Exercícios de Econometria

ISEG — Universidade de Lisboa

Artur Silva Lopes[1]

Capítulo 1 — Variáveis Explicativas Binárias ("Dummy")

Exercícios prioritários: 1, 5, 7, 9, 10, 12 e 14.

1. Exercício 7.1 do livro de Wooldridge, 6ª edição (**W**).

2. Exercício 7.2 de W, apenas as alíneas i) e ii).

3. Exercício 7C.1 de W, apenas as alíneas i) e ii).

4. (Do exame de EE de 7/9/2016.) Quando uma característica qualitativa dos elementos de uma população dá origem a dois grupos distintos — por exemplo, homens e mulheres —, é preferível introduzir apenas uma *dummy* e manter o termo independente do que introduzir as duas *dummies* e retirar esse termo porque ...

   ☐ ... o OLS não permite estimar modelos sem termo independente.

   ☐ ... a inferência estatística de interesse é mais simples de efectuar.

   ☐ ... a omissão do termo independente origina o aparecimento de heteroscedasticidade.

   ☐ ... a inclusão das duas *dummies* provocaria uma situação de autocorrelação perfeita ou exacta dos erros do modelo.

5. Exercício 7C.5 de W.

6. Exercício 7.5 de W.

7. (Exercício 2 do exame de ER de 25/6/2009.) Considerando os modelos

   (**A**) $log(wage) = \gamma_0\, female + \delta_0\, male + \beta_1\, educ + u,$

   (**B**) $log(wage) = \beta_0 + \alpha_0\, female + \beta_1\, educ + u,$

   (**C**) $log(wage) = \beta_0 + \theta_0\, female + \phi_0\, male + \beta_1\, educ + u,$

   Indique a afirmação que é **FALSA**:

   ☐ o modelo (B) é preferível ao modelo (A);

[1] Elaborado com base em Wooldridge, J. M. (2009, 2016), *Introductory Econometrics*, 4th and 6th eds., South-Western, Cengage Learning, e complementado com exercícios elaborados para os exames.

- ☐ o modelo (C) incorre na "armadilha das variáveis artificais" ;
- ☐ no modelo (A), as colunas da matriz X são linearmente independentes;
- ☐ no modelo (C), as colunas da matriz X são linearmente independentes;

8. Exercício 7C.2 de W, apenas as alíneas i) e iii).

9. (Do exame de EN de 4/1/2017) Suponha que se estimou um modelo explicativo dos salários dos trabalhadores de um sector de actividade (sal) usando apenas o número de anos de experiência ($exper$) e a *dummy* $licen$, com o valor 1 se o trabalhador é licenciado (e 0 no caso contrário), bem como a sua interacção ($exper \times licen$) como regressores, e que se obteve

$$\widehat{\mathrm{E}}(sal\,|exper, licen = 0) - \widehat{\mathrm{E}}(sal\,|exper, licen = 1) = -2.16 - 0.05\,exper.$$

Sabendo que o modelo estimado apenas com os dados dos trabalhadores não licenciados é $\widehat{sal} = 7.56 + 0.08 exper$, apresente, de forma justificada, o modelo estimado empregando apenas os dados dos trabalhadores licenciados. Com base no modelo especificado para todos os trabalhadores, indique ainda as hipóteses (H_0 e H_1) do teste estatístico que permite analisar se há diferenças no salário médio dos trabalhadores licenciados e não licenciados.

10. (Exercício 1 do exame de ER de 28/6/2011.) Suponha que se pretende explicar os salários dos indivíduos (sal) com base no número de anos da sua educação ($educ$) e também com base no género e na raça, distinguindo entre brancos, negros e asiáticos. Especifique um modelo que permita, simultaneamente:

    - que a componente salarial autónoma varie de acordo com o género;
    - que o rendimento da educação seja expresso em termos percentuais e que varie com a raça;
    - testar facilmente a igualdade do rendimento da educação entre asiáticos e brancos.

    Nota: explicite claramente a(s) variável(is) que necessita empregar.

11. Exercício 7.8 de W.

12. Exercício 7C.6 de W.

13. (Exercício 1 do exame de EN de 12/1/2011.) Pretende-se analisar eventuais diferenças regionais no comportamento das despesas em educação ($desed$) das famílias residentes no norte e no sul do país. O ficheiro de dados contém ainda as seguintes variáveis: $rend$ é o rendimento médio da família, $filhos$ é o número de filhos e sul uma variável dummy, que assume o valor 1 se a família reside no sul.

Utilizando o EViews, obteve-se:

$$\log(desed) = 1.26 + 0.13\log(rend) + 0.03 filhos + \hat{u}, \quad n = 310, \quad SSR = 189.4.$$

Com a instrução de EViews **if sul = 1**, obteve-se:

$$\log(desed) = 1.82 + 0.013\log(rend) + 0.08 filhos + \hat{u}, \quad n = 61, \quad SSR = 34.02.$$

Finalmente, com a instrução **if sul = 0**, obteve-se:

$$\log(desed) = 1.20 + 0.16\log(rend) + 0.01 filhos + \hat{u}, \quad n = 249, \quad SSR = 153.7.$$

a) Teste, ao nível de 5%, a hipótese de regressões idênticas para as famílias do norte e do sul.

b) Utilizando a mesma amostra, estimou-se ainda o seguinte modelo com $n = 310$:

$$\log(desed) = \beta_0 + \delta_0\, sul + \beta_1 \log(rend) + \delta_1\, \log(rend)\times sul + \beta_2 filhos + \delta_2\, filhos \times sul + u,$$

Então, as estimativas dos coeficientes δ_0, δ_1 e δ_2 são iguais a:

- ☐ $\hat{\delta}_0 = 1.26$, $\hat{\delta}_1 = 0.13$, $\hat{\delta}_2 = 0.03$;
- ☐ $\hat{\delta}_0 = 0.62$, $\hat{\delta}_1 = 0.173$, $\hat{\delta}_2 = 0.09$;
- ☐ $\hat{\delta}_0 = 0.62$, $\hat{\delta}_1 = -0.147$, $\hat{\delta}_2 = 0.07$;
- ☐ a informação é insuficiente para obter as estimativas pretendidas.

14. (Adaptação do exercício 1 do exame de EE de Setembro de 2017.) Suponha que, para investigar a influência do tempo gasto pelos jovens com idade inferior a 20 anos nas redes sociais sobre o seu rendimento escolar, se estimaram as regressões abaixo, com os dados de 250 jovens estudantes, onde as variáveis representam:
— *clamed*, a classificação média obtida no final do ano;
— $femin$, uma variável *dummy* se o estudante é do sexo feminino e,
— *horest*, número médio semanal de horas de estudo.

Para os $n = 250$ jovens obteve-se:

$$\widehat{clamed} = 3.24 + 1.10\, femin + 0.250\, horest, \quad R^2 = 0.850, \quad SSR = 1167.39.$$

Para os $n_1 = 127$ jovens que gastam, em média, menos de 3h por semana nas redes sociais:

$$\widehat{clamed} = 4.12 + 1.07\, femin + 0.249\, horest, \quad R^2 = 0.871, \quad SSR = 557.73.$$

Para os $n_2 = 123$ jovens que gastam, em média, 3 horas ou mais por semana nas redes sociais:

$$\widehat{clamed} = 2.27 + 1.01\, femin + 0.255\, horest, \quad R^2 = 0.871, \quad SSR = 429.90.$$

a) Que conclusão pode retirar desta informação?

b) Explique como procederia para tratar o problema da alínea anterior empregando variáveis artificiais. (Nota: indique o modelo, as hipóteses e a estatística de teste a considerar, bem como o seu valor.)

15. (Exercício 2 do exame de EN de 15/6/2012.) Com base num ficheiro de EViews que contém apenas as observações das variáveis *WAGE, FEMALE, EDUC* e *AGE* (idade) e pretendendo-se estimar o modelo

$$\log(wage) = \beta_0 + \alpha_0\, young + \gamma_0\, female + \beta_1\, educ + u,$$

onde

$$young_i = \begin{cases} 1, & \text{se a pessoa i tem idade inferior a 30 anos,} \\ 0, & \text{nos restantes casos,} \end{cases}$$

escrevemos, na janela (ou quadro) de estimação do EViews:

- ☐ log(wage) c young=1 if age<30 female educ.
- ☐ lwage c young female educ
- ☐ log(wage) c age<30 female educ.
- ☐ nenhuma das anteriores

∞∞∞

Capítulo 2 — Modelos para Variáveis Dependentes Binárias

Exercícios prioritários: 1,7, 9 e 10.

1. Exercício 7C.8 de W (6th edition), apenas as alíneas i) a iv).
2. Exercício 7.7 de W.
3. Exercício 8C.7 de W, apenas a alínea i).
4. Exercício 17.1 de W.
5. Exercício 17.2 de W.
6. Exercício 17C.1 de W.
7. Exercício 17C.2 de W.

8. (Do exame de EN de 4/1/2017.) No modelo estimado abaixo, $APROV$ tem o valor 1 se o empréstimo pedido por um indivíduo foi aprovado (e 0 no caso contrário), $RENDI$ representa o seu rendimento médio mensal e $CASADO$ e $PRECA$ são variáveis *dummy*, a primeira com o valor 1 se o indivíduo é casado e a segunda com o valor 1 se a situação profissional do indivíduo é precária (e 0 no caso contrário).

Dependent Variable: APROV
Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 840
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.041386	0.134106	0.308606	0.7576
CASADO	0.215588	0.163488	1.318676	0.1873
RENDI	0.026816	0.003559	7.534558	0.0000
PRECA	-0.059685	0.012127	-4.921847	0.0000

McFadden R-squared	0.081150	Mean dependent var	0.675000
S.D. dependent var	0.468654	S.E. of regression	0.440739
Akaike info criterion	1.168342	Sum squared resid	162.3939
Schwarz criterion	1.190882	Log likelihood	-486.7037
Hannan-Quinn criter.	1.176981	Restr. log likelihood	-529.6881
LR statistic	85.96881	Avg. log likelihood	-0.579409
Prob(LR statistic)	0.000000		

a) Das seguintes afirmações sobre este modelo, indique a que é **FALSA**:

☐ É possível afirmar que o efeito de uma variação positiva do rendimento médio mensal dos indivíduos sobre a probabilidade de terem os seus empréstimos aprovados é positiva.

☐ Os erros-padrão podem ser validamente empregues pois foram retirados da matriz de White.

☐ Não é por acaso que a *Log likelihood* do modelo é negativa.

☐ Os testes-F usuais não são válidos neste modelo.

b) Relativamente à "Log likelihood" do modelo $\Phi(\beta_0 + \beta_1\, casado)$, pode afirmar-se que:

☐ deve estar muito próxima do valor -486.7.

☐ deve ser bastante maior que -486.7.

☐ deve ser bastante menor que -486.7.

☐ não existe informação suficiente para escolher qualquer uma das 3 afirmações anteriores.

c) Indique as instruções de EViews que necessita empregar para estimar o efeito parcial médio da variável $RENDI$. Supondo que essa estimativa é igual a 0.0089, interprete-a.

9. (Exercício 3 do exame de ER de 25/6/2010.) Para analisar os determinantes do (in)cumprimento dos pagamentos associados aos seus cartões de crédito, uma instituição bancária estimou o modelo Probit apresentado abaixo, onde as variáveis têm o seguinte significado:

Dependent Variable: CUMP
Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 960
Convergence achieved after 3 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.021125	0.127902	0.165166	0.8688
SALDO	0.025659	0.003346	7.668841	0.0000
NCART	-0.057103	0.011214	-5.092315	0.0000
CASADO	0.293265	0.153204	1.914210	0.0556

McFadden R-squared	0.075422	Mean dependent var	0.679167
S.D. dependent var	0.467040	S.E. of regression	0.441063
Akaike info criterion	1.168672	Sum squared resid	185.9771
Schwarz criterion	1.188951	Log likelihood	-556.9624
Hannan-Quinn criter.	1.176394	Restr. log likelihood	-602.3962
LR statistic	90.86758	Avg. log likelihood	-0.580169
Prob(LR statistic)	0.000000		
Obs with Dep=0	308	Total obs	960
Obs with Dep=1	652		

- cump – variável *dummy* com o valor 1 se o indivíduo cumpriu (dentro do prazo) todos os pagamentos nos últimos 5 anos;
- saldo – saldo médio da conta à ordem do indivíduo, em centenas de euros;
- ncart – número total de cartões de crédito que o indivíduo possui (incluindo de outras instituições financeiras);
- casado – variável *dummy* com o valor 1 se o indivíduo é casado.

a) O estimador que foi empregue ...

☐ ... minimiza $SSR = \sum_{i=1}^{n}(cump_i - \beta_0 - \beta_1 saldo - \beta_2 ncart - \beta_3 casado)^2$.

☐ ... tem como expressão $\hat{\beta} = (X'X)^{-1}X'y$, com X a matriz das observações das variáveis explicativas.

☐ ... minimiza $\sum_{i=1}^{n}[y_i - \Phi(x_i\beta)]^2$, com $y_i = cump_i$ e $x_i = (1,\, saldo_i,\, ncart_i,\, casado_i)$.

☐ ... nenhuma das restantes respostas é correcta.

b) Indique as instruções de EViews que necessita empregar para obter a estimativa do efeito parcial médio da variável casado.

c) Suponha que, como resultado dessas instruções, obteve o valor 0.099. Interprete esta estimativa.

d) No *output* apresentado...
☐ ... a "LR statistic" é obtida com $2 \times (\mathcal{L}_R - \mathcal{L}_{UR})$.
☐ ... a "Restr. log likelihood" é obtida estimando um modelo Probit apenas com a constante, isto é, sem variáveis explicativas.
☐ ... a "LR statistic" é uma estatística dada por $LR = \frac{(R^2_{ur} - R^2_r)/(k-1)}{(1-R^2_{ur})/(n-k-1)}$.
☐ ... $\hat{\beta}_3 = 0.293$ significa que se estima que, em média, *ceteris paribus*, quando um indivíduo se casa, a probabilidade de passar a cumprir os pagamentos do seu cartão de crédito na referida instituição aumenta em 0.293.

10. (Exercício 3 do exame de ER de 3/2/2016). Para seguir o percurso profissional dos seus recém-licenciados, uma instituição de ensino superior estimou os 2 modelos apresentados na página seguinte, cujas variáveis são as seguintes:

- empre – variável *dummy* com o valor 1 se, 3 meses após a obtenção da licenciatura, o licenciado está empregado;
- media – clasificação média da licenciatura.
- mest – variável *dummy* com o valor 1 se o licenciado está inscrito num curso de mestrado;
- mulher – variável *dummy* com o valor 1 se o licenciado é mulher;

a) Formalize e efectue o teste estatístico que lhe permite averiguar se o segundo modelo é uma simplificação aceitável do primeiro.

b) Após a estimação do primeiro modelo e com as instruções

```
scalar k1=@cnorm(c(1)+c(2)*14+c(3)+c(4))
scalar k2=@cnorm(c(1)+c(2)*14+c(4))
```

obtiveram-se os valores 0.343 e 0.285, respectivamente[2]. Apresente as expressões algébricas representadas nos cálculos de EViews e calcule e interprete a sua diferença.

c) Numa tentativa para efectuar um teste da significância estatística conjunta de *mest* e de *mulher* no primeiro modelo, estimou-se também o modelo

$$\widehat{\mathrm{P}}(empre = 1|\, media) = 0.171 + 0.015\, media, \quad R^2 = 0.08, \quad \mathcal{L} = -182.97, \quad SSR = 61.82.$$

Pode então afirmar-se que:
☐ essa hipótese não é rejeitada por um teste estatístico com 5% de dimensão.
☐ essa hipótese é rejeitada por um teste estatístico com 5% de dimensão.
☐ nada se pode concluir pois este modelo não é um caso particular do primeiro.
☐ todas as respostas anteriores são incorrectas.

[2] Valores corrigidos com a ajuda de Inês C. Couto, a quem se agradece a colaboração.

Dependent Variable: EMPRE
Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 264
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.759932	0.400804	-1.896020	0.0580
MEDIA	0.037643	0.026292	1.431738	0.1522
MEST	0.163032	0.159497	1.022160	0.3067
MULHER	-0.335403	0.159111	-2.107975	0.0350

McFadden R-squared	0.021455	Mean dependent var	0.382576
S.D. dependent var	0.486939	S.E. of regression	0.482650
Akaike info criterion	1.332377	Sum squared resid	60.56716
Schwarz criterion	1.386558	Log likelihood	-171.8738
Hannan-Quinn criter.	1.354149	Restr. log likelihood	-175.6421
LR statistic	7.536672	Avg. log likelihood	-0.651037
Prob(LR statistic)	0.056624		

Dependent Variable: EMPRE
Method: ML - Binary Probit (Quadratic hill climbing)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.139710	0.108252	-1.290608	0.1968
MULHER	-0.334067	0.157912	-2.115523	0.0344

S.D. dependent var	0.486939	S.E. of regression	0.483713
Akaike info criterion	1.328739	Sum squared resid	61.30233
Schwarz criterion	1.355829	Log likelihood	-173.3935
Hannan-Quinn criter.	1.339625	Restr. log likelihood	-175.6421
LR statistic	4.497179	Avg. log likelihood	-0.656794
Prob(LR statistic)	0.033951		

11. (Exercício 2 do exame de EE de 3/9/2012.) Num estudo sobre a mobilidade dos estudantes universitários estimaram-se os modelos apresentados abaixo, onde as variáveis têm o seguinte significado:

- erasmus – variável *dummy* com o valor 1 se o estudante foi aluno Erasmus;
- rend – rendimento médio da família;
- nota – classificação média obtida antes de “ir para Erasmus”;
- ingles – variável *dummy* com o valor 1 se o estudante frequentou um curso de inglês ou se domina a língua inglesa.

$$\widehat{\mathrm{P}}(erasmus = 1|\, rend, nota, ingles) = \Phi(-2.265 + 0.544\, rend + 0.118\, nota + \widehat{\beta}_3\, ingles),$$

$$n = 753, \quad SSR = 151.821, \quad \mathcal{L} = -445.696.$$

$$\widehat{\mathrm{P}}(erasmus = 1|\, rend) = \Phi(-0.443 + 0.605\, rend)$$

$$n = 753, \quad SSR = 160.727, \quad \mathcal{L} = -467.196.$$

a) Empregando o primeiro modelo e usando as instruções

```
series x=@cnorm(c(1)+c(2)*rend+c(3)*nota+c(4))
series y=@cnorm(c(1)+c(2)*rend+c(3)*nota)
scalar z=@mean(x-y)
```

obteve-se o valor $\mathbf{z = 0.193}$.

Então, relativamente à estimativa $\widehat{\beta}_3$ do coeficiente da variável "*ingles*" pode-se concluir que:

☐ $\widehat{\beta}_3 > 0$; ☐ $\widehat{\beta}_3 < 0$; ☐ $\widehat{\beta}_3 = 0.193$;

☐ $\widehat{\beta}_3$ pode assumir qualquer valor não nulo.

b) Interprete o valor obtido para z (0.193).

c) Formalize, efectue e retire a conclusão apropriada do teste estatístico que conduziu à estimação do segundo modelo.

12. (Do exame de EE de 7/9/2016.) Para analisar a viabilidade do lançamento de um serviço de *streaming* de video (filmes e séries), recolheu-se informação sobre 514 indivíduos escolhidos ao acaso, convencionando-se atribuir o valor 1 à variável SSTREAM se a resposta à questão "Prefere um serviço de *streaming* à aquisição de DVDs?" fosse positiva (e 0 no caso contrário). No modelo estimado apresentado na página seguinte figuram ainda as variáveis REND, rendimento per capita do agregado do indivíduo, em euros, IDADE, a idade do indivíduo e LITO, uma *dummy* com o valor 1 se o indivíduo reside no litoral (e 0 no caso contrário).

    a) (Formalize e) Efectue o teste estatístico que emprega a LR statistic.

    b) Escreva agora o modelo **Logit genérico** que lhe permite estimar a probabilidade de obter uma resposta positiva à mesma questão mas apenas em função do rendimento *per capita* e da idade dos indivíduos. Supondo que a estimação desse modelo produzia um valor da *log-likelihood* de -270.421, poderia empregá-lo para efectuar a escolha entre o modelo inicial e este novo modelo? Justifique.

Dependent Variable: SSTREAM
Method: ML - Binary Probit (Quadratic hill climbing)
Sample: 1 514
Convergence achieved after 4 iterations
Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.698611	0.245847	-2.841647	0.0045
REND	0.000434	0.000163	2.658271	0.0079
IDADE	-0.011616	0.004425	-2.625237	0.0087
LITO	0.174720	0.125548	1.391664	0.1640

McFadden R-squared	0.027012	Mean dependent var	0.235529
S.D. dependent var	0.424753	S.E. of regression	0.419758
Akaike info criterion	1.078220	Sum squared resid	87.56976
Schwarz criterion	1.111885	Log likelihood	-266.0941
Hannan-Quinn criter.	1.091429	Restr. log likelihood	-273.4815
LR statistic	14.77477	Avg. log likelihood	-0.531126
Prob(LR statistic)	0.002020		

∞∞∞

Capítulo 3 — Séries Temporais: Análise de Regressão Básica

Exercícios prioritários: 3, 5, 9, 10, 11 e 13.

1. Exercício 10.3 de W.

2. Exercício 10.C10 de W.

3. (Exercício 5 do exame de ER de 25/6/2009.) Considerando que o modelo $y_t = \alpha + \delta_0\, z_t + \delta_1\, z_{t-1} + \delta_2\, z_{t-2} + u_t$ satisfaz a hipótese $\mathrm{E}(u_t|z_t, z_{t-1}, z_{t-2}) = 0$, $\forall t$, suponha que, no período 20, z tem uma variação transitória de 2 unidades, como na tabela seguinte.

t	...	18	19	20	21	22	23
z	...	5	5	7	5	5	5
$E(y_t\vert\, z_t, z_{t-1}, z_{t-2})$ (*)	...	10	10	11	12	11	10

(*) Nota: também pode considerar que são os valores de y quando $u = 0$.

Determine os valores de δ_0, δ_1 e δ_2, bem como do multiplicador de longo prazo.

4. (Do exame de EN de 4/1/2017.) Indique a afirmação correcta no contexto do seguinte modelo estimado, onde *custot* representa o custo total de produção e *sal* um índice de salários:

$$\log(\widehat{custot}) = 0.052{+}0.25\,\log(sal_t){+}0.15\,\log(sal_{t-1}){+}0.06\,\log(sal_{t-2}){+}0.05\,\log(sal_{t-3}).$$

☐ Estima-se que, em média, quando os salários têm uma variação transitória de uma unidade, o custo total tem uma variação de 0.25% ainda no mesmo período.

☐ Estima-se que, em média, quando os salários têm uma variação permanente de 1%, o custo total tem uma variação total, no longo prazo, de 0.51%.

☐ Estima-se que, em média, quando os salários têm uma variação permanente de 1%, o custo total tem uma variação contemporânea de 0.25% em torno da sua tendência.

☐ Nenhuma das afirmações anteriores é correcta.

5. (Do exame de EN de 4/1/2017.) Das seguintes afirmações respeitantes ao modelo clássico de regressão de séries temporais, indique a que é **FALSA**. A hipótese de exogeneidade estrita dos regressores ...

   ☐ ... não permite que variações de y tenham algum efeito sobre os valores futuros do(s) regressor(es).

   ☐ ... não é necessária para que os testes estatísticos sejam válidos.

   ☐ ... é mais forte que a hipótese de exogeneidade dos modelos de dados seccionais.

   ☐ ... exige, por exemplo, que $\text{Cov}(u_3,\, x_{45,2}) = 0$.

6. (Exercício 6 do exame de EE de 7/9/2016.) Considere um modelo que relaciona y com z e tal que, silenciando o efeito dos erros, satisfaz as condições:
   i) $\partial y_t / \partial z_t = 0$; ii) o multiplicador de longo prazo é 1.2; iii) $\partial y_{t+j} / \partial z_t = 0, j > 2$; iv) $\frac{y_{t+1} - y_{t-1}}{\Delta z_t} = 0.7$. Então, esse modelo é:

   ☐ $y_t = \beta_0 + 0.5\, z_t + 0.4\, z_{t-1} + 0.3\, z_{t-2} + u_t$.

   ☐ $y_t = \beta_0 + 0.7\, z_{t-1} + 0.5\, z_{t-2} + u_t$.

   ☐ $y_t = \beta_0 + 0.7\, z_{t-1} + 0.3\, z_{t-2} + 0.2\, z_{t-3} + u_t$.

   ☐ $y_t = \beta_0 + 0.7\, z_{t-1} + 0.3\, z_{t-2} + u_t$.

7. Exercício 10.2 de W.

8. Exercício 10.C1 de W.

9. (Exercício 6 do exame de ER de 28/6/2011.) Admita que se sabe que $lgas$, o logaritmo das vendas da gasolina (em preços constantes), e $lpreg$, o logaritmo do seu preço médio, são ambas estacionárias em tendência, com uma tendência crescente ao longo do tempo[3]. Suponha que se estimaram os seguintes modelos:

$$\widehat{lgas_t} = 2.34 + 0.675\, lpreg_t, \; R^2 = 0.543 \;\text{ e} \tag{1}$$

[3] Mais adiante será estudado o conceito de processo ou série estacionária em tendência. Por agora, é suficiente saber que o comportamento de uma série desse tipo é claramente dominado por uma tendência determinística; ou seja, em termos de modelo univariado, significa que a série é bem descrita por um modelo do tipo $y_t = \beta_0 + \beta_1\, t + v_t$, $t = 1, 2, \ldots, n$, com $v_t : \text{E}(v_t) = 0, \forall t$, e podendo estar autocorrelacionado, *ma non troppo*.

$$\widehat{lgas_t} = 2.315 + 0.002\,t - 0.082\,Q_{1t} + 0.026\,Q_{2t} + 0.065\,Q_{3t} - 0.427 lpreg_t,\ R^2 = 0.912, \quad (2)$$

onde $Q_{jt}, j = 1, \ldots, 4$, são *dummies* sazonais (trimestrais).

a) Relativamente à equação (1):

☐ a estimativa que ela proporciona para o coeficiente de *lpreg* é muito plausível.

☐ a heteroscedasticidade dos erros torna o estimador OLS enviesado.

☐ o facto de o R^2 ser mais baixo que o da equação (2) significa que o estimador OLS é enviesado.

☐ a omissão do termo de tendência torna o estimador OLS do coeficiente de *lpreg* enviesado, com um sinal inadequado.

b) Considerando a equação **(2)**, interprete as estimativas dos coeficientes de t e de Q_{3t}.

c) Com base na mesma amostra mas com as *dummies* sazonais Q_{2t}, Q_{3t} e Q_{4t}, estimou-se também o modelo

$$lgas_t = \alpha + \beta\,t + \alpha_2\,Q_{2t} + \alpha_3\,Q_{3t} + \alpha_4\,Q_{4t} + \beta_1 lpreg_t + u_t. \quad (3)$$

Por conseguinte, obteve-se:

☐ $\widehat{\alpha} = 2.233, \widehat{\alpha_2} = 0.108, \widehat{\alpha_3} = 0.147, \widehat{\alpha_4} = 0.082.$

☐ $\widehat{\alpha} = 2.233, \widehat{\alpha_2} = 0.147, \widehat{\alpha_3} = 0.108, \widehat{\alpha_4} = 0.082.$

☐ $\widehat{\alpha} = 2.315, \widehat{\alpha_2} = -0.108, \widehat{\alpha_3} = -0.147, \widehat{\alpha_4} = -0.082.$

☐ a informação disponível não é suficiente para indicar as estimativas da equação (3).

10. (Exercício 7 do exame de EN de 9/01/2013). Suponha que dispõe de observações trimestrais das variáveis y_t e z_t, ambas estacionárias em tendência e a primeira com sazonalidade. Especifique um modelo (explicitando todas as variáveis) que permita:

- comparar a evolução de y_t nos vários trimestres tendo por base o 1° trimestre;
- estimar a resposta percentual de y face a variações absolutas de z e
- evitar obter resultados espúrios de estimação e de inferência estatística.

11. Exercício 10.1 de W.

12. Exercício 10.C2 de W.

13. Exercício 10.C5 de W.

14. (Exercício 4 do exame de EN de 1/6/2010.) Para explicar as vendas (VENDAS) de determinada empresa em Portugal, foram estimadas várias regressões. As variáveis têm o seguinte significado: T = @trend+1, T2, T3 e T4 são as variáveis dummies sazonais e EURO é uma variável que assume o valor 1 a partir de 2002, data da entrada em circulação do euro.

Equação 1
Dependent Variable: LOG(VENDAS)
Sample (adjusted): 1979Q2 2009Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.193668	0.071944	-2.691916	0.0081
T	0.010073	0.001399	7.201710	0.0000
EURO	-0.055604	0.108291	-0.513466	0.6086
R-squared	0.510909	F-statistic		62.67661
Sum squared resid	13.43298	Prob(F-statistic)		0.000000

Equação 2
Dependent Variable: LOG(VENDAS)
Sample (adjusted): 1979Q2 2009Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.171641	0.090095	-1.905116	0.0592
T	0.010098	0.001409	7.167887	0.0000
T2	-0.027449	0.086290	-0.318105	0.7510
T3	-0.077242	0.086292	-0.895119	0.3726
T4	0.015010	0.086301	0.173923	0.8622
EURO	-0.058817	0.109090	-0.539157	0.5908
R-squared	0.516456	F-statistic		24.99272
Sum squared resid	13.28063	Prob(F-statistic)		0.000000

a) Sabendo que o ficheiro de dados de EViews contém apenas as variáveis VENDAS e ANO (1979 a 2009), diga como pode gerar a variável EURO. Os resultados da equação 1 permitem afirmar que a introdução do euro teve um impacto estatisticamente significativo sobre as vendas? Justifique a sua resposta.

b) Com base nos resultados da equação 2, estima-se que, depois de removida a sazonalidade:

☐ as vendas crescem, em média, aproximadamente mais 1% a partir de 2002;

☐ as vendas crescem, em média, aproximadamente 1% por trimestre, tendo em conta a introdução do euro;

☐ as vendas crescem, em média, aproximadamente 0.01% por trimestre, tendo em conta a introdução do euro e depois de removida a tendência;

☐ as vendas crescem, em média, aproximadamente mais 1% no primeiro trimestre que nos restantes.

c) Ao nível de 5%, e em relação à presença de sazonalidade em log(vendas), pode-se concluir que

☐ não existem provas estatísticas da sua presença;

☐ existem provas estatísticas da sua presença;

☐ neste caso, o teste de sazonalidade não faz sentido porque os dados são trimestrais;

☐ a informação dada não permite realizar o teste de sazonalidade.

15. (Exercício 6 do exame de ER de 29/1/2014.) Suponha que dispõe de dados trimestrais de uma variável de vendas (V_t) e de uma de preços (P_t) de um bem, ambas sem tendência, e que pretende analisar se a primeira tem sazonalidade (regular). Então, representando com $T_{jt}(j = 1, \ldots, 4)$ as variáveis *dummy* trimestrais e assumindo que as hipóteses do modelo clássico são satisfeitas, o melhor é estimar ...

☐ ... o modelo $V_t = \delta_0 + \delta_1 t + \delta_2 P_t + v_t$ e testar $H_0 : \delta_1 = 0$ vs. $H_1 : \delta_1 \neq 0$.

☐ ... o modelo $V_t = \beta_1 T_{t1} + \beta_2 T_{t2} + \beta_3 T_{t3} + \beta_4 T_{t4} + u_t$ e testar $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ vs. $H_1 : \exists \beta_j \neq 0,\ j = 2, 3, 4$.

☐ ... o modelo $V_t = \beta_0 + \alpha_1 T_{t1} + \alpha_2 T_{t2} + \alpha_3 T_{t3} + \gamma P_t + e_t$ e testar $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs. $H_1 : \exists \alpha_j \neq 0,\ j = 1, 2, 3$.

☐ ... o modelo $V_t = \gamma_1 T_{t1} + \gamma_2 T_{t2} + \gamma_3 T_{t3} + \gamma_4 T_{t4} + w_t$ e testar $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$ vs. H_1 : não H_0.

16. (Exercício 9 do exame de EE de 11/09/2008.) Considere o seguinte modelo para dados trimestrais

$$y_t = \beta_0 + \delta\, t + \gamma_1 Q_{t1} + \gamma_2 Q_{t2} + \gamma_3 Q_{t3} + \beta_1 x_{t1} + \beta_2 x_{t2} + u_t,$$

e as seguintes séries de resíduos:

$\hat{u}_{tA}$ – resíduos da regressão de y_t sobre c, t, Q_{t1}, Q_{t2} e Q_{t3};
$\hat{u}_{tx_1}$ – resíduos da regressão de x_{t1} sobre c, t, Q_{t1}, Q_{t2} e Q_{t3};
$\hat{u}_{tx_2}$ – resíduos da regressão de x_{t2} sobre c, t, Q_{t1}, Q_{t2} e Q_{t3};
$\hat{u}_{tB}$ – resíduos da regressão de y_t sobre c, x_{t1} e x_{t2}.

Então, as estimativas OLS de β_1 e β_2 podem ser obtidas:

☐ fazendo a regressão de $\hat{u}_{tB}$ sobre $\hat{u}_{tx_1}$ e $\hat{u}_{tx_2}$;

☐ fazendo a regressão de $\hat{u}_{tA}$ sobre $\hat{u}_{tx_1}$ e $\hat{u}_{tx_2}$;

☐ fazendo a regressão de $\hat{u}_{tx_1}$ sobre $\hat{u}_{tx_1}$ e $\hat{u}_{tx_2}$;

☐ fazendo a regressão de $\hat{u}_{tx_1}$ sobre $\hat{u}_{tB}$ e $\hat{u}_{tx_2}$.

17. (Do exame de EN de 4/1/2017.) A hipótese de amostragem aleatória não é assumida para os modelos de dados de séries temporais porque estes dados ...

- [ ] ... são geralmente muito certinhos, pouco aleatórios.
- [ ] ... são sempre muito heteroscedásticos.
- [ ] ... quase nunca (se é que alguma vez) são independentes.
- [ ] ... têm geralmente muita tendência e pouca dinâmica.

$\infty\infty\infty$

Capítulo 4 — Tópicos Adicionais sobre a Utilização do OLS

Exercícios prioritários: 2, 4, 9, 14 e 17.

1. Exercício 11.1 de W.

2. Exercício 11.2 de W.

3. (Exercício 8 do exame de ER de 25/01/2012.) Suponha que $x_t = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}$, com $e_t \sim iid(0, \sigma^2)$. Então:

- [ ] $\text{Var}(x_{t+1}) > \text{Var}(x_t)$.
- [ ] $\text{E}(x_t) = 1 + \alpha_1 + \alpha_2$.
- [ ] $\text{Var}(x_t) = 1 + \alpha_1 + \alpha_2$.
- [ ] $\text{Cov}(x_t, x_{t-1}) = \alpha_1 \sigma^2 (1 + \alpha_2)$.

4. (Exercício 7 do exame de ER de 28/6/2011.) Para que o processo $y_t = \alpha_0 + \alpha_1 t + u_t$ seja estacionário em tendência, é estritamente necessário que $\{u_t\}$ seja um processo ...

- [ ] ... ruído branco, isto é, de variáveis independentes e identicamente distribuídas (iid).
- [ ] ... AR(1), $u_t = \rho u_{t-1} + e_t$, $e_t \sim iid(0, \sigma^2)$, com $|\rho| < 1$.
- [ ] ... sempre negativo, para compensar a tendência crescente.
- [ ] ... estacionário e fracamente dependente.

5. (Exercício 6 do exame de ER de 27/01/2011.) Considere que $x_t = 2 + 0.5x_{t-1} + e_t$, com $e_t \sim iid(0, \sigma^2)$. Qual dos seguintes processos $\{y_t\}$ abaixo não é estacionário, nem sequer em tendência?[4]

- [ ] $y_t = \alpha_0 + \alpha_1 x_t + v_t$, $v_t = 0.8 v_{t-1} + e_t$.

[4] A propriedade seguinte pode ajudar: se $x_t \sim I(0)$ e $z_t \sim I(0)$, então $y_t = (x_t + z_t) \sim I(0)$.

☐ $y_t = \begin{cases} \gamma_0 + \gamma_1\, x_t + u_t, \text{ para } t = 1, 2, \ldots, 20, \\ \gamma_0 + \gamma_2\, x_t + u_t, \; u_t \sim iid(0, \sigma_u^2), \; \gamma_1 \neq \gamma_2, \text{ para } t = 21, 22, \ldots. \end{cases}$.

☐ $y_t = \delta_0 + \delta_1\, x_t + \delta_2\, t + w_t$, com $w_t \sim iid(0, \sigma_w^2)$.

☐ $y_t = \beta_0 + \beta_1\, x_t + \beta_2\, x_{t-1} + \beta_3\, x_{t-3} + e_t$.

6. (Exercício 8 do exame de ER de 25/6/2010.) Se um processo $\{x_t\}$ satisfaz a condição $Cov(x_t, x_{t+k}) = 0, \forall k \geq 2$, qual das seguintes possibilidades é mais plausível? Nota: $e_t \sim iid(0, \sigma^2)$.

☐ $x_t = \beta_0 + e_t + \alpha\, e_{t-1}$;

☐ $x_t = \gamma_0 + x_{t-1} + e_t$;

☐ $x_t = \delta_0 + \delta_1\, t + u_t, \; u_t = 0.5\, u_{t-1} + e_t$;

☐ $x_t = \alpha_0 + \alpha_1 x_{t-1} + e_t, \; |\alpha_1| < 1$.

7. (Exercício 6 do exame de ER de 25/1/2012.) Indique a afirmação que é **FALSA**:

☐ nos modelos de séries temporais também podem existir problemas de "armadilha de variáveis artificiais".

☐ se o modelo $\widehat{\log(y_t)} = 0.23 + 0.02\, t, \; t = 1, 2, \ldots n$, se adequa bem à série temporal anual é porque a sua taxa anual média de variação é, aproximadamente, de 2% por ano no período.

☐ a hipótese de exogeneidade contemporânea exclui a possibilidade de existir *feedback* de y para os valores futuros dos regressores.

☐ há séries económicas mensais e trimestrais que não têm sazonalidade.

8. (Exercício 8 do exame de ER de 2/7/2013.) Das seguintes afirmações, indique a que é **FALSA**. A hipótese de exogeneidade contemporânea ...

☐ não permite que existam regressores omitidos correlacionados contemporaneamente com os incluídos.

☐ permite *feedback* contemporâneo, implicando correlação entre o erro e o regressor do mesmo período de tempo.

☐ não exige nada sobre as correlações dos erros do modelo com os regressores de outros períodos de tempo.

☐ é mais fraca que a hipótese $\mathrm{E}(u_t \,|\, \mathbf{X}) = 0, \forall t$.

9. (Exercício 7 do exame de EN de 12/1/2011 modificado.) Considere o modelo

$$y_t = \beta_0 + \beta_1\, t + \beta_2\, T_2 + \beta_3\, T_3 + \beta_4\, T_4 + \beta_5\, y_{t-1} + u_t,$$

referente a uma série trimestral com tendência e onde T_2, T_3 e T_4 são *dummies* sazonais (trimestrais). Com base na informação dada pode-se concluir que o estimador OLS não é centrado porque ...

- [ ] ... os erros, u_t, são autocorrelacionados.
- [ ] ... y_t é uma série com tendência.
- [ ] ... os regressores não satisfazem a condição de exogeneidade estrita.
- [ ] ... existem provas estatísticas de sazonalidade na série y_t.

10. (Exercício 4 do exame de ER de 3/2/2016). O estimador OLS dos coeficientes do modelo

$$y_t = \beta_0 + \alpha\, y_{t-1} + \delta_0\, z_t + \delta_1\, z_{t-1} + u_t,$$

- [ ] pode ser centrado, se $\mathrm{E}(u_t|\, y_{t-1}, y_{t-2}, \ldots, z_t, z_{t-1}, \ldots) = 0, \forall t$, mas nunca poderá ser consistente.
- [ ] será sempre centrado e consistente desde que $\mathrm{Cov}(u_t, u_s|\, y_{-1}, z_t, z_{t-1}) = 0, \forall t, s, t \neq s$.
- [ ] nunca será centrado nem consistente.
- [ ] pode ser consistente, se $\mathrm{E}(u_t|\, y_{t-1}, z_t, z_{t-1}) = 0, \forall t$, mas nunca poderá ser centrado.

11. Exercício 11C.7 i) de W.

12. (Exercício 7 do exame de ER de 25/6/2009.) Considere que $y_t \sim I(0)$. Então, a hipótese $E(y_t|\, y_{t-1}, y_{t-2}, \ldots) = E(y_t)$ pode ser testada estimando a regressão

- [ ] $y_t = \beta_0 + \beta_1\, y_{t-1} + \beta_2\, y_{t-2} + u_t$ e testando $H_0 : \beta_0 = 0$.
- [ ] $y_t = \beta_0 + \beta_1\, y_{t-1} + \beta_2\, y_{t-2} + u_t$ e testando $H_0 : \beta_1 = \beta_2 = 0$.
- [ ] $y_t = \beta_0 + \beta_1\, t + \rho\, y_{t-1} + e_t$ e testando $H_0 : \rho = 1$.
- [ ] $y_t = \beta_0 + \beta_1\, y_{t-1} + \beta_2\, y_{t-2} + u_t$ e testando $H_0 : \beta_0 = \beta_1 = \beta_2 = 0$.

13. Exercício 11.6 i), ii) e iii) de W.

14. Exercício 11C.1 i), ii) e iv) de W.

15. Exercício 11C.5 de W.

16. Exercício 11C.9 de W.

17. (Exercício 9 do exame de ER de 25/06/2009.) Das seguintes afirmações, indique a que é **FALSA**. Sabendo que o modelo

$$y_t = \beta_0 + \alpha\, y_{t-1} + \beta_1\, z_{t-1} + u_t$$

é dinamicamente completo, então

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

☐ $\mathrm{Cov}(u_t, u_s) = 0, \forall t, s, t \neq s$;

☐ $\mathrm{E}(y_t|\, y_{t-1}, z_{t-1}, y_{t-2}, z_{t-2}, \ldots) = \mathrm{E}(y_t|\, y_{t-1}, z_{t-1})$;

☐ no modelo $y_t = \beta_0 + \alpha\, y_{t-1} + \beta_1\, z_{t-1} + \beta_2\, z_{t-2} + v_t$, β_2 é diferente de zero.

18. (Exercício 9 do exame de EN de 10/1/2012.) Considere o modelo $y_t = \gamma + \delta_0\, z_t + \delta_1\, z_{t-1} + u_t$. Para que este modelo possa ser considerado dinamicamente completo, é necessário que na equação estimada

$$\begin{array}{rllll} \widehat{y}_t = & 2.35+ & 0.62\, z_t+ & 0.21\, z_{t-1}+ & \mathbf{A}\, y_{t-1}, \\ & (1.01) & (0.27) & (0.10) & (0.20) \end{array}$$

entre os seguintes valores, **A** seja igual a

☐ $\mathbf{A} = 0.30$. ☐ $\mathbf{A} = 1.00$. ☐ $\mathbf{A} = 0.75$. ☐ $\mathbf{A} = -0.60$.

∞∞∞

Capítulo 5 — Autocorrelação e Heteroscedasticidade

Exercícios prioritários: 3, 4, 7 e 9.

1. Exercício 12.1 de W.

2. (Exercício 10 do exame de ER de 29/01/2014) Considere o modelo de regressão $y_t = \beta_0 + \beta_1 z_t + u_t$, e suponha que se sabe que u_t segue um processo AR(1) estacionário, $u_t = \rho\, u_{t-1} + e_t$, $e_t \sim iid(0, \sigma^2)$, $|\rho| < 1$. Então:

☐ apesar disso, os métodos de inferência usuais permanecem válidos.

☐ em geral, os testes de significância de β_1 tenderão a rejeitar H_0 mais frequentemente do que deveriam.

☐ em geral, esse problema implica que $se(\hat{\beta}_1) < \sqrt{\widehat{\mathrm{Var}(\hat{\beta}_1)}}$.

☐ se as restantes hipóteses do modelo clássico são satisfeitas, o estimador OLS continua a ser BLUE.

3. Exercício 12.5 i) de W.

4. (Exercício 10 do exame de ER de 27/1/2011.) Com 31 observações anuais, estimou-se o modelo

$$y_t = 0.893 + 0.778\, y_{t-1} + \hat{u}_t,\ R^2 = 0.64.$$

Empregando os resíduos OLS deste modelo ($\hat{u}_t$), obtiveram-se os seguintes resultados:

$$\hat{\hat{u}}_t = \underset{(0.177)}{0.287}\, \hat{u}_{t-1},\quad R^2 = 0.08,$$

$$\hat{\hat{u}}_t = \underset{(0.522)}{0.710} - \underset{(0.131)}{0.186}\, y_{t-1} + \underset{(0.220)}{0.473}\, \hat{u}_{t-1},\quad R^2 = 0.15\,.$$

Por conseguinte, relativamente à presença de autocorrelação de primeira ordem nos erros do modelo:

☐ não há informação suficiente para fazer inferência sobre a sua presença.

☐ encontram-se provas estatísticas da sua presença, empregando um teste com $\alpha = 0.05$.

☐ com base num teste com $\alpha = 0.05$, não se encontram provas estatísticas da sua presença.

☐ nenhuma das respostas anteriores é correcta.

5. a) Exercício 12C.1 de W.

   b) Acrescente como regressor a variável dependente desfasada e teste a presença de autocorrelação de primeira ordem nos erros da nova equação. Comente os resultados.

6. (Exercício 5c) do exame de 4/9/2017.) Admita que se sabe que $lenel_t$, o logaritmo do consumo de energia eléctrica e $lpre_t$, o logaritmo do seu preço médio (ambas em preços constantes), são ambas estacionárias em tendência, com uma tendência crescente com o tempo. Suponha que se estimaram os seguintes modelos:

$$\widehat{lenel_t} = 4.51 + 0.575\, lpre_t,\ R^2 = 0.49 \ \text{ e} \tag{4}$$

$$\widehat{lenel_t} = 4.46 + 0.007\, t + 0.186\, Q_{t1} - 0.012\, Q_{t2} + 0.095\, Q_{t3} - 0.321 lpre_t,\ R^2 = 0.81, \tag{5}$$

onde $Q_{tj}, j = 1, \ldots, 4$, são *dummies* sazonais (trimestrais).

Suponha que os regressores Q_{t1}, Q_{t2} e Q_{t3} do modelo da equação (2) são "altamente significativos" e que foram obtidas as estatísticas h-alt e BG(4) respeitantes à equação (1). Então, entre os seguintes pares de valores (h-alt, BG(4)), qual considera ser o mais plausível?

☐ (−1.2; 16.3). ☐ (5.2; 33.7). ☐ (0.8; 5.3). ☐ (4.8; 3.4).

7. (Exercício 4b) do exame de ER de 25/6/2010.) Para compreender a evolução das vendas de relógios em determinado país, estimou-se o modelo seguinte, com dados trimestrais respeitantes aos últimos 15 anos:

$$\begin{array}{lllllll} \widehat{lrelog} = & 2.054 & -0.017\,t & -0.708\,T1 & -0.322\,T2 & -0.724\,T3 & -0.601\,lpreco, \\ & (0.558) & (0.008) & (0.434) & (0.434) & (0.444) & (0.074) \end{array}$$

onde *lrelog* representa o logaritmo do número de relógios vendidos, $T1$, $T2$ e $T3$ são *dummies* sazonais e *lpreco* representa o logaritmo do preço médio dos relógios vendidos.

Com os resíduos do modelo ($\hat{u}_t$), e empregando as últimas 56 observações, obteve-se

$$\hat{\hat{u}}_t = -0.10 - 0.001\,t + 0.096\,T1 + 0.077\,T2 + 0.122\,T3 + 0.005\,lpreco + 0.412\,\hat{u}_{t-1} + 0.129\,\hat{u}_{t-2} - 0.255\,\hat{u}_{t-3} + 0.442\,\hat{u}_{t-4},$$

$$F - statistic = 3.133, \quad R^2 = 0.380, \quad \bar{R}^2 = 0.259, \quad h - alt = 1.222.$$

Formalize as hipóteses, efectue o teste estatístico que corresponde a esta regressão auxiliar e retire a conclusão adequada.

8. Exercício 12.6 de W.

9. a) Exercício 12C.6 i) e ii) de W.
   b) Considerando ainda o modelo de 12C.6 ii), teste a presença de autocorrelação nos erros até à ordem 2.

Capítulo 6 — Introdução aos Métodos para Dados de Painel

Exercícios prioritários: 4, 5, 6, 7, 8 e 9.

1. Exercício 13.2 de W.
2. Exercício 13.C1 de W.
3. Exercício 13.C3 de W.
4. Exercício 13.3 de W.
5. Exercício 13.C5 de W.
6. Exercício 13.4 de W.
7. Exercício 13.5 de W.

8. Exercício 13.6 de W.

9. Exercício 13.7 de W.

10. Exercício 13.C7 de W.

11. (Exercício do exame de 7/3/2018 modificado.) Com dados respeitantes aos anos de 1995 e 2010 das mesmas 25 freguesias das áreas metropolitanas de Lisboa e do Porto, e com informação sobre as variáveis

    – $PM2_{it}$: preço médio do metro quadrado da habitação construída;
    – $A2010_t$: variável *dummy* com o valor 1 para o ano de 2010 (e 0 no caso contrário);
    – MAR_i: variável *dummy* com o valor 1 se a freguesia se situa perto do mar ou do rio;
    – $IQUA_{it}$: índice de qualidade média da habitação construída na freguesia;
    – MET_{it}: variável *dummy* com o valor 1 se, entre 1995 e 2010, a freguesia passou a ter uma ou mais estações de metro,

    estimaram-se as equações:

    (1) com o OLS agregado (*"pooled"*) sobre as 50 observações:

$$\widehat{\log(PM2)}_{it} = \underset{(2.356)}{7.6155} + \underset{(0.111)}{0.615}A2010_t + \underset{(0.091)}{0.351}MAR_i + \underset{(0.061)}{1.154}\log(IQUA)_{it} - \underset{(0.101)}{0.031}MET_{it};$$

    (2) com o OLS sobre as 25 observações diferenciadas:

$$\Delta\widehat{\log(PM2)} = \underset{(0.034)}{0.112} + \underset{(2.651)}{0.391}\Delta\log(IQUA) + \underset{(0.103)}{0.342}\Delta MET.$$

    Classifique e comente (de forma justificada) as afirmações:

    a) *"Tendo em consideração a primeira equação, é óbvio que a segunda sofre de um problema de variáveis omitidas. ..."*.

    b) *... "E até a variável* $\log(IQUA)$ *acaba por ser estatisticamente irrelevante!"*.

ASL, 24/7/2018