MACROECONOMETRIA I

Mestrado em Econometria Aplicada e Previsão

Estimação OLS do Modelo de Regressão Linear com Séries Temporais

Artur C. B. da Silva Lopes

Versão 1.01, Março de 2009

Instituto Superior de Economia e Gestão

Universidade Técnica de Lisboa

Conteúdo

1 Introdução

Este texto de apoio resulta de notas de leitura que escrevi para apoiar a leccionação do primeiro capítulo de Macroeconometria I, do 1º ano do mestrado em Econometria Aplicada e Previsão, do ISEG, no ano lectivo de 2008/09. Trata-se de estudar a aplicação do principal "cavalo de trabalho" da Econometria, o modelo de regressão linear, aos dados de séries macroeconómicas.

Baseei-me sobretudo no livro de F. Hayashi, *Econometrics*, de 2000, mas estas notas não pretendem substituir a leitura desse livro, ou de outro bom livro de Econometria. De resto, o nível deste texto não é tão avançado nem tão detalhado quanto o desse livro. Também a notação empregue é algo diferente e, sobretudo, procurou-se adaptá-la aos dados de séries temporais.

Um pressuposto importante é de que o leitor já teve um primeiro contacto com o modelo de regressão linear, ao nível da licenciatura. Por outro lado, no mestrado em Econometria Aplicada e Previsão, Macroeconometria I é precedida da disciplina de Econometria, onde o modelo de regressão linear foi objecto de estudo mais aprofundado. Desta forma, como há muitos aspectos comuns com o modelo para dados seccionais independentes e identicamente distribuídos, a apresentação do modelo clássico, com regressores estritamente exógenos, é muito abreviada. A abordagem seguida aqui difere da dessa disciplina dado que a perspectiva passa a ser a da utilização do modelo de regressão linear no contexto de modelos com dados de séries temporais. Como se verá, este facto introduz algumas diferenças significativas. Por outro lado, uma vez que as especificidades das séries temporais emergem com muito maior importância, o modelo de regressão com regressores pré-determinados é objecto de atenção muito mais detalhada.

Como esta é a primeira versão deste texto, agradeço a todos aqueles que formularem críticas e sugestões e que indicarem erros, imprecisões, incorrecções e gralhas. Obviamente, assumo a responsabilidade por todos os erros e omissões que possam existir.

2 O Modelo Clássico de Regressão

Nesta secção estuda-se o modelo clássico de regressão linear com regressores estocásticos, isto é, as propriedades do estimador OLS em pequenas amostras. Começaremos por listar e discutir as hipóteses do modelo.

2.1 Hipóteses

[H1] – Linearidade. A relação entre a variável dependente e os regressores é linear,

$$y_t = \beta_1 + \beta_2\, x_{t2} + \ldots + \beta_K\, x_{tK} + u_t, \;\; t = 1, 2, \ldots, T, \tag{1}$$

onde os β's são parâmetros a estimar e u_t é o termo de erro ou variável residual, não observada. A equação (1) também se pode escrever

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

com x_t' a representar a t-ésima linha da matriz X, isto é, $x_t' = [1\ x_{t2}\ \ldots\ x_{tK}]$ e β o vector de coeficientes da regressão. Em termos matriciais, como é sabido, o modelo escreve-se

$$y = X\beta + u.$$

Um exemplo bem conhecido é o da "velha" função consumo keynesiana,

$$C_t = \beta_1 + \beta_2\, YD_t + u_t,$$

onde C representa o consumo privado[1] e YD o rendimento disponível das famílias (ambas em termos agregados para uma economia). Como é sabido, trata-se de um modelo de regressão simples.

Um exemplo bastante mais recente é o da chamada regra de Taylor para a política monetária, que relaciona a taxa de juro de curto prazo fixada pelo banco central, R_t (nos EUA, a taxa *fed funds*), com os desvios da inflação do objectivo e com o hiato do produto (*output gap*). Uma versão simplificada (e estática) é dada por

$$R_t = \alpha + g_\pi(\bar{\Pi}_t - \Pi_t^*) + g_y(y_t - y_t^*) + u_t,$$

[1] Recorda-se que no TSP não se pode representar nenhuma variável com C.

onde $\bar{\Pi}_t$ é uma média dos valores da taxa de inflação (anualizada, Π_t) sobre os últimos quatro trimestres, Π_t^* é o objectivo para a taxa de inflação, y_t é o logaritmo do produto e y_t^p é o logaritmo do produto potencial. Os coeficientes g_π e g_y representam a resposta das taxas de juro relativamente aos desvios da inflação do seu objectivo e do *output* do seu potencial, ou seja, constituem indicadores da agressividade dos bancos centrais relativamente a cada um desses objectivos. Embora objecto de debate, uma equação deste tipo parece representar bem o comportamento de alguns bancos centrais e, sobretudo da *Fed*. No seu trabalho original de 1993, Taylor sugeriu os coeficientes $\hat{g}_\pi = 1.5$ e $\hat{g}_y = 0.5$, o que significa que a Fed responderia a um aumento de 1% da taxa de inflação nos últimos 4 trimestres aumentando a taxa de curto prazo em 150 pontos base [2].

[H2] – Exogeneidade estrita. O valor esperado de cada erro condicional em todas as observações dos regressores é zero:

$$E(u_t|\, X) = 0, \quad t = 1, 2, \ldots, T. \tag{2}$$

Se esta hipótese for satisfeita diz-se que os regressores são estritamente exógenos. Ela é muitíssimo forte e, em geral, nos modelos seccionais não é necessário assumi-la sob esta forma porque a independência entre observações resultante da amostragem aleatória garante a independência entre o erro de uma equação e os regressores de outras observações. Todavia, nos modelos para séries temporais ela é crucial.

Dela decorrem várias implicações que tornam mais simples a sua interpretação.

a) O valor esperado não condicional de cada erro é zero, i.e.,

$$E(u_t) = 0, \quad t = 1, 2, \ldots, T.$$

Com efeito, usando a propriedade ou lei dos valores esperados totais [3] :

$$\begin{aligned} E(u_t) &= E\left[E(u_t|\, X)\right] \\ &= E(0) = 0. \end{aligned}$$

[2] Para um sumário de algumas estimativas apresentadas na literatura veja-se, por exemplo, Stock, J. e Watson, M. (2003), Has the business cycle changed? Evidence and explanations, documento de trabalho.

[3] Autores como Wooldridge (2002) consideram que esta é uma versão simples da propriedade dos valores esperados iterados. As propriedades dos valores esperados condicionais podem ser revistas em Wooldridge (2002), pp. 29-30.

b) Os regressores são ortogonais [4] aos erros de todas as observações, i.e.,

$$E(x_{tj}\, u_s) = 0, \;\; \forall t, s (= 1, 2, \ldots, T) \text{ e } \; j (= 1, 2, \ldots, K).$$

De facto, pela lei ou propriedade dos valores esperados iterados, tem-se

$$E(u_s |\, x_{tj}) = E\left[E(u_t |\, X) |\, x_{tj}\right] = 0.$$

Daqui resulta que

$$\begin{aligned} E(x_{tj}\, u_s) &= E\left[E(x_{tj}\, u_s |\, x_{tj})\right] \\ &= E\left[x_{tj}\, E(u_s |\, x_{tj})\right] = E(x_{tj}\, 0) = 0, \end{aligned}$$

onde a primeira igualdade decorre da lei dos valores esperados totais e a última da implicação anterior.

Ou seja, a hipótese de exogeneidade estrita dos regressores implica e, portanto, requer que os regressores sejam ortogonais não só ao erro contemporâneo, $E(x_{tj}\, u_t) = 0, \;\; \forall j, t$ — caso em que se fala apenas em exogeneidade contemporânea — mas também aos erros de todas as outras observações, $E(x_{tj}\, u_s) = 0$, $\forall j$ e $\forall t, s$, $t \neq s$.

c) Os regressores não estão correlacionados nem com os erros passados, nem com o erro corrente, nem com os erros futuros,

$$Cov(u_s,\, x_{tj}) = 0, \; \forall t, s, j.$$

Com efeito,

$$\begin{aligned} Cov(u_s,\, x_{tj}) &= E(x_{tj}\, u_s) - E(x_{tj}) E(u_s) \\ &= E(x_{tj}\, u_s) = 0, \end{aligned}$$

onde a última igualdade decorre da implicação anterior.

Esta é a implicação mais importante pois é a que fornece a interpretação prática mais intuitiva em aplicações: ausência total de correlação entre regressores e erros (passados, contemporâneo e futuro). Todavia, alguns autores preferem referir-se à implicação anterior, ou seja, à ortogonalidade entre regressores e erros passados, corrente e futuros.

[4] Se o valor esperado do produto de duas variáveis aleatórias, Z e W, é zero, $E(ZW) = 0$, dizemos que são ortogonais.

Usando esta implicação verifica-se, por exemplo, que um modelo autoregressivo não pode satisfazer a hipótese de exogeneidade estrita. Por exemplo, no caso do modelo AR(1),

$$y_t = \rho\, y_{t-1} + u_t,$$

com $u_t \sim iid(0,\, \sigma^2)$, não pode ter-se $Cov(u_s,\, y_{t-1}) = 0,\, \forall s$. Basta considerar $s = t-1$, caso em que se tem $Cov(u_{t-1},\, y_{t-1}) \neq 0$ (pois é muito fácil verificar que y_{t-1} depende de $u_t - 1$: $y_{t-1} = \rho\, y_{t-2} + u_{t-1}$.) Mais formalmente, note-se que

$$\begin{aligned} E(u_t|\, X) &= E(y_t - \rho\, y_{t-1}|\, X), \;\; t = 2, \ldots, T-1, \\ &= y_t - \rho\, y_{t-1}, \;\; t = 2, \ldots, T-1, \end{aligned}$$

pois tanto y_t como y_{t-1} são observações incluídas na matriz X. Ora, não existe nada que garanta que $y_t - \rho\, y_{t-1}$ seja zero. Alternativamente, também se pode calcular $cov(y_t, u_t)$ ou $E(y_t\, u_t)$ e verificar que são diferentes de zero.

Mais geralmente, como a hipótese requer a exogeneidade estrita de todos os regressores, nenhum modelo que tenha como regressor a variável dependente desfasada, isto é, que tenha uma componente autoregressiva, satisfaz a hipótese de exogeneidade estrita.

Recordam-se outras situações "clássicas" (estudadas no contexto dos modelos para dados seccionais) em que esta hipótese é violada:

a) má especificação da forma funcional da relação entre a variável dependente e as explicativas;

b) regressores omitidos correlacionados com os incluídos;

c) erros de medição nalgum dos regressores que está correlacionado com o termo de erro, e

d) simultaneidade, isto é, causalidade recíproca (ou inversa) simultânea.

Existe ainda um caso de violação da hipótese típico de modelos para séries temporais. Considere-se, por exemplo, o modelo

$$INF_t = \beta_1 + \beta_2\, r_t + u_t$$

onde INF representa a taxa de inflação na zona euro e r uma taxa de juro controlada pelo BCE. Para este modelo não é razoável assumir que $Cov(u_t,\, r_{t+1}) = 0$.

Com efeito, se, por exemplo, num determinado período a taxa de inflação está mais elevada que o esperado (dada a taxa de juro), no período seguinte o BCE tenderá a ajustar a taxa de juro, de forma a controlar a inflação. Ora, o valor "anormal" de INF só pode ser explicado por algum factor incluído no termo de erro da equação. Desta forma, é muito plausível que $Cov(u_t,\, r_{t+1}) \neq 0$. Note-se que este é um caso típico daquilo a que se chama de presença de um efeito de *feedback*: não só r influencia INF no período corrente, como é de esperar que INF_t tenha algum efeito sobre r_{t+1}.

Mais geralmente, quando algum dos regressores é uma variável de política, é muito improvável que a hipótese seja satisfeita: em geral, essa variável deve ser influenciada pelo passado da variável dependente, e esse efeito de *feedback* ou de retorno reflecte-se em correlação dos erros correntes com valores futuros do regressor.

Finalmente refira-se que alguns autores consideram, pelo menos implicitamente, que os modelos de desfasamentos distribuídos (*distributed lags*, DL) como, por exemplo, o modelo

$$y_t = \mu + \delta_0\, x_t + \delta_1\, x_{t-1} + \ldots + \delta_4\, x_{t-4} + u_t, \quad t = 1, 2, \ldots, T,$$

satisfazem a hipótese. Todavia, não existe nenhuma garantia que isso aconteça.

Note-se que, conjuntamente, as hipóteses H1 e H2 implicam que

$$E(y_t|X) = x_t'\beta, \quad t = 1, 2, \ldots, T. \tag{3}$$

Deixa-se a demonstração desta pequena proposição como exercício. Por outro lado, inversamente, se esta última equação for adoptada como hipótese, existe um vector de erros que satisfaz as duas hipóteses referidas. Ou seja, apenas com ela podem apresentar-se de forma compacta as duas primeiras hipóteses.

Com efeito, partindo desta equação, defina-se

$$u_t \equiv y_t - E(y_t|X).$$

Então, *por construção*, a hipótese H1 é satisfeita: $y_t = x_t'\beta + u_t$. Por outro lado, a hipótese H2 também é satisfeita. Tomando valores esperados na equação anterior

$$\begin{aligned} E(u_t|X) &= E(y_t|X) - E[\,E(y_t|X)|\,X] \\ &= E(y_t|X) - E(y_t|X) = 0. \end{aligned}$$

[H3] – Não colinearidade. A característica da matriz de observações dos regressores, de tipo $(T \times K)$, X, é K com probabilidade 1.

Ou seja, nenhuma das colunas de X pode ser expressa como combinação linear das restantes, o que significa que a matriz tem característica de colunas completa (*full column rank*). Como as K colunas não podem ser linearmente independentes se tiverem dimensão menor que K, a hipótese requer que $T \geq K$, isto é, o número de observações não pode ser inferior ao de regressores. Recorde-se que, se a hipótese for violada, se fala em colinearidade exacta ou perfeita.

[H4] – Homocedasticidade. A variância dos erros condicional em X é constante,

$$Var(u_t|X) = \sigma^2 > 0, \;\; t = 1, 2, \ldots, T.$$

Alternativamente, a hipótese também se pode escrever em termos dos segundos momentos dos erros:

$$E(u_t^2|X) = \sigma^2 > 0, \;\; t = 1, 2, \ldots, T.$$

Com efeito, empregando H2 tem-se

$$Var(u_t|X) = E(u_t^2|X) - E(u_t|X)^2 = E(u_t^2|X).$$

Embora a importância desta hipótese só recentemente tenha recebido a atenção que merece em modelos para séries temporais, também aqui ela não vai poder ser devidamente abordada.

[H5] – Ausência de autocorrelação. As autocorrelações dos erros, condicionais em X, são nulas,

$$Cov(u_t, \, u_s|X) = 0, \;\; \forall t, s (= 1, 2, \ldots, T), \; t \neq s.$$

Facilmente se verifica que esta hipótese também pode ser formalizada como

$$E(u_t \, u_s|X) = 0, \;\; \forall t, s (= 1, 2, \ldots, T), \; t \neq s.$$

No entanto, note-se que ela é muito frequentemente formalizada em termos não condicionais, isto é, como

$$Cov(u_t, \, u_s) = E(u_t \, u_s) = 0, \;\; \forall t, s (= 1, 2, \ldots, T), \; t \neq s.$$

o que facilita a sua interpretação e não implica perda de rigor significativa.

Estas duas últimas hipóteses podem ser compactadas numa só, usualmente chamada de esfericidade (da matriz de covariâncias) dos erros:

$$E(u\,u'|X) = Var(u|X) = \sigma^2\,I.$$

[H6] – Normalidade dos erros. A distribuição do vector de erros, u, condicional em X, é normal (multivariada), isto é, $u|X \sim N$.

Desta forma, conjugando com as hipóteses anteriores,

$$u|X \sim N(0,\,\sigma^2 I).$$

E como esta distribuição não depende de X, também se tem que $u \sim N(0,\,\sigma^2 I)$.

2.2 O estimador OLS

Nesta subsecção recorda-se a estimação OLS e alguns resultados algébricos associados a ela. Considera-se que estes aspectos deverão ser já bem conhecidos e, por esse facto, a exposição é muito abreviada.

Represente-se com $\tilde{\beta}$ um vector de estimativas para o vector de coeficientes, β. Com essas estimativas obtêm-se os resíduos para todas as observações:

$$y_t - x_t'\tilde{\beta}, \;\; t = 1, 2, \ldots, T.$$

Com estes, pode formar-se a soma dos quadrados dos resíduos,

$$SSR(\tilde{\beta}) = \sum_{t=1}^{T}(y_t - x_t'\tilde{\beta})^2 = (y - X\tilde{\beta})'(y - X\tilde{\beta}),$$

por vezes também representada com RSS. Como se sabe, o estimador OLS minimiza esta função:

$$\widehat{\beta} \equiv argmin\, SSR(\tilde{\beta}).$$

Recorda-se também que, resolvendo este problema de optimização, se obtêm as equações normais dos mínimos quadrados,

$$(X'X)\widehat{\beta} = X'y,$$

cuja solução, isto é, o estimador OLS (desde que H3 seja satisfeita) é dada por

$$\widehat{\beta} = (X'X)^{-1}X'y.$$

Como $(X'X)^{-1}X'y = \left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{T}X'y\right)$, o estimador OLS também pode ser escrito como

$$\widehat{\beta} = S_{xx}^{-1}s_{xy}, \tag{4}$$

onde

$$S_{xx} = \frac{1}{T}X'X = \frac{1}{T}\sum_{t=1}^{T} x_t x_t',$$

que representa a média amostral de $x_t x_t'$ e,

$$s_{xy} = \frac{1}{T}X'y = \frac{1}{T}\sum_{t=1}^{T} x_t y_t,$$

que representa a média amostral de $x_t y_t$.

Com as estimativas OLS podem obter-se os respectivos resíduos, representados com e:

$$e_t = y_t - x_t'\widehat{\beta}, \quad t = 1, 2, \ldots, T,$$

ou seja,

$$e = y - X\widehat{\beta}.$$

Rearranjando as equações normais, obtém-se $X'(y - X\widehat{\beta}) = 0$, isto é,

$$X'e = 0.$$

E esta expressão também se pode escrever como

$$\frac{1}{T}\sum_{t=1}^{T} x_t e_t = 0,$$

isto é, as equações normais podem ser vistas como o equivalente ou análogo amostral das condicões de ortogonalidade $E(x_t e_t) = 0$.

Recordam-se agora, de forma breve, alguns conceitos e resultados algébricos associados à estimação OLS:

1. O valor ajustado para a observação t é definido como $\widehat{y}_t \equiv x_t'\widehat{\beta}$. Desta forma, tem-se o vector de valores ajustados, $\widehat{y} = X\widehat{\beta}$. Assim, o vector de resíduos pode escrever-se como $e = y - \widehat{y}$.

2. As matrizes de projecção e aniquiladora são definidas, respectivamente, como,

$$P \equiv X(X'X)^{-1}X', \text{ e} \tag{5}$$

$$M \equiv I - P \equiv I - X(X'X)^{-1}X'. \tag{6}$$

   E elas possuem as seguintes propriedades:

   a) são ambas simétricas e idempotentes;

   b) $PX = X$ e

   c) $MX = 0$.

3. Usando estas matrizes, tem-se

$$\hat{y} = Py, \text{ e}$$

$$e = My = Mu.$$

   A segunda igualdade justifica o facto de a matriz M ser frequentemente chamada de matriz "fazedora de resíduos". Por exemplo, dada uma outra matriz de regressores, seja Z, a pré-multiplicação de y pela matriz $M_Z = I - Z(Z'Z)^{-1}Z'$ produz os resíduos OLS dessa regressão, $e_Z = M_Z\, y$.

4. A soma dos quadrados dos resíduos pode escrever-se como

$$SSR = RSS = e'e = u'Mu.$$

5. O estimador OLS de σ^2, representado com s^2 (ou com $\widehat{\sigma^2}$) é dado por

$$s^2 = \widehat{\sigma^2} = \frac{SSR}{T-K} = \frac{e'e}{T-K}.$$

6. A sua raiz quadrada, s ou $\widehat{\sigma}$, é chamada de erro padrão da regressão (SER, *standard error of the regression*), e é uma estimativa do desvio padrão dos erros.

7. $\widehat{\beta} - \beta$ é chamado de erro de amostragem. Recorde-se que se pode escrever

$$\widehat{\beta} - \beta = (X'X)^{-1}X'u.$$

2.3 Propriedades exactas do OLS

Também de forma muito abreviada e sem demonstração, veremos agora as propriedades exactas ou em pequenas amostras do estimador OLS de β.

Proposição 2.1 – Propriedades exactas do estimador OLS de β:

a) (não enviesamento) Sob as hipóteses H1 a H3, $E(\widehat{\beta}|X) = \beta$.

b) (matriz de covariâncias) Sob as hipóteses H1 a H5, $Var(\widehat{\beta}|X) = \sigma^2(X'X)^{-1}$.

c) (Teorema de Gauss-Markov) Sob as hipóteses H1 a H5, o estimador OLS de β é o mais eficiente na classe de estimadores lineares centrados, ou seja, é *BLUE*. Isto é, para qualquer estimador $\tilde{\beta}$ que seja centrado e linear em y, $Var(\tilde{\beta}|X) - Var(\widehat{\beta}|X)$ é uma matriz semi-definida positiva. Pode dizer-se que a matriz $Var(\widehat{\beta}|X)$ é menor que a matriz $Var(\tilde{\beta}|X)$ no sentido matricial.

d) (Ortogonalidade entre o estimador e os resíduos) Sob as hipóteses H1 a H5, $Cov(\widehat{\beta}, e|X) = 0$.

Algumas observações sobre estas propriedades:

1. Saliente-se que só se empregaram as hipóteses H1 a H5. A hipótese H6 não é necessária para obter estes resultados.

2. A hipótese H2, de exogeneidade estrita, é crucial, em particular para a propriedade de não enviesamento. Embora muito exigente, e raramente satisfeita em modelos para séries temporais, esta hipótese não pode ser enfraquecida. Isto é, para obter o não enviesamento, não é suficiente a hipótese de exogeneidade contemporânea, $E(u_t|x_t) = 0$, $\forall t$ ou $E(x_t u_t) = 0$, $\forall t$.

   Por outro lado, dessa propriedade também decorre o não enviesamento não condicional,
   $$E(\widehat{\beta}) = \beta,$$
   pela propriedade dos valores esperados totais.

3. Da mesma forma,
   $$Var(\tilde{\beta}) - Var(\widehat{\beta}) = H,$$
   com H matriz semi-definida positiva.

4. Uma das implicações mais importantes de c) é que

$$Var(\tilde{\beta}_j|X) \geq Var(\widehat{\beta}_j|X), \ \ \forall j(=1,2,\ldots,K),$$

ou seja, para qualquer um dos coeficientes da regressão, a variância do estimador OLS não é maior que a de qualquer outro estimador linear centrado.

Proposição 2.2 – não enviesamento de s^2 (ou $\widehat{\sigma^2}$).
Sob as hipóteses H1 a H4 e desde que $T > K$, $E(s^2|X) = \sigma^2$ e portanto $E(s^2) = \sigma^2$.

Desta forma, o estimador natural ou lógico para a matriz de covariâncias do estimador OLS é dado por

$$\widehat{Var(\hat{\beta}|X)} = s^2(X'X)^{-1}.$$

2.4 Testes de hipóteses

De novo de forma muito abreviada e sem apresentar qualquer demonstração, (re)-veremos em seguida os resultados empregues para efectuar testes estatísticos sobre os coeficientes do modelo de regressão. Como é bem sabido, agora a validade da hipótese H6 passa a ser crucial.

Começamos por considerar o caso em que se pretende testar uma hipótese sobre um único coeficiente,

$$H_0 : \beta_j = \beta_j^0,$$

onde β_j^0 representa o valor especificado pela hipótese nula. Introduza-se primeiro a notação

$$se(\widehat{\beta_j}) \equiv \sqrt{s^2[(X'X)^{-1}]_{jj}}$$

para representar o erro padrão (*standard error*) do estimador $\widehat{\beta_j}$.

Proposição 2.3 – Distribuição da estatística-t.
Sob as hipóteses H1 a H6 e quando H_0 é verdadeira, isto é, sob H_0, a estatística-t

$$t_j \equiv \frac{\widehat{\beta_j} - \beta_j^0}{se(\widehat{\beta_j})} \sim t_{(T-K)}.$$

Como é bem sabido, o teste estatístico pode ser efectuado com base nesta estatística de teste pois a sua distribuição não depende de qualquer parâmetro desconhecido. Se não fosse esse o caso, a esse(s) parâmetro(s) chamar-se-ia parâmetro(s) perturbador(es).

A lógica de utilização da estatística, a forma de tomada de decisão e a sua utilização são pressupostas conhecidas. O mesmo acontece com a sua utilização para construir intervalos de confiança e com o emprego de valores-p para tomar decisões (uma boa referência sobre este assunto é Davidson e MacKinnon, 1993, pp. 80-81).

Recorda-se ainda que, no caso em que se tem, $H_0 : \beta_0 = 0$, isto é, um teste de significância individual sobre um coeficiente, a estatística de teste reduz-se a

$$t_j = \frac{\widehat{\beta_j}}{se(\widehat{\beta_j})} \sim t_{(T-K)} \;\text{ sob } H_0,$$

e é usualmente designada de rácio-t.

Mais geralmente, poderemos estar interessados em testar restrições lineares sobre os coeficientes, escritas sob a forma de um sistema de equações lineares,

$$H_0 : \; R\beta = r,$$

onde a matriz R e o vector r são conhecidos, pois são especificados pela hipótese nula. Considera-se que o número de equações, isto é, o número de restrições lineares é J e, portanto, a matriz R é de tipo $(J \times K)$. Para assegurar que não há equações redundantes e que as restrições não são contraditórias, requer-se que as J linhas sejam linearmente independentes, isto é, que a característica de R é J $(r(R) = J)$.

Proposição 2.4 – Distribuição das estatísticas-F.
Sob as hipóteses H1 a H6 e sob $H_0 : R\beta = r$, com R matriz conhecida de tipo $(J \times K)$, com $r(R) = J$, tem-se

$$F \equiv (R\widehat{\beta} - r)'[R\widehat{Var(\beta}|X)R']^{-1}(R\widehat{\beta} - r)/J \sim F_{(J,T-K)}.$$

Recorde-se, ainda, que há uma forma muito mais simples de obter o valor desta estatística de teste. Representando com SSR_{UR} a soma dos quadrados dos resíduos do modelo estimado sem restrições e com SSR_R a soma dos quadrados dos resíduos do modelo estimado impondo as restrições, isto é, sob H_0, a estatística pode escrever-se

$$F = \frac{(SSR_R - SSR_{UR})/J}{SSR_{UR}/(T-K)},$$

onde o denominador é o estimador OLS de σ^2 do modelo sem restrições.

2.5 Estimação de máxima verosimilhança

Tal como nas subsecções anteriores, revêem-se aqui, de forma bastante sumária, os principais resultados associados à estimação de máxima verosimilhança (MV).

Com base nas hipóteses do modelo tem-se que

$$y|X \sim N(X\beta,\, \sigma^2 I).$$

Por conseguinte, a densidade do vector de observações y condicional em X, isto é, a função de verosimilhança é

$$f(y|X) = (2\pi\sigma^2)^{-T/2} \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right].$$

Logaritmizando, tem-se a função log-verosimilhança:

$$\log L(\beta, \sigma^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta).$$

Como se recorda, a maximização desta função concentrada relativamente a β equivale à minimização da soma dos quadrados dos resíduos. Desta forma, o estimador de máxima verosimilhança é o estimador OLS. Todavia, o estimador de MV de σ^2 não é o estimador OLS e é, portanto, enviesado.

Proposição 2.5 – Estimação de MV dos parâmetros.
Sob as hipóteses H1 a H6, o estimador de MV de β, $\widetilde{\beta}$, é o estimador OLS, $\widehat{\beta}$, e o estimador de MV de σ^2 é

$$\widetilde{\sigma^2} = \frac{e'e}{T} = \frac{SSR}{T} = \frac{T-K}{T}s^2.$$

Uma expressão que também pode ser útil é a do valor maximizado da função log-verosimilhança:

$$\log L(\widetilde{\beta}, \widetilde{\sigma^2}) = -\frac{T}{2}\log\left(\frac{2\pi}{T}\right) - \frac{T}{2} - \frac{T}{2}\log(SSR).$$

Como, por outro lado, se pode mostrar que a matriz de covariâncias do estimador OLS de β é igual à matriz de covariâncias mínima de estimadores centrados de β, isto é, ao limite inferior de Cramer-Rao, tem-se a seguinte proposição.

Proposição 2.6 – O estimador OLS é *BUE* (*best unbiased estimator*).
Sob as hipóteses H1 a H6, o estimador OLS $\widehat{\beta}$ é BUE, no sentido em que qualquer outro estimador centrado (mas não necessariamente linear) de β tem maior matriz de covariâncias no sentido matricial.

Como é evidente, este resultado é mais forte que o teorema de Gauss-Markov. Todavia, ao contrário deste, requer a satisfação da hipótese de normalidade.

2.6 Revisitando as hipóteses

Para cada resultado foram referidas as hipóteses necessárias à sua validade de forma breve, o que poderá levar a pensar que a sua importância é relativamente reduzida. Não é esse o caso e aconselha-se o leitor a rever as hipóteses e as proposições. A título de exemplo, e sem pretender esgotar estes aspectos, saliente-se que:

a) se pelo menos uma das duas primeiras hipóteses não for satisfeita, não só o estimador OLS não terá as "belas" propriedades referidas como os resultados obtidos para fazer inferência estatística deixarão de ser válidos (por exemplo, um teste estatístico poderá passar a ter uma dimensão real muito superior à nominal);

b) se a hipótese de homocedasticidade ou a de ausência de autocorrelação não forem satisfeitas, o estimador OLS continuará a ser centrado mas deixará de ser *BLUE* e, mais uma vez, os métodos de inferência associados às estatísticas de teste apresentadas deixam de ser válidos;

c) se a hipótese de normalidade não for satisfeita o estimador de MV deixa de ser o estimador OLS [5] e este deixa de ser *BUE*. Também os resultados distributivos exactos deixam de ser válidos.

Por outro lado, não é demais insistir na importância da hipótese H2, de exogeneidade estrita dos regressores. O modelo clássico permite ao OLS produzir "belos"

[5] E diz-se que o estimador OLS passa a ser um estimador de *quasi* ou pseudo-MV.

resultados de estimação e de inferência mas tem aplicabilidade muito reduzida devido a essa exigência, que raramente é satisfeita. Acresce que, como ela não é formulada no contexto de um modelo estatístico completo e como o estimador OLS assegura sempre a ortogonalidade amostral, $X'e = 0$, mesmo que a hipótese não seja satisfeita, ela é muito dificilmente testável.

Desta forma, é o raciocínio económico ou a análise económica da plausibilidade da hipótese que deverão ser empregues. Ou seja, é crucial considerar cuidadosamente o contexto económico da análise. Considerem-se alguns exemplos de ilustração.

1. Suponha-se que se pretende analisar a sensibilidade das exportações portuguesas para os EUA ao rendimento desse país empregando um modelo de regressão das primeiras sobre o segundo. Pode considerar-se que existe um efeito de *feedback*, isto é, de causalidade inversa, e que, pelo menos com dados anuais, este até é contemporâneo: uma redução das exportações portuguesas reduz o rendimento nacional, o qual tenderá a reduzir as importações portuguesas provenientes dos EUA e, assim, a reduzir o rendimento dos EUA. Todavia, a importância das importações portuguesas com origem nos EUA no rendimento deste país é muito pequena, pelo que é plausível tratar esta última variável como estritamente exógena.

2. Já não seria assim, no entanto, no caso de a variável dependente serem as exportações da UE para os EUA. Existe, neste caso, um efeito de causalidade inversa simultâneo com bastante significado, o que retira plausibilidade à satisfação da hipótese.

3. Uma situação do mesmo tipo caracteriza a análise da influência dos preços do petróleo sobre o rendimento ou sobre a inflação. Tradicionalmente, considerava-se que os preços do petróleo poderiam ser considerados como um regressor exógeno pois são determinados nos mercados internacionais e, em grande medida, por acção concertada dos países da OPEP. Todavia, como se tem constatado, os membros desta organização fixam as quotas de produção tomando em consideração as condições da economia mundial (entre outros factores). Desta forma, se a análise tem por objectivo o estudo do impacto dos preços sobre o rendimento ou sobre a inflação dos EUA, a variável não pode ser considerada (estritamente) exógena. Já não será assim se o objecto de estudo for a economia portuguesa, cujo peso na economia mundial é muitíssimo reduzido.

2.7 Um resultado algébrico importante

Nesta última subsecção apresenta-se, sem demonstração[6], um resultado algébrico importante para o estimador OLS do modelo de regressão linear. Trata-se de um resultado que se pode revelar muito útil para efectuar demonstrações sobre coeficientes individuais ou sobre grupos de coeficientes do modelo, bem como sobre estatísticas de teste associados a eles.

Proposição 2.7 – Teorema de Frisch-Waugh-Lovell (FWL).
Parta-se a matriz X em dois blocos, X_1 e X_2, de tipos $(T \times K_1)$ e $(T \times K_2)$, respectivamente, com $K_1 + K_2 = K$, $X = [X_1\ X_2]$. Faça-se a respectiva partição do vector de coeficientes β: $\beta' = [\beta_1'\ \beta_2']$. A regressão pode então escrever-se como

$$y = X_1\beta_1 + X_2\beta_2 + u.$$

Sejam:

$$P_1 = X_1(X_1'X_1)^{-1}X_1',$$

a matriz de projecção da regressão de y apenas sobre os regressores de X_1,

$$M_1 = I - P_1 = I - X_1(X_1'X_1)^{-1}X_1',$$

a matriz aniquiladora da mesma regressão,

$$\widetilde{X_2} = M_1X_2,$$

a matriz cujas colunas são os vectores de resíduos da regressão da respectiva variável de X_2 sobre os regressores de X_1, e

$$\widetilde{y} = M_1y,$$

o vector de resíduos da regressão de y apenas sobre os regressores de X_1. Note-se que, por exemplo, este vector $\widetilde{y}$ representa a variação de y expurgada da associação linear com os regressores de X_1, isto é, com essa parte da sua variação removida.

Então:

$$\widehat{\beta_2} = (\widetilde{X_2'}\,\widetilde{X_2})^{-1}\widetilde{X_2'}\,\widetilde{y},$$

[6]Convida-se o leitor a efectuar a demonstração. Para este efeito, encontra em Hayashi (2000), p. 72-74, várias sugestões.

isto é, as estimativas OLS de β_2 podem ser obtidas efectuando a regressão OLS dos resíduos $\widetilde{y}$ sobre a matriz de resíduos $\widetilde{X_2}$.

Por outro lado,

$$\widetilde{e_2} = \widetilde{M_2}\,\widetilde{y} = e,$$

ou seja, os resíduos OLS da regressão de $\widetilde{y}$ sobre $\widetilde{X_2}$ são idênticos aos resíduos da regressão de y sobre $X = [X_1\ X_2]$ (isto é, sobre todos os regressores).

Por outras palavras, as estimativas OLS de β_2 e os resíduos da regressão original e da regressão de $\widetilde{y}$ sobre $\widetilde{X_2}$ são iguais.

2.8 Exercícios

2.1 Mostre que as hipóteses H1 e H2 implicam a equação (3).

2.2 Mostre que as hipóteses H2, H4 e H5 implicam que

$$Var(u_t) = \sigma^2, \;\; t = 1, 2, \ldots, T, \quad \text{e}$$

$$Cov(u_t, u_s) = 0, \; \forall t, s (= 1, 2, \ldots T), \; t \neq s.$$

2.3 (Adaptado de Hayashi [2000], pp. 81-84.) O objectivo deste exercício é o de confirmar, com análise de simulação, as propriedades obtidas teoricamente para o modelo de regressão linear clássico, considerando um modelo muito simples.

Considere que o processo de geração de dados é

$$y_t = 1 + 0.5 x_t + u_t, \qquad u_t \sim iidN(0,\, 1), \quad t = 1, \ldots, 32,$$

com

$$x_t = 2 + 0.6\, x_{t-1} + \eta_t, \quad \eta_t \sim iidN(0,\, 1), \quad x_0 \sim N(5,\, 1.25).$$

a) Considerando 10000 réplicas, analise $E(\widehat{\beta}|X)$. Diga também como procederia para analisar $E(\widehat{\beta})$.

b) Ainda considerando 10000 réplicas, estime a probabilidade de rejeição do teste-t de $H_0 : \beta_2 = 0.5$ *vs.* $H_1 : \beta_2 \neq 0.5$. Interprete a estimativa obtida.

c) Efectue as análises das alíneas anteriores com 20000 e 50000 réplicas e compare os resultados.

d) Diga como procederia para estimar a potência do teste-t de $H_0 : \beta_2 = 0$ *vs.* $H_1 : \beta_2 \neq 0$, isto é, a potência do teste no ponto $\beta_2 = 0$.

e) Em vez dos números de réplicas das alíneas anteriores, Hayashi sugere a possibilidade de empregar 1 milhão de réplicas. Comente esta sugestão.

2.4 (Revisão de variáveis *dummy*.) Alguns investigadores sustentam que uma função consumo agregada simples das famílias, $C_t = \beta_1 + \beta_2 YD_t + u_t$, apresentou, para Portugal, um comportamento anómalo entre 1974 e 1977. Supondo que dispõe de dados anuais, que esse modelo satisfaz as hipóteses H1 a H6, e utilizando variáveis artificiais (*dummy*), diga como procederia para testar a validade dessa proposição:

a) relativamente ao consumo autónomo;

b) relativamente à propensão marginal a consumir, e

c) relativamente a ambos os coeficientes conjuntamente.

d) Será mesmo razoável considerar que as referidas hipóteses são satisfeitas por este modelo? Justifique devidamente, discutindo a validade de cada uma das hipóteses.

3 O Modelo de Regressão com Regressores Pré-Determinados

Na secção anterior apresentaram-se os resultados exactos ou de pequenas amostras para a estimação e inferência OLS do modelo de regressão linear. Como se referiu, a hipótese H2, de exogeneidade estrita, é pouco realista para a maioria das aplicações com séries temporais. Nesta secção essa hipótese será abandonada, o mesmo acontecendo com a hipótese de normalidade. Os resultados que se obterão são válidos apenas assimptoticamente, isto é, em grandes amostras.

3.1 Alguns resultados preliminares

Pressupõe-se que os vários conceitos de convergência bem como os resultados teóricos que os relacionam e que os permitem manipular algebricamente são conhecidos. Em Hayashi (2000), pp. 88-94, encontra-se um sumário dessas matérias e em Wooldridge (2002), pp. 35-46, um tratamento mais extenso. Aqui tratar-se-á quase exclusivamente de apresentar duas versões de leis dos grandes números e uma do teorema de limite central. Embora estes resultados sejam assumidos conhecidos, o objectivo principal é de os contrastar com outros mais adequados para séries temporais, que serão apresentados mais adiante.

Recorde-se que um estimador $\hat{\theta}$, consistente (para o parâmetro θ), é dito ser assimptoticamente normal se

$$\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow{d} N(0, \Sigma).$$

Recorde-se também que a multiplicação do erro de amostragem por $\sqrt{T}$ é usualmente necessária para que a distribuição limite não seja degenerada, isto é, para que se possa efectuar inferência (fazer testes de hipóteses e construir intervalos de confiança) sobre o parâmetro θ, válida assimptoticamente. A essa multiplicação chama-se, por vezes, de transformação estabilizadora da variância. De um estimador nestas condições também se diz que é $\sqrt{T} - consistent$, e diz-se que ele converge para o parâmetro à taxa (ou à velocidade) $\sqrt{T}$. Por outro lado, um acrónimo usado para um estimador consistente e assimptoticamente normal é CAN.

A matriz de covariâncias Σ é chamada de matriz de covariâncias assimptótica e, por vezes, é representada com $\text{Avar}(\hat{\theta}_T)$. Embora seja também essa a convenção que será seguida aqui, note-se que alguns autores usam a notação $\text{Avar}(\hat{\theta}_T)$ para representar $(1/T)\Sigma$.

Vejamos, agora, duas leis dos grandes números (LLNs, *laws of large numbers*). Comecemos por considerar uma sucessão de variáveis aleatórias $\{z_t\}$ e a sua média amostral, $\bar{z}_T = (1/T)\sum_{t=1}^{T} z_t$. As leis dos grandes números estabelecem as condições sob as quais a sucessão $\{\bar{z}_T\}$ converge em probabilidade ou quase certamente (a.s.)

Uma versão da LLN fraca de Chebychev

Se $\lim_{T\to\infty} E(\bar{z}_T) = \mu$ e $\lim_{T\to\infty} Var(\bar{z}_T) = 0$, então

$$\bar{z}_T \xrightarrow{p} \mu.$$

Segunda lei forte dos grandes números de Kolmogorov

Se a sucessão $\{z_t\}$ é *i.i.d.* com $E(z_t) = \mu$, então

$$\bar{z}_T \xrightarrow{a.s.} \mu.$$

Note-se que esta LLN requer que as variáveis da sucessão sejam *i.i.d.* mas não requer que tenham variância finita. Recorde-se, ainda, que a convergência quase certa implica a convergência em probabilidade.

Por outro lado, a única versão de teorema de limite central (CLT) que se recordará aqui refere-se a sucessões de vectores aleatórios *i.i.d.*.

CLT de Lindeberg-Levy

Seja a sucessão de vectores aleatórios $\{z_t\}$, *i.i.d.*, com $E(z_t) = \mu$ e $Var(z_t) = \Sigma$. Então

$$\sqrt{T}(\bar{z}_T - \mu) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(z_t - \mu) \xrightarrow{d} N(0, \Sigma).$$

Isto é, a sucessão de vectores aleatórios $\{\sqrt{T}(\bar{z}_T - \mu)\}$ converge em distribuição para um vector aleatório cuja distribuição é $N(0, \Sigma)$.

3.2 Conceitos básicos

Embora a maioria dos conceitos básicos de séries temporais sejam considerados conhecidos, nesta secção revisitam-se alguns deles. Possivelmente, alguns não serão ainda conhecidos. Tal como anteriormente, segue-se Hayashi (2000), que se baseia em Hamilton (1994), muito de perto.

Processos estritamente estacionários

Um processo estocástico $\{z_t\}$ é estritamente estacionário se, para qualquer inteiro finito r e para qualquer conjunto de índices t_1, $t_2, \ldots$, t_r, a distribuição conjunta de $(z_t, z_{t_1}, z_{t_2}, \ldots, z_{t_r})$ depende somente dos intervalos que separam as datas, $t_1 - t$, $t_2 - t$, $t_3 - t$, ..., $t_r - t$, mas não de t.

Por exemplo, a distribuição de (z_2, z_8) terá que ser igual à de (z_9, z_{15}). Só a posição relativa na sucessão interessa para a distribuição. Então, em particular, a distribuição de z_t não deverá depender de t, de forma que a média, a variância e outros momentos, se existirem, são os mesmos para todo o t. Um exemplo básico de um processo estritamente estacionário é o de uma sucessão de variáveis aleatórias *i.i.d.* (que, adicionalmente, nem sequer apresenta qualquer dependência temporal).

Refira-se, também, que qualquer transformação ou função contínua (mensurável) de um processo estritamente estacionário é, em geral, ainda estritamente estacionária. Por exemplo, se o processo vectorial $\{z_t\}$ é estritamente estacionário, $\{z_t\, z_t'\}$ também é.

Saliente-se que, se um processo vectorial $\{z_t\}$ for estacionário, então cada um dos seus elementos, $\{z_{it}\}$, também é. Todavia, o inverso não é verdadeiro: não basta que cada elemento de um processo vectorial seja estacionário para que o vector o seja. Veja-se Hayashi (2000), p. 99, para um exemplo.

Processos estacionários em covariância

Um processo estocástico $\{z_t\}$ é fracamente estacionário, ou em covariância, se:

i. $E(z_t)$ não depende de t, e

ii. $Cov(z_t, z_{t-j})$ existe (finita) e só depende de j, não de t (por exemplo, $Cov(z_3, z_7) = Cov(z_{15}, z_{19})$).

Ou seja, a média é constante ao longo tempo e só a distância entre datas interessa para as (auto)covariâncias de um processo estacionário em covariância. Como é sabido, a condição ii. também implica que a variância do processo (que existe, finita) seja constante ao longo do tempo.

Obviamente, se um processo é estritamente estacionário e se a variância e as covariâncias são finitas, então o processo é estacionário em covariância. Todavia, há processos que são estacionários em covariância mas não em sentido estrito.

Tal como na generalidade da literatura, quando se referir que um processo é estacionário deverá entender-se que é em covariância ou em sentido fraco.

Processos ruído branco

Em Hamilton (1994) e em Hayashi (2000) um processo é chamado de ruído branco se for estacionário, com média nula e sem autocorrelação:

$$E(z_t) = 0 \quad \text{e} \quad Cov(z_t, z_{t-j}) = 0, \forall t \text{ e } j \neq 0.$$

Adicionalmente, se além de possuir média nula e variância finita o processo for *i.i.d*, então é chamado de processo ruído branco independente. Aqui, todavia, não será seguida a definição acima nem esta última convenção e adoptar-se-á a definição mais corrente na literatura, bem como a notação usual, $\{\epsilon_t\}$:

$$\text{o processo } \{\epsilon_t\} \text{ é um processo ruído branco se } \epsilon_t \sim iid(0, \sigma^2).$$

Se, adicionalmente, $\epsilon_t \sim N$, isto é, $\epsilon_t \sim iid\, N(0,\, \sigma^2)$, falamos em ruído branco Gaussiano.

Ergodicidade

O principal problema da análise de séries temporais é que só observamos uma trajectória do processo estocástico, isto é, apenas uma das suas possíveis realizações. Se tivéssemos acesso a várias trajectórias poderíamos estimar, por exemplo, a média, para uma determinada data, do processo. A essa média estimada (da população) chama-se usualmente média de conjunto (*ensemble mean*). Todavia, ao contrário do que fazemos com as simulações de Monte Carlo, não podemos relançar inúmeras vezes a história. Como é sabido, os dados da macroeconomia não são experimentais. Ora o conceito de ergodicidade tem a ver com a possibilidade de estimar de forma consistente a média ou outros momentos de conjunto com base na única realização que dispomos do processo estocástico.

Mais concretamente, se a distribuição da variável económica permanecer inalterada ao longo do tempo — estacionaridade —, a particular série temporal de T observações pode ser vista como T diferentes observações da mesma distribuição. Além disso, se o processo não for muito (temporalmente) dependente — ergodicidade —, cada elemento da série conterá informação não existente noutros elementos e a média temporal da série será consistente para a média de conjunto. O mesmo se poderá dizer relativamente a outros momentos.

Não será fornecida qualquer definição formal e rigorosa de processo (estacionário) ergódico [7]. No seu lugar adopta-se uma definição intuitiva: um processo estacionário é ergódico se for assimptoticamente independente, isto é, se duas variáveis aleatórias muito distantes na sucessão forem quase independentes. Ou ainda, se à medida que a distância entre os dois elementos do processo aumenta a independência entre eles tem tendência para aumentar. É neste sentido que Wooldridge (2006) emprega a designação, bastante feliz, de "fracamente dependente", para se referir a um processo ergódico. Todavia, seguiremos a designação mais comum.

De forma ainda menos rigorosa, consideraremos que um processo estacionário é ergódico se

$$Cov(y_t, y_{t-j}) \to 0 \text{ quando } j \to \infty,$$

isto é, se for assimptoticamente não (auto)correlacionado. Adicionalmente, a convergência para zero deverá ocorrer ainda de forma relativamente rápida. Dito de outra forma, exige-se que a memória do processo decaia com o tempo e que o decaimento seja relativamente rápido.

Note-se ainda que:

a) ao contrário da estacionaridade, a ergodicidade não é verificável empiricamente;

b) a um processo estacionário que seja ergódico chama-se estacionário ergódico;

c) em muitas aplicações as condições de estacionaridade e de ergodicidade são idênticas, razão pela qual muitos autores se referem apenas à primeira;

d) todavia, existem exemplos de processos que são estacionários mas não ergódicos (veja-se Hamilton, 1994, p. 47 e Canova, F., 2007, *Methods for Applied Macroeconomic Research*, p. 11).

Consideraremos que estes exemplos são "patológicos" e, frequentemente, quando se referir que um processo é estacionário estar-se-á a considerar que é também ergódico.

Como exemplos simples de processos ergódicos estacionários considerem-se:

a) os processos ruído branco, $\epsilon_t \sim iid(0, \sigma^2)$;

[7] Existem várias na literatura; veja-se, por exemplo, Hayashi (2000), p. 101, Hendry (1995), pp. 730-2 ou White (1999), p. 44. Em Breimann, L. (1969), *Probability*, pp. 313-7, encontra-se uma discussão interessante.

b) os processos AR(1) estacionários: $y_t = \rho y_{t-1} + \epsilon_t$, com $|\rho| < 1$.

Em seguida apresenta-se um resultado que é uma generalização muito importante da LLN de Kolmogorov.

Teorema ergódico

Seja $\{z_t\}$ um processo estacionário e ergódico com $E(z_t) = \mu$. Então,

$$\bar{z}_T \equiv \frac{1}{T}\sum_{t=1}^{T} z_t \xrightarrow{a.s.} \mu.$$

Ou seja, note-se que a hipótese *i.i.d.*, altamente irrealista para séries temporais, requerida pela LLN de Kolmogorov, não é necessária. O processo pode ter dependência temporal desde que esta tenda a desaparecer no longo prazo.

Adicionalmente, se $\{z_t\}$ for estacionário ergódico e se $f(.)$ for uma função contínua (mensurável), $\{f(z_t)\}$ também é estacionário ergódico. Desta forma, se existir (finito) um qualquer momento de um processo estacionário ergódico, ele é estimado de forma consistente pelo respectivo momento amostral.

Em particular, se $\{z_t\}$ for estacionário ergódico e se existir $E(z_t z_t')$, então

$$\frac{1}{T}\sum_{t=1}^{T} z_t z_t' \xrightarrow{a.s.} E(z_t z_t').$$

Martingalas

Seja z_{it} um elemento do vector z_t. O processo escalar $\{z_{it}\}$ é uma martingala relativamente a $\{z_t\}$ se

$$E(z_{it}|\, z_{t-1}, z_{t-2}, \ldots, z_1) = z_{i,t-1} \text{ para } t \geq 2.$$

Ao conjunto de informação condicionadora $(z_{t-1}, z_{t-2}, \ldots, z_1)$ chama-se usualmente de conjunto de informação no ponto ou data $t-1$.

Se o conjunto de informação se limitar à sua própria história passada, o processo $\{z_{it}\}$ é chamado simplesmente de martingala:

$$E(z_{it}|\, z_{i,t-1}, z_{i,t-2}, \ldots, z_{i,1}) = z_{i,t-1} \text{ para } t \geq 2.$$

E naturalmente, o processo vectorial $\{z_t\}$ é uma martingala se

$$E(z_t|\, z_{t-1}, \ldots, z_1) = z_{t-1} \text{ para } t \geq 2.$$

Se o processo escalar $\{w_t\}$ for uma martingala relativamente ao processo vectorial $\{z_t\}$ e se for um elemento do segundo, então também é uma martingala:

$$E(w_t|\, w_{t-1}, \ldots, w_1) = w_{t-1} \;\text{ para }\; t \geq 2.$$

Deixa-se esta pequena proposição para demonstração (usando a propriedade dos valores esperados iterados).

Algumas teorias económicas implicam que certos processos estocásticos sejam martingalas. É esse o caso da versão de R. E. Hall da hipótese do rendimento permanente. Considerando que z_t representa um vector contendo informação sobre várias variáveis macro-económicas como o rendimento, a moeda, o consumo privado (C), etc., deverá ter-se, de acordo com essa hipótese,

$$E(C_t|\, z_{t-1}, z_{t-2}, \ldots, z_1) = C_{t-1},$$

isto é, as variações futuras do consumo privado deverão ser imprevisíveis. Naturalmente, daqui também decorre que,

$$E(C_t|\, C_{t-1}, C_{t-2}, \ldots, C_1) = C_{t-1}.$$

Também a versão da hipótese de eficiência dos mercados de activos formulada por E. Fama implica que os preços (ou os seus logaritmos) dos activos financeiros devem seguir uma martingala.

Passeios aleatórios

O exemplo mais simples e conhecido de martingala é o do processo de passeio aleatório. Sendo $\{g_t\}$ um vector ruído branco (independente), um passeio aleatório é uma sucessão de somas cumulativas:

$$z_1 = g_1, \quad z_2 = g_1 + g_2, \quad \ldots, \quad z_t = g_1 + g_2 + \ldots + g_t, \ldots,$$

provando-se facilmente que é uma martingala.

Conhecendo $\{z_t\}$, a sucessão ruído branco que lhe está subjacente, $\{g_t\}$, pode ser obtida facilmente por diferenciação:

$$g_1 = z_1, \quad g_2 = \Delta z_2 = z_2 - z_1, \quad \ldots, \quad g_t = \Delta z_t = z_t - z_{t-1}, \ldots,$$

isto é, a primeira diferença de um passeio aleatório é um processo ruído branco.

Diferenças de martingala

Um processo vectorial $\{g_t\}$ com média nula, $E(g_t) = 0$, é uma diferença de martingala (m.d.s., *martingale difference sequence*) se o valor esperado condicional nos seus valores passados também for zero:

$$E(g_t|\, g_{t-1}, g_{t-2} \ldots, g_1) = 0, \quad \text{para } t \geq 2.$$

O nome do processo advém do facto de a soma cumulativa, $\{z_t\}$, formada a partir dele ser uma martingala. Inversamente, se um processo é uma martingala, o processo das suas primeiras diferenças é uma m.d.s. .

Se um processo é uma m.d.s., então não tem autocorrelação, $Cov(g_t, g_{t-j}) = 0,\ \forall t, j \neq 0$. De facto, sem perda de generalidade considere-se $j \geq 1$ e note-se que, como a média é nula, basta provar que $E(g_t g'_{t-j}) = 0$. Tem-se

$$\begin{aligned} E(g_t g'_{t-j}) &= E[E(g_t g'_{t-j})|\, g_{t-j}] \\ &= E[E(g_t|\, g_{t-j})\, g'_{t-j}]. \end{aligned}$$

Ora, como $j \geq 1$, $g_{t-j} \in \{g_{t-1}, \ldots, g_{t-j}, \ldots, g_1\}$. Portanto, pela propriedade dos valores esperados iterados,

$$E(g_t|\, g_{t-j}) = E[E(g_t|\, g_{t-1}, \ldots, g_{t-j}, \ldots, g_1)\, |g_{t-j}] = E(0|\, g_{t-j}) = 0.$$

Podemos então concluir que a condição para que um processo seja uma m.d.s. é mais forte que a de ausência de autocorrelação: se um processo não tem autocorrelação não pode ser previsto com base numa função linear dos seus valores passados; mas se é uma m.d.s. não pode ser previsto com base em nenhuma função dos seus valores passados, seja a função linear ou não.

Todavia, a condição para que um processo seja uma m.d.s. é mais fraca que a de independência porque os momentos de ordem superior a um, como $E(g_t^2|g_{t-1}, \ldots, g_1)$, podem depender de valores passados. Por exemplo, com $\epsilon_t \sim iid\, N(0, \sigma^2)$, o processo $y = \epsilon_t \epsilon_{t-1}$ é uma m.d.s. mas não é independente.

Um processo ARCH [8] é um exemplo de uma m.d.s. . Em particular, considere-se o processo $\{g_t\}$, ARCH(1):

$$g_t = \sqrt{\alpha_0 + \alpha_1 g_{t-1}^2}\, \epsilon_t, \quad \text{com} \quad \epsilon_t \sim iid(0, 1).$$

[8] Estes processos são estudados de forma mais profunda nas disciplinas de séries temporais.

Com efeito,

$$\begin{aligned}
E(g_t|\, g_{t-1}, \ldots, g_1) &= E(\sqrt{\alpha_0 + \alpha_1 g_{t-1}^2}\, \epsilon_t|\, g_{t-1}, \ldots, g_1), \\
&= \sqrt{\alpha_0 + \alpha_1 g_{t-1}^2}\, E(\epsilon_t|\, g_{t-1}, \ldots, g_1), \\
&= \sqrt{\alpha_0 + \alpha_1 g_{t-1}^2}\, E(\epsilon_t), \\
&= \sqrt{\alpha_0 + \alpha_1 g_{t-1}^2}\, 0 = 0,
\end{aligned}$$

onde a antepenúltima desigualdade resulta do facto de ϵ_t ser independente de $(g_{t-1}, \ldots, g_1)$.

De forma semelhante, também se pode mostrar que

$$Var(g_t|\, g_{t-1}, \ldots, g_1) = E(g_t^2|\, g_{t-1}, \ldots, g_1) = \alpha_0 + \alpha_1 g_{t-1}^2,$$

de forma que o processo apresenta heterocedasticidade condicional própria, no sentido em que o segundo momento condicional é função do próprio passado do processo.

Um processo ARCH(1) é estritamente estacionário e ergódico se $|\alpha_1| < 1$ e se g_1 tiver uma certa distribuição ou se o processo tiver sido iniciado no passado muito distante. Nessas condições, também se pode mostrar que $E(g_t^2) = \alpha_0/(1 - \alpha_1)$.

Para assentar ideias:

a) um processo ruído branco (independente) é, obviamente, uma m.d.s. estacionária com variância finita;

b) uma m.d.s. não tem autocorrelação;

c) as condições para que um processo seja uma m.d.s. estacionária com variância finita são mais fracas que a de ruído branco (independente); por exemplo, um processo ARCH(1) com $|\alpha_1| < 1$ satisfaz as primeiras mas não as segundas.

3.3 Outro teorema do limite central

O CLT seguinte generaliza o CLT de Lindeberg-Levy a processos m.d.s. estacionários e ergódicos.

CLT para processos m.d.s. estacionários e ergódicos (Billingsley)

Seja $\{g_t\}$ um vector m.d.s. estacionário e ergódico com

$$E(g_t\, g_t') = \Sigma,$$

e seja $\bar{g} = \frac{1}{T}\sum_{t=1}^{T} g_t$. Então,

$$\sqrt{T}\,\bar{g} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} g_t \xrightarrow{d} N(0, \Sigma).$$

Em primeiro lugar, obviamente, as condições deste CLT são mais fracas que as do de Lindeberg-Levy: não é necessário que o processo seja i.i.d. . Note-se também que, como a média de uma m.d.s. é nula, não é necessário subtrair a média. Por esta razão e também porque o processo é estacionário, tem-se $Var(g_t) = \Sigma$, que não depende de t (e todos os elementos desta matriz deverão existir, finitos).

Note-se, contudo, que existe ainda um CLT com condições mais fracas que este.

3.4 As hipóteses do modelo

Veremos agora com bastante detalhe as hipóteses de um modelo que tem muito maior aplicabilidade prática que o modelo clássico da secção anterior, pois a hipótese de exogeneidade estrita será substancialmente enfraquecida. Também não será necessário empregar nenhuma hipótese sobre a distribuição dos erros. Todavia, como veremos também, deixaremos de ter as boas propriedades do estimador OLS em amostras finitas e passaremos a ter apenas propriedades assimptóticas [9].

[H1'] – Linearidade. Idêntica à H1 do modelo clássico.

[H2'] – Estacionaridade ergódica. O processo estocástico vectorial, de dimensão $(K+1)$, $\{y_t,\, x_t\}$, é conjuntamente estacionário e ergódico.

Obviamente, deve salientar-se que esta hipótese não era necessária no modelo clássico. Ela representa o principal preço a pagar pelo enfraquecimento da hipótese de exogeneidade estrita (que veremos a seguir).

Naturalmente, esta hipótese permite considerar os dados seccionais de amostragem casual como caso particular. Isto é, se $\{y_t,\, x_t\}$ são i.i.d. a hipótese é trivialmente satisfeita.

[9] Ao contrário de Hayashi (2000), p. 109, não se discute aqui em detalhe a distinção subtil e algo "filosófica" entre modelo e processo de geração de dados (DGP, *data generating process*). Como DGP entende-se uma especificação completa do processo que gerou a amostra finita de dados, incluindo valores específicos para os parâmetros, o que o permite simular num computador. Por outro lado, um modelo é um conjunto de DGPs. Para mais detalhes, veja-se, por exemplo, Davidson e MacKinnon (2004), pp. 30-1 e 86-7.

Note-se ainda que a satisfação desta hipótese implica que também o processo dos erros, $\{u_t\}$, é estacionário e ergódico. Basta notar que $u_t = y_t - x_t'\beta$. Por outro lado, daqui decorre também que, $E(u_t^2)$ não varia com t, ou seja, deverá existir homocedasticidade não condicional (ou marginal):

$$E(u_t^2) = \sigma^2,\ t = 1, 2, \ldots, T.$$

Todavia, sublinhe-se que poderá existir heterocedasticidade condicional, isto é, poderá acontecer que $E(u_t^2|\, x_t)$ seja uma função de x_t.

[H3'] – Regressores pré-determinados. Todos os regressores são pré-determinados, no sentido em que são ortogonais ao erro contemporâneo:

$$E(x_{tj}\, u_t) = 0,\ \forall t(= 1, 2, \ldots, T) \text{ e } j(= 1, 2, \ldots, K).$$

Também se pode escrever $E[x_t(y_t - x_t'\beta)] = 0$, ou ainda, com a convenção

$$g_t \equiv x_t u_t,$$

$$E(g_t) = 0.$$

Muito há a dizer sobre esta hipótese:

a) a definição adoptada de regressor pré-determinado não é universal pois exige apenas a ortogonalidade relativamente ao erro contemporâneo, isto é, da equação em que aparece o regressor (não necessariamente com o mesmo índice temporal). Historicamente, era usual dizer-se que o j-ésimo regressor era pré-determinado quando $E(x_{t-i,j}\, u_t) = 0,\ \forall i \geq 0$, e não somente para $i = 0$; ou seja, exigia-se ortogonalidade dos erros não só em relação aos regressores contemporâneos mas também relativamente aos valores passados dos regressores. No entanto, embora mais fraca, a condição adoptada é suficiente para obter os resultados assimptóticos. Por outro lado, talvez para evitar essa confusão, autores como Wooldridge (2006) não se referem a regressores pré-determinados mas sim a *contemporaneamente exógenos*, o que parece ser mais adequado. Todavia, é esta a definição adoptada em Hamilton (1994) e em Hayashi (2000).

b) Claramente, esta hipótese é mais fraca que H2: qualquer regressor estritamente exógeno é pré-determinado, mas o recíproco não é verdadeiro. Em particular, H2 não permitia a existência de correlação do erro u_t com os regressores futuros, x_{t+i}, com $i \geq 1$. Pelo contrário, a nova H3' permite a existência dessa correlação.

Assim, por exemplo, os processos AR(1) (com erros i.i.d.) satisfazem a nova hipótese. Note-se que, para estes processos, a exigência de ortogonalidade se refere a $E(y_{t-1}\, u_t) = 0$ (não, obviamente, a $E(y_t\, u_t) = 0$, impossível de satisfazer).

c) No lugar da condição de ortogonalidade poder-se-ia ter considerado a condição

$$E(u_t|\, x_t) = 0,$$

que é a condição de exogeneidade contemporânea de Wooldridge (2006). Todavia, esta condição é mais forte que a de ortogonalidade (reveja os comentários sobre H2, na secção anterior). Em particular, ela implica que, para qualquer função $f(.)$ (mensurável) de x_t, $f(x_t)$ é ortogonal a u_t:

$$E[f(x_t)u_t] = E[E(f(x_t)\, u_t|x_t] = E[f(x_t)\, E(u_t|x_t)] = 0.$$

Adicionalmente, a condição acima é equivalente a exigir-se

$$E(y_t|\, x_t) = x_t'\, \beta.$$

d) No caso em que o modelo tem termo independente — como tem vindo a ser assumido —, a hipótese implica que

$$E(u_t) = 0,\ \forall t \quad \text{e} \quad Cov(x_t, u_t) = 0,\ \forall t,$$

o que significa que se pode afirmar que os regressores não estão contemporaneamente correlacionados com os erros.

e) Obviamente, esta hipótese continua a excluir a possibilidade de existir algum regressor endógeno, isto é, algum regressor x_j tal que $E(x_{tj}u_t) \neq 0$ para algum t.

[H4'] – Condição de característica. A matriz quadrada de ordem K $E(x_t\, x_t')$ é regular (e, portanto, é finita). Convenciona-se representar esta matriz com Σ_{xx}.

Esta hipótese é a análoga de H3 (de ausência de colinearidade). De facto, considere-se o processo

$$S_{xx} = \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \frac{1}{T} X'X,$$

e note-se que se trata do processo dos momentos cruzados dos regressores. Com base nesta hipótese e em H2' e empregando o teorema ergódico, tem-se

$$S_{xx} \xrightarrow{a.s.} \Sigma_{xx}.$$

Ou seja, pelas hipóteses H2' e H4' e para T suficientemente grande, $S_{xx} \equiv (1/T)X'X$ é regular. Mas como $(1/T)X'X$ é regular se e só se $r(X) = K$, a hipótese H3 é satisfeita em probabilidade no limite (quando $T \to \infty$).

[H5'] – $\{\mathbf{g_t}\}$ é uma m.d.s. com segundos momentos finitos. $\{g_t\}$ é uma m.d.s. e a matriz quadrada de ordem K de momentos cruzados $E(g_t\, g_t')$ é regular.

Considerando o processo $\{\bar{g}\} = (1/T)\sum_{t=1}^{T} g_t$ convencione-se representar com S a matriz de covariâncias assimptótica da distribuição de $\sqrt{T}\,\bar{g}$, isto é,

$$S \equiv \text{Avar}(\bar{g}).$$

Desta forma, pela hipótese H2' e pelo CLT para processos m.d.s. estacionários e ergódicos está-se a convencionar que

$$S = E(g_t\, g_t'),$$

e note-se que esta é uma matriz de momentos de ordem 4. Com efeito, como $g_t = x_t u_t$, ela pode escrever-se como $E(u_t^2 x_t x_t')$ e o seu elemento da linha i e coluna j é $E(u_t^2 x_{ti} x_{tj})$. Para estimar a matriz S de forma consistente será necessário introduzir uma hipótese adicional (veja-se a subsecção 3.7, mais adiante).

Note-se, também, que esta hipótese é mais forte que H3'. Com efeito, basta recordar que, por definição, um processo m.d.s. tem média nula, para se ter imediatamente $E(x_t u_t) = 0$. No entanto, embora ela não seja necessária para mostrar a consistência do estimador OLS, sê-lo-á para obter a sua distribuição assimptótica.

A interpretação intuitiva desta hipótese não é simples mas pode considerar-se uma condição suficiente para que ela seja satisfeita, e que é mais simples, que é:

$$E(u_t |\, u_{t-1}, u_{t-2}, \ldots, u_1, x_t, x_{t-1}, \ldots, x_1) = 0,\ \forall t, \tag{7}$$

onde se nota a inclusão dos regressores contemporâneos (para além dos desfasados). Com efeito, note-se que, pela propriedade dos valores esperados iterados,

$$E(g_t|\, g_{t-1}, \ldots, g_1) = E[\, E(g_t|\, u_{t-1}, u_{t-2}, \ldots, u_1, x_t, x_{t-1}, \ldots, x_1) \,|\, g_{t-1}, \ldots, g_1],$$

pois há mais informação no conjunto informativo "de dentro" do que no "de fora". Mas então, como

$$E(g_t|\, g_{t-1}, \ldots, g_1) = E[\, x_t\, E(u_t|\, u_{t-1}, u_{t-2}, \ldots, u_1, x_t, x_{t-1}, \ldots, x_1)|\, g_{t-1}, \ldots, g_1],$$

é igual a zero se a condição da equação (7) for satisfeita, essa condição é suficiente para que $\{g_t\}$ seja uma m.d.s. .

Note-se, ainda, que a condição de (7) tem duas implicações muito importantes:

a) os erros não podem estar autocorrelacionados, e

b) os erros não podem estar correlacionados com os regressores contemporâneos e passados [10].

Por outro lado, quando o modelo tem termo independente, a própria condição da hipótese H5' implica (requer) que os erros não estão (estejam) autocorrelacionados. De facto, como o primeiro elemento do vector $g_t (\equiv x_t u_t)$ é simplesmente u_t, a hipótese implica que

$$E(u_t|\, g_{t-1}, g_{t-2}, \ldots, g_1) = 0.$$

Mas então, de novo pela propriedade dos valores esperados iterados, o processo $\{u_t\}$ é uma m.d.s. (escalar), isto é,

$$E(u_t|\, u_{t-1}, u_{t-2}, \ldots, u_1) = 0,$$

e, portanto, é um processo sem autocorrelação.

Finalmente, note-se ainda que é graças a esta hipótese que S é igual a $E(g_t g_t')$. Sem ela, a expressão de S seria mais complicada, pois envolveria as autocovariâncias de g_t.

[10] Deixa-se a demonstração ao cuidado do leitor e nota-se apenas que é muito semelhante à de que uma m.d.s. não tem autocorrrelação.

3.5 Propriedades assimptóticas

As bases para as propriedades assimptóticas da estimação e inferência OLS foram estabelecidas e falta apenas tratar do problema da estimação consistente de $S(\equiv \text{Avar}(\bar{g}) = E(g_t\, g_t') = E(u_t^2 x_t x_t'))$. O tratamento desse problema é relegado para uma fase posterior e, por agora, será assumido que existe um estimador consistente para essa matriz.

Mostrar-se-á então que o estimador OLS é consistente e assimptoticamente normal (CAN). Recorda-se que, embora não se empregue nenhum índice, o comportamento do estimador $\widehat{\beta}$ depende da dimensão da amostra, T.

Proposição 3.1 – Propriedades assimptóticas do estimador OLS de β.

a) (Consistência) Sob as hipóteses H1' a H4' (isto é, sem a hipótese H5'), tem-se

$$\operatorname{plim} \widehat{\beta} = \beta.$$

b) (Normalidade assimptótica) Sob as hipóteses H1', H2', H4' e H5', e com $\Sigma_{xx} \equiv E(x_t x_t')$, $g_t \equiv x_t u_t$ e $S = E(g_t g_t')$, tem-se

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \text{Avar}(\widehat{\beta})) \text{ quando } T \to \infty,$$

onde

$$\text{Avar}(\widehat{\beta}) = \Sigma_{xx}^{-1} S\, \Sigma_{xx}^{-1}.$$

c) (Estimação consistente de $\text{Avar}(\widehat{\beta})$) Suponha-se que existe um estimador consistente da matriz S, quadrada de ordem K, $\widehat{S}$. Então, sob a hipótese H2', a matriz $\text{Avar}(\widehat{\beta})$ pode ser estimada de forma consistente com

$$\widehat{\text{Avar}(\widehat{\beta})} = S_{xx}^{-1} \widehat{S}\, S_{ss}^{-1},$$

onde S_{xx} é a média amostral de $x_t x_t'$ (reveja-se (4), por exemplo):

$$S_{xx} = \frac{1}{T} \sum_{t=1}^{T} x_t x_t' = \frac{1}{T} X'X.$$

Para provar a) comecemos por escrever o erro de amostragem em termos de médias amostrais:

$$\begin{aligned}
\widehat{\beta} - \beta &= (X'X)^{-1} X'u \\
&= \left(\tfrac{1}{T} X'X\right)^{-1} \left(\tfrac{1}{T} X'u\right) \\
&= \left(\tfrac{1}{T} \textstyle\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\tfrac{1}{T} \textstyle\sum_{t=1}^{T} x_t u_t\right) \\
&= S_{xx}^{-1}\, \bar{g},
\end{aligned}$$

onde $\bar{g} = (1/T) \sum g_t$.

Ora, como pela hipótese H2' o processo $\{x_t x_t'\}$ é estacionário e ergódico, $S_{xx} \xrightarrow{p} \Sigma_{xx}$. E, como por H4', esta matriz é regular, tem-se $S_{xx}^{-1} \xrightarrow{p} \Sigma_{xx}^{-1}$. Também por H2' (e H1'), $\bar{g} \xrightarrow{p} E(g_t)$, que é um vector nulo, por H3'. Desta forma,

$$\text{plim}(\widehat{\beta} - \beta) = \Sigma_{xx}^{-1} 0 = 0,$$

e, portanto, $\text{plim}\widehat{\beta} = \beta$.

Por outro lado, para provar b) multiplique-se a última expressão obtida acima para o erro de amostragem por $\sqrt{T}$:

$$\sqrt{T}(\widehat{\beta} - \beta) = S_{xx}^{-1}(\sqrt{T}\,\bar{g}).$$

Ora, por H5', $\sqrt{T}\,\bar{g} \xrightarrow{d} N(0, S)$. Portanto, por uma das propriedades básicas para sucessões de variáveis aleatórias [11], tem-se

$$\sqrt{T}(\widehat{\beta} - \beta) \xrightarrow{d} N\left[0, \Sigma_{xx}^{-1} S\, (\Sigma_{xx}^{-1})'\right].$$

Mas como a matriz Σ_{xx} é simétrica, este resultado pode escrever-se sob a forma da proposição. Assim, a distribuição do erro de amostragem escalado por $\sqrt{T}$ pode ser aproximada, em grandes amostras, pela distribuição normal. O estimador OLS é, neste caso, $\sqrt{T}$-*consistent*.

Relativamente a c), trata-se de uma aplicação imediata da estacionaridade ergódica. Em resumo, note-se que estas demonstrações seguiram o padrão usual da teoria assimptótica clássica:

a) começar por escrever as magnitudes a tratar sob a forma de momentos amostrais;

b) aplicar a LLN e o CLT relevantes a esses momentos (neste caso, respectivamente, o teorema ergódico e o resultado de Billinsgley para m.d.s. estacionários e ergódicos);

c) empregar as propriedades básicas da teoria assimptótica.

Recorde-se que, pela hipótese H2', $E(u_t^2) = \sigma^2$ para todo o t, ou seja, existe homocedasticidade não condicional. Adicionalmente, quando o modelo tem termo independente, σ^2 representa a variância não condicional dos erros ($Var(u_t) =$

[11] Se $x_T \xrightarrow{d} x \sim N(0, \Sigma)$ e $A_T \xrightarrow{p} A$, então $A_T x_T \xrightarrow{d} N(0, A\Sigma A')$.

σ^2). A proposição seguinte estabelece a consistência do estimador usual OLS desse parâmetro [12].

Proposição 3.2 – Estimação consistente da variância dos erros.
Desde que $E(u_t^2)$ exista (finito), representando os resíduos OLS com $e_t = y_t - x_t'\beta (t = 1, 2, \ldots, T)$ e sob as hipóteses H1' a H4', tem-se que

$$s^2 \equiv \widehat{\sigma^2} \equiv \frac{1}{T-K}\sum_{t=1}^{T} e_t^2 \xrightarrow{p} E(u_t^2).$$

3.6 Inferência estatística

Nesta subsecção continua a assumir-se que dispomos de um estimador consistente de $S(\equiv E(g_t g_t'))$, $\widehat{S}$. Tal como na secção anterior, só o problema dos testes de hipóteses lineares será tratado [13].

Comecemos por considerar um teste sobre um único coeficiente, com H_0 formalizada como $H_0 : \beta_j = \beta_j^0$. De acordo com a proposição 3.1, sob H_0:

$$\sqrt{T}(\widehat{\beta}_j - \beta_j^0) \xrightarrow{d} N[0, \mathrm{Avar}(\widehat{\beta}_j)],$$

onde $\mathrm{Avar}(\widehat{\beta}_j)$ é o j-ésimo elemento da diagonal principal da matriz $\mathrm{Avar}(\widehat{\beta})$. Naturalmente, teremos que estimar esta magnitude de forma consistente:

$$\widehat{\mathrm{Avar}(\widehat{\beta}_j)} \xrightarrow{p} \mathrm{Avar}(\widehat{\beta}_j).$$

Por conseguinte, usando novamente a propriedade da nota de pé-de-página 11,

$$t_j \equiv \frac{\sqrt{T}(\widehat{\beta}_j - \beta_j^0)}{\sqrt{\widehat{\mathrm{Avar}(\widehat{\beta}_j)}}} = \frac{\widehat{\beta}_j - \beta_j^0}{se^*(\widehat{\beta}_j)} \xrightarrow{d} N(0, 1),$$

onde se nota que a $\sqrt{T}$ do numerador cancela com a do denominador e onde se usa a convenção

$$se^*(\widehat{\beta}_j) \equiv \sqrt{\frac{1}{T}\widehat{\mathrm{Avar}(\widehat{\beta}_j)}} \equiv \sqrt{\frac{1}{T}(S_{xx}^{-1}\widehat{S}\,S_{xx}^{-1})_{jj}},$$

chamado de *se* (erro padrão) consistente sob heterocedasticidade, ou *se* robusto à heterocedasticidade ou, ainda, mais popularmente, *se* de White. A explicação

[12] Pode ver-se a demonstração em Hayashi (2000), pp. 115-6.
[13] Veja-se Hayashi (2000), pp. 121-2, para os testes de hipóteses não lineares.

para esta designação é bastante simples: recorde-se que os erros podem ser condicionalmente heterocedásticos e que, por conseguinte, a matriz $\widehat{\text{Avar}(\widehat{\beta})}$ acomoda essa possibilidade, isto é, estima de forma consistente $\text{Avar}(\widehat{\beta})$ mesmo nessas condições.

Por outro lado, para a distinguir da estatística-t do modelo clássico, a estatística-t anterior é usualmente designada de estatística-t robusta à heterocedasticidade, pretendendo-se com esta designação dizer que ela é válida mesmo na presença desse problema, isto é, que se baseia num estimador de $\text{Avar}(\widehat{\beta})$ que permite (e acomoda) essa possibilidade.

Para o caso particular dos testes de significância individuais, com $H_0 : \beta_j = 0$, têm-se também os rácios-t robustos à heterocedasticidade,

$$t_j = \frac{\widehat{\beta_j}}{se^*(\widehat{\beta}_j)} \xrightarrow{d} N(0,1).$$

Mostrou-se, assim, a primeira parte da proposição seguinte.

Proposição 3.3 – Estatísticas t e de Wald robustas.
Suponha-se que as hipóteses H1' a H5' são satisfeitas e que dispomos de um estimador consistente de S, $\widehat{S}$, e seja

$$\widehat{\text{Avar}(\widehat{\beta})} \equiv S_{xx}^{-1} \widehat{S}\, S_{xx}^{-1}.$$

Então:

a) sob $H_0 : \beta_j = \beta_j^0$, a estatística-$t$ definida acima converge em distribuição para uma $N(0,1)$;

b) sob $H_0 : R\beta = r$, com R uma matriz de tipo $(J \times K)$, com $r(R) = J$, tem-se

$$W \equiv T(R\widehat{\beta} - r)'\{R\,[\widehat{Avar(\widehat{\beta})}]\,R'\}^{-1}(R\widehat{\beta} - r) \xrightarrow{d} \chi^2_{(J)}.$$

Embora relativamente pequena, omite-se a demonstração de b) (veja-se Hayashi, 2000, p. 119). A designação da estatística (W) resulta do facto de se tratar de uma aplicação do princípio de Wald, pois a estimação é feita sem restrições (tanto β como $\text{Avar}(\widehat{\beta})$ são estimadas sem impôr as restrições de H_0).

Como é sabido, uma vez que os testes são assimptóticos, a sua dimensão real ou exacta é igual à nominal (α) só aproximadamente, com a aproximação a melhorar à medida que $T \to \infty$. Na realidade, a situação mais frequente, mesmo para amostras de dimensão razoável, é que a dimensão real seja superior à nominal, caso em que se fala em *distorções de dimensão* ou em *sobre-rejeições* de H_0.

Por outro lado, desde que as hipóteses do modelo sejam satisfeitas, ambos os testes são consistentes, no sentido em que, a sua potência contra valores alternativos dos parâmetros tende para 1 quando $T \to \infty$. Esta propriedade é bastante simples de verificar no caso dos testes baseados nas estatísticas-t.

Com efeito, considere-se um DGP que satisfaz as hipóteses H1' a H5' mas não a hipótese a testar, $H_0 : \beta_j = \beta_j^0$, suponha-se que a alternativa é bilateral, $H_1 : \beta_j \neq \beta_j^0$, e represente-se com β_j^a ($\neq \beta_j^0$) o verdadeiro valor de β_j. Recorde-se que a potência do teste contra esta alternativa é obtida calculando

$$\text{Prob}(|t_j| > t_{\alpha/2} \,|\, \beta_j^a),$$

e recupere-se a expressão da estatística-t

$$t_j \equiv \frac{\sqrt{T}(\widehat{\beta_j} - \beta_j^0)}{\sqrt{\widehat{\text{Avar}(\widehat{\beta}_j)}}}.$$

Ora, como o denominador converge para $\sqrt{\text{Avar}(\widehat{\beta}_j)}$ e como o numerador tende para $+\infty$ ou $-\infty$ (à taxa $\sqrt{T}$) dado que $\widehat{\beta}_j \xrightarrow{p} \beta_j^a \neq \beta_j^0$, a probabilidade acima tende para 1.

3.7 Estimação consistente de S

Para estimar a matriz de covariâncias assimptótica $\text{Avar}(\widehat{\beta})$, até agora assumiu-se a disponibilidade de um estimador consistente de $S = E(g_t g_t') = E(u_t^2\, x_t x_t')$. Nesta subsecção trataremos desse problema.

Empregando a estacionaridade ergódica, se observássemos os erros poderíamos estimar essa matriz de forma consistente com o respectivo momento amostral

$$(1/T) \sum u_t^2 x_t x_t'.$$

A solução para o problema é a usual: podemos substituir os erros por estimadores, desde que estes sejam consistentes. Isto é, sob condições adicionais que veremos a seguir, podemos empregar um estimador do tipo

$$\widehat{S} = \frac{1}{T} \sum_{t=1}^{T} e_t^2 x_t x_t',$$

onde, apesar de se ter empregue a notação adoptada para os resíduos OLS, os e_t's representam os resíduos obtidos com algum estimador consistente de β, $e_t = y_t - x't\widehat{\beta}$ (onde $\widehat{\beta}$ representa um estimador consistente de β, e não necessariamente o estimador OLS).

Mas para que este estimador seja consistente, ainda é necessário introduzir uma hipótese adicional, sobre os quartos momentos dos regressores.

[H6'] – Momentos finitos de ordem quatro dos regressores
Os momentos $E[(x_{ti}x_{tj})^2]$ existem e são finitos para todo i e $j(= 1, 2, \ldots K)$.

Com esta hipótese adicional, é possível mostrar-se a proposição seguinte.

Proposição 3.4 – Estimação consistente de S.
Suponha-se que o estimador $\widehat{\beta}$, empregue para obter os resíduos e_t, é consistente e que $S = E(g_t g_t')$ existe e é finita. Suponha-se ainda que as hipóteses H1', H2' e H6' são satisfeitas. Nessas condições, o estimador $\widehat{S}$ anterior é consistente para S.

Para perceber a razão da introdução de H6' considere-se o caso do modelo de regressão simples, sem termo independente, isto é, com x_t um escalar. Nesse caso tem-se

$$\begin{aligned}
e_t^2 &= (y_t - \widehat{\beta}x_t)^2, \\
&= (\beta x_t + u_t - \widehat{\beta}x_t)^2, \\
&= u_t^2 - 2(\widehat{\beta} - \beta)x_t u_t + (\widehat{\beta} - \beta)^2\, x_t^2.
\end{aligned}$$

Multiplicando os dois membros da igualdade anterior por x_t^2 e somando em t, tem-se

$$\frac{1}{T}\sum_{t=1}^{T} e_t^2 x_t^2 - \frac{1}{T}\sum_{t=1}^{T} u_t^2 x_t^2 = -2(\widehat{\beta} - \beta)\frac{1}{T}\sum_{t=1}^{T} u_t x_t^3 + (\widehat{\beta} - \beta)^2 \frac{1}{T}\sum_{t=1}^{T} x_t^4.$$

Repare-se então que, com H6', como o quarto momento $E(x_t^4)$ é finito, pela estacionaridade ergódica o respectivo momento amostral do último termo da expressão converge para esse valor. E como o estimador de β é consistente, todo esse termo converge em probabilidade para zero. Pode ainda mostrar-se que, desde que as hipóteses anteriores sejam satisfeitas, a média amostral de $u_t x_t^3$ também converge em probabilidade para um valor finito, de forma que o primeiro termo do lado direito da equação anterior também converge em probabilidadade para zero. Desta forma, o primeiro termo do lado esquerdo converge em probabilidade para o segundo.

Assim, a matriz que se empregará para estimar de forma consistente a matriz de covariâncias assimptótica do estimador OLS é

$$\widehat{\text{Avar}}(\widehat{\beta}) = S_{xx}^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} e_t^2 x_t x_t' \right) S_{xx}^{-1}.$$

Notando ainda que $\widehat{S}$ se pode escrever como

$$\widehat{S} = \frac{1}{T} X'BX,$$

onde B é a matriz diagonal de ordem T cujos elementos diagonais são os quadrados dos resíduos OLS ($B = \text{diag}(e_1^2, e_2^2, \ldots, e_T^2)$), o estimador de S também pode ser escrito como

$$\widehat{\text{Avar}}(\widehat{\beta}) = T(X'X)^{-1}(X'BX)(X'X)^{-1}.$$

É esta matriz, sob esta forma ou sob a anterior, que é frequentemente chamada de matriz de White (por ter sido proposta por este econometrista, em 1980, num artigo que ficou célebre). Embora seja bastante conveniente em muitas ocasiões, a sua utilização em pequenas amostras pode conduzir, com alguma frequência, a distorções de dimensão dos testes (isto é, a sobre-rejeições da hipótese nula). Existem várias correcções ou variantes para melhorar esse comportamento de pequenas amostras. Veja-se, por exemplo, Davidson e MacKinnon (1993), pp. 552-6.

3.8 O caso de homocedasticidade condicional

A teoria assimptótica que acabou de se apresentar é válida mesmo quando existe heterocedasticidade condicional (nos regressores). Na prática, ela é empregue de forma muito rotineira mesmo quando, *a priori*, não há qualquer indício concreto da existência desse problema. Duas razões contribuem fortemente para esse facto:

a) a simplicidade e conveniência do procedimento, que é válido sob qualquer forma de heterocedasticidade condicional (nos regressores);

b) a disponibilidade generalizada do procedimento em programas de regressão, mesmo nos menos potentes.

Para reforçar a), note-se que, com um custo muito reduzido, o procedimento permite dispensar facilmente preocupações e procedimentos especiais para um problema que pode causar distorções potencialmente graves sobre os métodos de inferência.

No entanto, deve salientar-se que o procedimento é muito mais útil para dados seccionais do que para os modelos de séries temporais económicas. Isto deve-se ao facto de o tipo de heterocedasticidade condicional que é acomodado ser pouco frequente com estes dados. Ou seja, nos modelos com dados temporais, raramente a variância dos erros é função das variáveis consideradas como regressores. É mais comum que a heterocedasticidade apareça sob a forma de regimes distintos, com mudanças algo bruscas, ou que seja do tipo autoregressivo, como no caso dos modelos ARCH [14]. Desta forma, a utilização do procedimento estudado anteriormente está muito longe de ser uma panaceia e pode, até, dar uma falsa sensação de segurança ou de conforto ao investigador empírico.

Mas, e se há homocedasticidade condicional? As estatísticas de teste desta secção devem continuar a ser empregues, ou, pelo contrário, podemos voltar a recorrer às velhas t's e F's do modelo clássico? E, em caso afirmativo, sob que condições? Introduza-se, então, a hipótese seguinte.

[H7'] – Homocedasticidade condicional:

$$E(u_t^2|\, x_t) = \sigma^2 > 0,\ \forall t.$$

E, pela propriedade dos valores esperados totais, tem-se que também $E(u_t^2) = \sigma^2$, $\forall t$. De resto, já tínhamos visto que, se $\{y_t, x_t\}$ é estacionário ergódico, também o processo $\{u_t\}$ é e existe homocedasticidade não condicional, no sentido de $E(u_t^2)$ não depender de t.

Para responder às questões anteriores, comecemos por calcular a matriz de quartos momentos, S, neste caso de homocedasticidade condicional:

$$\begin{aligned} S &= E(g_t g_t') = E(x_t x_t' u_t^2) \\ &= E[\,E(x_t x_t' u_t^2|\, x_t)] \\ &= E[x_t x_t'\, E(u_t^2|\, x_t)] \\ &= E(x_t x_t'\, \sigma^2) \\ &= \sigma^2 E(x_t x_t') \\ &= \sigma^2 \Sigma_{xx}. \end{aligned}$$

Repare-se então que:

[14] Veja-se, por exemplo, Hamilton, J. D. (2008), Macroeconomics and ARCH, *NBER working paper 14151.*

a) como, de acordo com a hipótese H5', esta matriz S é regular, também a matriz Σ_{xx} é regular. Ou seja, a hipótese H4' (condição de característica) é satisfeita.

b) Uma vez que a matriz S é mais simples, também o seu estimador deverá mudar. Em particular, o estimador natural ou óbvio passa a ser

$$\widehat{S} = s^2 S_{xx},$$

onde s^2 representa o usual estimador OLS de σ^2.

c) Ainda pela mesma razão, repare-se que a hipótese H6', de quartos momentos finitos, deixa de ser necessária. Por outras palavras, pela estacionaridade ergódica, tem-se, como já vimos, que $S_{xx} \xrightarrow{p} \Sigma_{xx}$. Por outro lado, também já sabemos que s^2 é consistente para σ^2. Então, o estimador da alínea anterior é consistente para S, sem qualquer necessidade da hipótese sobre os quartos momentos:

$$\widehat{S} = s^2 S_{xx} \xrightarrow{p} \sigma^2 \Sigma_{xx} = S.$$

d) A mudança de S também implica, naturalmente, que $\mathrm{Avar}(\widehat{\beta})$ muda:

$$\mathrm{Avar}(\widehat{\beta}) = \sigma^2 \Sigma_{xx}^{-1}.$$

e) E, naturalmente, daqui também decorre que o novo estimador consistente dessa matriz passa a ser dado por

$$\widehat{\mathrm{Avar}}(\widehat{\beta}) = s^2 S_{xx}^{-1} = T\, s^2 (X'X)^{-1}.$$

f) E, por conseguinte, os *se*'s dos estimadores dos coeficientes também voltam a ser dados por

$$\sqrt{s^2[(X'X)^{-1}]_{jj}},\ j = 1, 2, \ldots K,$$

que são os *se*'s usuais do modelo clássico (para pequenas amostras; reveja-se a subsecção 2.4).

g) Logo, também as estatísticas-t regressam à forma que tinham no modelo clássico.

Para completar o parelelismo com o modelo clássico, falta apenas verificar o que acontece com as estatísticas W. Note-se então que:

$$\begin{aligned} W &= T(R\widehat{\beta} - r)'\{R[Ts^2(X'X)^{-1}]\,R'\}^{-1}(R\widehat{\beta} - r), \\ &= (R\widehat{\beta} - r)'\{R[s^2(X'X)^{-1}]\,R'\}^{-1}(R\widehat{\beta} - r), \\ &= (R\widehat{\beta} - r)'\{R[(X'X)^{-1}]\,R'\}^{-1}(R\widehat{\beta} - r)/s^2, \\ &= J \times F, \\ &= (SSR_R - SSR_{UR})/s^2, \end{aligned}$$

isto é, as estatísticas de Wald podem ser calculadas facilmente com $J.F$. Em resumo, temos então a seguinte proposição.

Proposição 3.5 – Propriedades assimptóticas de $\widehat{\beta}$, t e F sob homocedasticidade condicional.

Suponha-se que as hipóteses H1' a H5' e H7 são satisfeitas. Então:

a) (Distribuição assimptótica de $\widehat{\beta}$) O estimador OLS de β, $\widehat{\beta}$, é consistente e assimptoticamente normal, com $\text{Avar}(\widehat{\beta}) = \sigma^2\Sigma_{xx}^{-1}$.

b) (Estimação consistente da matriz de covariâncias assimptótica) Essa matriz é estimada de forma consistente com $\widehat{\text{Avar}(\widehat{\beta})} = s^2 S_{xx}^{-1} = Ts^2(X'X)^{-1}$.

c) (Distribuição assimptótica das estatísticas t e F) Sob $H_0 : \beta_j = \beta_j^0$, a estatística-$t$ usual do modelo clássico tem distribuição assimptótica $N(0, 1)$.

   E sob $H_0 : R\beta = r$, a estatística $W = J.F$ tem distribuição assimptótica $\chi^2_{(J)}$, onde F é a estatística usual do modelo clássico.

Desta forma, podemos concluir que o enfraquecimento da hipótese de exogeneidade estrita e a remoção da hipótese de normalidade dos erros têm como consequência:

a) a necessidade de passar a exigir a estacionaridade e a ergodicidade aos dados, bem como aos erros dos modelos;

b) o desaparecimento das boas propriedades exactas do estimador OLS;

c) o desaparecimento da validade exacta da inferência associada ao OLS.

Relativamente a c), refira-se, contudo, que muitos investigadores preferem continuar a usar as distribuições t e F, como se a amostra fosse muito grande, mesmo em aplicações com um número reduzido de observações. Relativamente à utilização da t, recorde-se que

$$t_{(m)} \xrightarrow{d} N(0,1), \text{ quando } m \to \infty.$$

Por outro lado, relativamente à F, recorde-se também que

$$F_{(J,m)} \xrightarrow{d} \chi^2_{(J)}/J, \text{ quando } m \to \infty,$$

isto é, em grandes amostras, a comparação da estatística $W = J.F$ com o valor da tabela da $\chi^2_{(J)}$ é equivalente à comparação da estatística F com o valor da tabela da $F_{(J,T-K)}$. Ora, em pequenas amostras, as distribuições t e F podem ser melhores aproximações às distribuições das estatísticas que as distribuições assimptóticas. Em particular, visa-se amortecer ou suavizar as distorções de dimensão que são frequentes quando se empregam estas últimas.

3.9 Exercícios

3.1 (Exercício de revisão) Suponha que $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2)$. Notando que

$$\hat{\theta} - \theta = \frac{1}{\sqrt{T}}\sqrt{T}(\hat{\theta} - \theta),$$

Mostre que $\hat{\theta} \xrightarrow{p} \theta$.

3.2 Considerando que o primeiro elemento de um processo de passeio aleatório é fixo, mostre que o processo não é estacionário.

3.3 Considere que o processo $\{x_t\}$ é uma martingala. Mostre que o processo obtido por diferenciação, $\{\Delta x_t\}$, é uma m.d.s. . Note que os conjuntos informativos de ambos os processsos têm a mesma informação.

3.4 Considere que o processo $\{g_t\}$ é dado por $g_t = \epsilon_t \epsilon_{t-1}$, com $\epsilon_t \sim iid(0, \sigma^2)$. Verifique que $\{g_t\}, (t = 2, 3, \ldots)$ é uma m.d.s. .

3.5 Comece por assumir que $E(y_t|x_t) = x_t'\beta$, isto é, que a regressão de y_t sobre x_t é uma função linear de x_t. Definindo $u_t \equiv y_t - x_t'\beta$, mostre que x_t é ortogonal a u_t.

3.6 Mostre que, se o modelo desta secção tem termo independente, então a variância dos erros é finita. Sugestão: analise o primeiro elemento de $E(g_t g_t')$.

3.7 Suponha que, no modelo de regressão com termo independente, se pretende testar a significância global da regressão, isto é, que $H_0 : \beta_2 = 0, \ldots \beta_K = 0$. Como sabe, para o modelo clássico de regressão, a estatística-F para este caso pode escrever-se

$$F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)}.$$

Para o modelo desta secção com homocedasticidade condicional, mostre que

$$T.R^2 \xrightarrow{d} \chi^2_{(K-1)}.$$

4 Princípios de Previsão

Esqueçam-se agora, temporariamente, os modelos das secções anteriores e passe-se a considerar que estamos "apenas" interessados em fazer previsão. Dito de outra forma, suponha-se que abandonamos a abordagem econométrica estrutural, de análise e quantificação de relações de causalidade entre variáveis económicas, para centrar a atenção na previsão do seu comportamento futuro. As questões a responder são as seguintes:

a) o que é que o estimador OLS aplicado a uma equação linear estima?

b) Em que condições é que o estimador OLS é consistente?

4.1 Previsões baseadas no valor esperado condicional

Suponha-se que nos encontramos no período t e que o objectivo consiste na previsão do valor de y_{t+1}. Suponha-se ainda que, para esse efeito, dispomos da informação sobre um conjunto de variáveis, observadas até essa data, e represente-se o vector com essa informação na data t com x_t (como anteriormente). Por exemplo, podemos ter $x_t' = [1\, y_t\, y_{t-1} \ldots y_{t-p}]$, isto é, um modelo autoregressivo de ordem p. Represente-se com $y^*_{t+1|t}$ uma previsão de y_{t+1} baseada em x_t.

Como usar da melhor forma a informação disponível? Para responder a esta questão é necessário introduzir uma função perca, que dependa dos erros de previsão. Como é usual, suponha-se que a função perca é quadrática, isto é, que escolhemos $y^*_{t+1|t}$ de forma a minimizar

$$MSE(y^*_{t+1|t}) = E(y_{t+1} - y^*_{t+1|t})^2,$$

isto é, o erro quadrático médio de previsão. Nestas condições, tem-se a seguinte (bem conhecida) proposição.

Proposição 4.1 – Optimalidade do valor esperado condicional.

A previsão que minimiza o erro quadrático médio (MSE) é o valor esperado de y_{t+1} condicional em x_t,

$$y^*_{t+1|t} = E(y_{t+1}|\, x_t).$$

Ou seja, o previsor óptimo em MSE é o valor esperado condicional na informação disponível. A demonstração desta proposição pode ver-se, por exemplo, em Hamilton (1994), p. 73, ou em Hayashi (2000), pp. 138-9.

Esta proposição constitui um importante suporte para a adopção de modelos de regressão — entendidos como modelos de valor esperado condicional —, mas é óbvio que os modelos de regressão linear das secções anteriores assumem, adicionalmente, que o valor esperado condicional da variável a prever é uma função linear das variáveis do conjunto informativo. Neste caso, supondo que o valor a prever é y_t e não y_{t+1}, quando se escreve

$$E(y_t|\, x_t) = x_t'\beta,$$

tem-se $u_t \equiv y_t - E(y_t|\, x_t)$ e, por construção, $E(u_t|x_t) = 0$, donde resulta que também a condição de ortogonalidade, $E(u_t x_t) = 0$, é satisfeita por construção.

4.2 Previsões baseadas na projecção linear

Suponha-se, agora, que se restringe a classe das previsões àquelas que são funções lineares de x_t, isto é, que são do tipo

$$y^*_{t+1|t} = x_t'\beta.$$

Projecção linear

Projecção linear de y_{t+1} sobre x_t é a previsão linear, $x_t'\beta$ (ou $\beta' x_t$), tal que o erro de previsão é ortogonal a x_t, isto é:

$$E[(y_{t+1} - \beta' x_t)x_t'] = 0', \quad \text{ou} \quad E(u_{t+1}x_t') = 0',$$

onde $u_{t+1} \equiv y_{t+1} - \beta' x_t$ representa o erro da projecção linear. O vector β é o vector de coeficientes da projecção linear.

Voltando a considerar, momentaneamente, que o valor a prever é y_t e não y_{t+1}, na medida em que no modelo de regressão linear se tem $E(u_t x_t) = 0$, ele é um caso particular da projecção linear. Todavia, o inverso não é verdadeiro pois o erro de projecção não satisfaz necessariamente a condição $E(u_t|\, x_t)$. Ou seja, a projecção linear é bem menos ambiciosa que a regressão linear: trata-se de uma simples fórmula de previsão, sem a pretensão de representar o valor esperado condicional.

Também relativamente às projecções lineares se tem uma importante proposição (veja-se Hamilton, 1994, p. 74, ou Hayashi, 2000, p. 139, para a demonstração).

Proposição 4.2 – Optimalidade da projecção linear.
A projecção linear minimiza o erro quadrático médio (MSE) na classe das regras de previsão lineares.

Ou seja, esta proposição fornece outra justificação para chegar à usual equação de regressão linear, ainda em termos de previsão. Repita-se, no entanto, que agora o objectivo é menos ambicioso e que a condição de ortogonalidade é assumida ou imposta desde o início.

Usando a notação $P(y_{t+1}|\, x_t)$ para representar a projecção linear, uma vez que o valor esperado condicional oferece a melhor previsão possível em MSE, tem-se

$$MSE[P(y_{t+1}|\, x_t)] \geq MSE[E(y_{t+1}|\, x_t)].$$

4.3 Projecções lineares e estimação OLS

Da definição de projecção linear resulta imediatamente que

$$E(y_{t+1}x_t') = \beta'\, E(x_t x_t'),$$

ou seja, $\beta' = E(y_{t+1}x_t')[E(x_t x_t')]^{-1}$, ou ainda,

$$\beta = [E(x_t x_t')]^{-1} E(x_t y_{t+1}), \tag{8}$$

que é definido de forma única quando $E(x_t x_t')$ é regular (reveja-se a hipótese H4').

Ora, para a projecção linear $y_{t+1}^* = x_t'\beta$, o estimador OLS é

$$\widehat{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1} \sum_{t=1}^{T} x_t y_{t+1} = \left(\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right)^{-1} \left(\frac{1}{T}\sum_{t=1}^{T} x_t y_{t+1}\right).$$

Compare-se esta última expressão com a da equação (8): enquanto que o vector de coeficientes da projecção, β, se obtém dos momentos da população, $\widehat{\beta}$ é construído com os correspondentes momentos amostrais. Ou seja, enquanto que β se refere às características da população, isto é, do processo estocástico $\{y_{t+1}, x_t\}$, $\widehat{\beta}$ sintetiza a informação correspondente contida na amostra. Por outras palavras, trata-se de outra aplicação do princípio da analogia.

Proposição 4.3 – Consistência do estimador OLS.
Se o processo estocástico $\{y_{t+1}, x_t\}$ é estacionário e ergódico, o estimador OLS é um estimador consistente para os coeficientes da projecção linear, isto é, $\widehat{\beta} \xrightarrow{p} \beta$.

Naturalmente, a demonstração resulta imediatamente da aplicação do teorema ergódico e, comparando com o modelo da secção anterior, repare-se que as únicas hipóteses requeridas são H2' (estacionaridade ergódica) e H4' (condição de característica, isto é, regularidade de $E(x_t x_t')$).

Esta proposição justifica a utilização do estimador OLS quando o objectivo é a previsão linear, isto é, quando a abordagem econométrica não pretende ter carácter estrutural e se restringe a previsão a funções lineares. Note-se, contudo, que:

a) melhores previsões poderão ser obtidas com base numa análise estrutural cuidadosa;

b) se os momentos das variáveis mudarem ao longo do tempo (não estacionaridade) e/ou se a sua dependência temporal não se atenuar rapidamente com a passagem do tempo (não ergodicidade), a consistência da estimação OLS não é garantida.

4.4 Exercício

4.1 Prove que, para a previsão que minimiza o MSE, $E(y_{t+1}|\, x_t)$, o erro de previsão, u_{t+1}, é ortogonal a qualquer função $\phi(x_t)$ de x_t, isto é, que $E[u_{t+1}\phi(x_t)] = 0$. Sugestão: empregue a propriedade dos valores esperados totais.

5 Um Exemplo Ilustrativo

Para ilustrar alguns dos aspectos que acabaram de ser apresentados, vai apresentar-se um pequeno exemplo de aplicação. O ponto de partida deste exemplo é o modelo de regressão

$$LC_t = \beta_1 + \beta_2\, LR_t + \beta_3\, LS_t + \beta_4\, INF_t + v_t$$

onde LC representa o logaritmo do consumo privado, LR o logaritmo do rendimento disponível e LS o logaritmo de um índice de salários reais, todas a preços constantes, e INF representa a taxa de inflação.

Este modelo não tem a pretensão de ser um modelo estrutural pois não tem alicerces firmes na teoria económica. Trata-se apenas de uma função consumo tradicional aumentada com as variáveis LS e INF, a primeira pretendendo traduzir o facto de o consumo das famílias depender fortemente da evolução dos salários reais e a segunda tentando acomodar alguma provas estatísticas de que a inflação também pode ser um determinante importante desse consumo [15]. Note-se, no entanto, que um modelo bastante mais sofisticado não incluiria directamente LS na equação de regressão, antes considerando que a propensão marginal do consumo varia com LS, isto é, um modelo com

$$\beta_{2t} = \alpha_1 + \alpha_2\, LS_t + w_t,$$

chamado de modelo com parâmetros variáveis no tempo (TVP, *time varying parameters*). Naturalmente, a hipótese de estabilidade dos parâmetros implícita em H1 e em H1' seria violada, mas existem técnicas relativamente comuns para acomodar este problema. Em resumo, a equação acima deve ser vista mais como uma equação de projecção linear do que como uma equação estrutural sólida.

Ainda assim, como veremos mais adiante, as séries envolvidas não deverão ser estacionárias nem ergódicas, pelo que, como ilustração da matéria anterior, se considerou como mais adequado adoptar um modelo nas primeiras diferenças das variáveis, incluindo ainda termo independente:

$$\Delta LC_t = \beta_1 + \beta_2\, \Delta LR_t + \beta_3\, \Delta LS_t + \beta_4\, \Delta INF_t + u_t,$$

convidando-se o leitor a reflectir sobre a satisfação das hipóteses do modelo com regressores pré-determinados (e, em particular, sobre H2').

[15] Espera-se que a influência da inflação sobre o consumo seja positiva, mas existe algum debate em torno desta questão.

Estimado com dados anuais para a economia portuguesa, apenas para o período de 1966 a 1995, o modelo produziu os seguintes resultados:

$$\begin{array}{lclllll} \widehat{\Delta LC_t} & = & 0.018 & +0.331\,\Delta LR_t & +0.376\,\Delta LS_t & +0.0029\,\Delta INF_t, \\ & & (0.005) & (0.100) & (0.086) & (0.0009) \\ & & [0.004] & [0.108] & [0.092] & [0.0007] \end{array}$$

onde os valores dentro de parênteses rectos representam os se's robustos à heterocedasticidade, usando a matriz de White [16]. Tem-se, ainda, $T = 30$, $s = 0.018$, $R^2 = 0.653$ e $\bar{R}^2 = 0.612$.

Por outro lado, têm-se ainda os seguintes valores para algumas estatísticas de teste:

$$WHITE = 8.43(0.491), \quad BP = 4.41(0.110), \quad LRHET = 2.47(0.116),$$

onde $WHITE$ representa a usual estatística do teste de White para heterocedasticidade, BP representa a estatística do teste de Breusch-Pagan também para heterocedasticidade, supondo que esta está relacionada com ΔLR e ΔLS, e $LRHET$ é uma estatística de teste de rácio de verosimilhanças ainda para heterocedasticidade, considerando a possibilidade de σ^2 ter uma alteração abrupta a partir do meio da amostra. Os valores dentro de parênteses representam os valores-p associados a cada uma das estatísticas. Não se encontram, portanto, indícios significativos da presença de heterocedasticidade, o que confirma a suspeita (de não violação da homocedasticidade) baseada na comparação dos se's acima.

Como, por outro lado, a estatística do teste $RESET$, obtida com os quadrados dos valores ajustados, tem o valor de 1.24, com um valor-p associado de 0.276, também não se encontram indícios estatísticos da presença de incorrecção da forma funcional.

Embora não se tenham encontrado provas estatísticas da presença de heterocedasticidade, para ilustrar a utilização da estatística de Wald robusta, considere-se que estamos interessados em testar

$$H_0 : \beta_2 = 2\beta_3 \quad vs \quad H_1 : \beta_2 \neq 2\beta_3,$$

empregando essa estatística. Para este efeito, as instruções de TSP podem ser as seguintes:

[16] Recorde-se que, para obter estes se's no TSP, a instrução OLSQ deve ser seguida da opção "(HCTYPE=0, ROBUSTSE)".

```
FORM EQ DLC C DLR DLS DINF;
GMM (HET,NMA=0,INST=(C,DLR,DLS,DINF)) EQ;
FRML REST BEQ1-2*BEQ2;
ANALYZ REST;
```

Note-se que:

a) o prefixo "D" nos nomes das variáveis representa o operador de diferenciação, Δ. Isto é, previamente deram-se instruções de diferenciação e usou-se esse prefixo na designação das novas séries (por exemplo, "DLC = LC-LC(-1);").

b) Para obter a estatística de teste é necessário simular que o modelo está a ser estimado pelo GMM. Todavia, como a lista de instrumentos é idêntica à lista de regressores, o método de estimação dos coeficientes é, de facto, o OLS; só a matriz de covariâncias do estimador OLS está a ser estimada de forma diferente da usual.

c) Em particular, com as opções "HET,NMA=0" obtém-se a chamada matriz de White, tal como foi apresentada na subsecção 3.7. O facto de se ter imposto "NMA=0" será compreendido mais adiante, quando se abordar o problema da inferência robusta à autocorrelação.

d) A restrição a testar é designada aqui com "REST" e note-se que o TSP atribui aos parâmetros da equação o prefixo "B" e que emprega os índices 0, 1, 2, ... (e não 1, 2, ...) para indexar os coeficientes.

e) A instrução "ANALYZ" permite obter o cálculo do valor da estatística de Wald (neste caso, robusta à heterocedasticidade).

Como resultado deste procedimento, além dos resultados de estimação idênticos aos do OLS, o TSP produz a estatística de Wald robusta, à qual chama, neste caso "CHISQ(1)", com valor de 2.84, e valor-p associado de 0.092. Desta forma, os dados fornecem algum (mas não muito) suporte à hipótese nula. Por outras palavras, H_0 não é rejeitada por um teste com dimensão $\alpha = 0.05$, mas é-o por um teste a 10%. Note-se ainda que, neste caso, o mesmo resultado poderia ter sido obtido com a utilização de uma estatística-t, desde que as estimativas das variâncias e das covariâncias dos estimadores OLS sejam retiradas da matriz de White (também calculada pelo TSP).

No anexo seguinte encontra-se ainda o teste da hipótese conjunta $H_0 : \beta_2 = 2\beta_3,\ \beta_4 = 0$. Admite-se que o leitor não terá dificuldades em compreender as instruções de TSP bem como o respectivo resultado.

Anexo

```
        PROGRAM
**************************************************************
1  OPTIONS CRT;
2  FREQ A;
3  SMPL 65 95;
4  READ (FILE='C:\tsp\consgive.WK1',FORM=LOTUS) lrd lcp inf lsr;
5
5  title 'EXEMPLO DA FUNÇÃO CONSUMO';
6  dlc=lcp-lcp(-1);
7  dlr=lrd-lrd(-1);
8  dinf=inf-inf(-1);
9  dls=lsr-lsr(-1);
10
10  regopt (stars,bplist=(c,dlr,dls)) all;
11  olsq dlc c  dlr dls dinf;
12  olsq (hctype=0, robustse) dlc c  dlr dls dinf;
13
13  FORM EQ DLC c DLR DLS DINF;
14  GMM (HET,NMA=0,INST=(C,DLR,DLS,DINF)) EQ;
15  FRML REST BEQ1-2*BEQ2;
16  ANALYZ REST;
17
17  FRML REST2 BEQ3;
18  ANALYZ REST REST2;
19
************************************************************************
```

EXEMPLO DA FUNÇÃO CONSUMO
========================

Equation 1
============

Method of estimation = Ordinary Least Squares

Dependent variable: DLC
Current sample: 1966 to 1995

```
        Mean of dep. var. = .033158
   Std. dev. of dep. var. = .029000
 Sum of squared residuals = .847058E-02
    Variance of residuals = .325792E-03
 Std. error of regression = .018050
                R-squared = .652678
            Durbin-Watson = 1.73047 [.079,.454]
                Chow test = .860326 [.503]
   LR het. test (w/ Chow) = 2.47449 [.116]
          White het. test = 8.43332 [.491]
  Breusch-Pagan het. test = 4.40698 [.110]
          Ramsey's RESET2 = 1.24085 [.276]
         F (zero slopes) = 16.2862 ** [.000]
           Schwarz B.I.C. = -73.2147
 Akaike Information Crit. = -76.0171
           Log likelihood = 80.0171
```

Variable	Estimated Coefficient	Standard Error	t-statistic		P-value
C	.018740	.460239E-02	4.07186	**	[.000]
DLR	.330561	.100103	3.30221	**	[.003]
DLS	.375609	.085867	4.37431	**	[.000]
DINF	.294196E-02	.853161E-03	3.44830	**	[.002]

Equation 2
============

Method of estimation = Ordinary Least Squares

Dependent variable: DLC
Current sample: 1966 to 1995

Variable	Estimated Coefficient	Standard Error	t-statistic		P-value
C	.018740	.369122E-02	5.07698	**	[.000]
DLR	.330561	.108065	3.05890	**	[.005]
DLS	.375609	.092207	4.07354	**	[.000]
DINF	.294196E-02	.664564E-03	4.42690	**	[.000]

Standard Errors are heteroskedastic-consistent (HCTYPE=0).
(No df correction)

GENERALIZED METHOD OF MOMENTS
=============================

WITH STARTING VALUES VIA:
NONLINEAR TWO STAGE LEAST SQUARES

EQUATIONS: EQ

INSTRUMENTS: C DLR DLS DINF

Parameter	Estimate	Standard Error	t-statistic		P-value
BEQ0	.018740	.369122E-02	5.07698	**	[.000]
BEQ1	.330561	.108065	3.05890	**	[.002]
BEQ2	.375609	.092207	4.07354	**	[.000]
BEQ3	.294196E-02	.664564E-03	4.42690	**	[.000]

Standard Errors computed from heteroscedastic-consistent matrix
(Robust-White)

```
Equation: EQ
Dependent variable: DLC

       Mean of dep. var. = .033158
  Std. dev. of dep. var. = .029000
Sum of squared residuals = .847058E-02
   Variance of residuals = .325792E-03
Std. error of regression = .018050
               R-squared = .652678
      Adjusted R-squared = .612603
           Durbin-Watson = 1.73047 [.079,.454]

                         Results of Parameter Analysis
                         =============================

                         Standard
Parameter  Estimate        Error       t-statistic       P-value
REST       -.420657      .249714       -1.68455          [.092]

Wald Test for the Hypothesis that the given set of Parameters are jointly zero:

CHISQ(1) =   2.8377104     ; P-value = 0.09208

                         Results of Parameter Analysis
                         =============================

                         Standard
Parameter  Estimate        Error       t-statistic       P-value
REST       -.420657      .249714       -1.68455          [.092]
REST2      .294196E-02   .664564E-03   4.42690       **  [.000]


Wald Test for the Hypothesis that the given set of Parameters are jointly zero:

CHISQ(2) =   20.027571     ; P-value = 0.00004

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

Bibliografia

Para além de algumas referências que foram indicadas localmente, empregou-se, sobretudo, a bibliografia que se segue.

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

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

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

Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton.

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

Ribeiro, C. S. (2007), *Apontamentos de Econometria*, *mimeo*, ISEG–UTL, Lisboa.

Stock, J. H. e Watson, M. W. (2007), *Introduction to Econometrics*, 2nd ed., Pearson Addison-Wesley, Boston.

White, H. (1999), *Asymptotic Theory for Econometricians*, revised edition, Academic Press, San Diego.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge Ma. .

Wooldridge, J. M. (2006), *Introductory Econometrics, a Modern Approach*, 3rd ed., Thomson South-Western, Mason.