t$;
 - (e) $(1 + 0,9L)y_t = (1 + 0,5L + 0,4L^2 + 0,3L^3)\epsilon_t$.
7. Calcule as autocorrelações dos modelos $MA(2)$, $AR(2)$ e $ARMA(1,1)$.
8. Considere o seguinte processo estocástico:

$$Y_t = \phi Y_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.N(0,1), \quad Y_0 = 0,$$

em que ϕ pode assumir os seguintes valores: 1,0; 0,9; 0,5. Simule 1.000 séries (com 100 observações cada) para cada um dos parâmetros teóricos de ϕ ; estime-os em seguida por MQO. Comente as propriedades do estimador.

Processos Estacionários

1. Considere o processo $AR(1)$ a seguir:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t.$$

- Defina os estimadores por OLS de ϕ_0 e ϕ_1 .
 - Considere que $\epsilon_t \sim i.i.N(0, \sigma^2)$. Suponha que se observem $\{y_1, y_2, \dots, y_T\}$. Tome a primeira observação y_1 como dada e obtenha a função de log-verossimilhança condicional das observações restantes [ou seja, de $p(y_2, y_3, \dots, y_T | y_1)$].
 - Mostre que o estimador por ML condicional resultado de (b) é equivalente ao estimador por OLS de (a). O que aconteceria se tivéssemos média móvel, ou seja, se quiséssemos estimar um $ARMA$? Discorra em linhas gerais.
2. Calcule (manualmente) as primeiras cinco autocorrelações para cada um dos seguintes processos:
- $Y_t = \epsilon_t + \theta \epsilon_{t-1}$, com $\theta = -0,5$;
 - $(1 - \phi L) Y_t = \epsilon_t$, com $\phi = 0,9$;
 - $(1 - \phi L) Y_t = \epsilon_t + \theta \epsilon_{t-1}$, com $\phi = 0,9$ e $\theta = -0,5$.
3. Considere o processo $AR(2)$ a seguir:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t,$$

em que $\phi_0 = 0$, $\phi_1 = 0,4$ e $\phi_2 = -0,5$. Calcule (manualmente) os primeiros valores da função de autocorrelação parcial.

4. No capítulo, simulou-se um processo $MA(2)$. Apesar de gerado um $MA(2)$, o correlograma, assim como os critérios de informação, indica que o processo que melhor se ajustaria seria um $MA(2)$ degenerado, ou seja, $y_t = \epsilon_t + \theta_2 \epsilon_{t-2}$. Discuta possíveis explicações para esse fenômeno.
5. Considere o seguinte modelo:

$$\hat{y}_t = \underset{(0,159843)}{0,2969} + \underset{(0,04252)}{0,803458} y_{t-1} + \hat{\epsilon}_t,$$

amostra de 200 observações;

(...) = desvio-padrão.

Calcule $\hat{\mu} = \frac{\hat{\epsilon}}{1-\hat{\phi}}$ e mostre que $\hat{\sigma}_{\hat{\mu}} = 0,7588$. Use o método delta.

6. Considere um modelo $MA(1)$: $y_t = \epsilon_t + \theta\epsilon_{t-1}$, $\epsilon_t \sim N(0, \sigma^2)$. Usando a metodologia de Box, Jenkins e Reinsel (1994) de *backforecasting*, determine $E(\epsilon_t|Y, \theta)$ de forma a obter a verossimilhança exata, isto é, determine sua formulação recursiva.

Processos Não Estacionários

1. Identifique e estime um processo $ARIMA$ para as séries a seguir. Proceda aos testes de raiz unitária ADF, ADF-GLS, PP, KPSS, ERS, NP e indique possíveis discrepâncias entre esses testes. Explícite os passos efetuados (por exemplo: "Observando a FAC e a FACP, a série pode ser um $ARMA(1,1)$, $ARMA(2,1)$, $AR(1)$..."). (Lembre-se dos passos: estacionaridade, identificação, estimação e verificação).
 - (a) IPCA (aplicar o $\ln()$ ao número índice).
 - (b) Produção industrial mensal do IBGE (aplicar o $\ln()$ ao número índice).
 - (c) Exportações brasileiras (aplicar o $\ln()$ à série) em US\$ FOB (código BCB: 2946).
2. Simule o seguinte modelo $ARMA$ com 300 observações: $y_t = y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$. Faça o teste de raiz unitária para $\theta = [-0,98; -0,95; -0,90; -0,85; -0,80; -0,50; 0,90; 0,95; 0,98]$. Você aceita a hipótese de raiz unitária para todos os valores de θ ? Se não, para quais valores você rejeita usando DF e usando PP? Interprete o resultado.
3. Utilizando a série do Ibovespa, mensal, aplique os testes de raiz unitária e avalie a presença de raiz unitária. A inferência estatística difere se forem utilizados testes distintos? A forma de cálculo da variância de longo prazo é relevante neste caso?

GMM

1. Hansen e Singleton (1983) estimam o modelo *consumption based asset pricing* por GMM. Após a maximização da utilidade intertemporal de um agente representativo, tendo como variável o consumo, sujeito à sua restrição orçamentária intertemporal, o problema se reduz a estimar os parâmetros equação de consumo e investimento ótimos:

$$u'(c_t) = \delta E_t [R_{t+1} u'(c_{t+1})] \text{ para todo } t,$$

em que R_{t+1} é o retorno bruto ($1 +$ taxa de retorno) do ativo em t , p_t é o preço do ativo em t , c_t é o consumo no período t , δ o desconto intertemporal e $E_t[\cdot]$ a esperança condicional [note que a taxa de retorno é $R_{t+1} = (P_{t+1} + D_{t+1}) / P_t$, em que D_t é o dividendo].

- (a) Veja os exemplos de apreçamento de ativos no capítulo sobre GMM e derive esse resultado.
 - (b) Reescreva a condição de equilíbrio tomando a função de utilidade $u(c_t) = c_t^\gamma / \gamma$.
 - (c) Se o objetivo é estimar os parâmetros (δ, γ) por GMM, defina os momentos condicionais da população (tome como dado o vetor de instrumentos z_t).
2. Utilizando os dados de consumo de bens não duráveis (*proxy* para c_t) dos Estados Unidos e o índice de retorno com *proxy* dos retornos sobre investimentos, estime no Eviews por GMM os parâmetros (δ, γ) . (Hansen e Singleton utilizam como instrumentos as defasagens em um, dois, quatro ou seis meses dos retornos e consumo.)
 - (a) Estime o modelo usando os instrumentos sugeridos. Compare com os resultados de Hansen e Singleton (1983).
 - (b) Compare como as estimativas variam quando o conjunto de instrumentos varia.
 - (c) Calcule a estatística “J”. Comente o resultado dessa estatística.
 - (d) Estime o mesmo modelo por mínimos quadrados não-lineares e compare os resultados. Os parâmetros estimados diferem? Comente.
 3. Em que casos a estimação por GMM gera resultados iguais à da estimação por OLS? Derive o resultado algebricamente.
 4. Nem sempre a estimação por GMM é desejável, mesmo sendo mais eficiente. Apresente pelo menos dois argumentos que comprovem essa afirmação.
 5. Issler e Piqueira (2002) estimam as equações de Euler utilizando três tipos de funções de utilidade para o caso brasileiro. Usando as ferramentas de GMM, reproduza os resultados dos três modelos a seguir:

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{i,t+1}) \right] = 1,$$

em que $i = (\text{ibov}, \text{títulopub})$.

$$E_t \left[\beta \left(\frac{c_t}{c_{t-1}} \right)^{-\kappa(\gamma-1)} \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + r_{i,t+1}) \right] \approx 1,$$

em que $i = (\text{ibov}, \text{títulopub})$.

$$E_t \left[\beta^\eta \left(\frac{c_{t+1}}{c_t} \right)^{\eta(\rho-1)} r_{\text{ibov},t+1}^{\eta-1} (1 + r_{\text{títulopub},t+1}) \right] = 1$$

$$E_t \left[\frac{1}{\eta} \left\{ \left[\beta \left(\frac{c_t}{c_{t-1}} \right)^{(\rho-1)} r_{\text{ibov},t+1} \right]^\eta - 1 \right\} \right] = 0.$$

Estime os parâmetros (β, γ) , (β, κ, γ) e $(\beta, \rho, \psi, \gamma)$, em que $\psi = (1 - \rho)^{-1}$, $\gamma = 1 - \alpha$, correspondentes a cada especificação acima. Apresente os valores das estatísticas dos parâmetros e a J . Utilizando o Eviews, veja se os resultados variam conforme o método de estimação adotado (Kernel e janelas). Compare os resultados com o artigo de Issler e Piqueira (2002) e com Cysne (2006).

6. Clarida, Galí e Gertler (2000) estimam a seguinte função de reação do Banco Central na determinação da taxa de juros norte-americana:

$$i_t = g_i i_{t-1} + (1 - g_i) (\mu + g_\pi \pi_{t,k} + g_x x_{t,q}) + \epsilon_t,$$

em que $\mu = i^* - (g_\pi - 1)\pi^*$, i_t é a taxa de juros efetiva, i^* é a meta da taxa de juros para t , π^* é a meta da taxa de inflação, $\pi_{t,k}$ é a inflação entre t e $t + k$, $x_{t,q}$ é o hiato do produto entre os períodos t e $t + q$.

- Use o artigo mencionado para explicar como os autores chegam à especificação anterior. Quais imposições são feitas sobre os dados?
- Defina os instrumentos e as condições de momentos. Aponte justificativas econômicas dessas condições.
- Estime os parâmetros (g_i, g_π, g_x, μ) por GMM, nos períodos: completo, pré-Volcker (antes de 1979 : 02 inclusive) e pós-Volcker.
- Comente como o parâmetro g_π estimado varia nestas três amostras.

Vetor Auto-Regressivo – VAR

1. Considere o seguinte modelo VAR estrutural bivariado:

$$\begin{pmatrix} 1 & -a_{12} \\ -a_{21} & 1 \end{pmatrix} \begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{pmatrix}.$$

- (a) Pressuponha que $E[\epsilon_{t,1}] = E[\epsilon_{t,2}] = 0$, $E[\epsilon_{t,1}\epsilon_{\tau,1}] = \sigma_1$ para $t = \tau$ e zero, caso contrário, $E[\epsilon_{t,2}\epsilon_{\tau,2}] = \sigma_2$ para $t = \tau$ e zero, caso contrário, e $E[\epsilon_{t,1}\epsilon_{\tau,2}] = 0$ para todo t e τ . Quais as implicações desses pressupostos? Explique brevemente.
- (b) Reescreva o VAR em sua estrutura original na forma reduzida. Discorra sobre a identificação do VAR estrutural (número de parâmetros, como atingir identificação).
- (c) Suponha que o VAR estrutural em (b) está exatamente identificado. Expresse os parâmetros estruturais na forma dos parâmetros reduzidos.

2. Considere o seguinte VAR estrutural:

$$\begin{aligned}y_{1,t} &= 0,5y_{2,t} - 0,1y_{1,t-1} - 0,2y_{2,t-1} + u_{1,t}; \\ y_{2,t} &= 0,8y_{1,t-1} + 0,4y_{2,t-1} + u_{2,t}.\end{aligned}$$

- (a) O modelo está identificado?
- (b) Reescreva o modelo na forma reduzida. O que se pode dizer sobre a causalidade de Granger das séries?
- (c) Calcule a função resposta ao impulso de $\epsilon_{t,1}$ sobre $y_{t,2}$ para os três primeiros períodos.
3. Considere o seguinte VAR:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} 0,1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0,8 & 0 \\ 0,2 & 0,4 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} e_{1,t} \\ e_{2,t} \end{pmatrix}$$

- (a) O modelo é estacionário, cointegrado ou possui duas raízes unitárias independentes?
- (b) Calcule o valor da função resposta ao impulso de um choque unitário em $u_{1,t}$ e $u_{2,t}$, para seis períodos.
- (c) Suponha que a matriz de covariância dos resíduos seja dada por:

$$\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}.$$

Decomponha essa matriz usando o método de André Luís Cholesky e recalcule a função resposta ao impulso.

Vetor de Correção de Erros – VECM

1. Modelos de cointegração são uma ferramenta importante para testar modelos econômicos que envolvam séries não estacionárias. Em particular, muitos estudos têm procurado estabelecer relações entre as taxas de câmbio de diversas moedas.
 - (a) Utilize seu conhecimento macroeconômico e estabeleça uma relação teórica entre as taxas de câmbio dos países.
 - (b) Descreva a metodologia proposta por Engle e Granger e aplique-a para testar seu modelo para a taxa de câmbio entre dólar e iene.
 - (c) Descreva brevemente a metodologia de Johansen. Teste a presença de cointegração e interprete as estatísticas λ_{\max} e λ_{trace} .
 - (d) O que se pode dizer das duas metodologias para esse caso? Alguma é preferível à outra?
2. Muito se tem dito sobre a taxa de câmbio e outras variáveis financeiras, porém praticamente nada sobre o nosso pão francês de cada dia. Este exercício e o seguinte propõem-se a preencher essa lacuna e aproximar a teoria econômica do nosso cotidiano matinal.

Sejam os preços da farinha de trigo ao consumidor em duas regiões metropolitanas, p_1 e p_2 , e o preço do trigo para o produtor, p_p . Tem-se o seguinte modelo VECM:

$$\begin{aligned}
 \begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \\ \Delta p_{p,t} \end{pmatrix} &= \begin{pmatrix} \phi_{11} & \phi_{12} & \phi_{13} \\ \phi_{21} & \phi_{22} & \phi_{23} \\ \phi_{31} & \phi_{32} & \phi_{33} \end{pmatrix} \begin{pmatrix} \Delta p_{1,t-1} \\ \Delta p_{2,t-1} \\ \Delta p_{p,t-1} \end{pmatrix} \\
 &+ \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{bmatrix} (p_1 - \beta_{10} - \beta_{11}p_2)_{t-1} \\ (p_2 - \beta_{20} - \beta_{21}p_p)_{t-1} \end{bmatrix} \\
 &+ \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} d_{1,t} + \begin{pmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{pmatrix},
 \end{aligned}$$

em que $d_{1,t}$ é uma *dummy* para o período de câmbio flutuante (a partir de janeiro de 1999).

- (a) Interprete o modelo anterior, explicando cada uma de suas partes.

- (b) Estime os parâmetros e interprete o significado deles. (Note que o vetor de correção de erros está restrito; atenção quando estimá-lo).
3. No exemplo do capítulo, foi criado um modelo para os índices de preços do trigo no atacado e da farinha e pão francês ao consumidor. Escolha alguma região metropolitana e:
- (a) verifique se há alguma relação de cointegração entre as séries, usando o teste de Johansen;
 - (b) especifique o modelo e estime-o;
 - (c) interprete os resultados.

Heterocedasticidade Condicional

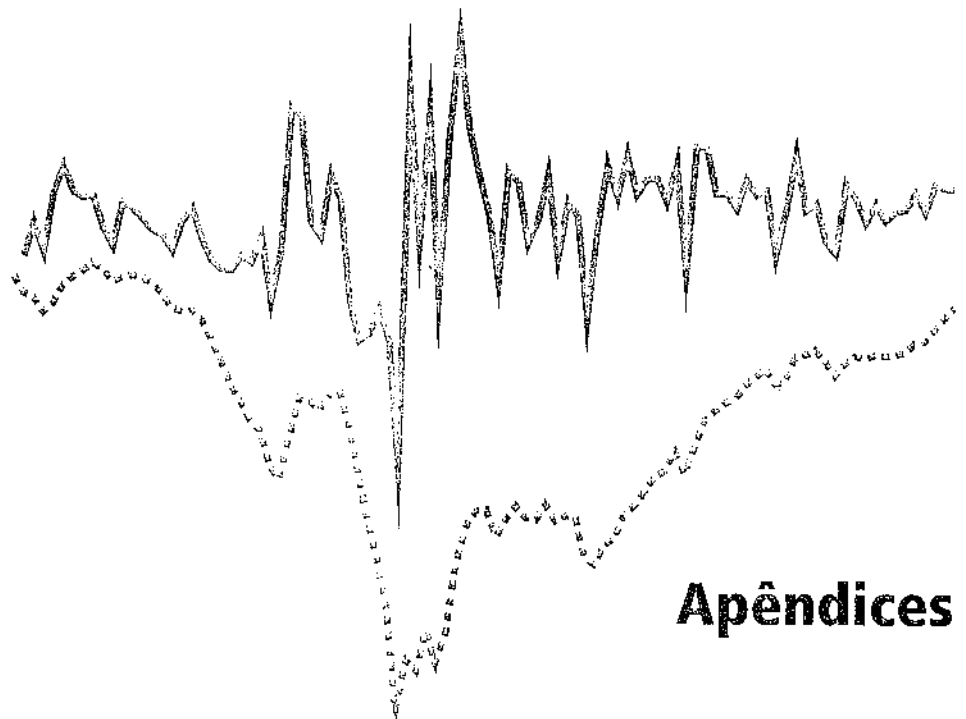
1. Seja $\epsilon_t \sim N(0, \sigma_t^2)$, em que $\sigma_t^2 = 0,2 + 0,09\epsilon_{t-1}^2 + 0,9\sigma_{t-1}^2$.
 - (a) Lembrando que $\sigma_t^2 = E[\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \dots]$, use a lei das expectativas iteradas para encontrar a variância não condicional de ϵ_t .
 - (b) Seja $e_t = \epsilon_t^2 - \sigma_t^2$ com $VAR(e_t) = \sigma_e^2$. Encontre a expressão para as cinco primeiras autocorrelações de ϵ_t^2 .
2. Usando a diferença do log da taxa de câmbio nominal R\$/US\$ diária, identifique e estime o modelo que melhor se encaixa aos dados (tanto a equação das médias como a equação da variância). Considere os modelos ARCH e GARCH.
3. Faça o solicitado no exercício anterior utilizando o índice Ibovespa e considere agora também modelos assimétricos TGARCH, EGARCH e PGARCH.
4. Seja o seguinte modelo BEKK bivariado:

$$\begin{bmatrix} \text{Ibov}_t \\ \text{Câmbio}_t \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}, \epsilon_t \sim N(0, H)$$

$$\begin{aligned}
\begin{bmatrix} \sigma_{11,t} & \sigma_{12,t} \\ \sigma_{21,t} & \sigma_{22,t} \end{bmatrix} &= \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix}' \begin{bmatrix} c_{11} & 0 \\ c_{21} & c_{22} \end{bmatrix} \\
&+ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \epsilon_{1,t-1}^2 & \epsilon_{1,t-1}\epsilon_{2,t-1} \\ \epsilon_{2,t-1}\epsilon_{1,t-1} & \epsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
&+ \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}' \begin{bmatrix} \sigma_{11,t-1} & \sigma_{12,t-1} \\ \sigma_{21,t-1} & \sigma_{22,t-1} \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}.
\end{aligned}$$

Transforme o modelo em um DBEKK (impondo diagonalidade às matrizes $[a_{ij}]$ e $[g_{ij}]$) e estime-o usando dados diários do log da diferença das séries Ibovespa e câmbio, de janeiro de 1999 a dezembro de 2006. Utilize o software que preferir. Faça o gráfico das variâncias e covariância.

5. Comente as principais diferenças entre os modelos ARCH, GARCH, ARCH-M, EGARCH e TARCH.



Apêndices

Equações a Diferenças Estocásticas

Esta seção revisa os modelos de equações a diferenças lineares. Aparentemente restritivos em relação a modelos não-lineares, na verdade são poderosos para previsão. A estratégia é começar da maneira mais simples e, paulatinamente, passar a modelos mais elaborados. Os modelos mais simples são importantes ao prover a intuição para as elaborações posteriores. O objetivo da revisão é resolver as equações a diferenças como função dos termos que são dados.

Uma equação a diferenças é tal que y_t depende do seu próprio passado, y_{t-1}, y_{t-2}, \dots . Se $y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p})$, diz-se que a equação a diferenças é de ordem p . Considere, então, a seguinte equação a diferenças de ordem 1:

$$y_t = \mu + \phi y_{t-1} + \epsilon_t.$$

O termo ϵ_t é um termo aleatório, ao qual se chama erro, que deve ter algumas propriedades a serem discutidas mais tarde. Como o objetivo é encontrar a solução dessa equação, é preciso caracterizar como é essa solução. É o que a definição a seguir faz.

Definição A.1 *Uma solução para uma equação a diferenças expressa o valor de y_t como uma função dos elementos da sequência $\{\epsilon_t\}$, do tempo t e das condições iniciais de y_t .*

Exemplo A.1

Para ilustrar a definição mencionada, considere a seguinte equação a diferenças, em que $\epsilon_t \sim N(0, 1)$ é o erro. Os possíveis valores de y_t podem ser calculados a partir da seguinte equação:

$$y_t = 0,5 + 0,8y_{t-1} + \epsilon_t,$$

em que $y_0 = 1$.

Tabela A.1 Valores de y_t

ϵ_t	$y_t = 0,5 + 0,8y_{t-1} + \epsilon_t$
0,03	1,33
-0,81	0,75
-1,49	-0,39
-0,08	0,10
1,14	1,73
1,22	3,11
0,09	3,08
1,46	4,42
1,23	5,27
0,65	5,37

Dados $y_0 = 1$, $\mu = 0,5$, $\phi = 0,8$ e a sequência $\{\epsilon_t\}_{t=1}^{10}$, pode-se reconstruir toda a série, representada na Figura A.1 a seguir.

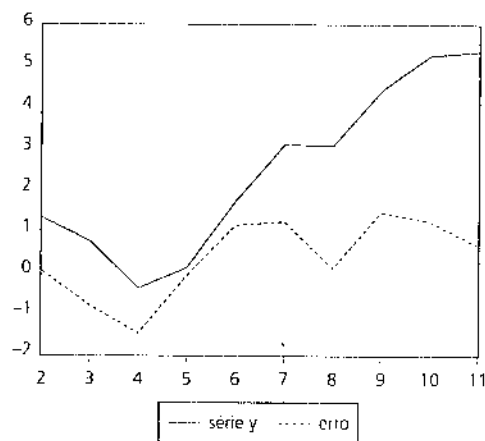


Figura A.1 Série y_t dados y_0 , μ , ϕ e a sequência $\{\epsilon_t\}_{t=1}^{10}$.

Substituição Recursiva

Esta seção encontra a solução geral da equação anterior por substituição recursiva. Essa solução tem de ser tal que dependa dos parâmetros que são dados: μ , ϕ , y_0 e a sequência $\{\epsilon_t\}$. Para isso, pode-se escrever a equação anterior como:

$$y_{t-1} = \mu + \phi y_{t-2} + \epsilon_{t-1}.$$

Substituindo esse resultado na equação anterior, obtém-se:

$$\begin{aligned} y_t &= \mu + \phi y_{t-1} + \epsilon_t \implies y_t = \mu + \phi (\mu + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \mu + \phi\mu + \phi^2 y_{t-2} + \epsilon_t + \phi\epsilon_{t-1}. \end{aligned}$$

Repetindo o procedimento $(t-1)$ vezes:

$$y_t = \mu \sum_{j=0}^{t-1} \phi^j + \phi^t y_0 + \sum_{j=0}^{t-1} \phi^j \epsilon_{t-j}. \quad (\text{A-1})$$

Dados os valores dos parâmetros do modelo, pode-se encontrar a solução. Os termos $\{\epsilon_j\}_{j=1}^t$ são aleatórios e dados pela natureza. Como ninguém tem controle sobre eles, importam apenas para definir os resultados numéricos do problema.

Algumas observações emergem dessa última equação. Primeiro, é fácil encontrar o efeito da perturbação ϵ_{t-j} sobre y_t :

$$\frac{\partial y_t}{\partial \epsilon_{t-j}} = \phi^j.$$

A importância desse fato refere-se à previsão. É muito comum querer estimar o efeito de um choque aleatório sobre a variável de interesse. Em particular, quer-se saber o efeito de um choque hoje daqui a j períodos. No caso da equação a diferença estocástica de ordem 1, o efeito será:

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = \phi^j.$$

Nota A.1 Observe que, se $|\phi| < 1$, então o efeito de um choque hoje será menor quanto mais próximo de zero estiver ϕ e quanto maior for o valor de j . Entretanto, se $|\phi| = 1$, o choque será permanente, refletindo-se totalmente no futuro, não interessando o número de períodos à frente. Além disso, será acumulado com os demais choques que porventura tiverem ocorrido. Finalmente, se $|\phi| > 1$, os efeitos serão crescentes, tendendo à explosão.

Definição A.2 Se $|\phi| < 1$, diz-se que o sistema é estável ou convergente. Se $|\phi| > 1$, diz-se que o sistema é divergente ou explosivo.

A Figura A.2 dá uma idéia do que acontece quando se desenha a função $f(j) = \phi^j$ para uma equação a diferenças de primeira ordem. A interpretação que se deve dar a j é a seguinte. O termo j é o horizonte de tempo de interesse do econometrista para saber o efeito de um choque hoje. A figura é capaz de mostrar o efeito do choque em cada período de tempo futuro para cada j , isto é, a dinâmica do choque ao longo do tempo.

A segunda observação é relacionada à hipótese de inexistência da condição inicial para y_0 . O que acontece se a condição inicial não puder ser definida? Nesse caso, podem-se iterar mais $(m+1)$ passos, para encontrar:

$$y_t = \mu \sum_{j=0}^{t+m} \phi^j + \phi^{t+m+1} y_{-m-1} + \sum_{j=0}^{t+m} \phi^j \epsilon_{t-j}.$$

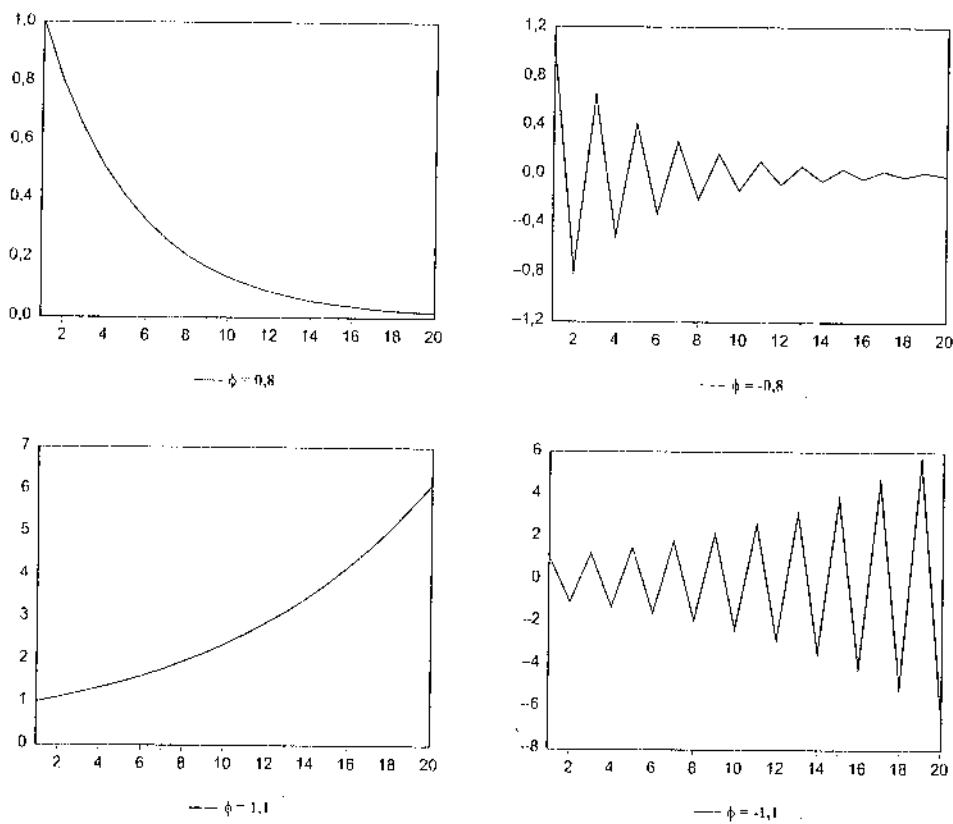


Figura A.2 Multiplicador dinâmico ou função resposta ao impulso: $\phi(j) = \phi^j$.

Ora, se $m \rightarrow \infty$ e se $|\phi| < 1$, então:

$$\phi^{t+m+1} \rightarrow 0 \quad \text{e} \quad \sum_{j=0}^{t+m} \phi^j \rightarrow \frac{1}{1-\phi},$$

resultando no seguinte:

$$y_t = \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}. \quad (\text{A-2})$$

Essa é a solução do problema, independentemente do valor inicial, y_0 , bastando, pois, dar os termos μ e ϕ . Além disso, é interessante ver aqui que o termo $\mu/(1-\phi)$ é a esperança não condicional de y_t . Isso pode ser visto simplesmente tomando a esperança de y_t , assumindo-se que $E(\epsilon_t) = 0, \forall t$. Outra observação é que a solução não é única, pois

$$y_t = a\phi^{t-1} + \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \quad (\text{A-3})$$

também é solução, dado $a \in \mathbb{R}$. Para ver isso, lembre-se de que $y_t = \mu + \phi y_{t-1} + \epsilon_t$, então ambos os lados da seguinte equação são idênticos, conforme o leitor poderá verificar se desenvolver a álgebra do lado direito:

$$a\phi^{t-1} + \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} = \mu + \phi \left[a\phi^{t-2} + \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-1-j} \right] + \epsilon_t.$$

Logo, como compatibilizar as equações (A-1) e (A-3)? É preciso encontrar um y_0 definido de tal forma que iguale ambas as equações. Tentativamente, considere a seguinte definição:

$$y_0 = a\phi^{-1} + \frac{\mu}{1-\phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{-j}$$

logo,

$$a = \phi y_0 - \frac{\phi\mu}{1-\phi} - \sum_{j=0}^{\infty} \phi^{j+1} \epsilon_{-j}.$$

Se y_0 é dado, a é totalmente determinado. E para ver se a definição está correta, é preciso substituir o valor encontrado na solução geral e verificar se a

solução recursiva gera o mesmo resultado. Portanto, substituindo a na equação (A-3), tem-se:

$$\begin{aligned}
 y_t &= \left(\phi y_0 - \frac{\phi \mu}{1 - \phi} - \sum_{j=0}^{\infty} \phi^{j+1} \epsilon_{-j} \right) \phi^{t-1} + \frac{\mu}{1 - \phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \Rightarrow \\
 &= \phi^t y_0 + \frac{1 - \phi^t}{1 - \phi} \mu + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} - \sum_{j=0}^{\infty} \phi^{t+j} \epsilon_{-j} \\
 &= \phi^t y_0 + \mu (1 - \phi^t) \sum_{j=0}^{\infty} \phi^j + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} - \sum_{j=t}^{\infty} \phi^j \epsilon_{-(j-t)} \\
 &= \phi^t y_0 + \mu \left(\sum_{j=0}^{\infty} \phi^j - \sum_{j=t}^{\infty} \phi^j \right) + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} - \sum_{j=t}^{\infty} \phi^j \epsilon_{t-j} \\
 y_t &= \phi^t y_0 + \mu \sum_{j=0}^{t-1} \phi^j + \sum_{j=0}^{t-1} \phi^j \epsilon_{t-j}.
 \end{aligned}$$

O resultado é idêntico à solução geral encontrada na equação (A-1). Conclusão: a solução de $y_t = \mu + \phi y_{t-1} + \epsilon_t$ não é única. Porém, para encontrar a solução mais geral, é preciso saber o valor de y_0 .

Um caso interessante surge quando $\phi = 1$. Sob essa condição, diz-se que y_t é integrado de ordem 1. Veja o porquê. Lembre-se de que:

$$y_t = \mu + y_{t-1} + \epsilon_t.$$

Conseqüentemente:

$$y_t = (t - 1) \mu + y_0 + \sum_{j=0}^{t-1} \epsilon_{t-j}.$$

Nota A.2 Observa-se que y_t é resultado do somatório dos resíduos. Como isso lembra integração em cálculo, dá-se o nome *integrado de ordem 1*.

Solução para Equações a Diferenças de Ordem 2

As idéias anteriores são interessantes para obter a intuição de equações a diferenças. O problema começa a ficar bem mais elaborado quando surge uma equação a diferenças de ordem maior do que 1, pois a complexidade algébrica torna derivações como as anteriores impraticáveis. Nesse caso, podem-se encontrar o

que se chama *solução homogênea* e *solução particular* de uma equação a diferenças. Encontradas essas soluções, a solução geral será:

$$\text{Solução geral} = \text{Solução homogênea} + \text{Solução particular}.$$

Para expressar essas idéias intuitivamente, considere uma equação a diferenças de ordem 2 determinística:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2}.$$

A solução particular é muito simples e sucede quando se faz, *particularmente*, $y_t = y_{t-1} = y_{t-2}$. Dessa forma:

$$y_t^p = \frac{\mu}{1 - \phi_1 - \phi_2}, \phi_1 + \phi_2 \neq 1,$$

em que o superíndice p representa particular.

A solução homogênea é obtida quando se ignora μ e, para cada defasagem de y_t , associa-se uma constante multiplicada por uma das raízes características¹ da equação a diferenças elevada a t . A combinação linear de cada uma das soluções homogêneas é também solução do problema. Isto é, tentativamente considere a seguinte solução para o problema:

$$y_t^h = a_1 \lambda_1^t + a_2 \lambda_2^t, \lambda_1 \neq \lambda_2.$$

Com isso, a solução geral do problema é:

$$y_t = y_t^p + y_t^h = \frac{\mu}{1 - \phi_1 - \phi_2} + a_1 \lambda_1^t + a_2 \lambda_2^t.$$

As constantes a_1 e a_2 serão determinadas a partir das condições iniciais. Para ver que essa solução funciona, substitua em $y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2}$.

A seguir, apresenta-se uma metodologia para obter a solução desse problema. A abordagem será informal, pois o objetivo é entender as razões que levam a obter séries estacionárias ou não estacionárias. O problema é encontrar os λ s que resolvem a equação a diferenças. Para isso, é preciso introduzir o operador defasagem.

¹ As raízes características são obtidas na sequência.

Operador Defasagem

Uma série temporal é uma sequência de observações indexada por sua data: $\{y_t\}_{t=1}^T = (y_1, y_2, \dots, y_T)$. Um operador de séries temporais transforma uma série temporal em uma nova série. Ele toma como insumo uma série, digamos, $\{x_t\}_{t=-\infty}^{\infty}$, e a transforma em outra, $\{y_t\}_{t=-\infty}^{\infty}$, segundo uma regra.

Exemplo A.2

O operador multiplicação é bem simples. Ele toma a série inicial e a multiplica por uma constante k , obtendo outra série multiplicada por k :

$$y_t = kx_t.$$

Em particular, o maior interesse deste trabalho é no operador defasagem que desloca a série original no tempo. Assim, se $y_t = x_{t-1}$, isso significa que o operador defasagem, L , foi aplicado à série original x_t :

$$Lx_t = x_{t-1}.$$

Note que:

$$L(Lx_t) = Lx_{t-1} = x_{t-2}.$$

Portanto, de forma geral:

$$L^j x_t = x_{t-j}.$$

Veja, a seguir, algumas propriedades do operador defasagem:

1. Seja $\mu \in \mathbb{R}$ uma constante, então $L\mu = \mu$;
2. $L(y_t + x_t) = y_{t-1} + x_{t-1}$;
3. $L(y_t x_t) = y_{t-1} x_{t-1}$.

É muito interessante observar o seguinte exemplo de uso do operador defasagem:

Exemplo A.3

Seja $y_t = \mu + \phi y_{t-1} + \epsilon_t$. Pode-se reescrever essa equação da seguinte forma:

$$(1 - \phi L) y_t = \mu + \epsilon_t.$$

Se $|\phi| < 1$, a progressão geométrica infinita vale para o operador defasagem, isto é:

$$\frac{1}{1 - \phi L} = 1 + \phi L + \phi^2 L^2 + \dots$$

Então, encontra-se

$$\begin{aligned} y_t &= \frac{\mu + \epsilon_t}{(1 - \phi L)} = \left(1 + \phi L + \phi^2 L^2 + \dots\right) (\mu + \epsilon_t) \\ &= \left(1 + \phi L + \phi^2 L^2 + \dots\right) \mu + \left(1 + \phi L + \phi^2 L^2 + \dots\right) \epsilon_t \\ &= \left(1 + \phi + \phi^2 + \dots\right) \mu + \left(1 + \phi L + \phi^2 L^2 + \dots\right) \epsilon_t \\ y_t &= \frac{\mu}{1 - \phi} + \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}. \end{aligned}$$

Isso foi exatamente o obtido com substituição recursiva.

Dito isso, é preciso retomar o problema original. Ignorando μ , pode-se usar o operador defasagem para descrever $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2}$ da seguinte forma:

$$\left(1 - \phi_1 L - \phi_2 L^2\right) y_t = 0.$$

O termo entre parênteses é usado para definir a *equação característica* do problema.

Definição A.3 *Seja a equação a diferenças definida por:*

$$\left(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p\right) y_t = 0.$$

A polinomial $(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$, quando igualada a zero, é chamada *equação característica da equação a diferenças de ordem p*.

Substituindo o operador defasagem pela incógnita z :

$$1 - \phi_1 z - \phi_2 z^2 = 0.$$

Uma forma análoga de ver a equação anterior é dividir ambos os lados por z^2 e definir $\lambda \equiv \frac{1}{z}$, para, então, encontrar as raízes da seguinte equação:

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0.$$

Achando as raízes dessa equação, encontra-se a solução do problema. Em seguida, dadas duas condições iniciais, podem-se encontrar os termos a_1 e a_2 da solução homogênea. Veja um exemplo bem simples para ter uma idéia do procedimento.

Exemplo A.4

(Enders, 2004) Seja a seguinte equação a diferenças:

$$y_t = 0,9y_{t-1} + 0,2y_{t-2} + 3,$$

dados $y_0 = 13$ e $y_1 = 11,3$. O objetivo é encontrar a solução geral para essa equação.

Primeiro, a solução particular:

$$y_t = \frac{3}{1 - 0,9 - 0,2} = 10.$$

Para a solução homogênea, deve-se montar a equação do segundo grau:

$$\begin{aligned}\lambda^2 - 0,9\lambda + 0,2 &= 0 \implies \\ (\lambda - 0,5)(\lambda - 0,4) &= 0.\end{aligned}$$

Então: $\lambda_1 = 0,5$ e $\lambda_2 = 0,4$. Portanto, a solução homogênea é:

$$y_t^h = a_1 0,5^t + a_2 0,4^t.$$

Para encontrar a_1 e a_2 , é preciso resolver o seguinte sistema:

$$\begin{aligned}a_1 + a_2 + 10 &= 13 \\ 0,5a_1 + 0,4a_2 + 10 &= 11,3.\end{aligned}$$

A solução do sistema é $a_1 = 1$ e $a_2 = 2$. Portanto, a solução geral é:

$$y_t = 0,5^t + 2 \times 0,4^t + 10.$$

Note, no exemplo anterior, que as raízes poderiam ser obtidas utilizando a fórmula de Báskara:

$$\begin{aligned}\lambda &= \frac{\phi_1 \pm \sqrt{\Delta}}{2}; \\ \Delta &= \phi_1^2 + 4\phi_2.\end{aligned}$$

Existem três casos possíveis a partir deste momento. O primeiro é quando o discriminante $\Delta > 0$, como no exemplo anterior. Nessa condição, há duas raízes reais distintas, $\lambda_1 = (\phi_1 + \sqrt{\Delta})/2$ e $\lambda_2 = (\phi_1 - \sqrt{\Delta})/2$, de modo que:

$$y_t^h = a_1 \lambda_1^t + a_2 \lambda_2^t.$$

Se $\Delta = 0$, $\lambda_1 = \lambda_2 = \phi_1/2$. Nesse caso, surge uma homogênea dada por $t(\phi_1/2)^t$, resultando na seguinte solução²:

$$y_t^h = a_1 \left(\frac{\phi_1}{2} \right)^t + a_2 t \left(\frac{\phi_1}{2} \right)^t.$$

Se $|\phi_1| < 2$, há convergência do sistema, já que o $(\phi_1/2)^t$ domina t . O sentido de “dominar” aqui quer dizer que, ainda que t esteja aumentando infinitamente, a força da redução desse aumento quando multiplicado por $(\phi_1/2)^t$, $|\phi_1| < 2$, é maior. Use uma planilha de cálculos para ver que isso acontece de fato.

Finalmente, existe um último caso; quando $\Delta < 0$, o resultado é um número complexo, da seguinte forma: $\lambda_1 = (\phi_1 + i\sqrt{-\Delta})/2$ e $\lambda_2 = (\phi_1 - i\sqrt{-\Delta})/2$. Assim, a solução é dada por:

$$y_t^h = \alpha R^t \cos(\beta + t\theta).$$

As constantes α e β são obtidas a partir das duas condições iniciais. Para existir convergência, é preciso que $|R| < 1$, em que

$$R = \sqrt{-\phi_2}; \quad \theta = \arccos \left(\frac{1}{2} \times \frac{\phi_1}{\sqrt{-\phi_2}} \right).$$

Exemplo A.5

(Enders, 2004) Seja $y_t = 1,6y_{t-1} - 0,9y_{t-2}$. Reescreva: $(1 - 1,6L + 0,9L^2)y_t = 0$. É preciso encontrar, pois, as raízes de:

$$\lambda^2 - 1,6\lambda + 0,9 = 0,$$

em que $\phi_1 = 1,6$ e $\phi_2 = -0,9$.

O discriminante é calculado por: $\Delta = (-1,6)^2 - 4 \times 0,9 = -1,04 < 0$.

Para obter R , calcule:

$$R = \sqrt{-\phi_2} = \sqrt{-(-0,9)} = 0,949.$$

O ângulo em radianos é dado por:

$$\begin{aligned} \cos(\theta) &= \frac{\phi_1}{2\sqrt{-\phi_2}} = \frac{1,6}{2 \times 0,949} = 0,8433 \Rightarrow \\ \theta &= 0,568. \end{aligned}$$

² Substitua esse resultado para verificar que funciona. Não é objetivo deste texto discutir completamente solução de equações a diferenças, por isso apenas mencionamos como seria a solução, caso as raízes da equação a diferenças de grau 2 fossem iguais.

Logo, a solução é:

$$y_t = \alpha (0,949)^t \cos(\beta + 0,568t).$$

A derivação da solução anterior está na próxima seção. Para entender a solução que se apresenta nessa situação, é preciso relembrar alguns conceitos de trigonometria usados para trabalhar com números complexos.

Trigonometria

Suponha um círculo de raio unitário em que o eixo vertical é denotado por y e o eixo horizontal, por x . Defina um ponto (x_0, y_0) na circunferência desse círculo. O ângulo em radianos que esse ponto faz com o eixo horizontal é θ . Logo, pode-se afirmar:

$$\text{sen}(\theta) = y_0;$$

$$\text{cos}(\theta) = x_0.$$

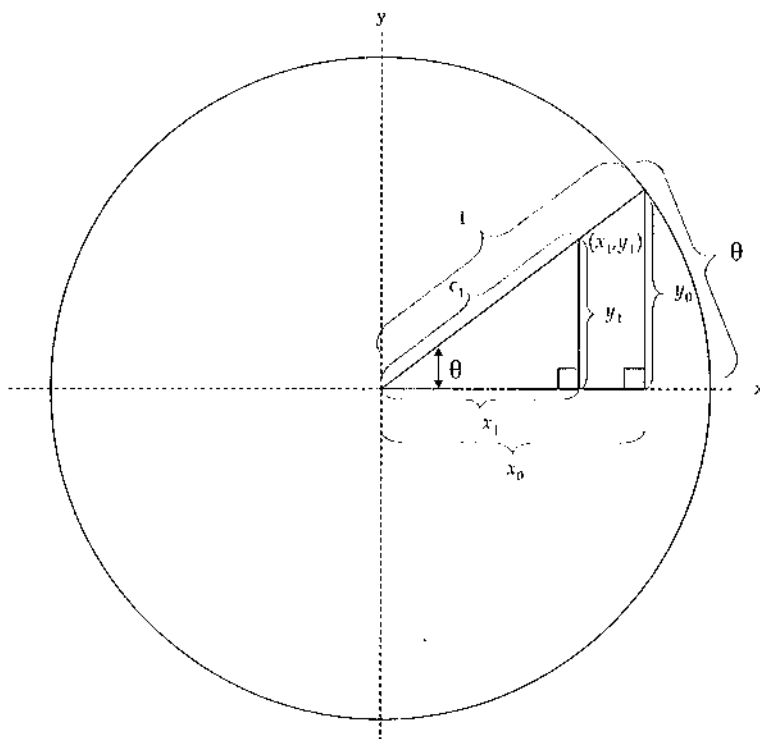


Figura A.3 Círculo unitário.

Admita um ponto (x, y) no segmento de reta entre a origem e (x_0, y_0) . Defina a distância entre a origem e (x, y) como R . Então, por semelhança de triângulos:

$$\frac{y}{R} = \frac{y_0}{1} \Rightarrow y = R \sin(\theta)$$

$$\frac{x}{R} = \frac{x_0}{1} \Rightarrow x = R \cos(\theta).$$

É fácil ver que, pelo teorema de Pitágoras:

$$R = \sqrt{x^2 + y^2}.$$

A Figura A.3 mostra graficamente o que foi verbalizado anteriormente.

Qualquer ponto no espaço (x, y) descrito como $(R \cos(\theta), R \sin(\theta))$ descreve as coordenadas polares desse ponto em termos de R e θ . Usam-se as mesmas idéias para recordar as propriedades dos números complexos na próxima seção.

Números Complexos

Definição A.4 O número imaginário, i , é tal que $i^2 = -1$.

Exemplo A.6

Se $x^2 = -1$, então $x = i$ ou $x = -i$, pois $i^2 = -1$ e $(-i)^2 = i^2 = -1$.

Definição A.5 O número complexo é sempre dado por $x + yi$, em que x e y são reais.

Pode-se representá-lo em uma circunferência assumindo o eixo x como o de números reais e o eixo y como o de números imaginários. O conjugado complexo de $x + yi$ é $x - yi$.

Com isso, pode-se entender que o módulo, R , de um número complexo é medido pela distância da origem ao número, ou seja:

$$R = |x + yi| = \sqrt{(x + yi)(x - yi)} = \sqrt{x^2 - y^2 i^2} = \sqrt{x^2 + y^2}.$$

Observe que $\cos(\theta) = x/R$ e que $\sin(\theta) = y/R$. Portanto, o número complexo $x + yi$ escrito em coordenadas polares é $R[\cos(\theta) + i \sin(\theta)]$.

Se essa circunferência tiver raio unitário, dentro dela estarão todos os complexos de módulo menor que 1. Além disso, observe que, se $y = 0$, então volta-se ao caso estudado anteriormente de números reais.

Agora, expanda por Taylor três funções: $\text{sen}(x)$, $\cos(x)$ e e^x , em torno de $x_0 = 0$.

$$\begin{aligned}\text{sen}(x) &= \text{sen}(0) + \cos(0)x - \frac{\text{sen}(0)x^2}{2!} - \frac{\cos(0)x^3}{3!} + \frac{\text{sen}(0)x^4}{4!} + \dots \\ &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \dots\end{aligned}$$

$$\begin{aligned}\cos(x) &= \cos(0) - \text{sen}(0)x - \frac{\cos(0)x^2}{2!} + \frac{\text{sen}(0)x^3}{3!} - \frac{\cos(0)x^4}{4!} + \dots \\ &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots\end{aligned}$$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

Assim, a fórmula de De Moivre é escrita como:

$$\begin{aligned}e^{i\theta} &= 1 + i\theta + \frac{(i\theta)^2}{2!} + \frac{(i\theta)^3}{3!} + \frac{(i\theta)^4}{4!} \dots \\ &= \left[1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \dots \right] + i \left[\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} \dots \right] \\ &= \cos(\theta) + i \text{sen}(\theta).\end{aligned}$$

Disso observe que:

$$\begin{aligned}\frac{(e^{i\theta} + e^{-i\theta})}{2} &= \frac{\cos(\theta) + i \text{sen}(\theta) + \cos(\theta) - i \text{sen}(\theta)}{2} = \cos(\theta); \\ \frac{(e^{i\theta} - e^{-i\theta})}{2i} &= \frac{\cos(\theta) + i \text{sen}(\theta) - \cos(\theta) + i \text{sen}(\theta)}{2i} = \text{sen}(\theta).\end{aligned}$$

Logo, pode-se exponenciar um número complexo usando coordenadas polares:

$$\begin{aligned}x + yi &= R [\cos(\theta) + i \text{sen}(\theta)] = R e^{i\theta} \implies \\ (x + yi)^k &= R^k e^{ik\theta} = R [\cos(k\theta) + i \text{sen}(k\theta)].\end{aligned}$$

Uma vez que se encontra $\Delta < 0$, a solução é λ_1 e λ_2 , tal que um é conjugado do outro: $\lambda_1 = (\phi_1 + i\sqrt{-\Delta})/2$ e $\lambda_2 = (\phi_1 - i\sqrt{-\Delta})/2$. Então, verifica-se que³:

$$\lambda_1 = R e^{i\theta} \text{ e } \lambda_2 = R e^{-i\theta}.$$

³ Lembrando que $\text{sen}(-\theta) = -\text{sen}(\theta)$ e que $\cos(-\theta) = \cos(\theta)$.

Logo a solução da homogênea é:

$$\begin{aligned} y_t^h &= a_1 \left(R e^{i\theta} \right)^t + a_2 \left(R e^{-i\theta} \right)^t \\ &= R^t \{ a_1 (\cos(t\theta) + i \operatorname{sen}(t\theta)) + a_2 (\cos(t\theta) - i \operatorname{sen}(t\theta)) \} \\ &= R^t [(a_1 + a_2) \cos(t\theta) + i(a_1 - a_2) \operatorname{sen}(t\theta)] \end{aligned}$$

Como y_t^h é um número real, é necessário que a_1 e a_2 sejam conjugados complexos. Seja, pois, $a_1 = (\alpha/2) [\cos(\beta) + i \operatorname{sen}(\beta)]$. Segue-se que:

$$\begin{aligned} a_1 + a_2 &= \frac{\alpha}{2} [\cos(\beta) + i \operatorname{sen}(\beta)] + \frac{\alpha}{2} [\cos(\beta) - i \operatorname{sen}(\beta)] = \alpha \cos(\beta); \\ a_1 - a_2 &= \frac{\alpha}{2} [\cos(\beta) + i \operatorname{sen}(\beta)] - \frac{\alpha}{2} [\cos(\beta) - i \operatorname{sen}(\beta)] = \alpha \operatorname{sen}(\beta) i. \end{aligned}$$

Portanto⁴:

$$\begin{aligned} y_t^h &= \alpha R^t [\cos(\beta) \cos(t\theta) - \operatorname{sen}(\beta) \operatorname{sen}(t\theta)] \in \mathbb{R} \implies \\ y_t^h &= \alpha R^t \cos(\beta + t\theta). \end{aligned}$$

As constantes α e β são obtidas a partir das duas condições iniciais. Para existir convergência, é preciso que $|R| < 1$. Lembrando que $R = \sqrt{x^2 + y^2}$, em que $x = \phi_1/2$ e $y = \sqrt{-\Delta}/2$, verifica-se, após algumas manipulações algébricas que:

$$R = \sqrt{-\phi_2}.$$

O ângulo θ é escolhido para satisfazer:

$$\cos(\theta) = \frac{x}{R} = \frac{\phi_1}{2\sqrt{-\phi_2}} \implies \theta = \arccos\left(\frac{\phi_1}{2\sqrt{-\phi_2}}\right).$$

Condições de Estabilidade

Enquanto o círculo unitário define as condições de estabilidade para todas as raízes características da polinomial que define a equação a diferenças, para o caso particular de uma equação a diferenças de ordem 2 podem-se expressar essas condições usando os próprios coeficientes da equação característica.

⁴ Lembrando que:

$$\begin{aligned} \operatorname{sen}(A \pm B) &= \operatorname{sen}(A) \cos(B) \pm \operatorname{sen}(B) \cos(A); \\ \cos(A \pm B) &= \cos(A) \cos(B) \mp \operatorname{sen}(A) \operatorname{sen}(B). \end{aligned}$$

Se $\Delta > 0$, é necessário que a maior raiz seja menor do que 1, e a menor raiz seja maior do que -1. Ou seja:

$$\begin{aligned}\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} < 1 &\Rightarrow \\ \sqrt{\phi_1^2 + 4\phi_2} < 2 - \phi_1 &\Rightarrow \\ \phi_1^2 + 4\phi_2 < 4 - 4\phi_1 + \phi_1^2 &\Rightarrow \\ \phi_2 < 1 - \phi_1.\end{aligned}$$

Analogamente,

$$\begin{aligned}\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} > -1 &\Rightarrow \\ 2 + \phi_1 > \sqrt{\phi_1^2 + 4\phi_2} &\Rightarrow \\ 4 + 4\phi_1 + \phi_1^2 > \phi_1^2 + 4\phi_2 &\Rightarrow \\ \phi_2 < 1 + \phi_1.\end{aligned}$$

No caso em que $\Delta = 0$, é necessário que $|\phi_1| < 2$, isto é, $-2 < \phi_1 < 2$. O caso 3 requer estabilidade quando $\sqrt{-\phi_2} < 1 \Rightarrow \phi_2 > -1$.

Em resumo, diz-se que um sistema é estável quando suas raízes encontram-se dentro do círculo unitário (Enders, 2004) e, no caso de um sistema de equações a diferenças de ordem 2, dentro do triângulo de estabilidade AOB, conforme mostra a Figura A.4.

Nota A.3 Note que, se fosse dada a solução do sistema $1 - \phi_1 z - \phi_2 z^2 = 0$, a condição de estabilidade requereria que as raízes z_1 e z_2 estivessem FORA do círculo unitário (Hamilton, 1994) já que $z = 1/\lambda$.

A solução de sistema de ordens superiores segue método análogo, pois a soma de uma solução particular com a combinação linear de cada uma das soluções homogêneas também é uma solução. Um sistema de ordem superior, a propósito, pode ser transformado em um sistema de equações a diferenças de ordem 1.

Em primeiro lugar, note que uma equação a diferenças de ordem 2 pode ser transformada em um sistema de ordem 1. Para isso, escreva:

$$\underbrace{\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix}}_{\equiv Y_t} = \underbrace{\begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix}}_{\equiv \Phi} \underbrace{\begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix}}_{\equiv Y_{t-1}} + \underbrace{\begin{bmatrix} \epsilon_t \\ 0 \end{bmatrix}}_{\equiv \epsilon_t}.$$

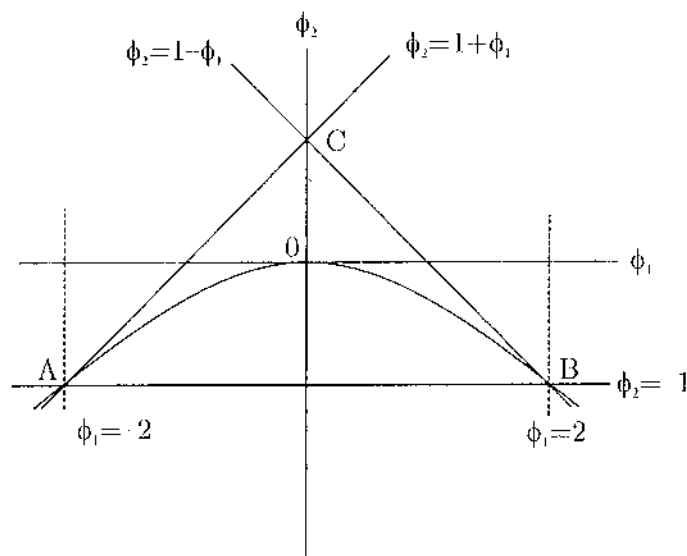


Figura A.4 Triângulo de estabilidade.

Compactamente:

$$Y_t = \Phi Y_{t-1} + \epsilon_t.$$

Definição A.6 Diz-se que a equação a diferenças estocástica $Y_t = \Phi Y_{t-1} + \epsilon_t$, em que Y_t é um vetor $n \times 1$, Φ é uma matriz $n \times n$ e $\epsilon_t \sim i.i.d. (0, \Sigma)$ é um vetor $n \times 1$ de variáveis aleatórias, é convergente se os autovalores de Φ estiverem dentro do círculo unitário.

Os autovalores de uma matriz qualquer, A , são obtidos pela solução do seguinte problema:

$$Ax = \lambda x,$$

ou, equivalentemente, do sistema:

$$zAx = x,$$

em que

λ e z são escalares recíprocos um do outro e x é um vetor não trivial.

Assim, desenvolvendo a primeira equação, tem-se:

$$\begin{aligned} Ax = \lambda x &\implies \\ (A - \lambda I)x &= 0, \end{aligned}$$

em que I é a matriz identidade.

Como x é não trivial, para satisfazer a igualdade é preciso que $(A - \lambda I)$ não tenha posto pleno. Isso quer dizer que o determinante dessa matriz tem de ser igual a zero. Portanto, deve-se ter:

$$\begin{aligned} 0 &= \left| \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 \\ 1 & -\lambda \end{vmatrix} \\ &= -(\phi_1 - \lambda)\lambda - \phi_2 = \lambda^2 - \phi_1\lambda - \phi_2. \end{aligned}$$

Nota A.4 *Veja que é uma equação idêntica ao do caso univariado.*

De modo equivalente, pode-se dizer que o sistema $(I - \Phi L)Y_t = \epsilon_t$, se as raízes características do determinante da matriz $(I - \Phi L)^5$ estiverem fora do círculo unitário. Observe:

$$\begin{aligned} 0 &= \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix} z \right| = \begin{vmatrix} 1 - \phi_1 z & -\phi_2 z \\ -z & 1 \end{vmatrix} \\ &= 1 - \phi_1 z - \phi_2 z^2. \end{aligned}$$

Exemplo A.7

Considere a seguinte equação a diferenças: $y_t = \phi_1 y_{t-1} + \phi_3 y_{t-3}$. Transforme-a em um sistema de equações a diferenças de ordem 1.

$$\begin{aligned} \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \end{bmatrix} &= \begin{bmatrix} \phi_1 & 0 & \phi_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \end{bmatrix} \Rightarrow \\ Y_t &= \Phi Y_{t-1}. \end{aligned}$$

É importante ter a idéia de transformar um modelo em um sistema de ordem 1 para facilitar o cálculo da função resposta ao impulso. Isso seria muito complicado usando um modelo univariado de ordem p . Por exemplo, pode-se escrever o choque hoje daqui a j períodos como:

$$\frac{\partial Y_{t+j}}{\partial \epsilon_t} = \Phi^j.$$

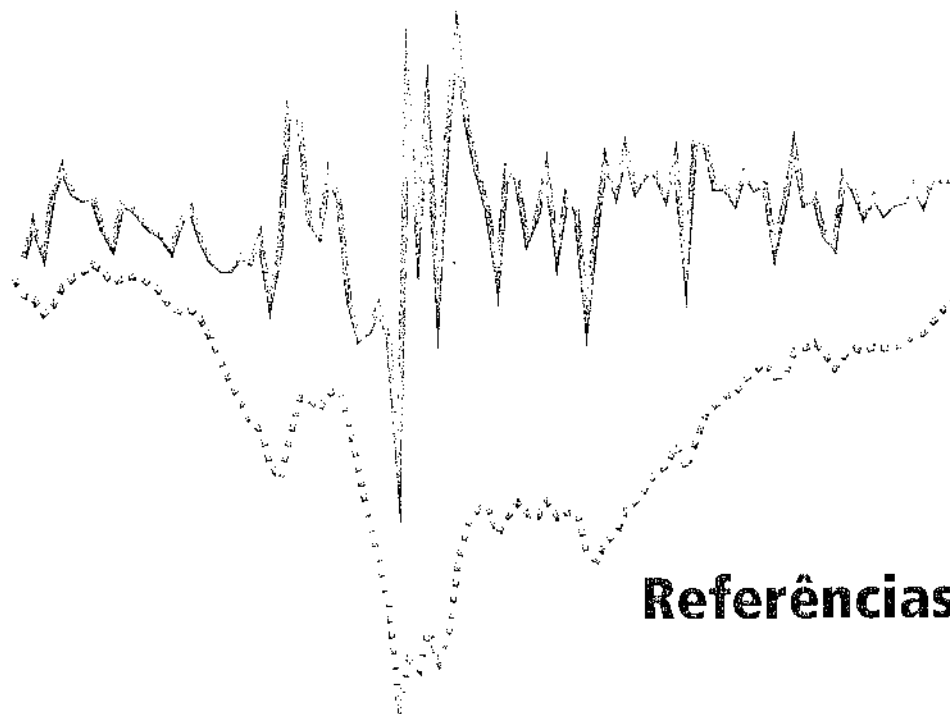
Basta substituir recursivamente para encontrar a solução apresentada.

⁵ Essa forma representa a segunda possibilidade de cálculo de autovalores.

Símbolos Usados

- AR** auto-regressivo
- $E(\cdot)$ operador esperança
- I_t conjunto de informação até $t - 1$
- H_0 hipótese nula
- H_1 hipótese alternativa
- $\text{var}(\cdot)$ operador variância
- $\text{cov}(\cdot)$ operador covariância
- N distribuição normal
- σ^2 variância
- ϵ_t ruído branco
- γ_{jt} coeficiente da autocovariância de defasagem j no período t
- ρ_{jt} coeficiente da autocorrelação de defasagem j no período t
- \mathbb{Z} conjunto de números reais
- μ média
- MA** média móvel
- i.i.d.* idêntica e independentemente distribuída
- $\psi(L)$ polinomial de médias móveis infinita
- $g(z)$ função geradora de autocovariância
- $h(\theta)$ espectro populacional
- $I(d)$ série integrada de ordem d
- I_m matriz identidade de dimensão m
- τ estatística teste Dickey-Fuller sem constante e sem tendência
- τ_μ estatística teste Dickey-Fuller com constante e sem tendência
- τ_τ estatística teste Dickey-Fuller com constante e com tendência
- Φ^i estatística teste Dickey-Fuller aumentado
- z_t estatística teste Philips-Perron para a estatística t sem constante e sem tendência
- $z_{t,\mu}$ estatística teste Philips-Perron para a estatística t com constante e sem tendência
- $z_{t,\tau}$ estatística teste Philips-Perron para a estatística t com constante e com tendência
- z_α estatística teste Philips-Perron para o coeficiente α sem constante e sem tendência

- $z_{\alpha,\mu}$ estatística teste Philips-Perron para o coeficiente α com constante e sem tendência
- $z_{\alpha,\tau}$ estatística teste Philips-Perron para o coeficiente α com constante e com tendência
- $\hat{\sigma}^2$ variância de longo prazo
- η estatística teste KPSS sem constante e sem tendência
- η_μ estatística teste KPSS com constante e sem tendência
- η_τ estatística teste KPSS com constante e com tendência
- Mz_α estatística teste Ng-Perron para o coeficiente α
- Mz_t estatística teste Ng-Perron para a estatística t
- MSB estatística teste Ng-Perron
- MP_T^{GLS} estatística teste Ng-Perron Point Optimal
- AIC estatística de Akaike
- BIC estatística de Schwarz
- HQ estatística de Hannan-Quinn
- FAC função de autocorrelação
- $FACP$ função de autocorrelação parcial
- \xrightarrow{d} convergência em distribuição
- Ψ vetor de parâmetros da estimação de máxima verossimilhança
- Ω matriz de covariância
- MSE** erro quadrático médio
- MAE** erro absoluto médio
- MAPE** erro absoluto percentual médio
- θ vetor de parâmetros no GMM
- $g(\cdot)$ vetor de funções reais no GMM
- w_t vetor de variáveis aleatórias no GMM
- W matriz de pesos dos momentos no GMM
- G_T jacobiano das condições de momento no GMM
- S variância de longo prazo dos momentos no GMM
- \xrightarrow{P} convergência em probabilidade
- Γ_j função de autocovariância multivariada
- RB ruído branco
- $VAR(p)$ vetor auto-regressivo de ordem p
- Σ matriz de covariância dos resíduos no modelo VAR
- vech** operador que empilha as colunas de uma matriz uma sobre as outras



Referências

- ALVES, Denisard C. O.; BUENO, Rodrigo D. L. S. Short-run, long-run and cross elasticities of gasoline demand in Brazil. *Energy Economics*, v. 25, n. 2, p. 191-199, 2003.
- ANDREWS, Donald W. K. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, v. 59, p. 817-858, 1991.
- BAILLIE, Richard T.; BOLLERSLEV, Tim. A multivariate generalized ARCH approach to modeling risk premia in forward rate markets. *Journal of International Money and Finance*, v. 9, p. 309-324, 1990.
- BALKE, Nathan; FOMBY, Thomas B. Large shocks, small shocks and economic fluctuations: outliers in macroeconomic time series. *Journal of Applied Econometrics*, v. 9, p. 181-200, 1994.
- BERA, Anil K.; HIGGINS, Matthew L. On ARCH models: properties, estimation and testing. In: OXLEY, Les et al. *Surveys in Econometrics*. Oxford: Blackwell, 1995.
- BEVERIDGE, Stephen; NELSON, Charles. A new approach to decomposition of economic time series into permanente and transitory components with particular attention to measurement of business cycle. *Journal of Monetary Economics*, v. 7, p. 151-174, 1981.
- BLAKE, David. The estimation of rational expectations models: a survey. *Journal of Economic Studies*, v. 18, n. 3, p. 31-70, 1991.
- BLANCHARD, Olivier; KAHN, C. M. The solution of linear difference models under rational expectations. *Econometrica*, v. 48, p. 1305-1311, 1980.

- BLANCHARD, Olivier J.; QUAH, Danny. The dynamic effects of aggregate demand and supply disturbances. *American Economic Review*, v. 79, n. 4, p. 655-673, 1989.
- BOLLERSLEV, Tim. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, v. 31, n. 3, p. 307-327, 1986.
- . Modelling coherence in the short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics*, v. 72, p. 498-505, 1990.
- BOLLERSLEV, Tim; CHOU, Ray Y.; KRONER, Kenneth F. ARCH modeling in finance. *Journal of Econometrics*, v. 52, n. 1, p. 5-59, 1992.
- BOLLERSLEV, Tim; ENGLE, Robert F.; NELSON, Daniel B. "Arch Models", ch. 49, In: ENGLE, Robert F.; MCFADDEN, Daniel L. *Handbook of econometrics*. v. 4. Amsterdam: Elsevier Science, 1994.
- BOLLERSLEV, Tim; ENGLE Robert F.; WOOLDRIDGE, Jeffrey M. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, v. 96, n. 1, p. 116-131, 1988.
- BOX, George E.; JENKINS, Gwilyn M.; REINSEL, Gregory C. *Time series analysis: forecasting and control*. 2. ed. Englewood Cliffs: Prentice Hall, 1994.
- BROCKWELL, Peter; DAVIS, Richard A. *Time series: theory and methods*. 2. ed. Nova York: Springer-Verlag, 1991.
- BUENO, Rodrigo D. L. S. *Soluções para o manual de econometria*. São Paulo: Atlas, 2001.
- . Ainda modelos GARCH. *Economia Aplicada*, v. 6, n. 2, p. 367-409, 2002.
- CAMPBELL, John Y.; PERRON, Pierre. Pitfalls and opportunities: what macroeconomists should know about unit roots. In: BLANCHARD, Olivier J.; FISCHER, Stanley (eds). *NBER Macroeconomics Annual*. The MIT Press, p. 141-201, 1991.
- CLARIDA, Richard; GALÍ, Jordi; GERTLER, Mark. Monetary policy rules and macroeconomic stability: evidence and some theory. *Quarterly Journal of Economics*, v. 115, n. 1, p. 147-180, 2000.
- COCHRANE, John H. *Asset pricing*. Revisado. Princeton: Princeton, 2005.
- DICKEY, David A.; FULLER, Wayne A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, v. 74, n. 366, p. 427-31, 1979.
- . Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, v. 49, n. 4, p. 1057-1073, 1981.
- DICKEY, David A.; PANTULA, Sastry G. Determining the order of differencing in autoregressive processes. *Journal of Business & Economic Statistics*, v. 5, n. 4, p. 455-61, 1987.

- DING, Zhuangxin; GRANGER, Clive W. J.; ENGLE, Robert F. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, v. 1, p. 83-106, 1993.
- DOLADO, Juan J.; JENKINSON, Tim. Cointegration and unit roots. *Journal of Economic Surveys*, v. 4, n. 3, p. 249-73, 1990.
- DOORNIK, Jurgen A.; HANSEN, Henrik. An omnibus test for univariate and multivariate normality. Oxford: manuscript, 1994.
- DUFOUR, J.-M.; RENAULT, E. Short run and long run causality in time series: theory. *Econometrica*, v. 18, p. 343-386, 1998.
- ELLIOT, Graham; ROTHENBERG, Thomas; SFOCK, James H. Efficient tests for an autoregressive unit root. *Econometrica*, v. 64, n. 4, p. 813-836, 1996.
- ENDERS, Walter. *Applied econometric time series*. 2. ed. Nova York: Wiley, 2004.
- ENGLE, Robert F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, v. 50, n. 4, p. 987-1007, 1982.
- . Dynamic conditional correlation: a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, v. 20, n. 3, p. 339-350, 2002.
- ENGLE, Robert F.; GRANGER, C. W. J. Cointegration and error correction: representation, estimation, and testing. *Econometrica*, v. 55, n. 2, p. 251-76, 1987.
- ENGLE, Robert F.; KRONER, Kenneth F. Multivariate simultaneous generalized ARCH. *Econometric Theory*, v. 11, p. 122-150, 1995.
- ENGLE, Robert F.; SHEPPARD, Kevin. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. *Journal of Business and Economic Statistics*, v. 20, n. 3, p. 339-350, 2002.
- ENGLE, Robert F.; LILIEN, David M.; ROBINS, Russel P. Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica*, v. 55, n. 2, p. 391-407, 1987.
- EViews 5 USER'S GUIDE. QMS: Irvine, 2004.
- FAMA, Eugene F.; FRENCH, Kenneth R. The capital asset pricing model: theory and evidence. *Journal of Economic Perspectives*, v. 18, n. 3, p. 25-46, 2004.
- FOX, A. J. Outliers in time series. *Journal of Royal Statistical Society*, n. 3, p. 350-363, 1972.
- GALLANT, A. Ronald. *An introduction to econometric theory*. Princeton: Princeton, 1997.
- GLOSTEN, L. R.; JAGANATHAN, R.; RUNKLE, D. On the relation between the expected value and the volatility of the normal excess return on stocks. *Journal of Finance*, v. 48, p. 1779-1801, 1993.
- GOURIÉROUX, Christian. *ARCH models and financial applications*. Nova York: Springer-Verlag, 1997.

- HALL, Alastair R. *Generalized method of moments*. Oxford: Oxford, 2005.
- HAMILTON, James D. *Time series analysis*. Princeton: Princeton, 1994.
- HANSEN, Lars P. Large sample properties of generalized method of moments estimators. *Econometrica*, v. 50, n. 4, p. 1029-1054, 1982.
- HANSEN, Lars P.; SARGENT, Thomas J. Formulating and estimating dynamic linear rational expectations models. In: LUCAS Jr., Robert & SARGENT, Thomas J. (eds.). *Rational expectation and econometric practice*. v. 1. Minneapolis: University of Minnesota, 1981.
- HANSEN, Lars P.; SINGLETON, Kenneth J. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *The Journal of Political Economy*, v. 91, n. 2, p. 249-265, 1983.
- HARRIS, Richard I. D. *Cointegration analysis in econometric modelling*. Englewood Cliffs: Prentice Hall, 1995.
- HATANAKA, Michio. *Time-series-based econometrics: unit roots and cointegration*. Nova York: Oxford, 1996.
- HEGERUD, Gustaf E. Specification tests for asymmetric GARCH. *Paper não publicado*. Department of Finance, Stockholm School of Economics, Jan./1997.
- HENTSCHEL, Ludger. All in the family nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics*, v. 39, p. 71-104, 1995.
- HEYNEN, Ronald C.; KAT, H. Volatility prediction: a comparison of the stochastic volatility, GARCH (1, 1), and EGARCH (1, 1) models. *Journal of Derivatives*, winter, p. 50-65, 1994.
- ISSLER, João V.; PIQUEIRA, Natália S. Estimating relative risk aversion, the discount rate, and the intertemporal elasticity of substitution in consumption for Brazil using three types of utility function. *Revista de Econometria*, v. 20, n. 2, 2002.
- JOHANSEN, Søren. Statistical analysis of cointegration vectors. *Journal of Economic Dynamic and Control*, v. 12, p. 231-254, 1988.
- . Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, v. 59, n. 6, p. 1551-1580, 1991.
- . Cointegration in partial systems and the efficiency of single-equation analysis. *Journal of Econometrics*, v. 52, n. 3, p. 389-402, 1992.
- . *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford, 1995.
- JOHNSTON, Jack; DINARDO, John. *Econometric methods*. 4. ed. Nova York: McGraw-Hill, 1997.
- KIM, Chang-jin; NELSON, Charles R. *State-space models with regime switching*. Cambridge: MIT, 1999.

- KWIATKOWSKI, Denis; PHILLIPS, Peter C. B.; SCHMIDT, Peter; SHIN, Yongcheol. Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics*, v. 54, p. 159-178, 1992.
- LEE, John H. H. A lagrange multiplier test for GARCH models. *Economic Letters*, v. 37, p. 265-271, 1991.
- LINTNER, John. The valuation and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economic and Statistics*, v. 47, p. 13-37, 1965.
- LUMSDAINE, Robin L. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1). *Econometrica*, v. 64, p. 561-573, 1996.
- LÜTKEPOHL, Helmut. *New introduction to time series analysis*. Berlin: Springer-Verlag, 2005.
- LÜTKEPOHL, Helmut; KRÄTZIG, Markus. *Applied time series econometrics*. Cambridge: Cambridge, 2004.
- MACKINNON, James G. Critical values for cointegration tests. In: ENGLE, R. F.; GRANGER, C. W. J. *Long-run economic relationship: readings on cointegration*. Oxford: Oxford, 1991.
- . Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, v. 11, p. 601-618, 1996.
- MACKINNON, James G.; HAUG, Alfred A.; MICHELIS, Leo. Numerical distribution functions of likelihood ratio tests for cointegration. *Journal of Applied Econometrics*, v. 14, p. 563-577, 1999.
- MADDALA, G. S.; KIM, In-Moo. *Unit roots, cointegration, and structural change*. Cambridge: Cambridge, 1998.
- MCCRORIE, J. Roderick; CHAMBERS, Marcus J. Granger causality and the sampling of economic processes. *Journal of Econometrics*, v. 131, p. 311-336, 2006.
- MINELLA, Andre; FREITAS, Paulo S.; GOLDFAJN, Ilan; MUINHOS, Marcelo K. Inflation targeting in Brazil. BCB, Working Paper Series 77, 2003.
- NELSON, Daniel B. Conditional heteroscedasticity in asset returns. *Econometrica*, v. 59, n. 2, p. 347-370, 1991.
- NELSON, Daniel B.; CAO, Charles Q. Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics*, v. 10, n. 2, p. 229-235, 1992.
- NEWKEY, Whitney; MCFADDEN, Daniel. Large sample estimation and hypothesis testing. In: ENGLE, Robert F.; MCFADDEN, Daniel. *Handbook of Econometrics*, v. IV. Elsevier Science B.V., p. 2113-2245, 1994.
- NEWKEY, Whitney; WEST, Kenneth. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, v. 61, p. 631-653, 1994.

- NG, Serena; PERRON, Pierre. Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, v. 69, n. 6, p. 1519-1554, 2001.
- PATTERSON, Kerry. *An introduction to applied econometrics: a time series approach*. Houndmills: MacMillan, 2000.
- PERRON, Pierre. *Lectures notes: Econ 513*. Princeton: Princeton University, 1990.
- . The great crash, the oil shock, and the unit root hypothesis. *Econometrica*, v. 57, p. 1361-1401, 1989.
- PERRON, Pierre; NG, Serena. Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies*, v. 63, p. 435-463, 1996.
- PHILLIPS, Peter; PERRON, Pierre. Testing for a unit root in time series regression. *Biometrika*, v. 75, n. 2, p. 335-346, 1988.
- PRIESTLEY, Maurice B. *Spectral analysis and time series*. v. 1/2. London: Academic Press, 1981.
- SAID, Said E.; DICKEY, David A. Testing for unit root in autoregressive-moving average models of unknown order. *Biometrika*, v. 71, n. 3, p. 599-607, 1984.
- SARGENT, Thomas J. *Macroeconomic theory*. 2. ed. San Diego: Academic Press, 1987.
- SCHOENBERG, Ronald. Simulation of bayesian posterior distributions of parameters of constrained models. *Paper não publicado*. The University of Washington, Aug. 1997.
- SCHWERT, G. W. Tests for unit roots: a Monte Carlo investigation. *Journal of Business and Economic Statistics*, v. 7, p. 147-159, 1989.
- SHARPE, William F. Capital asset prices: a theory of market equilibrium under conditions of risk. *The Journal of Finance*, v. 19, n. 3, p. 425-442, 1964.
- SHEPPARD, Kevin. *Generalized method of moments*, ch. 4. Manuscrito inédito, disponível em: <http://www.kevinshppard.com/>. Acessado em 2006.
- SILVERMAN, B. W. *Density estimation for statistics and data analysis*. Londres: Chapman; Hall, 1986.
- SIMS, Christopher; STOCK, James; WATSON, Mark. Inference in linear time series models with some unit roots. *Econometrica*, v. 58, p. 113-44, 1990.
- SIMS, Christopher. Macroeconomics and reality. *Econometrica*, v. 48, p. 1-49, 1980.
- SPANOS, Aris. *Statistical foundations of econometric modelling*. Cambridge: Cambridge, partes I e II, 1986.
- TAYLOR, S. *Modeling financial time series*. Nova York: John Wiley & Sons, 1986.
- TAYLOR, John B. Expectations, open market operations, and the changes in the federal funds rate. Federal Reserve Bank of Saint Louis, v. 83, n. 4, p. 33-58, 2001.

- TSAY, Ruey S. *Analysis of financial time series*. 2. ed. Nova York: Wiley, 2005.
- WHITE, Halbert. A heteroskedasticity consistent covariance estimator and a direct test for heteroskedasticity. *Econometrica*, v. 48, n. 4, p. 817-838, 1980.
- _____. Maximum likelihood estimation of misspecified models. *Econometrica*, v. 50, n. 1, p. 1-25, 1982.
- WOOLDRIDGE, Jeffrey M. *Introductory econometrics*. Mason: Thomson South-Western, 2006.
- ZAKOÏAN, J. M. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, v. 18, p. 931-944, 1994.

Bases de dados • eBooks • Coleções Digitais
Publicações periódicas acadêmicas • Livros impressos

Bases de Dados:

Academic OneFile

Bases de dados de periódicos eletrônicos, multidisciplinar e de perfil acadêmico que apresenta grande quantidade de artigos em texto completo. A interface é amigável e oferece tradutor on-line.

Informe Acadêmico

Coleção de periódicos em língua espanhola em todas as áreas do conhecimento e provenientes de diversas revistas publicadas pelas mais renomadas instituições acadêmicas da Iberoamérica.

Business & Management Practices

Valorosa ferramenta para estudo dos conceitos, processos, métodos e estratégia em administração de empresas.

Small Business Resource Center

Fonte informação para projetar, iniciar e operar pequenos negócios. Oferece informação sobre administração, financiamento, marketing, recursos humanos, franquias, contabilidade e impostos.

Business and Company Resource Center

Oferece perfis de empresas, marcas e produtos, preços das ações, relatórios de investimento, estatísticas industriais, notícias de indústria, artigos de revistas especializadas e análise de mercado.

eBooks:

Personal Money Management and Entrepreneurship
Economic Indicators Handbook
Everyday Finance: Economics
World Economic Prospects
Encyclopedia of Management
Encyclopedia of Small Business

Small Business Sourcebook
Advanced Project Management eBook Bundle
Encyclopedia of Business and Finance
Encyclopedia of Small Business
International Directory of Business Biographies

Livros impressos:

China and the Challenge of Economic Globalization
Information Technology and Economic Development
Everyday Finance: Economics
Economic & Business Handbook
Personal Money Management and
Entrepreneurship

Worldmark Encyclopedia of National Economies
Encyclopedia of Leadership
Management and Service
21st Century Management: A Reference Handbook
International Encyclopedia of Hospitality
Management

Para mais informações: www.galecengage.com
ou gale.brasil@cengage.com