

Apontamentos de Econometria Aplicada

João Sousa Andrade

Dezembro de 2001 - (Maio 2004)

Conteúdo

1	Apresentação do Modelo Geral Linear	7
1.1	Construção de Modelos	7
1.2	O Modelo de Regressão Clássico	10
1.2.1	As hipóteses do modelo	10
1.2.2	Alguns resultados algébricos	14
1.3	A Natureza do Processo Estocástico	15
1.3.1	Estacionaridade	15
1.3.2	Ergocidade	17
1.3.3	Conclusão	17
1.4	Regressões sem Sentido	17
1.4.1	Valores anormais ou sem sentido	19
1.5	Testes de aplicação mais corrente	20
1.5.1	Teste de normalidade dos erros	20
1.5.2	Teste LM	21
1.5.3	Teste LM de auto-correlação dos erros	21
1.5.4	Teste ARCH	22
1.5.5	Teste de especificação (Regression Specification Test)	22
1.5.6	Teste de Ljung-Box	22
1.5.7	Teste de Chow	23
1.5.8	Critérios de informação	24
1.5.9	Testes de restrição de coeficientes de regressão	24
2	Raízes Unitárias e Estacionaridade	27
2.1	Introdução	27
2.1.1	O operador de desfasamentos	27
2.1.2	Variáveis estacionárias em economia	28
2.2	Testes de Dickey-Fuller e Phillips-Perron	30
2.2.1	Procedimentos disponíveis no RATS	31
2.2.2	Diferentes comportamentos das séries	32
2.3	O Estudo de Raízes Unitárias em Séries Trimestrais	35
2.3.1	A metodologia HEGY	35

2.3.2	O procedimento do RATS	36
2.4	O Ratio de Cochrane e a Persistência das Inovações	36
2.5	Avaliação <i>ad hoc</i> de Processo AR	38
2.6	Teste de Perron a Alterações Estruturais	39
2.7	A Hipótese Nula de Estacionaridade	41
2.8	Exemplos de Aplicação no RATS	43
2.8.1	Séries com raiz unitária	43
2.8.2	Séries estacionárias	55
2.8.3	Séries com uma ruptura estrutural	61
2.8.4	Exemplo de séries trimestrais	72
3	Cointegração, Equilíbrio e Ajustamento	81
3.1	Exemplos Económicos	82
3.1.1	Procura de moeda	82
3.1.2	Função consumo	83
3.1.3	Eficiência em mercados cambiais	83
3.1.4	Paridade do poder de compra	84
3.1.5	Despesas do Estado	84
3.2	Equivalência do MCE e da Cointegração	85
3.2.1	Um cuidado adicional: ainda o caso de regressões espúrias	85
3.2.2	Equivalência MCE / Cointegração	86
3.3	Obtenção das Relações de Cointegração	88
3.3.1	Método de Engle-Granger	88
3.3.2	Cointegração à Johansen	89
4	Modelos VAR, VECM e Near-VAR(VECM)	121
4.1	Estabilidade de modelos auto-regressivos	122
4.1.1	Processo com dois desfasamentos	122
4.1.2	Processo com p desfasamentos	124
4.2	Apresentação de modelos VAR	124
4.2.1	Exemplo de modelo	125
4.2.2	Relação entre os erros dos modelos	125
4.2.3	Estabilidade do modelo	126
4.3	Identificação e estimação	127
4.3.1	A Sobre-parametrização dos modelos VAR	127
4.3.2	A escolha do número de desfasamentos	128
4.3.3	Apresentação alternativa de modelos VAR	129
4.3.4	Identificação e matriz de variâncias-covariâncias	130
4.3.5	Avaliação dos efeitos de choques e decomposição de Choleski	130
4.3.6	Um exemplo e uma regra prática	131

4.4	Decomposição da variância e análise de causalidade	132
4.4.1	Capacidade de previsão dos modelos VAR	132
4.4.2	Decomposição da variância dos erros	133
4.4.3	A exogeneidade por blocos de variáveis	134
4.4.4	Identificação do modelo e testes às restrições impostas	135
4.4.5	Decomposição histórica das séries	136
4.4.6	Programa para apresentação de alguns cálculos relaciona-	
	dos com um modelo VAR	136
4.5	Modelos VECM, Near-VAR e Near-VECM	143
5	Modelos ARCH	145
5.1	Apresentação geral da questão do ARCH	145
5.1.1	Variância condicional AR	146
5.1.2	Variância condicional ARMA	147
5.2	Apresentação do método de máxima verosimilhança	149
5.2.1	A utilização do RATS	151
5.3	Algumas observações adicionais sobre a pesquisa do tipo de	
	Variância condicionada	153
6	Métodos de Estimação em Painel Estático	157
6.1	Métodos de Estimação em Painel Estático	158
6.1.1	Teste aos efeitos individuais	159
6.1.2	Testes à variância dos erros individuais	159

Em 1993 a Texto Editora publicou os meus apontamentos de econometria, Andrade (1993), os quais constituíam o apoio a uma disciplina de opção na Licenciatura de economia e o apoio introdutório a uma disciplina de Mestrado.

Estes apontamentos pouco ou nada devem a esses outros. A única coisa em comum refere-se ao facto de continuar a utilizar, como apoio de cálculo, o programa da Estima, RATS. Na altura, o programa fazia a sua passagem da versão 2 para a 3 e agora encontra-se na versão 6.02.

Pretendo que estes novos apontamentos sejam uma base para o ensino da macro-economia aplicada. Não se trata de introduzir a econometria em estudos aplicados, mas de conhecer a econometria que deve começar por ser usada nos estudos aplicados de macro-economia a um nível não elementar. Entretanto, acabei por reproduzir no RATS os exemplos dados por Johston

e Dinardo na nova edição do livro. Alguns desses exemplos são hoje apresentados pela Estima a par com outros de alguns dos principais manuais de econometria.

Estes apontamentos começaram a ser feitos em Dezembro de 2001. O programa de texto utilizado na altura era o Scientific Word e depois de meados de 2003 passou a ser o WinEdt e, finalmente, o Kile como interface do LaTeX, em ambiente Windows e Linux.

O último capítulo apresentado acabou por não ser desenvolvido (Modelos ARCH). O curso do colega António Alberto Santos, sobre econometria das séries financeiras, retirou qualquer interesse à tentativa de desenvolver esse capítulo. Neste curso, iniciado em 2003, não é apenas a volatilidade determinista que é apresentada, este tipo de volatilidade “clássica” é mesmo criticada e estudada a volatilidade estocástica¹.

A econometria dos dados de painel em termos de análise estática é introduzida mas não a análise dinâmica desses mesmo dados.

¹Veja-se a sua página www4.fe.uc.pt/aasantos/analise_series_financeiras/analise_series_financeiras.htm.

Capítulo 1

Apresentação do Modelo Geral Linear

Neste capítulo iremos falar dos primeiros passos necessários à estimação de um modelo. Começaremos por ver como devemos encarar a ligação entre os elementos de análise e a construção de um modelo para ser estimado econometricamente. Em seguida trataremos da apresentação genérica do modelo clássico e das hipóteses que lhes estão subjacentes. Estas hipóteses são importantes porque determinarão a maior parte das exigências de testes às nossas bases de dados e aos nossos resultados.

Becker e Greene (2001) indicam um razoável número de livros de econometria, *undergraduate textbooks*, e também muitas páginas da internet onde se podem encontrar elementos de estudo de estatística e econometria. Para o objectivo aqui pretendido devemos juntar a esse livros o de Hayashi (2000), cuja apresentação do modelo clássico e a nova forma de o encarar acabámos por seguir.

Nos últimos meses os autores do RATS têm disponibilizado, na sua página, instruções adequadas aos exercícios que constam de um grupo razoável de livros de econometria.

1.1 Construção de Modelos

Em economia dispomos de encadeamentos lógicos entre variáveis que caracterizam o comportamento dos indivíduos ou de um dado conjunto de indivíduos. Estas variáveis podem, ou não, ser sujeitas a quantificação. No caso de o serem, é ainda possível que através de funções matemáticas acabemos por representar aqueles comportamentos. Trata-se do caso de uma

função consumo agregado, ou de uma função da procura de um bem por parte de um indivíduo, respectivamente de

$$C = C_A + c \cdot Y_d \quad (1.1)$$

ou de

$$Q_D = \beta_0 \cdot P^{-\beta_1} \quad (1.2)$$

Mas como sabemos, são muitos os casos em que temos relações entre variáveis em que apenas conhecemos com alguma certeza os sinais das derivadas e não a relação precisa entre elas. Até mesmo para a procura de moeda, podemos fazer

$$M_D = L(i, P, \dot{P}, Y^R) \quad (1.3)$$

onde $\frac{\partial M_D}{\partial i} < 0$, $\frac{\partial M_D}{\partial P} > 0$, $\frac{\partial M_D}{\partial \dot{P}} < 0$ e $\frac{\partial M_D}{\partial Y^R} > 0$.

E com esta forma geral procuramos ainda ter em conta que a taxa de juro, i , pode ser nominal ou real, taxa simples ou factor de capitalização. E mais, se retivermos a taxa de juro real não podemos usar (em geral) logaritmos. Os valores negativos desta taxa, para alguns períodos impossibilita-nos de o fazer.

Onde pretendemos chegar? Que a análise económica conduz-nos à verificação empírica de algumas das suas hipóteses de comportamento. Mas deixa à investigação propriamente empírica a investigação do tipo de relações, que podem existir entre variáveis, que traduzem esses comportamentos. De outra forma, a análise económica de natureza dedutiva não apresenta como produto final relações funcionais bem determinadas, para as quais apenas haverá que determinar os valores dos parâmetros. Por essa razão, o trabalho empírico envolve duas tarefas: - conhecer o tipo de relação funcional existente entre as variáveis; e - determinar os parâmetros que fazem parte dessa relação. Devemos retomar a lição de Neyman, Pearson e Wald (McCloskey e Ziliak (1996)¹) que a significância estatística não pode eliminar a significância económica. Um modelo só deverá ser retido se tiver um sentido económico, não apenas no que respeita aos sinais dos seus parâmetros, mas também no que respeita à grandeza desses parâmetros.

Pelo que acabámos de dizer acaba por ser bastante interessante a observação de Chow (1983) acerca de a econometria ser afinal uma “arte”. Uma

¹Penso que aqueles autores viviam um período em que as confirmações empíricas económicas não eram “arma de arremesso” entre escolas económicas. Quando isto acontece, a significância estatística tende a dominar a económica.

“arte” que não fez sentido sem os olhos que a possam apreciar, a “análise económica”, e uma “arte” que pode dar um significado preciso ao nosso raciocínio hipotético-dedutivo.

Normalmente o economista usa e abusa de modelos com elasticidades constantes, como o que está expresso na equação (1.2). Esse modelo tem uma representação linear bastante simples em logaritmos. O mesmo é dizer que “gostamos” de modelos lineares de variáveis previamente transformadas em logaritmos. Mas ficarmos por relações deste tipo é limitarmos excessivamente a nossa imaginação e capacidade. Por outro lado, e como iremos ver mais à frente, não podemos esquecer algumas características que são exigidas às observações que reunimos nas nossas variáveis e que nos obrigam a obter diferenças dos seus valores.

Exemplifiquemos o que queremos dizer com esta última observação. Só aparentemente um modelo de elasticidade constante é equivalente a um modelo linear em termos de taxas de crescimento. O primeiro será escrito em termos de logaritmos de observações em níveis e o segundo em primeiras diferenças desses logaritmos. Mas eles são profundamente diferentes do ponto de vista das variáveis que usamos... O economista sem conhecimentos de econometria não imagina as diferenças entre uma e outra forma de apresentar o que afinal até é o “mesmo”.

A formulação de uma relação económica é o primeiro passo na nossa investigação empírica. Mas este passo nunca deve ser tomado como irreversível pelo investigador. Por um lado existem alguns testes estatísticos que nos ajudam a saber se estamos perante um modelo bem especificado e por outro, temos a obrigação de testar (e tentar) várias formulações que não ofendam a teoria económica, pelo contrário que a confirmem.

Que não nos iludamos, o nosso trabalho econométrico é apenas de puro confirmacionismo. Mesmo que saíamos desse confirmacionismo e possamos descobrir algo de novo, apenas se esse “novo” for coerente com a análise económica o devemos reter. Descobrir o que não sabemos explicar ou compreender não faz sentido. Ao aceitarmos estas regras de jogo estamos afinal a reduzir os efeitos negativos do uso do método indutivo que a econometria implica. Mas ao fazermos isto estamos também a contribuir para imunizar as nossas teorias.

Ao começarmos o nosso trabalho econométrico devemos ter uma relação em condições de ser testada econometricamente. é a partir dela que toda a nossa investigação se vai desenvolver. Se porventura ela for substituída por uma outra é porque não pudemos excluir toda uma série de requisitos que consideramos essenciais a uma boa relação funcional. é disto que iremos tratar.

Se pretendemos que uma proposição geral tenha carácter científico, ela deve ser posta à prova da experiência. Qualquer experiência é sempre singular e não podemos tentar uma experiência como “a experiência geral”. A econometria lida com numerosos resultados de experiências. E como, depois de Popper, é certo que nenhum conhecimento científico deve ser tomado como absolutamente certo, a natureza estocástica dos resultados econométricos parece-nos adequada à análise científica em economia.

Façamos, finalmente, uma referência a uma falácia específica à econometria: a falácia da *regressão para a média*. Esta falácia foi apresentada por Hotelling (1933) e recentemente lembrada por Friedman (1992). Este último autor chama-lhe mesmo “armadilha”. Referindo-se ao estudo onde se “provava” a tendência para a média de diferentes empresas, Hotelling disse “The real test of a tendency to convergence would be in showing a consistent diminution of variance, not among means of groups, but among individual enterprises”. O mesmo tipo de falácia pode ser encontrado nos exemplos dados por Friedman, para o crescimento económico e para a função consumo baseada em dados *cross-section*. Esta falácia pode ser encarada como um resultado indirecto da ausência de estudo da “significância” económica em econometria.

1.2 O Modelo de Regressão Clássico

Em economia não podemos esperar que os dados que dispomos resultem de experiências que podem ser repetidas. Por esse motivo devemos tomar as variáveis a estudar como sendo o resultado de processos aleatórios, como constituindo um processo estocástico.

1.2.1 As hipóteses do modelo

Passemos em revista as hipóteses em que se fundamenta o modelo.

Hipótese 1 *A linearidade do modelo*

Tomemos o modelo simples

$$Y_t = \beta_0 + \beta_1 \cdot x_{t1} + \beta_2 \cdot x_{t2} + \dots + \beta_k \cdot x_{tk} + \varepsilon_t \quad (1.4)$$

o membro direito é afinal a função a estimar. Os β são os coeficientes de regressão. A variável ε_t corresponde ao que se convencionou designar por erros do modelo. Esta variável concentra, para além dos erros que derivam

de estarmos a lidar com um processo estocástico, os efeitos de outras variáveis não presentes no modelo, as quais ignoramos quais sejam, ou, tendo uma ideia de quais são, não apresentam uma relação estável com a variável dependente.

O modelo a estimar tem a configuração linear representada em (1.4). Em termos matriciais podemos fazer

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_N \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1k} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ x_{N1} & x_{N2} & \cdot & \cdot & \cdot & x_{Nk} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}$$

ou seja, de forma compacta

$$\mathbf{Y} = \mathbf{X} \cdot \beta + \varepsilon \quad (1.5)$$

Os modelos a estimar baseiam-se nesta regra de linearidade. é certo que a econometria não se limita a modelos lineares, mas o modelo clássico e todas as suas características de distribuição estatística dos estimadores se referem ao modelo (1.5) acima. Também é certo que o modelo referido é já muitas vezes o resultado de transformações que nos levam de relações não lineares a uma relação que é linear. O exemplo mais conhecido, e fácil de perceber, é o de aquelas variáveis serem logaritmos de outras. Estaríamos neste caso a lidar com um modelo de elasticidades constantes, mas cuja estimação se resume à de um modelo linear.

Hipótese 2 *Exogeneidade estrita*

Esta hipótese traduz-se na seguinte exigência expressa pelo valor esperado dos erros condicionados aos valores de \mathbf{X}

$$E[\varepsilon_t/\mathbf{X}] = 0, \quad t = 1, 2, N \quad (1.6)$$

O valor esperado da variável aleatória erros, sendo dados os valores das variáveis independentes, é uma constante que toma o valor nulo. Uma outra forma de apresentar esta hipótese é estabelecer a independência de ε e cada uma das variáveis em \mathbf{X} .

Esta hipótese de igualdade estrita acarreta algumas consequências.

1ª.

A média dos erros (não condicionados) é nula. Como podemos ver

$$E[E[\varepsilon_t/\mathbf{X}]] = E[\varepsilon_t] \wedge E[\varepsilon_t/\mathbf{X}] = 0 \Rightarrow E[\varepsilon_t] = 0 \quad (1.7)$$

2ª.

As variáveis independentes são ortogonais com o erro

$$E[x_{jk}\varepsilon_t] = 0 \quad (1.8)$$

Senão vejamos,

$$\begin{aligned} E[\varepsilon_t/x_{jk}] &= E[E[\varepsilon_t/\mathbf{X}]/x_{jk}] = 0 \Rightarrow \\ E[x_{jk}\varepsilon_t] &= E[E[x_{jk}\varepsilon_t/x_{jk}]] = E[x_{jk}E[\varepsilon_t/x_{jk}]] = 0 \end{aligned}$$

As variáveis independentes são assim ortogonais com os erros de observações idênticas, $E[x_{tk}\varepsilon_t]$, e ainda de observações diferentes, para o qual bastará que em $E[x_{jk}\varepsilon_t]$, $j \neq t$.

3ª.

Como resulta da 2ª consequência, a covariância dos erros e das variáveis independentes é nula.

$$\begin{aligned} Cov(\varepsilon_t, x_{jk}) &= E[x_{jk}\varepsilon_t] - E[x_{jk}] \cdot E[\varepsilon_t] \\ &= E[x_{jk}\varepsilon_t] = 0 \end{aligned} \quad (1.9)$$

No caso de $j = t$, as variáveis independentes não são correlacionadas com os erros.

Não é difícil perceber que a segunda daquelas consequências é demasiado exigente para modelos com séries temporais: as variáveis independentes são ortogonais com os erros para observações contemporâneas, passadas ou futuras destes. Isto invalidaria os nossos modelos auto-regressivos. Felizmente os nossos estimadores possuem boas características mesmo com esta violação da exogeneidade estrita.

Veja-se o caso de um processo AR1,

$$Y_t = \beta \cdot Y_{t-1} + \varepsilon_t$$

que afinal nos conduz a

$$\begin{aligned} E[Y_t\varepsilon_t] &= E[(\beta \cdot Y_{t-1} + \varepsilon_t)\varepsilon_t] \\ &= \beta \cdot E[Y_{t-1}\varepsilon_t] + E[\varepsilon_t^2] \\ &= \beta \cdot 0 + E[\varepsilon_t^2] \\ &= E[\varepsilon_t^2] \end{aligned}$$

e assim, mesmo com $E[\varepsilon_t] = 0$, é natural ter $E[Y_t\varepsilon_t] \neq 0$!

Hipótese 3 *Não multicolinearidade*

Esta hipótese pode ser expressa de uma forma bastante simples, exigindo-se que a característica da matriz \mathbf{X} seja igual a k .

$$R[\mathbf{X}] = k \quad (1.10)$$

Hipótese 4 *Homocedasticidade e não correlação dos erros*

A homocedasticidade implica que

$$E[\varepsilon_t^2/\mathbf{X}] = \sigma^2 \quad (1.11)$$

ou, de outra forma, em termos da variância dos erros

$$\begin{aligned} \text{Var}[\varepsilon_t/\mathbf{X}] &= E[\varepsilon_t^2/\mathbf{X}] - E[\varepsilon_t/\mathbf{X}]^2 \\ &= E[\varepsilon_t^2/\mathbf{X}] \\ &= \sigma^2 \end{aligned}$$

A não correlação das observações dos erros pode ser expressa como

$$E[\varepsilon_i\varepsilon_j/\mathbf{X}] = 0 \quad (1.12)$$

ou ainda como

$$\text{Cov}[\varepsilon_i\varepsilon_j/\mathbf{X}] = 0$$

Resulta destas duas hipóteses, (1.11) e (1.12) que

$$E[\varepsilon\varepsilon'/\mathbf{X}] = \sigma^2 \cdot I_N \quad (1.13)$$

ou, de outra forma

$$\text{Var}[\varepsilon/\mathbf{X}] = \sigma^2 \cdot I_N$$

A matriz das variâncias dos erros é assim uma matriz diagonal, sendo o valor da diagonal dado pela variância dos erros.

Hipótese 5 *Distribuição Normal dos erros*

A distribuição Normal dos erros é importante para o conhecimento das distribuições estatísticas dos coeficientes de regressão

$$\varepsilon/\mathbf{X} \sim N(0, \sigma^2 \cdot I_N) \quad (1.14)$$

De acordo com a nossa primeira hipótese e (1.14), temos o resultado a que nos referimos imediatamente acima

$$(\mathbf{b} - \beta) / X \sim N\left(0, \sigma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}\right) \quad (1.15)$$

1.2.2 Alguns resultados algébricos

Vamos passar em revista algumas fórmulas úteis que resumem a estimação do modelo linear.

1) Resulta da minimização do quadrado dos erros da equação (1.5) que os coeficientes de regressão são obtidos de acordo com a fórmula (1.16.a)

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \cdot \mathbf{X}' \cdot \mathbf{Y} \quad (1.16.a)$$

2) Os valores estimados do modelo são representados por $\hat{\mathbf{Y}}$ e são obtidos através do modelo usando os coeficientes de regressão obtidos. Desta forma podemos obter os valores dos erros do modelo, (1.16.b) e (1.16.c)

$$\hat{\mathbf{Y}} = \mathbf{X} \cdot \mathbf{b} \quad (1.16.b)$$

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} \quad (1.16.c)$$

3) A soma do quadrado dos erros *RSS* (*residual sum of squares*), (1.16.d), permite-nos calcular o valor do desvio padrão da estimação (1.16.e)

$$RSS = \mathbf{e}' \cdot \mathbf{e} \quad (1.16.d)$$

$$SEE = \sqrt{s^2} = \sqrt{\frac{RSS}{N - k}} \quad (1.16.e)$$

4) O valor do coeficiente de correlação não-centrado, (1.16.f), e centrado, ou coeficiente de determinação, (1.16.g), vêm dados por

$$R_{un}^2 = 1 - \frac{\mathbf{e}' \cdot \mathbf{e}}{\mathbf{Y}' \cdot \mathbf{Y}} \quad (1.16.f)$$

$$R_c^2 = 1 - \frac{\mathbf{e}' \cdot \mathbf{e}}{\tilde{\mathbf{Y}}' \cdot \tilde{\mathbf{Y}}} \quad (1.16.g)$$

com $\tilde{\mathbf{Y}} = \mathbf{Y} - \mathbf{E}[\mathbf{Y}]$.

5) O valor da estatística F para $(k - 1)$ e $(n - k)$ graus de liberdade vem dado por

$$F_{k-1, n-k} = \frac{\tilde{\mathbf{Y}}' \cdot \tilde{\mathbf{Y}} / (k - 1)}{\mathbf{e}' \cdot \mathbf{e} / (\mathbf{N} - \mathbf{k})} \quad (1.16.h)$$

e destina-se a testar a hipótese nula dos coeficientes do modelo, para além da constante, serem nulos.

1.3 A Natureza do Processo Estocástico

Como já dissemos acima, estamos perante uma sequência de variáveis aleatórias que seleccionámos. A partir desta escolha vamos procurar obter o modelo que as relaciona. Tomemos os valores do produto e da inflação de 1951 a 2000. Dispomos de 50 observações. Estas observações caracterizam o “produto” e a “taxa de inflação” em geral, ou apenas podem caracterizar o “produto” e a “taxa de inflação” daquele período? Outros períodos históricos e outras seriam as séries obtidas? Como obter outra amostra de 50 observações daqueles valores do produto e da taxa de inflação?

Poderemos tomar o valor de 1980 como representando a média dos valores do processo que gera o produto? E da taxa de inflação? Se a distribuição da taxa de inflação continuar inalterada² aquelas 50 observações são apenas valores da mesma distribuição. Se o fenómeno não for muito persistente³ cada observação conterà informação que não está disponível nas outras observações. Não é difícil perceber que estas questões que colocámos são mais complicadas de aplicar aos valores do produto que aos da inflação.

1.3.1 Estacionaridade

Passemos a esclarecer o que pretendemos com a característica de estacionaridade. Um processo $\{Z_t\}$ é fracamente estacionário (ou estacionário em covariância) se:

a) O valor esperado em qualquer momento não depender desse preciso momento. Ou seja, $E[Z_t]$, não depender de t .

²A variável for estacionária.

³Falamos de ergodicidade.

b) A variância dos seus valores, seja qual for o período a que nos reportemos, for constante e finita: $E [(Z_t - E[Z_t])^2] < \infty$

c) A covariância entre diferentes observações depender do intervalo entre essas observações mas não do período em que as tomamos: $Cov [Z_t, Z_{t-k}] = \gamma_k$. De outra forma:

$$Cov [Z_{t+121}, Z_{t+100}] = \gamma_{21} = Cov [Z_{t+22}, Z_{t+1}] \neq Cov [Z_{t+25}, Z_{t+2}].$$

Exemplos simples de variáveis estacionárias são os (i) erros, a que já fizemos referência, $\varepsilon \sim IID$, e que apresentam a ausência de relação auto-regressiva; e (ii) uma constante, que já apresenta o máximo de dependência auto-regressiva.

Uma forma prática de determinar se a variável é estacionária

Sem procurarmos avançar de imediato no estudo das características de estacionaridade podemos desde já apresentar uma regra prática e uma função com comportamento típico. A primeira consiste em dividir a amostra, por exemplo em duas partes, Z_{1t} e Z_{2t} , e ver se as respectivas médias e desvios padrão coincidem

$$\bar{Z}_1 = \bar{Z}_2 \wedge \sigma_{Z_1}^2 = \sigma_{Z_2}^2$$

Se tal acontecer existem fortes possibilidades de estarmos perante uma variável estacionária.

A segunda baseia-se na obtenção da função de auto-correlação da amostra

$$\rho_j = \frac{Cov(Z_t, Z_{t+j})}{Var(t)} \quad (1.17)$$

onde $\rho_0 = 1 \wedge \rho_{-k} = \rho_k$. Aquela expressão pode ser obtida através de

$$\rho_j = \frac{\sum_{t=1}^{N-j} (Z_t - \bar{Z}) \cdot (Z_{t+j} - \bar{Z})}{\sum_{t=1}^N (Z_t - \bar{Z})^2} \quad (1.17.a)$$

Ora, em caso de estacionaridade, o valor de ρ_j cai para 0 após $j = 1$. Por outro lado, sabemos que $\rho \sim N(o, \sigma_\rho) \wedge \sigma_\rho = \frac{1}{\sqrt{N}}$, pelo que podemos avaliar o grau de probabilidade de $\rho = 0$.

1.3.2 Ergodicidade

Se numa qualquer série Z , $(Z_{t+1}, Z_{t+2}, \dots, Z_{t+h})$ e $(Z_{t+k+1}, Z_{t+k+2}, \dots, Z_{t+k+h})$, com k suficientemente grande, forem independentes⁴, então o processo estacionário é dito que apresenta ergodicidade. Um importante teorema de uma série estacionária e ergótica é que se $\{Z_t\}$ representar essa variável, e $E[Z_t] = \mu$, então, para uma amostra N , teremos

$$\bar{Z}_N = \frac{1}{N} \sum_{t=1}^N Z_t \rightarrow \mu \quad (1.18)$$

a média daqueles valores converge assintoticamente para μ .

1.3.3 Conclusão

Se um processo estocástico for formado por variáveis que verifiquem as características de estacionaridade e ergodicidade então podemos com segurança tomar uma amostra dos valores desse processo e com base nela passarmos à representação económica (econométrica) pretendida. No exemplo acima, se a taxa de inflação for uma variável desse tipo, então podemos estar seguros de usar aquela amostra de valores para representar o fenómeno inflacionista - não temos necessidade de esperar mais 50 anos para vermos se o processo se repete ... -. Se a nossa amostra apenas correspondesse a um fenómeno particular a econometria seria de pouca valia na confirmação das nossas ideias (ou deduções) pela impossibilidade que teríamos de generalizar.

Como vemos, o economista deseja para as séries que vai usar características que lhe permitem mitigar⁵ o problema da “indução” em econometria. Ele pretende trabalhar com séries que mais não são que uma amostra do processo gerador dos valores dessas variáveis.

1.4 Regressões sem Sentido

Yule (1926) foi o primeiro autor a levantar a questão de regressões sem sentido e de regressões espúrias. No entanto, a influência do seu trabalho foi-se reduzindo ao longo do tempo e ficou mesmo esquecido até ao trabalho de Phillips (1986). Para uma apresentação dos diferentes tipos daquelas regressões veja-se Hendry e Juselius (2000), pp. 17-21. Mais à frente voltaremos à questão das regressões espúrias.

⁴Assintoticamente.

⁵E mesmo ultrapassar.

No caso de regressões sem sentido temos dois subtipos de regressões para as quais devemos estar atentos. No primeiro subtipo obtemos bons modelos que relacionam variáveis sem que essas relações expressem conteúdos económicos com significado. Obviamente que devemos evitar este tipo de resultados. As nossas munições para essa prevenção residem na análise económica. O segundo subtipo parte de variáveis, que pela sua natureza estatística, conduzem a falsas relações. Nestes dois casos, dizemos, hoje, que apesar dessas variáveis serem integradas, elas são mutuamente independentes. Vejamos este último caso.

Tomemos apenas duas variáveis, Y_t e X_t . Em modelos em que aquelas variáveis apresentam uma tendência ou não são estacionárias, os R^2 acabam por ser muito elevados e a estatística de Durbin-Watson apresenta valores muito reduzidos. Um dos problemas com tal regressão é que $\mathbf{X}'\mathbf{X} \cdot \frac{1}{N}$ não converge para um limite, e assim, o estimador dos mínimos quadrados não é convergente. Para além disso, as estatísticas habituais acabam por não ter as distribuições desejadas.

Tomemos dois casos. Começemos por admitir que as variáveis apresentam uma tendência. Assim

$$\begin{aligned} Y_t &= \alpha_1 + \alpha_2 \cdot t + \varepsilon_t \\ X_t &= \beta_1 + \beta_2 \cdot t + \mu_t \end{aligned} \tag{1.19a}$$

e que ε_t e μ_t não têm qualquer relação entre si. Se tomarmos os valores $Y_t - \alpha_1 = \tilde{Y}_t$ e $X_t - \beta_1 = \tilde{X}_t$, podemos ver que

$$\begin{aligned} E[\tilde{Y}_t] &= \alpha_2 \cdot t \\ E[\tilde{X}_t] &= \beta_2 \cdot t \end{aligned}$$

e finalmente

$$E[\tilde{Y}_t] = \frac{\alpha_2}{\beta_2} \cdot E[\tilde{X}_t]. \tag{1.19b}$$

O que acabámos de fazer pode ser traduzido no seguinte: tomemos duas variáveis sem qualquer relação entre elas, ε_t e μ_t , somemos a cada uma tendência determinista, acabamos por obter duas novas variáveis que irão apresentar uma relação óbvia entre elas. Como podemos saber se num modelo com duas variáveis (ou mais) estamos perante o caso aqui apresentado? Em primeiro lugar devemos saber se as variáveis apresentam, ou não, uma tendência determinista. Se esta hipótese for confirmada devemos retirar essa

tendência dessas variáveis e apenas depois passar à estimação definitiva do modelo?

Vejamos agora o caso em que temos duas variáveis que apresentam um processo *random walk* com *drift*

$$\begin{aligned} Y_t &= Y_{t-1} + \alpha_2 + \varepsilon_t \\ X_t &= X_{t-1} + \beta_2 + \mu_t \end{aligned} \quad (1.19c)$$

Façamos o desenvolvimento recursivo de uma delas

$$\begin{aligned} Y_1 &= Y_0 + \alpha_2 + \varepsilon_1 \\ Y_2 &= Y_0 + \alpha_2 \cdot (2) + (\varepsilon_1 + \varepsilon_2) \\ &\dots \end{aligned}$$

$$Y_t = Y_0 + \alpha_2 \cdot t + v_t \quad \left(v_t = \sum_{i=1}^t \varepsilon_i \right) \quad (1.19d)$$

Se tomarmos $Y_0 = X_0 = 0$, facilmente chegamos a uma relação idêntica a (1.19b). Nesta última equação encontramos a presença de uma tendência determinista ($\alpha_2 \cdot t$) e de uma tendência estocástica $\left(v_t = \sum_{i=1}^t \varepsilon_i \right)$.

Pelos exemplos que vimos, podemos ser levados a pensar que, nestes casos, retirando a tendência às séries envolvidas resolvemos os nossos problemas. Mas isso não é correcto, veja-se Hendry (1995), pp. 133-4.

1.4.1 Valores anormais ou sem sentido

Já vimos o problema de modelos sem sentido económico ou econométrico. Devemos também ter uma palavra para valores anormais de variáveis “sem sentido”. Se num modelo tivermos uma variável que apresenta valores fora do normal, exageradamente elevados ou reduzidos, estes valores acabam por dominar a estimação dos seus coeficientes de regressão.

Num modelo de séries temporais o problema pode ser resolvido, escolhendo um critério claro para a sua identificação, e substituindo esse valor por uma média centrada. Se a variável não for uma série temporal mas sim composta por valores de um “indivíduo” a questão é mais complicada. Não faz sentido substituir esse valor pela média de valores de outros indivíduos. Ou se exclui o indivíduo ou se altera a variável de forma a que esse problema

deixe de existir. Uma correcção possível, nalguns casos, é a passagem a uma variável de valores relativos.

Na resolução deste problema também podemos optar por uma opção automática, do tipo de excluir certas observações nas nossas estimações que verifiquem uma dada condição. Por exemplo, excluir as observações cujo valor absoluto do erro seja superior a um múltiplo do desvio-padrão dos erros.

1.5 Testes de aplicação mais corrente

Vamos apresentar um conjunto alargado de testes a diferentes hipóteses do modelo linear que se destinam a confirmar se as hipóteses retidas para a sua estimação são, ou não, confirmadas. Num estudo econométrico constitui uma boa prática a realização destes testes, assim como a sua divulgação.

1.5.1 Teste de normalidade dos erros

As estatísticas obtidas para o modelo linear geral estão dependentes desta característica. Se o modelo apresentar uma constante (intersecção) então a média dos erros é necessariamente nula. A ideia mais geral consiste em procurar ver se os valores dos erros apresentam (as)simetria e não são achatados⁶. Se tomarmos uma variável X , e admitirmos que temos N valores para a representar, podemos resumir um conjunto de dados importantes

Média	$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$
Variância	$s^2 = \frac{1}{N-1} \sum_{t=1}^N \left(X_t - \bar{X} \right)^2$
Desvio-padrão da média	$\frac{s}{\sqrt{N}}$
Estatística t para $\bar{X} = 0$	$\frac{\bar{X}\sqrt{N}}{s}$
Skewness	$Sk = \frac{N^2}{(N-1)(N-2)} \frac{m_3}{s^3}$
Estatística para $SK = 0$	$Sk \sqrt{\frac{(N-1)(N-2)}{6N}}$
Kurtosis	$Ku = \frac{N^2}{(N-1)(N-2)(N-3)} \frac{(N+1)m_4 - 3(N-1)m_2^2}{s^4}$
Estatística para $Ku = 0$	$Ku \sqrt{\frac{(N-1)(N-2)(N-3)}{24N(n+1)}}$
Jarque-Bera	$N \left(\frac{Ku^2}{24} + \frac{Sk^2}{6} \right) \sim \chi^2(2)$

Se tivermos uma distribuição Normal $N(0, 1)$, a média deverá ser nula, o desvio-padrão igual a 1 e a *skewness* e o *excesso de kurtosis* deverão ser ambos nulos. O teste de *Jarque-Bera* tem como $H(0)$ a distribuição Normal.

⁶Skewness e excesso de Kurtosis.

1.5.2 Teste LM

Pertence a um conjunto de testes bastante usados⁷. Baseia-se na comparação de dois modelos: um sem qualquer restrição (UR) e outro com restrição (R). Escrevamos os dois modelos, R e UR

$$Y_t = \beta_1^1 + \beta_2^1 X_{2t} + \dots + \beta_k^1 X_{kt} + \mu_t \quad (1.20.a)$$

$$Y_t = \beta_1^2 + \beta_2^2 X_{2t} + \dots + \beta_k^2 X_{kt} + \beta_{k+1}^2 X_{k+1t} + \dots + \beta_{k+m}^2 X_{k+mt} + \nu_t \quad (1.20.b)$$

O nosso objectivo consiste em saber se porventura os coeficientes de regressão $\beta_{k+1}^2, \dots, \beta_{k+m}^2$ são nulos (ou não). Do modelo (1.20.a) obtemos os valores estimados de $\hat{\mu}_t$ e fazemos a estimação do modelo seguinte

$$\hat{\mu}_t = \beta_1^* + \beta_2^* X_{2t} + \dots + \beta_k^* X_{kt} + \beta_{k+1}^* X_{k+1t} + \dots + \beta_{k+m}^* X_{k+mt} + \nu_t^* \quad (1.20.c)$$

O valor de $N \cdot R^2$ que resulta de (1.20.c) tem uma distribuição do χ^2 com m graus de liberdade. Se o seu valor for superior ao valor crítico escolhido (para uma dado grau de probabilidade), $\chi^2(m) > \chi_c^2(m)$, então a hipótese nula dos coeficientes $\beta_{k+1}^2, \dots, \beta_{k+m}^2$ deve ser excluída.

Trata-se de um teste muito robusto quando rejeita a hipótese nula daqueles coeficientes. Este teste acaba por ter bastantes aplicações. Vejamos para já uma das suas aplicações. Se tivermos um modelo como o da equação (1.20.a) podemos questionarmos se a inclusão de outras variáveis não melhoraria o nosso modelo, ou se afinal não estaríamos a excluir injustificadamente outras variáveis dele. Foi no fundo o que acabámos de ver ao comparar (1.20.a) com (1.20.b).

Vejamos de imediato outras aplicações deste tipo de teste.

1.5.3 Teste LM de auto-correlação dos erros

O problema que se nos coloca é saber se num modelo como (1.20.a) podemos excluir para os erros uma estrutura auto-regressiva de ordem r

$$\mu_t = \sum_{i=1}^r \alpha_i \mu_{t-i} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma_\epsilon) \quad (1.21.a)$$

O teste é em tudo semelhante ao feito em (1.20.c).

⁷Como o de Wald e o de Ratio de Verosimilhança.

$$\hat{\mu}_t = \beta_1^* + \beta_2^* X_{2t} + \dots + \beta_k^* X_{kt} + \beta_{k+1}^* \hat{\mu}_{t-1} + \dots + \beta_{k+r}^* \hat{\mu}_{t-r} + \nu_t^* \quad (1.21.b)$$

onde $N \cdot R^2 \sim \chi^2(r)$ na hipótese de não existir auto-correlação.

1.5.4 Teste ARCH

O famoso teste de *AutoRegressive Conditional Heteroscedasticity* baseia-se na possibilidade de a variância dos erros ter uma estrutura auto-regressiva. Assim, devemos testar a hipótese do seguinte comportamento dos erros dum qualquer modelo

$$\mu_t^2 = \sum_{i=1}^r \gamma_i \mu_{t-i}^2 + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma_\epsilon) \quad (1.21.c)$$

A estimação adicional

$$\hat{\mu}_t^2 = \beta_1^* + \beta_1^* \hat{\mu}_{t-1}^2 + \dots + \beta_r^* \hat{\mu}_{t-r}^2 + \nu_t^* \quad (1.21.d)$$

$N \cdot R^2$ apresenta a mesma distribuição que acima para a hipótese nula de ausência de processo ARCH.

1.5.5 Teste de especificação (Regression Specification Test)

Procuramos com este teste comparar um modelo com a alternativa de as potências do valor estimado pertencerem ao modelo. Normalmente ficamos pela potência de ordem 2. A questão que se coloca é se aquela nova variável, \hat{Y}_t^2 , faz ou não parte do modelo

$$Y_t = \beta_1^* + \beta_2^* X_{2t} + \dots + \beta_k^* X_{kt} + \beta_{k+1}^* \hat{Y}_t^2 + \nu_t^* \quad (1.21.e)$$

Na estimação de uma equação do tipo (1.21.b) o valor de $N \cdot R^2$ apresenta uma distribuição $\chi^2(1)$, na hipótese nula de correcta especificação do modelo original. Se a hipótese nula (H_0) for rejeitada então devemos pensar numa outra especificação do modelo em termos algébricos.

1.5.6 Teste de Ljung-Box

Trata-se de um teste de Box-Pierce adaptado a amostras pequenas, para valores elevados da ordem de auto-correlação, r . Se representarmos por $\hat{\rho}_k^2$ o coeficiente de correlação entre valores da mesma variável desfasados k períodos, a estatística de Ljung-Box virá dada por

$$Q(k) = N \cdot (N + 2) \cdot \sum_{i=1}^k \frac{1}{N - i} \cdot \hat{\rho}_i^2 \quad (1.21.f)$$

que tem uma distribuição $\chi^2(k)$ na hipótese nula (conjunta) dos coeficientes de correlação.

1.5.7 Teste de Chow

Este teste ganhou um outro valor após a crítica de Lucas à ausência de inclusão nos modelos económicos da modelação dos comportamentos de antecipações dos agentes. Estes últimos, ao reagirem a diferentes políticas provocarão alterações nos parâmetros de comportamento das variáveis da economia. Ou seja, devemos ter um cuidado particular com a estabilidade dos coeficientes do modelo que estimarmos.

Suponhamos que temos $N_1 + N_2$ observações e k parâmetros que nos permitem construir o seguinte modelo⁸

$$\mathbf{Y} = \mathbf{X} \cdot \beta + \varepsilon \quad (1.21.g)$$

Suponhamos também que acabamos por ter sérias dúvidas de que o modelo seja o mesmo para as primeiras N_1 observações e para as últimas N_2 observações. Por exemplo, sabemos que se registaram alterações importantes de um para o outro período.

Construindo os dois modelos

$$\begin{aligned} \mathbf{Y}_1 &= \mathbf{X}_1 \cdot \beta_1 + \varepsilon_1 \\ \mathbf{Y}_2 &= \mathbf{X}_2 \cdot \beta_2 + \varepsilon_2 \end{aligned}$$

queremos, no fundo, saber se $\beta_1 = \beta_2$. Para tal construamos o modelo UR (não restrito)

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (1.21.h)$$

Se retivermos o somatório do quadrado dos erros do modelo (1.21.g), RSS_R , e do modelo (1.21.h), RSS_{UR} podemos obter

$$\frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(N_1 + N_2 - 2k)} \sim F_{k, N_1 + N_2 - 2k}$$

que tem uma distribuição F na hipótese nula de igualdade dos coeficientes.

⁸Que passaremos a designar por R (restrito).

1.5.8 Critérios de informação

Não se trata agora de apresentar quaisquer testes estatísticos, mas antes o que é conhecido como critérios de informação que traduzem a qualidade de ajustamento de um modelo. Para estes indicadores a variável principal acaba por ser uma medida do valor absoluto dos erros. Eles giram à volta, da soma do quadrado dos erros, $RSS = \mathbf{e}'\mathbf{e}$. Vejamos os mais conhecidos.

Desvio-padrão da estimação (*see*)

$$see = \sqrt{\frac{RSS}{N - k}} \quad (1.21.i)$$

Critério de Akaike (*AIC*)

$$AIC = N \cdot \log(RSS) + 2 \cdot k \quad (1.21.j)$$

Critério de Schwarz (*BIC*)⁹

$$BIC = N \cdot \log(RSS) + k \cdot (\log(N)) \quad (1.21.k)$$

Comparação entre os diferentes critérios

Estes três critérios penalizam por ordem crescente o número de desfasamentos. Excluída que está a utilização do coeficiente de correlação para traduzir o nível de informação sobre a variável dependente de um dado modelo, restamos a escolha de um daqueles critérios. Talvez porque tenha a virtude de se situar ao centro, o *AIC* acaba por ser o mais utilizado.

1.5.9 Testes de restrição de coeficientes de regressão

A questão que agora colocamos é que temos um dado modelo¹⁰ e queremos saber se podemos admitir com um grau razoável de probabilidade certos valores para combinações lineares desses coeficientes. Lembremos que no se refere a um coeficiente de regressão isolado, esse teste é conduzido através da estatística *t*. O que normalmente acontece é que o teste *t* é feito (calculado) para a $H(0)$ de $E[\beta] = 0$. Logo, bastará substituir o valor esperado do coeficiente, para podermos fazer um novo teste.

Trata-se agora de impor como $H(0)$:

⁹Também designado Bayesian Information Criterion.

¹⁰Linear ou, mesmo, não linear.

$$\mathbf{R} \cdot \beta = \mathbf{r} \quad (1.21.l)$$

onde $\mathbf{R}_{q \times k}$, $q < k$ e $\mathbf{r}_{q \times 1}$.

Veamos três testes: o LM, o de Wald (W) e o do ratio de verosimilhança (LR). Estes três testes têm a seguinte relação entre eles: $W \geq LR \geq LM$ ¹¹. Uma nota importante deve ser feita a propósito da relação entre a estatística do $\chi^2(q)$ e a $F_{q,(N-k)}$. Quando $(N - k) \rightarrow \infty$, $q \cdot F_{q,(N-k)} \xrightarrow{\text{a}} \chi^2(q)$. Alguns autores indicam os valores em termos da estatística F e outros em termos da estatística do χ^2 .

Teste LM

A propósito de analisarmos algumas das hipóteses do modelo geral já vimos a lógica dos testes LM. Aplicada à hipótese expressa em (1.21.l) temos

$$LM = N \cdot \frac{RSS_R - RSS_{UR}}{RSS_R} \sim \chi^2(q) \quad (1.21.m)$$

Teste de Wald

Para este teste temos a expressão

$$W = N \cdot \frac{RSS_R - RSS_{UR}}{RSS_{UR}} \sim \chi^2(q) \quad (1.21.n)$$

Teste do Ratio de Verosimilhança

A expressão que nos conduz a esta estatística, também com distribuição do chi-quadrado, é

$$LR = N \cdot (\ln RSS_R - \ln RSS_{UR}) \sim \chi^2(q) \quad (1.21.o)$$

¹¹Embora sejam assintoticamente equivalentes.

Capítulo 2

Raízes Unitárias e Estacionaridade

Já falámos da necessidade de sabermos se as variáveis usadas num modelo são ou não estacionárias. No caso de o não serem a técnica econométrica a utilizar não pode ser a mesma que a utilizada quando as variáveis são estacionárias: as distribuições estatísticas dos estimadores não são as convencionais.

Como dissemos, o modelo clássico está construído para variáveis estacionárias. No que respeita às raízes unitárias, o livro de Hamilton (1994) passou a ser a referência obrigatória. No entanto, o livro de Maddala e Kim (1998), que apenas se dedica a esta questão, é bastante mais completo do que aquele. Veja-se também, em português, o livro de Marques (1998).

Começemos por apresentar um conceito útil, o de operador de defasamentos, e passaremos depois a justificar porque razão certas variáveis apresentam um interesse limitado para os economistas.

Os testes mais usados para o estudo da característica de estacionaridade das séries serão em seguida apresentados. Veremos os testes de ADF, Perron, HEGY, Cochrane e KPSS.

2.1 Introdução

2.1.1 O operador de defasamentos

O conceito de operador de defasamento é bastante usado em análise económica e econométrica pela “economia” que introduz na representação de fórmulas com defasamentos. A referência clássica continua a ser Dhrymes (1971).

Vejam os como o podemos caracterizar

$$\begin{aligned} L \cdot X_t &= X_{t-1} \\ L^h \cdot X_t &= X_{t-h} \\ \psi^{(k)}(L) \cdot \varepsilon_t &= (1 + \psi_1 \cdot L + \psi_2 \cdot L^2 + \dots + \psi_k \cdot L^k) \cdot \varepsilon_t = \\ &= \varepsilon_t + \psi_1 \cdot \varepsilon_{t-1} + \psi_2 \cdot \varepsilon_{t-2} + \dots + \psi_k \cdot \varepsilon_{t-k} \end{aligned}$$

Como vemos, a forma extremamente condensada da apresentação de um polinómio de defasamentos pode ser de grande utilidade.

2.1.2 Variáveis estacionárias em economia

Tomemos uma variável, Y , estacionária em covariância

$$Y_t = \mu + \varepsilon_t + \psi_1 \cdot \varepsilon_{t-1} + \psi_k \cdot \varepsilon_{t-k} + \dots = \mu + \psi(L) \cdot \varepsilon_t \quad (2.1)$$

onde $\sum_{j=0}^{\infty} |\psi_j| < \infty$ e as raízes de $\psi(z) = 0$ estão fora do círculo unitário.

A variável ε_t apresenta as características normais, desejadas para tal variável, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, com média nula e variância constante.

Uma variável como a apresentada em (2.1) é identificada simbolicamente por $I(0)$. Esta representação permite conhecer o número de diferenças temporais que foram feitas aos valores originais de uma série para que ela se tenha tornado estacionária. $I(d)$ significa que fizemos d diferenciações para obter uma variável do tipo (2.1). Ela será lida como variável integrada de ordem d .

A variável Y , acima descrita, apresenta ainda, como características

$$E[Y_t] = \mu \quad (2.1.a)$$

$$E[Y_{t+s}] = \hat{Y}_{t+s} = E[Y_{t+s}/Y_t, Y_{t-1}, \dots] \longrightarrow \mu \quad (2.1.b)$$

(i) uma média não condicionada constante; e (ii) o seu valor esperado em qualquer período de tempo, conhecida a história da série, tende para uma constante.

Estas características são de tal forma exigentes que levam ao desinteresse dos economistas por séries deste tipo. Acabamos por nos interessar por outros tipos de variáveis como as que passaremos a descrever em (2.2) e em (2.3). Trata-se de variáveis muito comuns no trabalho dos economistas.

Vejam os primeiro tipo de variáveis

$$Y_t = \mu + \delta \cdot t + \psi(L) \cdot \varepsilon_t \quad (2.2)$$

onde a variável t multiplicada por δ representa a tendência determinista. A variável Y é uma variável estacionária à volta de uma tendência (*trend*), como podemos ver melhor em baixo

$$(Y_t - \delta \cdot t) = \mu + \psi(L) \cdot \varepsilon_t$$

O segundo tipo de variáveis é dado por

$$(1 - L) \cdot Y_t = \delta + \psi(L) \cdot \varepsilon_t \quad (2.3)$$

onde Y apresenta uma raiz unitária. Devemos agora impor

$$\begin{aligned} \psi(1) &\neq 0 \\ \text{sendo } \psi(z) &= 1 + \psi_1 \cdot z^1 + \psi_2 \cdot z^2 + \dots \\ \text{com } \psi(1) &= \psi(z = 1) \end{aligned}$$

O significado de δ poderá ser visto mais à frente quando apresentarmos o ratio de Cochrane.

Acontece que estes dois tipos de variáveis têm comportamentos muito diferentes em termos de previsão, no que respeita ao erro quadrado médio MSE (*mean square error*)

$$\left[Y_{t+s} - \hat{Y}_{t+s/t} \right]^2.$$

Senão vejamos,

- para um processo estacionário em tendência, (2.2), o MSE tende para um valor finito quando o horizonte se distancia;
- para um processo de raiz unitária, (2.3), o MSE cresce de forma permanente com o horizonte de previsão.

No que respeita a efeitos de inovações, ε_t , eles são estacionários na primeira variável e os seus efeitos anulam-se; enquanto na segunda os seus efeitos são persistentes.

Vejamos o seguinte exemplo para uma variável y , que podemos identificar como sendo o produto

$$\phi(L) = 0.5 + \varepsilon_t$$

com

$$\phi(L) = 1 - 0.3 \cdot L - 0.12 \cdot L^2 + 0.11 \cdot L^3 + 0.08 \cdot L^4$$

Simplificando, para obtermos a equação do produto de longo prazo, podemos ver que a variação de uma unidade em ε_t arrasta uma variação total no produto de

$$\psi(1) = \frac{1}{\phi(1)} \cdot 0,5 = \frac{1}{1 - 0,3 - 0,12 + 0,11 + 0,08} \cdot 0,5 = 1,3 \cdot 0,5$$

onde $\phi(1) = \phi(L = 1)$.

Não restam dúvidas que os efeitos daquela inovação são persistentes no valor do produto ($\Delta Y_t = 1,3 \cdot \varepsilon_t$). Eles jamais desaparecem.

2.2 Testes de Dickey-Fuller e Phillips-Perron

Vejamos o tipo de testes mais popular, por vezes, conhecido apenas pelo nome dos primeiros (DF) Dickey-Fuller, e também o teste conhecido por ADF (*augmented DF*), devido à presença de termos desfasados da variável dependente na equação a ser estimada do teste DF, que se destina a corrigir a presença de auto-correlação dos erros¹.

Começemos em primeiro lugar pela apresentação dos testes e depois passemos às diferentes hipóteses a considerar para o comportamento das séries.

Tomemos um processo gaussiano AR(1)

$$Y_t = \rho \cdot Y_{t-1} + \varepsilon_t$$

com $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, $Y_{t=0} = 0$. O estimador OLS de ρ virá dado por

$$\hat{\rho} = \frac{\sum_{t=1}^N (Y_{t-1} \cdot Y_t)}{\sum_{t=1}^N Y_{t-1}^2}$$

Quando $|\rho| < 1$, e N suficientemente grande, teremos

$$\sqrt{N} \cdot (\hat{\rho}_N - \rho) \sim N(0, (1 - \rho^2))$$

Mas como estamos interessados na presença de uma raiz unitária naquela série devemos investigar se $\rho = 1$.

Se aquela distribuição se aplicasse, neste caso ($\rho = 1$), teríamos a sua variância reduzida a zero, ou seja, ela própria se resumiria a um único ponto.

¹Veja-se Dickey e Fuller (1979), Phillips (1987) e Phillips e Perron (1988).

Por esta razão, multiplicamos, na sua distribuição, o valor de $(\hat{\rho}_N - 1)$ não por \sqrt{N} , mas antes por N . Em consequência, obtemos

$$N \cdot (\hat{\rho} - 1) = \frac{\frac{1}{N} \cdot \sum_{t=1}^N (Y_{t-1} \cdot Y_t)}{\frac{1}{N^2} \cdot \sum_{t=1}^N Y_{t-1}^2} \quad (2.4)$$

A distribuição de $(\hat{\rho} - 1)$ não é a nossa conhecida *Normal*, mas antes um ratio que envolve uma $\chi^2(1)$, no numerador, e uma distribuição não estandardizada no denominador.

O teste nulo de $\rho = 1$ também pode ser conduzido via a estimação pontual habitual

$$t_N = \frac{\hat{\rho} - 1}{\hat{\sigma}_\rho} \quad (2.5)$$

Mas apesar de a estatística t ser calculada na sua forma habitual, ela não corresponde ao tipo de distribuição convencional quando $\rho = 1$ ².

O número de testes que giram à volta da raiz unitária não tem parado de aumentar. Citemos apenas o trabalho de Stock (1991) para a determinação dos intervalos de ρ a diferentes níveis de probabilidade e o de Bhargava (1986) para o teste de hipótese de passeio aleatório contra estacionaridade e de passeio aleatório contra um comportamento explosivo³.

2.2.1 Procedimentos disponíveis no RATS

No que respeita a procedimentos disponíveis para o RATS, o procedimento *adf.src* inclui a função geradora dos valores da estatística, enquanto que o procedimento *uradf.src* apenas tem valores da tabela próximos dos valores correspondentes às observações usadas no teste para os valores de probabilidade habituais⁴.

O procedimento *ppunit.src*, sem qualquer opção, apresenta como *Phillips-Perron Test*, o valor de $N \cdot (\hat{\rho} - 1)$ e com a opção *ttest* o valor de $(\hat{\rho} - 1)/\hat{\sigma}_\rho$.

O procedimento *adf.src* também indica o valor de $(\hat{\rho} - 1)/\hat{\sigma}_\rho$, mas com a indicação do valor crítico associado ao número de observações para o nível de 5%. Os testes mais completos são executados com o procedimento *uradf.src*.

²As tabelas apropriadas à verificação da hipótese nula encontram-se, por exemplo, nas pp. 762-4 de Hamilton (1994).

³Veja-se Andrade (2003) para uma aplicação de vários desses testes.

⁴A página www.estima.com é uma fonte importante de procedimentos para estudo de raízes unitárias e estacionaridade.

Neste procedimento, *Augmented Dickey-Fuller t-test* refere-se a $(\hat{\rho} - 1)/\hat{\sigma}_\rho$ e *Augmented Dickey-Fuller Z-test* refere-se a $N \cdot (\hat{\rho} - 1)$.

2.2.2 Diferentes comportamentos das séries

No que se segue iremos supor que não existe problema de auto-correlação dos erros. Se tal for o caso, devemos incluir desfasamentos da variável dependente para eliminar esse problema. E como saber se ele está presente e que foi eliminado? A minha preferência vai para um vulgar teste LM à auto-correlação dos erros. Se a $H(0)$ for excluída estamos perante a presença de auto-correlação e devemos incluir um desfasamento para a eliminar. Se o problema continuar a existir, $H(0)$ continuar a ser excluída, então devemos incluir mais um desfasamento para eliminar a auto-correlação. E assim sucessivamente ... até que o problema seja definitivamente resolvido. No RATS os procedimentos *adf.src* e *wradf.src* utilizam processos automáticos de eliminação do problema da auto-correlação através do teste LM.

Passemos em revista os três casos típicos de série com raiz unitária.

Série com *drift*

O processo em causa é caracterizado pelo seguinte comportamento

$$Y_t = \phi + \mu_t \text{ com } \mu_t = \rho \cdot \mu_{t-1} + \varepsilon_t \quad (2.6)$$

Para obtermos o modelo reduzido, determinamos aquela variável desfasada um período, Y_{t-1} , e subtraímos em (2.6)

$$\begin{aligned} Y_{t-1} &= \phi + \mu_{t-1} \\ Y_t - \rho \cdot Y_{t-1} &= \phi \cdot (1 - \rho) + \mu_t - \rho \cdot \mu_{t-1} \end{aligned}$$

pelo que chegamos a

$$Y_t = \phi \cdot (1 - \rho) + \rho \cdot Y_{t-1} + \varepsilon_t$$

e finalmente, para que possamos fazer o teste à hipótese de coeficiente nulo (raiz unitária)

$$Y_t - Y_{t-1} = \phi \cdot (1 - \rho) + (\rho - 1) \cdot Y_{t-1} + \varepsilon_t$$

que ainda pode ser escrito numa forma mais compacta

$$\Delta Y_t = \phi \cdot (1 - \rho) + (\rho - 1) \cdot Y_{t-1} + \varepsilon_t \quad (2.6.a)$$

Como podemos ver, se $\rho = 1$, a constante também deve tomar o valor nulo.

Série com *drift* e tendência

Suponhamos agora que temos o seguinte processo

$$Y_t = \phi + \gamma \cdot t + \mu_t \text{ com } \mu_t = \rho \cdot \mu_{t-1} + \varepsilon_t \quad (2.7)$$

A mesma metodologia de simplificação feita acima conduz-nos a

$$\begin{aligned} Y_{t-1} &= \phi + \gamma \cdot (t-1) + \mu_{t-1} \\ Y_t - \rho \cdot Y_{t-1} &= \phi \cdot (1-\rho) + \gamma \cdot t - \rho \cdot \gamma \cdot (t-1) + \mu_t - \rho \cdot \mu_{t-1} \\ Y_t - \rho \cdot Y_{t-1} &= \phi \cdot (1-\rho) + \gamma \cdot t - \rho \cdot \gamma \cdot t + \rho \cdot \gamma + \varepsilon_t \\ Y_t - \rho \cdot Y_{t-1} &= [\phi \cdot (1-\rho) + \rho \cdot \gamma] + \gamma \cdot (1-\rho) \cdot t + \varepsilon_t \end{aligned}$$

Passando a segunda parcela da esquerda para a direita e subtraindo o valor desfasado de Y , chegamos a

$$\Delta Y_t = [\phi \cdot (1-\rho) + \rho \cdot \gamma] + \gamma \cdot (1-\rho) \cdot t + (\rho-1) \cdot Y_{t-1} + \varepsilon_t \quad (2.7.a)$$

Vemos que agora, se tivermos $\rho = 1$, o coeficiente da tendência também se anula.

Tarefas de identificação

O processo mais conveniente para a identificação de uma raiz unitária numa série consiste em partir do caso geral de uma série com *trend* e testar se essa hipótese não pode ser anulada. Se excluirmos a presença de *trend* devemos passar ao caso de uma série com *drift*. Se a presença de *drift* puder ser excluída, então resta-nos o caso de uma série sem *drift*.

Vejam os diferentes casos que podemos encontrar. As tabelas e referências de página que faremos em baixo são do livro de Hamilton (1994).

Caso 1

Regressão estimada com base no processo

$$Y_t = \mu_t, \text{ com } \mu_t = \rho \cdot \mu_{t-1} + \varepsilon_t$$

Sendo o processo verdadeiro dado por

$$Y_t = \mu_t, \text{ com } \mu_t = \mu_{t-1} + \varepsilon_t$$

$N \cdot (\hat{\rho} - 1)$ tem a sua distribuição na Tabela B.5, Caso 1, p. 762

$(\hat{\rho} - 1)/\hat{\sigma}_\rho$ tem a sua distribuição na Tabela B.6, Caso 1, p. 763

Caso 2

Regressão estimada com base em

$$Y_t = \phi + \mu_t, \text{ com } \mu_t = \rho \cdot \mu_{t-1} + \varepsilon_t$$

Sendo o processo verdadeiro dado por

$$Y_t = \mu_t, \text{ com } \mu_t = \mu_{t-1} + \varepsilon_t$$

$N \cdot (\hat{\rho} - 1)$ tem a sua distribuição na Tabela B.5, Caso 2, p. 762

$(\hat{\rho} - 1)/\hat{\sigma}_\rho$ tem a sua distribuição na Tabela B.6, Caso 2, p. 763

O teste F para a hipótese conjunta, $\phi = 0 \wedge \rho = 1$ tem a sua distribuição na Tabela B.7, Caso 2, p. 764

Caso 3

Regressão estimada com base em

$$Y_t = \phi + \mu_t, \text{ com } \mu_t = \rho \cdot \mu_{t-1} + \varepsilon_t$$

Sendo o processo verdadeiro dado por

$$Y_t = \phi + \mu_t, \text{ com } \mu_t = \mu_{t-1} + \varepsilon_t$$

$(\hat{\rho} - 1)/\hat{\sigma}_\rho$ tem a sua distribuição aproximada pela $N(0, 1)$

Caso 4

Regressão estimada com base em

$$Y_t = \phi + \gamma \cdot t + \mu_t, \text{ com } \mu_t = \rho \cdot \mu_{t-1} + \varepsilon_t$$

Sendo o processo verdadeiro dado por

$$Y_t = \phi + \mu_t, \text{ com } \mu_t = \mu_{t-1} + \varepsilon_t$$

$N \cdot (\hat{\rho} - 1)$ tem a sua distribuição na Tabela B.5, Caso 4, p. 762

$(\hat{\rho} - 1)/\hat{\sigma}_\rho$ tem a sua distribuição na Tabela B.6, Caso 4, p. 763

O teste F para a hipótese conjunta, $\gamma = 0 \wedge \rho = 1$ tem a sua distribuição na Tabela B.7, Caso 4, p. 764

2.3 O Estudo de Raízes Unitárias em Séries Trimestrais

O economista trabalha muitas vezes com valores trimestrais, desta forma, pretende-se que os fenómenos de curto prazo sejam valorizados. Assim, o economista defronta com frequência a necessidade de conhecer as características de séries trimestrais quanto à presença de raízes unitárias (E. Ghysels e Noh (1994)), e não apenas as características de séries anuais. Como iremos ver, essas raízes podem ser de vários tipos o que vai ditar a forma da sua transformação para a obtenção de séries estacionárias.

Vejamos a metodologia conhecida, pelo nome dos seus autores, HEGY, S. Hylleberg, R. Engle, W. Granger e B. Yoo ⁵, e depois a sua aplicação no RATS.

2.3.1 A metodologia HEGY

Aqueles autores propuseram a análise de uma série gerada pelo seguinte processo

$$A(L) \cdot Y = \varepsilon_t$$

com $A(L)$ de ordem 4 de forma que

$$(1 - \alpha_1 \cdot L) \cdot (1 + \alpha_2 \cdot L) \cdot (1 - \alpha_3 \cdot i \cdot L) \cdot (1 + \alpha_4 \cdot i \cdot L) \cdot Y_t = \varepsilon_t \quad (2.8)$$

Fazendo o desenvolvimento da expressão acima como função de $\alpha_1, \alpha_2, \alpha_3$ e α_4 , através da aproximação em série de Taylor à volta do ponto $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1$, chegamos à equação

$$(1 - L^4) \cdot Y_t = \gamma_1 \cdot (1 + L + L^2 + L^3) \cdot Y_{t-1} - \gamma_2 \cdot (1 - L + L^2 - L^3) \cdot Y_{t-1} + (1 - L^2) \cdot (\gamma_5 - \gamma_6 \cdot L) \cdot Y_{t-1} + \varepsilon_t$$

que é uma fórmula em condições de ser estimada. Mas para o fazermos vamos simplificar, obtendo previamente as seguintes variáveis transformadas

$$\begin{aligned} Y_{1,t-1} &= (1 + L + L^2 + L^3) \cdot Y_{t-1} = Y_{t-1} + Y_{t-2} + Y_{t-3} + Y_{t-4} \\ Y_{2,t-1} &= (1 - L + L^2 - L^3) \cdot Y_{t-1} = Y_{t-1} - Y_{t-2} + Y_{t-3} - Y_{t-4} \\ Y_{3,t-1} &= (1 - L^2) \cdot Y_{t-1} = Y_{t-1} - Y_{t-3} \\ Y_{3,t-2} &= Y_{t-2} - Y_{t-4} \end{aligned}$$

⁵S. Hylleberg e Yoo (1990).

De forma que agora podemos fazer a estimação da equação

$$(1 - L^4) \cdot Y_t = \gamma_1 \cdot Y_{1,t-1} - \gamma_2 \cdot Y_{2,t-1} + \gamma_5 \cdot Y_{3,t-1} - \gamma_6 \cdot Y_{3,t-2} + \varepsilon_t \quad (2.8.a)$$

Esta última equação poderá incluir ainda uma constante, um *trend* e variáveis mudas sazonais e, não esqueçamos, deve ser estimada com inclusão de valores desfasados da variável dependente de forma a anular a auto-correlação dos erros.

2.3.2 O procedimento do RATS

Passemos a ver o significado daqueles coeficientes, com indicação do nome da variável incluída no procedimento *heggnewy.src*⁶, em termos das respectivas hipóteses nulas

$$\gamma_1 = 0 [= PI1]$$

raiz unitária não sazonal, $[(Y_t - Y_{t-1}) \sim I(0)]$

$$\gamma_2 = 0 [= PI2]$$

raiz unitária de frequência semi-anual, $[(Y_t - Y_{t-2}) \sim I(0)]$,
mas também $[(Y_t + Y_{t+1}) \sim I(0)]$

$$\gamma_5 = \gamma_6 = 0 [= PI3 = PI4 = (F34)]$$

raiz unitária de frequência anual, $[(Y_t - Y_{t-4}) \sim I(0)]$,
mas também $[(Y_t + Y_{t-2}) \sim I(0)]$.

Como vemos, não se trata apenas de estudar a presença de “uma” raiz unitária, mas da possibilidade de existência de tipos diferentes de raízes unitárias. E da não exclusão de um certo tipo de raiz unitária resultará uma forma adequada de diferenciação conducente à obtenção de uma nova variável estacionária.

2.4 O Ratio de Cochrane e a Persistência das Inovações

Antes de apresentarmos o ratio proposto por Cochrane (1988) e Campbell e Mankiw (1987), convém lembrar que se uma variável segue um processo do

⁶Incluimos os valores críticos para 5% das diferentes estatísticas e para 100 observações, como constam do artigo citado.

2.4. O RATIO DE COCHRANE E A PERSISTÊNCIA DAS INOVAÇÕES³⁷

tipo $Y_t = \beta_0 + \alpha \cdot t$, então as primeiras diferenças dessa variável $\Delta Y_t = \alpha$, traduzem um crescimento de Y a taxa constante. E da mesma forma podemos dizer que se uma variável registra um crescimento a taxa constante o seu comportamento pode ser traduzido por um processo de tendência determinista.

Tomemos a seguinte definição para a variável Y

$$\Delta Y_t = \alpha + \mu_t \text{ com } \mu_t = \sum_{j=0}^{\infty} \psi_j \cdot \varepsilon_{t-j} \quad (2.9)$$

Se Y_t for estacionária em tendência, uma inovação não terá efeitos permanentes sobre os seus valores futuros. Nesse caso, e em termos da sua derivada⁷

$$\lim_{s \rightarrow \infty} \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = \psi(1), \text{ com } \psi(1) = 0$$

Cochrane explorou essa característica e procurou medir a presença dos efeitos da inovação na série.

Começemos por ver a alteração em Y passados s períodos depois de uma inovação

$$Y_{t+s} - Y_t = \alpha \cdot s + \underbrace{\mu_{t+s} + \mu_{t+s-1} + \dots + \mu_{t+1}}_{\text{erros acumulados}}$$

e assim

$$\frac{Y_{t+s} - Y_t}{s} = \alpha + s^{-1} \cdot \underbrace{(\mu_{t+s} + \mu_{t+s-1} + \dots + \mu_{t+1})}_{\text{valor da média de } s \text{ observações de } \mu}$$

A segunda parcela do membro direito representa uma média de s observações aleatórias da variável μ , que apresenta a seguinte propriedade

$$\lim_{s \rightarrow \infty} s \cdot Var(\dots) = \sigma^2 \cdot [\psi(1)]^2$$

O estimador α vem dado, como é normal, por

$$\hat{\alpha} = N^{-1} \cdot \sum_{t=1}^n (Y_t - Y_{t-1})$$

A estimativa da variância da diferença entre o valor tomado por Y em $t + s$ e t virá dada por

⁷Ver notação já atrás utilizada.

$$\hat{J}_N(s) = \frac{\sum_{t=0}^{N-s} (Y_{t+s} - Y_t - \hat{\alpha} \cdot s)^2}{N}$$

Para valores de N suficientemente grandes obteremos

$$\begin{aligned} J(s) &= E[(Y_{t+s} - Y_t - \alpha \cdot s)^2] = E[(\mu_{t+s} + \mu_{t+s-1} + \dots + \mu_{t+1})^2] \quad (2.1) \\ &\implies \lim_{s \rightarrow \infty} \frac{1}{s} \cdot J(s) = \sigma^2 \cdot [\psi(1)]^2 \end{aligned}$$

Cochrane propôs-se calcular $\hat{J}_N(s)$ para diferentes valores de s . A característica interessante nesta estatística é que se a variável Y for estacionária, ou estacionária à volta de uma tendência, aquele valor cai para 0 para valores elevados de s . Esta estatística obriga a que os valores de s sejam bastante mais pequenos que os de n . O que por vezes é difícil nas séries que dispomos. Ao mesmo tempo, se Y é $I(1)$, aqueles valores dão-nos a importância dos efeitos permanentes da inovação à medida que o tempo passa.

Por vezes é usada a seguinte simbologia $A(1) = \psi(1)$ e $V(k) = J(s)$. Nesta nova simbologia, um valor de $A(1)=0,32$ para $k=100$ significa que 100 períodos após a existência de uma inovação de 1% a variável ainda retém 0,32% daquele valor.

2.5 Avaliação *ad hoc* de Processo AR

Pivetta e Reis (2002) apresentam um conjunto variado de alternativas para analisar a persistência de inovações numa série. Nestas alternativas sobressai um método de grande simplicidade: trata-se de obter uma representação auto-regressiva de uma dada variável e somar os coeficientes dos termos desfasados. Desta forma, ficamos com uma ideia do “arrastamento” dos valores de uma variável. Suponhamos que temos o seguinte processo

$$\phi(L) \cdot Y_t = \beta_0 + \epsilon_t$$

O valor de $\frac{1}{1 - \phi(1)}$, fácil de calcular, dá-nos o grau de persistência da variável Y . Ficamos assim com uma ideia da retenção das inovações sobre esta variável.

2.6 Teste de Perron a Alterações Estruturais

Vamos apresentar a análise de Perron (1989) e Perron (1997) em conjunto com o procedimento elaborado para o RATS por *G. Colletaz* e *F. Serranito*.

O objectivo deste teste consiste em testar a presença de uma raiz unitária em variáveis com tendência determinista. A hipótese nula consiste na presença de uma raiz unitária. Este teste procura dar resposta ao problema de termos séries que são estacionárias à volta de uma tendência, mas que sofreram um choque e, em consequência, somos levados a concluir, para a totalidade do período, que ela tem uma raiz unitária. E esta nossa dedução é obviamente incorrecta.

A forma de resolver o problema, proposta em 1997 por *P. Perron*, consiste em determinar endogenamente o período do choque. Para tal devemos ter uma ideia das consequências do choque, ou seja, do tipo de choque. Identifiquemos por “modelo” os diferentes tipos de alterações

modelo	IO1	alteração na intersecção
modelo	IO2	alteração na intersecção e na inclinação
modelo	AO	alteração na inclinação sem descontinuidade na tendência

No modelo a estimar, a presença dos desfasamentos das diferenças da variável dependente, de ordem k , será escolhida tendo em conta que o valor de t leva a excluir a hipótese nula do último coeficiente e a não excluir a do coeficiente seguinte $k + 1$ ⁸. O nível de significância que vamos reter será, como habitualmente, o de 5%.

O método de determinação endógena do período de ruptura obedece ao princípio de pesquisa do período que conduz ao valor de $t_{\alpha=1}$ mínimo⁹. Uma vez que o teste se destina a levantar a hipótese de uma série aparentemente de raiz unitária ser de facto estacionária, ele é bastante robusto na exclusão dessa hipótese de raiz unitária.

Em termos do procedimento do RATS, temos *perron97.src*, e para cada um daqueles modelos devemos fazer

```
@perron97(model=io1,signif=0,05);
@perron97(model=io2,signif=0,05); e
@perron97(model=ao,signif=0,05).
```

O output do procedimento deve ser lido com cuidado. Em primeiro lugar temos a informação da data de ruptura na série, ou data do choque nela verificado. Depois devemos olhar para o valor de t associado a *alpha* igual à unidade ($t_{\alpha=1}$). Um valor superior ao valor crítico (por exemplo 5%) significa

⁸Foi essa a escolha preferida por Colletaz e Serranito.

⁹O coeficiente α é o coeficiente do termo desfasado da variável a estudar.

a exclusão dessa hipótese nula e portanto a variável deve ser tomada como estacionária. Mas isto não significa que não devamos olhar para o valor do t associado ao coeficiente do *trend*. Se este último não excluir a hipótese nula não há razão alguma para tomarmos este tipo de procedimento que pressupõe a presença de uma tendência na variável.

Assim, apenas devemos reter a hipótese de variável estacionária à volta de uma tendência com rupturas temporais se a existência dessa tendência for um facto, de outra forma este procedimento não deve ser utilizado.

Não esqueçamos também, que este teste deve ser sobretudo ensaiado para o caso de termos chegado à conclusão, com outros métodos (testes) que a variável tinha uma raiz unitária.

Vejam os modelos estimados para as diferentes hipóteses de ruptura. Para o primeiro caso, onde ensaiamos uma alteração na intersecção em TB , o modelo é o seguinte

$$Y_t = \mu + \theta \cdot DU_t + \beta \cdot t + \delta \cdot DTB_t + \alpha \cdot Y_{t-1} + \sum_{i=1}^k c_i \cdot \Delta Y_{t-i} + \varepsilon_t$$

onde tomamos

$DU_t = 0$ para $t \leq TB$ e $DU_t = 1$ para $t > TB$ e ainda

$DTB_t = 1$ para $t = TB + 1$, $DTB_t = 0$ para $t > TB + 1$ e $DTB_t = 0$ para $t < TB$

Para o segundo caso, em que ensaiamos uma alteração não só na intersecção como na inclinação, devemos ter o seguinte modelo

$$Y_t = \mu + \theta \cdot DU_t + \beta \cdot t + \gamma \cdot DT_t + \delta \cdot DTB_t + \alpha \cdot Y_{t-1} + \sum_{i=1}^k c_i \cdot \Delta Y_{t-i} + \varepsilon_t$$

onde devemos tomar

$DU_t = 0$ para $t \leq TB$ e $DU_t = 1$ para $t > TB$

$DTB_t = 1$ para $t = TB + 1$, $DTB_t = 0$ para $t > TB + 1$ e $DTB_t = 0$ para $t < TB$ e ainda

$DT_t = 0$ para $t \leq TB$ e $DT_t = t - TB$ para $t > TB$.

Finalmente para o terceiro caso considerado, de alteração na inclinação não havendo descontinuidade na curva de tendência, temos um processo em duas etapas. Na primeira temos

$$Y_t = \mu + \beta \cdot t + \gamma DT_t + \tilde{Y}_t$$

onde DT_t é nossa conhecida.

E na segunda

$$\tilde{Y}_t = \alpha \cdot \tilde{Y}_{t-1} + \sum_{i=1}^k c_i \cdot \Delta \tilde{Y}_{t-i} + \varepsilon_t.$$

Estas três hipóteses deverão ser verificadas uma a uma. Voltemos a insistir no facto de se a variável a estudar não apresenta qualquer *trend*, ou seja, não excluimos a hipótese nula do coeficiente associado a t , então não devemos continuar com os testes de investigação de raiz unitária, de $\alpha = 1$.

As distribuições apropriadas, a cada uma das formas de detectar endogeneamente a ruptura temporal, e aplicadas a $t_{\alpha=1}$, encontram-se no artigo de Perron (1997)¹⁰.

No caso de os testes aqui apresentados nos indicarem a exclusão da raiz unitária, a exclusão da nossa $H(0)$, então devemos ter algum cuidado com o tipo de ruptura que vamos reter. O nosso conhecimento da variável em estudo deve ser então utilizado. Se os três tipos de modelos nos derem datas de ruptura diferentes, talvez o melhor seja passar ao procedimento exógeno de determinação das possíveis rupturas, com base no período que à partida nos parece mais propício a esse fenómeno.

2.7 A Hipótese Nula de Estacionaridade

Os testes até aqui analisados tomam como hipótese nula a presença de uma raiz unitária. O teste apresentado por Denis Kwiatkowski; Peter C. B. Phillips; Peter Schmidt and Yongcheol Shin¹¹, e conhecido pelas iniciais dos seus autores, tomam como hipótese nula a estacionaridade. Ou seja, a não exclusão da hipótese nula, para o nível de informação disponível sobre a variável, leva-nos a aceitar a característica de estacionaridade. Por este motivo, é frequente o seu uso em face de dificuldades encontradas na interpretação dos resultados do teste ADF.

Tomemos a seguinte decomposição de uma variável y

$$y_t = \beta \cdot t + r_t + \mu_t$$

onde r constitui um *random walk*, $r_t = r_{t-1} + \varepsilon_t$, tendo os erros as propriedades habituais, $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. A componente μ é tomada como sendo

¹⁰Pp. 362-3.

¹¹Schmidt e Shin (1992).

estacionária, $\mu \sim I(0)$. Tomemos o primeiro valor de $r_{t(=0)}$ como sendo fixo, r_0 .

Quando $\sigma_\varepsilon^2 = 0$, virá $r_t = r_{t-1} = r_0$ (r_0 constante). Passaremos neste caso a ter

$$y_t = r_0 + \beta \cdot t + \mu_t$$

em que y é estacionário ao longo de uma tendência, ou quando $\beta = 0$,

$$y_t = r_0 + \mu_t$$

em que y é estacionário em redor de uma constante.

O problema que se coloca consiste afinal em saber como avaliar se podemos reter $\sigma_\varepsilon^2 = 0$. Tomemos os erros da estimação do modelo

$$y_t = \beta_0 + \beta_1 \cdot t + e_t$$

onde teremos $e_t = y_t - \bar{y}$ no caso de estudarmos um processo estacionário em redor de uma constante.

Defina-se a soma parcial dos erros como $S_t = \sum_{i=1}^t \hat{e}_i$, para $t = 1, 2, \dots, T$, vindo a variância de longo prazo dada por

$$\sigma^2 = \lim_{T \rightarrow \infty} E [S_T^2]$$

sendo o seu estimador eficiente (Newey-West) dado por

$$\hat{\sigma}^2 = \frac{1}{T} \cdot \sum_{t=1}^T (\hat{e}_t)^2 + \frac{2}{T} \cdot \sum_{s=1}^l \left(1 - \frac{s}{1+l}\right) \cdot \sum_{t=s+1}^T \hat{e}_t \cdot \hat{e}_{t-s}$$

O valor da estatística KPSS é definido por

$$KPSS = \frac{1}{T^2} \cdot \frac{\sum_{t=1}^T S_t^2}{\hat{\sigma}^2}.$$

O problema que se coloca de imediato é o da escolha do parâmetro l . Ele deve ser escolhido de forma a que $\hat{\sigma}^2$ estabilize. Uma possível solução é usar os desfasamentos do teste ADF que eliminam a presença de AR.

2.8 Exemplos de Aplicação no RATS

Optamos pela apresentação de um conjunto alargado de exemplos.

O nosso objectivo é provar que a aplicação dos testes que acabámos de ver pode não conduzir a “certezas” acerca das características das séries, ou não estivéssemos nós no campo dos fenómenos “estocásticos”. No que se segue, as instruções do RATS encontram-se em itálico. Depois de uma primeira apresentação dos resultados completos de um teste procurámos eliminar o que já poderia ser redundante. Mantivemos, no entanto, a informação de natureza estatística. Apesar de termos usado uma base de dados criada por nós, incluímos também a forma como as séries foram originalmente criadas. Se bem que tivéssemos optado pela instrução “%invnormal(%uniform(0,1))” os autores do programa insistem na correcção do uso simples de “%ran(1)”, valores com distribuição Normal de média nula e desvio padrão unitário.

```
* _____
end 1
Normal Completion. Halt at 1
cal
all 0 500
open data basecd.rat
data(format=rats) /
table
Series Obs Mean Std Error Minimum Maximum
Y_1 500 11.4308091 13.0212173 -11.5384738 39.4031874
Y_2 500 255.5191409 147.1280748 1.0000000 523.4755801
Y_3 500 444.5352434 328.3178184 0.8024026 1114.8879613
Y_4 500 -0.0233314 0.9675172 -3.0437533 2.9480151
Y_5 500 100.0010792 0.9482761 97.2244322 103.1935670
Y_6 500 13.5365366 7.3489073 -0.6660893 28.0716595
set trend = t
source(noecho) kpss.src
source(noecho) uradf.src
source(noecho) cochrans2.src
```

2.8.1 Séries com raiz unitária

Serão aqui apresentados os diferentes casos de média nula, média não nula e de média não nula com tendência.

Raiz unitária sem intersecção e sem tendência

* com $\beta_0 = 0.0$

* com $\alpha = 1$

```

* set y_1 = beta_0
* dofor i = 2 to 500
* com y_1(i) = beta_0 + alpha*y_1(i-1)+ %invnormal(%uniform(0,1))
* end dofor
graph(header='Series contains a unit root with zero drift') 1
# y_1 100 500

```



```
@uradf(sclags=2,crit=lmtest) y_1 100 500
```

```

*****
* TESTING THE NULL HYPOTHESIS OF A UNIT ROOT IN Y_1 *
* Using data from 100 to 500 *
* Choosing the optimal lag length for the ADF regression *
* by adding lags until a Lagrange Multiplier test fails to *
* reject no residual serial correlation at level0.050. *
*****

```

Adding lag 0

Lagrange multiplier test for residual serial correlation of order 2

Test Statistic: 0.23186 Significance Level: 0.89054

```
*****
```

* Augmented Dickey-Fuller t-test with 0 lags: -0.8974 *

* 1% 5% 10% *

* -3.44 -2.87 -2.57 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -1.3773 *

* 1% 5% 10% *

* -20.5 -14.0 -11.2 *

* *

```

* Coefficient and T-Statistic on the Constant: *
*   0.08113   1.1007 *
* *
* Joint test of a unit root and no constant: 0.6195 *
*   1%   5%   10% *
*   6.47  4.61  3.79 *
*****
@uradf(sclags=2,crit=lmtest,det=none) y_1 100 500
*****
* TESTING THE NULL HYPOTHESIS OF A UNIT ROOT IN Y_1 *
* Using data from 100 to 500 *
* Choosing the optimal lag length for the ADF regression *
* by adding lags until a Lagrange Multiplier test fails to *
* reject no residual serial correlation at level0.050. *
*****
Adding lag 0
Lagrange multiplier test for residual serial correlation of order 2
Test Statistic: 0.00000 Significance Level: 1.00000
*****
* Augmented Dickey-Fuller t-test with 0 lags: -0.1655 *
*   1%   5%   10% *
*  -2.58  -1.95  -1.62 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -0.1792 *
*   1%   5%   10% *
*  -13.7  -8.0  -5.7 *
*****
@kpss(lmax=0) y_1 100 500
ETA(mu) Values:
Critical Level: 0.10  0.05  0.025  0.01
Critical Value: 0.347  0.463  0.574  0.739
For lag parameter l = 0 ETA(mu) = 33.22934
ETA(tau) Values:
Critical Level: 0.10  0.05  0.025  0.01
Critical Value: 0.119  0.146  0.176  0.216
For lag parameter l = 0 ETA(tau) = 4.74329
@cochran2(kmax=50) y_1 100 500
Procedure COCHRAN2. Written by Paco Goerlich
DESCRIPTIVE STATISTICS: DIFFERENCED SERIES
Statistics on Series DY_1
Observations 400
Sample Mean 0.03549046473 Variance 1.084411
Standard Error 1.04135057653 SE of Sample Mean 0.052068

```

t-Statistic 0.68162 Signif Level (Mean=0) 0.49587233

Skewness -0.01338 Signif Level (Sk=0) 0.91333310

Kurtosis -0.30703 Signif Level (Ku=0) 0.21404545

Jarque-Bera 1.58311 Signif Level (JB=0) 0.45314071

Correlations of Series DY_1

Autocorrelations

1: 0.0130180 -0.0203363 0.0446545 -0.0411355 0.0020879 -0.0417136

7: -0.0076704 0.0853379 -0.0145517 0.1100158 -0.0002016 0.0312013

13: 0.0286271 -0.0864040 0.0346841 0.0424573 0.0385783 0.0721545

19: -0.0365911 0.0308145 -0.0074737 -0.0788482 -0.0246001 0.0218464

25: 0.0462269 -0.0383788 0.0006904 0.0220878 0.0594751 0.0260752

31: -0.0879134 -0.0735294 -0.0002132 0.0277538 0.1061334 0.0548844

37: 0.1098404 -0.0851167 -0.0368970 -0.0282977 0.0610112 -0.0529606

43: 0.0345585 0.0725667 0.0321708 0.0148328 0.0014473 -0.0653047

49: 0.0013556 -0.0708737

COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK

CALCULATIONS ARE BIASED CORRECTED BY FACTOR: NOBS/(NOBS-

K)

Window size = 1 V 1.01690 Asymptotic SD 0.08303

A1 1.00850

Window size = 2 V 1.01305 Asymptotic SD 0.10131

A1 1.00659

Window size = 3 V 1.03598 Asymptotic SD 0.11963

A1 1.01792

Window size = 4 V 1.03334 Asymptotic SD 0.13340

A1 1.01662

Window size = 5 V 1.03339 Asymptotic SD 0.14614

A1 1.01664

.....
Window size = 45 V 1.62591 Asymptotic SD 0.63667

A1 1.27522

Window size = 46 V 1.63728 Asymptotic SD 0.64805

A1 1.27967

Window size = 47 V 1.64775 Asymptotic SD 0.65910

A1 1.28376

Window size = 48 V 1.65383 Asymptotic SD 0.66839

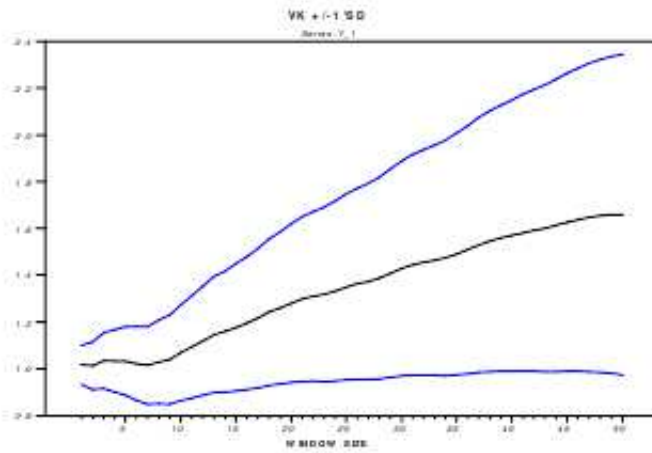
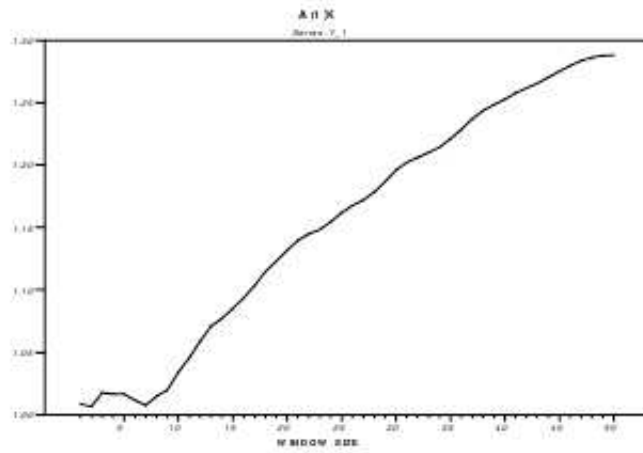
A1 1.28612

Window size = 49 V 1.65814 Asymptotic SD 0.67693

A1 1.28780

Window size = 50 V 1.65928 Asymptotic SD 0.68414

A1 1.28824



Raiz unitária com intersecção e sem tendência

```

* com beta_0 = 1
* com alpha = 1
* set y_2 = beta_0
* dofor i = 2 to 500
* com y_2(i)=beta_0+alpha*y_2(i-1)+(2*%invnormal(%uniform(0,1)))
* end dofor
graph(header='Series contains a unit root with drift') 1
# y_2 100 500

```



```
@uradf(sclags=2,crit=lmtest) y_2 100 500
```

```

*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: 0.3816 *
* 1% 5% 10% *
* -3.44 -2.87 -2.57 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: 0.1214 *
* 1% 5% 10% *
* -20.5 -14.0 -11.2 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.96854 3.7340 *
* *
* Joint test of a unit root and no constant: 63.7560 *
* 1% 5% 10% *

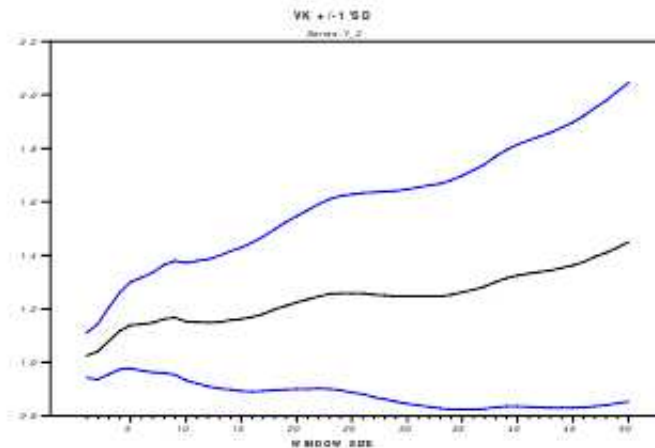
```



```

*   6.47   4.61   3.79 *
*****
@uradf(sclags=2,crit=lmtest,det=trend) y_2 100 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -1.3483 *
*   1%   5%   10% *
*  -3.98  -3.42  -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -4.2284 *
*   1%   5%   10% *
*  -28.9  -21.5  -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
*   0.92616   3.5505 *
* Coefficient and T-Statistic on the Linear Trend: *
*   0.01115   1.3942 *
* *
* Joint test of a unit root and no linear trend 1.0448 *
*   1%   5%   10% *
*   8.34   6.30   5.36 *
*****
@kpss(lmax=0) y_2 100 500
ETA(mu) Values:
Critical Level:  0.10  0.05  0.025  0.01
Critical Value:  0.347  0.463  0.574  0.739
For lag parameter l = 0 ETA(mu) = 39.82439
ETA(tau) Values:
Critical Level:  0.10  0.05  0.025  0.01
Critical Value:  0.119  0.146  0.176  0.216
For lag parameter l = 0 ETA(tau) = 5.98995
@cochran2(kmax=50) y_2 100 500
Procedure COCHRAN2. Written by Paco Goerlich
COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK
.....
Window size = 50   V   1.44845 Asymptotic SD 0.59721
A1 1.20380
.....

```



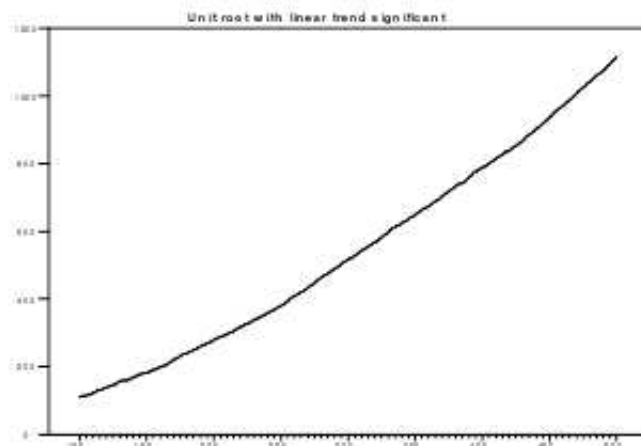
Raiz unitária com intersecção e tendência

Depois de efectuados os testes normais para as observações de 100 a 500, iremos dividir a amostra com observações de 100 a 200, 300 a 400 e 300 a 500. Como podemos verificar, as nossas conclusões, acerca de uma variável, poderão, infelizmente, estar dependentes da sub-amostra retida.

```

* com beta_0 = 1
* com alpha = 1
* set y_3 = beta_0
* dofor i = 2 to 500
* com y_3(i)=beta_0+.005*trend(i)+alpha*y_3(i-1)+%invnormal(%uniform(0,1))
* end dofor
graph(header='Unit root with linear trend significant') 1
# y_3 100 500
@uradf(sclags=2,crit=lmtest,det=trend) y_3 100 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -0.8334 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -0.5889 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *

```



```
* 0.59864 1.4761 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.00901 2.0217 *
* *
* Joint test of a unit root and no linear trend 73.0668 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
```

@kpss(lmax=0) y_3 100 500

```
ETA(mu) Values:
Critical Level: 0.10 0.05 0.025 0.01
Critical Value: 0.347 0.463 0.574 0.739
For lag parameter l = 0 ETA(mu) = 39.87341
ETA(tau) Values:
Critical Level: 0.10 0.05 0.025 0.01
Critical Value: 0.119 0.146 0.176 0.216
For lag parameter l = 0 ETA(tau) = 8.84511
```

@cochran2(kmax=50) y_3 100 500

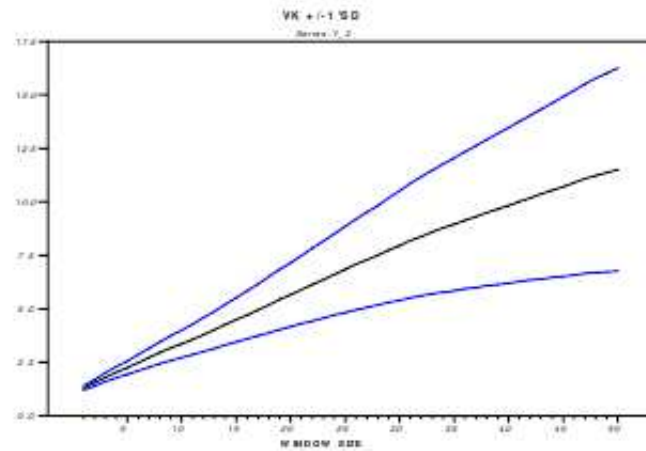
```
Procedure COCHRAN2. Written by Paco Goerlich
COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK
```

```
.....
Window size = 50 V 11.50650 Asymptotic SD 4.74425
A1 3.55936
```

```
***** power of our tests applied to our last serie:
* *****
```

@uradf(sclags=2,crit=lmtest,det=trend) y_3 100 200

```
*****
```



```

.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -1.3569 *
* 1% 5% 10% *
* -3.99 -3.43 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -3.7295 *
* 1% 5% 10% *
* -28.4 -21.3 -18.0 *
* *
* Coefficient and T-Statistic on the Constant: *
* -1.98867 -1.0540 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.07031 1.5306 *
* *
* Joint test of a unit root and no linear trend 3.5345 *
* 1% 5% 10% *
* 8.43 6.34 5.39 *
*****

```

```
@uradf(sclags=2,crit=lmttest,det=trend) y_3 300 400
```

```

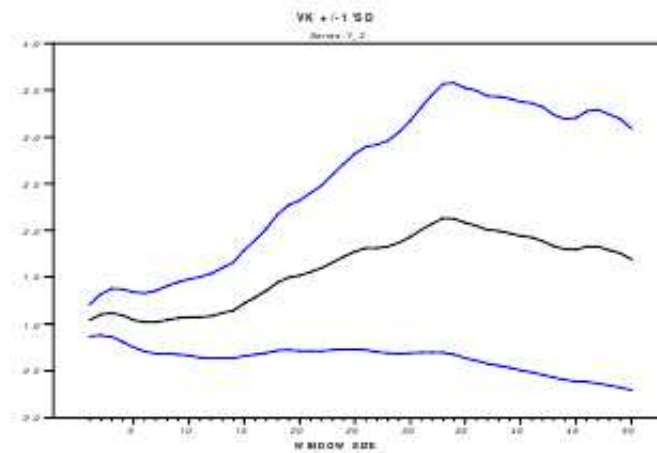
*****
.....
* Augmented Dickey-Fuller t-test with 1 lags: -3.3614 *
* 1% 5% 10% *
* -3.99 -3.43 -3.13 *
* *

```

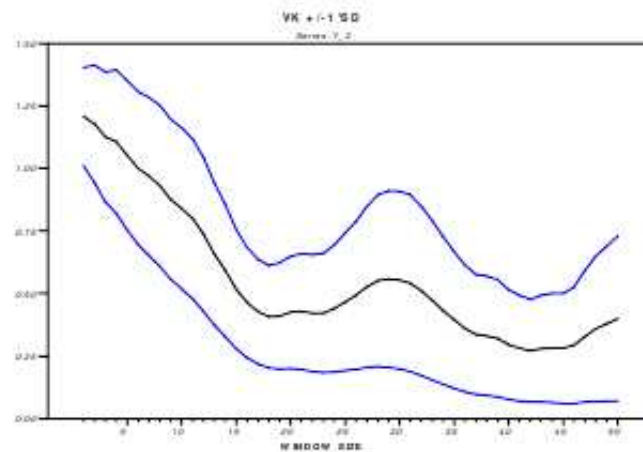
```

* Augmented Dickey-Fuller Z-test with 1 lags: -23.3141 *
* 1% 5% 10% *
* -28.4 -21.3 -18.0 *
* *
* Coefficient and T-Statistic on the Constant: *
* -49.36761 -3.2557 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.45708 3.3765 *
* *
* Joint test of a unit root and no linear trend 5.8377 *
* 1% 5% 10% *
* 8.43 6.34 5.39 *
*****
@uradf(sclags=2,crit=lmtest,det=trend) y_3 300 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 2 lags: 0.8936 *
* 1% 5% 10% *
* -3.99 -3.43 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 2 lags: 1.3821 *
* 1% 5% 10% *
* -28.4 -21.3 -18.0 *
* *
* Coefficient and T-Statistic on the Constant: *
* 3.16698 1.0565 *
* Coefficient and T-Statistic on the Linear Trend: *
* -0.01414 -0.6237 *
* *
* Joint test of a unit root and no linear trend 9.7798 *
* 1% 5% 10% *
* 8.43 6.34 5.39 *
*****
@cochran2(kmax=50) y_3 100 200
Procedure COCHRAN2. Written by Paco Goerlich
.....
Window size = 50 V 1.69476 Asymptotic SD 1.39753
A1 1.30279
@cochran2(kmax=50) y_3 300 400
Procedure COCHRAN2. Written by Paco Goerlich
COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK
.....

```



Window size = 50 V 0.39964 Asymptotic SD 0.32955
A1 0.64392

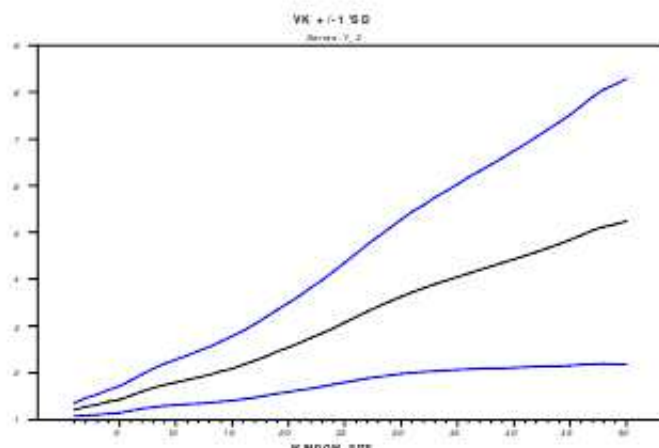


@cochran2(kmax=50) y_3 300 500

Procedure COCHRAN2. Written by Paco Goerlich
COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK

.....

Window size = 50 V 5.23656 Asymptotic SD 3.05342
A1 2.34242

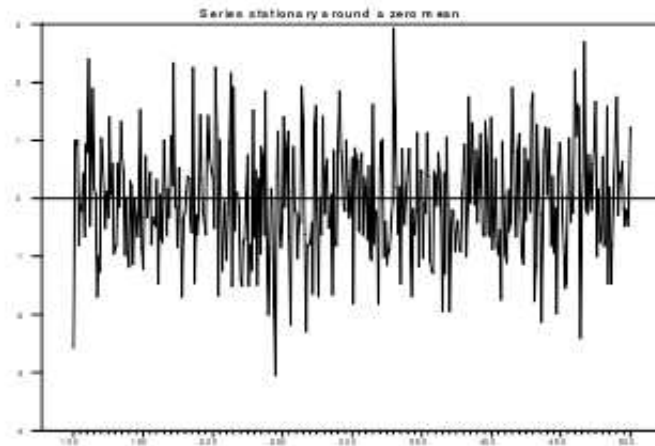


2.8.2 Séries estacionárias

Tal como foi feito para as séries com raiz unitária, vamos agora apresentar um conjunto de séries geradas de forma a apresentarem a característica de estacionaridade.

Série sem intersecção e tendência

```
* com beta_0 = 0.0
* set y_4 = beta_0
* dofor i = 2 to 500
* com y_4(i) = beta_0 + %invnormal(%uniform(0,1))
* end dofor
graph(header='Series stationary around a zero mean') 1
# y_4 100 500
@uradf(sclags=2,crit=lmtest) y_4 100 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -18.8006 *
* 1% 5% 10% *
* -3.44 -2.87 -2.57 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -377.4715 *
* 1% 5% 10% *
* -20.5 -14.0 -11.2 *
* *
* Coefficient and T-Statistic on the Constant: *
```



```
* -0.03262 -0.6690 *
```

```
* *
```

```
* Joint test of a unit root and no constant: 176.7357 *
```

```
* 1% 5% 10% *
```

```
* 6.47 4.61 3.79 *
```

```
*****
```

```
@uradf(sclags=2,crit=lmtest,det=none) y_4 100 500
```

```
*****
```

```
.....
```

```
*****
```

```
* Augmented Dickey-Fuller t-test with 0 lags: -18.8019 *
```

```
* 1% 5% 10% *
```

```
* -2.58 -1.95 -1.62 *
```

```
* *
```

```
* Augmented Dickey-Fuller Z-test with 0 lags: -376.9317 *
```

```
* 1% 5% 10% *
```

```
* -13.7 -8.0 -5.7 *
```

```
*****
```

```
@kps(lmax=0) y_4 100 500
```

```
ETA(mu) Values:
```

```
Critical Level: 0.10 0.05 0.025 0.01
```

```
Critical Value: 0.347 0.463 0.574 0.739
```

```
For lag parameter l = 0 ETA(mu) = 0.12475
```

```
ETA(tau) Values:
```

```
Critical Level: 0.10 0.05 0.025 0.01
```

```
Critical Value: 0.119 0.146 0.176 0.216
```

```
For lag parameter l = 0 ETA(tau) = 0.06762
```



```
@cochran2(kmax=50) y_4 100 500
```

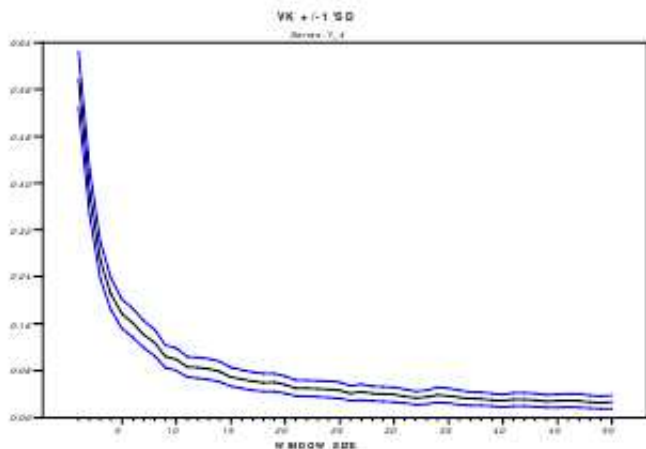
```
Procedure COCHRAN2. Written by Paco Goerlich
```

```
COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK
```

```
.....
```

```
Window size = 50 V 0.02659 Asymptotic SD 0.01096
```

```
A1 0.17967
```



Série estacionária com intersecção e sem tendência

```
* com beta_0 = 100
```

```
* set y_5 = beta_0
```

```
* dofor i = 2 to 500
```

```
* com y_5(i) = beta_0 + %invnorm(%uniform(0,1))
```

```
* end dofor
```

```
graph(header='Series stationary around a non-zero mean') 1
```

```
# y_5 100 500
```

```
@uradf(sclags=2,crit=lmtest) y_5 100 500
```

```
*****
```

```
.....
```

```
*****
```

```
* Augmented Dickey-Fuller t-test with 0 lags: -19.5165 *
```

```
* 1% 5% 10% *
```

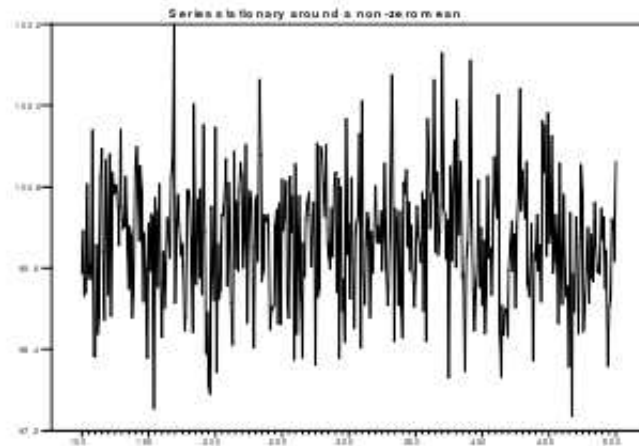
```
* -3.44 -2.87 -2.57 *
```

```
* *
```

```
* Augmented Dickey-Fuller Z-test with 0 lags: -391.8534 *
```

```
* 1% 5% 10% *
```

```
* -20.5 -14.0 -11.2 *
```



* *

* Coefficient and T-Statistic on the Constant: *

* 97.72481 19.5167 *

* *

* Joint test of a unit root and no constant: 190.4538 *

* 1% 5% 10% *

* 6.47 4.61 3.79 *

@uradf(sclags=2,crit=lmtest,det=trend) y_5 100 500

.....

* Augmented Dickey-Fuller t-test with 0 lags: -19.5125 *

* 1% 5% 10% *

* -3.98 -3.42 -3.13 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -392.2820 *

* 1% 5% 10% *

* -28.9 -21.5 -18.1 *

* *

* Coefficient and T-Statistic on the Constant: *

* 97.91123 19.5064 *

* Coefficient and T-Statistic on the Linear Trend: *

* -0.00027 -0.6388 *

* *

* Joint test of a unit root and no linear trend 190.3689 *

* 1% 5% 10% *

```
* 8.34 6.30 5.36 *
```

```
*****
```

```
@kps(lmax=0) y_5 100 500
```

```
ETA(mu) Values:
```

```
Critical Level: 0.10 0.05 0.025 0.01
```

```
Critical Value: 0.347 0.463 0.574 0.739
```

```
For lag parameter l = 0 ETA(mu) = 0.12788
```

```
ETA(tau) Values:
```

```
Critical Level: 0.10 0.05 0.025 0.01
```

```
Critical Value: 0.119 0.146 0.176 0.216
```

```
For lag parameter l = 0 ETA(tau) = 0.10176
```

```
@cochran2(kmax=50) y_5 100 500
```

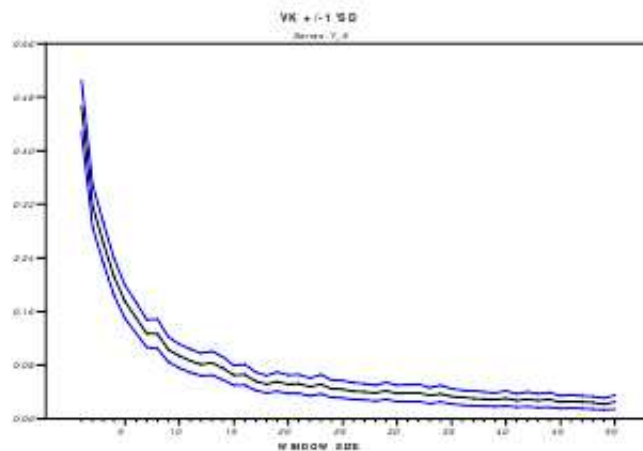
```
Procedure COCHRAN2. Written by Paco Goerlich
```

```
COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK
```

```
.....
```

```
Window size = 50 V 0.02435 Asymptotic SD 0.01004
```

```
A1 0.18474
```



Série estacionária à volta de uma tendência determinista

Depois da análise da totalidade da amostra podemos estudar os sub-períodos de 100 a 200, 300 a 400 e 300 a 500 e verificar que as características da totalidade se conservam.

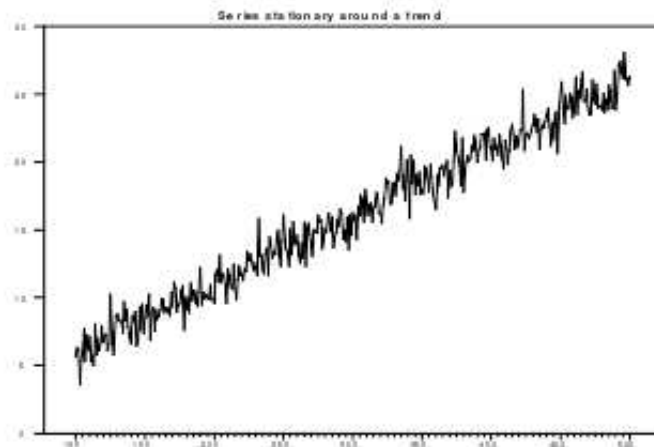
```
* com beta_0 = 1
```

```
* set y_6 = beta_0
```

```
* dofor i = 2 to 500
```

```
* com y_6(i)=beta_0+.05*trend(i)+%invnorm(%uniform(0,1))
```

```
* end dofor
graph(header='Series stationary around a trend') 1
# y_6 100 500
```



```
@uradf(sclags=2,crit=lmtest,det=trend) y_6 100 500
```

```
*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -18.6276 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -372.9703 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.84409 5.6589 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.04689 18.3298 *
* *
* Joint test of a unit root and no linear trend 173.4963 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
```

```
@kpss(lmax=0) y_6 100 500
```

```
ETA(mu) Values:
```

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.347 0.463 0.574 0.739

For lag parameter $l = 0$ $\text{ETA}(\mu) = 38.89179$

ETA(tau) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.119 0.146 0.176 0.216

For lag parameter $l = 0$ $\text{ETA}(\tau) = 0.04474$

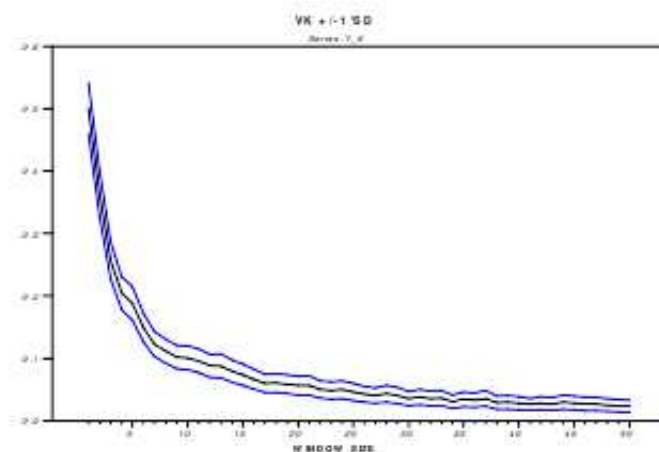
@cochran2(kmax=50) y_6 100 500

Procedure COCHRAN2. Written by Paco Goerlich

COCHRANE (1988 - JPE) MEASURE OF PERSISTENCE: VK

.....
Window size = 50 V 0.02313 Asymptotic SD 0.00954

A1 0.17580



2.8.3 Séries com uma ruptura estrutural

Iremos analisar séries com rupturas provocadas de acordo com os casos considerados por Perron para as alterações de comportamento tendencial.

source(noecho) perron97.src

Alteração na intersecção

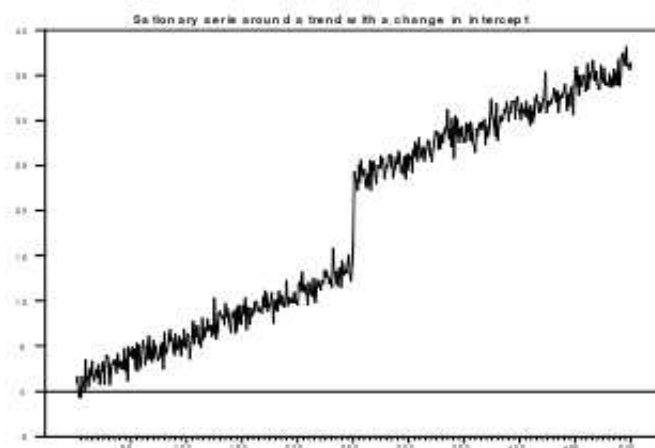
set y_7 1 250 = y_6

set y_7 251 500 = y_6 + 10

graph(header='Stationary serie around a trend with a change in intercept') 1

y_7

@uradf(sclags=2,crit=lmttest,det=trend) y_7 1 250



```

*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -16.7834 *
* 1% 5% 10% *
* -3.99 -3.43 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -269.2289 *
* 1% 5% 10% *
* -28.4 -21.3 -18.0 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.98029 7.0523 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.05454 16.2760 *
* *
* Joint test of a unit root and no linear trend 140.8820 *
* 1% 5% 10% *
* 8.43 6.34 5.39 *
*****

```

```
@uradf(sclags=2,crit=lmtest,det=trend) y_7 250 500
```

```

*****
.....
*****
* Augmented Dickey-Fuller t-test with 2 lags: -11.0074 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *

```

```

* *
* Augmented Dickey-Fuller Z-test with 2 lags: -151.8923 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
* 7.57827 10.4282 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.03502 10.1281 *
* *
* Joint test of a unit root and no linear trend 61.7732 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
@uradf(sclags=2,crit=lmtest,det=trend) y_7 1 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 4 lags: -2.1427 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 4 lags: -9.9249 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.10711 0.8904 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.00372 2.0551 *
* *
* Joint test of a unit root and no linear trend 2.3116 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
@perron97(model=IO1,lagmax=0,signif=.05) y_7
-----
break date TB = 249 statistic t(alpha==1) = -23.14654
critical values at 1% 5% 10% 50% 90% 95% 99%
for 100 obs. -5.70 -5.10 -4.82 -3.87 -3.05 -2.75 -2.22
infinite sample -5.41 -4.80 -4.58 -3.75 -2.99 -2.77 -2.32
-----
number of lag retained : 0

```

explained variable : Y_7

coefficient student

CONSTANT	0.91373	8.37138
DU	9.93441	21.46539
D(Tb)	-7.36952	-6.68097
TIME	0.04980	22.06153
Y_7 {1}	0.00947	0.22120

print 247 253 y_7

ENTRY Y_7

247 13.06744938978

248 12.32065786018

249 14.14908566475

250 16.06326899359

251 24.25605780010

252 23.55802349582

253 23.30761681966

Série com raiz unitária e tendência e alteração da intersecção

Tomamos como ponto de partida a série já utilizada y_3

set y_8 1 250 = y_3

set y_8 251 500 = y_3 + 100

graph(header='Unit root with trend and a change in intercept') 1

y_8

@uradf(sclags=2,crit=lmtest,det=trend) y_8 1 250

.....

* Augmented Dickey-Fuller t-test with 0 lags: -0.8505 *

* 1% 5% 10% *

* -3.99 -3.43 -3.13 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -1.0443 *

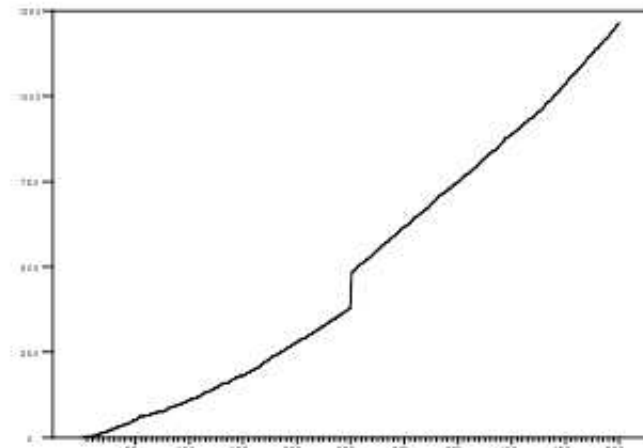
* 1% 5% 10% *

* -28.4 -21.3 -18.0 *

* *

* Coefficient and T-Statistic on the Constant: *

* 0.68368 3.3608 *



```

* Coefficient and T-Statistic on the Linear Trend: *
* 0.01199 1.5831 *
* *
* Joint test of a unit root and no linear trend 19.4894 *
* 1% 5% 10% *
* 8.43 6.34 5.39 *
*****
@uradf(sclags=2,crit=lmtest,det=trend) y_8 250 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 1 lags: -6.8634 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 1 lags: -50.2800 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
* -45.08393 -5.9180 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.56183 6.7857 *
* *
* Joint test of a unit root and no linear trend 24.1000 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
    
```

```

*****
@uradf(sclags=2,crit=lmtest,det=trend) y_8 1 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -1.9921 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -3.5543 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.02183 0.0329 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.02360 2.5628 *
* *
* Joint test of a unit root and no linear trend 9.3487 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
@perron97(model=IO1,lagmax=0,signif=.05) y_8
-----
break date TB = 249 statistic t(alpha==1) = -3.61783
critical values at 1% 5% 10% 50% 90% 95% 99%
for 100 obs. -5.70 -5.10 -4.82 -3.87 -3.05 -2.75 -2.22
infinite sample -5.41 -4.80 -4.58 -3.75 -2.99 -2.77 -2.32
-----
number of lag retained : 0
-----
explained variable : Y_8
-----
coefficient student
-----
CONSTANT -0.20333 -0.30822
DU 3.49358 3.72052
D(Tb) -4.40872 -0.95587
TIME 0.03255 3.45457
Y_8 {1} 0.98523 241.30729
-----

```

Série com alteração na inclinação e intersecção

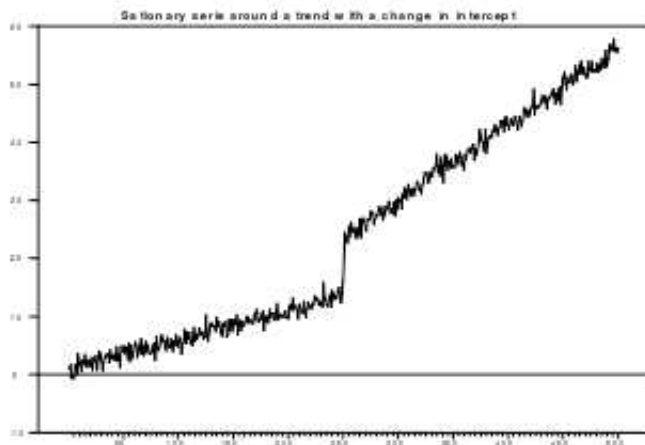
A série que vamos construir apresenta algumas características interessantes. Tendo em conta a ruptura temporal ela pode ser retida como estacionária ao longo de uma tendência. Analisada de forma usual ela apresenta uma raiz unitária para a totalidade da amostra. no entanto, ela é estacionária para as observações de 1 a 250 e de 250 a 500.

```
set y_9 1 250 = y_6
```

```
set y_9 251 500 = y_6 + 10 + .08*(trend-250)
```

```
graph(header='Stationary serie around a trend with a change in intercept') 1
```

```
# y_9
```



```
@uradf(sclags=2,crit=lmtest,det=trend) y_9 1 250
```

```
*****
```

```
.....
```

```
*****
```

```
* Augmented Dickey-Fuller t-test with 0 lags: -16.7834 *
```

```
* 1% 5% 10% *
```

```
* -3.99 -3.43 -3.13 *
```

```
* *
```

```
* Augmented Dickey-Fuller Z-test with 0 lags: -269.2289 *
```

```
* 1% 5% 10% *
```

```
* -28.4 -21.3 -18.0 *
```

```
* *
```

```
* Coefficient and T-Statistic on the Constant: *
```

```
* 0.98029 7.0523 *
```

```
* Coefficient and T-Statistic on the Linear Trend: *
```

```
* 0.05454 16.2760 *
```

* *

* Joint test of a unit root and no linear trend 140.8820 *

* 1% 5% 10% *

* 8.43 6.34 5.39 *

@uradf(sclags=2,crit=lmtest,det=trend) y_9 250 500

.....

* Augmented Dickey-Fuller t-test with 2 lags: -10.9369 *

* 1% 5% 10% *

* -3.98 -3.42 -3.13 *

* *

* Augmented Dickey-Fuller Z-test with 2 lags: -151.8559 *

* 1% 5% 10% *

* -28.9 -21.5 -18.1 *

* *

* Coefficient and T-Statistic on the Constant: *

* -6.26858 -8.3683 *

* Coefficient and T-Statistic on the Linear Trend: *

* 0.09053 10.6883 *

* *

* Joint test of a unit root and no linear trend 60.9824 *

* 1% 5% 10% *

* 8.34 6.30 5.36 *

@uradf(sclags=2,crit=lmtest,det=trend) y_9 1 500

.....

* Augmented Dickey-Fuller t-test with 4 lags: -1.8346 *

* 1% 5% 10% *

* -3.98 -3.42 -3.13 *

* *

* Augmented Dickey-Fuller Z-test with 4 lags: -5.5309 *

* 1% 5% 10% *

* -28.9 -21.5 -18.1 *

* *

* Coefficient and T-Statistic on the Constant: *

* -0.04332 -0.2843 *

* Coefficient and T-Statistic on the Linear Trend: *

* 0.00373 2.0805 *

* *

```

* Joint test of a unit root and no linear trend 2.5737 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****

```

```
@perron97(model=IO2,lagmax=0,signif=.05) y_9
```

```

-----
break date TB = 249 statistic t(alpha==1) = -23.11577
critical values at 1% 5% 10% 50% 90% 95% 99%
for 100 obs. -6.21 -5.55 -5.25 -4.22 -3.35 -3.13 -2.63
infinite sample -5.57 -5.08 -4.82 -3.98 -3.25 -3.06 -2.72
-----

```

```
number of lag retained : 0
```

```
explained variable : Y_9
```

```
coefficient student
```

```

-----
CONSTANT 0.92150 6.73669
DU -9.90679 -17.16078
D(Tb) -7.36391 -6.66410
TIME 0.04975 21.31848
DT 0.07937 21.48080
Y_9 {1} 0.00935 0.21806
-----

```

Alteração da inclinação

```

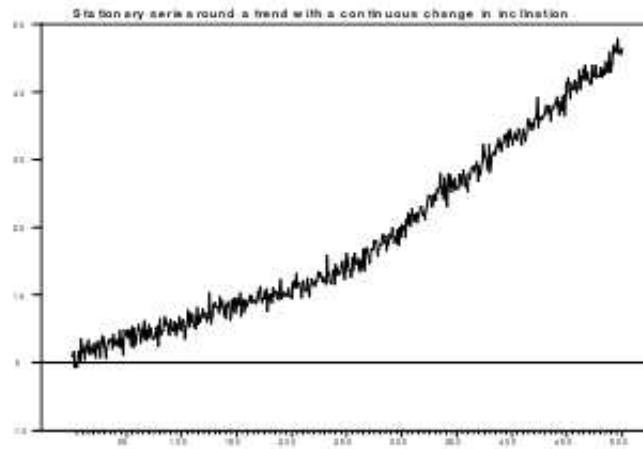
set y_10 1 250 = y_6
set y_10 251 500 = y_6 + .08*(trend-250)
graph(header='Stationary serie around a trend with a continuous change in incli-
nation') 1
# y_10
@uradf(sclags=2,crit=lmtest,det=trend) y_10 1 250

```

```

*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -16.7834 *
* 1% 5% 10% *
* -3.99 -3.43 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -269.2289 *
* 1% 5% 10% *
* -28.4 -21.3 -18.0 *

```



* *

* Coefficient and T-Statistic on the Constant: *

* 0.98029 7.0523 *

* Coefficient and T-Statistic on the Linear Trend: *

* 0.05454 16.2760 *

* *

* Joint test of a unit root and no linear trend 140.8820 *

* 1% 5% 10% *

* 8.43 6.34 5.39 *

@uradf(sclags=2,crit=lmtest,det=trend) y_10 250 500

.....

* Augmented Dickey-Fuller t-test with 0 lags: -13.9380 *

* 1% 5% 10% *

* -3.98 -3.42 -3.13 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -220.3016 *

* 1% 5% 10% *

* -28.9 -21.5 -18.1 *

* *

* Coefficient and T-Statistic on the Constant: *

* -16.64110 -13.2890 *

* Coefficient and T-Statistic on the Linear Trend: *

* 0.11420 13.8597 *

* *

```

* Joint test of a unit root and no linear trend 97.1382 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
@uradf(sclags=2,crit=lmtest,det=trend) y_10 1 500
*****
.....
*****
* Augmented Dickey-Fuller t-test with 8 lags: -0.5212 *
* 1% 5% 10% *
* -3.98 -3.42 -3.13 *
* *
* Augmented Dickey-Fuller Z-test with 8 lags: -0.8439 *
* 1% 5% 10% *
* -28.9 -21.5 -18.1 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.10359 0.8265 *
* Coefficient and T-Statistic on the Linear Trend: *
* 0.00206 1.3379 *
* *
* Joint test of a unit root and no linear trend 6.9125 *
* 1% 5% 10% *
* 8.34 6.30 5.36 *
*****
@perron97(model=AO,lagmax=0,signif=.05) y_10
-----
break date TB = 250 statistic t(alpha==1) = -21.72345
critical values at 1% 5% 10% 50% 90% 95% 99%
for 200 obs. -5.28 -4.65 -4.38 -3.32 -2.48 -2.27 -1.90
infinite sample -4.91 -4.36 -4.07 -3.13 -2.32 -2.12 -1.78
-----

number of lag retained : 0
-----

explained variable : Y_10
-----

coefficient student
-----

CONSTANT 0.91256 7.47835
TIME 0.05042 70.64384
DT 0.07988 62.71360
-----

Y_10 {1} 0.02686 0.59964

```

```
*****
*****
```

2.8.4 Exemplo de séries trimestrais

end 1

Normal Completion. Halt at 1

cal 1970 1 4

all 2000:4

GRPARM(BOLD) HEADER 30 SUBHEADER(italic) 25

env noecho

source uradf.src

source kpss.src

source hegyqnew.src

source cochrans2.src

sea s

open data cnt2001.rat

data(format=rats) / pib pibr95

table

.....

set p = pib/pibr95

com base = (p(1970:1)+p(1970:2)+p(1970:3)+p(1970:4))/4

*set ip = 100*p/base*

set lip = log(ip)

diff(sdiffs=1) lip / d4lip

set lq = log(pibr95)

diff lq / dlq

```
*****
```

graph(header='Inflation rate',subheader='1971:1-2000:4') 1

d4lip

@uradf(sclags=2,crit=lmtest) d4lip

```
*****
```

.....

```
*****
```

* Augmented Dickey-Fuller t-test with 0 lags: -2.0640 *

* 1% 5% 10% *

* -3.46 -2.88 -2.57 *

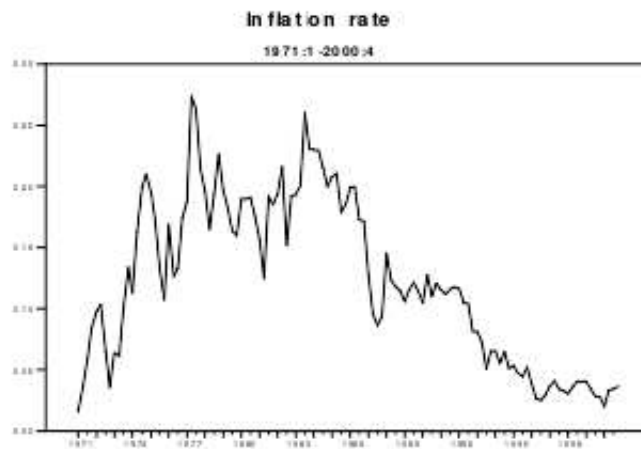
* *

* Augmented Dickey-Fuller Z-test with 0 lags: -7.8902 *

* 1% 5% 10% *

* -20.3 -14.0 -11.2 *

* *



* Coefficient and T-Statistic on the Constant: *

* 0.00833 1.8503 *

* *

* Joint test of a unit root and no constant: 2.1332 *

* 1% 5% 10% *

* 6.52 4.63 3.81 *

@uradf(sclags=2,crit=lmtest,det=none) d4lip

.....

* Augmented Dickey-Fuller t-test with 0 lags: -0.9087 *

* 1% 5% 10% *

* -2.58 -1.95 -1.62 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -1.6808 *

* 1% 5% 10% *

* -13.6 -8.0 -5.7 *

@kpss(lmax=0) d4lip

ETA(mu) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.347 0.463 0.574 0.739

For lag parameter l = 0 ETA(mu) = 5.40821

ETA(tau) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.119 0.146 0.176 0.216

For lag parameter $l = 0$ $\text{ETA}(\tau) = 1.87035$

Lin(noprint) d4lip / res

constant s{-2 to 0}

@uradf(sclags=2,crit=lmtest) res

.....

* Augmented Dickey-Fuller t-test with 0 lags: -2.0672 *

* 1% 5% 10% *

* -3.46 -2.88 -2.57 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -7.8971 *

* 1% 5% 10% *

* -20.3 -14.0 -11.2 *

* *

* Coefficient and T-Statistic on the Constant: *

* 0.00023 0.1054 *

* *

* Joint test of a unit root and no constant: 2.1402 *

* 1% 5% 10% *

* 6.52 4.63 3.81 *

@uradf(sclags=2,crit=lmtest,det=none) res

.....

* Augmented Dickey-Fuller t-test with 0 lags: -2.0749 *

* 1% 5% 10% *

* -2.58 -1.95 -1.62 *

* *

* Augmented Dickey-Fuller Z-test with 0 lags: -7.8928 *

* 1% 5% 10% *

* -13.6 -8.0 -5.7 *

@kpss(lmax=0) res

ETA(mu) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.347 0.463 0.574 0.739

For lag parameter $l = 0$ $\text{ETA}(\mu) = 5.40664$

ETA(tau) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.119 0.146 0.176 0.216

For lag parameter $l = 0$ $\text{ETA}(\tau) = 1.87023$

```

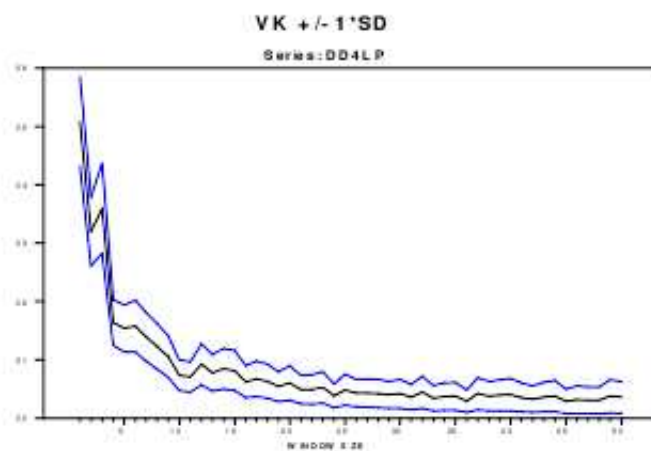
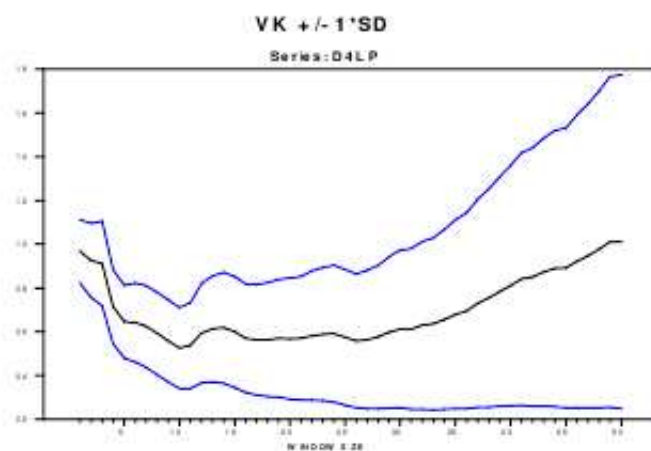
@hegyqnew(criterion=nocrit,nlags=9) lip
***** HEGY-tests *****
Testing for seasonal integration in series: LIP
Effective sample: 1973:02 to 2000:04
Aux.regr. ' t1 ' ' t2 ' ' t3 ' ' t4 ' ' F3&4 ' LM-sign Lags - -
0.388 -2.274 -1.083 -1.133 1.230 0.098 123456789
I -4.511 -2.228 -1.275 -1.103 1.424 0.529 123456789
I,SD -4.463 -2.205 -1.587 -0.459 1.359 0.567 123456789
I,Tr -1.084 -2.211 -1.269 -1.093 1.405 0.490 123456789
I,SD,Tr -1.049 -2.188 -1.576 -0.454 1.338 0.522 123456789
*****
Hylleberg et al.((1990) critical values for 100 obs
(5%) PI1 PI2 PI3 PI4 F34
None -1.97 -1.92 -1.90 -4.02 3.12
I only -2.88 -1.95 -1.90 -3.92 3.08
I,SD -2.95 -2.94 -3.44 -5.31 6.57
I,Tr -3.47 -1.94 -1.89 -3.90 2.98
I,SD,Tr -3.53 -2.94 -3.48 -5.29 6.60
*****
* we must insist with d4lip
@hegyqnew d4lip
***** HEGY-tests *****
Testing for seasonal integration in series: D4LIP
Effective sample: 1973:02 to 2000:04
Aux.regr. ' t1 ' ' t2 ' ' t3 ' ' t4 ' ' F3&4 ' LM-sign Lags
- - 0.564 -6.248 -4.143 -7.072 45.477 0.107 123456
I -0.917 -5.742 -4.240 -4.071 21.128 0.348 12345678
I,SD -0.914 -5.629 -4.236 -4.032 21.020 0.307 12345678
I,Tr -3.602 -6.443 -6.453 -7.209 47.629 0.225 12345
I,SD,Tr -3.597 -6.378 -6.406 -7.202 47.318 0.221 12345
*****
Hylleberg et al.((1990) critical values for 100 obs
(5%) PI1 PI2 PI3 PI4 F34
None -1.97 -1.92 -1.90 -4.02 3.12
I only -2.88 -1.95 -1.90 -3.92 3.08
I,SD -2.95 -2.94 -3.44 -5.31 6.57
I,Tr -3.47 -1.94 -1.89 -3.90 2.98
I,SD,Tr -3.53 -2.94 -3.48 -5.29 6.60
*****
* we can retain a sonseasonal unit root
diff d4lip / dd4lip
@uradf(sclags=2,crit=lmtest) dd4lip
*****

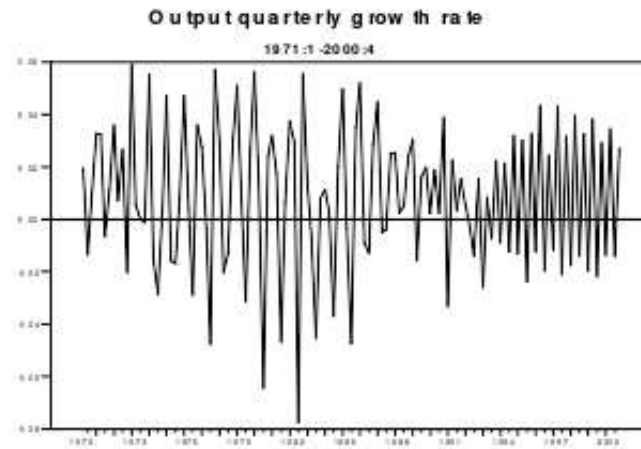
```

```

.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -11.3113 *
* 1% 5% 10% *
* -3.46 -2.88 -2.57 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -123.4169 *
* 1% 5% 10% *
* -20.3 -14.0 -11.2 *
* *
* Coefficient and T-Statistic on the Constant: *
* 0.00001 0.0039 *
* *
* Joint test of a unit root and no constant: 63.9756 *
* 1% 5% 10% *
* 6.52 4.63 3.81 *
*****
@uradf(sclags=2,crit=lmtest,det=none) dd4lip
*****
.....
*****
* Augmented Dickey-Fuller t-test with 0 lags: -11.3602 *
* 1% 5% 10% *
* -2.58 -1.95 -1.62 *
* *
* Augmented Dickey-Fuller Z-test with 0 lags: -123.4167 *
* 1% 5% 10% *
* -13.6 -8.0 -5.7 *
*****
* See what happen to the Cochrane ratio
@cochran2(kmax=50) d4lip
.....
@cochran2(kmax=50) dd4lip
.....
*****
graph(header='Output quarterly growth rate',subheader='1971:1-2000:4') 1
# dlq
@uradf(sclags=2,crit=lmtest) dlq
*****
.....
*****
* Augmented Dickey-Fuller t-test with 4 lags: -2.9853 *
* 1% 5% 10% *

```





```

*   -3.46  -2.88  -2.57 *
* *
* Augmented Dickey-Fuller Z-test with 4 lags: -23.1508 *
*   1%   5%   10% *
*  -20.3  -14.0  -11.2 *
* *
* Coefficient and T-Statistic on the Constant: *
*   0.00609   2.3814 *
* *
* Joint test of a unit root and no constant: 4.4749 *
*   1%   5%   10% *
*   6.52  4.63  3.81 *
*****

```

@kps(*lmax=4*) *dlq*

ETA(mu) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.347 0.463 0.574 0.739

For lag parameter $l = 0$ ETA(mu) = 0.03215

For lag parameter $l = 1$ ETA(mu) = 0.04740

For lag parameter $l = 2$ ETA(mu) = 0.07714

For lag parameter $l = 3$ ETA(mu) = 0.17442

For lag parameter $l = 4$ ETA(mu) = 0.09029

ETA(tau) Values:

Critical Level: 0.10 0.05 0.025 0.01

Critical Value: 0.119 0.146 0.176 0.216

For lag parameter $l = 0$ ETA(tau) = 0.01975

For lag parameter $l = 1$ ETA(tau) = 0.02919

For lag parameter l = 2 ETA(tau) = 0.04777

For lag parameter l = 3 ETA(tau) = 0.11021

For lag parameter l = 4 ETA(tau) = 0.05646

@hegyqnew lq

***** HEGY-tests *****

Testing for seasonal integration in series: LQ

Effective sample: 1972:02 to 2000:04

Aux.regr. 't1' 't2' 't3' 't4' 'F3&4' LM-sign Lags

- 2.926 -1.206 -1.748 0.055 1.530 0.811 1234

I -0.421 -1.190 -1.740 0.065 1.517 0.796 1234

I,SD -0.387 -1.714 -2.069 -0.066 2.142 0.727 1234

I,Tr -3.819 -1.701 -2.140 0.109 2.297 0.113 123

I,SD,Tr -3.487 -2.580 -2.687 0.291 3.657 0.119 12

Hylleberg et al.((1990) critical values for 100 obs

(5%) PI1 PI2 PI3 PI4 F34

None -1.97 -1.92 -1.90 -4.02 3.12

I only -2.88 -1.95 -1.90 -3.92 3.08

I,SD -2.95 -2.94 -3.44 -5.31 6.57

I,Tr -3.47 -1.94 -1.89 -3.90 2.98

I,SD,Tr -3.53 -2.94 -3.48 -5.29 6.60

* Perhaps it's better to transform ...

diff(sdiffs=1) lq / d4lq

@uradf(sclags=2,crit=lmtest) d4lq

.....

* Augmented Dickey-Fuller t-test with 1 lags: -3.0225 *

* 1% 5% 10% *

* -3.46 -2.88 -2.57 *

* *

* Augmented Dickey-Fuller Z-test with 1 lags: -18.0055 *

* 1% 5% 10% *

* -20.3 -14.0 -11.2 *

* *

* Coefficient and T-Statistic on the Constant: *

* 0.00636 2.4143 *

* *

* Joint test of a unit root and no constant: 4.5854 *

* 1% 5% 10% *

* 6.52 4.63 3.81 *

Capítulo 3

Cointegração, Equilíbrio e Ajustamento

Vamos dedicar este capítulo ao estudo da cointegração entre variáveis, ao seu significado económico e à forma como várias relações com implicações temporais diferentes pode ser deduzidas. Procuramos analisar a representação de variáveis em termos de longo prazo e o mecanismo de ajustamento do curto prazo a estes valores de equilíbrio.

A cointegração apresenta-se como um capítulo exemplar da econometria onde o significado económico das relações obtidas nunca pode ser descurado em favor de aspectos mais estatísticos: “a mindless attempt at finding cointegrated relations without knowing what they mean is not going to be fruitful, so I believe that the econometrician has to carefully choose the variables that should enter the study, and carefully discuss the economic theory that motivates this” (Johansen (1995), p. 6). A significância económica e estatística são aqui tomadas a par, “It must be emphasized that a cointegration analysis cannot be the final aim of an econometric investigation, but it is our impression that as an intermediate step a cointegrations analysis is a useful tool in the process of gaining understanding of the relation between data and theory, which should help in building a relevant econometric model” (Johansen (1995), p.8).

Começemos por apresentar o significado económico de relação, ou vector, de cointegração, através de alguns exemplos. Passaremos depois à obtenção de relações de equilíbrio de longo prazo, à ausência de significado de relações entre variáveis em certos casos e à equivalência entre os processos que envolvem um mecanismo de correcção dos erros e a cointegração. Apresentaremos o método de *Engle-Granger* para cálculo da relação de co-itegração. As limitações desta última metodologia vão conduzir-nos ao método de *Johansen*. Na exposição do seu método veremos também como podemos impor algumas restrições às relações de cointegração e aos vectores de ajustamento e testá-las.

3.1 Exemplos Económicos

A utilização da ideia de equilíbrio e ajustamento no curto prazo àqueles valores de equilíbrio de longo prazo é bastante frequente em economia. Em praticamente todos os seus capítulos temos exemplos de ajustamentos de curto e de longo prazo e do caminho do curto ao longo prazo. O economista está habituado a distinguir um e outro tipo de equilíbrio. Os pais da economia ensinaram os economistas a analisar unicamente o longo prazo, com as suas leis tendenciais, ou naturais. Apenas com o aparecimento de *Keynes*, e a sua contestação às ideias da época, os economistas deixaram de ser escravos desse longínquo tempo sem fim. Infelizmente alguns economistas confundem o Keynes quantitativista, que denunciava que no longo prazo estaríamos todos mortos, com o Keynes já keynesiano. Talvez hoje assistamos a algum exagero inverso, que leva os economistas a preocuparem-se exclusivamente com o curto (e o muito curto) prazo.

As relações de longo prazo, ou de equilíbrio, são vistas em termos da nova econometria como relações entre os níveis das variáveis integradas de ordem 1. A relação que daí resulta tem as características de estacionaridade, ou seja, de desvios relativamente ao equilíbrio. Funcionando a relação de equilíbrio como um atractor, as variáveis presentes nesse equilíbrio vão reagir num período subsequente ao do desequilíbrio para o corrigir. O que nos leva às relações entre primeiras diferenças das variáveis do modelo e os valores de desequilíbrio. Ao qual chamamos comportamento de curto prazo, porque traduz a evolução, período a período, das alterações das variáveis para atingirem a situação descrita na relação de longo prazo.

Vejam, pois, alguns exemplos retirados da análise económica aplicados apenas a uma relação de longo prazo: o que significa que valorizamos modelos ECM. Mais à frente veremos a equivalência entre processo ECM e cointegração.

3.1.1 Procura de moeda

As diversas formulações da procura de moeda, com diferentes implicações macroeconómicas e o seu comportamento aparentemente instável, entusiasmaram muito os economistas nos finais dos setenta e nos anos oitenta. Os investigadores não dispunham na altura dos conhecimentos econométricos que apenas mais tarde seriam conhecidos e popularizados. O livro de Hoffman e Rasche (1996) é um marco na aplicação das novas metodologias à procura de moeda. Também a ideia de uma detenção de encaixes monetários no sentido de posse de um *buffer stock* (Mizen (1994)) levou a modelar a procura de moeda em termos de comportamentos de equilíbrio e de ajustamento a esse equilíbrio.

Vejam, uma forma comum a esta ideia de expressar a procura de moeda,

sendo as variáveis tomadas em logaritmos

$$\begin{aligned} \Delta M_t = & \alpha_0 + \lambda (M_{t-1} - \beta_1 \cdot Y r_{t-1} - \beta_2 \cdot P_{t-1} - \beta_3 \cdot r_{t-1}) + \\ & + \alpha_1(L) \cdot \Delta M_{t-1} + \alpha_2(L) \cdot \Delta Y r_{t-1} + \alpha_3(L) \cdot \Delta P_{t-1} + \\ & + \alpha_1(L) \cdot \Delta r_{t-1} + \varepsilon_t \end{aligned} \quad (3.1)$$

Os $\alpha_j(L)$ são polinómios de desfasamento de ordem não fixa, a determinar.

A parte da equação que está entre parênteses traduzirá o comportamento de longo prazo da procura de moeda, enquanto que a equação em si traduz o comportamento de curto prazo, ou de ajustamento ao longo prazo. A equação de longo prazo pode ser assim escrita

$$M_t = \beta_1 \cdot Y r_t + \beta_2 \cdot P_t + \beta_3 \cdot r_t \quad (3.1.a)$$

onde temos as diferentes elasticidades constantes¹. O valor de λ deve ser negativo ($\lambda < 0$). O que traduz o processo de correcção de M , (ΔM), ao seu valor de equilíbrio de longo prazo. Se $M_t - \hat{M}_t > 0$, então o valor de ΔM_{t+1} (< 0) deve corrigir essa diferença positiva².

3.1.2 Função consumo

O trabalho pioneiro na nova formulação empírica da função consumo deve-se a J. Davidson e Yeo (1978) e envolveu um estudo empírico aplicado ao Reino Unido.

A separação entre a componente permanente e transitória do consumo agregado pode levar-nos à seguinte formulação

$$\Delta C = \beta_0 + \lambda \cdot (C - \beta_P \cdot Y)_{-1} + \alpha_1(L) \cdot \Delta C_{-1} + \alpha_2(L) \cdot \Delta Y_{-1} \quad (3.2)$$

onde fazemos uso de uma outra forma, também convencional, de indicar os períodos.

No caso da equação (3.2) podemos ver que o longo prazo se caracteriza por uma função com propensão média idêntica à propensão marginal a consumir.

3.1.3 Eficiência em mercados cambiais

A simplificação que vamos fazer é talvez excessiva. Veja-se a este propósito a apresentação consagrada em Hallwood e MacDonald (2000), pp. 255-63. Voltemos a utilizar logaritmos. Seja f_t a cotação a prazo (*forward*) de uma divisa e s_t o seu valor à vista (*spot*). Fazendo uso do conceito de valor esperado podemos escrever

$$E_t [s_{t+1}] = f_t \quad (3.3)$$

¹Veremos mais à frente que não devemos tomar aqueles coeficientes como elasticidades constantes em modelos com várias equações. Obviamente que estamos a admitir que as variáveis estão em logaritmos.

²A presença de outras relações de longo prazo também é possível. Mas é preferível usar exemplos simples.

Se considerarmos que as antecipações são correctas, então os agentes, ao usarem toda a informação disponível, obtêm valores que podem ser expressos como

$$s_{t+1} - E_t [s_{t+1}] = \varepsilon_{t+1} \quad (3.3.a)$$

onde ε_t tem as características habituais de um ruído branco.

Podemos, assim, testar a hipótese de eficiência estudando a relação de equilíbrio

$$s_{t+1} = f_t + \varepsilon_{t+1}$$

que pode ser integrada numa outra que se refira explicitamente aos ajustamentos de curto prazo da taxa de câmbio.

3.1.4 Paridade do poder de compra

Da mesma forma, a paridade do poder de compra na determinação das taxas de câmbio pode ser testada pela relação

$$s_t + p_t^* - p_t = \varepsilon_t \quad (3.4)$$

incluída em

$$\Delta s_t = \beta_0 + \lambda \cdot \varepsilon_{t-1} + \alpha_1(L) \cdot \Delta p_t^* + \alpha_2(L) \cdot \Delta p_t + \mu_t \quad (3.4.a)$$

onde p se refere aos preços internos e p^* aos preços externos e as variáveis continuam em logaritmos.

Esta formulação deve ser encarada como a mais adequada, na medida em que a respectiva teoria nunca defendeu a sua verificação no curto prazo, surgindo mais como uma restrição de longo prazo às relações entre taxas de câmbio.

3.1.5 Despesas do Estado

As despesas do Estado também podem ser modeladas procurando ter em conta um comportamento de longo prazo à *Wagner* e uma dinâmica de curto prazo

$$\Delta G_t = \beta_0 + \lambda \cdot (G_{-1} - \beta_G \cdot Y_{-1}) + \alpha_1(L) \cdot \Delta G + \alpha_2(L) \cdot \Delta Y + \mu \quad (3.5)$$

Pensamos que já tenha ficado claro que o coeficiente λ representa, em equações deste tipo, a velocidade de ajustamento ao equilíbrio.

Imaginemos um sistema como o que se segue

$$\begin{aligned} \Delta G_t &= \beta_0 + \lambda_G \cdot (G_{-1} - \beta_G \cdot Y_{-1}) + \alpha_1(L) \cdot \Delta G + \alpha_2(L) \cdot \Delta Y + \mu \\ \Delta Y_t &= \beta_0^* + \lambda_Y \cdot (G_{-1} - \beta_G^* \cdot Y_{-1}) + \alpha_1^*(L) \cdot \Delta G + \alpha_2^*(L) \cdot \Delta Y + \mu \end{aligned} \quad (3.2)$$

Se neste sistema, para os ajustamentos das duas variáveis, tivéssemos os coeficientes $\lambda_G < 0 \wedge \lambda_Y = 0$, então poderíamos defender que G não determinava Y . Os afastamentos da relação de longo prazo não afectavam esta variável (Y) e, por isso, G não causaria Y .

3.2 Equivalência do MCE e da Cointegração

A nossa ideia de equilíbrio, e de erro (ou afastamento) de equilíbrio, leva-nos a definir o equilíbrio, fazendo uso da representação vectorial, como

$$\beta \cdot \mathbf{x}_t = 0 \quad (3.3)$$

e o erro de equilíbrio como

$$\beta \cdot \mathbf{x}_t = \mathbf{e}_t \quad (3.4)$$

onde \mathbf{e}_t tem as características de ruído branco normais.

As variáveis pertencentes ao vector \mathbf{x} são variáveis cointegradas de ordem d, b se forem integradas, isoladamente, de ordem d e se existir uma combinação linear entre elas que seja integrada de ordem $(d - b)$, com b positivo

$$\begin{aligned} \mathbf{x}_t &\sim CI(d, b) \\ x_{it} &\sim I(d) \\ \exists \beta &: \beta \cdot \mathbf{x}_t \sim I(d - b), b > 0 \end{aligned}$$

O vector β é designado por vector de cointegração e, para nós, representa o vector de equilíbrio de longo prazo das variáveis presentes no modelo.

A nossa exigência quanto à cointegração é assim sempre dupla.

1º queremos o significado económico da relação entre as variáveis, e

2º o respeito pelas características econométricas da relação que as envolve.

A primeira observação a fazer refere-se à propriedade de homogeneidade, à qual os economistas também se habituaram: se β é um vector de cointegração, $(c \cdot \beta)$, também é vector de cointegração, sendo c um escalar.

3.2.1 Um cuidado adicional: ainda o caso de regressões espúrias

O problema das regressões espúrias resume-se ao facto de regressões entre variáveis integradas, sem qualquer relação entre elas, conduzirem a valores das estatísticas t e F , que excluem a hipótese nula da inclinação, levando-nos a concluir pela existência de relacionamento quando tal não deveria acontecer.

Sigamos o exemplo de Granger e Newbold (1974). Tomemos o caso em que o nosso modelo de regressão envolve variáveis não estacionárias. Tomemos duas variáveis, x e y , *random walk*

$$\begin{aligned} y_t &= y_{t-1} + \mu_t \\ x_t &= x_{t-1} + \varepsilon_t \end{aligned} \quad (3.5)$$

com as seguintes características, $\mu_t \sim i.i.d.(0, \sigma_\mu^2)$, $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$, $E[\mu_t \varepsilon_k] \forall t, k$, $E[\mu_t \mu_{t-k}] = E[\varepsilon_t \varepsilon_{t-k}] = 0 \forall k \neq 0$.

O modelo com y , como variável dependente, virá

$$y_t = \beta_0 + \beta_1 \cdot x_t + v_t \quad (3.6)$$

Como y_t e x_t não estão correlacionados, y não exerce nenhuma influência sobre x , e esta última variável também não exerce qualquer influência sobre y , acreditamos que β_1 seja nulo e que R^2 tenda para zero. Pois bem, não é este o caso.

Se duas variáveis tiverem valores crescentes, então o R^2 afasta-se de zero, ainda que os motivos que levam uma variável a crescer nada tenham a ver com os motivos que levam a outra variável também a crescer e os incrementos em cada uma delas não estejam correlacionados. De notar que em (3.6) não só é errado tomar $\beta_1 \neq 0$, como $\beta_1 = 0$ ($y = \beta_0 + v_t$). Também o teste conjunto ($\beta_0 = \beta_1 = 0$) não deve ser aplicado no caso das variáveis como (3.5) (Phillips (1986)).

Normalmente, um bom indicador de uma regressão espúria é o valor muito reduzido que se obtém para a estatística de Durbin-Watson. Granger e Newbold (1974) propõem mesmo que tomemos as regressões onde $R^2 > DW$ como sendo regressões com muito forte probabilidade de serem espúrias.

Foi frequente, no passado, a obtenção de equações deste tipo sem que os economistas percebessem a ausência de sentido do que acabavam de obter, mas que lhes agradava por confirmar as suas posições teóricas. Como veremos, a econometria de variáveis não estacionárias, mais propriamente, de variáveis com raiz unitária, resolveu este tipo de problema de forma adequada.

3.2.2 Equivalência MCE / Cointegração

Procuramos agora comparar um sistema de ajustamento baseado no princípio de correção dos erros com um outro baseado numa relação de cointegração. Utilizemos duas variáveis (y e z) com o mesmo número de desfasamentos nas equações em primeiras diferenças. Engle e Granger (1987) provaram que a cointegração conduz a um efeito de *feedback* negativo nos valores desfasados das variáveis em níveis e um efeito negativo de *feedback* dessas variáveis implica a existência de cointegração.

Tomemos pois um sistema com duas variáveis integradas de ordem um [$y, x \sim I(1)$] e limitemos a um os desfasamentos a introduzir nas equações de ajustamento de curto prazo

$$\Delta y_t = \beta_{10} + \lambda_1 \cdot (y_{t-1} - \beta_{11} \cdot z_{t-1}) + \beta_{12} \cdot \Delta y_{t-1} + \beta_{13} \cdot \Delta z_{t-1} + \varepsilon_{yt} \quad (3.7)$$

$$\Delta z_t = \beta_{20} + \lambda_2 \cdot (y_{t-1} - \beta_{21} \cdot z_{t-1}) + \beta_{22} \cdot \Delta y_{t-1} + \beta_{23} \cdot \Delta z_{t-1} + \varepsilon_{zt} \quad (3.8)$$

Este sistema, com uma representação MCE, pode tomar uma forma diferente, mas equivalente

$$\begin{aligned}\Delta y_t &= \beta_{10} + \lambda_1 \cdot y_{t-1} - \lambda_1 \cdot \beta_{11} \cdot z_{t-1} + \beta_{12} \cdot \Delta y_{t-1} + \beta_{13} \cdot \Delta z_{t-1} + \varepsilon_{yt} \\ \Delta z_t &= \beta_{20} + \lambda_2 \cdot y_{t-1} - \lambda_2 \cdot \beta_{21} \cdot z_{t-1} + \beta_{22} \cdot \Delta y_{t-1} + \beta_{23} \cdot \Delta z_{t-1} + \varepsilon_{zt}\end{aligned}$$

Se atendermos a que podemos representar por vectores e matrizes os seguintes parâmetros

$$\begin{aligned}\mathbf{M}_1 &= \begin{bmatrix} \lambda_1 & -\lambda_1 \cdot \beta_{11} \\ \lambda_2 & -\lambda_2 \cdot \beta_{21} \end{bmatrix}, \mathbf{M}_2 = \begin{bmatrix} \beta_{12} & \beta_{13} \\ \beta_{22} & \beta_{23} \end{bmatrix}, \varepsilon_t = \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}, \\ \mathbf{m}_o &= \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} \text{ e } \mathbf{x}_t = \begin{bmatrix} y_t \\ z_t \end{bmatrix}\end{aligned}$$

e então podemos condensar aquele sistema da seguinte forma

$$\Delta \mathbf{x}_t = \mathbf{m}_o + \mathbf{M}_1 \cdot \mathbf{x}_{t-1} + \mathbf{M}_2 \cdot \Delta \mathbf{x}_{t-1} + \varepsilon_t \quad (3.9)$$

ou ainda

$$\mathbf{M}_1 \cdot \mathbf{x}_{t-1} = \Delta \mathbf{x}_t - \mathbf{m}_o - \mathbf{M}_2 \cdot \Delta \mathbf{x}_{t-1} - \varepsilon_t \quad (3.10)$$

Sendo para o caso geral de p desfazamentos

$$\mathbf{M}_1 \cdot \mathbf{x}_{t-1} = \Delta \mathbf{x}_t - \mathbf{m}_o - \sum_{j=2}^p \mathbf{M}_j \cdot \Delta \mathbf{x}_{t-j+1} - \varepsilon_t \quad (3.11)$$

No membro direito de (3.11) encontramos uma combinação linear entre variáveis estacionárias pelo que a característica de estacionaridade está assegurada. E esta característica de estacionaridade do membro direito garante a estacionaridade do membro esquerdo. Se tivermos $\mathbf{M}_1 = 0$, então temos um modelo VAR com variáveis estacionárias. No caso geral, temos $\mathbf{M}_1 \neq 0$, e é a esta situação que aplicamos as nossas ideias de equilíbrio e de afastamento do equilíbrio.

Recapitulemos. Como temos

$$\mathbf{M}_1 \cdot \mathbf{x}_{t-1} \sim I(0)$$

e sabendo que cada uma das variáveis \mathbf{x} , (y, z) , é integrada de ordem um

$$\mathbf{x} \sim I(1)$$

podemos afirmar que estamos perante variáveis cointegradas

$$\mathbf{x} \sim CI(1, 1)$$

Este resultado significa que a representação em termos de mecanismo de correcção dos erros (MCE) é equivalente à representação em termos de variáveis cointegradas. Falamos afinal do mesmo, num e noutro caso.

3.3 Obtenção das Relações de Cointegração

Começamos por apresentar a metodologia proposta por *Engle-Granger* e depois de tecermos alguns comentários acerca dos seus limites, passaremos à exposição da metodologia de *Johansen*.

3.3.1 Método de Engle-Granger

Iremos fazer a apresentação nos dois passos tradicionais. Começamos por supor que as variáveis y e z são variáveis integradas de ordem 1. Sendo assim, podemos fazer

$$y_t = \beta_0 + \beta_1 \cdot z_t + \varepsilon_t \quad (3.12)$$

de forma a obter valores para os erros que sejam estacionários. Se, porventura, esses erros, $\hat{\varepsilon}_t$, apresentarem uma raiz unitária as variáveis não são cointegradas³. Como esta nova variável é obtida pela regressão acima, Engle e Granger (1987) calcularam as tabelas apropriadas para aquele tipo de raiz unitária para o caso de duas variáveis e Engle e Yoo (1987) para o caso de mais de duas variáveis.

Se $y, z \sim CI(1, 1)$ então podemos, depois de estimar (3.12), passar à estimação do sistema

$$\begin{aligned} \Delta y_t = & \alpha_1^{(1)} \cdot \left(y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 \cdot z_{t-1} \right) + \sum_{i=1}^{k_1} \alpha_{2i}^{(1)} \cdot \Delta y_{t-i} + \\ & + \sum_{j=1}^{k_2} \alpha_{3j}^{(1)} \cdot \Delta z_{t-j} + \varepsilon_{1t} \end{aligned} \quad (3.13)$$

$$\begin{aligned} \Delta z_t = & \alpha_1^{(2)} \cdot \left(y_{t-1} - \hat{\beta}_0 - \hat{\beta}_1 \cdot z_{t-1} \right) + \sum_{i=1}^{k_3} \alpha_{2i}^{(2)} \cdot \Delta y_{t-i} + \\ & + \sum_{j=1}^{k_4} \alpha_{3j}^{(2)} \cdot \Delta z_{t-j} + \varepsilon_{2t} \end{aligned} \quad (3.14)$$

onde os coeficientes $\hat{\beta}_0, \hat{\beta}_1$ já foram determinados por (3.12).

Um problema que deve ser resolvido é, obviamente, o da auto-correlação que poderá estar presente naquelas equações. O remédio, como sempre, reside na dinamização adequada do modelo. Não esqueçamos também que os erros de uma

³A possibilidade de $\beta_1 = 0$ é excluída pela hipótese de y e z serem, cada uma, integradas de ordem 1, $I(1)$.

das equações podem estar correlacionados com os erros da outra equação, se admitirmos efeitos contemporâneos entre Δy e Δz .

Os parâmetros $\alpha_1^{(1)}, \alpha_1^{(2)}$ medem as velocidades de ajustamento das variáveis aos respectivos valores de equilíbrio. Não faz sentido, em geral, que estes valores sejam muito elevados. Se porventura $\alpha_1^{(2)}$ for nulo, podemos concluir que a variável y não exerce influência sobre a variável z . Esta variável z é dita fracamente exógena.

Este processo, que acabamos de descrever, evolui em vários passos, e está sujeito a algumas críticas.

- Num modelo a duas variáveis que erros tomar: os da função $y(z)$ ou $z(y)$? Sabemos que para um número infinito de observações é indiferente um ou outro começo, mas em economia as nossas observações são, em geral, em número reduzido.
- Os problemas complicam-se com três ou mais variáveis. Este método não fornece uma metodologia precisa para estes casos.
- E, como em todos os processos em dois passos, os erros introduzidos no primeiro passo, logicamente que ficam presentes no segundo.

Hoje, não é muito complicado resolver algumas das questões postas pelas críticas à resolução (simplificada) daqueles processos não lineares. Por métodos não lineares podemos estimar directamente o sistema acima, num único passo. Mas obviamente que não respondemos à questão da complexidade que deriva do uso de três e mais variáveis. Felizmente que o método de Johansen(1988), utilizando a técnica de máxima verosimilhança, resolve com eficácia todos estes problemas.

3.3.2 Cointegração à Johansen

Iremos seguir Johansen (1995), Hansen e Juselius (1995) e Bação (1999).

Tomemos um processo \mathbf{x} de raiz unitária, do tipo

$$\mathbf{x}_t = \mathbf{A}_1 \cdot \mathbf{x}_{t-1} + \varepsilon_t \quad (3.15)$$

ao qual podemos dar a seguinte configuração

$$\begin{aligned} \Delta \mathbf{x}_t &= \mathbf{A}_1 \cdot \mathbf{x}_{t-1} - \mathbf{x}_{t-1} + \varepsilon_t \\ &= (\mathbf{A}_1 - \mathbf{I}) \cdot \mathbf{x}_{t-1} + \varepsilon_t \end{aligned}$$

e finalmente

$$\Delta \mathbf{x}_t = \mathbf{\Pi} \cdot \mathbf{x}_{t-1} + \varepsilon_t \quad (3.16)$$

A característica de $\mathbf{\Pi}$ dar-nos-á o número de vectores de cointegração presentes entre as variáveis do vector \mathbf{x} .

No caso extremo dessa característica ser nula,

$$\text{Rank}(\mathbf{\Pi}) = 0$$

não teremos vectores cointegrados.

A expressão acima, (3.16), pode tomar a forma geral

$$\Delta \mathbf{x}_t = \mathbf{\Pi} \cdot \mathbf{x}_{t-1} + \mathbf{A}_0 + \varepsilon_t \quad (3.17)$$

no caso de \mathbf{A}_0 ser diferente de zero, temos a presença de variáveis deterministas na explicação dos valores do vector \mathbf{x} . Estas variáveis podem ser uma constante e uma tendência temporal, por exemplo.

Suponhamos que a característica daquela matriz é igual à unidade. Neste caso, temos um vector de cointegração e, como já recordámos atrás, todos os outros vectores possíveis limitam-se a combinações lineares deste. Vejamos como podemos representar o sistema acima condensado

$$\begin{aligned} \Delta x_{1t} &= \Pi_{11} \cdot x_{1t-1} + \Pi_{12} \cdot x_{2t-1} + \dots + A_{01} + \varepsilon_{1t} \\ \Delta x_{2t} &= s_2 (\Pi_{11} \cdot x_{1t-1} + \Pi_{12} \cdot x_{2t-1} + \dots) + A_{02} + \varepsilon_{2t} \\ &\dots\dots\dots \\ \Delta x_{kt} &= s_k (\Pi_{11} \cdot x_{1t-1} + \Pi_{12} \cdot x_{2t-1} + \dots) + A_{0k} + \varepsilon_{kt} \end{aligned}$$

No caso específico de $A_{0j} = s_j \cdot A_{01}$, a constante passa para dentro do parênteses, o que leva a retirar dos níveis das variáveis \mathbf{x} o comportamento de tendência temporal.

A obtenção dos vectores de cointegração

Tomemos um processo auto-regressivo de ordem p para k variáveis

$$\mathbf{x}_t = \mathbf{A}_1 \cdot \mathbf{x}_{t-1} + \mathbf{A}_2 \cdot \mathbf{x}_{t-2} + \dots + \mathbf{A}_p \cdot \mathbf{x}_{t-p} + \varepsilon_t \quad (3.18)$$

a subtracção de \mathbf{x}_{t-1} de cada membro conduz-nos a

$$\Delta \mathbf{x}_t = (\mathbf{A}_1 - \mathbf{I}) \cdot \mathbf{x}_{t-1} + \mathbf{A}_2 \cdot \mathbf{x}_{t-2} + \dots + \mathbf{A}_p \cdot \mathbf{x}_{t-p} + \varepsilon_t$$

se agora somarmos e subtrairmos $(\mathbf{A}_1 - \mathbf{I}) \cdot \mathbf{x}_{t-2}$ chegamos a

$$\Delta \mathbf{x}_t = (\mathbf{A}_1 - \mathbf{I}) \cdot \Delta \mathbf{x}_{t-1} + (\mathbf{A}_2 + \mathbf{A}_1 - \mathbf{I}) \cdot \mathbf{x}_{t-2} + \dots + \mathbf{A}_p \cdot \mathbf{x}_{t-p} + \varepsilon_t$$

se continuarmos e somarmos e subtrairmos $(\mathbf{A}_2 + \mathbf{A}_1 - \mathbf{I}) \cdot \mathbf{x}_{t-3}$

$$\begin{aligned} \Delta \mathbf{x}_t &= (\mathbf{A}_1 - \mathbf{I}) \cdot \Delta \mathbf{x}_{t-1} + (\mathbf{A}_2 + \mathbf{A}_1 - \mathbf{I}) \cdot \Delta \mathbf{x}_{t-2} + \\ &\quad + (\mathbf{A}_3 + \mathbf{A}_2 + \mathbf{A}_1 - \mathbf{I}) \cdot \mathbf{x}_{t-3} + \dots + \mathbf{A}_p \cdot \mathbf{x}_{t-p} + \varepsilon_t \end{aligned}$$

e finalmente, se generalizarmos, obtemos

$$\Delta \mathbf{x}_t = \sum_{i=1}^{p-1} \mathbf{\Pi}_i \cdot \Delta \mathbf{x}_{t-i} + \mathbf{\Pi} \cdot \mathbf{x}_{t-p} + \varepsilon_t \quad (3.19)$$

onde

$$\mathbf{\Pi} = - \left(\mathbf{I} - \sum_{i=1}^p \mathbf{A}_i \right)$$

e

$$\mathbf{\Pi}_i = - \left(\mathbf{I} - \sum_{j=1}^i \mathbf{A}_j \right)$$

A característica de $\mathbf{\Pi}$ dá-nos o número de vectores de cointegração. No caso de ser nula, estamos perante um VAR normal⁴. Se tivermos o valor k , idêntico ao número das variáveis do modelo, então o vector das variáveis é estacionário e se tivermos um valor entre 1 e k , teremos esse número de vectores independentes de cointegração.

O valor da característica daquela matriz é o número de valores próprios associados à matriz que são diferentes de zero. Sabemos que os valores de λ_i se obtêm da resolução de $|\mathbf{\Pi} - \lambda \cdot \mathbf{I}|$ e uma raiz nula implica que $|\mathbf{\Pi}|$ seja nulo, pelo que pelo menos uma fila será não independente das restantes. Tomemos λ_i como representando o valor próprio i e ordenemos os diferentes valores próprios por ordem decrescente

$$\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_k$$

Se a característica for nula, todos os λ_i serão nulos ou, de outra forma mais útil,

$$\ln(1 - \lambda_i) = 0$$

Se a característica for igual à unidade, então $0 < \lambda_1 < 1$, e assim

$$\ln(1 - \lambda_1) < 0$$

sendo neste caso

$$\ln(1 - \lambda_j) = 0, \text{ para } \forall j \neq 1$$

para todas os outros valores próprios.

O problema que temos de resolver é saber quantos valores próprios são diferentes de zero, ou quantos obedecem à condição

$$(1 - \lambda_i) \neq 1$$

Dois testes foram propostos para responder a tal questão. O primeiro vem dado por

$$\lambda_{traço}(r) = -N \cdot \sum_{i=r+1}^k \ln(1 - \hat{\lambda}_i) \quad (3.20)$$

⁴Como já atrás dissemos.

- que testa a H_0 de o número de vectores de cointegração distintos ser em número inferior ou igual a r . Quanto mais afastados de zero estiverem os valores de λ_i , mais elevado será o valor daquela estatística.

O segundo teste é dado por

$$\lambda_{\max}(r, r+1) = -N \cdot \ln \left(1 - \hat{\lambda}_{r+1} \right) \quad (3.21)$$

- que testa a H_0 de o número de vectores de cointegração ser r contra a hipótese alternativa de $r+1$.

Por simulação, diversos autores obtiveram tabelas para aquelas estatísticas. Nomeadamente Osterwald-Lenum (1992), Johansen e Juselius (1990) e Soren Johansen e Niesen (2000).

Alguns testes na relação de cointegração

Devido à sua importância devemos começar por fazer o teste de exclusão de uma constante no vector de cointegração contra a sua não presença, ou não restrição da constante. Estimemos o modelo para os dois casos. Representemos $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k$ os valores próprios associados à presença da constante fora do vector, o que designamos por não restrição da constante, e por $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \dots, \hat{\lambda}_k^*$ os valores associados à integração da constante no vector de cointegração. Assintoticamente, temos a estatística

$$-N \cdot \sum_{i=r+1}^k \left[\ln \left(1 - \hat{\lambda}_i^* \right) - \ln \left(1 - \hat{\lambda}_i \right) \right] \sim \chi^2(k-r)$$

onde r foi previamente retido. Para conhecermos os graus de liberdade que devemos usar veja-se Johansen (1995). A hipótese nula consiste na restrição da constante no espaço de cointegração.

Os valores reduzidos da expressão levam-nos a não excluir a H_0 e assim não excluir a constante do vector de cointegração. Ou, de outra forma, se $\chi^2(k-r) > \chi^{2^C}(k-r)$ devemos aceitar a sua exclusão do vector de integração e em contrapartida admitir a presença de tendência temporal nas variáveis em estudo.

Vejam como também podemos impor outras restrições às relações de cointegração. Apresentamos assim alguns testes que envolvem certos valores dos parâmetros. Tomemos

$$\mathbf{\Pi} = \alpha \cdot \beta'$$

onde $\beta_{(k \times r)}$ é uma matriz de parâmetros de cointegração e $\alpha_{(k \times r)}$ se compõe dos pesos com que cada vector de cointegração entra nas equações do sistema estimado (3.19), ou seja, temos aqui os valores, já atrás designados, das velocidades de ajustamento.

Podemos pois, a partir daquele sistema (3.19), fazer

$$\Delta \mathbf{x}_t = \sum_{i=1}^{p-1} \mathbf{\Pi}_i \cdot \Delta \mathbf{x}_{t-i} + \alpha \cdot \beta' \cdot \mathbf{x}_{t-p} + \varepsilon_t$$

O que, para $r = 1$, e normalizando para a primeira variável, $\beta_1 = 1$, nos permite escrever

$$\beta = \begin{bmatrix} 1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} \text{ e } \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_k \end{bmatrix}$$

As restrições que pretendamos impor em β e α têm um tratamento idêntico às anteriores. Tomemos os valores próprios do modelo não restringido $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k$ e os novos valores associados à restrição imposta $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \dots, \hat{\lambda}_k^*$. Assintoticamente, temos a estatística

$$-N \cdot \sum_{i=1}^r \left[\ln \left(1 - \hat{\lambda}_i^* \right) - \ln \left(1 - \hat{\lambda}_i \right) \right] \sim \chi^2 \text{ (restrições em } \beta \text{ ou } \alpha \text{)}$$

que, como indicamos, pode ser aproximada pela estatística do chi-quadrado com um número de graus de liberdade igual às restrições impostas.

Exemplifiquemos com um modelo de procura de moeda. Se estivermos interessados em testar se a elasticidade preço é, no longo prazo, igual à unidade, devemos fazer um teste deste tipo, $\beta_P = 1$, e se, porventura, o valor obtido daquela estatística for inferior ao seu valor crítico, não excluimos a restrição ! Da mesma forma devemos fazer para o parâmetro α . Não esquecendo que a nossa hipótese nula consistirá no valor que impusermos a α , por exemplo $\alpha = 0$. Tenhamos também em atenção que os testes em α são verdadeiros testes sobre exogeneidade fraca, como já atrás referimos.

Exemplificação da obtenção de relações de cointegração no RATS

O reconhecimento da importância da característica de estacionaridade das séries conduziu ao desenvolvimento da econometria de variáveis não estacionárias. A análise de séries com raiz unitária, ou análise da cointegração, revolucionou os nossos anteriores conhecimentos. As suas consequências para os economistas são de profundo alcance. Ao mesmo tempo permitiu que alterássemos a hipótese, de veras irrealista, de os valores das variáveis independentes serem tomados como constantes.

As observações que aqui serão feitas não eliminam o estudo do manual Hansen e Juselius (1995) e que se destina a apresentar a metodologia e o programa *CATS in RATS*.

O procedimento para a *CI* à *Johansen* começou por ser distribuído com o *RATS* e era da autoria de *K. Juselius*. Mais tarde, a sua designação passou a ser a actual, *CATS*, da autoria de *H. Hansen* e a estar disponível na Internet. Finalmente passou a ser vendido à parte do *RATS* pela empresa sua proprietária (Hansen e Juselius (1995)). Sempre que fizermos referência ao directório “c:cats” referimo-nos a esta última versão. Se tivermos “c:oldcat”, então trata-se do penúltimo procedimento. Este difere do mais actual por um resultado diferente na opção “rank”, para o primeiro tipo de vector de *CI*.

As instruções e resultados que se seguem demonstram uma forma possível de estudo da presença de *CI* entre várias variáveis. Obviamente que não esgotam as possibilidades de cálculos e testes que são possíveis utilizando directamente os comandos da nova janela do *RATS* e que é designada por *CATS*.

As tabelas estatísticas baseadas no Traço, para os cinco modelos (Johansen (1995), p. 212) constam das pp. 214-8. Aninda Banerjee e Hendry (1993) apresentam para os modelos 1 (Quadro 8.1, p. 269), 2 (Quadro 8.7, p. 276) e 3 (Quadro 8.5, p. 274) tabelas também baseadas no valor próprio máximo e que foram retiradas de Osterwald-Lenum (1992), sendo por isso mais precisas que as precedentes.

Passemos ao nosso exemplo com um modelo que podemos apelidar do tipo IS-LM. Base de dados: massa monetária em sentido restrito (M1), PIB real a preços constantes, preços implícitos no PIB e taxa de juro das operações bancárias activas de 181 dias a um ano a empresas não financeiras. O período vai de 1997 a 2000 e tem periodicidade trimestral. à excepção das taxas de juro, que continuam em valores decimais, todas as variáveis foram transformadas em índices de base 100 para 1995. As variáveis foram depois transformadas em logaritmos. A simbologia é a seguinte: M, Q, P e R.

Começamos por questionar o tipo de variáveis deterministas que devemos incluir no estudo. Ou o que equivale ao mesmo, que tipo de modelo devemos considerar. Para isso usamos o antigo programa *CATS* (oldcat).

```
end 1
cal 1970 1 4
all 2000:4
open data base.rat
data(format=rats) /
dofor i = q p m r
log i
end dofor i
source(noecho) catsmain.src
open copy a:temp.out
@cats(proc=rank,lags=5,season=4)
# m q p r
# 'money' 'output' 'prices' 'interest'
```

Como vemos, indicamos para além das variáveis sazonais, cinco desfasamentos.

Estes desfasamentos foram escolhidos porque ao ensaiarmos modelos com cinco e seis desfasamentos chegávamos à conclusão que podíamos rejeitar a presença de auto-correlação de erros, o que não acontecia com desfasamentos inferiores. Johansen (1995) (p. 21) diz-nos que na presença de auto-correlação é preferível ensaiar o acréscimo de novas variáveis a aumentar os desfasamentos, devido à perda de graus de liberdade que isso representa. Num sistema com 4 variáveis um desfasamento a mais equivale a 16 graus de liberdade perdidos. Apesar disso quisemos estudar a hipótese de nulidade dos coeficientes quando passamos de 5 para 6 desfasamentos. Os resultados para um e outro tipo de modelo ($p_0 = 5$ e $p_1 = 6$), para o período efectivo de 1978:3 a 2000:4 foram os seguintes⁵

	p_0	p_1
$\log(\Omega)$	-31.83559	-32.11987
SC	-27.03580	-26.52011
HQ	-28.62699	-28.37650

Os critérios de Schwarz e Hannan-Quinn levam-nos a aceitar o modelo a 5 desfasamentos (p_0) contra os 6 desfasamentos (p_1). A fórmula do ratio de verosimilhança vem dada por $(N - c) \cdot (-31.83559 + 32.11987)$, onde $N - c$ vem igual a 62. Sims (1980) propôs aquele valor de c , igual ao número de parâmetros por equação em p_1 , para corrigir o uso de amostras pequenas, que é em geral o nosso caso. O valor da estatística, com $k^2 \cdot (p_1 - p_0)$ graus de liberdade, virá assim

Chi-Squared(16)= 17.625360 with Significance Level 0.34628661

ou seja, não excluimos a hipótese nula dos coeficientes $p_1 - p_0$.

Já explicámos a razão porque tomámos 5 desfasamentos, mas essa escolha fundamentou-se num tipo de modelo. Regressamos pois à análise dos resultados para os diferentes modelos, como resultam das instruções acima do RATS

Os resultados foram os seguintes

Nr.	Parametrization	Interpretation
mu(t) =		
1.	0	No deterministic components
2.	0 + a*b_0'	Intercept in the cointegration relations
3.	mu_0 +	Deterministic trends in the levels
4.	mu_0 + a*b_1*t +	Trends in the cointegration relations
5.	mu_0 + mu_1*t +	Quadratic trends in the levels

Some of the models might be excluded in advance. Therefore, please input the first and the last parametrization you wish to include in the test. Thank you. I am working on it.

The eigenvalues

rp-r		model_1	model_2	model_3	model_4	model_5
0	4	0.2238	0.2798	0.2709	0.2949	0.2908
1	3	0.1006	0.2067	0.1803	0.2122	0.1745

⁵Modelo 3. Ver mais abaixo.

2	2	0.0533	0.1002	0.0902	0.1499	0.1394
3	1	0.0008	0.0425	0.0352	0.0586	0.0043
The l-max test						
	rp-r	model_1	model_2	model_3	model_4	model_5
0	4	23.0561	29.8682	28.7562	31.7913	31.2687
1	3	9.6475	21.0723	18.0956	21.7083	17.4556
2	2	4.9825	9.6098	8.6003	14.7733	13.6622
3	1	0.0735	3.9485	3.2610	5.4988	0.3965
The trace test						
	rp-r	model_1	model_2	model_3	model_4	model_5
0	4	37.7595	64.4987	58.7131	73.7718	62.7829
1	3	14.7034	34.6305	29.9569	41.9804	31.5142
2	2	5.0559	13.5583	11.8613	20.2721	14.0586
3	1	0.0735	3.9485	3.2610	5.4988	0.3965

Fazendo uso das tabelas incluídas em Johansen (1995) (pp. 214-6) vemos que, para um valor crítico a 95%

Modelo 5. Aceitamos um vector de cointegração.

Modelo 4. Continuamos a aceitar um.

Modelo 3. Passamos a aceitar dois vectores.

Modelo 2. Aceitamos de novo apenas um vector.

Modelo 1. Não aceitamos nenhum vector.

Seguindo a metodologia de Johansen (1995) (pp. 98-100) somos levados a aceitar a presença de dois vectores e a usar o modelo 3.

As instruções adequadas à estimação do modelo 3 com 5 defasamentos são as seguintes (novo CATS)

```
source(noecho) e:catsmain.src
@cats(dettrend=drift,lags=5,season=4,rec) 1977:1 2000:4
# m q p r
```

e o resultado principal vem dado assim

```
COINTEGRATION ANALYSIS
Endogeneous series :
    M    Q    P    R
Deterministic series :
    Unrestricted constant
    3 centered seasonal dummies
Effective sample : 1978:02 TO 2000:04
Lag(s) in VAR-model : 5
No. of observations : 91
Obs.- no.of variables: 67
I(1) ANALYSIS
Eigenv. L-max Trace H0: r p-r L-max90 Trace90
0.2709 28.76 58.71 0 4 17.14 43.84
```


0.1803	18.10	29.96	1	3	13.39	26.70
0.0902	8.60	11.86	2	2	10.60	13.31
0.0352	3.26	3.26	3	1	2.71	2.71

Os valores das tabelas transcritas em Aninda Banerjee e Hendry (1993) (pp. 269-76) são os seguintes para 95%

L-Max95	Trace95
27.07	47.21
20.97	29.68
14.07	15.41
3.76	3.76

Estes valores são mais precisos que os indicados Johansen (1995). O que nos leva a reter dois vectores de cointegração pela estatística do traço e apenas um pela estatística do valor próprio máximo. No que se segue tomaremos em primeiro lugar dois vectores e depois um só vector.

Caso A: 2 vectores Normalizando para a equação da oferta de moeda e do produto, obtemos

The matrices based on 2 cointegration vectors				
BETA (transposed)				
M	Q	P	R	
1.000	-1.332	-0.790	0.194	
0.192	1.000	-0.288	0.131	
ALPHA		T-VALUES FOR ALPHA		
DM	-0.423	-0.067	-3.637	-1.483
DQ	0.108	-0.087	1.723	-3.568
DP	-0.103	-0.069	-1.334	-2.282
DR	1.030	0.024	3.652	0.220

A primeira equação pode ser reescrita como $M = 1.332 \cdot Q + 0.790 \cdot P - 0.194 \cdot R$ e a segunda como $Q = 0,288 \cdot P - 0.192 \cdot M - 0.131 \cdot R$. Num modelo destes podemos dizer que a primeira equação representa o equilíbrio monetário e a segunda o equilíbrio real. Por vezes existe a tendência para lermos estes resultados de longo prazo em termos de elasticidades. Eles devem ser lidos como atractores (Johansen (1995), p. 41) para os quais os agentes adaptam os seus comportamentos e cujas reacções aos desequilíbrios são dadas pelos valores dos α . Isto, porque num modelo com variáveis cointegradas, "A shock to one variable implies a shock to all variables in the long run, and hence the coefficients do not in general allow a *ceteris paribus* interpretation", como defendeu, nomeadamente, Lutkepohl (1994) (ver Johansen (1995), p. 50).

Podemos constatar que os coeficientes daquelas duas relações não são des-tituídos de significado económico e que os valores dos α são para cada uma delas negativos (-3.637 e -3.568) como postula a teoria. Olhando para os restantes α que são diferentes de zero, podemos ainda ver que a inflação se reduz quando o

produto é superior ao seu valor de equilíbrio e a variação da taxa de juro é positiva quando a quantidade de moeda é mais elevada que a de equilíbrio, renunciando antecipações inflacionistas.

A "análise dos resíduos" do modelo produz os seguintes resultados

```
MULTIVARIATE STATISTICS
LOG(DET(SIGMA)) = -31.66202
INFORMATION CRITERIA: SC = -27.10159
                        HQ = -28.61593
TRACE CORRELATION = 0.67552
TEST FOR AUTOCORRELATION
L-B(22), CHISQ(280) = 313.168, p-val = 0.08414
LM(1), CHISQ(16) = 19.423, p-val = 0.24734
LM(4), CHISQ(16) = 26.107, p-val = 0.05254
TEST FOR NORMALITY
CHISQ(8) = 13.220, p-val = 0.1045
UNIVARIATE STATISTICS
...
```

Onde, de forma independente, precisámos os níveis de significância dos valores do chi-quadrado através da instrução "cdf chisqr valor graus" no RATS.

Não apresentamos os restantes resultados quanto a cada uma das equações no que respeita à exclusão de processo ARCH (de ordem 5) e da normalidade (Shenton-Bowman, ver Doornik e Hansen (1994)) dos erros (ordem 2) porque optámos⁶ por obter os valores de significância das estatísticas. Em baixo estão esses valores.

	ARCH(5)	NS	Normalidade	NS
Equação de M	10.722	0.057	1.885	0.390
Equação de Q	2.239	0.815	1.217	0.544
Equação de P	3.484	0.626	8.349	0.015
Equação de R	10.019	0.075	1.277	0.528

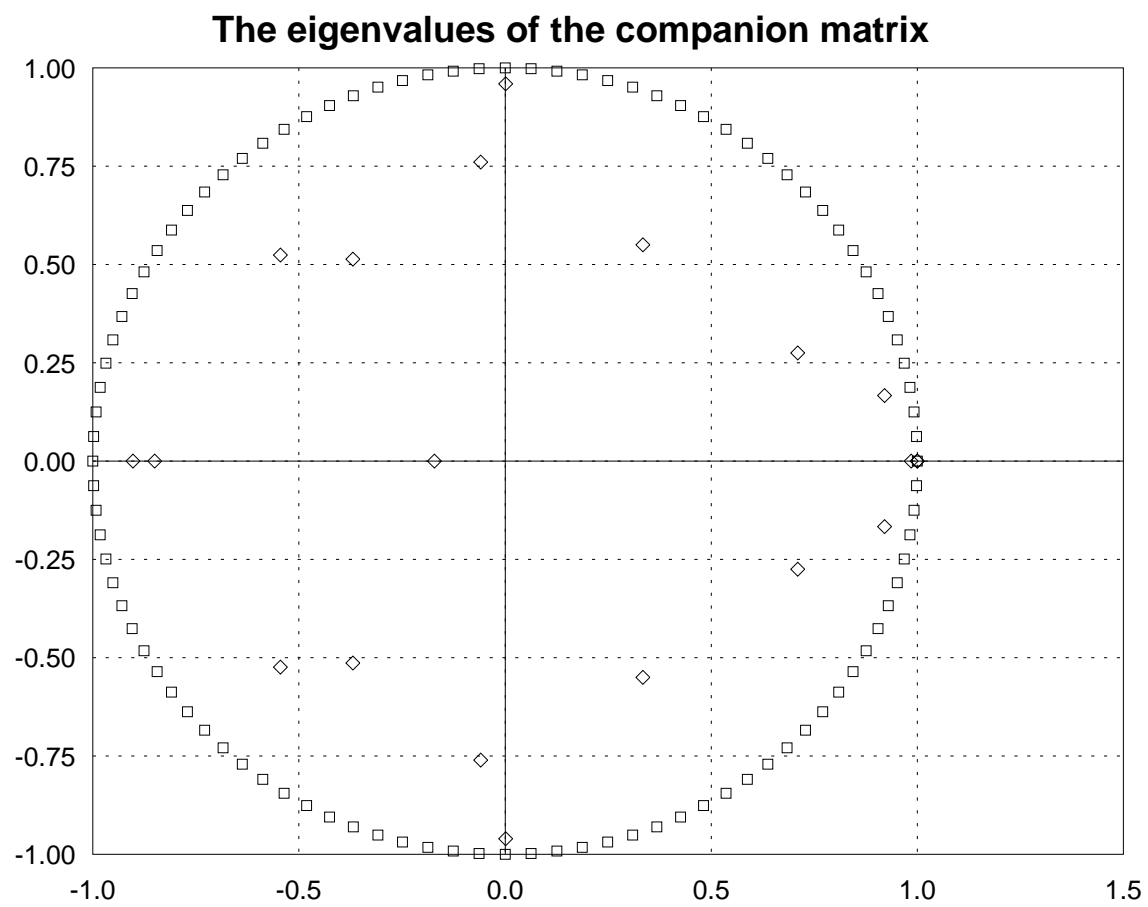
Como podemos ver, apenas na equação 3, dos preços, a normalidade dos erros não está garantida. Em todas as outras equações podemos rejeitar a presença de processo ARCH e aceitar a distribuição Normal dos erros.

Os valores próprios da matriz A, "matriz acompanhante", correspondem ao inverso das raízes do polinómio característico. O seu gráfico é o seguinte

$$\text{A matriz } A \text{ é formada por } \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & I_k & 0 \end{bmatrix}$$

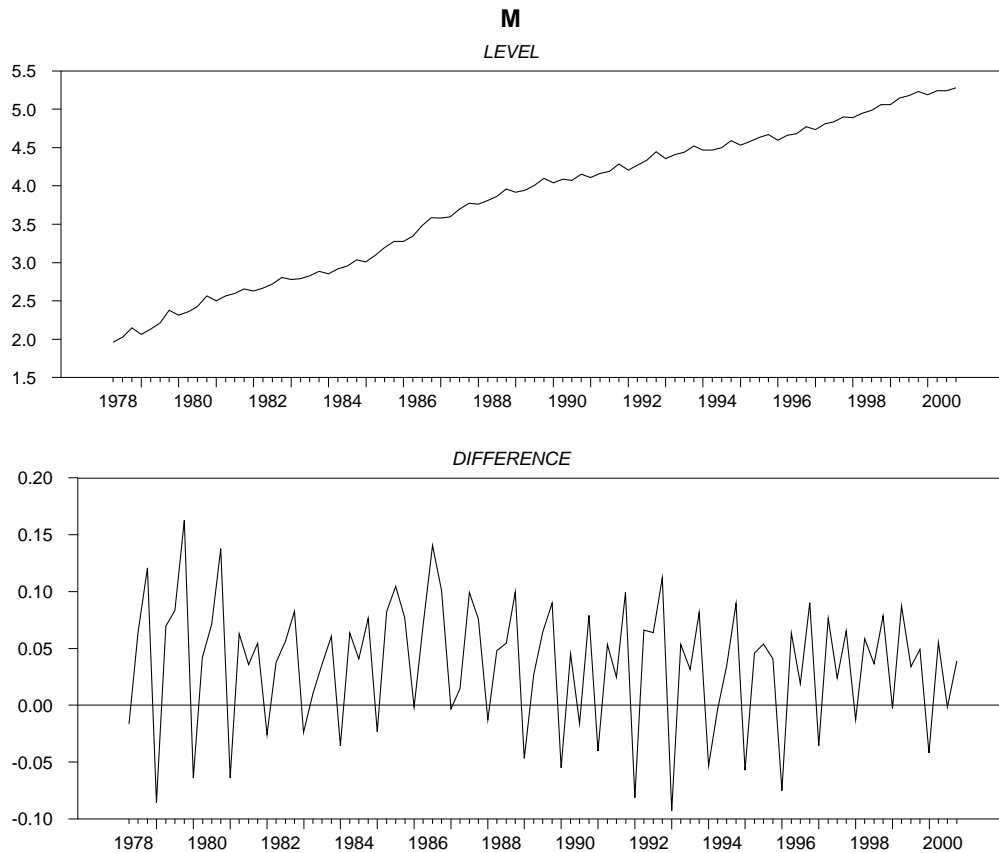
onde I_k representa a matriz identidade de ordem k .

⁶Ver a nota anterior.



Como podemos ver, todos os valores estão dentro, ou no, círculo unitário, o que traduz um processo não explosivo representado pelo modelo com as variáveis que escolhemos.

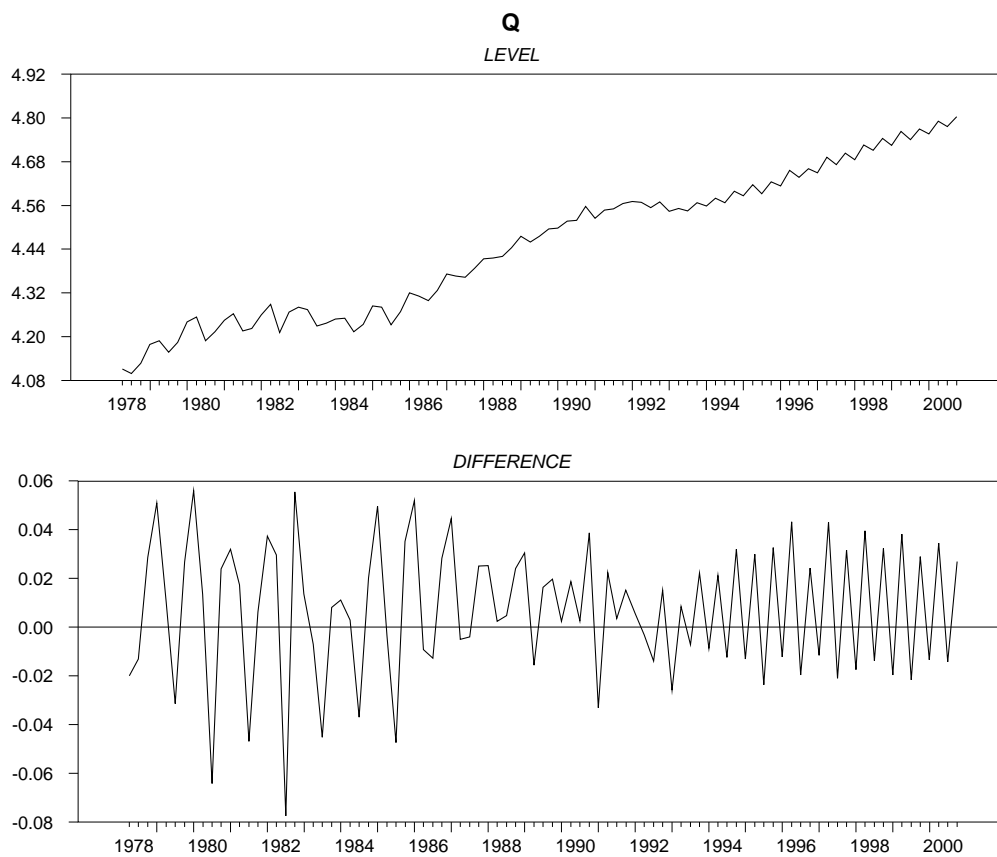
As variáveis que utilizamos confirmam a ideia de $I(1)$ em níveis e $I(0)$ em primeiras diferenças, como podemos ver nas figuras em baixo.

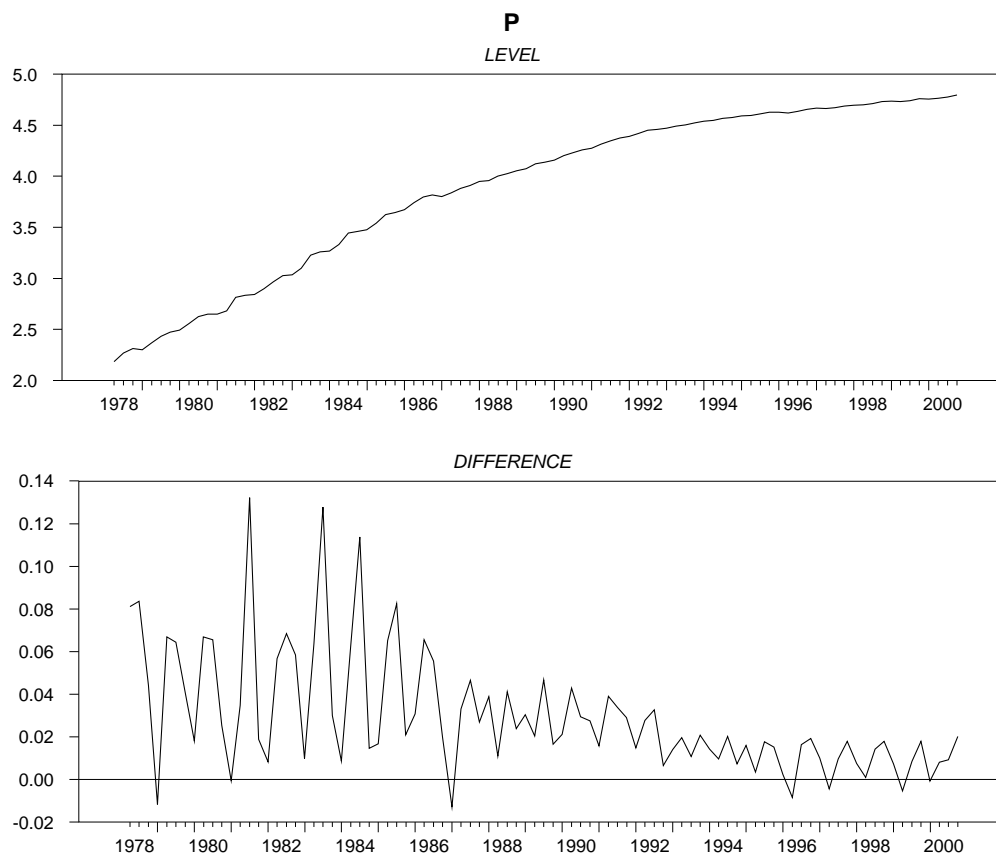


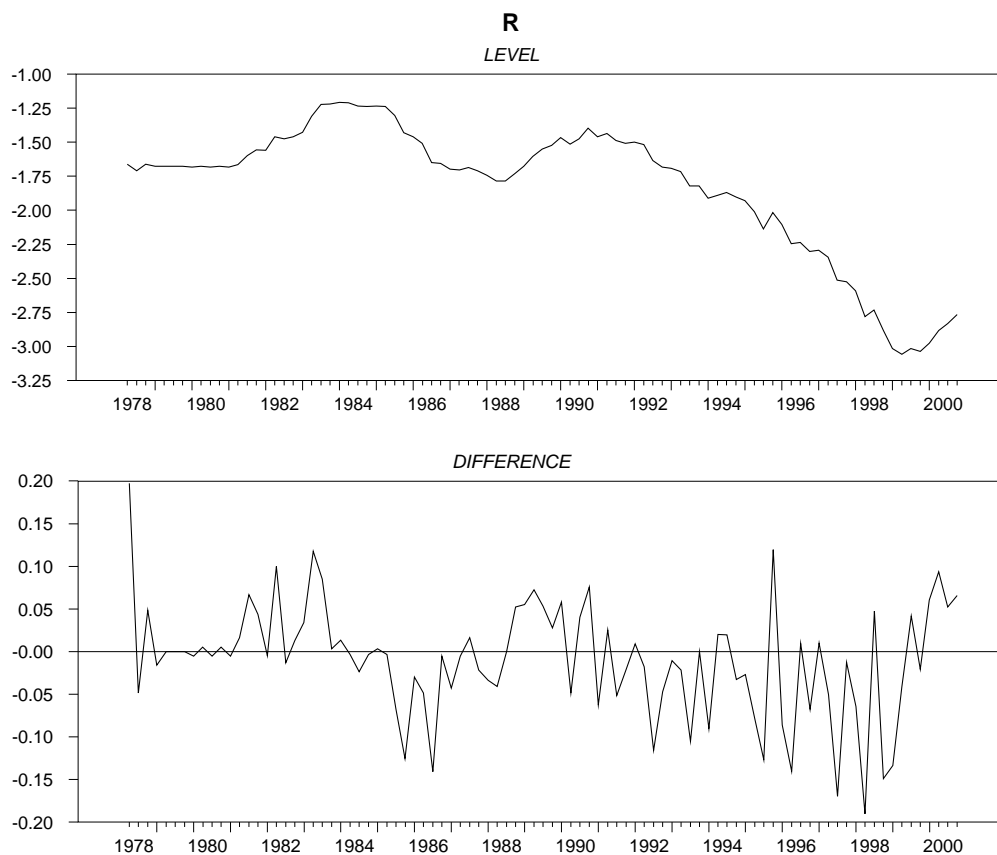
A leitura gráfica das séries usadas num modelo pode revelar-se muito importante. Ao olharmos para estas 4 séries ficamos com uma ideia clara do processo de estabilização da inflação e da redução da incerteza associada à sua evolução, assim como da convergência da taxa de juro, e ainda da “estranha” evolução do produto - cuja elucidação apenas os Santos protectores do INE poderão conhecer ...

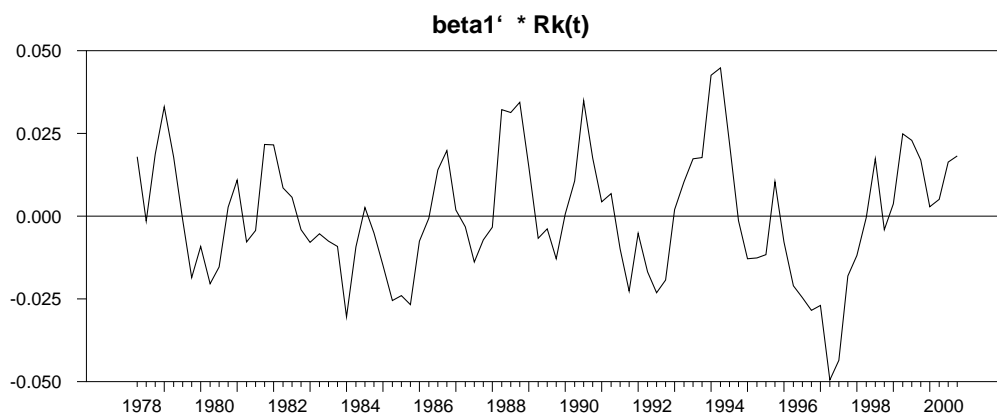
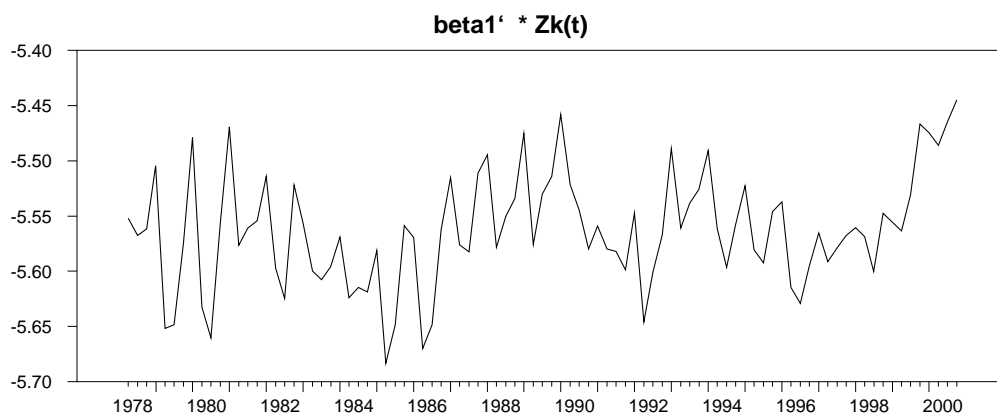
Os valores de desequilíbrio de um e outro vector traduzem resultados esperados.

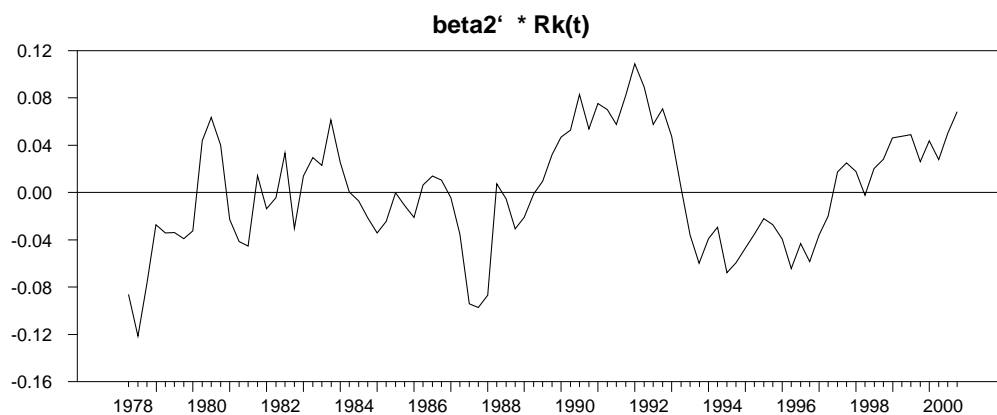
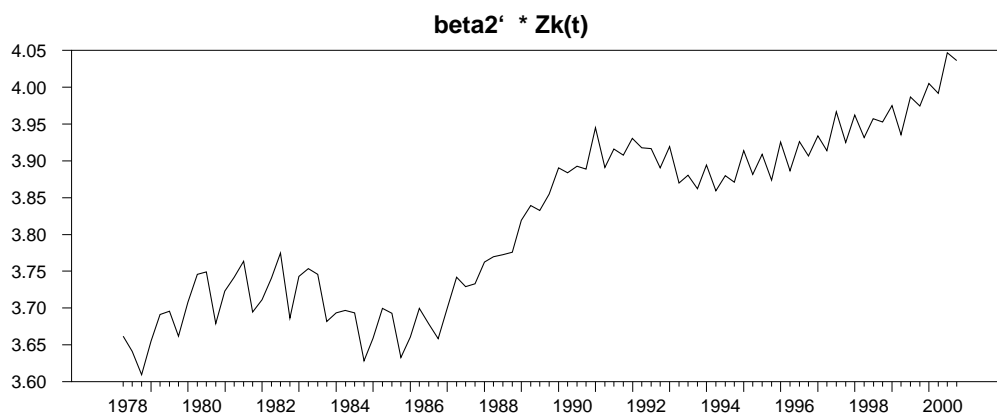
Em cada um dos gráficos a segunda curva, que é corrigida dos efeitos de curto prazo e ainda sazonais, tem uma aparência clara de variável estacionária.



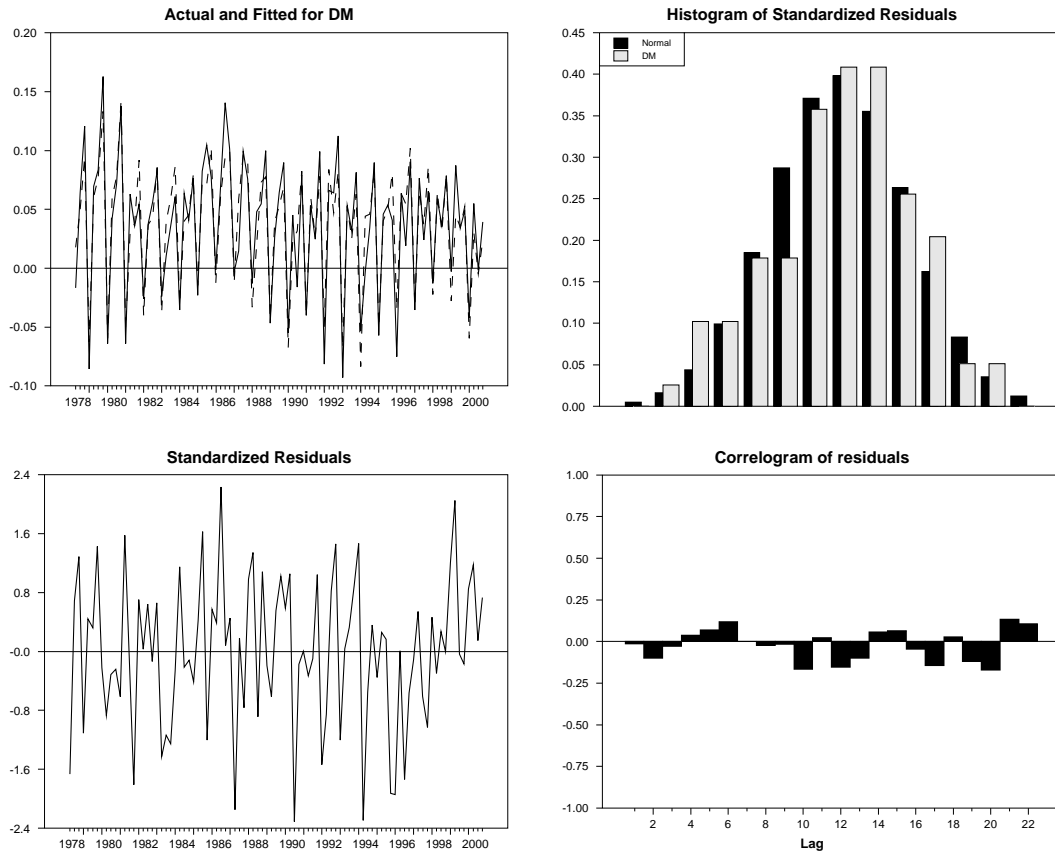






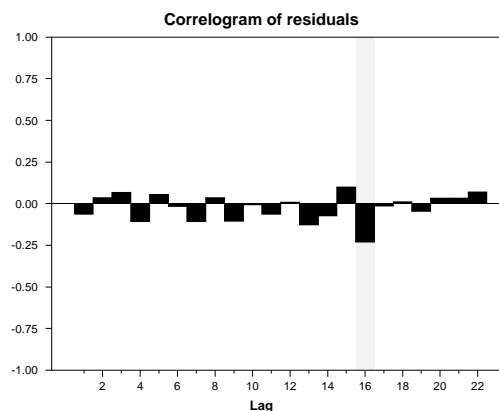
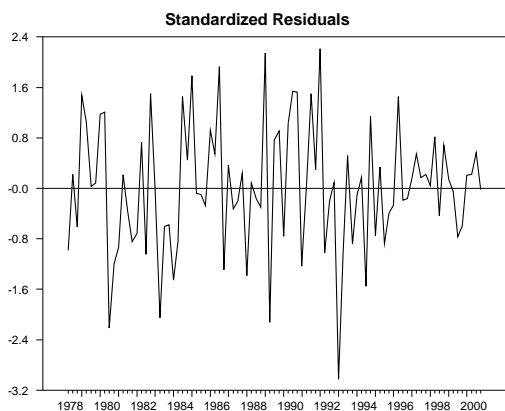
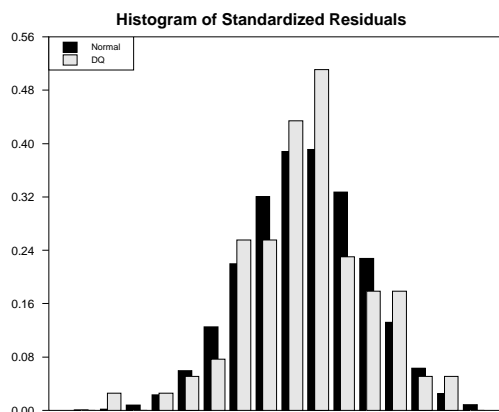
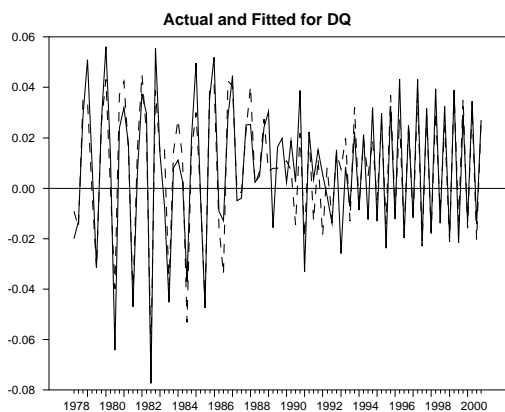


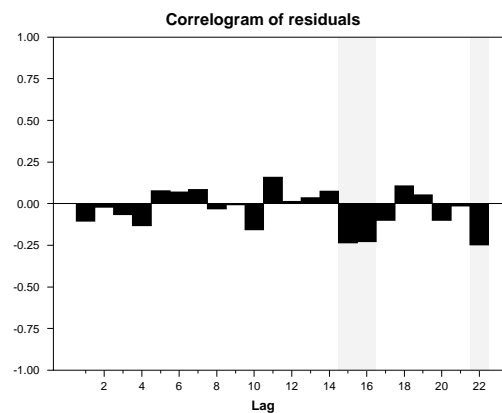
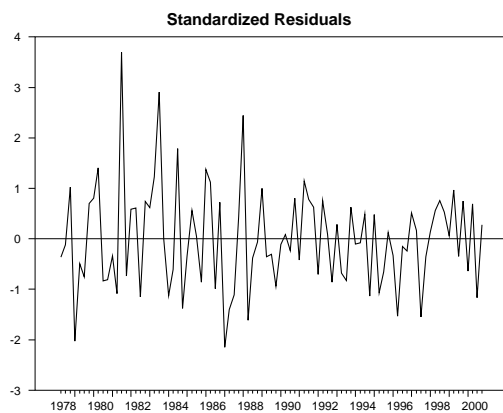
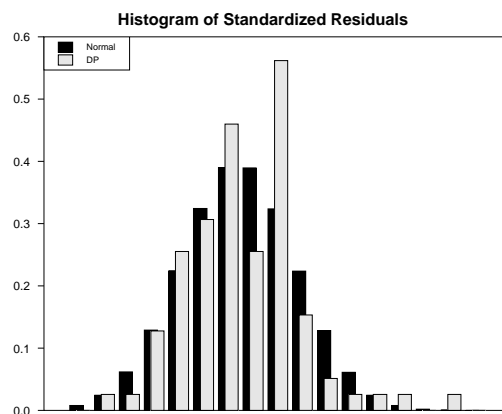
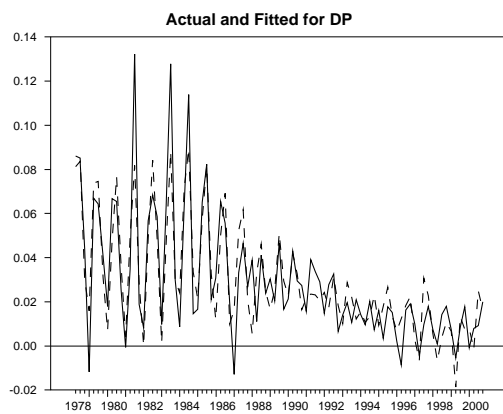
Vejam os gráficos com os valores efectivos (em primeiras diferenças) e estimados, e os erros, sua evolução, histograma e correlograma. Neste último é sombreada a situação em que o valor da correlação é superior a $|2/\sqrt{N}|$. O que acontece para os erros das três últimas equações, sem que no entanto traduzam um padrão que pudéssemos reter.

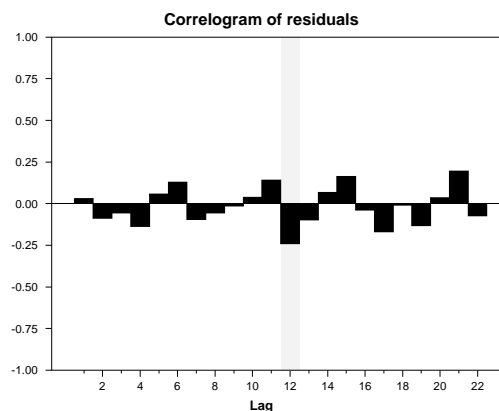
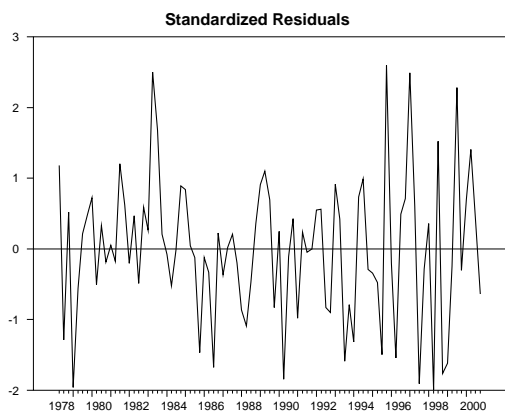
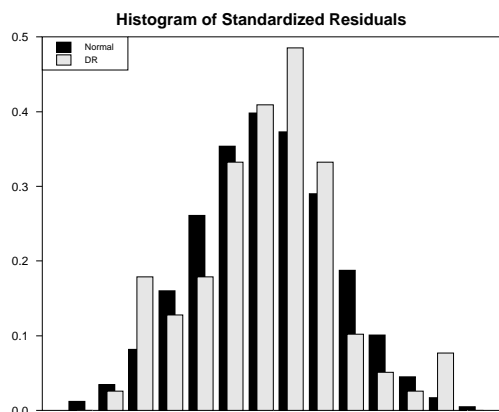
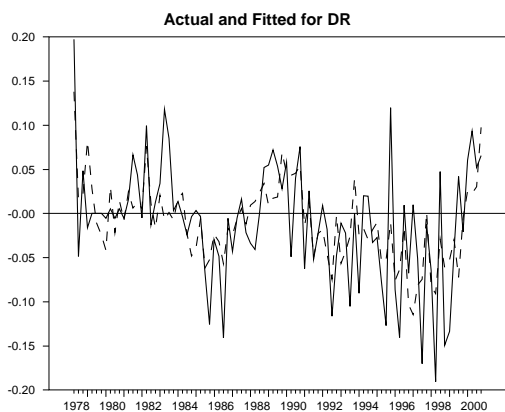


Confirmamos aquilo que tínhamos visto mais acima acerca da natureza não Normal dos erros da equação dos preços (ΔP). O R^2 das diferentes equações tem os seguintes valores: 0.86; 0.84; 0.77; e 0.44. O que é perfeitamente visível na menos boa qualidade do ajustamento da última equação (ΔR). De notar os bons resultados obtidos com o modelo para as taxas de variação da oferta de moeda e mesmo do produto, apesar do comportamento estranho desta série.

Voltemos aos vectores obtidos para o nosso modelo: $M = 1.332 \cdot Q + 0.790 \cdot P - 0.194 \cdot R$ e $Q = 0,288 \cdot P - 0.192 \cdot M - 0.131 \cdot R$. Ao olharmos para uma equação como a primeira somos levados, pelos nossos conhecimentos de análise







económica, a colocar a possibilidade de o valor do coeficiente dos preços ser igual à unidade, $\beta_{21} = 1$, assim como o do produto, $\beta_{11} = 1$. No primeiro caso teríamos um comportamento do equilíbrio monetário referido a valores reais da oferta de moeda. E se verificassem aqueles dois casos, poderíamos falar num comportamento da velocidade de circulação da moeda apenas dependente da taxa de juro. Em ambos os casos temos teorias que justificam tais comportamentos. Mas será assim para o caso em apreciação? A seguinte observação é de ter em conta: "although one's economic theory may be fine, the data chosen may not illustrate this. Hence a careful statistical analysis helps to support the economic conclusions.", (Johansen (1995), p. 5).

Este tipo de restrições de que falamos podem ser testadas de uma forma directa através da reparametrização do modelo, como $\beta^* = \mathbf{H} \cdot \varphi$, onde $\mathbf{H}_{p \times s}$ é nossa conhecida, sendo determinada $\varphi_{s \times r}$, com $(p - s)$ a representar o número de restrições a impor. No caso da reparametrização indirecta bastará fazer $\mathbf{R}'\beta = 0$.⁷

No caso de $\beta_{11} = 1$, pretendemos que os coeficientes β associados à moeda e ao produto venham dados por $\beta'_i = (a_i, -a_i, *, *)$, onde "*" representa valor a determinar. Começemos por utilizar a opção de defeito do CATS, ou seja a construção da matriz \mathbf{H}' . Assim no comando CATS escolhemos a opção 2 (Restrictions on subsets of beta), e indicamos para "Input the number of different groups", o valor 1 e para "Input the number of restrictions" ainda o valor 1. A matriz transposta \mathbf{H} deve ser de seguida construída

$$\mathbf{H}' = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

De imediato apenas teremos de indicar as variáveis para normalizar os vectores de cointegração assim obtidos, e que são a primeira e a segunda, M e Q respectivamente. O resultado obtido foi o seguinte

The LR test, CHISQ(2) = 10.35, p-value = 0.01

BETA (transposed)

M	Q	P	R
1.000	-1.000	-0.836	0.225
-1.000	1.000	1.026	-0.397

ALPHA

T-VALUES FOR ALPHA

DM	-0.425	0.035	-4.079	1.185
DQ	-0.027	-0.018	-0.443	-1.062
DP	-0.138	-0.055	-2.020	-2.851
DR	0.874	-0.095	3.412	-1.310

A restrição deve ser rejeitada. Em equilíbrio o coeficiente do produto não é o simétrico do da moeda. Para além dessa informação ainda surge um problema. No equilíbrio real, no segundo vector, passa a existir uma associação positiva entre

⁷Johansen (1995) apresenta um conjunto útil e imaginativo de restrições, nos coeficientes β e α , em dois modelos diferentes, pp. 73-8 e 114-120.

o produto e a taxa de juro. O que não acontecia atrás. Entretanto a inflação reduz-se quando a quantidade de moeda é superior à de equilíbrio, o que também é, no mínimo, estranho. Veja-se ainda que a variação das quantidades não depende dos desequilíbrios monetários ou reais.

Vejam agora a hipótese de o coeficiente dos preços ser o simétrico do da moeda. A matriz \mathbf{H}' vem agora dado por

$$\mathbf{H}' = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

E o resultado em termos do teste do ratio de verosimilhança vem dado por

The LR test, CHISQ(2) = 18.46 , p-value = 0.00

pelo que também rejeitamos esta restrição. Insistindo apesar de tudo na ideia e impondo assim os coeficientes do produto e dos preços com o mesmo valor e igual ao simétrico do coeficiente da moeda, teríamos que indicar o valor 2 para a pergunta "Input the number of restrictions" e viria para \mathbf{H}'

$$\mathbf{H}' = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Com o seguinte resultado para a restrição

The LR test, CHISQ(4) = 30.29 , p-value = 0.00

pelo que mais uma vez voltamos a rejeitar a restrição estudada.

Que conclusões retiramos destas imposições feitas ao modelo? Parece-me que em termos dos resultados estatísticos e da coerência económica devemos rejeitar a igualdade dos coeficientes do produto e dos preços, em conjunto e isoladamente, com o simétrico da moeda. O que significa que o equilíbrio monetário traduz um fenómeno de moeda como bem de luxo (coeficiente do produto superior à unidade) e ainda a presença de ilusão monetária (coeficiente dos preços inferior à unidade).

Uma palavra sobre a reparametrização indirecta. Na opção 4 do CATS deve-se alterar o método escolhido por defeito para o método $\mathbf{R}'\beta = 0$. Apenas temos de ter cuidado com a introdução dos valores para \mathbf{R}' . Os diferentes valores virão para aqueles três casos anteriores, construídos da seguinte forma

$$\begin{aligned} \mathbf{R}' &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

Uma vez que temos um modelo com variáveis nominais, à excepção do produto, podemos perguntar-nos se a variável produto não deveria juntar-se aos preços para obtermos uma variável produto nominal. O teste a efectuar em termos de reparametrização indirecta seria agora, para 1 grupo, 1 restrição

$$\mathbf{R}' = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}$$

O valor do teste vem dado por

The LR test, CHISQ(2) = 12.37 , p-value = 0.00

pelo que devemos rejeitar essa alteração e manter o modelo com aquelas variáveis nominais e a variável real.

Uma outra possibilidade a ter em consideração⁸ é a da exclusão de cada uma das variáveis naquele modelo. Respondemos assim à questão se não podemos rejeitar nenhuma variável das relações de cointegração - de uma relação de longo prazo. Pelo que devemos fazer para 1 grupo, 1 restrição

$$\mathbf{H}' = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R}' = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$$

para a reparametrização directa ou indirecta, respectivamente, da hipótese nula da primeira variável, M . E assim sucessivamente para as restantes variáveis. Os resultados foram os seguintes

The LR test, CHISQ(2) = 19.30 , p-value = 0.00

The LR test, CHISQ(2) = 19.25 , p-value = 0.00

The LR test, CHISQ(2) = 18.22 , p-value = 0.00

The LR test, CHISQ(2) = 13.78 , p-value = 0.00

para M , Q , P e R , respectivamente. Nenhuma das variáveis pode ser rejeitada.

Pelo que até aqui vimos, devemos manter este modelo e reter os coeficientes não restringidos que calculámos de início. Aceitando esta conclusão passemos a verificar a hipótese de alguma daquelas variáveis ser fracamente exógena. Não nos devemos esquecer que sendo a economia portuguesa uma pequena economia aberta, é possível que preços e taxa de juro, sobretudo, possam ter essa característica. Iremos fazer o teste dessa restrição a α sem impor qualquer outra restrição aos valores dos vectores de cointegração, β , que daí resultarem. Trata-se agora de definir a matriz β' , na restrição $\beta' \cdot \alpha = 0$. Faremos o ensaio variável a variável, ou seja, 1 restrição de cada vez. Para a primeira variável devemos fazer $\beta' = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}$, e assim sucessivamente. Os resultados foram⁹

The LR test, CHISQ(2) = 10.28 , p-value = 0.006

The LR test, CHISQ(2) = 9.34 , p-value = 0.009

The LR test, CHISQ(2) = 4.24 , p-value = 0.120

⁸Que já deveríamos ter levantado, mas que por motivos de exposição apenas agora referimos.

⁹Ver nota mais acima sobre o obtenção de valores mais precisos para os níveis de significância.

The LR test, $\text{CHISQ}(2) = 10.80$, $p\text{-value} = 0.005$

O que nos diz que não podemos rejeitar a hipótese (da equação) dos preços ser fracamente exógena. Estaríamos à espera de uma resultado destes mais para a taxa de juro. Talvez isto seja o resultado de na nossa pequena economia aberta os preços terem uma determinação externa, pela taxa de câmbio do escudo e da inflação externa, muito importante.

O novo modelo apresenta agora os seguintes parâmetros

The LR test, $\text{CHISQ}(2) = 4.24$, $p\text{-value} = 0.12$

BETA (transposed)

	M	Q	P	R
	1.000	-1.385	-0.789	0.194
	1.132	1.000	-1.268	0.501
	ALPHA		T-VALUES FOR ALPHA	
DM	-0.387	-0.066	-3.439	-2.409
DQ	0.136	-0.052	2.232	-3.480
DP	0.000	0.000	0.000	0.000
DR	1.070	0.098	3.877	1.458

Terminemos a apresentação do nosso modelo com o gráfico do valor da função de máxima verosimilhança calculado de forma regressiva a partir de 1990:1. O valor da função pode ser decomposto em duas partes (Hansen e Juselius (1995), p. 55-6) que correspondem aos quatro primeiros gráficos. Numa primeira leitura somos levados a pensar que os valores não são estáveis, mais ou menos constantes, mas tal deve-se apenas à escala utilizada nos gráficos. Com os valores totais da função, nos dois últimos gráficos, temos ainda os valores do intervalo a 95% ($\pm 2 \cdot \sqrt{2 \cdot p/N}$) e como podemos ver aquele valor está bem dentro deste intervalo.

Da mesma forma, os valores dos valores próprios apresentam uma grande estabilidade quando os calculamos a partir de 1990:1.

E desta forma damos praticamente por terminado o nosso trabalho de investigação de um modelo de longo prazo aplicado à economia portuguesa de 1977:1 a 2000:4, quando aceitamos a presença de dois vectores que traduzem relações de equilíbrio monetário e real.

O trabalho que aqui descrevemos procura seguir uma evolução lógica, do ponto de vista da análise económica. Mas podemos assumir uma atitude mais pragmática e usar a possibilidade dada pelo CATS de, pós conhecimento do tipo de modelo retido, obter informação sobre a possível exclusão da relação de longo prazo de variáveis do modelo, sobre a estacionaridade de cada série tomada isoladamente, e ainda sobre a exogeneidade fraca presente no modelo. A instrução é a seguinte

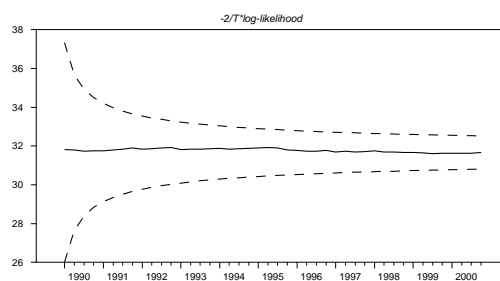
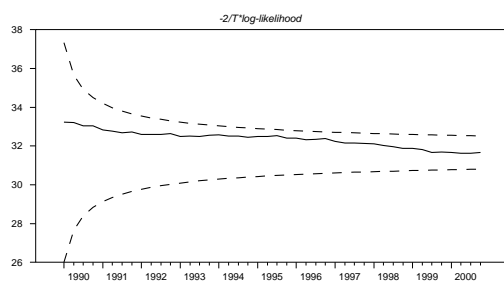
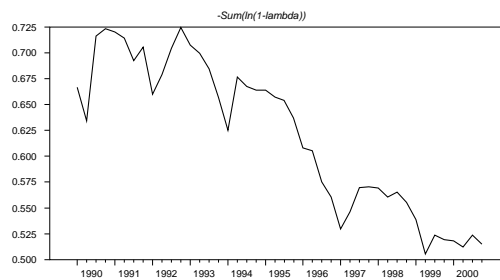
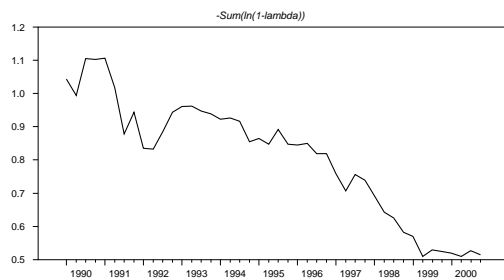
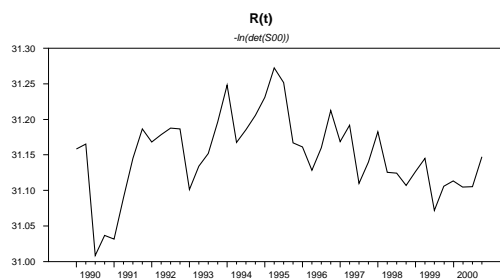
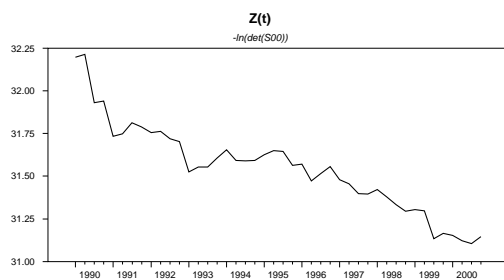
```
source(noecho) catsmain.src
```

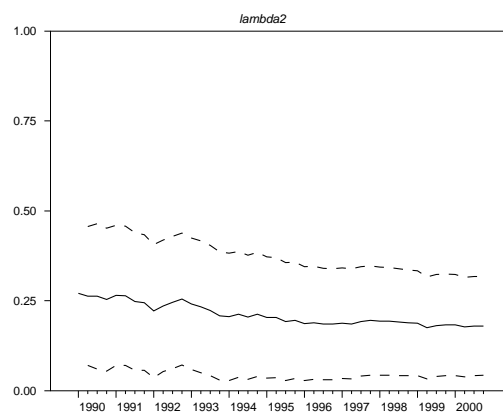
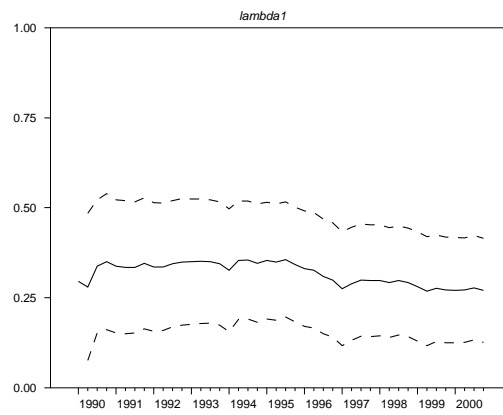
```
@cats(dettrend=drift,lags=5,season=4,proc=tsprop) 1977:1 2000:4
```

```
# m q p r
```

Sendo o resultado dado por

COINTEGRATION ANALYSIS





Endogeneous series : M Q P R
 Deterministic series : Unrestricted constant
 3 centered seasonal dummies
 Effective sample : 1978:02 TO 2000:04
 Lag(s) in VAR-model : 5
 No. of observations : 91
 Obs.- no.of variables: 67
 Batch tests of the time series properties of the individual series

TEST FOR EXCLUSION: LR TEST CHISQ(r)

r	DGF	CHISQ_5	M	Q	P	R
1	1	3.84	10.60	9.83	10.44	9.59
2	2	5.99	19.30	19.25	18.22	13.78
3	3	7.81	24.21	23.30	23.43	18.53

TEST FOR STATIONARITY: LR TEST CHISQ(p-r)

r	DGF	CHISQ_5	M	Q	P	R
1	3	7.81	17.34	12.87	20.26	22.90
2	2	5.99	6.69	2.21	9.65	13.28
3	1	3.84	2.10	0.35	3.68	3.94

TEST FOR WEAK-EXOGENEITY: LR TEST CHISQ(r)

r	DGF	CHISQ_5	M	Q	P	R
1	1	3.84	6.93	1.19	0.99	10.12
2	2	5.99	10.28	9.34	4.24	10.80
3	3	7.81	14.54	12.58	8.36	11.55

Os resultados, à exceção do caso da estacionaridade, são nossos conhecidos. Para 2 vectores de cointegração¹⁰ não temos nenhuma razão para excluir qualquer variável da representação de longo prazo. No que se refere à exogeneidade fraca já sabíamos que os preços assim poderiam ser considerados quando retivéssemos 2 vectores. Se retivéssemos 3 vectores nenhuma das variáveis poderia ser tomada como fracamente exógena. Finalmente no que respeita à estacionaridade, o produto pode ser tomado como estacionário no modelo com 2 vectores de cointegração. Os problemas seriam de impossível solução no caso de 3 vectores porque apenas a variável R seria nesse caso I(1) e como sabemos precisamos pelo menos de 2 variáveis I(1) para fazer o modelo da nossa economia usando a metodologia da cointegração.

Caso B: 1 só vector No caso de uma única relação de cointegração, e de acordo com os resultados imediatamente acima

- podemos rejeitar a ausência de uma qualquer daquelas variáveis;
- podemos rejeitar a estacionaridade de qualquer uma daquelas variáveis; e finalmente vemos que
- Q e P são fracamente exógenas.

¹⁰O mesmo podíamos dizer para 3 vectores.

Vejam os melhor este último resultado. Quando tínhamos dois vectores de cointegração, um dos vectores representando o equilíbrio monetário e o outro o equilíbrio real, os preços eram fracamente exógenos. Ao retirarmos do nosso modelo a equação do equilíbrio real passamos a ter o produto também fracamente exógeno. O resultado final não é pois estranho.

O resultado da escolha de $r = 1$ vem agora

```
EIGENVECTOR(S) (transposed)
  M      Q      P      R
53.8437 -71.7283 -42.5577 10.4638
The matrices based on 1 cointegration vectors
BETA (transposed)
  M      Q      P      R
1.000 -1.332 -0.790 0.194
  ALPHA      T-VALUES FOR ALPHA
DM -0.423 -3.594
DQ 0.108 1.614
DP -0.103 -1.297
DR 1.030 3.651
```

A equação do equilíbrio monetário pode ser escrita como $M = 1.332 \cdot Q + 0.790 \cdot P - 0.194 \cdot R$.

Uma vez que se trata de um novo sistema procurámos ver se os 5 desfasamentos se justificavam relativamente aos 6. Em termos dos critérios de informação (SC e HQ) o modelo escolhido seria o de 5 desfasamentos. O teste do ratio de verosimilhança, estimando o modelo para o período 1978:3 a 2000:4, tem o seguinte valor

Chi-Squared(16)= 18.593520 with Significance Level 0.29030404
 pelo que também este teste nos aponta para os 5 desfasamentos pela não exclusão da hipótese nula do sexto desfasamento em cada uma das equações do modelo.

A análise dos erros para o conjunto e para cada uma das equações, fazendo como atrás, produz os seguintes valores

```
MULTIVARIATE STATISTICS
LOG(DET(SIGMA)) = -31.46317
INFORMATION CRITERIA: SC = -27.15059
                        HQ = -28.58263

TRACE CORRELATION = 0.65574
TEST FOR AUTOCORRELATION
L-B(22), CHISQ(284) = 318.970, p-val = 0.08
LM(1), CHISQ(16) = 16.108, p-val = 0.45
LM(4), CHISQ(16) = 25.092, p-val = 0.07
TEST FOR NORMALITY
CHISQ(8) = 16.616, p-val = 0.03
UNIVARIATE STATISTICS
...
```

	ARCH(5)	NS	Normalidade	NS
Equação de M	12.787	0.025	2.304	0.316
Equação de Q	0.630	0.987	3.103	0.212
Equação de P	3.992	0.551	9.842	0.007
Equação de R	9.637	0.086	1.269	0.530

Parece-nos que podemos estar seguros acerca da ausência de auto-correlação dos erros, mas no que toca à presença de processo ARCH ou à ausência de normalidade dos erros os resultados já não são tão satisfatórios. Não podemos excluir um processo ARCH para a equação da oferta de moeda e também não podemos excluir uma distribuição não normal para os erros, para o conjunto, e para a equação dos preços em particular.

Se porventura usássemos 6 desfasamentos o problema da auto-correlação estaria resolvido; não teríamos processos ARCH; a normalidade conjunta estaria garantida (embora para um nível de 0,06), mas não a normalidade para a equação dos preços. O vector de cointegração teria um coeficiente para Q mais elevado e mais reduzido para P. O que do ponto de vista económico não desperta grande interesse. Por outro lado deveríamos abandonar este vector como de equilíbrio monetário, uma vez que o valor de α viria agora positivo para a equação da moeda. O que significava que deveríamos procurar outro significado para essa equação. Em suma, o ganho de passarmos para 6 desfasamentos não seria nem relevante, nem estatisticamente justificado.

Passemos a analisar algumas das restrições de coeficientes da relação de longo prazo.

Teste de simetria entre o coeficiente de M e de Q

The LR test, CHISQ(1) = 3.27 , p-value = 0.07

Teste de simetria entre o coeficiente de M e de P

The LR test, CHISQ(1) = 10.30 , p-value = 0.00

Teste de simetria entre o coeficiente de M e de P e Q

The LR test, CHISQ(2) = 17.70 , p-value = 0.00

é interessante verificar que podemos não rejeitar a hipótese de o coeficiente do produto ser igual à unidade na relação do equilíbrio monetário.

No que respeita à exogeneidade fraca já sabemos que o produto ($\alpha_2 = 0$) e os preços ($\alpha_3 = 0$) podem ser tomados como tal, isoladamente. O teste de restrição conjunta ($\alpha_2 = \alpha_3 = 0$) tem o valor

The LR test, CHISQ(2) = 3.22 , p-value = 0.20

O que confirma a exogeneidade daquelas variáveis, agora em conjunto. Para executarmos este teste devemos escolher de "Restrictions on alpha" e indicar duas restrições.

Mantendo a exogeneidade fraca para Q e P, ensaiámos a restrição da simetria dos coeficientes de M e Q

The LR test, CHISQ(3) = 5.64 , p-value = 0.13

BETA (transposed)

M	Q	P	R
---	---	---	---

	1.000	-1.000	-0.850	0.236
		ALPHA	T-VALUES FOR ALPHA	
DM	-0.447		-4.144	
DQ	0.000		0.000	
DP	0.000		0.000	
DR	1.058		3.992	

Como vemos a restrição conjunta, de $\alpha_2 = \alpha_3 = 0$ e $\beta_1 + \beta_2 = 0$, não pode ser excluída. E a nova equação do equilíbrio monetário é muito interessante: existe agora uma maior sensibilidade à taxa de juro e aos preços. Neste último caso podemos dizer que a ilusão monetária vem agora menor. O teste foi conduzido mantendo a anterior restrição nos α e impondo de seguida uma restrição nos β .

O maior valor do coeficiente de P levou-nos a impor também a restrição de simetria do coeficiente de M com P, mas o resultado é elucidativo

The LR test, $\text{CHISQ}(4) = 23.17$, p-value = 0.00

Tal hipótese deve ser rejeitada. Tratou-se agora de ainda manter a restrição em α e impor 1 grupo 2 restrições para β .

O modelo com exogeneidade fraca de Q e P e simetria do coeficiente de M e Q leva a resultados mais fracos no que respeita aos erros do modelo e das suas equações. Voltamos por isso ao modelo onde apenas impomos a exogeneidade de Q e P deixando livre os coeficientes da equação de equilíbrio. A análise dos erros vem neste caso dada por

MULTIVARIATE STATISTICS
 LOG(DET(SIGMA)) = -31.43637
 INFORMATION CRITERIA: SC = -27.12379
 HQ = -28.55583
 TRACE CORRELATION = 0.65711
 TEST FOR AUTOCORRELATION
 L-B(22), $\text{CHISQ}(284) = 313.381$, p-val = 0.111
 LM(1), $\text{CHISQ}(16) = 16.692$, p-val = 0.406
 LM(4), $\text{CHISQ}(16) = 26.032$, p-val = 0.054
 TEST FOR NORMALITY
 $\text{CHISQ}(8) = 16.982$, p-val = 0.030
 UNIVARIATE STATISTICS

...		ARCH(5)	NS	Normalidade	NS
Equação de M	13.915	0.016	1.723	0.422	
Equação de Q	0.290	0.998	3.326	0.190	
Equação de P	4.030	0.545	10.035	0.007	
Equação de R	10.909	0.053	1.491	0.474	

Como vemos, não temos problemas de auto-correlação dos erros, mas temos um problema de ARCH na primeira equação (M) e um problema de ausência de normalidade na terceira equação (P). Situação semelhante à encontrada no modelo sem restrição alguma.

Algumas opções úteis do CATS. Descrevamos, para finalizar este ponto, as possibilidades de representação incluídas no CATS, tal como figuram no procedimento "cats.src"

```

/*
CATS for RATS
March 1995
Henrik Hansen, Soren Johansen, Katarina Juselius
*****
SYNTAX :
@CATS(OPTIONS) START END
# ENDOGENOUS VARIABLES
# EXOGENOUS VARIABLES * OPTIONAL WITH EXO
# DUMMY SERIES * OPTIONAL WITH DUM
OPTIONS:
LAGS= INTEGER [2] * LAGS IN VAR-MODEL
DETTREND= NONE/CIMEAN/[DRIFT]/CIDRIFT * TREATMENT OF CONSTANT
SEASON= INTEGER [0] * CENTERED SEASONAL DUMMIES
EXO/[NOEXO] * INCLUSION OF EXOGENOUS I(1)
DUM/[NODUM] * CONDITIONING ON DUMMY SERIES
PROC= RANK/TSPROP/[I1] * PROCEDURES
TABLES/[NOTABLES] * SHOW THE TABLES OF CRITICAL V.
[MISC]/NOMISC * INCLUDE MISC. PROCEDURES
REC/[NOREC] * INCLUDE THE RECURSIVE PROCEDURE
GNAME= STRING [ ] * ADD A PREFIX TO PLOT FILES
BATCH/[NOBATCH] * SWITCH FOR RUN IN EDITOR OR BATCH
*/

```


Capítulo 4

Modelos VAR, VECM e Near-VAR(VECM)

Podemos dizer que o criador dos modelos VAR (vector auto regressive) foi o economista Christopher Sims (Sims (1980)). A primeira ideia associada a estes modelos, que procuravam responder a deficiências dos modelos estruturais de natureza keynesiana, era a não necessidade de estarem ancorados numa teoria económica. Podiam assim ser a-teóricos. No que respeitava às limitações dos modelos estruturais tínhamos agora um tipo de modelos que não necessitava de condições especiais de identificação e que resolvia o problema das antecipações de forma original, sem recurso a modelações duvidosas. As variáveis de um modelo VAR eram afinal variáveis endógenas, para os valores do período corrente, e exógenas para os valores desfasados. A presença dos valores desfasados tinha assim em conta o passado da economia e desta forma as antecipações não tinham de ter uma representação especial. à primeira ideia de "a-teoria" depressa se sucedeu a ideia que a teoria não podia estar afastada desses modelos. é fácil verificar que assim teria de ser: a escolha das variáveis reflecte sempre a orientação teórica de um autor, a menos que as escolha de olhos fechados e se limite a apontar ... E este último comportamento não é muito frequente ... No fundo, o que se pretendia dizer é que passávamos a ter uma classe de modelos aos quais não tínhamos de impor restrições de natureza keynesiana ou monetarista para fazer o estudo dessa representação da economia.

Os modelos VAR são uma representação da economia¹. E como representação da economia podem:

- fornecer-nos uma visão do comportamento passado da economia;
- ajudar-nos a conhecer a dinâmica de comportamento da economia;
- identificar relações de causalidade; e
- indicar-nos como comportamentos não esperados podem influenciar a economia e assim a fundamentar as acções de política.

¹Walter Enders escreveu um livro de apoio para o RATS onde os modelos deste tipo são especialmente considerados. Veja-se Enders (2003).

Logo no início do sucesso dos VAR os autores bayseanos desenvolveram o que ficou conhecido por BVAR (Thomas Doan e Sims (1984) e Litterman (1986)), modelos VAR bayseanos que procuravam responder ao problema da sobre-parametrização que afecta os modelos VAR. Os desenvolvimentos continuaram (Watson (1994), Amisano e Giannini (1997) e Stock e Watson (2001)) e hoje os VAR são usados em previsão e em simulação de políticas com grande frequência.

Vamos apresentar neste capítulo a família de modelos VAR. Levantaremos o problema da estabilidade dos modelos, começando com exemplos a uma só equação e apresentando de seguida o caso de várias equações; passaremos aos modelos VAR propriamente ditos; e depois veremos, ainda que de forma breve, as outras variantes da família. O essencial sobre os modelos aqui tratados será apresentado para os primeiros. Procuramos também apresentar um exemplo de um modelo VAR usando o programa RATS.

4.1 Estabilidade de modelos auto-regressivos

Sabemos já pelo que vimos atrás, que estamos interessados em aplicar os métodos normais da econometria a séries estacionárias. Dito de outra maneira, o princípio que desejamos para uma variável é que a influência dos choques sobre essa variável tendam a anular-se à medida que o tempo passa, tendam a desaparecer na história da série. O mesmo acontecendo para um modelo, desejamos que um choque sobre uma qualquer variável tenha efeitos, sobre ela própria e as restantes, que tendam a desaparecer.

4.1.1 Processo com dois desfasamentos

O que vamos dizer não é novo, sendo-o, no entanto, a forma como o faremos. Exemplifiquemos com um processo $AR(2)$ para uma variável y

$$y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \omega_t \quad (4.1)$$

que pode tomar uma forma mais conveniente aos nossos intuitos

$$(1 - \phi_1 \cdot L - \phi_2 \cdot L^2) \cdot y_t = \omega_t \quad (4.2)$$

Nesta última expressão,(4.2), o polinómio de desfasamentos, por factorização, pode ser re-escrito da seguinte forma

$$(1 - \phi_1 \cdot L - \phi_2 \cdot L^2) = (1 - \lambda_1 \cdot L) \cdot (1 - \lambda_2 \cdot L) \quad (4.3)$$

O que nos leva a apresentar o polinómio dos desfasamentos em termos das novas variáveis

$$(1 - \phi_1 \cdot L - \phi_2 \cdot L^2) = 1 - (\lambda_1 + \lambda_2) \cdot L + (\lambda_1 \cdot \lambda_2) \cdot L^2 \quad (4.4)$$

Esta igualdade, (4.4), para que seja verdadeira, conduz-nos à seguinte relação entre as variáveis dos membros esquerdo e direito

$$\begin{aligned}(\lambda_1 + \lambda_2) &= \phi_1 \\ (\lambda_1 \cdot \lambda_2) &= -\phi_2\end{aligned}\tag{4.5}$$

O que significa que se tivermos os valores $\phi_1 = 0,6$ e $\phi_2 = -0,05$, então devemos ter $\lambda_1 = 0,5$ e $\lambda_2 = 0,1$, ou seja

$$(1 - 0,6 \cdot L + 0,05 \cdot L^2) = (1 - 0,5 \cdot L) \cdot (1 - 0,1 \cdot L)$$

Em vez de estarmos a utilizar como variável, naqueles polinómios, o operador de desfasamentos L , passemos a utilizar a variável ∂ . A relação acima (4.3) virá, agora, dada por

$$(1 - \phi_1 \cdot \partial - \phi_2 \cdot \partial^2) = (1 - \lambda_1 \cdot \partial) \cdot (1 - \lambda_2 \cdot \partial)\tag{4.6}$$

Se dividirmos (4.6) por ∂^2 obtemos

$$(\partial^{-2} - \phi_1 \cdot \partial^{-1} - \phi_2) = (\partial^{-1} - \lambda_1) \cdot (\partial^{-1} - \lambda_2)\tag{4.7}$$

que será de grande utilidade se fizermos $\lambda = \partial^{-1}$, porque, assim, teremos

$$(\lambda^2 - \phi_1 \cdot \lambda - \phi_2) = (\lambda - \lambda_1) \cdot (\lambda - \lambda_2)\tag{4.8}$$

O membro direito anula-se para $\lambda = \lambda_1$ e $\lambda = \lambda_2$, o que leva a fazer para o membro esquerdo

$$\begin{aligned}\lambda_1 &= \frac{\phi_1 + \sqrt{\phi_1^2 + 4 \cdot \phi_2}}{2} \\ \lambda_2 &= \frac{\phi_1 - \sqrt{\phi_1^2 + 4 \cdot \phi_2}}{2}\end{aligned}\tag{4.9}$$

Voltando a tomar $\phi_1 = 0,6$ e $\phi_2 = -0,05$, obtemos os valores $\lambda_1 = 0,5$ e $\lambda_2 = 0,1$.

Podemos já concluir que o modelo apresentado, para ser estável, deverá apresentar as raízes daquela equação do segundo grau, em módulo, inferior à unidade, ou seja, $|\lambda_1| < 1 \wedge |\lambda_2| < 1$.

Como $\lambda = 1/\partial$, tomar o modelo (4.1) e dizer que y é estável quando as raízes de $\lambda^2 - \phi_1 \cdot \lambda - \phi_2 = 0$ estão dentro do círculo unitário é o mesmo que dizer que as raízes do polinómio $1 - \phi_1 \cdot \partial - \phi_2 \cdot \partial^2 = 0$ estão fora do círculo unitário, então, as duas afirmações são equivalentes do ponto de vista da estabilidade de y .

Uma forma prática de verificarmos se um processo é estacionário é vermos os efeitos de um qualquer choque sobre o comportamento da variável (ou variáveis).

4.1.2 Processo com p defasamentos

Apresentemos este mesmo resultado para um processo auto-regressivo de ordem p . O modelo virá agora representado por

$$y_t = \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \dots + \phi_p \cdot y_{t-p} + \omega_t \quad (4.10)$$

Façamos uma representação deste modelo em termos matriciais. Com

$$\zeta_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{t-p+1} \end{bmatrix} \mathbf{F} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} v_t = \begin{bmatrix} \omega_t \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (4.11)$$

virá

$$\zeta_t = \mathbf{F} \cdot \zeta_{t-1} + v_t \quad (4.12)$$

que de forma desenvolvida equivale a

$$\begin{aligned} y_t &= \phi_1 \cdot y_{t-1} + \phi_2 \cdot y_{t-2} + \dots + \phi_p \cdot y_{t-p} + \omega_t \\ y_{t-1} &= y_{t-1} + 0 \\ y_{t-p+1} &= y_{t-p+1} + 0 \end{aligned}$$

Como podemos ver, em (4.12) temos as mesmas relações que acima, em (4.10). Os valores próprios de \mathbf{F} são os valores de λ para os quais se verifica

$$|\mathbf{F} - \lambda \cdot \mathbf{I}_p| = 0$$

Vindo o polinómio característico da matriz \mathbf{F} dado por

$$\lambda^p - \phi_1 \cdot \lambda^{p-1} - \dots - \phi_p = 0$$

o que nos leva a impor como condição de estabilidade que as raízes sejam, em módulo, inferiores à unidade. O que coincide com a imposição de valores próprios de \mathbf{F} dentro do círculo unitário.

4.2 Apresentação de modelos VAR

Passemos agora à exposição de modelos com várias equações auto-regressivas para as diferentes variáveis presentes no modelo.

4.2.1 Exemplo de modelo

Designamos por modelo VAR estrutural um modelo do seguinte tipo

$$\begin{aligned} y_t &= b_{10} - b_{12} \cdot z_t + \gamma_{11} \cdot y_{t-1} + \gamma_{12} \cdot y_{t-2} + \varepsilon_{yt} \\ y_t &= b_{20} - b_{22} \cdot y_t + \gamma_{21} \cdot z_{t-1} + \gamma_{22} \cdot z_{t-2} + \varepsilon_{zt} \end{aligned} \quad (4.13)$$

Neste caso, temos apenas um VAR de ordem 1 com duas variáveis onde y e z são variáveis $I(0)$ e ε_{yt} e ε_{zt} são variáveis *white noise* e não correlacionadas. Uma vez que os valores correntes de cada uma das variáveis influenciam a outra, não estamos perante uma forma reduzida de um modelo, ou modelo reduzido, apesar da sua aparência. Veja-se Bernanke (1986), Blanchard e Watson (1986), Sims (1986) e Stock e Watson (2001) para uma apresentação deste tipo de modelos.

Façamos, antes, a seguinte apresentação (equivalente) do modelo acima

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \cdot \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \cdot \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \quad (4.14)$$

que, em termos de representação vectorial, se reduz a

$$\mathbf{B} \cdot \mathbf{x}_t = \mathbf{\Gamma}_0 + \mathbf{\Gamma}_1 \cdot \mathbf{x}_{t-1} + \varepsilon_t \quad (4.15)$$

Se pre-multiplicarmos por \mathbf{B}^{-1} obtemos

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1 \cdot \mathbf{x}_{t-1} + \mathbf{e}_t \quad (4.16)$$

onde

$$\begin{aligned} \mathbf{A}_0 &= \mathbf{B}^{-1} \cdot \mathbf{\Gamma}_0 \\ \mathbf{A}_1 &= \mathbf{B}^{-1} \cdot \mathbf{\Gamma}_1 \\ \mathbf{e}_t &= \mathbf{B}^{-1} \cdot \varepsilon_t \end{aligned}$$

O modelo VAR assim construído é um modelo VAR estandardizado. Este modelo toma a seguinte forma desenvolvida

$$\begin{aligned} y_t &= a_{10} - a_{11} \cdot y_{t-1} + a_{12} \cdot z_{t-1} + e_{1t} \\ z_t &= a_{20} - a_{21} \cdot y_{t-1} + a_{22} \cdot z_{t-1} + e_{2t} \end{aligned} \quad (4.17)$$

4.2.2 Relação entre os erros dos modelos

As relações entre os erros entre um e outro sistema levam-nos neste caso a

$$\begin{aligned} e_{1t} + b_{12} \cdot e_{2t} &= \varepsilon_{yt} & e_{1t} &= \frac{\varepsilon_{yt} - b_{12} \cdot \varepsilon_{zt}}{1 - b_{12} \cdot b_{21}} \\ b_{21} \cdot e_{1t} + e_{2t} &= \varepsilon_{zt} & e_{2t} &= \frac{\varepsilon_{zt} - b_{21} \cdot \varepsilon_{yt}}{1 - b_{12} \cdot b_{21}} \end{aligned} \quad (4.18)$$

Podemos facilmente perceber que os "novos" erros mantêm as características desejadas, não só para o seu valor esperado como para a sua variância. Para uma e outra e daquelas variáveis temos

$$E[e_{1t}] = \frac{E[\varepsilon_{yt}] - b_{12} \cdot E[\varepsilon_{zt}]}{1 - b_{12} \cdot b_{21}} = 0$$

$$E[e_{2t}^2] = \frac{\sigma_{zt}^2 - b_{21}^2 \cdot \sigma_{yt}^2}{(1 - b_{12} \cdot b_{21})^2} \longrightarrow \text{não depende do tempo}$$
(4.19)

No entanto, deparamos, agora, com uma característica entre aqueles erros que estava ausente nos erros anteriores

$$E[e_{1t} \cdot e_{2t}] = \frac{E[(\varepsilon_{yt} - b_{12} \cdot \varepsilon_{zt}) \cdot (\varepsilon_{zt} - b_{21} \cdot \varepsilon_{yt})]}{(1 - b_{12} \cdot b_{21})^2} = \frac{-(b_{21} \cdot \sigma_y^2 + b_{12} \cdot \sigma_z^2)}{(1 - b_{12} \cdot b_{21})^2} \neq 0$$
(4.20)

Este último resultado diz-nos que os choques em y e z passaram a estar correlacionados. A menos que os valores contemporâneos das variáveis não pertençam ao modelo, ou seja, $b_{12} = b_{21} = 0$.

4.2.3 Estabilidade do modelo

Tomemos agora um modelo com k variáveis, cuja ordem auto-regressiva seja de p . Utilizando o conceito de operador de defasamentos, na formulação (4.16) temos

$$\mathbf{x}_t = \mathbf{A}_0 + \mathbf{A}_1 \cdot \mathbf{x}_{t-1} + \mathbf{A}_2 \cdot \mathbf{x}_{t-2} + \dots + \mathbf{A}_p \cdot \mathbf{x}_{t-p} + \mathbf{e}_t$$

$$\mathbf{x}_t = \mathbf{A}_0 + (\mathbf{A}_1 \cdot L + \mathbf{A}_2 \cdot L^2 + \dots + \mathbf{A}_p \cdot L^p) \cdot \mathbf{x}_t + \mathbf{e}_t$$

$$(\mathbf{I} - \mathbf{A}_1 \cdot L - \mathbf{A}_2 \cdot L^2 - \dots - \mathbf{A}_p \cdot L^p) \cdot \mathbf{x}_t = \mathbf{A}_0 + \mathbf{e}_t$$
(4.21)

que pode tomar a forma resumida, em termos matriciais

$$\mathbf{A}(L) \cdot \mathbf{x}_t = \mathbf{A}_0 + \mathbf{e}_t$$
(4.22)

e cuja condição de estabilidade impõe que as raízes de $\mathbf{A}(L) = \mathbf{0}$ estejam fora do círculo unitário. O que é o mesmo que dizer que as raízes de

$$|\mathbf{I}_k \cdot \lambda^p - \mathbf{A}_1 \cdot \lambda^{p-1} - \mathbf{A}_2 \cdot \lambda^{p-2} - \dots - \mathbf{A}_p| = 0$$

devem cair dentro do círculo unitário.

Uma outra forma de apresentarmos a questão da estabilidade do modelo levamos a tomar o valor médio das variáveis

$$\mu = E[\mathbf{x}_t], \forall t$$

$$\mu = \mathbf{A}_0 + \mathbf{A}_1 \cdot \mu + \mathbf{A}_2 \cdot \mu + \dots + \mathbf{A}_p \cdot \mu$$
(4.23)

$$\mathbf{x}_t - \mu = \mathbf{A}_1 \cdot (\mathbf{x}_{t-1} - \mu) + \mathbf{A}_2 \cdot (\mathbf{x}_{t-2} - \mu) + \dots + \mathbf{A}_p \cdot (\mathbf{x}_{t-p} - \mu) + \mathbf{e}_t$$

e a expressarmos de forma matricial esta última diferença

$$\zeta_t = \mathbf{F} \cdot \zeta_{t-1} + v_t \quad (4.24)$$

onde

$$\zeta_{t(kp \times 1)} = \begin{bmatrix} \mathbf{x}_t - \mu \\ \mathbf{x}_{t-1} - \mu \\ \mathbf{x}_{t-2} - \mu \\ \dots \\ \mathbf{x}_{t-p+1} - \mu \end{bmatrix} \mathbf{F}_{(kp \times kp)} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_k & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I}_k & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbf{I}_k & 0 \end{bmatrix}$$

$$v_{t(kp \times 1)} = \begin{bmatrix} \mathbf{e}_t \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

De forma que mais uma vez chegamos à imposição de os valores próprios de \mathbf{F} deverem cair no círculo unitário para que o processo (4.24) seja estacionário (em covariância). Não esqueçamos que os valores próprios de \mathbf{F} satisfazem

$$|\mathbf{I}_k \cdot \lambda^p - \mathbf{A}_1 \cdot \lambda^{p-1} - \mathbf{A}_2 \cdot \lambda^{p-2} - \dots - \mathbf{A}_p| = 0$$

4.3 Identificação e estimação

Passaremos a analisar as questões que se colocam directamente à estimação de modelos VAR. O primeiro problema a colocar é precisamente o da identificação, ou, dito de outra forma, da sobre-parametrização dos modelos VAR.

4.3.1 A Sobre-parametrização dos modelos VAR

No modelo (4.21) temos k variáveis e p defasamentos. Cada matriz \mathbf{A} contém k^2 coeficientes, pelo que teremos de estimar $k + p \cdot k^2$ parâmetros num modelo deste tipo. Obviamente que se trata de um exagero. O modelo é pois sobre-parametrizado.

Este problema é relevante se o modelo (4.21) for usado para fazer previsões. Sendo o modelo utilizado para conhecer as relações dinâmicas entre as variáveis aí presentes, então aquele problema perde importância. Para estimar o modelo basta-nos a aplicação do método de OLS. E avance-se, desde já, que não adianta utilizar a metodologia SUR porque as variáveis da direita são as mesmas em todas as equações.

Como já dissemos a utilidade da metodologia VAR resulta de uma série de problemas que os modelos tradicionais colocavam à representação da economia. Voltemos a lembrar algumas das vantagens dos VARs:

- não exigem a divisão entre variáveis endógenas e exógenas,
- não temos de impor restrições nulas de forma abusiva, e
- podemos esquecer a teoria económica sobre a qual assenta o nosso modelo, apenas nos temos de preocupar com a escolha das variáveis.

Os modelos VAR levam-nos a dividir os seus adeptos em dois grupos. O primeiro grupo defende o que até aqui temos vindo a desenvolver e que leva a exigir que as variáveis do modelo sejam $I(0)$. Este grupo impõem ainda que apresentemos num modelo de variáveis em diferenças ($I(0)$), pelo menos, tantas variáveis deterministas como as que representam os ECMs desfasados entre as variáveis $I(1)$. Neste último caso, estamos a supor que as variáveis aí presentes, enquanto variáveis $I(1)$, são cointegradas.

O segundo grupo de economistas recusa a diferenciação e defende mesmo que se possam utilizar variáveis $I(1)$ num VAR, uma vez que se trata de obter a dinâmica de relacionamento entre essas variáveis. Estes autores defendem, também, que não se utilize uma tendência que afinal acaba por ser dada pelo próprio comportamento das variáveis $I(1)$. Para estes autores, a diferenciação elimina toda uma série de informações sobre o relacionamento das variáveis que deveriam ser retidas.

4.3.2 A escolha do número de desfasamentos

A escolha do número de desfasamentos a reter num modelo de k variáveis pode ser feita utilizando a ratio de verosimilhança. Tomemos um exemplo para $k = 2$ com 50 observações e onde pretendemos seleccionar entre 3 (p_o) e 4 (p_1) desfasamentos. Como, no máximo, temos 4 desfasamentos acabamos por ter apenas 46 observações úteis, $T = 50 - 4$, $T = N - p_1$. A fórmula do rácio virá

$$T \cdot \left\{ \log \left| \widehat{\Omega}_0 \right| - \log \left| \widehat{\Omega}_1 \right| \right\} \sim \chi^2_{(\text{restrições em } H_0)} \quad (4.25)$$

onde $\widehat{\Omega}_0$ é a matriz das variâncias-covariâncias do modelo p_o
onde $\widehat{\Omega}_1$ é a matriz das variâncias-covariâncias do modelo p_1

A nossa hipótese nula consiste em impor $k \cdot (p_1 - p_o)$ restrições em cada equação. Pelo que, para todo o sistema, teremos $k \cdot [k \cdot (p_1 - p_o)] = k^2 \cdot (p_1 - p_o)$ graus de liberdade.

No caso acima temos $2^2 \cdot (1) = 4$. Se, por exemplo

$$\widehat{\Omega}_0 = \begin{bmatrix} 2 & 1 \\ 1 & 2,5 \end{bmatrix} \widehat{\Omega}_1 = \begin{bmatrix} 1,8 & 0,9 \\ 0,9 & 2,2 \end{bmatrix}$$

$\log \left| \widehat{\Omega}_0 \right| = 1,386$ e $\log \left| \widehat{\Omega}_1 \right| = 1,147$, pelo que $46 \cdot (1,386 - 1,147) = 10,99$. Os graus de liberdade para o teste são 4. Ora $10,99 > \chi_4^2(= 9,49)^{5\%}$, pelo que a hipótese nula é rejeitada. O modelo a quatro desfasamentos é preferível sobre o modelo a três desfasamentos, uma vez que podemos rejeitar a hipótese nula dos coeficientes dos termos auto-regressivos de ordem quatro.

Sims sugeriu uma correcção para pequenas amostras em que o teste viria

$$(T - c) \cdot \left\{ \log \left| \widehat{\Omega}_0 \right| - \log \left| \widehat{\Omega}_1 \right| \right\} \sim \chi^2_{(\text{restrições em } H_0)} \quad (4.26)$$

onde c é o número de parâmetros estimados por equação, $1 + k \cdot p_1$. No caso acima levaria ao valor $(46 - 9) \cdot (1,386 - 1,147) = 8,84$ e, como vemos, a conclusão seria justamente a inversa: passaríamos a não rejeitar a hipótese nula daqueles coeficientes e assim a reter um modelo de ordem 3. As correcções estatísticas de amostras pequenas não são assim sem consequências ...

4.3.3 Apresentação alternativa de modelos VAR

Desenvolvamos uma forma alternativa de apresentar um modelo deste tipo. Tomemos o modelo (4.16) e façamos o seu desenvolvimento recursivo

$$\begin{aligned} \mathbf{x}_t &= \mathbf{A}_0 + \mathbf{A}_1 \cdot (\mathbf{A}_0 + \mathbf{A}_1 \cdot \mathbf{x}_{t-2} + \mathbf{e}_{t-1}) + \mathbf{e}_t \\ \mathbf{x}_t &= (\mathbf{I} + \mathbf{A}_1) \cdot \mathbf{A}_0 + \mathbf{A}_1^2 \cdot \mathbf{x}_{t-2} + \mathbf{A}_1 \cdot \mathbf{e}_{t-1} + \mathbf{e}_t \end{aligned}$$

Ao fim de n vezes acabamos por chegar a

$$\mathbf{x}_t = (\mathbf{I} + \mathbf{A}_1 + \mathbf{A}_1^2 + \dots + \mathbf{A}_1^n) \cdot \mathbf{A}_0 + \sum_{i=0}^n \mathbf{A}_1^i \cdot \mathbf{e}_{t-1} + \mathbf{A}_1^{n+1} \cdot \mathbf{x}_{t-n-1} \quad (4.27)$$

Esta última expressão, no caso de verificação das condições já expostas de estabilidade, leva a $\lim_{n \rightarrow \infty} \mathbf{A}_1^n = 0$ e, assim, obtemos

$$\mathbf{x}_t = (\mathbf{I} + \mathbf{A}_1 + \mathbf{A}_1^2 + \dots + \mathbf{A}_1^n) \cdot \mathbf{A}_0 + \sum_{i=0}^{\infty} \mathbf{A}_1^i \cdot \mathbf{e}_{t-1}$$

Que, atendendo à regra da inversão de matrizes por potências em série, pode ser escrita como

$$\mathbf{x}_t = (\mathbf{I} - \mathbf{A}_1)^{-1} \cdot \mathbf{A}_0 + \sum_{i=0}^{\infty} \mathbf{A}_1^i \cdot \mathbf{e}_{t-1} \quad (4.28)$$

Para termos uma ideia da primeira parcela presente em (4.28) calculámo-la para o caso de duas variáveis, como correspondendo a (4.17), e chegamos a

$$(\mathbf{I} - \mathbf{A}_1)^{-1} \cdot \mathbf{A}_0 = \begin{bmatrix} \frac{a_0 \cdot (1 - a_{22}) + a_{20} \cdot a_{12}}{(1 - a_{11}) \cdot (1 - a_{22}) - a_{12} \cdot a_{21}} \\ \frac{a_{10} \cdot a_{21} + a_0 \cdot (1 - a_{11})}{(1 - a_{11}) \cdot (1 - a_{22}) - a_{12} \cdot a_{21}} \end{bmatrix}$$

Insistindo na representação matricial podemos escrever

$$\mathbf{x}_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}_1^i \cdot \mathbf{e}_{t-1}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad (4.29)$$

Esta expressão será bastante útil para expressar o modelo equivalente ao modelo VAR que estamos a apresentar.

Para obtermos a matriz das variâncias-covariâncias deste modelo, lembremos que

$$E[\mathbf{e}_t^2] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}, \quad E[\mathbf{e}_t, \mathbf{e}_{t-1}] = 0$$

e, assim, virá

$$E[\mathbf{x}_t - \mu]^2 = E\left[\sum_{i=0}^{\infty} \mathbf{A}_1^i \cdot \mathbf{e}_{t-1}\right]^2 = (\mathbf{I} + \mathbf{A}_1^2 + \mathbf{A}_1^4 + \dots) \cdot \Sigma = (\mathbf{I} - \mathbf{A}_1^2)^{-1} \cdot \Sigma \quad (4.30)$$

onde Σ representa a matriz das variâncias-covariâncias dos erros.

4.3.4 Identificação e matriz de variâncias-covariâncias

Levantámos, acima, a questão da identificação a propósito da apresentação destes modelos. Voltemos, agora, a esse problema com o exemplo de duas variáveis e com uma solução que é bastante vulgar nestes modelos.

Tomemos os modelos (4.15) e (4.16). No primeiro caso, no modelo estrutural, temos oito coeficientes a serem determinados e os dois desvios-padrões dos erros. No segundo caso, no modelo estandardizado, temos 6 coeficientes, mais os dois desvios-padrão dos erros e a covariância entre estes. Ou seja, para o primeiro modelo temos 10 parâmetros e para o segundo apenas 9. O modelo é sub-identificado.

A solução pode ser a sugerida por *Sims*, através de um modelo recursivo em que $b_{21} = 0$. Com esta solução o modelo passou a ser identificado

$$\begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \cdot \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$$

Como

$$B^{-1} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix}$$

os erros estimados terão a seguinte relação com os choques a que estão sujeitos as variáveis do modelo

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -b_{12} \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \quad (4.31)$$

4.3.5 Avaliação dos efeitos de choques e decomposição de Choleski

Como podemos ver, os valores correntes da variável y não determinam os valores correntes de z , enquanto que os valores correntes de z determinam os valores de y . Ao nível dos choques nas variáveis, os choques correntes de y e

z afectam os valores correntes de y , enquanto que os valores correntes de z são afectados apenas pelos choques correntes de z . Os choques correntes de y apenas afectam os valores de z um período mais tarde. Aquela equação (4.31), com matriz triangular, corresponde à decomposição de *Choleski* e equivale a atribuir um certo comportamento ao modelo, como veremos de imediato. A equação (4.29) pode tomar a seguinte formulação

$$\mathbf{x}_t = \mu + \sum_{i=0}^{\infty} \mathbf{A}_1^i \cdot \mathbf{B}^{-1} \cdot \varepsilon_{t-1} \quad (4.32)$$

Se fizermos $\theta_i = \mathbf{A}_1^i \cdot \mathbf{B}^{-1}$ podemos passar a obter a representação equivalente do VAR em média móvel

$$\mathbf{x}_t = \mu + \sum_{i=0}^{\infty} \theta_i \cdot \varepsilon_{t-i} \quad (4.33)$$

Esta formulação é bastante útil porque, com relativa facilidade, obtemos os valores que resultam de choques aleatórios nas variáveis y e z . E obtemos esses valores para o que podemos chamar curto prazo e longo prazos. Chamamos normalmente a este tipo de análise a avaliação dos impulsos que resultam de choques não esperados sobre as variáveis do nosso modelo. Infelizmente o investigador não conhece, como dissemos atrás a propósito da sub-identificação do modelo. Mas, e agora felizmente, também dissemos que o problema poderia ser resolvido através da imposição de restrições, de que a associada à decomposição de *Choleski* é um exemplo. Assim, os valores correntes de y não determinariam os valores correntes de z , $b_{21} = 0$ o que, em termos do modelo a 2 variáveis e de ordem 1, conduz a

$$\begin{aligned} e_{1t} &= \varepsilon_{yt} - b_{12} \cdot \varepsilon_{zt} \\ e_{2t} &= \varepsilon_{zt} \end{aligned}$$

Os erros observados (e_{2t}) são atribuídos exclusivamente aos choques ε_{zt} . Uma vez conhecidos ε_{zt} podemos passar ao conhecimento de ε_{yt} . Mas se os valores correntes de y não afectam os valores correntes de z , os valores passados de y afectam os valores correntes de z e, assim, a sua influência acaba por se verificar, embora de forma indirecta.

Esta decomposição de *Choleski* assume assim uma hierarquia de efeitos e uma assimetria de choques

$$\varepsilon_{zt} \longrightarrow \begin{matrix} y_t \\ z_t \end{matrix} \text{ e } \varepsilon_{yt} \text{ não } \longrightarrow z_t$$

pelo que podemos dizer que z é uma variável anterior a y .

4.3.6 Um exemplo e uma regra prática

Consideremos o seguinte processo

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} 0,5 & 0,2 \\ 0,2 & 0,5 \end{bmatrix} \cdot \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

Por aqui vemos que o processo é estável e que as variáveis convergem para zero. Admitamos ainda que

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} 1,0 & 0,8 \\ 0 & 1,0 \end{bmatrix} \cdot \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{yt} + 0,8 \cdot \varepsilon_{zt} \\ \varepsilon_{zt} \end{bmatrix}$$

- um choque sobre z de uma unidade afecta y no montante de 0,8 e z em 1. No segundo momento, para conhecermos os efeitos sobre y e z , devemos olhar para o processo auto-regressivo acima. O mesmo para os períodos subsequentes até que os valores de y e z praticamente se anulam.

- se o choque for sobre y , de uma unidade, o efeito sobre y é obviamente de uma unidade e sobre z de zero. Apenas no momento seguinte esta variável se altera, de acordo com o processo auto-regressivo acima.

- em ambos os casos a nossa suposição foi que no período imediatamente a seguir ao choque os valores de e são nulos.

A principal questão que se coloca é, obviamente, saber se a ordem escolhida, entre aquelas variáveis é a adequada? Dois princípios podem ser seguidos

- teoricamente assim se justifica;

- o valor do coeficiente de correlação entre os erros estimados \hat{e}_{1t} e \hat{e}_{2t} é significativo, ($|\rho_{12}| > 0,2$), *rule of thumb*, e neste caso devemos estudar diferentes alternativas de ordenação.

4.4 Decomposição da variância e análise de causalidade

Passemos de imediato à análise das potencialidades de previsão de modelos VAR, assim como à leitura da informação neles contida sobre relações de causalidade entre as suas variáveis.

4.4.1 Capacidade de previsão dos modelos VAR

Infelizmente a sobre-parametrização destes modelos leva-os a fazerem más previsões. Mas, como também dissemos, estamos mais interessados no conhecimento das interdependências entre variáveis do modelo.

Utilizemos (4.16) e façamos a previsão do modelo para $t+1$, $t+2$ e $t+n$

$$\begin{aligned} E_t[\mathbf{x}_{t+1}] &= \mathbf{A}_0 + \mathbf{A}_1 \cdot \mathbf{x}_t \\ E_t[\mathbf{x}_{t+2}] &= \mathbf{A}_0 + \mathbf{A}_1 \cdot E_t[\mathbf{x}_{t+1}] \\ &= (\mathbf{I} + \mathbf{A}_1) \cdot \mathbf{A}_0 + \mathbf{A}_1^2 \cdot \mathbf{x}_t \\ E_t[\mathbf{x}_{t+n}] &= (\mathbf{I} + \mathbf{A}_1 + \dots + \mathbf{A}_1^{n-1}) \cdot \mathbf{A}_0 + \mathbf{A}_1^n \cdot \mathbf{x}_t \end{aligned}$$

De acordo com esta expressão, os erros de previsão para um período vêm dados por

$$\mathbf{x}_{t+1} - E_t[\mathbf{x}_{t+1}] = \mathbf{e}_{t+1}$$

e para dois períodos, uma vez que

$$\mathbf{x}_{t+2} = \mathbf{A}_0 + \mathbf{A}_1 \cdot (\mathbf{A}_0 + \mathbf{A}_1 \cdot \mathbf{x}_t + \mathbf{e}_{t+1}) + \mathbf{e}_{t+2}$$

virá

$$\mathbf{x}_{t+2} - E_t[\mathbf{x}_{t+2}] = \mathbf{e}_{t+2} + \mathbf{A}_1 \cdot \mathbf{e}_{t-1}$$

A partir desta última expressão passamos ao resultado para n períodos

$$\mathbf{e}_{t+n} + \mathbf{A}_1 \cdot \mathbf{e}_{t+n-1} + \mathbf{A}_1^2 \cdot \mathbf{e}_{t+n-2} + \dots + \mathbf{A}_1^{n-1} \cdot \mathbf{e}_{t+1} \quad (4.34)$$

Mas, uma vez que o processo (4.16) é equivalente a (4.33), para $t+n$ podemos fazer

$$\mathbf{x}_{t+n} = \mu + \sum_{i=0}^{\infty} \theta_i \cdot \varepsilon_{t+n-i}$$

e o erro de previsão virá dado por

$$\mathbf{x}_{t+n} - E_t[\mathbf{x}_{t+n}] = \sum_{i=0}^{n-1} \theta_i \cdot \varepsilon_{t+n-i} \quad (4.35)$$

Com esta última expressão, podemos calcular a variância do erro n períodos no futuro

$$\sigma_{(n)}^2 = \sum_{i=0}^{n-1} \theta_i^2 \cdot \sigma_{\varepsilon}^2 \quad (4.36)$$

onde, para o caso de duas variáveis

$$\sigma_{\varepsilon}^2 = \begin{bmatrix} \sigma_{\varepsilon_{yt}}^2 \\ \sigma_{\varepsilon_{zt}}^2 \end{bmatrix}$$

Esta última formulação é bastante ilustrativa sobre a capacidade de previsão dos modelos VAR: os erros aumentam com o afastamento do período de previsão, com n . Estamos perante modelos que devem ser usados com muito cuidado em previsões.

4.4.2 Decomposição da variância dos erros

Continuando a utilizar o resultado para duas variáveis, temos para essas variáveis

$$\begin{aligned} \sigma_{y(n)}^2 &= \sigma_{\varepsilon_{yt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{11}(i)^2 + \sigma_{\varepsilon_{zt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{12}(i)^2 \\ \sigma_{z(n)}^2 &= \sigma_{\varepsilon_{yt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{21}(i)^2 + \sigma_{\varepsilon_{zt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{22}(i)^2 \end{aligned} \quad (4.37)$$

Mas mais que os valores da variância dos erros, ao longo do período de previsão, estamos interessados na proporção nesses valores dos choques das diferentes variáveis. Para isso devemos fazer a seguinte desagregação

	efeitos dos choques de $y(\varepsilon_{yt})$	efeitos dos choques de $z(\varepsilon_{zt})$
para a Var dos erros de y	$\frac{\sigma_{\varepsilon_{yt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{11}(i)^2}{\sigma_{y(n)}^2}$	$\frac{\sigma_{\varepsilon_{zt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{12}(i)^2}{\sigma_{y(n)}^2}$
para a Var dos erros de z	$\frac{\sigma_{\varepsilon_{yt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{21}(i)^2}{\sigma_{z(n)}^2}$	$\frac{\sigma_{\varepsilon_{zt}}^2 \cdot \sum_{i=0}^{n-1} \theta_{22}(i)^2}{\sigma_{z(n)}^2}$

Se o contributo de ε_{zt} sobre y for desprezível, ($\rightarrow 0$), então podemos dizer que y é exógena face a z . Neste caso, a variável y evolui sem que seja afectada pelos choques de z . Podemos esquecer z no estudo de y .

Mas, mais uma vez, defrontamos um "pequeno" problema: o conhecimento de ε_{yt} e de ε_{zt} , através de e_{1t} e de e_{2t} . Neste caso, a decomposição de *Choleski* leva-nos a estudar a decomposição da variância dos erros, mas devemos excluir as primeiras observações porque é justamente nestas que existe maior incidência da hierarquia que impusemos às variáveis do modelo. Não esqueçamos que a importância da ordenação aumenta com os valores da correlação entre os desvios estimados, e_{1t} e e_{2t} .

4.4.3 A exogeneidade por blocos de variáveis

Para além da ideia que pode ser dada pela decomposição da variância dos erros, devemos, também, chamar a atenção que os testes de exogeneidade devem ser estatisticamente ensaiados. Para isso, podemos fazer uso da ratio de máxima verosimilhança onde, agora, passamos a ter

$$(T - c) \cdot \left[\log \left| \widehat{\Omega}_r \right| - \log \left| \widehat{\Omega}_u \right| \right] \chi^2_{(\text{restrições em } H_0)} \tag{4.38}$$

onde r e u se aplicam aos modelos restringidos e não restringidos. Por exemplo, r é o nosso modelo inicial e passamos a u juntando a cada uma das suas equações os desfasamentos da nova variável cuja inclusão queremos estudar. Os graus de liberdade da estatística correspondem ao número de desfasamentos vezes as equações do modelo inicial. Com este teste podemos, também, ensaiar a presença de sazonalidade no modelo e ainda outras variáveis deterministas.

Por vezes, também se utiliza o grau de informação do sistema

$$\begin{aligned} AIC &= T \cdot \log \left| \widehat{\Omega} \right| + 2 \cdot k_s \\ SBC &= T \cdot \log \left| \widehat{\Omega} \right| + k_s \cdot \log(T) \end{aligned} \tag{4.39}$$

onde k_s se refere ao número de parâmetros estimados em todo o sistema. O primeiro é um critério de *Akaike* e o segundo de *Schwartz*.

4.4.4 Identificação do modelo e testes às restrições impostas

A identificação do modelo estrutural é fundamental para compreendermos o significado da decomposição de Choleski. De (4.15) passámos a (4.16) com $\mathbf{e}_t = \mathbf{B}^{-1} \cdot \varepsilon_t$.

O uso de modelos estruturais deve levar-nos a utilizar a análise económica na definição apropriada dos erros. Isto é, a análise económica deve permitir que recuperemos as inovações estruturais dos resíduos ε_t .

Retomemos um VAR de ordem unitária com k variáveis.

$$\mathbf{x}_t = \mathbf{B}^{-1} \cdot \mathbf{\Gamma}_0 + \mathbf{B}^{-1} \cdot \mathbf{\Gamma}_1 \cdot \mathbf{x}_{t-1} + \mathbf{B}^{-1} \cdot \varepsilon_t$$

a matriz das variâncias-covariâncias deste modelo virá

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \dots & \dots & \dots & \dots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_k^2 \end{bmatrix}, \sigma_{1j} = \frac{\sum_{i=0}^{n-1} e_{it} \cdot e_{jt}}{T}$$

Ora, $\mathbf{\Sigma}$ é simétrica com $(k^2 + k)/2$ elementos, a nossa matriz \mathbf{B} que apresenta, na sua matriz diagonal principal o valor unitário, tem $k^2 - k$ elementos não conhecidos e não esqueçamos que não conhecemos as variâncias dos choques das nossas k variáveis [$var(\varepsilon_{it})$]. Em suma, queremos conhecer $k^2 - k + k = k^2$ parâmetros. A identificação destas k^2 incógnitas deverá ser feita com os $(k^2 + k)/2$ elementos de $\mathbf{\Sigma}$. Para isso temos de impor $k^2 - [(k^2 + k)/2] = (k^2 - k)/2$ restrições no sistema. Este último resultado aplica-se a qualquer ordem (p) de um VAR.

A solução tipo *Choleski* consistia em anular os elementos de \mathbf{B} abaixo da diagonal principal. Essas restrições de nulidade são justamente em número suficiente para a identificação do sistema.

Sims (1986) e Bernanke (1986) utilizaram outras técnicas. O primeiro, ao utilizar 6 variáveis, impôs 17 restrições nulas. Com $17 > 15$, o sistema passou a ser sobre-identificado. Em casos destes, o teste à nulidade da sobre-identificação consiste num teste do chi-quadrado do tipo

$$\chi_{(\text{restrições sobre-identificadas})}^2 = \left| \widehat{\mathbf{\Omega}}_r \right| - \left| \widehat{\mathbf{\Omega}} \right|$$

onde, no caso de $\chi^2 < \chi^{(c)}$, não rejeitamos as restrições que acabamos de impor.

Se, porventura, quisermos ensaiar dois tipos de restrições, em que o número de restrições $R_2 > R_1$ [$\geq k^2 - k)/2$], podemos testar R_2 contra R_1 utilizando

$$\chi_{(R_2 - R_1)}^2 = \left| \widehat{\mathbf{\Omega}}_{R_2} \right| - \left| \widehat{\mathbf{\Omega}}_{R_1} \right| \quad (4.40)$$

4.4.5 Decomposição histórica das séries

De acordo com a fórmula (4.33), que aqui repetimos, podemos fazer a seguinte decomposição

$$\mathbf{x}_{T+j} = \mu + \sum_{i=0}^{\infty} \theta_i \cdot \varepsilon_{T+j-i} = \sum_{i=0}^{j-1} \theta_i \cdot \varepsilon_{T+j-i} + \left[\mu + \sum_{i=j}^{\infty} \theta_i \cdot \varepsilon_{T+j-i} \right] \quad (4.41)$$

Na primeira parcela temos os valores das componentes de \mathbf{x}_{T+j} devidos às inovações de $T+1$ a $T+j$, e que são tantas quantas as variáveis do modelo VAR. Dentro do parêntese temos a previsão de \mathbf{x}_{t+j} sendo dada a informação disponível em T , $\mathbf{x}_{T+j}/\mathbf{I}_T$.

Este tipo de desagregação é útil para conhecermos a influência que as inovações das várias variáveis acabam por ter na determinação dos valores de cada uma das variáveis do modelo.

4.4.6 Programa para apresentação de alguns cálculos relacionados com um modelo VAR

Retomemos o exemplo de um modelo macroeconómico para a economia portuguesa. Na análise da cointegração tínhamos chegado a duas possibilidades quando a equações de equilíbrio. Na primeira tínhamos uma equação de equilíbrio monetário e outra de equilíbrio real. Na segunda, apenas uma equação de equilíbrio monetário.

Vamos apresentar um modelo VAR com aqueles dois tipos de equilíbrio (real e monetário).

Modelo VAR com duas relações de cointegração

A primeira coisa a reter num modelo do tipo VECM respeita à atenção que deve ser dada às relações de longo prazo. Nunca as devemos perder de vista. A segunda é que nesta parte do trabalho vamos apenas modelar comportamentos de curto prazo à volta daquelas relações de longo prazo.

Definição de variáveis ECM Devemos começar a construção do nosso modelo com a definição das variáveis ECM a que chegámos na análise de cointegração

```
set ecm1 = m - 1.332*q - .790*p + .194*r
set ecm2 = q - .288*p + .131*r + .192*m
```

Investigação sobre a ordem do VECM Como o nosso trabalho se segue ao da cointegração conhecemos a ordem do VAR. Mas se por acaso ainda tivéssemos de determinar a dimensão do modelo VAR, em termos dos desfasamentos, por exemplo entre 6 e 5 desfasamentos, o teste do ratio de verosimilhança a fazer, depois de definidas as variáveis, seria o seguinte

4.4. DECOMPOSIÇÃO DA VARIÂNCIA E ANÁLISE DE CAUSALIDADE 137

```
diff m / dm
diff q / dq
diff p / dp
diff r / dr
system 1 to 4
var dm dq dp dr
lags 1 to 6
det constant ecm1{1} ecm2{1} s{-2 to 0}
end(system)
estimate(noftest,noprint) 1978:4 *
system 1 to 4
var dm dq dp dr
lags 1 to 5
det constant ecm1{1} ecm2{1} s{-2 to 0}
end(system)
estimate(noftest,noprint) 1978:4 *
ratio(degrees=16,mcrr=30) 1978:4 *
# 1 2 3 4
# 5 6 7 8
```

onde na instrução "ratio" indicamos os graus de liberdade, 1 desfasamento por cada uma das 4 variáveis nas 4 equações, e a correção correspondente ao número máximo de parâmetros dos dois modelos (4 variáveis \times 6 desfasamentos + 6 variáveis deterministas = 30). Este exemplo serve como ilustração para todos os casos em que tenhamos necessidade de fazer um teste (LR) de restrição a alguns coeficientes de um modelo VAR. De referir a identificação do mesmo período para os dois modelos VAR a serem estimados.

Mas no caso em apreciação sabemos que devemos reter 5 desfasamentos para cada variável.

Estimação simples do modelo A indicação e estimação do modelo é feita da seguinte forma

```
system(model=ISLM)
vars dm dq dp dr
lags 1 to 5
det constant ecm1{1} ecm2{1} s{-2 to 0}
end(system)
estimate(noftest,noprint,outsigma=v) * *
```

De notar que não estamos interessados, neste caso, nem nos valores dos parâmetros (noprint) nem dos testes de exclusão das diferentes variáveis (noftest). No entanto queremos que seja retida pelo RATS o valor da matriz Ω para cálculos subsequentes.

Decomposição da variância de cada variável do modelo Retendo aquele modelo vamos decompor a variância associada aos choques das diferentes variáveis, para podermos ter uma ideia da importância dos choques verificados nela própria e nas restantes sobre a evolução de cada uma das variáveis. As instruções são as seguintes

```
list ieqn = 1 2 3 4
errors(impulses) 4 40 v
cards ieqn * * ieqn
```

O facto de não termos identificado as equações do modelo leva-nos a usar a sua ordem (1, 2, 3 e 4). O resultado resumido vem dado por

Decomposition of Variance for Series DM

Step	Std Error	DM	DQ	DP	DR
1	0.019648187	100.000	0.000	0.000	0.000
10	0.024238918	73.364	12.071	7.459	7.106
...
30	0.025154355	68.668	11.845	10.528	8.958

Decomposition of Variance for Series DQ

Step	Std Error	DM	DQ	DP	DR
1	0.010434567	1.383	98.617	0.000	0.000
10	0.015915196	11.845	56.817	25.531	5.807
...
30	0.016855986	12.240	53.833	28.216	5.711

Decomposition of Variance for Series DP

Step	Std Error	DM	DQ	DP	DR
1	0.013298920	0.449	0.657	98.894	0.000
10	0.017319575	2.347	9.716	76.208	11.728
...
30	0.018328353	3.430	11.196	73.967	11.407

Decomposition of Variance for Series DR

Step	Std Error	DM	DQ	DP	DR
1	0.049419557	8.697	4.444	11.758	75.100
10	0.060998041	13.987	11.383	18.461	56.169
...
30	0.062511503	13.513	11.411	19.770	55.305

Estes resultados traduzem o comportamento de curto prazo da economia como está representada no VECM que estamos a usar. As variações do produto e dos preços têm uma influência parecida na variação da oferta de moeda e ligeiramente superior à importância da taxa de juro. Praticamente 69% das suas variações se devem ao seu próprio comportamento. Quanto aos outros resultados haverá a realçar a importância das variações de preços na explicação da evolução da produção; o fraco papel da oferta de moeda na explicação da inflação; e o papel da inflação, mais importante que o da oferta de moeda, na explicação das variações da taxa de juro. Pelos resultados obtidos podemos confirmar a ideia que nenhuma daquelas variáveis deve ser tomada como exógena no modelo.

Simulação de choques exógenos nas variáveis do modelo Em seguida fizemos cada uma das variáveis sofrerem um choque e vemos como todas evoluíram em resultado desse choque. Começamos por um choque na oferta de moeda. Chamemos a atenção para o facto de as nossas variáveis de ajustamento estarem definidas em termos de primeiras diferenças, pelo que somámos os seus valores para termos efeitos acumulados. Os gráficos representam assim influências sobre as variáveis em níveis. As instruções são as seguintes

```
clear resp1; clear resp2; clear resp1_a; clear resp2_a
clear resp3; clear resp4; clear resp3_a; clear resp4_a
compute nstep = 30
impulse(noprint,input) 4 nstep
# 1 resp1
# 2 resp2
# 3 resp3
# 4 resp4
# 1.0 0.0 0.0 0.0
set resp1_a 1 nstep = resp1
set resp2_a 1 nstep = resp2
set resp3_a 1 nstep = resp3
set resp4_a 1 nstep = resp4
*
do for i = 2 to nstep
compute resp1_a(i) = resp1(i) + resp1_a(i-1)
compute resp2_a(i) = resp2(i) + resp2_a(i-1)
compute resp3_a(i) = resp3(i) + resp3_a(i-1)
compute resp4_a(i) = resp4(i) + resp4_a(i-1)
end do for i
label resp1_a resp2_a resp3_a resp4_a
# 'Resp Ac de Moeda' 'Resp Ac de Produto' 'Resp Ac de Preços' 'Resp Ac de
Juro'
spgraph(vfields=2,hfields=2,header='Choque unitário de Moeda')
do for i = resp1_a resp2_a resp3_a resp4_a
graph(header=%l(i),nodates) 1
# i
end do for
spgraph(done)
```

O resultado em termos gráficos vem dado por

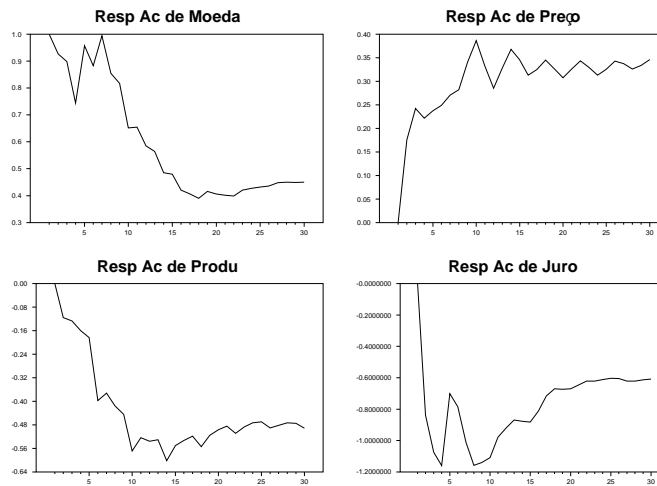
O choque obtido no produto deverá ser representado pela instrução

```
# 0.0 1.0 0.0 0.0
```

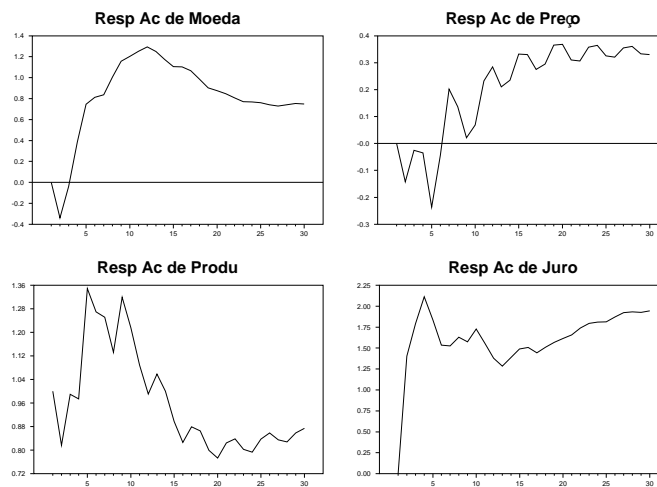
no conjunto de instruções sobre os impulsos. E assim sucessivamente. Os resultados viriam dados como se segue

O comportamento de curto prazo, de resposta a choques nas variáveis endógenas, está assim representado no nosso modelo. No primeiro gráfico vemos que um cho-

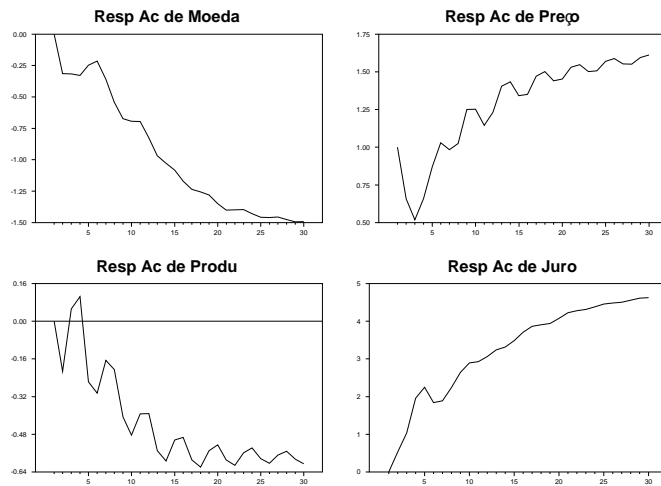
Choque unitário de Moeda



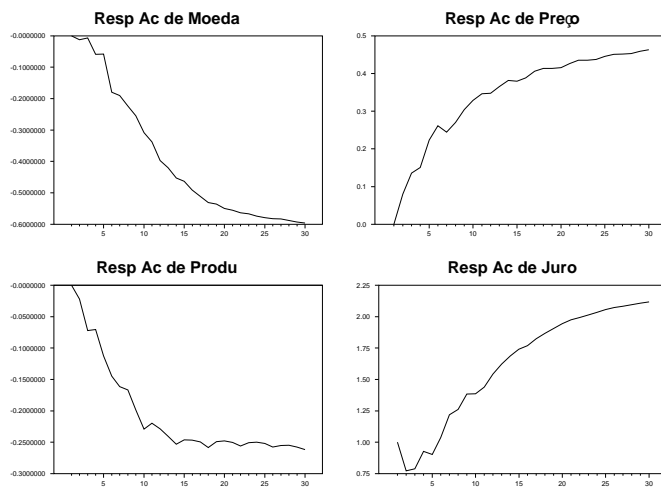
Choque unitário do Produto



Choque unitário de Preços



Choque unitário de Juro



que unitário da oferta de moeda (taxa de variação da oferta de moeda) leva a um efeito acumulado de 0,45, estabilizando a sua evolução à volta do 17º trimestre. Antes desse período, por volta do 18º estabilizou a influência negativa sobre produto, cujo valor total ronda os -0,49. Sobre os preços temos um efeito negativo que estabiliza por volta do 12º trimestre e que no total soma os 0,35, valor inferior ao acumulado pela oferta de moeda. O efeito negativo sobre a taxa de juro é nítido, estabilizando o efeito por volta do 18º trimestre e somando no final um valor próximo de -0,61.

Os valores que resultam dos choques podem ser lidos como variações percentuais uma vez que as variáveis estavam representadas em primeiras diferenças de logaritmos.

Um choque sobre o produto leva a que 30 trimestres depois o seu efeito ainda seja de 0,87, deixando a sua evolução de ter grandes flutuações a partir do 16º trimestre. Os efeitos sobre a oferta de moeda são também substanciais, ao fim de dois anos o efeito estabiliza nos 0,75. Os preços aumentam até aos 0,33, com algumas flutuações à volta desse valor a partir do 15º trimestre. A taxa de juro aumenta de forma muito sensível e muito rapidamente.

Um choque sobre a taxa de inflação de 100% arrasta os preços para uma subida de 166% e uma queda na produção de 61%. A moeda decresce, chegando a -1,60, por efeito do importante acréscimo sobre a taxa de juro que passa para 4,91. Mais do que os resultados prévios, estes resultados reflectem de forma clara a posição da economia portuguesa, uma economia muito pequena e aberta.

Vejam os finalmente o efeito de um choque sobre a taxa de juro. Este efeito arrasta os preços à subida (até 0,49), o que por sua vez actua sobre a própria taxa de juro (2,22). Estes efeitos em conjunto reduzem a quantidade de moeda (-0,64) e a quantidade produzida (-0,27). Os efeitos sobre esta última são relativamente rápidos, estando praticamente realizados ao fim de 14 trimestres.

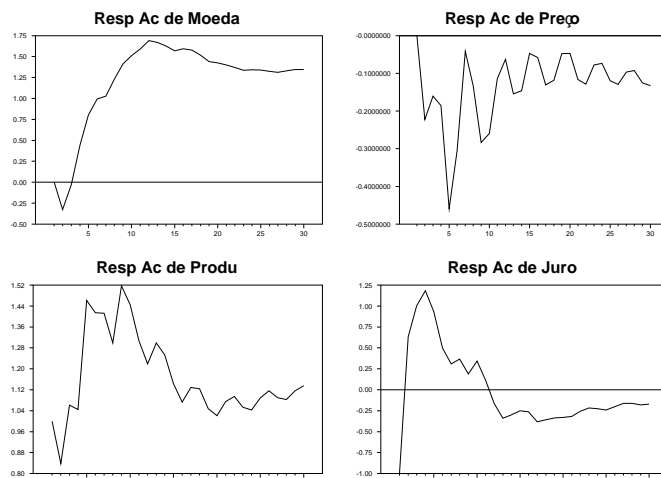
Tomemos a possibilidade de escolha de choques para impormos um choque positivo sobre o produto e um choque negativo sobre a taxa de juro. A alteração a fazer nas instruções seria agora

```
# 0.0 1.0 0.0 -1.0
```

Os efeitos constam do gráfico em baixo

Como vemos, durante três anos a oferta de moeda crescerá, estabilizando apenas no final do 24º trimestre. A influência sobre o produto será de realçar, sendo máxima no final do 9º trimestre e estabilizando em 1,12, valor superior ao seu próprio choque. Apesar do crescimento do produto o efeito sobre a taxa de juro é ainda negativo no final (-0,20), apesar dos valores positivos do 3º ao 12º trimestres. O efeito final sobre os preços é negativo, -0,13. Estes têm uma evolução inicial em tudo semelhante a um choque. Uma economia que por algum motivo possa sofrer do exterior um choque positivo da produção e negativo sobre o valor da taxa de juro, pode ainda praticar uma política no curto prazo de expansão da oferta. Se perante aqueles choques ainda aumentasse a oferta de moeda de 0,32, o valor final do produto (dos três choques) viria de 0,96, para ausência de efeitos

Choque unitário do Produto e de Juro



sobre os preços e redução da taxa de juro de 0,39.

Duas observações adicionais sobre o que fizemos até aqui. Estes choques são analisados tendo em conta as relações de curto prazo e por isso não se tratam de valores obtidos no longo prazo, mas antes de valores na suposição de equilíbrio de longo prazo. Se é instrutivo supormos choques unitários, estes choques podem não ter relação alguma com a natureza aleatória do modelo e da variável em estudo. Por isso somos levados muitas vezes a tomar os valores dos choques iguais ao valor do desvio padrão do erro da equação respectiva. Exemplifiquemos com o choque combinado no produto e na taxa de juro. A instrução a dar seria agora

$\# 0 \ v(3,3) \ 0 \ -v(4,4)$

e os resultados finais para a moeda, produto, preços e juro seriam dados por

0,002; 0,001; -0,001; e -0,005

como vemos, os valores são muito reduzidos, da ordem das permilagens.

4.5 Modelos VECM, Near-VAR e Near-VECM

O que distingue um modelo VAR de um modelo VECM é que o segundo apresenta como variáveis, deterministas os ECMs correspondentes às relações de cointegração entre as diferentes variáveis. Eventualmente um único ECM correspondente à única relação de cointegração entre as diferentes variáveis.

Um modelo Near-VAR é um modelo que abandona a característica de idêntico número de defasamentos para todas as variáveis do modelo. Procuramos neste caso reduzir o número de parâmetros através da investigação de quais as melhores equações para a construção do modelo. No caso destes modelos devemos aplicar

métodos de estimação como SUR e FIML.

Finalmente um Near-VECM baseia-se na filosofia dos modelos Near-VAR com a inclusão de vector, ou vectores, ECM (s).

Capítulo 5

Modelos ARCH

Os modelos ARCH surgiram a propósito do estudo da inflação na corrente de estudos das antecipações racionais. Os modelos da inflação acabaram por revelar um comportamento interessante no que respeitava aos seus erros. Como Engle (1982) provou, aqueles modelos apresentavam uma estrutura auto-regressiva da variância dos erros. Não só a variância, afinal, não era constante, como tinha aquele tipo de comportamento temporal. A variância dos erros de um modelo por ser encarada como a incerteza associada aos valores médios da previsão. Os modelos que incorporam este comportamento dos erros acabaram por ter uma importância crescente no estudo de fenómenos de natureza financeira devido à relação normal entre rendimentos de activos e incerteza (variância dos seus rendimentos).

5.1 Apresentação geral da questão do ARCH

Uma apresentação não complexa é feita no Capítulo 21 de Hamilton (1994). Devemos salientar também a colecção de *readings* em Engle (1995) e a apresentação feita em Engle (2001).

Quando temos um modelo auto-regressivo de ordem 1,

$$y_t = \alpha_0 + \alpha_1 \cdot y_{t-1} + \varepsilon_t \quad (5.1)$$

assumimos o seguinte comportamento para a variância de y

$$Var [y_t/y_{t-1}] = E_{t-1} [(y_t - \alpha_0 - \alpha_1 \cdot y_{t-1})^2] = E_{t-1} [\varepsilon_t^2] = \sigma^2 \quad (5.2)$$

Engle chamou a atenção para o facto de em variáveis monetárias e financeiras, a hipótese expressa em (5.2) não ser realista. A volatilidade de uma série, dada pela sua variância, apresenta muitas vezes um comportamento auto-regressivo, ou um comportamento de outro tipo, que não nos permite tomar a hipótese da variância constante como realista ou, até mesmo, desejável.

5.1.1 Variância condicional AR

Tomemos assim uma expressão que traduza um comportamento auto-regressivo da variância

$$\widehat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \cdot \widehat{\varepsilon}_{t-1}^2 + \dots + \alpha_q \cdot \widehat{\varepsilon}_{t-q}^2 + v_t \quad (5.3)$$

$$v_t \sim i.i.d. (0, \sigma^2) \quad (5.4)$$

ora, apenas no caso de termos

$$\alpha_1 = \alpha_2 = \dots = \alpha_q = 0 \quad (5.5)$$

encontramos o resultado do modelo clássico

$$E_t [\sigma_\varepsilon^2] = \alpha_0 \quad (5.6)$$

A determinação de um modelo pela forma clássica é feita na suposição que (5.5) se verifica. Ou seja, dito de outra forma, a presunção de (5.6), que nos leva a estudar a exclusão de heteroscedasticidade e de auto-correlação, deve também levar ao estudo da exclusão de (5.3), que é designada sugestivamente por heteroscedasticidade condicionada auto-regressiva (ARCH).

Nesta hipótese, o valor esperado da variância vem dado por:

$$E_t [\widehat{\varepsilon}_{t+1}^2] = \alpha_0 + \alpha_1 \cdot \widehat{\varepsilon}_t^2 + \dots + \alpha_q \cdot \widehat{\varepsilon}_{t-q+1}^2 + v_t \quad (5.7)$$

O problema da estimação de um modelo para y consiste no facto de necessitarmos dos valores do quadrado dos erros. Num primeiro momento podemos usar os valores estimados dos erros e através de um teste LM¹ excluir, ou não, um processo como (5.7). Mas não podemos esquecer que o uso de dois passos, como é vulgar na resolução dos problemas de heteroscedasticidade, não é eficiente. Por esse motivo é preferível que o modelo de y seja estimado pelo método de máxima verosimilhança². Sendo assim, não devemos tomar v_t , mas antes $\log(v_t)$. Engle (1982) propôs que se tomasse

$$\varepsilon_t = v_t \cdot \sqrt{\alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2} \quad (5.8)$$

onde $\sigma_v^2 = 1$.

Teremos assim para a variância

$$\sigma_\varepsilon^2 = \varepsilon_t^2 = \frac{\alpha_0}{1 - \alpha_1} \quad (5.9)$$

¹O teste LM de adição de p variáveis apresenta para $T.R^2$ uma distribuição do Chi-quadrado com p desfasamentos.

²á qual voltaremos mais à frente.

Devido a (5.9) e à exigência de estabilidade de (5.8) devemos impor que

$$\alpha_0 > 0 \wedge 0 < \alpha_1 < 1 \quad (5.9)$$

Entretanto, temos para a média condicionada dos erros do modelo (5.1)

$$E[\varepsilon_t / \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] = E[v_t] \cdot E[(\alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2)^{1/2}] = 0 \quad (5.10)$$

uma vez que, por hipótese, $E[v_t] = 0$ e $v_t \wedge \varepsilon_{t-1}$ são independentes.

Um modelo ARCH(1) é afinal um modelo cuja variância condicional dos erros vem dada por

$$E[\varepsilon_t^2 / \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots] = \alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2 \quad (5.11)$$

e que apresenta as seguintes características

- a média condicional é nula,
- a variância condicional segue um processo AR(1).

De notar, que no caso de processos ARCH, ε_t não é um processo autocorrelacionado, mas os seus valores não são independentes porque os seus segundos momentos estão relacionados.

Um processo ARCH geral, que designaremos por ARGH(q), corresponderá ao processo seguinte

$$\varepsilon_t = v_t \cdot \sqrt{\alpha_0 + \sum_{i=1}^q \alpha_i \cdot \varepsilon_{t-i}^2} \quad (5.12)$$

5.1.2 Variância condicional ARMA

Bollerslev (1986) avançou com uma outra hipótese de comportamento da variância, abandonando a limitação do processo auto-regressivo.

Tomemos a seguinte representação para os erros

$$\varepsilon_t = v_t \cdot \sqrt{h_t} \quad (5.13)$$

ainda com $\sigma_v^2 = 1$, mas agora com a seguinte representação para h

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \cdot \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \cdot h_{t-j} \quad (5.14)$$

Como vemos, em (5.14), temos um processo ARMA, que será designado por GARCH(q,p). Nesta formulação [com (5.13) e (5.14)], temos

$$\begin{aligned} E[\varepsilon_t/\varepsilon_{t-1}, \dots] &= 0 \\ E_{t-1}[\varepsilon_t^2] &= h_t \end{aligned} \quad (5.10)$$

A expressão de um GARCH(q,p) acaba por englobar a anterior, uma vez que, por exemplo, um GARCH(1,0) é idêntico a um ARCH(1), ou ARCH(AR=1). Obviamente, que devemos ter cuidado com a estabilidade implícita em (5.14). Assim, as raízes características de (5.14) devem implicar convergência.

Nalgumas situações podemos estar interessados numa variância condicional que apresente raiz unitária sendo assim divergente para infinito. Bastará para tal impor que os coeficientes α e β somem a unidade.

De (5.14) deduzimos o valor esperado da variância

$$E_t[\varepsilon_t^2] = \alpha_0 + \sum_{i=1}^q \alpha_i \cdot \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \cdot h_{t-j}^2 \quad (5.11)$$

de onde naturalmente retiramos as seguintes questões

- como pesquisar a aplicação às variâncias de um modelo AR ou ARMA?
- no primeiro caso, como determinar AR ? e
- no segundo como obter os graus de AR e MA ?

Uma hipótese consiste em estudar as variâncias estimadas e fazer o estudo da auto-correlação simples e parcial entre os valores obtidos.

Quando a variância condicional afecta a média

Robert Engle e Robins (1987) levantaram a hipótese de nas séries financeiras³ o facto de o risco ser elevado, ou seja, as variâncias serem elevadas, conduzir à alteração da média dos valores da série. Assim, o aumento do risco associado a um título levará ao aumento do rendimento do mesmo. A hipótese é perfeitamente lógica. Este comportamento da variância condicionada é designado por ARCH-M.

A equação do excesso de rendimento de um título vem dada por

$$R_{it} = \bar{R} + y_t$$

onde y é o excesso de rendimento do activo i sobre uma taxa de rendimento isenta de risco, \bar{R} . Em circunstâncias normais esperamos ter para o valor do excesso de rendimentos

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ E_{t-1}[y_t] &= \mu_t \end{aligned} \quad (5.12)$$

³Era nelas que os autores pensavam.

5.2. APRESENTAÇÃO DO MÉTODO DE MÁXIMA VEROSIMILHANÇA 149

onde μ_t representa o prémio de risco.

A ideia de Engle traduziu-se em admitir que podíamos ter

$$\mu_t = \beta + \delta \cdot h_t \quad (\delta > 0) \quad (5.13)$$

com h a poder ser representado por

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \cdot \varepsilon_{t-i}^2 \quad (5.14)$$

Naturalmente que podemos ter outras representações mais gerais para h .

Mais uma vez, verificando-se (5.5) temos, $\mu_t = \beta + \delta \cdot \alpha_0$, e assim retornamos ao caso clássico de estimação.

Exemplo de um modelo ARCH-M

Para dados trimestrais obteve-se: $y_t = 0,142 + \varepsilon_t$

Vamos estudar $h_t = \alpha_0 + \alpha_1 \cdot (0,4 \cdot \varepsilon_{t-1}^2 + 0,3 \cdot \varepsilon_{t-2}^2 + 0,2 \cdot \varepsilon_{t-3}^2 + 0,1 \cdot \varepsilon_{t-4}^2)$

Onde os pesos das variâncias são decrescentes.

A sua estimação leva à exclusão da hipótese nula de acordo com o teste LM $T \cdot R^2 \sim \chi_1^2$ e chegamos a $\alpha_0=0,0023$ e $\alpha_1=1,64$.

Finalmente obtemos por máxima verosimilhança:

$$y_t = -0,0241 + 0,687 \cdot h_t$$

(1,29) (5,15)

$$h_t = 0,0023 + 1,64 \cdot (0,4 \cdot \varepsilon_{t-1}^2 + 0,3 \cdot \varepsilon_{t-2}^2 + 0,2 \cdot \varepsilon_{t-3}^2 + 0,1 \cdot \varepsilon_{t-4}^2)$$

(1,08) (6,30)

5.2 Apresentação do método de máxima verosimilhança

Tomemos o seguinte modelo geral

$$\varepsilon_t = y_t - \beta \cdot x_t \quad (5.15)$$

Fazendo uso da hipótese habitual de distribuição Normal, formamos a equação de log likelihood da observação t , como

$$-\frac{1}{2} \cdot \log(2 \cdot \pi) - \frac{1}{2} \cdot \log(\sigma^2) - \frac{1}{2 \cdot \sigma^2} \cdot (y_t - \beta \cdot x_t)^2 \quad (5.16)$$

que nos leva para T observações a

$$\log(L) = -\frac{T}{2} \cdot \log(2 \cdot \pi) - \frac{T}{2} \cdot \log(\sigma^2) - \frac{1}{2 \cdot \sigma^2} \cdot \sum_{t=1}^T (y_t - \beta \cdot x_t)^2 \quad (5.17)$$

O que pretendemos com este método é minimizar este valor, (5.17), para obtermos os parâmetros desejados. As derivadas de primeira ordem vêm dadas por

$$\begin{aligned}\frac{\partial \log(L)}{\partial \sigma^2} &= -\frac{T}{2 \cdot \sigma^2} + \frac{1}{2 \cdot \sigma^4} \cdot \sum_{t=1}^T (y_t - \beta \cdot x_t)^2 \\ \frac{\partial \log(L)}{\partial \beta} &= \frac{1}{\sigma^2} \cdot \sum_{t=1}^T (y_t \cdot x_t - \beta \cdot x_t^2)\end{aligned}\quad (5.18)$$

Para termos as condições de primeira ordem basta igualar a zero aquelas equações (2.4). O que nos conduz ao valor dos parâmetros a estimar

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{T} \cdot \sum_{t=1}^T \varepsilon_t^2 \\ \hat{\beta} &= \frac{\sum_{t=1}^T x_t \cdot y_t}{\sum_{t=1}^T x_t^2}\end{aligned}\quad (5.19)$$

O nosso problema é que as condições acima são lineares e quando estamos perante um ARCH as condições de primeira ordem não são lineares.

Não temos agora uma relação como (5.15), mas antes como (5.12). O que significa que (5.16) virá agora

$$-\frac{1}{2} \cdot \log(2 \cdot \pi) - \frac{1}{2} \cdot \log(h_t) - \frac{1}{2 \cdot h_t} \cdot (y_t - \beta \cdot x_t)^2 \quad (5.20)$$

e assim, para a totalidade das observações passamos a ter para a função do logaritmo de máxima verosimilhança

$$\log(L) = -\frac{T-1}{2} \cdot \log(2 \cdot \pi) - \frac{1}{2} \cdot \sum_{t=2}^T \log(h_t) - \frac{1}{2} \cdot \sum_{t=2}^T \frac{(y_t - \beta \cdot x_t)^2}{h_t} \quad (5.21)$$

onde definimos h , a variância, como

$$h_t = \alpha_0 + \alpha_1 \cdot (y_{t-1} - \beta \cdot x_{t-1})^2 \quad (5.22)$$

5.2.1 A utilização do RATS

Pelo facto de o RATS executar instruções que organizamos de forma apropriada torna-se bastante adequado à estimação deste tipo de modelos. Começemos por ver o significado da instrução FRML e MAXIMIZE antes de nos preocuparmos com a estimação do próprio modelo.

As instruções FRML e MAXIMIZE

A instrução FRML destina-se a definir uma equação. Por isso é usada, por vezes mais de uma vez, até chegarmos a uma equação cujo valor queremos maximizar. A instrução MAXIMIZE aplica-se a uma equação FRML previamente definida e identificada pelo nome a ela atribuído. Ao mesmo tempo deve ser indicado o período para o qual queremos a estimação. Aqui, o cuidado a ter relaciona-se com o início, sobretudo quando temos valores desfasados nas instruções FRML anteriormente definidas.

O comando de maximização é assim composto, no essencial, por
MAXIMIZE(opções) FRML start end

As opções referem-se ao método que usamos, sendo por defeito BFGS. Se não houver convergência na maximização devemos alterá-lo para *method=BHHH*. O RATS também admite o uso do *SIMPLEX*. No caso, frequente, em que a estimação envolve valores desfasados, devemos incluir a opção *RECURSIVE*. As instruções subsequentes *TEST* e *RESTRICT*, bastante úteis, podem ser utilizadas com *MAXIMIZE*. Podemos usar o método *SIMPLEX* para uma primeira aproximação aos valores dos parâmetros a estimar.

Tomemos o modelo já atrás apresentado (2.1) e (2.3). As instruções devem ser

- (i) *NONLIN b var*
- (ii) *FRML L=-log(var)-((y-b*x)**2)/var*
- (iii) *COMPUTE b=?, var=?*
- (iv) *MAXIMIZE(method=BHHH,recursive) L * **

Os pontos de interrogação referem-se a valores a serem por nós atribuídos. Vejamos alguns casos interessantes de alteração daquelas instruções para uma utilização mais flexível do programa. Eliminámos em L as constantes porque não afectam o resultado.

- a) A instrução da linha (ii) pode ser substituída por estas duas
 - (ii a) *FRML e=y-b*x*
 - (ii b) *FRML L=-log(var)-(e(t)**2)/var*
- b) Ainda em (ii) podemos fazer o seguinte, para o caso de um ARCH(1)
 - (ii a) *FRML e=y-b*x*
 - (ii b) *FRML v=a_0+a_1*e{1}**2*
 - (ii c) *FRML L=-.5*(log(v)+(e(t)**2)/v*
- c) Mas as três definições acima podem ser reduzidas a
 - (ii a) *FRML e=y-b*x*
 - (ii b) *FRML L=(v=a_0+a_1*e{1}**2),-.5*(log(v)+(e**2)/v)*

De notar a extrema condensação permitida na última definição, em que usamos duas equações em simultâneo. Podemos usar em vez de duas as que forem necessárias para condensar as instruções. Antes da definição, propriamente dita, temos condições de igualdade a verificarem-se.

d) O cuidado principal, quando temos desfasamentos, consiste em uma variável não poder ser definida à sua própria custa. Por exemplo $e=y-b*e\{1\}$, constitui uma formulação incorrecta. A solução reside na prévia definição de uma variável cujo valor seja usado para o cálculo da primeira observação de e .

*SET temp=0.0 ; * definição que será abandonada*

NONLIN b var

*FRML e=y-b*temp{1}; depois do primeiro cálculo será abandonada*

FRML L=(temp=e),.5(log(var)+e(t)**2)/var*

...

Esta regra pode ser aplicada a outras variáveis que necessitem de um valor prévio desfasado.

Ainda a propósito de MAXIMIZE devemos chamar a atenção para a instrução NLPAR com a sua opção do número máximo de SUBITERATIONS e as relacionadas com os critérios usados para encontrar a solução óptima e determinar, antes disso, o caminho para o óptimo. Em MAXIMIZE, ITERATIONS controla o número máximo de iterações do programa a serem executadas.

A programação no RATS

Passemos a ver com mais cuidado como podemos usar o RATS na resolução dos problemas colocados por estruturas condicionadas da variância dos erros.

Lembremos que o nosso modelo de partida é o seguinte

$$\begin{aligned} y_t &= \beta \cdot x_t + \varepsilon_t \\ \varepsilon_t &= v_t \cdot \sqrt{\alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2} \end{aligned}$$

A forma de instruir o RATS para a estimação deste modelo é constituída pelas seguintes instruções

(i) *NONLIN b a_0 a_1*

(ii) *FRML e=y-b*x*

(iii) *FRML h=a_0+a_1*e{1}**2*

(iv) *FRML LIKELIHOOD=0.5*(log(h)+(e(t)**2)/h)*

(v) *LIN(NOPRINT) y*

(vi) *# x*

(vii) *COMPUTE b=%beta(1), a_0=%seesq, a_1=0.0*

(viii) *MAXIMIZE LIKELIHOOD 2 end*

O quadro interior constitui o bloco principal de instruções. De notar que iniciamos os cálculos com a segunda linha de comandos.

Considerando um modelo ARMA(1,4), a variável vem representada por

5.3. ALGUMAS OBSERVAÇÕES ADICIONAIS SOBRE A PESQUISA DO TIPO DE VARIÂNCIA CO

$$y_t = \alpha_0 + \alpha_1 \cdot y_{t-1} + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \beta_3 \cdot \varepsilon_{t-3} + \beta_4 \cdot \varepsilon_{t-4}$$

Para o qual podemos admitir a seguinte variância condicional

$$h_t = \alpha_0 + \alpha_1 \cdot (0,4 \cdot \varepsilon_{t-1}^2 + 0,3 \cdot \varepsilon_{t-2}^2 + 0,2 \cdot \varepsilon_{t-3}^2 + 0,1 \cdot \varepsilon_{t-4}^2)$$

Desta forma, devemos substituir a linha (ii) e (iii) do quadro acima, por

(ii) FRML $e=y-a_0-a_1*y\{1\}-b_1*e\{1\}-b_2*e\{2\}-b_3*e\{3\}-b_4*e\{4\}$

(iii) FRML $h=a_0+a_1*(.4*e\{1\}**.2+.3*e\{2\}**.2+.2*e\{3\}**.2+.1*e\{4\}**.2)$

depois de termos feito as alterações adequadas na indicação das variáveis a calcular em (i).

E no caso da variância condicional tipo GARCH(1,1)

$$h_t = \alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2 + \beta_1 \cdot h_{t-1}$$

temos para (iii)

(iii) FRML $h=a_0+a_1*e\{1\}**.2+b_1*h\{1\}$

No caso apresentado atrás de Robert Engle e Robins (1987), em que a média da variável obedece a um processo ARCH, passaremos a ter par (ii) e (iii)

(ii) FRML $e=y-a_0-a_1*h$

(iii) FRML $h=a_0+a_1*(.4*e\{1\}**.2+.3*e\{2\}**.2+.2*e\{3\}**.2+.1*e\{4\}**.2)$

Lembremos que num modelo GARCH(q,p) também podemos incluir variáveis exógenas, seja na definição da média, seja na definição da variância. Assim, na definição de h haveria que juntar, por exemplo no caso de uma variável, $c(L) \cdot Z_t$, onde Z representa uma variável exógena e $c(L)$ é o polinómio de defasamentos.

5.3 Algumas observações adicionais sobre a pesquisa do tipo de Variância condicional

Talvez a forma mais usual de conhecermos o tipo de variância defronte da qual nos encontramos seja determinar os resíduos e a partir deles fazer a nossa investigação.

Suponhamos que os resíduos são identificados por RES, então podemos fazer

**Supomos dados mensais*

CLEAR RES

LIN(NOPRINT) y / RES

CONSTANT

*SET RES2=RES*RES*

** Para estudarmos a estrutura AR dos erros*

COR(QSTATS,NUMBER=24,SPAN=4,DFC=1) RES

** Para a estrutura auto-regressiva da variância*

COR(PARTIAL=PACF,QSTATS,NUMBER=24,SPAN=4,DFC=1) RES2

```

* suspeitando de AR(1) para as variâncias
LIN(NOPRINT) RES2
# CONSTANT RES{1}
COMPUTE TRSQ=%NOBS*%RSQUARED
CDF CHISQR TRSQ 1
* se houvesse exclusão da H0:
NONLIN b a_0 a_1
FRML e=y-b*x
FRML v=a_0+a_1*e(t-1)**2
FRML L=.5*(log(v)+(e(t)**2)/v)
LIN(NOPRINT) y
# x
COMPUTE b=%beta(1), a_0=%seesq, a_1=0.0
MAXIMIZE L start-1 end

```

Tomemos agora o caso de possível verificação de um modelo ARMA(1,1) para y e que justifica pelas regras anteriores um ARCH(4). As instruções vêm agora dadas por

```

SET u=0.0; * definição transitória
NONLIN b_0 b_1 b_2 a_0 a_1
FRML e=y-b_0-b_1*y{1}-b_2*u{1}
FRML v=a_0+a_1*e(t-4)**2
FRML L=(u=e),.5*(log(v)+(e(t)**2)/v)
BOXJENK(noprint,constant,ar=1,ma=1) y
COMPUTE b_0=%beta(1), b_1=%beta(2), b_2=%beta(3)
COMPUTE a_0=%seesq, a_1=0.0
MAXIMIZE(iter=1000) L 5 *

```

Admita-se ainda que temos

$$\begin{aligned}
 y_t &= \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot \varepsilon_{t-1} + \beta_3 \cdot \varepsilon_{t-4} \\
 v_t &= \alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2 + \alpha_2 \cdot v_{t-1}
 \end{aligned}$$

um ARMA(AR=1,MA=1) para y combinado com um GARCH(1,1).

```

SET w=0.0
SET u=0.0
NONLIN b_0 b_1 b_2 b_3 a_0 a_1 a_2
FRML e=y-b_0-b_1*y{1}-b_2*u{1}-b_3*u{4}
FRML v=a_0+a_1*e{1}**2+a_2*w{1}
FRML L=(u=e),(w=v),.5*(log(v)+(e(t)**2)/v)
BOXJENK(noprint,constant,ar=1,ma=1) y
COMPUTE b_0=%beta(1), b_1=%beta(2), b_2=%beta(3), b_3=%beta(4)
COMPUTE a_0=%seesq, a_1=.1, a_2=.2
MAXIMIZE(iter=1000) L 4 *

```

5.3. ALGUMAS OBSERVAÇÕES ADICIONAIS SOBRE A PESQUISA DO TIPO DE VARIÂNCIA CO

Para vermos se os erros do modelo apresentam as características desejadas devemos fazer agora

```
SET RES=0.0
```

```
SET RES 4 * = y-%beta(1)-%beta(2)*y{1}-%beta(3)*RES{1}-%beta(4)*RES{4}
```

* e depois devemos estudar as auto-correlações das médias e das variâncias ...

e investigar os resultados obtidos.

Um resultado interessante em termos de processos de variância condicionada é o do conhecimento dos desvios-padrão. Tomando a penúltima caixa devemos fazer

```
SET DVPQ = 0.0
```

```
SET DVPQ * * = %beta(5)+%beta(6)*RES{1} **2+%beta(7)*DVPQ{1}
```

```
SET UPPER=y+2*DVPQ**.5
```

```
SET LOWER=y-2*DVPQ**.5
```

```
GRAPH(header='y e D-P condicionados',KEY=upleft,PATTERNS) 3
```

```
# y
```

```
# UPPER
```

```
# LOWER
```

A investigação dos processos da variância pode ser feita com utilização do procedimento BJIDENT, para além do conhecimento das auto-correlações parciais.

Ilustremos por fim um caso que nos conduziu a estimar duas hipóteses de comportamento da média alternativas

$$y_t = \beta_0 + \beta_1 \cdot Var_t + \varepsilon_t$$

$$Var_t = \alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2$$

ou

$$y_t = \beta_0 + \beta_1 \cdot Var_t + \varepsilon_t + \beta_2 \cdot \varepsilon_{t-3}$$

$$Var_t = \alpha_0 + \alpha_1 \cdot \varepsilon_{t-1}^2$$

Temos para o primeiro caso

```
SET u=0.0
```

```
NONLIN a_0 a_1 b_0 b_1
```

```
FRML var=a_0+a_1*u{1} **2
```

```
FRML e=y-b_0-b_1*var(t)
```

```
FRML L=(u=e),-.5*(log(var(t))+(e(t)**2)/var)
```

```
LIN(noprint) y
```

```
# constant
```

```
COMPUTE b_0=%beta(1), b_1=0.0
```

```
COMPUTE a_0=%seesq, a_1=1.0
```

```
MAXIMIZE L 2 *
```

e para o segundo

```
SET u=0.0
```

```

SET w=0.0
NONLIN a_0 a_1 b_0 b_1 b_2
FRML var=a_0+a_1*u{1}**2
FRML e=y-b_0-b_1*var(t)-b_2*w{3}
FRML L=(u=e),(w=e),-.5*(log(var(t))+(e(t)**2)/var)
BOXJENK(noprint,constant,ma=3) y
COMPUTE b_0=%beta(1), b_2=%beta(2), a_0=%seesq
COMPUTE b_1=0.0, a_1=0.0
MAXIMIZE L 3 *

```

Vejam os para finalizar estas observações o caso em que a série, por exemplo dos rendimentos, exibe variância condicional assimétrica. Supomos um GARCH(1,1) e colocaremos em itálico a transformação sugerida por Lawrence Glosten e Runkle (1993)

```

Declare series u
Declare series w
NONLIN b_0 a_0 a_1 a_2
NONLIN b_0 a_0 a_1 a_2 a_3
FRML e=y-b_0
FRML h=a_0+a_1*w{1}+a_2*u{1}**2
FRML h=a_0+a_1*w{1}+a_2*u{1}**2+%if(u{1};0.0,a_3*u{1}**2,0.0)
FRML L=(w=h),(u=e),-.5*(log(h(t))+(e(t)**2)/h(t))
LIN(noprint) y / u
# constant
COMPUTE b_0=%beta(1)
COMPUTE a_0=%seesq, a_1=.05, a_2=.05
SET w=%seesq
MAX(method=bhhh,recurs,iter=1000) L start end

```

O recurso à definição dos vectores *u* e *w* corresponde a uma alternativa elegante da definição dessas variáveis como o fizemos atrás.

Refira-se ainda que em *Maximize* podemos indicar uma opção para a estimação da matriz das variâncias/covariâncias, **ROBUSTERRORS**.

Também a hipótese apontada mais acima de uma processo de variância condicional infinita pode ser calculado deste que a restrição seja indicada no conjunto das instruções de cálculo. Bollerslev (1986) designaram um processo deste tipo por IGARCH (integrated GARCH).

Capítulo 6

Métodos de Estimação em Painel Estático

A estimação econométrica em painel consiste em aplicar os modelos económicos a dados de natureza *cross-section* e de sucessão cronológica. Usamos dados respeitantes a “indivíduos”, “famílias”, “unidades de produção” ou “países” para um dado número de observações temporais. Estas bases constituem, em geral, uma maior variabilidade de valores que os constantes de simples séries *cross-section* ou sucessões cronológicas, pelo que permitirão a obtenção de estimadores mais eficientes. Por outro lado, o estudo em painel permite resolver o problema da reduzida dimensão temporal das nossas amostras, para além de possibilitar um melhor conhecimento da heterogeneidade individual.

Vamos apresentar alguns dos métodos de estimação de modelos lineares não dinâmicos. A exposição sobre o método de efeitos aleatórios, *random effects*, não focará o tipo de metodologia usada. Diferentes processos de mínimos quadrados generalizados, *GLS*, foram propostos para solucionar o problema da matriz desconhecida das variâncias dos erros¹.

Os modelos dinâmicos incluem a variável dependente desfasada como uma das variáveis explicativas. Surge então um problema de não convergência dos estimadores dos mínimos quadrados (m.q.o.) devido à correlação entre os erros e as variáveis explicativas. O problema é tanto mais grave quanto mais reduzida, do ponto de vista temporal, for a nossa base de dados².

Após a apresentação dos diferentes métodos de estimação de modelos de painel estático exporemos os testes aos efeitos individuais e à variância dos erros. Estes testes são importantes para a selecção do tipo de modelos que nos interessa.

Quando construímos uma base de dados, um dos primeiros passos que devemos

¹Veja-se Baltagi (2001) sobre os diferentes métodos propostos. Estes autor defende a pouca importância da escolha de um ou outro método (p.19).

²Veja-se a este propósito a metodologia de Arellano e Bond. Arellano e Bond (1988), Arellano e Bond (1991) e mais recentemente Arellano (2003).

ter em conta respeito à justificação da “junção” dos nossos dados. Será que se justifica juntarmos os “indivíduos” para os quais possuímos dados?³ Esta questão é em tudo semelhante à questão da estabilidade de um modelo em termos de análise temporal: justificar-se-à a junção de períodos diferentes na nossa base? O que significa que o vulgar teste de Chow deve esclarecer a situação. O problema é que para dados modelos o teste de Chow recusa a junção dos dados mesmo quando esta se justifica⁴.

6.1 Métodos de Estimação em Painel Estático

Admitamos que o modelo geral a testar é do tipo $y = \mathbf{x}' \cdot \beta + \epsilon$. Então o modelo de base para *pooling* é o seguinte

$$y_{it} = \mathbf{x}' \cdot \beta + \mathbf{z}' \cdot \alpha_i + \epsilon_{it} \quad (6.1)$$

onde temos N indivíduos e T períodos.

A partir deste modelo (equação(6.1)) definimos as diferentes hipóteses de estudo consequentes à junção de dados.

- Quando \mathbf{z}' contém apenas um termo constante, os m.q.o. conduzem a estimadores convergentes e eficientes.
- Se \mathbf{z}'_{it} for constituído por variáveis não observáveis mas correlacionadas com \mathbf{x}'_{it} , o estimador de m.q.o. de β é enviesado e não convergente (equivalente à situação de variável omitida). Neste caso podemos ter um modelo com $\alpha_i = \mathbf{z}'_i \cdot \alpha$, onde impomos que aquele α agrupa todos os efeitos individuais e representa uma média condicionada. Este modelo é conhecido por modelo de efeitos fixos, ou também como modelo de m.q.o. com variáveis mudas (LSVD).
- No caso de aquelas variáveis não observáveis não estarem correlacionadas com \mathbf{x}'_{it} , estamos perante um efeito aleatório atribuído a cada grupo da nossa base. Este modelo é designado por modelo de efeitos aleatórios, *random effects*. Este modelo pode ser visto como se as suas unidades (indivíduos) resultassem de uma tiragem aleatória de uma população mais alargada. Balestra e Nerlove (1966) propuseram o seguinte modelo

$$y_{it} = \mathbf{x}'_{it} \cdot \beta + \alpha_i + \gamma_t + \epsilon_{it} \quad (6.2)$$

onde à formulação inicial se juntou um efeito temporal comum a cada indivíduo.

³Não esqueçamos que o “economista” tem permanentemente uma grande sede de dados.

⁴O que acontece quando a especificação correcta de um modelo é de efeitos aleatórios (ver mais abaixo), Baltagi (1981). Veja-se Baltagi (2002), p. 319.

O modelo conjunto (*pooled*) pode ser estimado de três maneiras diferentes:

a) de acordo com a estimação genérica original;

b) usando os desvios da média

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \cdot \beta + \epsilon_{it} - \bar{\epsilon}_i$$

que é conhecido pelo modelo *within*;

c) usando as médias individuais

$$\bar{y}_i = \bar{\mathbf{x}}_i' \cdot \beta + \alpha + \bar{\epsilon}_i$$

que é conhecido pelo modelo entre indivíduos, *between*. Por vezes também é designado por estimador da média individual.

Os valores estimados de β são os mesmos nos casos a) e b).

6.1.1 Teste aos efeitos individuais

O primeiro teste respeita à hipótese nula dos termos constantes individuais e o segundo à hipótese nula de serem iguais.

Com o primeiro teste queremos saber se existe justificação para incluir variáveis mudas individuais para os diferentes indivíduos estudados. O problema resume-se a um vulgar teste F ao conjunto dos coeficientes com base num modelo LSDV. Um problema existirá se os coeficientes forem estimados com o modelo *within*. Neste caso, o teste vulgarmente fornecido pelos programas informáticos normais deverá ser corrigido⁵.

Com o segundo teste pretendemos saber se devemos insistir na utilização de um modelo LSDV ou se o simples *pooling* de dados num modelo com constante comum, com estimação pelo método dos m.q.o., é adequado. O teste F vem neste caso dado por⁶

$$F_{N-1, N \cdot T - N - K} = \frac{(R_{LSDV}^2 - R_{Pooled}^2) / (N - 1)}{(1 - R_{LSDV}^2)}$$

6.1.2 Testes à variância dos erros individuais

O estimador β no caso do modelo RE e no caso dos m.q.o. é equivalente a uma média ponderada dos valores estimados com o modelo *within* e o modelo *between*. Quando a variância do erro associado aos indivíduos é nula não faz sentido não utilizar o modelo LSDV. O teste da presença de efeitos aleatórios deve ser conduzido em dois passos (a) e b)), que significam outros tantos testes. Vejamos em que consistem.

⁵Veja-se Baltagi (2002), p. 311.

⁶Veja-se Greene (2003), p. 289.

- a) Breusch e Pagan (1980) propuseram um teste a partir dos erros de um modelo de simples *pooling*, em que a hipótese nula consiste na nulidade da variância dos efeitos aleatórios individuais. A estatística vem dada por

$$LM = \frac{N \cdot T}{2 \cdot (T - 1)} \cdot \left(\frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right)^2$$

onde $e_i = \sum_{t=1}^T e_{it}$. A exclusão da hipótese nula significa o abandono do modelo de *pooling* simples com uma única constante.

- b) Claro que a questão a que ainda devemos responder respeita à retenção de um modelo LSDV ou RE (de efeitos fixos ou aleatórios). O problema principal reside no facto de o modelo LSDV reduzir imenso os graus de liberdade e de o modelo RE implicar que os efeitos individuais não estão correlacionados com as restantes variáveis independentes, o que não é razoável. Hausman (1978) propôs o teste conhecido pelo seu nome, e que se baseia no seguinte: se aqueles efeitos não estiverem correlacionados, os estimadores usados com LSDV e RE são convergentes mas os m.q.o., usados no primeiro, são ineficientes. O teste proposto para a hipótese nula de não correlação, de retenção do modelo RE, é um teste Wald dado por

$$W = \hat{q}' \cdot [\text{var}(\hat{q})]^{-1} \cdot \hat{q}$$

onde $\hat{q} = \hat{\beta}_{RE} - \tilde{\beta}_{Within}$ e $\text{var}(\hat{q}) = \text{var}(\tilde{\beta}_{Within}) - \text{var}(\hat{\beta}_{RE})$, tendo W uma distribuição do χ^2 com $(K - 1)$ graus de liberdade. Os chapéus $\hat{\cdot}$ e $\tilde{\cdot}$ representam os habituais valores estimados e K é a dimensão do vector β .

Bibliografia

- Amisano, G., e C. Giannini (1997), *Topics in Structural VAR Econometrics*. Springer verlag, Berlin, 2nd edn.
- Andrade, J. S. (1993), *Análise Econométrica, Uma Introdução Para Economistas e Gestores*. Texto Editora, Lisboa.
- (2003), “Régimes Monétaires et Théorie Quantitative du Produit Nominal au Portugal (1854-1998),” *Estudos* 6, G.E.M.F., Faculdade de Economia, Coimbra.
- Aninda Banerjee, Juan Dolado, J. G., e D. Hendry (1993), *Co-Integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford University Press, Oxford.
- Arellano, M. (2003), *Panel Data Econometrics*. Oxford University Press, Oxford.
- Arellano, M., e S. R. Bond (1988), “Dynamic Panel Data Estimation Using DPD-A Guide for Users,” Working Paper 88/15, Institute for Fiscal Studies.
- (1991), “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies*, 58 277–97.
- Bação, P. (1999), “Nota Sobre a Estimação de Vectores de Cointegração Com Os Programas CATS in RATS, PCFIML e EVIEWS,” *Working Paper, 9, GEMF, Faculdade de Economia, Coimbra*, Working Paper, 9.
- Balestra, P., e M. Nerlove (1966), “Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas,” *Econometrica*, 34 585–612.
- Baltagi, B. (1981), “Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Errors Components Model,” *Journal of Econometrics*, 17 21–49.
- (2001), *Econometric Analysis of Panel Data*. John Wiley, Chichester, 2nd edn.

- (2002), *Econometrics*. Springer-Verlag, Berlin, 3rd edn.
- Becker, W., e W. Greene (2001), "Teaching Statistics and Econometrics to Undergraduates," *Journal of Economic Perspectives*, 15, 169–82.
- Bernanke, B. (1986), "Alternative Explanations of the Money-Income Correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25 49–99.
- Bhargava, A. (1986), "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies*, 53 369–84.
- Blanchard, O., e M. Watson (1986), *The American Business Cycle: Continuity and Change*. Are Business Cycles All Alike?, pp. 123–56. University of Chicago Press.
- Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307–27.
- Breusch, T. S., e A. R. Pagan (1980), "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *Review of Economic Studies*, 47 239–253.
- Campbell, J., e G. Mankiw (1987), "Are Output Fluctuations Transitory," *Quarterly Journal of Economics*, 102, 857–80.
- Chow, G. (1983), *Econometrics*. McGraw-Hill, New York.
- Cochrane, J. (1988), "How Big Is the Random Walk in GNP?," *Journal of Political Economy*, 96, 893–920.
- Dhrymes, P. (1971), *Distributed Lags, Problems of Estimation and Formulation*. Holden-Day, San Francisco.
- Dickey, D., e W. Fuller (1979), "Distribution of the Estimators for Time Series Regressions with a Unit Root," *Journal of the American Statistical Association*, 74 427–31.
- Doornik, J., e H. Hansen (1994), "A Practical Test for Univariate and Multivariate Normality," *Discussion Paper, Nuffield College*.
- E. Ghysels, H. L., e J. Noh (1994), "Testing for Unit Roots in Seasonal Time Series," *Journal of Econometrics*, 62 415–42.
- Enders, W. (2003), "RATS Programming Manual," e-book, Estima.
- Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.
- (ed.) (1995), *ARCH, Selected Readings*. Oxford University Press, Oxford, reprint 2000 edn.

- (2001), “GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics,” *Journal of Economic Perspectives*, 15, 157–68.
- Engle, R., e C. Granger (1987), “Co-Integration and Error Correction: Representation, Estimation, and Testing,” *Econometrica*, 55, 251–76.
- Engle, R., e B. Yoo (1987), “Forecasting and Testing in Co-Integrated Systems,” *Journal of Econometrics*, 35, 143–59.
- Friedman, M. (1992), “Do Old Fallacies Ever Die?,” *Journal of Economic Literature*, XXX, 2129–32.
- Granger, C. W., e P. Newbold (1974), “Spurious Regressions in Econometrics,” *Journal of Econometrics*, 2, 111–20.
- Greene, W. (2003), *Econometric Analysis*. Prentice Hall, Prentice, 5rd edn.
- Hallwood, P., e R. MacDonald (2000), *International Money and Finance*. Blackwell, Oxford, 3rd edn.
- Hamilton, J. (1994), *Time Series Analysis*. Princeton University Press, Princeton.
- Hansen, H., e K. Juselius (1995), *CATS in RATS, Cointegration Analysis of Time Series*. Estima, Evanston, 1st edn.
- Hausman, J. A. (1978), “Specification Tests in Econometrics,” *Econometrica*, 46 1251–71.
- Hayashi, F. (2000), *Econometrics*. Princeton University Press, Princeton.
- Hendry, D. (1995), *Dynamic Econometrics*. Oxford University Press, Oxford.
- Hendry, D., e K. Juselius (2000), “Explaining Cointegration Analysis: Part,” *The Energy Journal*, 21, 1–42.
- Hoffman, D., e R. Rasche (1996), *Aggregate Money Demand Functions, Empirical Applications in Cointegrated Systems*. Kluwer Academic Pub., Boston.
- Hotelling, H. (1933), “Review of The Triumph of Mediocrity in Business, by Horace Secrist,” *Journal American Statistical Association*, 28, 463–5.
- J. Davidson, D. Hendry, F. S., e S. Yeo (1978), “Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers’ Expenditures and Income in the United Kingdom,” *Economic Journal*, 88 661–92.
- Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

- Johansen, S., e K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration - With Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169–210.
- Lawrence Glosten, R. J., e D. Runkle (1993), "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–801.
- Litterman, R. (1986), "Forecasting with Bayesian Vector Autoregressions-Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38.
- Lutkepohl, H. (1994), "Interpretation of Cointegration Relations-Comments on "Estimating Systems of Trending Variables."," *Econometric Reviews*, 13, 391–4.
- Maddala, G. S., e I.-M. Kim (1998), *Unit Roots, Cointegration and Structural Change*. Cambridge University Press, Cambridge.
- Marques, C. R. (1998), *Modelos Dinâmicos, Raízes Unitárias e Cointegração*. EDI-NOVA, Lisboa.
- McCloskey, D., e S. Ziliak (1996), "The Standard Error of Regressions," *Journal of Economic Literature*, XXXIV, 97–114.
- Mizen, P. (1994), *Buffer Stock Models and the Demand for Money*. Macmillan, London.
- Osterwald-Lenum, M. (1992), "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, 461–72.
- Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.
- Perron, P. (1997), "Further Evidence on Breaking Trend Functions in Macroeconomic Variables," *Journal of Econometrics*, 80 355–85.
- Phillips, P. (1986), "Understanding Spurious Regressions in Econometrics," *Journal of Econometrics*, 33, 311–40.
- (1987), "Time Series Regression with a Unit Root," *Econometrica*, 55 277–301.
- Phillips, P., e P. Perron (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75 335–46.
- Pivetta, F., e R. Reis (2002), "The Persistence of Inflation in the United States," Working paper, Harvard University.

- Robert Engle, D. L., e R. Robins (1987), "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model," *Econometrica*, 55, 391–407.
- S. Hylleberg, R. Engle, W. G., e B. Yoo (1990), "Seasonal Integration and Cointegration," *Journal of Econometrics*, 44 215–38.
- Schmidt, D. K. P. C. B. P. P., e Y. Shin (1992), "Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?," *Journal of Econometrics*, 54, 159–78.
- Sims, C. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- (1986), "Are Forecasting Models Usable for Policy Analysis?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 2–16.
- Soren Johansen, R. M., e B. Niesen (2000), "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend," *Econometrics Journal*, 3, 216–49.
- Stock, J. (1991), "Confidence Intervals for the Largest Autoregressive Root in U. S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28 435–59.
- Stock, J., e M. Watson (2001), "Vector Autoregressions," *Journal of Economic Perspectives*, 15, 101–15.
- Thomas Doan, R. L., e C. Sims (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.
- Watson, M. (1994), *Handbook of Econometrics* vol. IV, chap. Vector Autoregressions and Cointegration, pp. 2844–915. Elsevier Science Ltd., Amsterdam.
- Yule, G. U. (1926), "Why Do We Sometimes Get Nonsense Correlations Between Time-Series?," *Journal of the Royal Statistical Society*, 89 2–9, 30–41.