ECONOMETRIA II

Modelos DL e ADL, Raízes Unitárias e Cointegração: uma Introdução

Artur C. B. da Silva Lopes

Instituto Superior de Economia e Gestão

Universidade Técnica de Lisboa

Prefácio

Este texto de apoio resulta basicamente da agregação de notas de leitura que escrevi para apoiar a leccionação de algumas das matérias da disciplina de Econometria II, do 3º ano da licenciatura em Economia do ISEG, no ano lectivo de 1997/98.

Como não podia deixar de ser dada a finalidade a que se destinavam, e como o próprio título indica, estas notas têm apenas carácter introdutório relativamente às matérias abordadas. Assim, sugere-se que o leitor interessado prossiga o seu estudo com base nos trabalhos de Campbell e Perron [1991], Davidson e MacKinnon [1993] cap. 20, Boswijk [1994], Hendry [1995], Maddala e Kim [1998] ou Marques [1998], para citar apenas alguns exemplos. Ainda a título de exemplo, mas requerendo um nível de conhecimentos mais avançado, sugerem-se também as obras de Banerjee *et al.* [1993], Hamilton [1994] ou Johansen [1995].

Finalmente, aproveito a oportunidade para agradecer a todos os colegas que formularam críticas e sugestões à primeira versão destas notas. Obviamente, assumo a responsabilidade por todos os erros e omissões que possam ainda subsistir.

ISEG, Lisboa, Março de 1999

Artur C. B. da Silva Lopes

Índice

1 Estimação de Modelos DL(s) 5

1.1 A determinação de s 5
1.2 A estimação dos coeficientes de desfasamento 6
1.2.1 Estimação OLS e transformações lineares 8
1.2.2 Um exemplo simples: aplicação ao modelo DL(1) 9
1.2.3 Generalização: aplicação aos modelos DL(s) 10

2 O Modelo ADL(1,1) e o Modelo de Correcção de Erros 12

2.1 A generalidade do modelo ADL(1,1) 12
2.2 Transformações lineares do modelo ADL(1,1) 14
2.3 O modelo de correcção de erros 16

3 O Modelo ADL(r, s), o MCE e a Modelização do Geral para o Particular 18

3.1 A reparametrização em MCE dos modelos ADL(r, s) 18
3.2 A modelização do geral para o particular 21
3.3 A abordagem do particular para o geral 23

4 Raízes Unitárias: uma Introdução 25

4.1 Introdução: conceitos básicos e motivações 26
4.1.1 Processos estacionários e processos integrados 26
4.1.2 Séries económicas: estacionárias em tendência (TSP) ou por diferenciação (DSP)? 30
4.1.3 Regressões espúrias 34
4.2 Testes de raízes unitárias: Testes DF e ADF 36
4.2.1 Testes DF 37
4.2.2 Testes ADF 40

5 Cointegração: uma Introdução 42

5.1 Definição 42
5.2 Exemplos 43
5.3 Estimação OLS 45
5.4 Cointegração e MCE: o Teorema de Representação de Granger 48
5.5 Testes de cointegração 50

5.5.1 Testes de Engle-Granger 50
5.5.2 O teste t-MCE 51
5.6 Estimação do MCE condicional 53
5.6.1 Método dos dois passos de Engle e Granger 53
5.6.2 Método num só passo 53

1 Estimação de Modelos DL(s)

Actualmente, a utilização de modelos de desfasamentos distribuídos finitos (DL(s), com s a representar o comprimento de desfasamento ou desfasamento máximo) na modelização de fenómenos económicos não tem, praticamente, qualquer significado. Com efeito, reconhece-se cada vez mais que a modelização adequada dos efeitos dinâmicos requer a introdução de desfasamentos da variável dependente no conjunto dos regressores, conduzindo aos modelos ADL (*autoregressive distributed lag*). Todavia, os problemas de colinearidade destes encontram-se também presentes nos modelos DL(s), pelo que podem ser abordados no seu contexto.

A estimação de modelos DL(s) do tipo

$$y_t = \mu + \delta_0 x_t + \delta_1 x_{t-1} + ... + \delta_s x_{t-s} + u_t,\ t = 1, 2, ..., T,$$

ou

$$y_t = \mu + \mathrm{D}(L) x_t + u_t,\ t = 1, 2, ..., T,$$

onde $\mathrm{D}(L) = \delta_0 + \delta_1 L + ... + \delta_s L^s$, coloca dois problemas distintos:

1) a determinação de s e

2) a estimação dos coeficientes de desfasamento.

Enquanto que o primeiro desses problemas será abordado brevemente, o segundo será objecto de atenção detalhada.

1.1 A determinação de s

Para a determinação da ordem máxima de desfasamento encontram-se disponíveis vários métodos, todos eles baseados na fixação de um valor máximo inicial, S, $s \leq S$. Para a fixação deste valor não há regras gerais rígidas: o investigador deverá basear-se em informação obtida *a priori* e ter em consideração a dimensão da amostra disponível e a frequência dos dados. Assim, por exemplo, para dados anuais, considerar $S = 6$ parece ser algo excessivo, sobretudo se a dimensão da amostra fôr pequena, mas o mesmo valor poderá revelar-se insuficiente para dados trimestrais e para uma amostra de dimensão razoável.

Uma possibilidade consiste em utilizar critérios de discriminação ou de selecção de modelos. Por exemplo, escolher s de forma a maximizar o valor do coeficiente de determinação ajustado ou corrigido, $\bar{R}^2(s)$. Como é sabido, esta regra de decisão é equivalente à de escolha do modelo que minimiza a estimativa OLS da variância dos erros, $\hat{\sigma}^2 = e'e/(T-k)$. Em alternativa a esta estatística poder-se-á utilizar o *critério de informação de Akaike*:

$$\text{AIC} = -\frac{2}{T}\ln\hat{L} + \frac{2k}{T},$$

onde $\hat{L}$ representa o valor maximizado da função de verosimilhança (obtido assumindo que os erros são iid e normais; recorde-se que, nestas condições, o estimador OLS se identifica com o estimador de máxima verosimilhança). Tal como o $\bar{R}^2$, também este critério procura reflectir o conflito entre maximização da "bondade de ajustamento", por um lado, e obediência ao *princípio da parcimónia*, isto é, minimização do número de parâmetros empregues, por outro. Neste caso, como se compreende facilmente, a regra de decisão consiste em seleccionar o modelo (o valor de s) que minimiza a estatística. Ainda uma outra alternativa neste domínio consiste em seleccionar o valor de s que minimiza a estatística (Bayesiana) de Schwarz:

$$\text{SBIC} = \ln\frac{e'e}{T} + \frac{k\ln T}{T}.$$

Muitos outros critérios existem ainda com o mesmo propósito. Para além de serem os mais comuns, estes podem ser facilmente empregues quando se utiliza o TSP.

Uma outra possibilidade baseia-se em testes-F sequenciais encaixados sobre os últimos coeficientes de desfasamento, iniciando-se o procedimento com o modelo DL(S). Ou seja, a análise é iniciada pelo modelo mais geral que é considerado e prossegue na direcção de modelos mais simples, seguindo uma estratégia que é conhecida como de *modelização do geral para o particular.* Neste caso, o valor seleccionado para s obtém-se quando se rejeita a hipótese nula de que os últimos coeficientes são conjuntamente nulos (para mais pormenores veja-se, por exemplo, Judge *et al.* [1988], pp. 723-7).

1.2 A estimação dos coeficientes de desfasamento

Escolhido o valor de s, enfrenta-se o problema da estimação dos coeficientes de desfasamento. Como é sabido, desde que x_t seja estritamente exógena e sob a condição de os erros serem iid [1] o estimador OLS é BLUE.

[1] Refira-se que, no contexto destes modelos, esta hipótese é bastante "heróica"; em geral, para que o modelo não apresente sintomas de autocorrelação residual, será necessário considerar modelos ADL, isto é, incluir também desfasamentos da variável y_t como regressores.

Todavia, como as séries económicas variam muito lentamente, isto é, como geralmente se encontram muito autocorrelacionadas (positivamente), as correlações entre x_t e os seus valores desfasados (e entre estes) devem ser elevadas. Por outras palavras: na estimação de modelos DL(s) são muito vulgares os problemas de (multi)colinearidade (grave) entre os regressores.

Em particular, a precisão com que é estimada a estrutura de desfasamentos é, em geral, reduzida. Por outro lado, a probabilidade de cometer erros de tipo II [2] quando se efectuam testes do tipo de $H_0 : \delta_i = 0$ *vs* $H_1 : \delta_i \neq 0$ $(i = 0, 1, 2, ..., s)$ tende a ser inflacionada (devido à "inflação" dos denominadores dos rácios-t), isto é, a potência dos testes tende a ser reduzida. Ou seja, caso a escolha de s de baseie neste tipo de testes, é bastante provável que se acabe por chegar a um modelo subparametrizado.

Uma consequência adicional, também comum nas regressões "atacadas pela colinearidade", é a da incerteza acrescida sobre os coeficientes: ao excluir um desfasamento com base na aceitação de uma hipótese nula como a de cima, as estimativas dos coeficientes dos desfasamentos não excluídos alteram-se substancialmente, provocando no investigador grande incerteza sobre a estrutura dos efeitos dinâmicos.

Note-se contudo que, apesar de todos estes problemas, pode provar-se que, em geral, o multiplicador de longo prazo pode continuar a ser estimado com bastante precisão. Por outras palavras, o estimador desse multiplicador não é, geralmente, afectado pelos problemas decorrentes da colinearidade entre os regressores. O leitor interessado poderá encontrar em Davidson, R. e MacKinnon, J. G. [1993], pp. 673-4, uma explicação para este facto.

Todavia, muito frequentemente, o principal interesse do investigador reside na estimação precisa da estrutura de desfasamentos, isto é, dos δ_i, e não apenas na do multiplicador total ou de longo prazo. A discussão que se segue visa tratar este problema. Refira-se que, para esse efeito, se encontram disponíveis duas metodologias distintas, abordando-se com detalhe apenas uma delas. A outra, baseada na imposição de restrições (polinomiais) sobre os coeficientes, teve o seu apogeu nas décadas de 60 e de 70. Trata-se do método dos desfasamentos distribuídos polinomiais (PDL) ou desfasamentos de Almon (veja-se, por exemplo, Stewart [1991], pp. 181-6 ou Judge *et al.* [1988], pp. 729-34).

[2] Recorde-se que o erro de tipo II de um teste estatístico consiste em aceitar a hipótese nula quando ela é, na realidade, falsa.

1.2.1 Estimação OLS e transformações lineares

A abordagem cujo estudo será privilegiado baseia-se em transformações lineares do modelo original, isto é, na reparametrização do modelo inicial com base em transformações lineares. Para proceder a esse estudo, convirá começar por salientar um resultado segundo o qual o estimador OLS de modelos de regressão lineares é invariante a transformações lineares. Por outras palavras: é indiferente estimar o modelo (linear) inicial pelo OLS ou estimar o modelo reparametrizado e obter as estimativas dos parâmetros do modelo inicial usando as transformações lineares da reparametrização. Ora, a estimação com base no modelo reparametrizado pode revelar-se vantajosa por permitir atenuar os problemas de colinearidade do modelo original.

De facto, considere-se o modelo de regressão linear com k regressores

$$y_t = x_t'\beta + u_t,\ t = 1, 2, ..., T,$$

e sejam H uma matriz qualquer, $k \times k$, regular e conhecida e γ um vector conhecido, $k \times 1$. Subtraindo à equação anterior $x_t'\gamma$ (que é uma função linear conhecida de x_t), tem-se

$$(y_t - x_t'\gamma) = x_t'(\beta - \gamma) + u_t.$$

Em seguida, como $HH^{-1} = I_k$, pode escrever-se

$$(y_t - x_t'\gamma) = x_t'H[H^{-1}(\beta - \gamma)] + u_t,$$

ou seja,

$$y_t^* = {x^*}'_t\beta^* + u_t,$$

onde $\beta^* = H^{-1}(\beta - \gamma)$. Note-se que o modelo inicial foi apenas reparametrizado: continuamos a ter k coeficientes para estimar (o vector β^*), de k regressores ($x_t^{*\prime}$) e o erro inicial (u_t). É indiferente começar por estimar β e obter o estimador de β^* indirectamente (com $\tilde{\beta}^*$) ou estimar directamente β^* do modelo transformado (com $\hat{\beta}^*$): tanto as estimativas dos coeficientes como as matrizes de covariâncias dos estimadores são numericamente iguais.

Para mostrar que $\hat{\beta}^* = \tilde{\beta}^*$ começa por notar-se que

$$\hat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*,$$

e que

$$\tilde{\beta}^* = H^{-1}(\hat{\beta} - \gamma).$$

Ora,

$$\begin{aligned}\hat{\beta}^* &= (X^{*\prime}X^*)^{-1}X^{*\prime}y^* = (H'X'XH)^{-1}H'X'(y - X\gamma) \\ &= H^{-1}(X'X)^{-1}X'(y - X\gamma) \\ &= H^{-1}[(X'X)^{-1}X'y - (X'X)^{-1}X'X\gamma \\ &= H^{-1}(\hat{\beta} - \gamma) = \tilde{\beta}^*.\end{aligned}$$

Por outro lado, também se mostra facilmente que

$$\text{Var}[\hat{\beta}^*] = \text{Var}[\tilde{\beta}^*],$$

que pode ser uma matriz diagonal apesar de a matriz (X'X) do modelo inicial ser uma matriz quase singular (devido à colinearidade). Ou seja, no modelo transformado pode existir ortogonalidade entre os regressores, sendo o problema inicial da colinearidade eliminado através das transformações lineares. Deste facto conclui Hendry que "a colinearidade não é uma propriedade do conjunto de regressores mas sim da parametrização do modelo".

1.2.2 Um exemplo simples: aplicação ao modelo DL(1)

Para ilustrar as potencialidades das transformações lineares nos modelos DL, comecemos por considerar um simples modelo DL(1), onde nem sequer se inclui termo independente,

$$y_t = \delta_0 x_t + \delta_1 x_{t-1} + u_t \tag{1}$$

e onde se supõe que, como é usual nas variáveis económicas, x_t se encontra muito autocorrelacionada (embora possa ser estacionária, por exemplo, $x_t = 0.9x_{t-1} + \epsilon_t,\ \epsilon_t \sim iid$). Nestas circunstâncias, é bastante provável que:

— os estimadores dos coeficientes tenham desvios-padrões (variâncias) muito elevados(as);

— (como consequência) um dos coeficientes (ou até mesmo ambos) pode não aparecer como estatisticamente significativo, o que levará à sua exclusão indevida do modelo;

— (como consequência) a exclusão desse regressor conduzirá a uma grande alteração na estimativa do coeficiente do outro [3], provocando grande incerteza no investigador.

[3] Recorde-se que a reduzida robustez das estimativas a ligeiras alterações da especificação é uma das consequências usuais da colinearidade entre os regressores.

Neste caso, a resolução do problema é bastante simples: basta substituir x_t por $x_{t-1} + \Delta x_t$ (ou somar e subtrair $\delta_0 x_{t-1}$ ao lado direito da equação), obtendo-se

$$y_t = \delta_0 \Delta x_t + (\delta_0 + \delta_1) x_{t-1} + u_t, \tag{2}$$

modelo que possui as seguintes vantagens:

a) se x_t e x_{t-1} estiverem altamente correlacionadas então, em geral, Δx_t e x_{t-1} estarão fracamente correlacionadas (serão quase ortogonais), o que permitirá atenuar substancialmente os problemas de colinearidade e, em particular, a estimativa de $\delta_0 + \delta_1$ será praticamente insensível à presença (ou ausência) de Δx_t na regressão;

b) obtém-se imediatamente uma estimativa do multiplicador de longo prazo ($\delta_0 + \delta_1$, através do coeficiente de x_{t-1}), bem como da variância do seu estimador.

Da subsecção anterior sabemos que as estimativas OLS serão exactamente iguais às do modelo (1). Todavia, as vantagens da reparametrização justificam a utilização preferencial de (2).

1.2.3 Generalização: aplicação aos modelos DL(s)

Para proceder de forma semelhante relativamente ao modelo geral DL(s) é necessário começar por considerar um resultado preliminar sobre decomposição de polinómios no operador de desfasamento, que se apresenta em seguida sem demonstração [4].

P1 Considere-se um polinómio de ordem p em L, $a(L) = \sum_{j=0}^{p} a_j L^j = a_0 + a_1 L + a_2 L^2 + ... + a_p L^p$ e definam-se os coeficientes $d_i = -\sum_{j=i+1}^{p} a_j$, $i = 0, 1, ..., p-1$ e $d_p = 0$. Então, pode escrever-se:

$$a(L) = a(1) + \sum_{i=0}^{p-1} d_i L^i (1 - L), \tag{3}$$

ou seja,

$$a(L) = a(1) + d(L)(1 - L), \tag{4}$$

onde $d(L)$ é um polinómio de ordem $p-1$ em L.

[4] Para a qual se convida o leitor mais interessado.

Usando o resultado anterior sobre o modelo DL(s) inicial, tem-se:

$$\begin{aligned} y_t = & \; D(L)x_t + u_t \\ = & \; [D(1) + \textstyle\sum_{i=0}^{s-1} \beta_i L^i (1-L)]x_t + u_t \\ = & \; \delta x_t + \textstyle\sum_{i=0}^{s-1} \beta_i \Delta x_{t-i} + u_t, \end{aligned}$$

onde se fez $\delta = D(1)$ para representar o multiplicador de longo prazo e $\Delta = (1 - L)$. Mais uma vez a estimação OLS deste modelo produz as mesmas estimativas dos coeficientes originais que a do modelo inicial. Aliás, se se pretender obter os coeficientes de desfasamento basta considerar que

$$\beta_i = -\sum_{j=i+1}^{s} \delta_j, \; i = 0, 1, ..., s-1.$$

Todavia, o modelo reparametrizado oferece várias vantagens: a) permite-nos obter imediatamente uma estimativa do multiplicador de longo prazo e da variância do seu estimador; b) os problemas de colinearidade do modelo original deverão ser substancialmente atenuados, uma vez que tanto a correlação de x_t com os valores desfasados de Δx_t como as correlações entre estes deverão ser bem menores que entre os regressores do modelo original.

Alternativamente, a reparametrização pode basear-se noutro resultado sobre decomposição de polinómios em L.

P2 Considere-se um polinómio de ordem p em L, $a(L) = \sum_{j=0}^{p} a_j L^j = a_0 + a_1 L + a_2 L^2 + ... + a_p L^p$ e definam-se os coeficientes $c_0 = a_0$, $c_i = -\sum_{j=i+1}^{p} a_j$, $i = 1, ..., p-1$ e $c_p = 0$. Então, pode escrever-se:

$$a(L) = a(1)L + \sum_{i=0}^{p-1} c_i L^i (1-L), \tag{5}$$

ou seja,

$$a(L) = a(1)L + c(L)(1-L), \tag{6}$$

onde $c(L)$ é um polinómio de ordem $p-1$ em L.

Neste caso tem-se:

$$\begin{aligned} y_t = & \; D(L)x_t + u_t \\ = & \; [D(1)L + \textstyle\sum_{i=0}^{s-1} \alpha_i L^i (1-L)]x_t + u_t \\ = & \; \delta x_{t-1} + \textstyle\sum_{i=0}^{s-1} \alpha_i \Delta x_{t-i} + u_t, \end{aligned}$$

onde $\alpha_0 = \delta_0$ e $\alpha_i = -\sum_{j=i+1}^{s} \delta_i,\ i = 1, 2, ..., s-1$, com as mesmas vantagens que no caso anterior, mas aparecendo agora o multiplicador de longo prazo como coeficiente de x_{t-1}.

Finalmente, refira-se que a aplicação deste tipo de reparametrização a um modelo com várias variáveis (assumidas como) estritamente exógenas também parece ser recomendável. De facto, em geral, embora as variáveis em níveis (não diferenciadas) se encontrem muito fortemente correlacionadas, o mesmo não acontece com as suas primeiras diferenças. A utilização de P1 ou de P2 sobre cada um dos polinómios em L permitirá, em princípio, mitigar os problemas de colinearidade, uma vez que só o nível corrente de cada uma das variáveis, no caso de P1, ou o respectivo valor desfasado um período, no de P2, permanecerão no modelo; todos os restantes regressores serão obtidos por diferenciação das variáveis em nível.

2 O Modelo ADL(1,1) e o Modelo de Correcção de Erros

A abordagem desta secção visa essencialmente dois objectivos. Em primeiro lugar, visa-se introduzir os modelos ADL (*autoregressive distributed lag*) como instrumentos importantes da modelização econométrica causal de séries temporais, estudando um dos seus casos particulares mais simples. Em segundo lugar, trata-se de apresentar o modelo de correcção de erros (MCE), preparando o caminho para o estudo a empreender mais adiante. De resto, como se verá posteriormente, este tipo de modelo será privilegiado na análise empírica multivariada de séries temporais integradas.

2.1 A generalidade do modelo ADL(1,1)

Este ponto visa salientar a generalidade dos modelos ADL, considerando como exemplo um dos seus representantes mais simples, o modelo ADL(1,1), incorporando observações de uma única variável (assumida como) exógena:

$$y_t = \mu + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t, \tag{7}$$

onde $\epsilon_t \sim \mathrm{iid}(0, \sigma^2)$ e para o qual se assume a usual condição de estabilidade, $|\alpha_1| < 1$. O estudo deste modelo reveste-se de particular importância porque:

1) muitos modelos são seus casos particulares e, por conseguinte, em vez de os tomar como ponto de partida no trabalho de especificação, parece ser mais

correcto iniciar a análise com o modelo geral, o ADL(1,1), e deixar que os dados escolham o modelo mais adequado (através dos testes das restrições que permitem obter os modelos particulares a partir do modelo geral);

2) os resultados determinados para o ADL(1,1) também são válidos, com modificações e/ou adaptações, para o modelo mais geral, ADL(r, s), contendo várias variáveis exógenas.

Para justificar a primeira afirmação, repare-se na diversidade de modelos que se obtêm como casos particulares do modelo ADL(1,1) impondo restrições sobre os seus parâmetros [5]:

1) a equação de regressão (simples) estática (de Econometria I), $y_t = \mu + \beta_0 x_t + \epsilon_t$, obtém-se impondo as restrições $\alpha_1 = \beta_1 = 0$;

2) o modelo simples autoregressivo de primeira ordem (AR(1)) obtém-se facilmente com $\beta_0 = \beta_1 = 0$;

3) o modelo nas primeiras diferenças das variáveis, $\Delta y_t = \mu + \beta_0 \Delta x_t + \epsilon_t$, resulta das restrições $\beta_1 = -\beta_0$ e $\alpha_1 = 1$ (mas note-se que esta restrição viola a condição de estabilidade);

4) o modelo de "indicador avançado", em que o comportamento de x antecipa o de y com um período de avanço (neste caso), isto é, $y_t = \mu + \beta_1 x_{t-1} + \epsilon_t$, obtém-se através das restrições $\alpha_1 = \beta_0 = 0$;

5) o (outrora popular) modelo de ajustamento parcial, $y_t = \mu + \alpha_1 y_{t-1} + \beta_0 x_t + \epsilon_t$, resulta simplesmente de $\beta_1 = 0$;

6) o modelo de regressão estático com erros AR(1) (um modelo que é chamado de "factores comuns"), $y_t = \theta + \beta_0 x_t + u_t,\ u_t = \alpha_1 u_{t-1} + \epsilon_t$, obtém-se impondo as restrições não lineares $\mu = \theta(1 - \alpha_1)$ e $\beta_1 = -\alpha_1 \beta_0$;

7) o modelo DL(1), $y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$, resulta de (7) quando $\alpha_1 = 0$;

8) o modelo que Hendry designa de "dead start", e cuja parte sistemática ou principal se identifica com a do modelo de expectativas adaptativas, $y_t = \mu + \alpha_1 y_{t-1} + \beta_1 x_{t-1} + \epsilon_t$, resulta de $\beta_0 = 0$.

[5] Esta apresentação baseia-se exclusivamente em Hendry [1995].

No capítulo 7 de Hendry [1995] são investigadas as propriedades de todos estes modelos quando o "verdadeiro modelo" (o PGD, processo de geração de dados) é o modelo ADL(1,1). As consequências são as já esperadas: inconsistência dos estimadores e sintomas diversos de má especificação (autocorrelação residual, instabilidade dos parâmetros, etc.).

Todavia, como veremos mais adiante, a parametrização mais importante do ADL(1,1) é a de modelo de correcção de erros, dada por

$$\Delta y_t = (\alpha_1 - 1)(y_{t-1} - \lambda_0 - \lambda_1 x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t,$$

onde $\lambda_0 = \mu/(1-\alpha_1)$ e $\lambda_1 = (\beta_0 + \beta_1)/(1-\alpha_1)$, mas onde não se impõe qualquer restrição. Se, contudo, se impuser a restrição $\alpha_1 + \beta_0 + \beta_1 = 1$, tem-se o "MCE homogéneo", dado por

9) $\Delta y_t = (\alpha_1 - 1)(y_{t-1} - \lambda_0 - x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t.$

2.2 Transformações lineares do modelo ADL(1,1)

Para se compreender melhor as vantagens de algumas transformações lineares ou reparametrizações (sem a imposição de restrições) do modelo ADL(1,1), considere-se uma situação de equilíbrio de estado estacionário, em que deixou de haver qualquer variação das variáveis, encontrando-se estas nos seus valores médios (representados com um asterisco). Isto é, de (7) resulta

$$y^* = \mu + \alpha_1 y^* + \beta_0 x^* + \beta_1 x^*,$$

ou seja, resolvendo em ordem a y^* obtém-se a solução de equilíbrio (de longo prazo):

$$y^* = \frac{\mu}{1-\alpha_1} + \frac{\beta_0 + \beta_1}{1-\alpha_1} x^*,$$

ou

$$y^* = \lambda_0 + \lambda_1 x^*$$

onde $\lambda_0 = \mu/(1-\alpha_1)$ e $\lambda_1 = (\beta_0+\beta_1)/(1-\alpha_1)$ representa o multiplicador total, ou de longo prazo ou de equilíbrio [6]. Estamos agora em condições de proceder às referidas reparametrizações.

---

[6] Recorde-se que este multiplicador também se obtém facilmente com $B(1)/A(1)$, com $A(L) = 1 - \alpha_1 L$ e $B(L) = \beta_0 + \beta_1 L$ neste caso.

Em primeiro lugar, subtraindo y_{t-1} a ambos os membros de (7) obtém-se

$$\Delta y_t = \mu + (\alpha_1 - 1)y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t, \tag{8}$$

parametrização que não oferece nenhuma vantagem especial. Todavia, somando e subtraindo $\beta_0 x_{t-1}$ ao lado direito, obtém-se a chamada *"forma de Bardsen"*:

$$\Delta y_t = \mu + (\alpha_1 - 1)y_{t-1} + \beta_0 \Delta x_t + (\beta_0 + \beta_1)x_{t-1} + \epsilon_t, \tag{9}$$

e note-se que, para obter a estimativa do multiplicador de longo prazo, basta tomar o simétrico do quociente da estimativa do coeficiente de x_{t-1} pela do coeficiente de y_{t-1}.

Se, por outro lado, somarmos e subtrairmos $\beta_1 x_t$ ao lado direito de (8), obtemos

$$\Delta y_t = \mu + (\alpha_1 - 1)y_{t-1} - \beta_1 \Delta x_t + (\beta_0 + \beta_1)x_t + \epsilon_t, \tag{10}$$

onde, para obter a estimativa do multiplicador de longo prazo, basta dividir o simétrico da estimativa do coeficiente de x_t pela do coeficiente de y_{t-1}.

A forma mais interessante (a forma MCE) obtém-se de (9), colocando $(\alpha_1 - 1)$ em evidência:

$$\Delta y_t = (\alpha_1 - 1)[y_{t-1} - \frac{\mu}{1-\alpha_1} - \frac{\beta_0 + \beta_1}{1-\alpha_1}x_{t-1}] + \beta_0 \Delta x_t + \epsilon_t, \tag{11}$$

ou seja,

$$\Delta y_t = (\alpha_1 - 1)[y_{t-1} - \lambda_0 - \lambda_1 x_{t-1}] + \beta_0 \Delta x_t + \epsilon_t, \tag{12}$$

equação que será retomada mais adiante. Por outro lado, a "forma homogénea" (por parecer que o multiplicador de longo prazo é unitário) é dada por

$$\Delta y_t = \mu + (\alpha_1 - 1)[y_{t-1} - x_{t-1}] + \beta_0 \Delta x_t + (\beta_0 + \beta_1 + \alpha_1 - 1)x_{t-1} + \epsilon_t, \tag{13}$$

onde se nota o termo adicional em x_{t-1} para "quebrar a homogeneidade".

Finalmente, a chamada "forma de Bewley", obtém-se subtraindo $\alpha_1 y_t$ a ambos os membros do ADL(1,1), somando e subtraindo $\beta_1 x_t$ no lado direito e fazendo $\gamma = (1-\alpha_1)^{-1}$:

$$y_t = \gamma\mu - \gamma\alpha_1 \Delta y_t + \gamma(\beta_0 + \beta_1)x_t - \gamma\beta_1 \Delta x_t + \gamma\epsilon_t, \tag{14}$$

onde o multiplicador de longo prazo é dado imediatamente pelo coeficiente de x_t. Todavia, note-se que o estimador OLS desta equação é inconsistente. (Porquê?)

Antes de avançar para o estudo do MCE note-se que, como procedemos apenas a simples transformações lineares, sem a imposição de qualquer restrição, todas estas reparametrizações do ADL(1,1) são equivalentes, fornecendo o método dos mínimos quadrados as mesmas estimativas para os parâmetros (excepto a forma de Bewley) e possuindo todas elas a mesma capacidade para explicar os dados.

2.3 O modelo de correcção de erros

Retome-se a equação (12) (ou (11)) e repare-se que o termo dentro de parêntesis rectos mede o erro ou desvio de equilíbrio ou desequilíbrio no período anterior, isto é, a extensão em que o equilíbrio não foi satisfeito nesse período. Assim, essa equação dá-nos uma relação dinâmica de curto prazo (uma vez que a variável dependente representa a variação de y em cada período), em que Δy_t é "conduzido" pela variação de x (Δx_t) e pelo desequilíbrio do período anterior. Por outras palavras, a relação de equilíbrio (de longo prazo) também é importante para explicar a evolução de curto prazo da variável y. Assim, mesmo que $\epsilon_t = 0$ e $\Delta x_t = 0$ durante um período de tempo prolongado, Δy_t não se anulará até que a solução de equilíbrio seja satisfeita, isto é, até que $y = \lambda_0 + \lambda_1 x$.

O coeficiente $\alpha_1 - 1$ é um indicador da velocidade de ajustamento de curto prazo (para o equilíbrio) [7] e mede a proporção do erro de equilíbrio que se reflecte em y no período seguinte. Aliás, uma das vantagens da parametrização MCE do ADL(1,1) é que esse coeficiente aparece de forma explícita. O termo $(\alpha_1 - 1)[y_{t-1} - \lambda_0 - \lambda_1 x_{t-1}]$ é usualmente designado de termo de correcção de erro e a expressão dentro de parêntesis rectos, nos níveis das variáveis no período anterior (que pode não conter o coeficiente λ_0), é por vezes chamada de mecanismo de correcção de erros [8].

Para se perceber a designação atribuída ao modelo (de correcção de erros), comece-se por atentar que a satisfação da condição de estabilidade ($|\alpha_1| < 1$) garante que $-2 < \alpha_1 - 1 < 0$, isto é, que o coeficiente de ajustamento é negativo (mas maior que -2). Suponha-se agora que, por exemplo, no período $t-1$ o erro de equilíbrio foi positivo, isto é, que $y_{t-1} > \lambda_0 + \lambda_1 x_{t-1}$ [9]. Então, como $\alpha_1 - 1 < 0$, *ceteris paribus*, no período seguinte (t) y terá tendência para retornar ao valor de equilíbrio pois

[7] O seu papel é semelhante ao do parâmetro de ajustamento no modelo de ajustamento parcial.

[8] E daí por vezes o modelo ser chamado de modelo com mecanismo de correcção de erros.

[9] Supondo que y representa o consumo privado e x o rendimento, corresponde a pensar-se que as famílias terão consumido para além do nível de equilíbrio permitido pelo rendimento, sendo obrigadas a recorrer ao endividamento.

$\Delta y_t < 0$ [10]. Ou seja, o modelo incorpora um efeito de *feedback* negativo que tenta corrigir desequilíbrios de períodos anteriores para alcançar a relação de equilíbrio (de longo prazo). Assim, a equação de correcção de erros é basicamente uma equação de ajustamento dinâmico de curto prazo, mas em que o ajustamento também é guiado pela relação de equilíbrio (de longo prazo). Note-se, então, que para testar a existência de um modelo (ou mecanismo) de correcção de erros, efectua-se o teste de

$$H_0 : \alpha_1 - 1 = 0 \ \ vs. \ \ H_1 : \alpha_1 - 1 < 0,$$

correspondendo a rejeição da hipótese nula a evidência no sentido da presença desse tipo de efeito (e note-se que este teste também pode ser visto como testando uma condição necessária mas não suficiente de estabilidade).

Relativamente à estimação, se o multiplicador de longo prazo (e λ_0) for(em) conhecidos(s), isto é, dado(s) pela teoria económica, a estimação de (12) pode ser efectuada imediatamente pelo OLS. Em geral, contudo, essa informação não se encontrará disponível. Então, reparando na forma como os parâmetros aparecem em (12) ou em (11), parece ser necessário recorrer à estimação pelo método dos mínimos quadrados não lineares (NLLS). Todavia, qualquer uma das parametrizações do ADL(1,1) poderá fornecer as estimativas pretendidas (incluindo a forma de Bewley, estimada com o método IV). A forma privilegiada será, contudo, a de Bardsen:

1) por fornecer imediatamente uma estimativa para o coeficiente de ajustamento de curto prazo;

2) pelo facto de o rácio-t para esse coeficiente permitir testar imediatamente a presença do efeito de correcção de erros;

3) por permitir obter facilmente uma estimativa para o multiplicador de longo prazo através da forma descrita atrás e

4) por ser preferível, por exemplo, ao ADL(1,1) devido ao facto de os problemas de colinearidade entre os regressores virem atenuados.

Para terminar, refira-se que os modelos de correcção de erros têm obtido grande sucesso empírico na modelização de vários fenómenos macroeconómicos observados

[10] Ou seja, *ceteris paribus*, o consumo das famílias terá tendência a diminuir, regressando ao nível de equilíbrio, pois elas terão que usar parte do seu rendimento para amortizar as dívidas contraídas anteriormente.

ao longo do tempo, constituindo os exemplos mais importantes os casos das funções consumo e de procura de moeda. De resto, como veremos na última secção, quando trabalharmos com séries temporais não estacionárias, seremos frequentemente conduzidos a trabalhar com MCEs, uma vez que estes se encontram estreitamente ligados com a teoria da cointegração.

3 O Modelo ADL(r, s), o MCE e a Modelização do Geral para o Particular

Nesta secção generaliza-se a reparametrização em modelos de correcção de erros (MCE) dos modelos ADL(r, s), anteriormente abordada apenas para o caso particular do ADL(1,1) e, em seguida, fornecem-se alguns argumentos que favorecem a utilização da estratégia de modelização do geral para o particular (em detrimento da oposta).

3.1 A reparametrização em MCE dos modelos ADL(r, s)

[11]

Inicia-se este ponto com um resultado preliminar que não é mais que um corolário de P2 do Anexo 2, obtido para o caso em que $a_0 = 1$.

P3 Considere-se um polinómio de ordem p em L, $a(L) = \sum_{j=0}^{p} a_j L^j$ com $a_0 = 1$, isto é, $a(L) = 1 - a_1 L - a_2 L^2 - ... - a_p L^p$ e definam-se os coeficientes $f_0 = a_0 = 1$, $f_i = -\sum_{j=i+1}^{p} a_j$, $i = 1, ..., p-1$ e $f_p = 0$. Então, pode escrever-se:

$$a(L) = a(1)L + (1 - L)(1 - \sum_{i=1}^{p-1} f_i L^i), \tag{15}$$

ou seja,

$$a(L) = a(1)L + (1 - L)f(L), \tag{16}$$

onde $f(L)$ é um polinómio de ordem $p - 1$ em L.

[11] Este ponto inspira-se sobretudo em Boswijk, H. P. [1994].

Considere-se agora um modelo ADL(r, s) incorporando observações de uma única variável (assumida como) exógena e sem termo independente (uma vez que, para os efeitos da análise que se segue, a presença deste é irrelevante):

$$y_t = \sum_{i=1}^{r} \alpha_i y_{t-i} + \sum_{i=0}^{s} \beta_i x_{t-i} + \epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma^2), \tag{17}$$

ou seja,

$$A(L) y_t = B(L) x_t + \epsilon_t, \tag{18}$$

com $A(L) = 1 - \sum_{i=1}^{r} \alpha_i L^i$ e $B(L) = \sum_{i=0}^{s} \beta_i L^i$.

Usando P3, podemos escrever

$$A(L) = A(1)L + (1 - L)D(L)$$

onde

$$D(L) = 1 - \sum_{i=1}^{r-1} \delta_i L^i = 1 - \sum_{i=1}^{r-1} (- \sum_{j=i+1}^{r} \alpha_j) L^i.$$

Por outro lado, usando P2, tem-se

$$B(L) = B(1)L + (1 - L)G(L),$$

onde

$$G(L) = \sum_{i=0}^{s-1} \gamma_i L^i = \gamma_0 + \sum_{i=0}^{s-1} (- \sum_{j=i+1}^{s} \beta_j) L^i.$$

Empregando estas duas decomposições polinomiais em (18) obtém-se

$$[A(1)L + (1 - L)D(L)] y_t = [B(1)L + (1 - L)G(L)] x_t + \epsilon_t,$$

isto é,

$$A(1) y_{t-1} + (1 - \sum_{i=1}^{r-1} \delta_i L^i) \Delta y_t = B(1) x_{t-1} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t,$$

da qual se obtém facilmente a forma de Bardsen:

$$\Delta y_t = -A(1) y_{t-1} + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + B(1) x_{t-1} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t. \tag{19}$$

Finalmente, a representação MCE obtém-se colocando $-A(1) = \phi$ em evidência:

$$\Delta y_t = \phi [y_{t-1} - \frac{B(1)}{A(1)} x_{t-1}] + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t, \tag{20}$$

ou seja, representando com $\lambda = B(1)/A(1)$ o multiplicador de longo prazo, tem-se

$$\Delta y_t = \phi[y_{t-1} - \lambda x_{t-1}] + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t, \tag{21}$$

onde, saliente-se, o parâmetro ϕ é negativo (caso contrário não se terá um MCE) desde que a condição necessária (embora não suficiente) de estabilidade seja satisfeita (isto é, desde que $\sum_{i=1}^{r} \alpha_i < 1$, ou seja, $A(1) > 0$).

Note-se que, tal como no caso do ADL(1,1), também aqui a forma usualmente empregue para a estimação (pelo OLS) é a de Bardsen, dada pela equação (19), obtendo-se facilmente a estimativa do multiplicador de longo prazo a partir de $\lambda = B(1)/A(1) = -B(1)/\phi$.

Repare-se de novo nas vantagens de (19), onde se tem um ADL$(r-1, s-1)$ nas primeiras diferenças das variáveis, aumentado com os seus níveis desfasados um período como regressores, sobre (18):

1) permite obter imediatamente uma estimativa do coeficiente de ajustamento para o equilíbrio (desde que este exista), $-A(1) = \phi$, ou coeficiente do termo de correcção de erro;

2) o rácio-t desse coeficiente pode ser usado para testar a presença do efeito de correcção de erros ($H_0 : \phi = 0$ *vs.* $H_1 : \phi < 0$), ou seja, a condição necessária de estabilidade;

3) permite obter facilmente uma estimativa para o multiplicador de longo prazo e

4) permite atenuar os problemas decorrentes da colinearidade entre regressores, quase certamente presentes no modelo inicial.

A extensão ao caso de várias (k) variáveis exógenas é simples. Agora, x_t passa a representar um vector, $k \times 1$, de observações dessas variáveis ($x_t' = [x_{t1}\ x_{t2}\ ...\ x_{tk}]$), cujos comprimentos de desfasamento são representados com $s_1, s_2, ..., s_k$ (isto é, note-se que não é necessário que os desfasamentos máximos sejam todos iguais). Continuamos, contudo, a referir-nos a um modelo ADL(r, s), com $s = \max\{s_1, s_2, ..., s_k\}$. Em vez de (17) temos

$$y_t = \sum_{i=1}^{r} \alpha_i y_{t-i} + \sum_{i=0}^{s} \beta_i' x_{t-i} + \epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma^2), \tag{22}$$

onde os β_i's são vectores, $k \times 1$, de coeficientes de desfasamento, isto é,

$$\beta_i' = [\beta_{i1}\ \beta_{i2}\ ...\ \beta_{ik}],\ i = 0, 1, ...s.$$

Por seu turno, o análogo de (18) escreve-se agora

$$A(L)y_t = B(L)'x_t + \epsilon_t, \tag{23}$$

onde $B(L) = \sum_{i=0}^{s} \beta_i L^i$ representa agora um vector, $k \times 1$, de polinómios em L.

Procedendo de forma semelhante à anterior, chega-se à forma de Bardsen:

$$\Delta y_t = -A(1)y_{t-1} + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + B(1)'x_{t-1} + \sum_{i=0}^{s-1} \gamma_i' \Delta x_{t-i} + \epsilon_t, \tag{24}$$

onde $B(1)'$ e γ_i' representam vectores, $1 \times k$, de coeficientes. Finalmente, representando com λ o vector, $k \times 1$, de multiplicadores de longo prazo, $\lambda' = A(1)^{-1}B(1)' = -\phi^{-1}B(1)'$, tem-se a representação em MCE:

$$\Delta y_t = \phi[y_{t-1} - \lambda' x_{t-1}] + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + \sum_{i=0}^{s-1} \gamma_i' \Delta x_{t-i} + \epsilon_t. \tag{25}$$

Note-se que, no caso de o modelo ser estável, a solução de equilíbrio (de longo prazo) é dada por

$$y^* = \lambda' x^* = \lambda_1 x_1^* + \lambda_2 x_2^* + \ ...\ + \lambda_k x_k^*,$$

onde o "*" tem o mesmo significado que anteriormente. Note-se ainda que, neste caso, dada a presença de várias variáveis exógenas, as vantagens da reparametrização sob a forma de Bardsen em termos de atenuação dos problemas de colinearidade tendem a ser ainda maiores que no caso anterior.

3.2 A modelização do geral para o particular

Actualmente considera-se que a melhor estratégia para chegar a um modelo dinâmico adequado consiste em iniciar o trabalho de especificação considerando um modelo bastante geral — um ADL(r, s) de ordem elevada, de preferência na forma reparametrizada que vimos anteriormente —, e prosseguir "testando para baixo", até obter um modelo mais simples mas satisfatório sob vários pontos de vista. É esta estratégia de modelização, que é preconizada sobretudo por investigadores britânicos (como D. Hendry e G. Mizon), que é conhecida como modelização "do geral para o particular" (ou "para o específico"). As principais características desta estratégia são descritas em seguida.

1. Desde o início não há qualquer intenção ou objectivo de identificar um modelo simples ou parcimonioso, ou seja, há uma sobreparametrização assumida e deliberada. Esta é uma das características que tem recebido mais críticas pois, muito frequentemente, os modelos mais simples podem ser considerados preferíveis (por exemplo, para efectuar previsão ou porque são mais fáceis de compreender).

2. Desde a primeira especificação e em cada fase do processo de simplificação os modelos são submetidos a uma bateria de testes para detecção de erros de especificação. A saber, testes de heterocedasticidade e de autocorrelação dos erros, RESET, de instabilidade de estrutura, de previsão, etc. [12].

3. A teoria económica empregue para a especificação é entendida em sentido lato e as restrições que ela implica são sistematicamente testadas sobre os dados, isto é, ela só é usada para indicar que variáveis estão envolvidas e qual é a forma geral da relação de equilíbrio de longo prazo mas não, por exemplo, a forma da estrutura dinâmica de curto prazo. Esta é determinada pelos dados, que podem ou não aceitar as restrições decorrentes da teoria económica, e o mesmo acontece com quaisquer restrições respeitantes à solução de equilíbrio.

Um exemplo simples permite ilustrar a lógica empregue. Pergunta-se: porquê iniciar o trabalho de especificação com um modelo de ajustamento parcial se esse é um caso particular do ADL(1,1)? Que sentido tem impôr restrições não testadas previamente? Logicamente, parece ter mais sentido iniciar o estudo com o ADL(1,1) e considerar apenas o modelo de ajustamento parcial se as restrições que ele implica não forem rejeitadas pelos dados (e se ele não apresentar sintomas de má especificação). Assim, visa-se deixar que os dados "falem livremente", impondo apenas as restrições que eles não rejeitem ou, até, que eles próprios sugiram (por exemplo, excluindo regressores com coeficientes não significativos).

Em termos simples, esta estratégia envolve as seguintes etapas ou fases:

1. Especificação inicial de um modelo dinâmico com ordens de desfasamento elevadas (um ADL(r, s), de preferência reparametrizado sob a forma de Bardsen),

[12] Este "etc." refere-se a outros testes que não foram estudados, como os testes de normalidade dos erros ou os de detecção de efeitos ARCH (heterocedasticidade condicional autoregressiva) nos erros.

que seja consistente com a relação de equilíbrio dada pela teoria económica e que não imponha restrições sobre a dinâmica de curto prazo.

2. Simplificação do modelo excluindo regressores não significativos ou impondo outras restrições que não violem os dados e que não provoquem o aparecimento de sintomas de erros de especificação. Em geral, são os coeficientes de desfasamento de ordens mais elevadas que tenderão a ser os mais pequenos, pelo que se começa por testar a sua significância, só em seguida se "descendo" para os desfasamentos de ordem mais baixa. Para além dos testes estatísticos estudados, instrumentos importantes também nesta fase são as estatísticas de "bondade de ajustamento" ($\bar{R}^2$, $\hat{\sigma}_u^2$, AIC, SBIC, etc.), pois o processo de simplificação não deverá permitir que esta se deteriore substancialmente [13].

3. Avaliação final do modelo seleccionado com base na teoria económica e nos "testes de má especificação" (*misspecification tests*), mas usando também testes para hipóteses não encaixadas contra especificações rivais sugeridas na literatura (recorde-se o que se estudou sobre o "princípio da envolvência").

3.3 A abordagem do particular para o geral

A estratégia de modelização do particular para o geral é a abordagem tradicional que predominou até finais da década de 70, mas que continua ainda hoje na base da maior parte dos livros de texto de Econometria. Segundo uma perspectiva algo caricatural, para esta abordagem o modelo econométrico é uma especificação bastante restrita dada pela, ou melhor, por uma teoria económica, e esse modelo é o "verdadeiro modelo" (ou PGD). Ou seja, não há problemas de especificação: esta é assumida como conhecida desde o princípio, pelo que o papel do econometrista se limita à estimação com base em métodos eficientes, procurando "remediar" os "problemas" que surgem (autocorrelação, heterocedasticidade, sinais dos coeficientes incorrectos, R^2 baixo, etc.). Outro investigador, usando outra teoria mas o mesmo conjunto de dados e a mesma metodologia pode chegar a um modelo radicalmente diferente, no qual se baseará para afirmar que encontrou evidência empírica que a suporta. Assim, a Econometria seria apenas um instrumento que serve para estimar e

[13] Outro instrumento não estudado mas também empregue nesta fase consiste na chamada "análise COMFAC", de testes de restrições de factores comuns nos polinómios de desfasamento das várias variáveis.

confirmar ou corroborar as relações dadas pela teoria económica, mas não para a pôr verdadeiramente à prova, testando as suas implicações mais fortes sobre os dados.

As principais críticas que são usualmente apontadas a esta metodologia são as seguintes:

a) Iniciando a investigação empírica pelo modelo simples, cada teste de hipóteses está condicionado por hipóteses iniciais arbitrárias que, se não forem válidas, contaminam todo o processo de especificação. Por exemplo, suponha-se que se inicia a análise por um modelo subparametrizado (M_1), o qual, por esse facto, apresenta sintomas de autocorrelação residual. A "correcção da autocorrelação" (como é vulgar, através de um método de estimação admitindo erros AR(1)), conduz-nos ao modelo M_2 (mais geral). Por sua vez, este apresenta problemas de instabilidade dos coeficientes, o que nos conduz ao modelo M_3 (por exemplo, introduzindo variáveis artificiais). Mas uma vez que também se descobriram problemas em M_2, que valor tem a descoberta do primeiro problema que nos levou até ele? Como nota Hendry [1995, p. 270], em cada passo do processo de generalização as conclusões são potencialmente erradas, pois em fases posteriores poderemos descobrir novos problemas. Logo, "todo o estudo colapsa". Ou seja, esta estratégia é, muito frequentemente, ineficiente.

b) Mais geralmente, volta a salientar-se que os testes estatísticos usuais não são válidos em modelos com variáveis omitidas (como tenderão a ser os modelos simples, empregues como ponto de partida). Assim, iniciar a investigação com um modelo suficientemente geral permitirá, em princípio, evitar as armadilhas de inferência e as incoerências lógicas.

c) Não é possível controlar o nível ou dimensão global real da sequência de testes do particular para o geral. Pelo contrário, numa sequência de testes encaixados, do geral para o particular, esse nível global pode ser controlado [14].

d) Além de ineficiente, a estratégia do simples para o geral poderá até ocultar a descoberta de um modelo adequado. O exemplo mais simples desta possibilidade é o da estimação de um modelo estático com erros AR(1) quando, na realidade, o "modelo verdadeiro" é dado por um ADL(1,1). De facto, pode mostrar-se que

[14] Pode provar-se que o verdadeiro nível ou dimensão do j-ésimo teste, numa sequência de testes encaixados, é dado por $\alpha^* = 1 - (1-\alpha_1)(1-\alpha_2)...(1-\alpha_j)$, onde os α_i, $i = 1, 2, ..., j$ representam os níveis nominais individuais.

a estimação de um modelo estático com erros AR(1), tão usada no passado, equivale a estimar o modelo ADL(1,1) mas impondo certas restrições sobre os parâmetros sem sequer testar a sua validade previamente. Se essas restrições não forem válidas, não só se acaba por ficar com um modelo mal especificado como, ainda, o estimador empregue é inconsistente. Pelo contrário, a estratégia oposta tem obtido algum sucesso empírico na modelização de alguns fenómenos macroeconómicos [15].

4 Raízes Unitárias: uma Introdução

Na última quinzena de anos, a metodologia econométrica para as séries temporais sofreu uma revolução com a descoberta de que, ao contrário do que se supunha, muitas variáveis económicas observadas ao longo do tempo não parecem corresponder a processos estocásticos estacionários (em tendência). Pelo contrário, tais séries temporais são, frequentemente, melhor aproximadas por modelos autoregressivos não estacionários mas integrados, isto é, com uma raiz unitária no seu polinómio autoregressivo.

Sabe-se agora que a presença das raízes unitárias autoregressivas pode criar grandes problemas ao trabalho empírico mas que, por outro lado, também veio abrir novas oportunidades e pistas de investigação para esse trabalho. Assim, por um lado, se é verdade que em muitas regressões com séries económicas os estimadores OLS nem sequer são consistentes para os verdadeiros valores dos parâmetros, noutras, pelo contrário, eles convergem tão rapidamente para esses valores que, em certas fases da modelização, os podem substituir sem perda significativa da qualidade dos métodos de inferência. Todavia, geralmente, em ambos os casos a teoria assimptótica habitual dos testes de hipóteses não é válida, o que constitui um obstáculo importante à realização de inferências.

As notas que se seguem constituem uma introdução muito simples e breve a estes problemas. Os leitores interessados poderão prosseguir o estudo destas matérias com base na leitura dos principais textos que lhes serviram de base, a saber: Davidson, R. e MacKinnon, J. G. [1993], cap. 20, Boswijk, H. P. [1994] e Campbell, J. Y. e Perron, P. [1991].

[15] Os exemplos mais popularizados são os do modelo DHSY, da função consumo para o Reino Unido, e os de modelos para a função de procura de moeda.

4.1 Introdução: conceitos básicos e motivações

Nesta subsecção apresentam-se apenas alguns conceitos básicos e tenta-se motivar o leitor para a importância do estudo do tema. Os testes de raízes unitárias só serão abordados na subsecção seguinte.

4.1.1 Processos estacionários e processos integrados

Começa por recordar-se que as variáveis de um processo estocástico estacionário (em covariância ou de segunda ordem ou em sentido fraco) têm média e variância (finitas e) constantes ao longo do tempo e têm covariâncias que só dependem da distância no tempo a que encontram. Para contrastar o comportamento das séries [16] estacionárias com o das não estacionárias mas integradas de primeira ordem vai considerar-se o exemplo do modelo AR(1):

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t, \; \epsilon_t \sim \text{iid}(0, \sigma_\epsilon^2).$$

No caso em que o processo é estacionário ($|\rho| < 1$), repare-se que ele incorpora um mecanismo de correcção de erros (MCE). De facto, subtraindo y_{t-1} a ambos os membros tem-se

$$\Delta y_t = \alpha + (\rho - 1)y_{t-1} + \epsilon_t,$$

ou seja,

$$\Delta y_t = (\rho - 1)(y_{t-1} - \mu_y) + \epsilon_t, \tag{26}$$

onde $\mu_y = \alpha/(1-\rho)$. Ora, como $-2 < \rho - 1 < 0$ (porque $|\rho| < 1$), a equação (1) contém de facto um processo de ajustamento com mecanismo de correcção de erros, com μ_y a representar a solução de equilíbrio: se, por exemplo, $y_{t-1} - \mu_y < 0$, no período t, *ceteris paribus* deverá observar-se $\Delta y_t > 0$, de modo a que o equilíbrio seja restaurado. Dado o efeito de ϵ_t, pode acontecer que y_t se mantenha acima ou abaixo da sua média durante algum tempo. Todavia, a tendência é para que y_t acabe por regressar à situação de equilíbrio, isto é, à sua média. É este comportamento que é típico de qualquer série estacionária (e não apenas do AR(1) estacionário) que é chamado de comportamento de reversão ou de regressão para a média (*mean reversion* ou *mean regression*).

[16] De forma pouco rigorosa, usar-se-ão indistintamente os termos “processo estocástico" e “série temporal".

O caso em que $|\rho| > 1$ é pouco interessante sob o ponto de vista económico e é facilmente detectável com base em meios gráficos pois o processo é explosivo; por exemplo, com $\rho > 1$ o processo cresce de forma explosiva.

Por outro lado, o caso de $\rho = -1$ também é pouco interessante para a generalidade das séries económicas [17]: a série tenderia a oscilar em torno de dois valores simétricos. Note-se, contudo, que este é um dos casos em que o polinómio autoregressivo tem uma raiz de módulo unitário ($\Phi(z) = 0 \Leftrightarrow 1 + z = 0 \Leftrightarrow z = -1$). Simplesmente, esta raiz unitária raramente tem relevância económica.

O caso que parece caracterizar (aproximadamente) algumas séries económicas é o de $\rho = 1$, vulgarmente conhecido como *passeio aleatório*. Comece-se por analisar o passeio aleatório sem deriva (*drift*), isto é, quando $\alpha = 0$:

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

Por substituição recursiva e dado um valor inicial para o processo, y_0, tem-se:

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_i \tag{27}$$

onde o último termo, a acumulação de todos os choques aleatórios, é vulgarmente chamado de *tendência estocástica*, pretendendo-se com esta designação salientar que o comportamento de longo prazo do processo muda de forma lenta e aleatória, afectando também as previsões. De facto, pode provar-se que a previsão univariada óptima (em erro quadrático médio) de y_{t+s} formulada no final do período t é dada por

$$\hat{y}_{t+s|t} = E_t(y_{t+s}) = E(y_{t+s}|y_t, y_{t-1}, ...) = y_t.$$

Todavia, no final do período $t+s-1$, a previsão óptima de y_{t+s} passa a ser dada por

$$\hat{y}_{t+s|t+s-1} = E_{t+s-1}(y_{t+s}) = E(y_{t+s}|y_{t+s-1}, y_{t+s-2}, ...) = y_{t+s-1} = y_t + \epsilon_{t+1} + ... + \epsilon_{t+s-1},$$

isto é, a previsão óptima mudou de forma estocástica, com os efeitos dos choques aleatórios ou inovações (os ϵ's) ocorridos entre $t+1$ e $t+s-1$ a repercutirem-se integralmente na previsão. Note-se também que o valor esperado condicional não é constante ao longo do tempo. Por outro lado, note-se ainda que de (27), e assumindo que o valor inicial é fixo, tem-se

$$\text{Var}[y_t] = t\sigma_\epsilon^2,$$

[17] A não ser nalguns casos, de variáveis com sazonalidade.

isto é, o processo não tem variância (com esta a crescer linearmente com o tempo).

Por outro lado, para o caso do passeio aleatório com deriva,

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

o procedimento de substituição recursiva permite obter

$$y_t = y_0 + \alpha t + \sum_{i=1}^{t} \epsilon_i \tag{28}$$

onde, para além da tendência estocástica, se observa agora também a presença de uma tendência determinística (αt). Neste caso, nem sequer a média condicional em y_0 é constante, pois cresce linearmente com o tempo:

$$E(y_t|y_0) = y_0 + \alpha t.$$

Repare-se agora que para ambos os tipos de passeios aleatórios se nota a ausência de mecanismo de correcção de erros (pois $\rho = 1$). Para o caso sem deriva tem-se

$$\Delta y_t = \epsilon_t, \tag{29}$$

enquanto que quando $\alpha \neq 0$ se obtém

$$\Delta y_t = \alpha + \epsilon_t. \tag{30}$$

Ou seja, tanto num caso como no outro as séries não são atraídas por um qualquer valor fixo (que desempenhe o papel que a média tem no caso dos processos estacionários), isto é, não apresentam o comportamento de reversão para a média. Todavia, estas equações também revelam que, embora não estacionários, ambos os processos se tornam estacionários quando diferenciados uma vez. Por esse facto, são chamados de *integrados de primeira ordem*, empregando-se a notação $y_t \sim I(1)$. Na realidade, o passeio aleatório é o exemplo mais simples de um processo I(1).

Mais geralmente diz-se que um processo estocástico, y_t, sem componentes determinísticos, é integrado de ordem d, $y_t \sim I(d)$, quando só após ser diferenciado d vezes se torna estacionário (não o sendo ainda após $d-1$ diferenciações). Assim, emprega-se a notação $x_t \sim I(0)$ para representar um processo que não necessita ser diferenciado nenhuma vez para se tornar estacionário. Note-se ainda que a presença de componentes determinísticos é irrelevante; por exemplo, para saber a ordem de integração da série dada por

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t,$$

a presença do polinómio em t é absolutamente irrelevante: é a natureza de u_t que determina a ordem de integração de y_t.

O sentido da expressão *raiz unitária* também deverá ter ficado claro da discussão anterior: em ambos os passeios aleatórios, o polinómio autoregressivo contém uma raiz unitária (e não apenas em módulo). Mais geralmente se, excluindo os termos determinísticos, o processo y_t se escrever como

$$\Phi(L)y_t = \Theta(L)\epsilon_t,$$

onde $\Phi(L)$ e $\Theta(L)$ representam, respectivamente, os polinómios autoregressivo e de média móveis, e se o primeiro tiver d raízes unitárias, com as restantes fora do círculo unitário, então pode escrever-se

$$(1-L)^d\Gamma(L)y_t = \Theta(L)\epsilon_t,$$

onde todas as raízes de $\Gamma(L)$ se encontram fora do círculo unitário, isto é,

$$\Gamma(L)\Delta^d y_t = \Theta(L)\epsilon_t.$$

Assim, determinar a ordem de integração de uma série é o mesmo que determinar o número de raízes unitárias do seu polinómio autoregressivo, e daí a designação de raízes unitárias autoregressivas e de testes de raízes unitárias (autoregressivas). Em geral, há evidência de que muitas séries económicas são I(1); algumas serão I(0) e muito poucas, apenas algumas séries nominais como as dos índices de preços ou do *stock* de moeda, aparentam ser I(2). Na análise que se segue, só as hipóteses I(1) e I(0) serão consideradas.

Para terminar este ponto, procede-se a uma comparação entre um AR(1) estacionário e um passeio aleatório sem deriva, que recorre sobretudo a representações gráficas. Assim, as séries estacionárias (em torno de zero) como a do gráfico 1:

a) têm um comportamento muito pouco suave ou muito errático, com flutuações bem marcadas;

b) apresentam um comportamento de reversão para a média;

c) como possuem variância (finita e constante), flutuam sempre dentro de um intervalo de amplitude constante e

d) têm memória temporária, pois o efeito dos choques ou inovações vai-se diluindo com o decorrer do tempo; de facto, continuando a tomar o exemplo do AR(1) estacionário, por substituição recursiva obtém-se

$$y_t = \rho^t y_0 + \sum_{i=0}^{t-1} \rho^i \epsilon_{t-i},$$

com o efeito dos choques a diluir-se à medida que o tempo vai decorrendo pois $|\rho| < 1$.

Pelo contrário, as séries I(1) (sem deriva), como a do gráfico 2:

a) têm um comportamento muito mais suave, com flutuações lentas;

b) raramente retornam a qualquer valor, incluindo o valor inicial, isto é, não se sentem atraídas por nenhum valor fixo;

c) como a variância cresce com o tempo, tendem a flutuar dentro de intervalos de amplitude crescente e

d) têm memória permanente, pois os choques têm um efeito altamente persistente, que não se vai diluindo com o tempo, como se pode observar, por exemplo, a partir da equação (27) [18].

4.1.2 Séries económicas: estacionárias em tendência (TSP) ou por diferenciação (DSP)?

Embora útil para o estudo que se vai seguir, a análise empreendida no ponto anterior não é a mais relevante para séries económicas como as do produto, do consumo, do investimento, do índice de preços, do rendimento, etc. . Todas estas séries têm uma característica comum: uma tendência crescente com o tempo. Então, obviamente, como não têm média constante ao longo do tempo, não são estacionárias.

Todavia, dois modelos muito distintos estão disponíveis para aproximar este tipo de comportamento: o modelo (ou processo) de estacionaridade em tendência (TSP, *trend stationary process*) e o de estacionaridade por diferenciação (DSP, *difference stationary process*) ou de raiz unitária (RU).

[18] Esta é uma das características mais contestadas relativamente à hipótese I(1) e, de facto, não parece razoável admitir que todos os choques que a economia sofre a afectem de forma permanente. Contudo, a discussão desta questão afastar-nos-ia dos pontos principais.

Gráfico 1 Série estacionária: $x_t = 0.70x_{t-1} + \epsilon_t,\ \epsilon_t \sim nid(0, \sigma^2)$

Gráfico 2 Série I(1): $y_t = y_{t-1} + \epsilon_t,\ \epsilon_t \sim nid(0, \sigma^2)$ (passeio aleatório sem deriva)

Segundo a hipótese TSP, a série y_t pode exprimir-se como

$$y_t = f(t) + u_t,$$

onde $f(t)$ é uma função do tempo e u_t representa um processo estacionário. Por exemplo, assumindo uma tendência linear determinística, que é a hipótese mais vulgar, ter-se-ia

$$y_t = \beta_0 + \beta_1 t + u_t. \tag{31}$$

Todavia, como as séries económicas têm, em geral, uma taxa de crescimento aproximadamente constante, o modelo de tendência exponencial revela-se, frequentemente, mais adequado:

$$y_t = \exp[\beta_0 + \beta_1 + u_t],$$

donde resulta

$$\ln(y_t) = \beta_0 + \beta_1 t + u_t. \tag{32}$$

Em ambos os casos, o aspecto a realçar é que os desvios em relação à tendência (u_t) são assumidos como estacionários. Ora esta visão corresponde à perspectiva tradicional dos ciclos económicos: a tendência determinística seria determinada pelo crescimento quantitativo e qualitativo da população, pelo progresso tecnológico, etc.. Os ciclos económicos representariam apenas perturbações transitórias relativamente a essa tendência. Por outras palavras, os efeitos dos choques económicos dissipar-se-iam mais ou menos rapidamente com o tempo.

Pelo contrário, teorias mais recentes como a dos ciclos económicos reais — que atribuem a totalidade das flutuações do produto aos choques de produtividade —, ou do rendimento permanente, por exemplo, salientam a forte persistência dos choques sobre variáveis como o produto e o consumo, respectivamente. Assim, a melhor maneira de representar estatisticamente essas séries seria através de processos integrados, isto é, estacionários por diferenciação (DSP) ou de raiz unitária (RU).

De facto, recorde-se que, por exemplo, um passeio aleatório com deriva

$$y_t = \beta_1 + y_{t-1} + \epsilon_t$$

também consegue explicar facilmente a presença de uma tendência crescente (desde que $\beta_1 > 0$; a título de exemplo veja-se o gráfico 3). Como a equação (28) mostra, a grande diferença relativamente ao modelo de estacionaridade em tendência (TS) é que a presença da RU introduz também uma tendência estocástica. Ou seja, os

Gráfico 3 Série I(1): $z_t = \beta_1 + z_{t-1} + \epsilon_t,\ \epsilon_t \sim nid(0, \sigma^2)$ (passeio aleatório com deriva) desvios relativamente à tendência determinística são, segundo a hipótese DSP ou de raiz unitária, não estacionários.

A questão a responder é, pois, a seguinte: os desvios em relação à tendência determinística são meramente transitórios, existindo como que uma "mão invisível" que volta a colocar a economia no caminho da tendência determinística ou, pelo contrário, os choques económicos têm um efeito permanente?

Na fase inicial da "revolução" foram sobretudo estas motivações decorrentes do debate (político-)económico que a alimentaram. Todavia, os dois modelos têm também implicações estatísticas claramente dissemelhantes. Sem desenvolver o tema, refira-se, por exemplo, que os procedimentos para obter as melhores previsões são radicalmente distintos para os dois tipos de processos. Por outro lado, como é óbvio, o procedimento adequado para a estacionarização depende da natureza da série: a) se esta fôr TS deve remover-se a tendência determinística, usando-se os resíduos OLS da regressão (31) ou (32); b) no caso DSP, obviamente, deve diferenciar-se a série [19].

4.1.3 Regressões espúrias

Uma motivação adicional para o estudo das propriedades univariadas das séries económicas — isto é, para os testes de raízes unitárias — reside nos problemas de inferência que podem surgir nos modelos de regressão que envolvem variáveis integradas. Por outras palavras, a realização prévia dessa análise univariada para cada uma das séries envolvidas é indispensável para a (boa) modelização e inferência multivariadas.

A situação que melhor ilustra essa importância é a das regressões espúrias ou sem sentido (termo usado pela primeira vez por Yule na década de 20). Em termos simples, uma regressão espúria é uma regressão sem significado económico mas em que a utilização dos instrumentos estatísticos tradicionais se revela enganadora. Isto é, a utilização de estatísticas como o R^2 (e, em geral, os coeficientes de correlação), ou as estatísticas t e F sugere a existência de relações de causalidade entre as variáveis, relações essas que, de facto, não existem.

Embora existam vários tipos de regressões espúrias, o exemplo mais comum e perturbador é o que envolve dois passeios aleatórios sem deriva e independentes, isto é:

$$y_t = y_{t-1} + \epsilon_{t1},$$

[19] Se esta tiver sido previamente logaritmizada, o parâmetro α da equação (30) representa aproximadamente a sua taxa média de variação.

$$x_t = x_{t-1} + \epsilon_{t2},$$

onde ϵ_{ti} e ϵ_{t2} são processos ruído branco independentes. Todavia, o investigador desconhece a ausência de qualquer relação e estima o seguinte modelo de regressão, envolvendo os níveis das variáveis [20]:

$$y_t = \alpha + \beta x_t + u_t. \tag{33}$$

Então, é bastante provável que venha a obter resultados estatísticos enganadores, que sugerem a existência de uma relação de causalidade. Em particular, o R^2 poderá assumir valores moderados ou altos e, muito frequentemente, o rácio-t para testar $H_0 : \beta = 0$ *vs* $H_1 : \beta \neq 0$ poderá cair na região crítica do teste. Ou seja, apesar de se determinar o valor crítico da forma habitual, das tabelas da t (ou da normal) correspondendo (nominalmente) ao nível de 5%, o teste rejeitará H_0 muito mais frequentemente que os 5% assumidos. Para se ter uma ideia mais clara da dimensão do problema, geraram-se em computador 10000 séries como as indicadas acima e efectuaram-se as respectivas (10000) regressões. Apesar da independência dos processos, em 65.7% dos casos a hipótese $H_0 : \beta = 0$ foi rejeitada quando $T = 50$. Por outro lado, o aumento da dimensão da amostra só veio agravar o problema: quando se usaram amostras com $T = 300$ a percentagem de rejeições incorrectas passou a ser de 86.3% (mais longe ainda dos 5% que deveriam ter ocorrido).

Dada a independência das séries, esperar-se-ia que $R^2 \to 0$, que $t_\beta \to 0$ e que $\hat{\beta} \to 0$ (o símbolo '$\to$' representando convergência em probabilidade). Todavia, neste caso pode mostrar-se que:

a) o R^2 converge em distribuição para uma variável aleatória que assume valores em $[0; 1]$; ou seja, com alguma frequência o R^2 assume valores moderados ou elevados;

b) o rácio-t de β diverge (para $+\infty$ ou $-\infty$), ou seja, não existem valores críticos válidos assimptoticamente e a utilização dos valores críticos usuais conduzirá a uma percentagem de rejeições incorrectas cada vez maior à medida que a dimensão da amostra aumenta;

c) o estimador OLS de β é inconsistente, não convergindo em probabilidade para o verdadeiro valor do parâmetro ($\beta = 0$).

[20] Note-se que cada uma das séries incorpora uma tendência estocástica. Todavia, as duas tendências estocásticas são independentes.

Porque falha tão estrondosamente a teoria assimptótica habitual? Porque os seus pilares fundamentais caem por terra neste caso. Em particular:

a) como $x_t \sim I(1)$, a hipótese $\text{plim}(T^{-1}X'X) = Q$, matriz finita (definida positiva), não é satisfeita e,

b) sobretudo, porque os erros da equação (33) têm um comportamento muito longe do iid; na realidade, dada a ausência de uma relação entre y_t e x_t pode mostrar-se que $u_t \sim I(1)$ [21].

E numa regressão múltipla, envolvendo várias séries I(1) independentes? Todos estes problemas subsistem e, adicionalmente, a estatística-F do teste de significância global da regressão (de $H_0 : \beta_2 = ... = \beta_k = 0$) também diverge (para $+\infty$), não existindo valores críticos correctos, nem mesmo assimptoticamente.

Como detectar então se os resultados obtidos são ou não espúrios? Para esse efeito existem disponíveis vários procedimentos. Alguns só serão abordados mais adiante [22]. Dos restantes, referem-se apenas dois. Como em (33), $u_t \sim I(1)$, então a estatística d (ou DW) deve assumir um valor próximo de zero. O problema é que não existem valores críticos suficientemente gerais que permitam avaliar a sua proximidade de zero. Assim, o procedimento mais popularizado consiste em diferenciar as séries a incluir nos modelos de regressão. Ou seja, diferenciando (33) obtém-se a regressão

$$\Delta y_t = \beta \Delta x_t + \epsilon_{t1},$$

onde as séries envolvidas são I(0) e $\epsilon_{t1} \sim$ iid. Então, de novo no "mundo estacionário" e com "erros bem comportados", os métodos de inferência usuais, associados ao OLS, deverão permitir descobrir que, de facto, $\beta = 0$. No entanto, como veremos mais tarde, nem sempre esta é uma boa solução. De facto, se as séries forem I(1) mas estiverem relacionadas, a regressão envolvendo os seus níveis será preferível a esta.

4.2 Testes de raízes unitárias: Testes DF e ADF

A mensagem a reter da subsecção anterior é clara: dados os riscos que se correm quando trabalhamos com séries integradas, é conveniente estudar previamente as

[21] Ou seja, o problema pode ser visto como o caso limite dos problemas de inferência que ocorrem quando os erros da regressão são autocorrelacionados.

[22] Para referência futura, o leitor deverá notar que neste caso as séries envolvidas são I(1) e *não cointegradas.*

propriedades univariadas das séries individuais antes de "embarcarmos" em análises multivariadas. Isto é, antes de tentarmos fazer inferência com base em modelos de regressão, devemos efectuar testes de raízes unitárias a todas as séries envolvidas.

4.2.1 Testes DF

Inicia-se a análise com o teste mais comum, para escolher entre o modelo TS

$$y_t = \beta_0 + \beta_1 t + u_t,$$

com u_t estacionário, e o modelo DSP (o AR(1) com deriva e $\rho = 1$, isto é, com uma RU):

$$y_t = \beta_1 + y_{t-1} + \epsilon_t.$$

Como os dois modelos são não encaixados, é necessário formular um modelo mais geral de que ambos resultem como caso particular. Pode provar-se que esse modelo mais geral é do tipo

$$y_t = \alpha + \beta t + \rho y_{t-1} + \epsilon_t, \tag{34}$$

com $\epsilon_t \sim$ iid, obtendo-se o TSP quando $\rho < 1$ e o DSP quando $\rho = 1$.

Trata-se então de testar a hipótese nula de uma RU com base na equação (34), isto é,

$$H_0 : \rho = 1 \Leftrightarrow y_t \sim I(1),$$

contra a alternativa de estacionaridade em tendência,

$$H_1 : \rho < 1 \Leftrightarrow y_t \sim I(0),$$

e note-se que o teste é unilateral esquerdo: a região crítica será toda definida na aba esquerda da distribuição da estatística de teste. Esta é uma estatística-t, representada com $\tau_{ct} = (\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}}$.

Todavia, a estatística de teste calcula-se ainda mais facilmente se subtrairmos y_{t-1} a ambos os membros de (34):

$$\Delta y_t = \alpha + \beta t + (\rho - 1) y_{t-1} + \epsilon_t, \tag{35}$$

isto é,

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \epsilon_t, \tag{36}$$

onde se fez $\phi = (\rho - 1)$. Ou seja, trata-se agora simplesmente de testar

$$H_0 : \phi = 0 \;\; vs \;\; H_1 : \phi < 0,$$

teste que se pode basear no rácio-t de ϕ: $\tau_{ct} = \hat{\phi}/\hat{\sigma}_{\hat{\phi}}$.

O problema é que, como mostraram Dickey e Fuller, a distribuição da estatística de teste (sob H_0) não é uma *t-student* nem uma normal (nem mesmo assimptoticamente). A distribuição assimptótica é algo complicada, envolvendo os chamados processos de Wiener, não simétrica e com mediana e moda menores que zero, e foi primeiramente tabelada por Dickey e Fuller (usando métodos de simulação). Daí que o nome adoptado para a distribuição tenha sido o desses investigadores, o mesmo acontencendo com os testes, vulgarmente conhecidos como *testes DF*.

Todavia, mais recentemente, MacKinnon [1991] propôs tabelas de valores críticos assimptóticos mais precisas mas que, sobretudo, permitem obter um valor crítico (estimado) adequado à particular dimensão da amostra empregue (T). A obtenção desses valores críticos baseia-se na fórmula

$$\hat{C}(p,\, T) = \hat{\beta}_\infty + \hat{\beta}_1 T^{-1} + \hat{\beta}_2 T^{-2},$$

onde p representa a dimensão ou nível do teste (o usual α), $\hat{\beta}_\infty$ é o valor crítico assimptótico e a expressão que se segue serve para efectuar a correcção de acordo com a dimensão da amostra (isto é, representa a correcção para amostras finitas). Assim, por exemplo, para o caso que temos vindo a analisar, se pretendermos obter o valor crítico a 5% para uma amostra com $T = 50$ ter-se-á (do caso $N = 1$, "with trend"):

$$\hat{C}(0.05,\, 50) = -3.4126 - 4.039 \times 50^{-1} - 17.83 \times 50^{-2} \approx -3.50.$$

Isto é, a região crítica será dada por RC $= \{\tau_{ct} : \tau_{ct} < -3.50\}$. Assim, só se o rácio-$t$ de ϕ fôr inferior a -3.50 é que a hipótese DSP, isto é, a presença da RU, será rejeitada.

Todavia, nem sempre se justifica a inclusão do termo de tendência determinística (t) na equação de teste. É esse o caso, por exemplo, para séries como as das taxas de inflação ou de desemprego. Para casos como esse a questão é: serão os desvios em relação a uma constante estacionários?

Ou seja, em vez de (36), a regressão de teste é agora

$$\Delta y_t = \alpha + \phi y_{t-1} + \epsilon_t, \tag{37}$$

continuando a testar-se

$$H_0 : \phi = 0 \;\; vs \;\; H_1 : \phi < 0,$$

com base no rácio-t de ϕ, agora representado com $\tau_c = \hat{\phi}/\hat{\sigma}_{\hat{\phi}}$. Note-se, contudo, que também a distribuição da estatística de teste muda, isto é, os valores críticos não são os mesmos que para o caso anterior mas sim mais pequenos em valor absoluto. Para $\alpha = 0.05$ e $T = 50$ obtém-se agora das tabelas de MacKinnon ($N = 1$, "no trend"):

$$\hat{C}(0.05, 50) = -2.8621 - 2.738 \times 50^{-1} - 8.36 \times 50^{-2} \approx -2.92,$$

ou seja, a presença da RU só é rejeitada se $\tau_c < -2.92$.

Finalmente, se a série oscilar em torno de zero (o que raramente deverá ocorrer com séries económicas), nem sequer a constante é incluída na equação de teste, que passa a ser dada simplesmente por

$$\Delta y_t = \phi y_{t-1} + \epsilon_t, \tag{38}$$

representando-se com $\tau_{nc} = \hat{\phi}/\hat{\sigma}_{\hat{\phi}}$ a estatística de DF. Como se verifica facilmente das tabelas de MacKinnon ($N = 1$, "no constant"), os valores críticos são ainda mais pequenos em valor absoluto que no caso anterior.

Na prática, o problema da escolha da regressão de teste a usar ((36), (37) ou (38)) tem implicações mais importantes do que alguns julgam. Assim, por um lado, incluir regressores determinísticos desnecessários reduz a potência do teste. Isto é, por exemplo, se a regressão adequada fôr a de (37) mas fôr empregue a de (36), fornecendo a estatística τ_{ct}, o teste tem menos potência que o baseado na primeira (ou seja, na estatística τ_c). Por outro lado, a exclusão de regressores determinísticos necessários também reduz a potência dos testes mas, neste caso, de forma ainda mais grave que no anterior, pois a potência do teste tende para zero quando $T \to \infty$ [23].

Para ilustrar esta última situação, efectuou-se um estudo de simulação (de Monte Carlo) em que se geraram 5000 séries com base no seguinte PGD:

$$y_t = 2 + 0.1t + 0.7y_{t-1} + \epsilon_t,\ \epsilon_t \sim nid(0, 1),$$

isto é, considerando a equação (34), trata-se de uma série estacionária em tendência (TS) pois o coeficiente de y_{t-1} é inferior a um. Por outras palavras, não existe nenhuma raiz unitária na série. A presença de uma tendência determinística na série

[23] Um teste cuja potência contra a hipótese alternativa tende para a unidade quando $T \to \infty$ diz-se consistente. Em geral, os testes estatísticos são consistentes. Todavia, neste caso, tem-se uma forma extrema de inconsistência.

implica que a equação de teste adequada seja a de (36). Todavia, para cada uma das (5000) séries geradas, efectuou-se o teste da presença de uma RU usando (36) e (37). No quadro seguinte apresentam-se as percentagens de rejeições da hipótese nula (falsa) de RU, para diferentes dimensões da amostra, nas 5000 réplicas, isto é, estimativas das potências dos testes baseados em τ_{ct} e em τ_c.

T	30	50	100	200
τ_{ct}	36.5%	71.4%	99.5%	100.0%
τ_c	18.5%	2.3%	0.0%	0.0%

Como se pode constatar, o comportamento da potência do teste baseado na regressão correcta pode ser considerado normal, se bem que para o caso $T = 30$ seja relativamente baixa. Pelo contrário, a omissão incorrecta do termo em t na regressão (37), não só produz um teste com menor potência para qualquer T como, ainda, com um comportamento bastante estranho à medida que T aumenta, pois é claro que a potência tende para zero quando $T \to \infty$ (bastando neste caso que $T = 100$ para que a potência se anule).

Como se sugeriu implicitamente, em geral, a escolha entre (36), (37) e (38) deve basear-se numa representação gráfica da série sob análise. Todavia, sobretudo em caso de dúvida, o raciocínio económico deve prevalecer sobre a análise gráfica. Assim, por exemplo, uma análise gráfica da série trimestral da taxa de juro, observada apenas sobre os últimos 10 anos, sugere a presença de uma tendência determinística decrescente com o tempo. Todavia, essa é apenas uma característica da amostra observada, não fazendo sentido que o modelo adequado para a taxa de juro contenha um termo de tendência determinística. Assim, a regressão de teste a empregar deve ser a de (37) (e não a de (36)). Por outro lado, deve procurar-se trabalhar com amostras que correspondam a um espaço temporal bem maior que os 10 anos deste exemplo.

4.2.2 Testes ADF

Os testes da subsecção anterior poderão, no entanto, não ser válidos. De facto, repare-se que, em todos os casos, o modelo assumido sob a hipótese nula é muito restritivo: trata-se de um simples AR(1), que pode ser insuficiente para capturar a evolução da série em função dos seus valores passados. Por outras palavras, pode ser necessário recorrer a um AR(2) ou, mais geralmente, a um AR(p), com $p > 1$, para conseguir uma aproximação razoável ao comportamento univariado da série. Ora, nesse caso, pode mostrar-se que, ao contrário do assumido, os erros das equações de teste serão autocorrelacionados, o que torna a inferência baseada nas estatísticas DF inválida. Por outras palavras, nessas circunstâncias as estatísticas DF não têm (sob H_0) a distribuição de Dickey-Fuller.

A solução proposta por Dickey e Fuller [24] para esse problema consiste em usar uma autoregressão de ordem suficientemente elevada, que permita que os resíduos das equações de teste não apresentem sintomas de autocorrelação. Assim, por exemplo, para o caso DSP *vs* TSP, em vez de (36) deve empregar-se a equação de teste

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta y_{t-i} + \epsilon_t, \tag{39}$$

onde k é determinado de forma a que os erros tenham um comportamento próximo do de um processo ruído branco (e continuando a testar-se $H_0 : \phi = 0$ *vs* $H_1 : \phi < 0$ [25]). A justificação para o nome destes testes, "augmented DF", ou melhor, ADF(k), é clara: as equações dos testes DF ((36), (37) ou (38)) são aumentadas com desfasamentos da variável dependente.

Todavia, emerge um novo problema: o da escolha de k. Muito frequentemente, os resultados dos testes serão muito sensíveis a esse valor. Pode também mostrar-se que, com um k mal escolhido errando-se por defeito, os testes tenderão a ter nível real superior ao nominal (usualmente 5% [26]) mas que, por outro lado, um valor para k mais elevado que o estritamente necessário retirará potência aos testes. (Porquê?)

A estratégia adoptada inicialmente para a resolução do problema consistia em começar a análise com $k = 0$, efectuando a regressão dos testes DF, e em ir aumentando a equação com desfasamentos de Δy_t até que os sintomas de autocorrelação residual (avaliados com as estatísticas de Breusch-Godfrey) desaparecessem. Todavia, mais recentemente, a estratégia inversa, do geral para o particular, tem obtido mais argumentos favoráveis, sobretudo quando o objectivo principal reside no controlo da dimensão dos testes. Assim, inicia-se o processo com um k suficientemente elevado (k_{max}) e vai-se tentando simplificar a autoregressão com testes-t individuais sobre os coeficientes de desfasamento mais elevados, até se obter uma rejeição. Eventualmente poder-se-á chegar a ADF(0), isto é, a uma estatística DF. Todavia, se a eliminação de um desfasamento insignificante fizer surgir sintomas de autocorrelação residual até aí

[24] E estendida por Said e Dickey para o caso dos processos ARMA.

[25] Para o caso de um modelo AR(p), usando um dos resultados sobre decomposição de polinómios em L estudados anteriormente, pode mostrar-se que $k = p - 1$.

[26] Isto é, tenderão a rejeitar H_0 quando ela é verdadeira mais frequentemente que nos 5% dos casos assumidos.

inexistentes, esse desfasamento é reintroduzido na equação de teste. Resta referir que o valor escolhido para k_{max} depende, em geral, da dimensão da amostra. Assim, por exemplo, para as séries anuais disponíveis para a economia Portuguesa, com $T \approx 40$, poder-se-á iniciar o processo com $k_{max} = 4$ ou 5.

5 Cointegração: uma Introdução

Associada à revolução das raízes unitárias encontra-se a da cointegração, esta no plano da análise multivariada. O texto que se segue é apenas uma introdução muito breve e simples ao estudo do tema.

5.1 Definição

Comecemos por considerar algumas propriedades simples sobre combinações lineares de processos I(0) e I(1). Sendo a e b constantes não nulas,

a) se $x_t \sim I(0)$, então $(a + bx_t) \sim I(0)$;
   se $x_t \sim I(1)$, então $(a + bx_t) \sim I(1)$;

b) se $x_t,\ y_t \sim I(0)$, então $(ax_t + by_t) \sim I(0)$;

c) se $x_t \sim I(1)$ e $y_t \sim (0)$, então $(ax_t + by_t) \sim I(1)$ (isto é, "I(1) é uma propriedade dominante");

d) *em geral*, se $x_t,\ y_t \sim I(1)$, então $(ax_t + by_t) \sim I(1)$.

A excepção a esta última regra é precisamente o caso especial da cointegração. Assim, se $x_t,\ y_t \sim I(1)$ mas existe uma combinação linear,

$$y_t\ (-\alpha) - \lambda x_t = u_t \sim I(0),$$

(e que tem média nula), então x_t e y_t dizem-se cointegradas, escrevendo-se $(x_t, y_t) \sim CI(1,1)$, e o vector $[1 \quad -\lambda]'$ é chamado de vector de cointegração. Note-se que a presença do termo independente não é estritamente necessária. Ele serve apenas para que u_t tenha média nula (o que facilita a sua interpretação como erro de equilíbrio). O que é importante é que se a combinação linear das séries I(1) faz desaparecer a tendência estocástica presente em cada uma delas é porque essa tendência estocástica é comum às duas séries. Ou seja, elas partilham essa tendência estocástica, isto é, o seu comportamento de longo prazo está ligado ou relacionado.

Por outro lado, para que exista uma relação de cointegração, pode ser necessário considerar várias séries I(1) e não apenas duas. Nesse caso pode escrever-se

$$y_t'\beta = y_{t1}\beta_1 + y_{t2}\beta_2 + ... + y_{tk}\beta_k = u_t \sim I(0), \tag{40}$$

onde y_t é um vector de séries I(1) e β é o (ou melhor, é um dos) vector(es) de cointegração. De facto, dadas duas séries I(1) mas cointegradas, pode provar-se que existe um único vector de cointegração linearmente independente. Todavia, dadas k séries I(1) e cointegradas, podem existir $k-1$ vectores de cointegração linearmente independentes. Para simplificar, a análise que se segue assume que, mesmo neste último caso, só existe um vector de cointegração (linearmente independente).

5.2 Exemplos

A teoria económica sugere frequentemente a existência de uma relação de equilíbrio de longo prazo entre duas ou mais variáveis económicas. Se observarmos as variáveis apenas durante períodos temporais curtos, é natural que se registem desequilíbrios. Todavia, no longo prazo, as variáveis tenderão a mover-se de forma a satisfazer aproximadamente a referida relação. Ou seja, em situações de desequilíbrio, existem forças ou mecanismos económicos que actuam de forma a reestabelecer o equilíbrio, para o qual o sistema económico tenderá a convergir, no longo prazo. Por outras palavras, uma relação de equilíbrio de longo prazo tende a observar-se de forma aproximada, "em média", sobre períodos temporais longos [27].

O conceito de cointegração traduz, estatisticamente, a noção de equilíbrio de longo prazo. De facto, se as séries individuais são I(1), elas tendem a divergir quando $T \to \infty$, não sendo atraídas por nenhum valor fixo. Todavia, sendo cointegradas, o seu comportamento obedece à relação de equilíbrio de longo prazo, ou seja, o seu movimento comum obedece a uma relação sistemática. Assim, sendo u_t a combinação linear das séries I(1), $u_t \sim I(0)$, ela pode ser interpretada como o erro ou desvio relativamente ao equilíbrio, pois:

a) não apresenta qualquer tendência sistemática para divergir ao longo do tempo mas, pelo contrário, flutua sempre dentro de certos limites e

[27] Note-se que a situação de equilíbrio aqui mencionada não é a de equilíbrio estacionário estudada anteriormente. De facto, as séries envolvidas podem nem sequer ter média, pelo que não faz sentido falar em equilíbrio estacionário envolvendo os níveis das séries.

b) apresenta um comportamento de reversão para a média (nula), isto é, os movimentos das séries são atraídos pela situação de equilíbrio.

Considerem-se os seguintes exemplos:

1. Sejam f_t e s_t os logaritmos dos preços a prazo ("forward"), a um período, e "à vista" ("spot") de uma moeda estrangeira, respectivamente. Segundo a hipótese da eficiência dos mercados (cambiais, neste caso), tem-se

$$f_t = E_t(s_{t+1}),$$

onde E_t representa a expectativa de mercado formulada no período t. Por outro lado, segundo a hipótese das expectativas racionais, os erros de expectativas são estacionários e têm média nula, isto é,

$$s_{t+1} - E_t(s_{t+1}) = \epsilon_{t+1} \sim I(0), \text{ com } E_t(\epsilon_{t+1}) = 0, \forall t.$$

Assim, das duas equações resulta que

$$s_{t+1} - f_t = \epsilon_{t+1} \sim I(0).$$

Ora, se s_{t+1} e f_t são ambas I(1), tem-se uma combinação linear estacionária de duas séries I(1). O vector de cointegração é $[1 \ -1]'$.

2. Sejam p_t, p_t^* e e_t os logaritmos dos níveis de precos interno e externo e do preço em escudos da moeda estrangeira (taxa de câmbio), respectivamente. Ora, segundo a hipótese da paridade dos poderes de compra (PPP), a combinação linear $e_t + p_t^* - p_t = u_t$ deve ser estacionária. Então, se e_t, p_t^* e p_t são I(1), a PPP requer que elas sejam cointegradas, com vector de cointegração $[1\ 1\ -1]'$.

3. Pelo contrário, a função habitual de procura de moeda não impõe qualquer restrição de igualdade sobre os parâmetros da relação de equilíbrio de longo prazo:

$$m_t = \beta_0 + \beta_1 y_t + \beta_2 r_t + u_t,$$

onde m_t representa o *stock* real de moeda, y_t o rendimento real e r_t uma taxa de juro. Se todas estas séries forem I(1) mas $u_t \sim I(0)$, então a equação anterior exprime uma relação de cointegração.

Outros exemplos de relações de cointegração poderão ser dados pelas relações entre preços e salários, entre taxa de juro nominal e inflação, entre consumo e rendimento disponível, etc. .

Note-se ainda que:

a) obviamente, não há cointegração entre séries I(0);

b) uma regressão espúria é, de alguma forma, o oposto de uma relação de cointegração: no primeiro caso não existe nenhuma combinação linear das séries I(1) que seja I(0) (as tendências estocásticas das séries individuais são independentes).

Nos gráficos 4 a 7 apresentam-se exemplos de séries I(1) não cointegradas e cointegradas e dos resíduos (OLS) das regressões entre elas, os quais, no segundo caso, são estimativas dos erros de equilíbrio. Repare-se, por exemplo, no comportamento tipicamente não estacionário desses resíduos no caso de inexistência de relação de cointegração (gráfico 5) e no paralelismo aproximado, de longo prazo, da evolução das duas séries I(1) mas cointegradas (gráfico 6).

5.3 Estimação OLS

Neste momento colocam-se dois problemas: 1) como testar a existência de uma relação de cointegração? e 2) como estimar o vector (assumido único) de cointegração? Começa-se pela abordagem deste último assumindo que existe cointegração entre as séries envolvidas. Assim, retome-se a equação (40) mas normalize-se o coeficiente de y_{t1} (fazendo $\beta_1 = 1$), por exemplo, escrevendo a relação sob a forma de uma equação de regressão,

$$y_{t1} = \lambda_1 + \lambda_2 y_{t2} + ... + \lambda_k y_{tk} + u_t, \tag{41}$$

onde se incluiu também um termo independente. De resto, para que além de ser I(0) u_t possua também média nula (e não seja apenas estacionário em tendência), pode ser necessário incluir ainda uma tendência determinística em (41) [28]. Em geral, é tomada para y_{t1} a variável dependente da relação dada pela teoria económica; todavia, não tem que ser forçosamente assim.

[28] Uma solução não óptima mas muito frequentemente usada em aplicações empíricas consiste em experimentar a inclusão de um termo de tendência (desde que a variável usada como dependente o justifique), retendo-o na especificação desde que a evidência proporcionada pelos testes de cointegração seja mais favorável a essa hipótese quando ele se encontra presente.

Gráfico 4 Inexistência de cointegração entre séries I(1): passeios aleatórios independentes

Gráfico 5 Resíduos da regressão entre duas séries I(1) não cointegradas

Gráfico 6 Séries I(1) cointegradas

Gráfico 7 Resíduos da regressão entre duas séries I(1) cointegradas

O método mais simples para estimar $\lambda = [\lambda_1\ \lambda_2\ ...\ \lambda_k]'$ da equação (41), a chamada regressão de cointegração, é o OLS. Além disso, ao contrário do que sucede nas regressões espúrias, neste caso (de cointegração), as propriedades do estimador OLS são particularmente boas. Na realidade, pode provar-se que, nestas circunstâncias, o estimador OLS é "super-consistente": $\text{plim}\hat{\lambda} = \lambda$ e a convergência dá-se a uma velocidade ainda maior que no caso de as séries envolvidas serem estacionárias. Ou seja, pode provar-se que para se obter uma distribuição não degenerada não basta escalar $(\hat{\lambda} - \lambda)$ com $T^{1/2}$; a chamada "transformação estabilizadora da variância" tem agora que ser efectuada usando a multiplicação por T pois os erros de estimação convergem muito rapidamente para zero quando $T \to \infty$. Assim, mesmo que os regressores de (41) sejam endógenos (contemporaneamente correlacionados com u_t), $\hat{\lambda}$ não sofre de qualquer problema de inconsistência; em particular, desde que exista cointegração, o problema do "enviesamento de simultaneidade" do estimador OLS desaparece.

Todavia, deve salientar-se que: a) a endogeneidade dos regressores torna o estimador OLS enviesado e o enviesamento pode ser grande em pequenas amostras e b) o mesmo problema torna os métodos de inferência usuais (baseados nas estatísticas-t e F) não válidos. Ou seja, só no caso em que os regressores de (41) forem estritamente exógenos e os erros iid é que esses métodos serão válidos (assimptoticamente) [29]. Num contexto de cointegração, essa exigência só muito raramente poderá ser satisfeita pelo que, em geral, não se deverá tentar efectuar inferência sobre os parâmetros de (41) usando os procedimentos habituais. Por outro lado, apesar de ser I(0), u_t só muito raramente será não autocorrelacionada pois representa o erro de uma equação de regressão estática.

5.4 Cointegração e MCE: o Teorema de Representação de Granger

Os problemas mencionados relativamente aos métodos de inferência parecem sugerir que, mesmo no caso de cointegração, parece ser preferível trabalhar com regressões envolvendo as séries I(1) previamente diferenciadas do que com essas séries em nível: no primeiro caso, o "ambiente" de estacionaridade permitiria utilizar os métodos de inferência habituais, os quais, em geral, não são válidos no segundo. Todavia, a

[29] Mas note-se que a teoria assimptótica a que é necessário recorrer para provar este resultado não é a teoria *standard* pois a hipótese $\text{plim}(T^{-1}X'X) = Q$, matriz de constantes, finita, não é satisfeita.

utilização de modelos de regressão apenas com as séries I(1) diferenciadas despreza a informação sobre a existência de uma relação de equilíbrio de longo prazo, a qual envolve os níveis das séries.

Ora, já sabemos que o modelo de correcção de erros permite conciliar o ajustamento dinâmico de curto prazo com a relação de equilíbrio de longo prazo e, que nesse modelo, se encontram presentes também os níveis das variáveis. Não deve assim constituir surpresa a existência de uma relação estreita entre a cointegração e o MCE. Essa relação é dada pelo Teorema de Representação de Granger, do qual se apresenta aqui apenas uma versão muito simples e envolvendo somente duas séries. A generalização para o caso de um maior número de séries deverá ser clara.

Suponha-se que $(y_t, x_t) \sim CI(1,1)$ com vector de cointegração $[1 \;\; -\lambda]'$ e que o vector (y_t, x_t) admite uma representação VAR(p) [30]. Então, existe uma representação sob a forma MCE, do tipo

$$\Delta y_t = \delta_1 + \alpha_1(y_{t-1} - \lambda x_{t-1}) + \sum_{i=1}^{p-1} \gamma_{1i} \Delta y_{t-i} + \sum_{i=1}^{p-1} \theta_{1i} \Delta x_{t-i} + \epsilon_{t1}, \tag{42}$$

$$\Delta x_t = \delta_2 + \alpha_2(y_{t-1} - \lambda x_{t-1}) + \sum_{i=1}^{p-1} \gamma_{2i} \Delta y_{t-i} + \sum_{i=1}^{p-1} \theta_{2i} \Delta x_{t-i} + \epsilon_{t2}, \tag{43}$$

onde pelo menos um dos α's é diferente de zero e ϵ_{t1} e ϵ_{t2} são processos ruído branco. O inverso também é verdadeiro: se o vector $(y_t,\, x_t) \sim I(1)$ mas admite uma representação em MCE, então $(y_t, x_t) \sim CI(1,1)$.

Repare-se, então, que (42) e (43) são equações nas primeiras diferenças das séries mas aumentadas com o termo de correcção de erros, isto é, os modelos contendo apenas as primeiras estão mal especificados pois omitem o efeito do erro de equilíbrio do período anterior na dinâmica de curto prazo em pelo menos uma das equações.

As equações (42) e (43) podem ser vistas como fornecendo a distribuição conjunta de (y_t, x_t). Todavia, muito frequentemente a equação (43) não tem interesse sob o ponto de vista económico, concentrando-se a atenção apenas na estimação de (42), o MCE de y_t condicional em x_t (e sendo desprezada a informação contida em (43)).

[30] Os modelos VAR ou autoregressões vectoriais são a extensão dos modelos AR ao caso multivariado. Por exemplo, o modelo VAR(1) para o vector $z_t = [y_t \; x_t]'$ é dado por

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{t1} \\ \epsilon_{t2} \end{bmatrix},$$

onde $\epsilon_t = [\epsilon_{t1} \, \epsilon_{t2}]' \sim \text{iid}(0, \Omega)$.

Embora seja essa também a abordagem que será aqui prosseguida por razões de simplicidade, deve notar-se que nem sempre ela é adequada. Na realidade, pode mostrar-se que ela só fornece estimadores eficientes desde que $\alpha_2 = 0$ (caso em que se diz que x_t é "fracamente exógena" para os parâmetros de (42)).

5.5 Testes de cointegração

Para testar a existência de uma relação de cointegração existem inúmeros métodos, sendo o que se baseia na estimação de um modelo VAR um dos mais potentes. Aqui, só abordaremos dois dos métodos mais simples: os testes (já clássicos) de Engle-Granger e o teste t-MCE.

5.5.1 Testes de Engle-Granger

Estes testes são os mais populares pois decorrem imediatamente da definição de cointegração e só requerem a utilização do OLS. Assim, dado um conjunto de séries I(1), como em (40), trata-se apenas de testar se u_t, de (40) ou de (41), pode ser considerado I(0). Ora, para esse efeito, já dispomos dos testes DF (e ADF) onde, recorde-se, é na hipótese alternativa que se especifica a possibilidade I(0). Ou seja, a formalização das hipóteses é dada por:

$$H_0 : u_t \sim I(1) \Leftrightarrow \text{no cointegrao},$$

$$H_1 : u_t \sim I(0) \Leftrightarrow \text{cointegrao}.$$

Ou seja, não sendo adoptada como H_0, a hipótese de cointegração não é protegida. Assim, se o vector de cointegração fôr conhecido, como nos exemplos 1 e 2 da subsecção 5.2, u_t é observável e podem efectuar-se imediatamente sobre ela os testes de raízes unitárias, agora desempenhando o papel de testes de cointegração.

Contudo, a situação em que o vector de cointegração é conhecido não é a mais vulgar. Uma possibilidade consiste então em obter estimativas dos potenciais erros de equilíbrio (dos u_t's) estimando, com o OLS, os parâmetros da potencial equação de cointegração. Ou seja, sendo $e_t = \hat{u}_t = y_{t1} - \hat{\lambda}_1 - \hat{\lambda}_2 y_{t2} - \ldots - \hat{\lambda}_k y_{tk}$, trata-se agora de efectuar testes de raízes unitárias sobre os resíduos OLS, por exemplo com base na equação

$$\Delta\hat{u}_t = \phi\hat{u}_{t-1} + \epsilon_t, \tag{44}$$

usada para testar $H_0 : \phi = 0$ *vs* $H_1 : \phi < 0$ (empregando o rácio-t de ϕ). Note-se que não é necessário, agora, incluir um termo independente e uma tendência determinística nas equações de teste: os resíduos deverão ter média nula pois esses regressores terão sido já incluídos em (41). Contudo, tal como nos testes ADF, para que os testes sejam válidos, pode ser necessário incluir desfasamentos da variável dependente:

$$\Delta\hat{u}_t = \phi\hat{u}_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta\hat{u}_{t-i} + \epsilon_t. \tag{45}$$

Todavia, quando se usam os resíduos OLS, a distribuição (assimptótica) das estatísticas de teste já não é a de DF mas sim a de Engle-Granger (EG), assim chamada por terem sido estes os primeiros autores a obter tabelas de valores críticos. Na realidade, os valores críticos são agora ainda mais exigentes (maiores em valor absoluto), e são-no tanto mais quanto maior fôr o número de regressores estocásticos incluídos. Isso acontece porque os resíduos OLS tendem a apresentar um comportamento "mais estacionário" que os verdadeiros erros, tendência esta que se acentua quando o referido número aumenta. Assim, de (44) tem-se o teste EG e de (45) tem-se o teste AEG(k) (mas note-se que em muitos textos estes testes continuam a ser chamados DF e ADF(k), respectivamente).

Nas tabelas de MacKinnon usa-se agora a zona inferior, de acordo com o número total de variáveis (aleatórias) I(1) envolvidas (N, incluindo a variável que é tomada como regressando de (41)). Assim, por exemplo, para uma equação com dois regressores estocásticos I(1) e incluindo termo independente mas não tendência determinística tem-se ($N = 3$, "no trend"), quando $T = 50$:

$$\hat{C}(0.05,\, 50) = -3.7429 - 8.352 \times 50^{-1} - 13.41 \times 50^{-2} \approx -3.92\,.$$

5.5.2 O teste t-MCE

Com muita frequência, os testes anteriores não proporcionam evidência de suporte à teoria económica, isto é, à existência de uma relação de equilíbrio de longo prazo. Tal facto decorre, por um lado, da sua reduzida potência, característica herdada dos testes DF. Assim, se u_t fôr bem aproximada por um AR(1) estacionário mas com um parâmetro autoregressivo próximo da fronteira de não estacionaridade (por exemplo, $u_t = 0.9u_{t-1} + \epsilon_t$), para terem um bom comportamento em termos de potência os testes necessitam de amostras de grande dimensão, que raramente se encontram disponíveis. Por outro lado, recorde-se que, ao contrário do que deveria ocorrer, a formalização adoptada para as hipóteses não permite proteger as proposições dadas pela teoria económica.

Embora também padecendo deste último problema, o teste t-MCE possui, em geral, melhor comportamento em termos de potência, facto que resulta de se basear numa especificação dinâmica. A lógica do teste é simples: trata-se de empregar a segunda parte do Teorema de Representação de Granger, testando a presença de um mecanismo de correcção de erros. De acordo com esse teorema, evidência favorável à presença do MCE equivale a evidência de suporte à existência da relação de cointegração.

Embora o teste seja mais frequentemente apresentado com base na reparametrização em "MCE homogéneo", considerar-se-á aqui a parametrização do modelo ADL sob a forma de Bardsen. Assim, tomando como exemplo o caso de um ADL(2,2) em $y_t\,(\equiv y_{t1})$, $x_t\,(\equiv y_{t2})$ e $z_t\,(\equiv y_{t3})$ reparametrizado sob essa forma, tem-se (veja-se a equação (24)):

$$\begin{aligned}\Delta y_t = \; & \mu - A(1)y_{t-1} + \delta_1\Delta y_{t-1} + B_1(1)x_{t-1} + B_2(1)z_{t-1} \\ & +\gamma_{10}\Delta x_t + \gamma_{11}\Delta x_{t-1} + \gamma_{20}\Delta z_t + \gamma_{21}\Delta z_{t-1} + \epsilon_t,\end{aligned}$$

ou, de forma equivalente, fazendo $\alpha = -A(1)$, $\theta_1 = B_1(1)$ e $\theta_2 = B_2(1)$,

$$\Delta y_t = \mu+\alpha y_{t-1}+\delta_1\Delta y_{t-1}+\theta_1 x_{t-1}+\theta_2 z_{t-1}+\gamma_{10}\Delta x_t+\gamma_{11}\Delta x_{t-1}+\gamma_{20}\Delta z_t+\gamma_{21}\Delta z_{t-1}+\epsilon_t, \tag{46}$$

(equações que poderão conter ainda um termo em t). Assim, trata-se de testar

$$H_0 : \alpha = 0 \Leftrightarrow \text{no cointegrao},$$

$$H_1 : \alpha < 0 \Leftrightarrow \text{cointegrao},$$

com base em t-MCE $= t_\alpha = \hat{\alpha}/\hat{\sigma}_{\hat{\alpha}}$. Mais uma vez, contudo, o problema reside no facto de (sob H_0) a distribuição da estatística de teste não ser *standard*. Aliás, neste caso o problema nem sequer tem uma solução geral. Uma solução conservadora consiste em empregar os valores críticos da distribuição de EG. Todavia, em geral, sabe-se que os valores críticos adequados não serão tão exigentes, isto é, tão grandes em valor absoluto quanto os da distribuição de EG. Assim, uma solução menos conservadora mas muito grosseira poderá consistir em usar essa distribuição mas considerando uma dimensão de 10% para os testes.

5.6 Estimação do MCE condicional

Nalguns casos, o trabalho empírico encerra com os testes de cointegração; por exemplo, nos casos dos exemplos 1 e 2 da subsecção 5.2, em princípio só interessará testar a existência da relação de equilíbrio de longo prazo postulada pela teoria económica. Noutros casos, adicionalmente, também poderemos estar interessados na modelização da dinâmica de curto prazo de y_t (y_{t1}). Ora, em caso de cointegração, o modelo apropriado para esse efeito é o MCE (condicional). Dos métodos disponíveis para a sua estimação só dois serão abordados: o método clássico de Engle-Granger, em dois passos, e o método num só passo.

5.6.1 Método dos dois passos de Engle e Granger

Como exemplo, considere-se de novo um modelo envolvendo $(y_t, x_t, z_t) \sim CI(1,1)$, um ADL(2,2) reparametrizado sob a forma MCE:

$$\Delta y_t = \alpha(y_{t-1} - \lambda_1 - \lambda_2 x_{t-1} - \lambda_3 z_{t-1}) + \delta_1 \Delta y_{t-1} + \gamma_{10} \Delta x_t + \gamma_{11} \Delta x_{t-1} + \gamma_{20} \Delta z_t + \gamma_{21} \Delta z_{t-1} + \epsilon_t, \tag{47}$$

cuja parametrização parece requerer a utilização do método dos mínimos quadrados não lineares (NLLS). Todavia, dada a "superconsistência" do estimador OLS no caso de cointegração, Engle e Granger propuseram um método de estimação em dois passos. No primeiro, é estimado o vector de cointegração, isto é, estimam-se os multiplicadores de longo prazo, com base na regressão estática:

$$y_t = \lambda_1 + \lambda_2 x_t + \lambda_3 z_t + u_t. \tag{48}$$

No segundo passo, as estimativas obtidas são inseridas em (47), isto é, os resíduos OLS do primeiro passo são empregues como estimativas dos erros de equilíbrio:

$$\Delta y_t = \alpha \hat{u}_{t-1} + \delta_1 \Delta y_{t-1} + \gamma_{10} \Delta x_t + \gamma_{11} \Delta x_{t-1} + \gamma_{20} \Delta z_t + \gamma_{21} \Delta z_{t-1} + \epsilon_t,$$

e note-se que nesta equação, como todas as variáveis envolvidas são estacionárias, os métodos de inferência habituais são válidos (desde que $\epsilon_t \sim$ iid).

5.6.2 Método num só passo

Ao contrário do anterior, neste método a estimação dos parâmetros de longo prazo não é separada da dos parâmetros de ajustamento dinâmico de curto prazo: todos os parâmetros são estimados conjuntamente. De resto, este método já é conhecido:

trata-se de empregar a forma de Bardsen do modelo ADL. Ou seja, trata-se simplesmente de estimar, por exemplo, a equação (46):

$$\Delta y_t = \mu + \alpha y_{t-1} + \delta_1 \Delta y_{t-1} + \theta_1 x_{t-1} + \theta_2 z_{t-1} + \gamma_{10} \Delta x_t + \gamma_{11} \Delta x_{t-1} + \gamma_{20} \Delta z_t + \gamma_{21} \Delta z_{t-1} + \epsilon_t.$$

Recordando que este modelo também pode ser escrito como

$$\begin{aligned} \Delta y_t = \; & \alpha[y_{t-1} - (-\mu/\alpha) - (-\theta_1/\alpha)x_{t-1} - (-\theta_2/\alpha)z_{t-1}] + \delta_1 \Delta y_{t-1} \\ & + \gamma_{10} \Delta x_t + \gamma_{11} \Delta x_{t-1} + \gamma_{20} \Delta z_t + \gamma_{21} \Delta z_{t-1} + \epsilon_t, \end{aligned}$$

torna-se claro que as estimativas dos parâmetros de cointegração são obtidas indirectamente a partir da estimação de (46): $\hat{\lambda}_1 = -\hat{\mu}/\hat{\alpha}$, $\hat{\lambda}_2 = -\hat{\theta}_1/\hat{\alpha}$ e $\hat{\lambda}_3 = -\hat{\theta}_2/\hat{\alpha}$. Em muitos casos, as estimativas assim obtidas para os parâmetros de cointegração são melhores que as produzidas pelo estimador da regressão estática de (48). Pode ainda mostrar-se que, caso exista de facto uma relação de cointegração, mesmo os testes-t individuais sobre os coeficientes das variáveis em níveis (isto é, sobre α, θ_1 e θ_2) são válidos assimptoticamente. Recorde-se, no entanto, que a distribuição t-student não é válida para testar $H_0 : \alpha = 0$ *vs* $H_1 : \alpha < 0$, isto é, para efectuar o teste t-MCE. Por outro lado, os testes conjuntos sobre esses coeficientes com base em estatísticas-F não são válidos (nem sequer assimptoticamente).

Referências

Baneerjee, A., Dolado, J., Galbraith, J. W. e Hendry, D. F. [1993], *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press.

Boswijk, H. P. [1994], *Unit Roots and Cointegration: Statistical Analysis and Asymptotic Theory*, Tinbergen Institute and University of Amsterdam.

Campbell, J. Y. e Perron, P. [1991], Pitfalls and opportunities: what macroeconomists should know about unit roots, *NBER Macroeconomics Annual*, pp. 141-219.

Davidson, R. e MacKinnon, J. G. [1993], *Estimation and Inference in Econometrics*, Oxford University Press.

Hamilton. J. D. [1994], *Time Series Analysis*, Princeton University Press.

Hendry, D. F. [1995], *Dynamic Econometrics*, Oxford University Press.

Johansen, S. [1995], *Likelihood-Based Inference in Cointegrated Vector-Autoregressive Models*, Oxford University Press.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lutkepohl, H. e Lee, T.-C. [1988], *Introduction to the Theory and Practice of Econometrics*, 2nd ed., John Wiley & Sons.

MacKinnon, J. G. [1991], Critical values for cointegration tests, *in* Engle, R. F. e Granger, C. W. J. (eds.), *Long-run Economic Relationships*, pp. 267-76.

Maddala, G. S. e Kim, I.-M. [1998], *Unit Roots, Cointegration and Structural Change*, Cambridge University Press.

Marques, C. R. [1998], *Modelos Dinâmicos, Raízes Unitárias e Cointegração*, Edinova.

Stewart, J. [1991], *Econometrics*, Philip Allan.