Modelos Dinâmicos Lineares Uni-equacionais – uma introdução

Artur Silva Lopes

Versão 1.01, 2 de Abril de 2020

ISEG, ULisboa

Prefácio

Este texto de apoio destina-se sobretudo a ajudar os alunos de Macroeconometria I do ISEG no estudo dos modelos dinâmicos lineares uni-equacionais. Também pode ser empregue com outras finalidades, obviamente, e não se considera que o nível de dificuldade seja muito elevado. Existem contudo alguns pressupostos que se consideram satisfeitos como, por exemplo, algum conhecimento prévio de manipulações polinomiais no operador de desfasamento (aqui representado com L mas que na literatura de séries temporais é habitualmente representado com B, de *backward shift*).

A esmagadora maioria das notas em que este texto se baseia são bastante antigas, pois tiveram poucas alterações nos últimos anos. Aliás, em parte substancial deste texto transcrevo exactamente o que escrevi há cerca de 22 anos, para um texto feito para apoiar o estudo dos alunos de licenciatura de então, no ano lectivo de 1997/98 (Lopes, 1999). É o que se passa com as secções 3, 4 e 6, que constituem o núcleo central deste texto, bastante inspiradas no trabalho de Hendry (1995, por exemplo) e Boswijk (1994), este no que se refere à reparametrização dos modelos ADL e à sua relação com os modelos de correcção de erros. Procurei mudar algum desse texto mas, na maior parte das vezes, considerei que modificar o texto original só pioraria a sua clareza.

Não se pense, no entanto, que esta matéria está datada, no sentido de ultrapassada. Pelo contrário, a abordagem aqui apresentada é bastante mais moderna que a de Wooldridge (2016), cujo título faz apelo explícito à modernidade da abordagem. É também ainda actual, pois não é raro encontrar-se trabalho de investigação recente recorrendo aos modelos ADL, ou ARDL, como também são conhecidos. Aliás, boa parte desta matéria é importante para compreender um dos métodos empregues na estimação de vectores de cointegração, o chamado método "num só passo", que será abordado no último capítulo, o mais moderno e avançado do programa.

30 de Março, 2020

1 Introdução

Em economia é sabido que a modelação adequada da generalidade dos fenómenos exige, muito frequentemente, a inclusão, na mesma equação, de variáveis referidas a períodos de tempo distintos. Só desta maneira se tomam em consideração os "lags" ou desfasamentos de tempo na interacção entre variáveis. Existem inúmeros exemplos clássicos: funções de investimento, de procura, etc.. Por exemplo, na função consumo privado ou das famílias, agregada para a economia, o comportamento dos consumidores não é determinado exclusivamente pela sua situação de rendimento no período mas resulta também de hábitos do passado, de expectativas sobre o futuro (que muitas vezes dependem sobretudo da história passada), etc..

De uma forma geral, a necessidade de incluir variáveis referidas a períodos de tempo distintos numa mesma equação resulta designadamente de:

a) persistência de hábitos;

b) restrições (tecnológicas, institucionais, etc.) que não permitem que certos ajustamentos necessários ou desejados se efectuem imediatamente;

c) efeitos de expectativas dos agentes económicos, etc..

Mais geralmente, como nota Stewart (1991), não há razão para supôr que os dados são observados num estado de equilíbrio da economia, isto é, que os efeitos das alterações ocorridas no passado já se esgotaram ou que se esgotam no mesmo período em que ocorrem (como é assumido nos modelos estáticos). Parece mais adequado ver a economia em cada momento como estando a tentar-se ajustar aos choques que recebeu no passado, tentando mover-se para o equilíbrio, mas simultaneamente a receber novos choques que a afastam do equilíbrio, pelo menos temporariamente. Ora os modelos de regressão dinâmicos são os modelos apropriados pois são aqueles em que existem variáveis referidas a períodos de tempo distintos; que envolvem relações não contemporâneas entre variáveis.

2 Dinâmica sistemática

Além de algumas noções de dinâmica sistemática, nesta secção serão apresentados também alguns modelos dinâmicos comuns.

2.1 Modelos DL(s) e multiplicadores

Os modelos DL(s) são os modelos de desfasamento distribuído ("distributed lag") pois os efeitos das variações das variáveis *distribuem-se* por vários períodos de tempo. Considere-se, por exemplo, o seguinte modelo DL(2), sem termo independente para simplificar:

$$y_t = 0.4x_t + 1.0x_{t-1} + 0.6x_{t-2} + u_t.$$

Se em t dermos uma variação unitária a x que, no entanto, não se mantém, isto é, uma variação temporária ou transitória, o que acontece a y? Abstraindo de algum eventual efeito dos erros, ou seja, silenciando-os, tem-se a seguinte trajectória para (o valor esperado de) y:

$$\Delta y_t = y_t - y_{t-1} = 0.4, \quad y_{t+1} - y_{t-1} = 1, \;\; y_{t+2} - y_{t-1} = 0.6, \;\; y_{t+3} - y_{t-1} = 0, \ldots$$

O efeito contemporâneo da variação é de 0.4. Trata-se do multiplicador de impacto ou de curto prazo. Mas o efeito total ou acumulado da referida variação transitória, a soma das variações nos vários períodos, é 2.0. Trata-se do multiplicador total ou de longo prazo. O "longo prazo" é aqui muito curto pois o modelo só recua até ao desfasamento 2. O que significa que passados mais de 2 periodos sobre a referida variação de x, y já não varia mais, regressando ao valor que tinha antes da variação. É a esta análise de efeitos que se chama de análise de dinâmica sistemática, pois é desprezada a influência dos erros (só a parte principal do modelo é empregue).

Neste caso $s = 2$ apenas, é o chamado comprimento de desfasamento ou desfasamento máximo ou parâmetro de truncatura dos desfasamentos, é baixo; ou seja, o modelo é dinâmico mas não muito. No passado eram frequentemente usados modelos DL(∞) mas com restrições sobre os coeficientes, os chamados modelos de desfasamento geométrico ou de Koyck.

Note-se que o modelo anterior se pode escrever como

$$y_t = B(L)x_t + u_t,$$

com $B(L) = 0.4 + L + 0.6L^2$, com L a representar o habitual operador de desfasamento ($L^k y_t = y_{t-k}$). Os referidos modelos DL(∞) eram obtidos através do rácio de dois polínómios

$$y_t = \frac{B(L)}{A(L)} x_t + u_t,$$

com $D(L) = B(L)/A(L)$ em geral um polinómio infinito.

Estes modelos podem ser ainda mais aumentados considerando que os erros seguem um processo ARMA estacionário, $\phi(L)u_t = \theta(L)\epsilon_t$:

$$y_t = \frac{B(L)}{A(L)}x_t + \frac{\theta(L)}{\phi(L)}\epsilon_t,$$

chamado modelo de função transferência. Estes modelos são muito populares na análise de séries temporais mas não os estudaremos devido a três tipos de justificações:

a) a primeira é pragmática pois é conveniente que os erros sejam ruído branco para que o OLS se mantenha adequado para a estimação;

b) a segunda é conhecida do capítulo anterior: a dinâmica deve estar na parte sistemática dos modelos, não nos seus erros, que não devem ter qualquer tipo de comportamento sistemático;

c) os problemas de identificação destes modelos tornam-nos pouco adequados para fenómenos económicos; as funções de transferência são bons instrumentos apenas para fenómenos físicos.

2.2 Modelos ADL: introdução

Os modelos cujo estudo será privilegiado são os modelos ADL (*autoregressive distributed lag*) que, além de uma componente de desfasamento distribuído (DL), têm também uma parte autoregressiva. Considere-se, como exemplo, o ADL(1,1) com um único regressor (assumido exógeno):

$$y_t = \mu + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2),$$

onde se deve notar que o facto de os erros serem ruído branco é uma característica importante destes modelos. Este modelo também se pode escrever como

$$(1 - \alpha_1 L)y_t = \mu + (\beta_0 + \beta_1 L)x_t + \epsilon_t,$$

isto é,

$$A(L)y_t = \mu + B(L)x_t + \epsilon_t, \tag{1}$$

com $A(L) = 1 - \alpha_1 L$ e $B(L) = \beta_0 + \beta_1 L$, respectivamente.

Para exemplificar a riqueza de efeitos dinâmicos permitida por estes modelos suponha-se, novamente, que damos a x um incremento unitário apenas no período t, mantendo-se $x_{t+1}, x_{t+2}, \ldots, x_{t+j}$ nos mesmos valores, ou seja, que existe apenas uma variação transitória ou temporária unitária, $\Delta x_t = 1$. Silenciando o efeito dos erros, quais são os efeitos desta variação? De forma bastante óbvia, tem-se, ainda no mesmo período, $\Delta y_t = \beta_0$, que é o multiplicador de curto prazo ou de impacto. Como o modelo tem x_{t-1} como explicativa, no período seguinte (e em relação ao período anterior à variação, $t-1$), y vai ainda variar, por essa via via, β_1 unidades. Mas o modelo tem ainda y_{t-1} como explicativa e, no período anterior y variou β_0 unidades; então, por essa via, y vai variar $\alpha_1 \beta_0$. Ou seja, $y_{t+1} - y_{t-1} = \beta_1 + \beta_0 \alpha_1$. No período seguinte, o canal dos desfasamentos distribuídos já não tem qualquer contribuição mas y vai continuar a a variar porque variou no período anterior: $y_{t+2} - y_{t-1} = \beta_0 \alpha_1^2 + \beta_1 \alpha_1, \ldots$.

Uma forma alternativa de chegar aos mesmos resultados é bastante mais sofisticada e emprega operações com os polinómios de desfasamento. Retomando a equação (1) e resolvendo-a em ordem a y_t, multiplicando ambos os membros pelo inverso do seu polinómio autoregressivo,

$$y_t = A(L)^{-1}\mu + A(L)^{-1}B(L)x_t + A(L)^{-1}\epsilon_t,$$

que simplificamos algo grosseiramente retirando a componente em ϵ_t pois só estamos interessados nos efeitos de dinâmica sistemática,

$$y_t = A(1)^{-1}(\mu) + (1 - \alpha_1 L)^{-1}B(L)x_t,$$

onde o polinómio a operar sobre μ passou a ser calculado no ponto 1 ($A(1)$) pois todos os valores desfasados de μ são iguais a ele próprio, dado ser uma constante. Ou seja, invertendo o polinómio autoregressivo (veja-se o apêndice) tem-se

$$\begin{aligned} y_t &= A^{-1}(1)\mu + (1 + \alpha_1 L + \alpha_1^2 L^2 + \ldots)(\beta_o + \beta_1 L)x_t, \\ &= A^{-1}(1)\mu + [\beta_0 + (\beta_1 + \alpha_1\beta_0)L + (\beta_1\alpha_1 + \alpha_1^2\beta_0)L^2 + \ldots]x_t, \\ &= A^{-1}(1)\mu + \beta_0 x_t + (\beta_1 + \alpha_1\beta_0)x_{t-1} + (\beta_1\alpha_1 + \alpha_1^2\beta_0)x_{t-2} + \ldots, \end{aligned}$$

onde se reconhecem exactamente os mesmos multiplicadores que já tínhamos obtido. Genericamente estes *multiplicadores dinâmicos* podem ser representados com derivadas parciais, $\partial y_{t+j}/\partial x_t$, pois demos uma variação unitária apenas a x_t, mantendo todos os restantes nos mesmos valores. Vistos como função do desfasamento, j, estes multiplicadores são chamados de função de resposta a impulsos (ou de impulso-resposta) pois dão-nos a resposta de y a um impulso único, transitório e unitário, de x. É usual apresentar a sua representação gráfica como forma simples e rápida de perceber a forma como estes efeitos dinâmicos se distribuem no tempo.

Multiplicador de longo prazo. Tal como nos modelos DL(s), tem-se também um multiplicador total, ou de longo prazo, ou de equilíbrio que, em geral, é dado pela soma da série

$$\mathrm{MLP} = \sum_{j=0}^{\infty} \frac{\partial y_{t+j}}{\partial x_t},$$

que mede, portanto, o efeito acumulado total em y da variação transitória unitária de x. Antes de aplicar o conceito ao modelo ADL anterior, repare-se que no modelo DL(2) da subsecção anterior esse multiplicador pode obter-se simplesmente como $B(1) = 0.4 + 1 + 0.6 = 2$. No caso do ADL com que temos vindo a trabalhar

$$\sum_{j=0}^{\infty} \frac{\partial y_{t+j}}{\partial x_t} = \beta_0 + (\beta_0\alpha_1 + \beta_1) + (\beta_0\alpha_1^2 + \beta_1\alpha_1) + \ldots.$$

Desde que $|\alpha_1| < 1$ a série é convergente e o multiplicador total, que será representado com λ é dado por

$$\lambda = \frac{\beta_0 + \beta_1}{1 - \alpha_1},$$

resultado que se poderá demonstrar facilmente considerando a série que se inicia no segundo termo acima como uma série geométrica de razão α_1 e somando a sua soma com o primeiro termo, β_0. Ou seja, uma expressão que se revelará muito útil é, portanto,

$$\mathrm{MLP} = \lambda = \frac{B(1)}{A(1)}.$$

O multiplicador de longo prazo pode ainda ser calculado de forma muito diferente, e tem também uma interpretação bastante diversa da que lhe foi dada até agora. Suponha-se que a economia (ou o subsistema em análise) se encontra num hipotético estado de equilíbrio estacionário (estático) em que todas as variáveis se encontram no seu estado de equilíbrio, representado pela sua média no caso estacionário:

$$y^* = \mathrm{E}(y_t), \quad x^* = \mathrm{E}(x_t), \quad \text{e} \quad \epsilon^* = \mathrm{E}(\epsilon_t) = 0.$$

Tem-se, então

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

ou seja, a solução de equilíbrio estático (estacionário) é

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

Se x tiver uma variação permanente de uma unidade, no novo estado de equilíbrio e desde que o sistema seja estável ter-se-á

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

ou seja, y^* terá variado numa magnitude que é o valor do MLP. Assim, neste caso, o multiplicador de longo prazo representa a variação de equilíbrio de y, de longo prazo, correspondente a uma variação permanente unitária do valor de equilíbrio de x. Ou seja, como nota Hamilton (1994), neste caso o multiplicador de longo prazo é visto como

$$\text{MLP} = \lim_{j\to\infty}\left(\frac{\partial y_{t+j}}{\partial x_t} + \frac{\partial y_{t+j}}{\partial x_{t+1}} + \ldots + \frac{\partial y_{t+j}}{\partial x_{t+j}}\right).$$

Condições de estabilidade. A condição $|\alpha_1| < 1$, que se empregou na primeira derivação do MLP, é uma condição de estabilidade da equação às diferenças de primeira ordem. De facto, com $|\alpha_1| < 1$ os efeitos da variação transitória de x acabam por se tornar desprezíveis pois tendem para zero, e a economia (ou o seu sub-sistema em análise) voltará para um novo ponto de equilíbrio, com um valor finito de y (mesmo que a variação de x seja permanente).

Com $|\alpha_1| > 1$ o (sub-)sistema económico seria explosivo. Por exemplo, se $\beta_0 > 0, \beta_1 > 0$ e $\alpha_1 > 1$, y não convergiria para nenhum valor finito mas tenderia para $+\infty$ à medida que $t \to \infty$.

O caso $\alpha_1 = 1$ (e $\alpha_1 = -1$) também parece pouco razoável: passado um qualquer número de períodos sobre a variação temporária de x, y continuará a variar $\beta_0+\beta_1$ unidades, não convergindo para nenhum valor finito (e com $\alpha_1 = -1$ teria até um comportamento oscilatório).

Em suma, constatamos que a análise de estabilidade do (sub-)sistema só tem a ver com a parte autoregressiva do modelo.

Faça-se a generalização desta análise considerando um ADL geral, de ordens r e s, mas com uma única variável exógena

$$y_t = \mu + \alpha_1 y_{t-1} + \ldots + \alpha_r y_{t-r} + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_s x_{t-s} + \epsilon_t,$$

ou seja,

$$A(L)y_t = \mu + B(L)x_t + \epsilon_t,$$

com

$$A(L) = 1 - \alpha_1 L - \ldots - \alpha_r L^r \quad \text{e} \quad B(L) = \beta_0 + \beta_1 x_t + \ldots + \beta_s x_{t-s},$$

os polinómios autoregressivo e de desfasamento distribuído, respectivamente.

A condição necessária mas não suficiente de estabilidade é que

$$\sum_{i=1}^{r} \alpha_i < 1,$$

ou seja, que $\sum_{i=1}^{r} \alpha_i - 1 < 0$, ou ainda, de forma equivalente, que

$$-A(1) < 0,$$

que, no modelo ADL(1,1) com que trabalhámos anteriormente, corresponde simplesmente a $\alpha_1 < 1$. Na prática, será esta a condição que nos preocupará pois, em economia, é a mais importante.

A condição necessária e suficiente de estabilidade já é conhecida da análise de séries temporais e envolve as raízes da equação característica associada com o polinómio autoregressivo $A(L)$: os valores de z: $A(z) = 0 \Leftrightarrow 1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_r z^r = 0$, representados com z_i, $i = 1, 2, \ldots, r$, ou seja, as raízes de $A(z)$ devem estar fora do circulo unitário, isto é, sejam reais ou complexas o seu módulo deve ser superior a um $(|z_i| > 1, \forall i)$. Na prática, como referido, bastará que nos preocupemos com a satisfação da condição necessária, muito mais relevante em economia. Desde que o (sub-)sistema seja estável, tem-se neste caso geral

$$\text{MLP} = \lambda = \frac{B(1)}{A(1)} = \frac{1 + \beta_0 + \ldots + \beta_s}{1 - \alpha_1 - \ldots - \alpha_r}.$$

2.3 Breves observações

O que foi visto anteriormente também é adaptável a outros casos, a outras forma funcionais. Por exemplo, se todas as variáveis do modelo tiverem sido logaritmizadas previamente, temos elasticidades, de curto e de longo prazos, no lugar dos multiplicadores. E no caso dos modelos log-lin, teremos as semi-elasticidades.

Há ainda outros conceitos de dinâmica sistemática que não veremos, como os dos pesos dos desfasamentos, os dos desfasamentos médio e mediano, que são indicadores da velocidade dos ajustamentos, os multiplicadores interim, etc. . Uma referência útil sobre estes tópicos é Hendry (1995).

3 Estimação OLS e transformações lineares[1]

Um problema importante, tanto para os modelos DL como para os ADL, sobretudo se tiverem um grande número de desfasamentos, é o da estimação precisa dos seus coeficientes. A validade dos métodos de inferência usuais também requer que os erros dos modelos sejam iid, o que é uma hipótese "heróica" para os modelos DL e, regra geral, leva a preferir os modelos ADL. Todavia, nalguns casos pode pretender-se trabalhar com os mais simples modelos DL – que, por si sós, não excluem a hipótese de exogeneidade estrita — e a abordagem que se segue é útil para ambos os tipos de modelos.

3.1 Problemas de colinearidade

Como as variáveis macroeconómicas se comportam com enorme inércia, variando muito lentamente de um período para o seguinte, elas encontram-se regra geral muito autocorrelacionadas (positivamente): as correlações entre qualquer x_t e os seus valores desfasados (e entre estes) devem ser elevadas (o mesmo acontecendo com os desfasamentos de y_t). Daqui decorrem, como se sabe, problemas sérios de colinearidade: a precisão com que são estimadas as estruturas de desfasamentos é, em geral, reduzida. Por outro lado, as probabilidades de cometer erros de tipo II quando se efectuam testes de significância individuais dos coeficientes tendem a vir inflacionadas, precisamente devido à inflação dos denominadores dos rácios-t. Ou seja, a potência destes testes tende a ser reduzida. Assim, por exemplo, caso a escolha do número de desfasamentos se baseie neles, é bastante provável que se chegue a modelos subparametrizados.

Uma consequência adicional da colinearidade é a da incerteza acrescida sobre os coeficientes: ao excluir um desfasamento com base na sua insignificância aparente, as estimativas dos coeficientes dos desfasamentos não excluídos alteram-se substancialmente, provocando no investigador grande incerteza sobre a estrutura dos efeitos dinâmicos.

Apesar destes problemas, pode mostrar-se que, em geral, o multiplicador de longo prazo continua a poder a ser estimado com bastante precisão. Por outras palavras, o estimador desse multiplicador não é, geralmente, afectado pelos prob-

[1] O material desta secção e o da seguinte baseia-se fortemente em Lopes (1998). Por vezes, transcrever-se-á mesmo esse texto, que não ficou desactualizado, mas indicar-se-ão também as suas fontes.

lemas decorrentes da colinearidade. Em Davidson e MacKinnon (1993, pp. 673-4) pode encontrar-se a justificação para este facto.

Todavia, frequentemente o principal interesse do investigador reside na estimação precisa da estrutura de desfasamentos, dos vários coeficientes de desfasamento e não apenas na do multiplicador total ou de longo prazo. No que se segue abordar-se-á este problema empregando transformações lineares. Uma abordagem alternativa, que emprega 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).

3.2 Transformações lineares[2]

A abordagem privilegiada baseia-se em transformações lineares do modelo original, isto é, na reparametrização (sem imposição de restrições) do modelo inicial com base em transformações lineares. Para fazer esse estudo, convirá começar por salientar um resultado segundo o qual o estimador OLS de modelos de regressão linear é invariante a transformações lineares. Ou seja: 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 a partir do modelo reparametrizado pode ser vantajosa por permitir atenuar os problemas de colinearidade do modelo original.

Considere-se o modelo de regressão linear com k regressores, $y_t = x_t'\beta + u_t$, 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

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

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

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

ou seja,

$$y_t^* = x_t^{*\prime}\beta^* + u_t,$$

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 ($x_t^{*\prime}$)

[2] Este ponto baseia-se em Hendry (1995).

e o erro inicial (u_t). É indiferente começar por estimar β e obter o estimador de β^* indirectamente (com $\tilde{\beta}^* = H^{-1}(\hat{\beta} - \gamma)$), ou estimar directamente β^* do modelo transformado (com $\hat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$): tanto as estimativas dos coeficientes como as matrizes de covariâncias dos estimadores são idênticas.

Com efeito,

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

Por outro lado, também se mostra facilmente que

$$\mathrm{Var}(\hat{\beta}^*|X) = \mathrm{Var}(\tilde{\beta}^*|X),$$

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 (ou quase) entre os regressores, sendo o problema inicial da colinearidade eliminado através das transformações lineares. Daqui conclui Hendry (1995, p. 277) que "*a colinearidade não é uma propriedade do conjunto de regressores mas sim da parametrização do modelo*".

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

Para ilustrar as potencialidades das transformações lineares no caso dos modelos DL, comecemos por considerar um modelo DL(1) simples, que nem inclui termo independente,

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

e onde se supõe que, como é usual com variáveis macro-económicas, x_t se encontra muito autocorrelacionada (embora possa ser estacionária). Neste contexto, é muito provável que:

- os estimadores dos coeficientes tenham desvios-padrões (variâncias) muito elevados(as);
- como consequência, um dos coeficientes (ou até ambos) pode(m) não aparecer como estatisticamente significativo(s), o que levará à sua exclusão indevida do modelo;

- consequentemente, a exclusão desse regressor levará a uma grande alteração na estimativa do coeficiente do outro [3], provocando grande incerteza no investigador.

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

$$y_t = \delta_0 \Delta x_t + (\delta_0 + \delta_1) x_{t-1} + u_t, \tag{3}$$

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 (e portanto do erro padrão, se). Daqui também decorre que se pode efectuar imediatamente um teste de significância sobre esse multiplicador, sem necessidade de qualquer cálculo adicional.

Da subsecção anterior sabemos que as estimativas OLS serão exactamente iguais às do modelo (2). Todavia, as vantagens da reparametrização justificam a utilização preferencial de (3).

Exemplo empírico. Como pequeno mas ilustrativo exemplo das vantagens das reparametrizações lineares retomem-se os dados sobre a função consumo (privado) para a economia portuguesa e o modelo muito simples

$$LCP_t = \alpha + \delta_0 LRD_t + \delta_1 LRD_{t-1} + u_t,$$

onde, recorde-se, LCP e LRD representam os logaritmos do consumo privado e do rendimento disponível, respectivamente. Estimado com dados de 1960 a 1995, este modelo vem

$$\begin{array}{rlll} \widehat{LCP}_t = & 0.526+ & 0.556 LRD_t + & 0.323 LRD_{t-1} \\ & (0.163) & (0.296) & (0.283) \end{array}$$

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

ou seja, ao nível habitual de 5%, nenhum dos coeficientes individuais é estatisticamente significativo, com os rácios-t iguais a 1.878 e 1.140, respectivamente para δ_0 e δ_1 (os valores-p dados pelo TSP são, respectivamente, 0.069 e 0.263). A estimativa da elasticidade de longo prazo (ELP) é, portanto, 0.879 (= 0.556 + 0.323) e, para efectuar o teste da sua significância, será necessário recorrer à matriz de covariâncias estimada do estimador OLS, para retirar a estimativa da covariância entre os dois estimadores.

Se procedermos de forma ingénua e automática, retirando do modelo LRD_{t-1} por ser o regressor menos significativo, obtemos

$$\begin{array}{rll} \widehat{LCP}_t = & 0.416+ & 0.895 LRD_t, \\ & (0.130) & (0.021) \end{array}$$

com o único regressor agora muito "altamente significativo" (o rácio-t é 42.87). Note-se, contudo, que a estimativa da elasticidade de curto prazo se alterou muito, provocando incerteza. Por outro lado, a mesma estimativa estima também a ELP e esta pouco mudou. O facto de os parâmetros de longo prazo poderem ser bem estimados com modelos estáticos — o que parece não ter sentido — é um assunto que será retomado no último capítulo.

A reparametrização linear produz

$$\begin{array}{rlll} \widehat{LCP}_t = & 0.526+ & 0.556 \Delta LRD_t + & 0.879 LRD_{t-1}, \\ & (0.163) & (0.296) & (0.025) \end{array}$$

suavizando substancialmente os problemas iniciais de colinearidade: agora só o coeficiente estimado de ΔLRD_t não aparece como estatisticamente significativo ao nível de 5%, com um rácio-t de 1.878, exactamente igual ao inicial de LRD_t (tal como acontece com os coeficientes, como era esperado, pois representam o mesmo, neste caso, a elasticidade de curto prazo). A estimativa da ELP é obtida imediatamente, 0.879, e imediatamente também se verifica que é altamente significativa, com um rácio-t de 34.71; ou seja, a precisão da sua estimação parece ser enorme. As vantagens desta parametrização parecem ser bem claras. E até um eventual impulso para excluir algum regressor vem mitigado pelo facto de o regressor ΔLRD_t ser quase estatisticamente significativo ao nível usual de 5%.

3.4 Generalização: aplicação aos modelos DL(s)

Mais geralmente, considere-se o modelo DL(s), sem termo independente, para simplificar

$$\begin{aligned} y_t = & \ \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_s x_{t-s} + u_t \\ = & \ B(L) x_t + u_t, \end{aligned}$$

com $B(L) = \beta_0 + \beta_1 L + \ldots + \beta_s L^s$ um polinómio de ordem s em L. A transformação linear a operar aqui requer a apresentação prévia de uma proposição sobre decomposições polinomiais.

Proposição. Considere-se um polinómio de ordem p em L, seja

$$a(L) = \sum_{j=0}^{p} a_j L^j = a_0 + a_1 L + a_2 L^2 + ... + a_p L^p,$$

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:

$$a(L) = a(1)L + \sum_{i=0}^{p-1} c_i L^i (1 - L), \tag{4}$$

ou seja,

$$a(L) = a(1)L + c(L)(1 - L), \tag{5}$$

onde $c(L)$ é um polinómio de ordem $p-1$ em L. A demonstração desta proposição, bem como de outra que também pode ser útil, pode ser vista no apêndice.

Aplicando este resultado ao modelo DL(s) tem-se:

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

onde $\delta_0 = \beta_0$ (ou seja, o multiplicador de curto prazo não é afectado), $\delta_i = -\sum_{j=i+1}^{s} \beta_i$, $i = 1, 2, ..., s-1$ e $\Delta = 1 - L$. A estimação deste modelo é equivalente à do modelo inicial pois a reparametrização é linear, isto é, as estimativas dos coeficientes são exactamente as mesmas. Todavia, este último modelo oferece várias vantagens:

a) permite-nos obter imediatamente uma estimativa do multiplicador de longo prazo (como coeficiente de x_{t-1})[4] e da variância do seu estimador, ou seja, do erro-padrão;

b) como consequência, podemos efectuar imediatamente um teste à significância desse parâmetro, bem como construir facilmente um intervalo de confiança para ele e, sobretudo,

c) todos os problemas de colinearidade do modelo inicial 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 desse modelo.

No modelo reparametrizado, com excepção do multiplicador de curto prazo, os coeficientes iniciais vêm misturados. Se se pretender recuperá-los, basta resolver as equações que os ligam aos do modelo reparametrizado.

Neste caso simples considerou-se apenas uma única variável (assumida como) exógena. Mas a aplicação deste tipo de reparametrização a modelos com várias dessas variáveis é ainda mais recomendável. Com efeito, embora em geral as variáveis em níveis se encontrem muito fortemente correlacionadas, o mesmo não acontece com as suas primeiras diferenças. A utilização da proposição acima sobre cada um dos polinómios em L permitirá, em princípio, suavizar os problemas de colinearidade, uma vez que só o valor desfasado um período de cada variável em nível permanecerá no modelo; todos os restantes regressores serão obtidos por diferenciação das variáveis originais e estarão, portanto, muito menos correlacionados entre si e com essas variáveis em nível que os regressores originais.

4 O modelo ADL(1,1) e o modelo de correcção de erros

Nesta secção visam-se essencialmente dois objectivos:

[4] Existem outros resultados semelhantes que fazem aparecer o multiplicador de longo prazo como coeficiente de outra variável, x_t ou x_{t-s}, por exemplo. A opção é quase matéria de gosto pessoal mas é mais comum fazê-lo aparecer em x_{t-1}.

i) introduzir os modelos ADL[5] como instrumentos importantes da modelação econométrica causal de séries temporais, estudando um dos seus casos particulares mais simples, e

ii) apresentar o modelo de correcção de erros (MCE[6]), preparando o caminho para o estudo a empreender mais adiante.

Como se verá mais à frente, este tipo de modelo será privilegiado na análise empírica multivariada de séries temporais não estacionárias mas que, diferenciadas uma vez, ficam estacionárias (integradas de ordem 1, I(1)).

4.1 A generalidade do modelo ADL(1,1)

Visa-se aqui 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:

$$y_t = \mu + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma^2) \tag{6}$$

a satisfazer a usual condição de estabilidade, $|\alpha_1| < 1$. O estudo deste modelo tem particular importância porque:

a) 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 pelo 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);

b) os resultados obtidos para o ADL(1,1) também são válidos, com modificações e/ou adaptações, para o modelo mais geral, o 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 ADL(1,1), apenas impondo restrições sobre os seus coeficientes [7]:

[5] Nalguma literatura chamados ARDL.

[6] David Hendry, o principal pioneiro na utilização destes modelos, prefere chamar-lhes "modelos de correcção de equilíbrio" (e.g. Hendry, 2010), embora o acrónimo seja o mesmo. A sua ascendência, contudo, remonta ao trabalho de A. William Phillips nos anos 50.

[7] Esta apresentação baseia-se exclusivamente em Hendry (1995).

1) a equação de regressão simples, estática, $y_t = \mu + \beta_0 x_t + \epsilon_t$, 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 com $\alpha_1 = \beta_0 = 0$;

5) o (antigamente muito 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), que é o modelo 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), $y_t = \mu + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$, obtém-se com $\alpha_1 = 0$;

8) o modelo que Hendry (1995) 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 $\beta_0 = 0$.

No capítulo 7 de Hendry (1995) são investigadas as propriedades de todos estes modelos quando o processo de geração de dados (DGP) ou "verdadeiro modelo" é o 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) $\Delta y_t = (\alpha_1 - 1)(y_{t-1} - \lambda_0 - x_{t-1}) + \beta_0 \Delta x_t + \epsilon_t$, onde o multiplicador de longo prazo é unitário.

4.2 Transformações lineares do modelo ADL(1,1)

Para compreender a flexibilidade do ADL(1,1) bem como as vantagens de algumas das suas reparametrizações lineares, considere-se uma situação ideal, de longo prazo [8], de equilíbrio estacionário, em que deixou de haver qualquer variação das variáveis, encontrando-se estas nos seus valores de equilíbrio, as suas médias. Recorde-se que na subsecção 2.2 se obteve $y^* = \mu/(1-\alpha_1) + [(\beta_0 + \beta_1)/(1-\alpha_1)]x^*$ ou $y^* = \lambda_0 + \lambda_1 x^*$, com $\lambda_0 = \mu/(1-\alpha_1)$ e $\lambda_1 = (\beta_0 + \beta_1)/(1-\alpha_1)$ a representar o multiplicador total, ou de longo prazo ou de equilíbrio [9]. É esta a base para proceder às referidas reparametrizações.

Em primeiro lugar, subtraindo y_{t-1} a ambos os membros do ADL(1,1), de (6) obtém-se

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

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"*:

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

onde se deve notar 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}. Como se verá adiante, será esta a forma a privilegiar para efeitos de estimação.

Se, por outro lado, somarmos e subtrairmos $\beta_1 x_t$ ao lado direito de (7), obtemos

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

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}.

---

[8] Ideal e de longo prazo porque é necessário que não ocorra nenhum choque, nem sob a forma de um valor diferente de zero para o erro nem sob o de uma variação de x, durante bastante tempo, para que a economia convirja para esse estado. De longo prazo também porque se conhecermos o valor de x num futuro longínquo, por exemplo, daqui a 20 anos, podemos inseri-lo na equação e obter o correspondente valor de y (de equilíbrio) nesse período.

[9] 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.

A forma mais interessante (a “forma MCE”) obtém-se de (8), colocando $(\alpha_1 - 1)$ em evidência:

$$\Delta y_t = (\alpha_1 - 1)\left[y_{t-1} - \frac{\mu}{1-\alpha_1} - \frac{\beta_0+\beta_1}{1-\alpha_1}x_{t-1}\right] + \beta_0\Delta x_t + \epsilon_t, \tag{10}$$

ou seja,

$$\Delta y_t = (\alpha_1 - 1)[y_{t-1} - \lambda_0 - \lambda_1 x_{t-1}] + \beta_0\Delta x_t + \epsilon_t, \tag{11}$$

equação que será retomada mais adiante.

Por outro lado, a “forma homogénea” (por parecer que o multiplicador de longo prazo é unitário), obtém-se de (8) subtraindo e somando $(\alpha_1 - 1)x_{t-1}$ é dada por

$$\Delta y_t = \mu + (\alpha_1 - 1)[y_{t-1} - x_{t-1}] + \beta_0\Delta x_t + (\beta_0 + \beta_1 + \alpha_1 - 1)x_{t-1} + \epsilon_t, \tag{12}$$

onde se deve notar o termo adicional em x_{t-1} para “quebrar a homogeneidade” (ou seja, não há realmente a imposição de qualquer restrição).

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 $\gamma = (1-\alpha_1)^{-1}$:

$$y_t = \gamma\mu - \gamma\alpha_1\Delta y_t + \gamma(\beta_0 + \beta_1)x_t - \gamma\beta_1\Delta x_t + \gamma\epsilon_t, \tag{13}$$

que tem a vantagem de o multiplicador de longo prazo ser dado imediatamente pelo coeficiente de x_t. Tem, no entanto, uma enorme desvantagem em relação às anteriores: 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. Logo, com excepção da forma de Bewley[10], o estimador OLS fornece com todas elas as mesmas estimativas para os parâmetros, e todas elas têm a mesma capacidade para explicar os dados.

4.3 O modelo de correcção de erros

Retome-se a equação (11) (ou (10)) e repare-se que o termo dentro de parêntesis rectos mede o erro ou desvio de equilíbrio ou desequilíbrio do período anterior, isto

[10] Todavia, pode mostrar-se que se se empregarem como instrumentos os regressores do ADL(1,1), o método IV fornece as mesmas estimativas que o OLS; veja-se Banerjee et al. (1993, pp. 55 e seguintes).

é, a extensão em que o equilíbrio não foi satisfeito nesse período. Essas equações dão-nos relações de dinâmica de curto prazo, dado 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. Mas a dinâmica de curto prazo é também conduzida pela relação de equilíbrio (de longo prazo), que entra nela através desse desequilíbrio ou erro de equilíbrio, que aparece a explicar a evolução de curto prazo da variável y. Por conseguinte, mesmo que não haja quaisquer tipo de choques durante um período de tempo prolongado, com $\epsilon_t \equiv 0$ e $\Delta x_t \equiv 0$ para muitos t, Δy 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) 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 [11].

Para se perceber a designação atribuída ao modelo (de correcção de erros), comece-se por notar 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 $y_{t-1} > \lambda_0 + \lambda_1 x_{t-1}$. Então, como $\alpha_1 - 1 < 0$, *ceteris paribus*, no período seguinte (t) y terá tendência para retornar ao valor de equilíbrio pois $\Delta y_t < 0$; ou seja, na ausência de choques ter-se-á $\Delta y_t < 0$, para que se regresse à situação de equilíbrio. Por exemplo, supondo que y representa o consumo privado e x o rendimento disponível, corresponde a supôr que, em $t-1$, as famílias terão consumido para além do nível de equilíbrio permitido pelo rendimento, tendo recorrido ao endividamento para esse efeito. Então, o funcionamento do mecanismo tenderá a fazer com que a variação do consumo no período seguinte seja negativa, para haver retorno ao nível de equilíbrio; isto é, para pagar os juros e amortizar a dívida, as famílias terão que reduzir o seu nível de consumo.

Assim, o modelo incorpora um efeito de *feedback* negativo que tenta corrigir desequilíbrios de períodos anteriores para restabelecer a relação de equilíbrio (de

[11] E daí por vezes o modelo ser chamado de modelo com mecanismo de correcção de erros. Isto é, o acrónimo MCE tanto pode designar o modelo como o mecanismo.

longo prazo). A equação de correcção de erros é basicamente uma equação de ajustamento dinâmico de curto prazo, mas em que esse ajustamento também é guiado pela relação de equilíbrio (de longo prazo).

Note-se, então que, para testar a presença de um mecanismo de correcção de erros, efectua-se o teste de

$$H_0 : \alpha_1 - 1 = 0 \Leftrightarrow -A(1) = 0, \quad vs. \quad H_1 : \alpha_1 - 1 < 0 \Leftrightarrow -A(1) < 0,$$

correspondendo a rejeição da hipótese nula a evidência no sentido da existência desse tipo de efeito. Repare-se que este teste também pode ser visto como testando a condição necessária mas não suficiente de estabilidade.

Relativamente à estimação, se λ_0 e o multiplicador de longo prazo forem conhecidos, isto é, dados pela teoria económica, a estimação de (11) pode ser efectuada imediatamente pelo OLS. Em geral, contudo, essa informação não se encontra disponível. Então, reparando na forma como os parâmetros aparecem em (11) ou em (10), 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:

a) por fornecer imediatamente uma estimativa para o coeficiente de ajustamento de curto prazo $(\alpha_1 - 1)$;

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

c) por permitir obter facilmente uma estimativa para o multiplicador de longo prazo, através da forma que vimos, e

d) por ser preferível, por exemplo, ao ADL(1,1) inicial 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 um historial significativo de sucesso empírico na modelação de vários fenómenos macroeconómicos observados ao longo do tempo, sendo os exemplos mais importantes os casos das funções consumo e de procura de moeda. De resto, como veremos no último capítulo do programa, 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.

5 O MCE com variáveis não estacionárias

A interpretação estrita anterior do MCE como um modelo que contém uma relação de equilíbrio entre variáveis estacionárias é muito limitada em macroeconomia, onde muitas variáveis são não estacionárias. Nesta subsecção salienta-se que o MCE continua a ser interessante para variáveis não estacionárias pois, além de continuar a ser um modelo em que a dinâmica de curto é conduzida por uma relação de equilíbrio de longo prazo entre variáveis não estacionárias, o MCE incorpora também uma solução de crescimento estacionário ("*steady growth*"). Por outro lado, ainda, o MCE permite incorporar efeitos de crescimento: o nível de uma variável a depender do de outra mas também da taxa de crescimento desta outra.

Recorre-sa a uma convenção muito frequente, que é a de representar com maiúsculas as próprias variáveis em nível, e com minúsculas os seus logaritmos: $y = \log(Y)$ e $x = \log(X)$. Recorde-se, também, que as primeiras diferenças de variáveis previamente logaritmizadas constituem uma aproximação à sua variação relativa, isto é, à sua taxa de crescimento:

$$\Delta y_t = \Delta \log(Y_t) \approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}.$$

Para manter a análise simples, considerem-se apenas duas variáveis não estacionárias mas que se tornam estacionárias após uma diferenciação (integradas de ordem 1, I(1)). Para y_t considere-se

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

um passeio aleatório com deriva μ(com uma raiz unitária), que também representaremos com

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

Considerando que também a segunda variável segue o mesmo tipo de processo, podemos usar a representação

$$\Delta y_t = g_y + \epsilon_t \quad \text{e} \quad \Delta x_t = g_x + \eta_t,$$

com $\eta_t \sim iid(0, \sigma^2_\eta)$.

O ADL(1,1) pode continuar a ser reparametrizado sob a forma de MCE, $\Delta y_t = (\alpha_1 - 1)[y - \lambda_0 - \lambda_1 x]_{t-1} + \beta_0 \Delta x_t + \epsilon_t$, com o desequilíbrio do período anterior bem explicitado, a representar agora um desequilíbrio entre variáveis não estacionárias, correspondente ao equilíbrio de longo prazo $y = \lambda_0 + \lambda_1 x$, onde continua a não se usar subescrito pelo facto de a relação ser intertemporal, mas onde não se empregam os asteriscos pois agora as variáveis podem não ter média (constante).

A relação de equilíbrio de longo prazo entre as variáveis originais, não logaritmizadas, é

$$\begin{aligned} Y = & \exp(\lambda_0 + \log X^{\lambda_1}) \\ = & \exp(\lambda_0)\exp(\log X^{\lambda_1}) \\ = & k_0 X^{\lambda_1}, \end{aligned}$$

com $k_0 = \exp(\lambda_0)$.

Se diferenciarmos o ADL(1,1), tem-se

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

e, passando aos valores esperados (que existem tanto para Δy_t como para Δx_t por serem estacionárias), tem-se

$$g_y = \alpha_1 g_y + (\beta_0 + \beta_1) g_x$$

isto é,

$$g_y = \frac{\beta_0 + \beta_1}{1 - \alpha_1} g_x,$$

ou $g_y = \lambda_1 g_x$, que é a solução de crescimento estacionário (*steady growth*): a taxa média de crescimento de Y é proporcional à taxa média de crescimento de X e a constante de proporcionalidade é a elasticidade de longo prazo, λ_1.

Por outro lado, voltando ao ADL(1,1) mas substituindo as variáveis estacionárias pela sua média tem-se, desde logo

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

onde a igualdade já não é precisa pois trata-se já de uma aproximação. Continuando da mesma forma, faça-se:

$$y_{t-1} \approx y_t - g_y \quad \text{e} \quad x_{t-1} \approx x_t - g_x.$$

Substituindo na equação acima, tem-se

$$y_t \approx \mu + \alpha_1 y_t - \alpha_1 g_y + \beta_0 x_t + \beta_1 x_t - \beta_1 g_x,$$

ou seja,

$$y_t \approx \frac{\mu}{1-\alpha_1} + \lambda_1 x_t + \frac{-\alpha_1 g_y - \beta_1 g_x}{1-\alpha_1}.$$

Como $\beta_1 = \lambda_1(1-\alpha_1) - \beta_0$ e usando a solução de crescimento estacionário na equação anterior e simplificando, obtém-se

$$y_t \approx \frac{\mu}{1-\alpha_1} + \lambda_1 x_t + \frac{\beta_0 - \lambda_1}{1-\alpha_1} g_x,$$

que mostra que o nível de y_t depende não só do de x_t como, ainda, da taxa de crescimento de X no estado de crescimento estacionário, g_x. Note-se que, para que não exista este "efeito de crescimento", para que ele seja nulo, é necessário que $\beta_0 = \lambda_1$, isto é, a igualdade das elasticidades de curto e de longo prazos, uma restrição muito forte, raramente suportada pelos dados.

6 O modelo ADL(r, s) e o MCE

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

6.1 A reparametrização em MCE dos modelos ADL(r, s)[12]

Em primeiro lugar e obviamente, a proposição sobre decomposições polinomiais que vimos antes também se pode aplicar a polinómios autoregressivos. Assim, se o polinómio de ordem p em L, fôr $a(L) = \sum_{j=0}^{p} a_j L^j$ com $a_0 = 1$, isto é, $a(L) = 1 - a_1 L - a_2 L^2 - ... - a_p L^p$ e definindo os coeficientes $f_0 = a_0 = 1$, $f_i = -\sum_{j=i+1}^{p} a_j$, $i = 1, ..., p-1$ e $f_p = 0$, pode escrever-se:

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

ou seja,

$$a(L) = a(1)L + (1-L)f(L),$$

[12] Este ponto baseia-se fortemente em Boswijk (1994).

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

Considere-se agora um modelo ADL(r,s) com observações de uma única variável (assumida como) exógena[13] e sem termo independente (para simplificar e dado que, para os efeitos pretendidos, a presença deste é irrelevante):

$$y_t = \sum_{i=1}^{r} \alpha_i y_{t-i} + \sum_{i=0}^{s} \beta_i x_{t-i} + \epsilon_t, \quad \epsilon_t \sim \mathrm{iid}(0,\sigma^2), \tag{14}$$

ou seja,

$$A(L)y_t = B(L)x_t + \epsilon_t, \tag{15}$$

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

Usando a decomposição do polinómio autoregressivo, podemos escrever

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

onde $D(L)$ tem uma ordem inferior numa unidade à inicial

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

Por outro lado, usando a decomposição do polinómio $B(L)$, 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 (15) obtém-se

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

isto é (recordando novamente que $(1-L) = \Delta$),

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

[13] Por agora pode considerar-se que a variável é contemporâneamente exógena ou pré-determinada; mais tarde, no entanto, será precisada a noção de exogeneidade que é necessária aqui.

da qual se obtém facilmente a forma de Bardsen:

$$\Delta y_t = -A(1)y_{t-1} + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + B(1)x_{t-1} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t, \tag{16}$$

um ADL$(r-1, s-1)$ em Δy_t e Δx_t aumentado com y_{t-1} e x_{t-1}, os dois únicos regressores em nível.

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

$$\Delta y_t = -A(1)\left[y_{t-1} - \frac{B(1)}{A(1)}x_{t-1}\right] + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t, \tag{17}$$

ou seja, representando com ϕ o coeficiente de ajustamento, $\phi = -A(1)$ e, como é usual, $\lambda = B(1)/A(1)$ o multiplicador de longo prazo, tem-se

$$\Delta y_t = \phi\left[y_{t-1} - \lambda x_{t-1}\right] + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + \sum_{i=0}^{s-1} \gamma_i \Delta x_{t-i} + \epsilon_t, \tag{18}$$

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

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 (15), que permite obter facilmente a estimativa do multiplicador de longo prazo a partir de $\lambda = B(1)/A(1) = -B(1)/\phi$.

Repare-se novamente nas vantagens de (16), onde a maioria dos regressores são desfasamentos das primeiras diferenças de x_t e de y_t:

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 também imediatamente para efectuar o chamado teste t-MCE (t-*ECM*), isto é, 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 só requer pequenas adaptações. Agora, x_t passa a representar um vector, $k \times 1$, de observações dessas variáveis ($x_t' = [x_{t1}\ x_{t2}\ ...\ x_{tk}]$), com comprimentos de desfasamento dados por $s_1, s_2, ..., s_k$; ou seja, não é necessário que os comprimentos de desfasamento sejam todos iguais. O modelo continuará, no entanto, a ser referido como ADL(r, s), com $s = \max\{s_1, s_2, ..., s_k\}$. Em vez de (14) passamos a ter

$$y_t = \sum_{i=1}^{r} \alpha_i y_{t-i} + \sum_{i=0}^{s} \beta_i' x_{t-i} + \epsilon_t, \quad \epsilon_t \sim \text{iid}(0, \sigma^2), \tag{19}$$

com a ligeira mudança de notação a significar que os β_i's são agora 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 (15) escreve-se agora

$$A(L)y_t = B(L)'x_t + \epsilon_t, \tag{20}$$

onde $B(L) = \sum_{i=0}^{s} \beta_i L^i$ representa agora um vector, $k \times 1$, de polinómios em L, isto é, $B(L)' = [B_1(L)\ B_2(L)\ \ldots\ B_k(L)]$, com $B_j(L), j = 1, \ldots, k$, os polinómios em L das várias variáveis.

Procedendo da mesma forma que anteriormente, chega-se à forma de Bardsen:

$$\Delta y_t = -A(1)y_{t-1} + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + B(1)'x_{t-1} + \sum_{i=0}^{s-1} \gamma_i' \Delta x_{t-i} + \epsilon_t, \tag{21}$$

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:

$$\Delta y_t = \phi[y_{t-1} - \lambda' x_{t-1}] + \sum_{i=1}^{r-1} \delta_i \Delta y_{t-i} + \sum_{i=0}^{s-1} \gamma_i' \Delta x_{t-i} + \epsilon_t. \tag{22}$$

No caso de o modelo ser estável, a solução de equilíbrio estacionário (de longo prazo) é agora dada por

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

onde o símbolo "*" tem o mesmo significado que anteriormente. No caso de as variáveis envolvidas não serem estacionárias convém usar a convenção anterior de omitir os asteriscos. Em qualquer caso, recorde-se que o índice de observação também é omitido.

Saliente-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. Da mesma foram, os problemas iniciais de colinearidade deverão ser mais severos que os dos casos anteriores.

6.2 A modelação do geral para o particular

A melhor estratégia para chegar a um modelo dinâmico adequado parece consistir em iniciar o trabalho de especificação considerando um modelo bastante geral — um ADL(r, s) de ordens elevadas, 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 a estratégia de modelação que é preconizada sobretudo por investigadores britânicos (como David Hendry e Graham Mizon), que é conhecida como modelação "do geral para o particular" (ou "para o específico", em inglês e abreviadamente, GTS) e que está associada ao passado da London School of Economics (LSE) e ao orientador de muitos econometristas desta escola, Dennis Sargan.

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 inicial assumida e deliberada. Esta é uma das características que tem recebido mais críticas pois, muito frequentemente, os modelos mais simples podem ser considerados preferíveis para certas finalidades (por exemplo, para efectuar previsão ou porque são mais fáceis de compreender). No entanto, convém que o modelo final não tenha regressores redundantes.

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 heteroscedasticidade e de autocorrelação dos erros, RESET, de estabilidade dos coeficientes, de previsão, etc.

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 da 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 e que podem levar a enviesamento e inconsistência da estimação? 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 bem como 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, que seja coerente 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, BIC, etc.), pois o processo de simplificação não deverá permitir que esta se deteriore substancialmente. Também os testes de factores comuns (COMFAC) podem ser empregues nesta fase e programas como o Pc-Give incorporam-nos de forma bastante simples.

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[14].

6.3 A abordagem do particular para o geral

A estratégia de modelação do particular para o geral é a abordagem tradicional que predominou até um passado distante, mas que continua ainda hoje na base de muitos 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 possam surgir (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 confirmar ou corroborar as relações dadas pela teoria económica, mas não para a pôr verdadeiramente à prova, testando pelo menos 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

[14] Segundo o "princípio da envolvência", para que um modelo possa ser considerado bom ou adequado, ele deve conseguir explicar também as principais características dos modelos rivais e deve, portanto, permitir rejeitá-los nestes testes mas não ser rejeitado por eles.

inicia a análise por um modelo subparametrizado ($\mathcal{M}_1$), o qual, por esse facto, apresenta sintomas de autocorrelação residual. A "correcção da autocorrelação" (feita com o FGLS ou EGLS, como é usual, admitindo erros AR(1)), conduz-nos ao modelo $\mathcal{M}_2$, mais geral. Por sua vez, suponha-se que este apresenta problemas de instabilidade dos coeficientes, o que nos conduz ao modelo $\mathcal{M}_3$ (por exemplo, introduzindo variáveis artificiais). Mas uma vez que também se descobriram problemas em $\mathcal{M}_2$, 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, saliente-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.

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, como vimos, a estimação de um modelo estático com erros AR(1), tão usada no passado, equivale a estimar o modelo ADL(1,1) mas impondo restrições não-lineares 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 modelação de alguns fenómenos macroeconómicos.

A estratégia de modelação GTS (ou "GETS") não está também isenta de críticas[15]. Em particular, ela é muito exigente em termos da informação inicial a recolher

[15] Veja-se, por exemplo, Faust e Whiteman (1997), que contém uma crítica irónica mas pouco construtiva.

e pode não ser realista. Em particular, é muito difícil que toda a informação relevante possa ser recolhida logo no início do processo; é frequente o investigador deparar-se com problemas resultantes de pormenores inesperados que o podem levar a empregar, pelo menos temporariamente, uma estratégia do simples para o geral. Como nota Lütkepohl (2007), é também uma estratégia pouco adequada para o caso da modelação vectorial (multivariada e multi-equacional).

7 Exemplo empírico

Como ilustração empírica, retomam-se os dados respeitantes à função consumo agregada para a economia portuguesa, de 1965 a 1995 e o modelo que se considerou ser mais do tipo de projecção linear do que modelo estrutural sólido. Recordem-se as convenções empregues para representar as variáveis: $\mathrm{DLC}_t = \Delta LC_t = LCP_t - LCP_{t-1}$ representa a primeira diferença do logaritmo do consumo das famílias, $\mathrm{DLR}_t = \Delta LRD_t = LRD_t - LRD_{t-1}$, a primeira diferença do rendimento disponível, $\mathrm{DINF}_t = INF_t - INF_{t-1}$ a primeira diferença da inflação e $\mathrm{DLS}_t = LSR_t - LSR_{t-1}$ a primeira diferença de um índice de salários reais.

Como os dados são anuais e a amostra é curta, considerou-se suficiente iniciar o trabalho com um ADL(3,3) nos níveis das variáveis, parametrizado sob a forma de Bardsen. Refiram-se algumas ideias adicionais sobre o percurso de simplificação:

a) o resultado final é obtido após inúmeras escolhas que podem ser diferentes entre investigadores. Ou seja, não existe um único caminho ou percurso possível pelo que se pode chegar a especificações finais distintas.

b) Os testes de especificação privilegiados foram os testes de autocorrelação, para tentar garantir que a dinâmica do modelo é suficiente mas também se deu bastante importância aos testes de previsão, sobretudo nas especificações finais. Aliás, a constante do modelo foi retida apesar de não se revelar significativa, não só por ser usual em qualquer modelo mas também pelo facto de a sua remoção implicar uma deterioração significativa da qualidade das previsões[16].

c) Apesar de não se revelar significativa aos níveis usuais, uma variável em nível e desfasada um período, INF_{t-1}, foi retida porque a sua exclusão do modelo

[16] A presença da constante é ainda conveniente para justificar a utilização de um resultado que só veremos no último capítulo.

implicaria a sua exclusão da relação de equilíbrio de longo prazo, o que se considerou ser uma decisão drástica. A justificação completa para a sua inclusão só será vista no capítulo final, de cointegração.

Apesar de reparametrizada sob a forma de Bardsen, a primeira equação (Equation 1) apresenta um panorama quase desolador no que respeita à significância dos regressores, com apenas um significativo ao nível usual de 5% e com muitos dos restantes coeficientes estimados muito longe da significância. Retenha-se, no entanto, a significância global do modelo quase ao nível de 1% e, sobretudo, a ausência de quaisquer problemas visíveis de especificação e, em particular, a insignificância das estatísticas de autocorrelação e do teste RESET. Por razões de espaço, também não se reportam outras estatísticas que, no entanto, serão apresentadas mais adiante (como a AIC e a BIC bem como o $\bar{R}^2$).

```
                                Equation   1
                                ============
                     Method of estimation = Ordinary Least Squares
Dependent variable: DLC
Current sample:  1968 to 1995
         Mean of dep. var. = .032314
     Variance of residuals = .338888E-03
  Std. error of regression = .018409
                 R-squared = .831171
              LM het. test = 2.49818 [.114]
Breusch/Godfrey LM: AR/MA1 = .074112 [.785]
Breusch/Godfrey LM: AR/MA3 = 1.94985 [.583]
           Ramsey's RESET2 = 1.42562 [.258]
           F (zero slopes) = 3.93853 * [.011]
           Estimated    Standard
Variable  Coefficient     Error        t-statistic       P-value
C          .065127       .128018       .508735          [.620]
LCP(-1)   -.287277       .254321      -1.12958          [.281]
LRD(-1)    .261714       .240360       1.08884          [.298]
LSR(-1)    .083553       .087219       .957966          [.357]
INF(-1)   -.818099E-03   .280483E-02  -.291675          [.776]
DLC(-1)    .395331       .356604       1.10860          [.289]
DLC(-2)    .154389       .399052       .386890          [.706]
```

DLR	.380204	.167385	2.27144	*	[.042]
DLR(-1)	-.083260	.218713	-.380681		[.710]
DLR(-2)	-.290114	.202744	-1.43094		[.178]
DLS	.286027	.247846	1.15405		[.271]
DLS(-1)	.027639	.299438	.092304		[.928]
DLS(-2)	-.767700E-02	.263578	-.029126		[.977]
DINF	.875192E-03	.270116E-02	.324006		[.752]
DINF(-1)	.729020E-04	.234304E-02	.031114		[.976]
DINF(-2)	.353931E-03	.141999E-02	.249248		[.807]

Embora formalmente não se devam efectuar tantos testes, refira-se também a insignificância das estatísticas Q^* de ordens 1 e (até) 3. Também os testes CUSUM e CUSUMSQ (não apresentados) são insignificantes, bem como a estatística de Jarque-Bera e um teste para efeitos ARCH de primeira ordem.

Tendo o cuidado de manter a mesma amostra[17], fez-se um primeiro teste-F de simplificação, testando a nulidade conjunta dos coeficientes de todos os regressores (diferenciados e) desfasados 2 períodos, os das variáveis mais afastadas do presente e, portanto, os que provavelmente são menos importantes. A estatística-F assim obtida atingiu o valor de 3.46, com um valor-p (aproximado) de 0.042, o que leva à rejeição da hipótese nula. Não se puderam, portanto, impôr de uma só vez as quatro restrições da exclusão dos desfasamentos de ordem 2.

Observando com atenção as estimativas obtidas para esses coeficientes, nota-se que a origem da rejeição poderá residir em ΔLR_{t-2} que não está muito longe da significância. Removendo do teste anterior a nulidade desse coeficiente passa a obter-se uma $F = 0.76$ com valor$-p$ de 0.538 que agora não permite rejeitar a nulidade dos três coeficientes. Desta forma, as variáveis ΔLC_{t-2}, ΔLS_{t-2} e ΔINF_{t-2} são removidas do modelo.

Procedeu-se em seguida à reestimação do modelo incluindo também a observação de 1967, que anteriormente não era possível devido à insuficiência de observações desfasadas de ΔLS_t. Ainda assim, o único regressor significativo continua a ser ΔLR_t, mas esta equação também não é apresentada.

Em seguida, com base na observação dos rácios-t (o que não é estritamente correcto) testou-se a significância conjunta dos coeficientes de $\Delta LR_{t-1}, \Delta INF_t$ e

[17] Este cuidado é aqui importante pois a exclusão do segundo desfasamento de todas as variáveis liberta uma observação da amostra e a estimação do modelo com as restrições pode ser iniciada em 1967, o que não só não interessa para já como não se deve fazer, para preservar a amostra que fornece a magnitude que entra no cálculo da F.

ΔINF_{t-1} e não se rejeitou a nula conjunta, pelo que essas três variáveis foram excluídas. A equação assim obtida (também não apresentada) continua a não apresentar evidência de problemas de especificação e agora já contém vários regressores estatisticamente significativos aos níveis usuais. Para além da constante e de INF_{t-1}, não excluídas pelas razões já referidas, agora só ΔLS_{t-1} parece estar muito longe da significância, pelo que se excluiu da equação, obtendo-se a "Equation 6".

```
                                    Equation   6
                                    ============

Dependent variable: DLC
Current sample:  1966 to 1995
  Std. error of regression = .014674
                 R-squared = .814592
              LM het. test = .140255 [.708]
Breusch/Godfrey LM: AR/MA1 = .178087E-02 [.966]
Breusch/Godfrey LM: AR/MA3 = 3.53389 [.316]
                 Chow test = 2.27653 [.093]
           Ramsey's RESET2 = .661594 [.426]
           F (zero slopes) = 11.5330 ** [.000]
```

Variable	Estimated Coefficient	Standard Error	t-statistic		P-value
C	.048776	.082447	.591607		[.560]
LCP(-1)	-.266245	.071495	-3.72395	**	[.001]
LRD(-1)	.243360	.065785	3.69935	**	[.001]
LSR(-1)	.095209	.039934	2.38416	*	[.027]
INF(-1)	-.919432E-03	.585899E-03	-1.56927		[.132]
DLC(-1)	.456655	.135524	3.36956	**	[.003]
DLR	.338044	.105065	3.21748	**	[.004]
DLR(-2)	-.247299	.110304	-2.24199	*	[.036]
DLS	.225807	.079575	2.83767	**	[.010]

Repare-se que, com as duas excepções mencionadas, todos os regressores são agora claramente significativos. Mais importante, a equação não parece apresentar quaisquer problemas de especificação, se bem que se preferisse obter uma estatística de Chow (automática, com partição no meio da amostra) mais "saudável".

Em termos de bondade do ajustamento e tomando em consideração o número de parâmetros, o modelo final é superior ao inicial, como mostram as estatísticas do quadro seguinte, novamente todas obtidas com a mesma amostra, de 1968 a 1995, para que a comparação tenha significado.

Estatísticas de bondade do ajustamento

	mod. inicial	mod. final
R^2	0.831	0.814
$\bar{R}_2$	0.620	0.735
AIC	-67.99	-73.60
BIC	-57.33	-67.60
$\widehat{\sigma}^2$	0.00034	0.00024

Relativamente à presença de um mecanismo de correcção de erros, representando com $\phi = -A(1)$ o coeficiente de LCP_{t-1}, tem-se $t_{MCE} = -3.72$ que, pelos critérios usuais é muito significativo mas que, como veremos no último capítulo, deixa um pouco a desejar para uma finalidade ainda não apresentada. De qualquer forma, o coeficiente é claramente negativo, embora não seja muito elevado em valor absoluto: para que exista retorno ao equilíbrio estima-se que sejam necessários praticamente 4 períodos, isto é, 4 anos, o que é um ajustamento lento.

As estimativas das elasticidades de longo prazo são $\lambda^{LR} = -\frac{0.24336}{-0.266244} = 0.914$, que parece ser bastante pausível e $\lambda^{LS} = -\frac{0.095209}{-0.266244} = 0.358$. Já a estimativa da semi-elasticidade de longo prazo em relação à inflação é muito baixa, mas não colide com a informação *a priori*: $\lambda^{INF} = -\frac{0.00092}{-0.266244} = -0.00034$ (com um sinal de acordo com o esperado segundo as teorias explicativas do consumo mais recentes). Por outro lado, a estimativa da constante na relação de equilíbrio de longo prazo é $\lambda_0 = -\frac{0.048774}{-0.266244} = 0.183$. Tem-se assim, como estimativa da relação de equilíbrio de longo prazo a equação

$$\widehat{LCP} = 0.183 + 0.914LRD + 0.358LSR - 0.00034INF,$$

onde se nota que a elasticidade estimada de longo prazo relativamente ao rendimento é muito superior à de curto prazo (apenas 0.338).

Finalmente, escreva-se o MCE estimado

$$\begin{aligned}\widehat{\Delta LC}_t = \quad & -0.266(LCP_{t-1} - 0.183 - 0.914LRD_{t-1} - 0.358LSR_{t-1} + 0.00034INF_{t-1}) \\ & +0.338\Delta LR_t + 0.457\Delta LC_{t-1} + 0.226\Delta LS_t - 0.247\Delta LR_{t-2}.\end{aligned}$$

Este modelo produziu boas previsões para os primeiros anos a seguir ao final da amostra mas os sintomas de instabilidade dos coeficientes começaram a agravarse pouco depois e, em meados da década de 2000, com as inovações financeiras a alterarem substancialmente os comportamentos de muitos consumidores, deixou de ser razoável.

8 Apêndices

8.1 Inverso de um polinómio em L

Dado um polinómio em L, $A(L) = \alpha_0 + \alpha_1 L + \ldots + \alpha_p L^p$, desde que $\alpha_0 \neq 0$, pode definir-se o o seu inverso, $A(L)^{-1}$ ou $1/A(L)$ como sendo

$$A(L)A(L)^{-1} = 1.$$

Por exemplo, o inverso de $A(L) = 1 - \alpha L$ é $A(L)^{-1} = 1 + \alpha L + \alpha^2 L^2 + \ldots$ porque

$$\begin{aligned} (1-\alpha L)(1+\alpha L+\alpha^2 L^2+\ldots) &= (1+\alpha L+\alpha^2 L^2+\ldots)-(\alpha L+\alpha^2 L^2+\ldots) \\ &= 1 \end{aligned}$$

8.2 Decomposições polinomiais em L

As relações entre os novos coeficientes, c_i, e os coeficientes do polinómio inicial, a_i, permitem escrever $c_0 = -a_1 - a_2 - \ldots - a_p$, $c_1 = -a_2 - a_3 - \ldots - a_p, \ldots, c_{p-1} = -a_p$. Tem-se, então

$$a_i = c_i - c_{i-1} \quad \text{e} \quad a(1) = a_0 - c_0.$$

Assim, o polinómio $a(L)$ pode escrever-se como

$$\begin{aligned} a(L) = &\ a_0 + \textstyle\sum_{i=1}^{p}(c_i - c_{i-1})L^i \\ = &\ a_0 + \textstyle\sum_{i=1}^{p} c_i L^i - \sum_{i=0}^{p-1} c_i L^{i+1} \\ = &\ a_0 - c_0 + \textstyle\sum_{i=0}^{p-1} c_i L^i - \sum_{i=0}^{p-1} c_i L^{i+1} \\ = &\ a(1) + \textstyle\sum_{i=0}^{p-1} c_i L^i (1-L). \end{aligned}$$

Ou seja, tem-se um primeiro resultado que pode ser útil

$$a(L) = a(1) + c(L)(1-L). \tag{23}$$

Somando e subtraindo $a(1)L$ ao lado direito obtém-se a decomposição da proposição

$$a(L) = a(1)L + [a(1) + c(L)](1-L).$$

9 Leituras adicionais

- A um nível básico, e de forma bastante complementar a este texto, aconselha-se a leitura de Stock e Watson (2015), cap. 14 que, apesar de muito elementar, já "entra" pela matéria das raízes unitárias

- A um nível mais avançado e complementando este texto na abordagem metodológica da LSE, recomendo o artigo de Hendry, Pagan e Sargan (1984) escrito para o *Handbook of Econometrics*, vol. II.

Bibliografia

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.

Davidson, R. MacKinnon, J. G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press.

Davidson, R. MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford University Press.

Faust, J. e Whiteman, C. H. (1997), General-to-specific procedures for fitting a data-admissible, theory-inspired, congruent, parsimonious, encompassing, weakly-exogenous, identified, structural model to the DGP: a translation and critique, *Carnegie-Rochester Conference Series on Public Policy*, 47, pp. 121-61.

Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press, Princeton.

Harvey, A. (1990), *The Econometric Analysis of Time Series*, Philip Allan.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press, Oxford.

Hendry, D. F. (2010), Equilibrium correction models, *in* Durlauf, S. N. e Blume, L. E. (eds.), *Macroeconomics and Time Series Analysis*, Palgrave Macmillan, pp. 76-89.

Hendry, D. F., Pagan, A. R. e Sargan, J. D. (1984), Dynamic Specification, *in* Grilliches, Z. e Intriligator, M. D. (eds.), *Handbook of Econometrics*, vol. II, cap. 18 (reimpresso em Hendry, D. F. (1993), *Econometrics: Alchemy or Science*, Blackwell, cap. 4).

Johnston, J. e DiNardo, J. (1997), *Econometric Methods*, McGraw-Hill.

Lopes, A. S. (1999), Modelos DL e ADL, raízes unitárias e cointegração: uma introdução, CEMAPRE, Texto de Apoio 15.

Lütkepohl, H. (2007), General-to-specific or specific-to-general modelling? *Journal of Econometrics*, 136, 319-24.

Pesaran, M. H. (2015), *Time Series and Panel Data Econometrics*, Oxford Univerrsity Press.

Stewart, J. (1991), *Econometrics*, Prentice Hall.

Stewart, J. e Gil, L. (1998), *Econometrics*, 2nd ed., Prentice Hall.

Stock, J. H. e Watson, M. W. (2015), *Introduction to Econometrics*, 3rd ed., Pearson Addison-Wesley.

Wooldridge, J. M. (2016), *Introductory Econometrics, a Modern Approach*, 6th ed., Cengage Learning.

Exercícios

1. (De um exame antigo.) Explique, de forma justificada, como poderá estimar os coeficientes do modelo $y_t = \delta_0\, x_t + \delta_1\, x_{t-1} + \delta_2\, x_{t-2} + \epsilon_t$, com $\epsilon_t \sim iid(0, \sigma^2)$, de forma a suavizar os problemas de colinearidade.

2. (De um exame antigo.) Explique porque razão(ões) e sob que condição se chama à seguinte reparametrização do modelo ADL(1,1)

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

com $\lambda = (\beta_0 + \beta_1)/(1 - \alpha_1)$, modelo de correcção de erros (MCE).

3. (Adaptado de um exame antigo.) Classifique, de forma justificada, a validade da seguinte afirmação: *Os modelos ADL não podem ser analisados no contexto do modelo clássico.*

4. Explique com clareza duas das vantagens da forma de Bardsen de um modelo ADL sobre a sua forma original.

5. O ficheiro moeda.xls contém dados trimestrais, de 1977:1 a 1995:4, para estimar um modelo linear dinâmico para a procura de moeda em Portugal. A função a estimar é do tipo

$$\log(Mr) = f[\log(PIB), r, inf],$$

onde Mr representa a moeda (M1) real, deflacionada com base no IPC. Para a taxa de inflação deve empregar o deflator do PIB [18]. Adicionalmente, como os dados são trimestrais, deverá incluir as *dummies* sazonais. (No EViews as dummies são @SEAS(1), ..., @SEAS(4), desde que o programa reconheça os dados como trimestrais).

Deve iniciar a análise com um modelo ADL(6,6) e deverá considerar também um termo de tendência determinística (no EViews deverá fazer "t=@trend+1"). Empregando a estratégia *GTS* (ou *gets*) simplifique-o tanto quanto for possível empregando também testes de especificação.

Nota: por razões que ficarão claras mais adiante, convém também que, além de ser negativa, a estatística t-MCE tenha um valor absoluto elevado.

Teste a presença de um MCE e, caso se justifique, indique a relação de equilíbrio de longo prazo estimada.

6. Num estudo de simulação empregou-se o seguinte PGD (ou DGP):

$$\begin{aligned} x_t &= 0.6\, x_{t-1} + \epsilon_{1t}, \\ y_t &= 0.5 + \gamma y_{t-1} + x_t + \epsilon_{2t},\ \gamma = 0, 0.1, 0.2, 0.3, \ldots, 1.0, \end{aligned}$$

com,

$$\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} \sim iid\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

Por outro lado, o modelo estimado em cada réplica é

$$y_t = \beta_1 + \beta_2\, x_t + u_t,$$

[18] Isto é, deve calcular a taxa de inflação como se a série do deflator fosse a do IPC.

sendo estimada também a regressão auxiliar $e_t = \phi\, e_{t-1} + \varepsilon_t$, e efectuado o teste-$t$ de $H_0 : \phi = 0$ vs. $H_1 : \phi \neq 0$.

As percentagens de rejeições dessa hipótese nula no estudo (efectuado com 10 000 réplicas) são:

Percentagens de rejeições

γ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
percent.	5.06	9.5	34.3	69.7	93.2	99.4	≈ 100	100	100	100	100

Descreva sucintamente o PGD, explique o objectivo do estudo e apresente as respectivas conclusões.