

## Exercícios em Computador

### CAPÍTULO 2

---

**2.10** Os dados em 401K.RAW são um subconjunto de dados analisados por Papke (1995) para estudar a relação entre a participação em um plano de pensão 401(k) dos Estados Unidos e a generosidade do plano. A variável *taxap* é a porcentagem de trabalhadores com uma conta ativa; essa é a variável que gostaríamos de explicar. A medida de generosidade é a taxa de complementação do plano, *taxcont*. Essa variável dá a quantidade média com a qual a firma contribui, em cada plano do trabalhador, para cada \$ 1 de contribuição do trabalhador. Por exemplo, se  $taxcont = 0,50$ , a contribuição do trabalhador é complementada por uma contribuição de 50 cents pela firma.

- (i) Ache a taxa de participação média e a taxa de complementação média na amostra de planos.
- (ii) Agora, estime a equação de regressão simples

$$taxap = \hat{\beta}_0 + \hat{\beta}_1 taxcont,$$

e relate seus resultados juntamente com o tamanho da amostra e o *R*-quadrado.

- (iii) Interprete o intercepto de sua equação. Interprete o coeficiente de *taxcont*.
- (iv) Ache o *taxap* predito quando  $taxcont = 3,5$ . Essa predição é razoável? Explique o que está acontecendo.
- (v) Quanto da variação em *taxap* é explicado por *taxcont*? Em sua opinião, isso é bastante?

**2.11** Os dados em CEOSAL2.RAW contêm informações sobre chefes-executivos (CEOs) de corporações dos Estados Unidos. A variável *salário* é a compensação anual, em milhares de dólares, e *permceo* é o número de anos na condição de CEO na companhia.

- (i) Ache o salário médio e a permanência média da amostra.
- (ii) Quantos CEOs estão em seu primeiro ano na posição de CEO (isto é,  $permceo = 0$ )?
- (iii) Estime o modelo de regressão simples

$$\log(\text{salário}) = \beta_0 + \beta_1 permceo + u,$$

e relate seus resultados na forma usual. Qual é o aumento da percentagem predita (aproximado) no salário, dado um ano a mais como CEO?

**2.12** Use os dados em SLEEP75.RAW, de Biddle e Hamermesh (1990), para estudar se há um *tradeoff* entre o tempo gasto dormindo por semana e o tempo gasto em um trabalho pago. Poderíamos usar outra variável como variável dependente. Estime o modelo

$$dormir = \beta_0 + \beta_1 trabtot + u,$$

em que *dormir* corresponde a minutos gastos dormindo à noite por semana, e *trabtot* é o total de minutos trabalhados durante a semana.

- (i) Reporte seus resultados na forma de equação, juntamente com o número de observações e  $R^2$ . O que o intercepto significa nessa equação?
- (ii) Se *trabtot* aumenta em duas horas, em quanto se estima que *dormir* irá cair? Você acha que isso é um efeito grande?

**2.13** Utilize os dados em WAGE2.RAW para estimar uma regressão simples que explique o salário mensal (*salário*) em termos do escore do QI (*QI*).

- (i) Ache o salário médio e o QI médio da amostra. Qual é o desvio-padrão de *QI*? (Os escores do QI são padronizados, de modo que a média é 100, na população, com um desvio-padrão igual a 15.)
- (ii) Estime um modelo de regressão simples em que um aumento de um ponto em *QI* faça com que *salário* varie em uma quantidade constante em dólar. Use esse modelo para achar o aumento predito no salário a partir de um aumento de 15 pontos em *QI*. O *QI* explica muito da variação em *salário*?
- (iii) Agora, estime um modelo em que cada aumento de um ponto em *QI* tenha o mesmo efeito percentual sobre *salário*. Se *QI* aumenta em 15 pontos, qual é o aumento percentual aproximado no *salário* predito?

**2.14** Para a população de firmas da indústria química, seja *pq* os gastos anuais em pesquisa e desenvolvimento, e seja *vendas* as vendas anuais (ambos estão em milhares de dólares).

- (i) Escreva um modelo (e não uma equação estimada) que implique uma elasticidade constante entre *pq* e *vendas*. Qual parâmetro é a elasticidade?
- (ii) Agora, estime o modelo usando os dados em RDCHEM.RAW. Escreva a equação estimada na forma usual. Qual é a elasticidade estimada de *pq* com respeito a *vendas*? Explique, em palavras, o que essa elasticidade significa.

## CAPÍTULO 3

---

**3.13** Um problema de interesse das autoridades da saúde (e outras) é determinar os efeitos que fumar durante a gravidez exerce sobre a saúde do recém-nascido. Uma medida da saúde do recém-nascido é o peso de nascimento; um peso de nascimento muito baixo pode atribuir à criança o risco de contrair várias doenças. Como outros fatores que afetam o peso de nascimento, além de fumar cigarros, estão provavelmente correlacionados com o fumo, devemos levar em consideração tais fato-

res. Por exemplo, uma renda maior geralmente permite acesso a pré-natais melhores, bem como a uma nutrição melhor da mulher. Uma equação que reconhece isso é

$$pesonas = \beta_0 + \beta_1cigs + \beta_2rendfam + u.$$

- (i) Qual é o sinal mais provável de  $\beta_2$ ?
- (ii) Você acha que *cigs* e *rendfam* estão, provavelmente, correlacionados? Explique por que a correlação pode ser positiva ou negativa.
- (iii) Agora, estime a equação com e sem *rendfam*, usando os dados em BWGHT.RAW. Relate os resultados na forma de uma equação, incluindo o tamanho da amostra e o *R*-quadrado. Discuta seus resultados, dando ênfase ao fato de acrescentar *rendfam* mudar ou não substancialmente o efeito estimado de *cigs* sobre *pesonas*.

**3.14** Use os dados em HPRICE1.RAW para estimar o modelo

$$preço = \beta_0 + \beta_1mquad + \beta_2banhos + u,$$

em que *preço* é o preço da residência medido em milhares de dólares.

- (i) Escreva os resultados na forma de uma equação.
- (ii) Qual é o aumento estimado no preço para uma casa com um banheiro a mais, mantendo constante o metro quadrado?
- (iii) Qual é o aumento estimado no preço para uma casa com um banheiro adicional, a qual tem 140 metros quadrados de tamanho? Compare sua resposta à parte (ii).
- (iv) Qual é a percentagem da variação no preço que é explicada pelo metro quadrado e pelo número de banheiros?
- (v) A primeira casa na amostra tem *mquad* = 2.438 e *banhos* = 4. Ache o preço de venda predito para essa casa a partir da reta de regressão de MQO.
- (vi) O preço de venda real da primeira casa na amostra foi de \$ 300.000 (assim, *preço* = 300). Ache o resíduo para essa casa. Isso sugere que o comprador pagou mais ou menos por ela?

**3.15** O arquivo CEOSAL2.RAW contém dados de 177 diretores, os quais podem ser utilizados para examinar os efeitos do desempenho da firma sobre o salário do CEO.

- (i) Estime um modelo que relacione o salário anual às vendas da firma e ao seu valor de mercado. Faça um modelo de elasticidade constante para ambas as variáveis independentes. Escreva os resultados na forma de uma equação.
- (ii) Acrescente *lucros* ao modelo da parte (i). Por que essa variável não pode ser incluída na forma logarítmica? Você diria que as variáveis de desempenho dessa firma explicam muito da variação nos salários do CEO?
- (iii) Acrescente a variável *perceo* ao modelo da parte (ii). Qual é o retorno percentual estimado para um ano a mais da permanência do CEO no emprego atual, mantendo fixos os outros fatores?

- (iv) Ache o coeficiente de correlação amostral entre as variáveis  $\log(\text{valmerc})$  e  $\text{lucros}$ . Essas variáveis são altamente correlacionadas? O que isso diz sobre os estimadores de MQO?

**3.16** Use os dados em ATTEND.RAW para esse exercício.

- (i) Obtenha os valores mínimo, máximo e médio das variáveis  $\text{taxafreq}$ ,  $\text{priGPA}$  e  $\text{ACT}$ .  
 (ii) Estime o modelo

$$\text{taxafreq} = \beta_0 + \beta_1 \text{priGPA} + \beta_2 \text{ACT} + u,$$

e escreva os resultados em forma de equação. Interprete o intercepto. Ele tem um significado útil?

- (iii) Discuta os coeficientes de inclinação estimados. Há alguma surpresa?  
 (iv) Qual é a  $\text{taxafreq}$  predita se  $\text{priGPA} = 3,65$  e  $\text{ACT} = 20$ ? O que você extrai desse resultado? Há algum estudante na amostra com esses valores nas variáveis explicativas?  
 (v) Se Estudante A tem  $\text{priGPA} = 3,1$  e  $\text{ACT} = 21$ , e Estudante B tem  $\text{priGPA} = 2,1$  e  $\text{ACT} = 26$ , qual é a diferença predita em suas taxas de frequência?

**3.17** Confirme a interpretação de imparcialização das estimativas de MQO fazendo explicitamente a imparcialização para o Exemplo 3.2. Isso requer, em primeiro lugar, que você faça a regressão de  $\text{educ}$  sobre  $\text{exper}$  e  $\text{perm}$  e salve os resíduos,  $\hat{r}_1$ . Então, regrida  $\log(\text{salário})$  sobre  $\hat{r}_1$ . Compare o coeficiente de  $\hat{r}_1$  com o coeficiente de  $\text{educ}$  da regressão de  $\log(\text{salário})$  sobre  $\text{educ}$ ,  $\text{exper}$  e  $\text{perm}$ .

**3.18** Use o banco de dados em WAGE2.RAW para este problema. Como de costume, esteja seguro de que todas as seguintes regressões contenham um intercepto.

- (i) Rode uma regressão de  $\text{QI}$  sobre  $\text{educ}$  para obter o coeficiente de inclinação — digamos  $\tilde{\delta}_1$ .  
 (ii) Rode a regressão de  $\log(\text{salário})$  sobre  $\text{educ}$  e obtenha o coeficiente de inclinação,  $\tilde{\beta}_1$ .  
 (iii) Rode a regressão múltipla de  $\log(\text{salário})$  sobre  $\text{educ}$  e  $\text{QI}$  e obtenha os coeficientes de inclinação,  $\hat{\beta}_1 + \hat{\beta}_2$  e, respectivamente.  
 (iv) Verifique que  $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$ .

## CAPÍTULO 4

**4.12** O modelo seguinte pode ser usado para estudar se os gastos de campanha afetam os resultados da eleição:

$$\text{votoA} = \beta_0 + \beta_1 \log(\text{gastoA}) + \beta_2 \log(\text{gastoB}) + \beta_3 (\text{forpartA}) + u,$$

em que  $\text{votoA}$  é a porcentagem de votos recebidos pelo Candidato A,  $\text{gastoA}$  e  $\text{gastoB}$  são os gastos de campanha dos Candidatos A e B, e  $\text{forpartA}$  é uma medida da força do partido do Candidato A (a porcentagem dos mais recentes votos presidenciais que foram para o partido de A).

- (i) Qual é a interpretação de  $\beta_1$ ?

- (ii) Em termos dos parâmetros, formule a hipótese nula de que um aumento de 1% nos gastos de A é compensado por um aumento de 1% nos gastos de B.
- (iii) Estime o modelo dado usando os dados em VOTE1.RAW e relate os resultados na forma usual. Os gastos de A afetam o resultado? E os gastos de B? Você pode usar esses resultados para testar a hipótese da parte (ii)?
- (iv) Estime um modelo que dê diretamente a estatística  $t$  para testar a hipótese da parte (ii). O que você conclui? (Use uma alternativa bilateral.)

**4.13** Use os dados em LAWSCH85.RAW para este exercício.

- (i) Usando o mesmo modelo do Problema 3.4, formule e teste a hipótese nula de que o *ranking* das escolas de direito não tem efeito *ceteris paribus* sobre o salário mediano inicial.
- (ii) As características da classe nova de estudantes – a saber, *LSAT* e *GPA* – são individualmente ou conjuntamente significativas para explicar *salário*?
- (iii) Teste se o tamanho da classe iniciante (*tamclas*) ou o tamanho da faculdade (*tamfac*) precisam ser acrescentados a essa equação; faça um único teste. (Esteja atento com os dados de *tamclas* e *tamfac* que faltam.)
- (iv) Quais fatores devem influenciar o *ranking* da escola de direito que não estão incluídos na regressão do salário?

**4.14** Consulte o Problema 3.14. Agora, use o log do preço da casa como a variável dependente:

$$\log(\text{preço}) = \beta_0 + \beta_1 \text{arquad} + \beta_2 \text{qtdorm} + u.$$

- (i) Você está interessado em estimar e obter um intervalo de confiança da variação percentual do preço quando um quarto de 150 pés quadrados é acrescentado à casa. Na forma decimal, temos  $\theta_1 = 150\beta_1 + \beta_2$ . Use os dados em HPRICE1.RAW para estimar  $\theta_1$ .
- (ii) Escreva  $\beta_2$  em termos de  $\theta_1$  e  $\beta_1$ , e coloque isso na equação do  $\log(\text{preço})$ .
- (iii) Use a parte (ii) para obter um erro-padrão de  $\hat{\theta}_1$  e use esse erro-padrão para construir um intervalo de confiança de 95%.

**4.15** No Exemplo 4.9, a versão restrita do modelo pode ser estimada usando todas as 1.388 observações na amostra. Calcule o  $R$ -quadrado da regressão de *pesnasc* sobre *cigs*, *ordnas* e *rendfam* usando todas as observações. Compare esse resultado com o  $R$ -quadrado informado pelo modelo restrito do Exemplo 4.9.

**4.16** Use os dados em MLB1.RAW para este exercício.

- (i) Use o modelo estimado na equação (4.31) e retire a variável *rbisyrr*. O que acontece à significância estatística de *hrunsyr*? E quanto à magnitude do coeficiente de *hrunsyr*?
- (ii) Acrescente as variáveis *runsyr*, *fldperc* e *sbasesyr* ao modelo da parte (i). Quais desses fatores são individualmente significativos?
- (iii) No modelo da parte (ii), teste a significância conjunta de *rebmed*, *fldperc* e *sbasesyr*.

**4.17** Use os dados em WAGE2.RAW para este exercício.

- (i) Considere a equação padrão do salário

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + u.$$

Formule a hipótese nula de que um ano a mais de experiência geral da força de trabalho tem o mesmo efeito sobre  $\log(\text{salário})$  que um ano a mais de permanência com o empregador atual.

- (ii) Teste a hipótese nula da parte (i) contra a alternativa bilateral, ao nível de significância de 5%, construindo um intervalo de confiança de 95%. O que você conclui?

**4.18** Consulte o exemplo usado na Seção 4.4. Você usará os dados em TWOYEAR.RAW.

- (i) A variável *pensmed* é o percentil no ensino médio da pessoa. (Um número maior é melhor. Por exemplo, 90 significa que sua classificação é melhor que 90 por cento de sua turma de graduação.) Ache o *pensmed* menor, maior e médio da amostra.
- (ii) Acrescente *pensmed* à equação (4.26) e informe as estimativas de MQO na forma usual. O *pensmed* é estatisticamente significativo? Quanto vale, em termos salariais, dez pontos percentuais na classificação do ensino médio?
- (iii) Acrescentar *pensmed* a (4.26) mudou substancialmente as conclusões sobre os retornos das faculdades de dois e quatro anos? Explique.
- (iv) O conjunto de dados contém uma variável chamada *id*. Explique por que, ao acrescentar *id* à equação (4.17) ou (4.26), você espera que ele seja insignificante. Verifique que ele é insignificante.

**4.19** O conjunto de dados 401KSUBS.RAW contém informações sobre a riqueza financeira líquida (*finliq*), a idade do respondente da pesquisa (*idade*), a renda familiar anual (*rend*), o tamanho da família (*tamfam*), e informações sobre a participação em certos planos de pensão para pessoas dos Estados Unidos. As variáveis de riqueza e renda são registradas em milhares de dólares. Para esta questão, use somente os dados para pessoas solteiras (portanto, *tamfam* = 1).

- (i) Quantas pessoas solteiras há no conjunto de dados?
- (ii) Use MQO para estimar o modelo

$$\text{finliq} = \beta_0 + \beta_1 \text{rend} + \beta_2 \text{idade} + u,$$

e relate os resultados usando o formato habitual. Esteja seguro de usar somente as pessoas solteiras da amostra. Interprete os coeficientes de inclinação. Há alguma surpresa nas estimativas de inclinação?

- (iii) O intercepto da regressão da parte (ii) tem um significado interessante? Explique.
- (iv) Ache o *p*-valor para o teste  $H_0: \beta_2 = 1$  contra  $H_0: \beta_2 < 1$ . Você rejeita  $H_0$  ao nível de 5%?
- (v) Se você fizer uma regressão simples de *finliq* sobre *rend*, o coeficiente estimado de *rend* é muito diferente do estimado na parte (ii)? Por quê?

## CAPÍTULO 5

---

**5.5** Use os dados em WAGE1.RAW para este exercício.

- (i) Estime a equação

$$\text{salário} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + u.$$

Salve os resíduos e faça um histograma.

- (ii) Repita a parte (i), mas com  $\log(\text{salário})$  como a variável dependente.  
 (iii) Você diria que a Hipótese RLM.6 está mais próxima de ser satisfeita para o modelo nível-nível ou para o modelo log-nível?

**5.6** Use os dados em GPA2.RAW para este exercício.

- (i) Usando todas as 4.137 observações, estime a equação

$$\text{colgpa} = \beta_0 + \beta_1 \text{hsperc} + \beta_2 \text{sat} + u$$

e informe os resultados na forma padrão.

- (ii) Estime novamente a equação da parte (i), usando as primeiras 2.070 observações.  
 (iii) Ache a razão dos erros-padrão sobre *hsperc* das partes (i) e (ii). Compare isso com o resultado de (5.10).

**5.7** Na equação (4.42) do Capítulo 4, calcule a estatística *LM* para testar se *educm* e *educp* são conjuntamente significativos. Ao obter os resíduos do modelo restrito, esteja seguro de que o modelo restrito seja estimado usando somente aquelas observações para as quais todas as variáveis do modelo irrestrito estejam disponíveis (veja Exemplo 4.9).

## CAPÍTULO 6

---

**6.8** Utilize os dados contidos no arquivo KIELMC.RAW, somente do ano de 1981, para responder as questões que seguem. Os dados são de imóveis vendidos em 1981 em North Andover, Massachusetts; 1981 foi o ano em que foi iniciada a construção de um incinerador de lixo local.

- (i) Para estudar os efeitos da localização do incinerador sobre os preços dos imóveis, considere o seguinte modelo de regressão simples

$$\log(\text{preço}) = \beta_0 + \beta_1 \log(\text{dist}) + u.$$

onde *preço* é o preço do imóvel em dólares e *dist* a distância entre o imóvel e o incinerador. Interpretando essa equação de forma causal, que sinal você espera para  $\beta_1$  se a presença do incinerador desvalorizar o imóvel? Estime esta equação e interprete os resultados.

- (ii) Ao modelo de regressão simples na parte (i), adicione as variáveis  $\log(\text{intst})$ ,  $\log(\text{arquad})$ ,  $\log(\text{tamterr})$ , *qtdorm*, *banhoss*, e *idade*, nas quais *intst* é a distância do

imóvel até a rodovia interestadual, *arquad* é a área construída do imóvel, *tamterr* é o tamanho do terreno, *qtdorm* é o número de quartos, *banhos* é o número de banheiros, e *idade* é a idade do imóvel em anos. Qual sua conclusão sobre os efeitos do incinerador agora? Explique por que (i) e (ii) produzem resultados conflitantes.

- (iii) Adicione  $[\log(intst)]^2$  ao modelo da parte (ii). Agora o que acontece? Qual sua conclusão sobre a importância da forma funcional?
- (iv) O quadrado de  $\log(dist)$  é significativo quando você adiciona essa variável ao modelo da parte (iii)?

**6.9** Utilize os dados contidos no arquivo WAGE1.RAW para fazer este exercício.

- (i) Utilize o MQO para estimar a equação

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + u$$

e descreva os resultados usando o formato habitual.

- (ii) A variável  $\text{exper}^2$  é estatisticamente significativa no nível de 1%?
- (iii) Utilizando a aproximação

$$\% \Delta \hat{\text{salário}} \approx 100(\hat{\beta}_2 + 2\hat{\beta}_3 \text{exper}) \Delta \text{exper},$$

encontre o retorno aproximado do quinto ano de experiência. Qual é o retorno aproximado do vigésimo ano de experiência?

- (iv) Em que valor de *exper* a experiência adicional reduz de fato o valor previsto de  $\log(\text{salário})$ ? Quantas pessoas possuem mais experiência nesta amostra?

**6.10** Considere um modelo no qual o retorno da educação depende do tempo de experiência de trabalho (e vice-versa):

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{educ} \cdot \text{exper} + u.$$

- (i) Mostre que o retorno de mais um ano de educação (na forma decimal), mantendo *exper* fixo, é  $\beta_1 + \beta_3 \text{exper}$ .
- (ii) Especifique a hipótese nula de que o retorno da educação não depende do nível de *exper*. O que você pensa que seja a hipótese alternativa apropriada?
- (iii) Utilize os dados contidos no arquivo WAGE2.RAW para testar a hipótese nula em (ii) contra sua hipótese alternativa proposta.
- (iv) Fazendo  $\theta_1$  representar o retorno da educação (na forma decimal), quando  $\text{exper} = 10$ :  $\theta_1 = \beta_1 + 10\beta_3$ . Obtenha  $\hat{\theta}_1$  e um intervalo de confiança de 95% de  $\theta_1$ . (Sugestão: Escreva  $\beta_1 = \theta_1 - 10\beta_3$  e incorpore essa expressão na equação, reorganizando-a em seguida. Isso produzirá a regressão para obter o intervalo de confiança de  $\theta_1$ .)



**6.11** Utilize os dados contidos no arquivo GPA2.RAW para fazer este exercício.

- (i) Estime o modelo

$$sat = \beta_0 + \beta_1 tamclas + \beta_2 tamclas^2 + u,$$

onde *tamclas* é o tamanho da classe no curso de graduação (em centenas), e escreva os resultados na forma habitual. O termo quadrático é estatisticamente significativo?

- (ii) Usando a equação estimada na parte (i), qual é o tamanho “ótimo” do ensino médio? Justifique sua resposta.
- (iii) Esta análise é representativa do desempenho acadêmico de todos os formados no ensino médio? Explique.
- (iv) Encontre o tamanho ótimo do ensino médio, usando  $\log(sat)$  como a variável dependente. Ele é muito diferente do que você obteve na parte (ii)?

**6.12** Utilize os dados dos preços dos imóveis contidos no arquivo HPRICE1.RAW para fazer este exercício.

- (i) Estime o modelo

$$\log(preço) = \beta_0 + \beta_1 \log(tamterr) + \beta_2 \log(arquad) + \beta_3 qtdorm + u$$

e descreva os resultados no formato MQO habitual.

- (ii) Encontre o valor previsto de  $\log(preço)$ , quando  $tamterr = 20.000$ ,  $arquad = 2.500$  e  $qtdorm = 4$ . Utilizando os métodos da Seção 6.4, encontre o valor estimado de *preço* nos mesmos valores das variáveis explicativas.
- (iii) Para explicar a variação em *preço*, decida se você prefere o modelo da parte (i) ou o modelo

$$preço = \beta_0 + \beta_1 tamterr + \beta_2 arquad + \beta_3 qtdorm + u.$$

**6.13** Utilize os dados contidos no arquivo VOTE1.RAW para fazer este exercício.

- (i) Considere um modelo com uma interação entre os gastos:

$$votoA = \beta_0 + \beta_1 forpartA + \beta_2 gastoA + \beta_3 gastoB + \beta_4 gastoA \cdot gastoB + u.$$

Qual é o efeito parcial de *gastoB* sobre *votoA*, mantendo *forpartA* e *gastoA* fixos? Qual é o efeito parcial de *gastoA* sobre *votoA*? O sinal esperado de  $\beta_4$  é óbvio?

- (ii) Estime a equação da parte (i) e descreva os resultados da forma habitual. O termo de interação é estatisticamente significativo?
- (iii) Encontre a média de *gastoA* na amostra. Fixe *gastoA* em 300 (representando 300.000 dólares). Qual é o efeito estimado de mais 100.000 dólares gastos pelo Candidato B sobre *votoA*? Esse efeito é grande?

- (iv) Agora, fixe  $gastoB$  em 100. Qual é o efeito estimado de  $\Delta gastoA = 100$  sobre  $votoA$ ? Isso faz sentido?
- (v) Agora, estime um modelo que substitua a interação por  $partA$ , a percentagem de participação do Candidato A nos gastos totais de campanha. Faz sentido manter tanto  $gastoA$  quanto  $gastoB$  fixos quando se altera  $partA$ ?
- (vi) (Requer cálculo) No modelo da parte (v), encontre o efeito parcial de  $gastoB$  sobre  $votoA$ , mantendo  $forpartA$  e  $gastoA$  fixos. Avalie isto com  $gastoA = 300$  e  $gastoB = 0$  e comente os resultados.

**6.14** Utilize os dados contidos no arquivo ATTEND.RAW para fazer este exercício.

- (i) No modelo do Exemplo 6.3, sustente que

$$\Delta respad / \Delta nmgradp \approx \beta_2 + 2\beta_4 nmgradp + \beta_6 taxafreq.$$

Use a equação (6.19) para estimar o efeito parcial quando  $nmgradp = 2,59$  e  $taxafreq = 0,82$ . Interprete sua estimativa.

- (ii) Mostre que a equação pode ser escrita como

$$\begin{aligned} respad = & \theta_0 + \beta_1 taxafreq + \theta_2 nmgradp + \beta_3 tac + \beta_4 (nmgradp - 2,59)^2 \\ & + \beta_5 tac^2 + \beta_6 nmgradp (taxafreq - 0,82) + u, \end{aligned}$$

onde  $\theta_2 = \beta_2 + 2\beta_4(2,59) + \beta_6(0,82)$ . (Note que o intercepto mudou, mas isso não é importante.) Use esta equação para obter o erro-padrão de  $\hat{\theta}_2$  da parte (i).

- (iii) Suponha que, em lugar de  $nmgradp(taxafreq - 0,82)$ , você use  $(nmgradp - 2,59)(taxafreq - 0,82)$ . Agora, como você interpreta o coeficiente da  $taxafreq$  e da  $nmgradp$ ?

**6.15** Utilize os dados contidos no arquivo HPRICE1.RAW para fazer este exercício.

- (i) Estime o modelo

$$preço = \beta_0 + \beta_1 tamterr + \beta_2 arquad + \beta_3 qtdorm + u$$

e descreva os resultados da forma habitual, incluindo o erro-padrão da regressão. Obtenha o preço previsto quando são inseridos  $tamterr = 10.000$ ,  $arquad = 2.300$ , e  $qtdorm = 4$ ; arredonde este preço para o inteiro mais próximo.

- (ii) Compute uma regressão que possibilite a você colocar um intervalo de confiança de 95% em torno do valor previsto da parte (i). Note que sua previsão será um pouco diferente devido ao erro de arredondamento.
- (iii) Seja  $preço^0$  o preço futuro desconhecido do imóvel com as características usadas nas partes (i) e (ii). Encontre um IC de 95% de  $preço^0$  e comente a amplitude desse intervalo de confiança.

**6.16** O conjunto de dados contidos no arquivo NBASAL.RAW contém informações a respeito de salários e estatísticas sobre a carreira de 269 jogadores de basquete da National Basketball Association (NBA) dos EUA.

- (i) Estime um modelo relacionando pontos por jogo (*pontos*) com anos como jogador profissional (*exper*), *idade*, e anos jogados na faculdade (*anuniv*). Inclua um termo quadrático em *exper*; as outras variáveis devem aparecer na forma de nível. Descreva os resultados da maneira habitual.
- (ii) Mantendo fixos os anos jogados na faculdade e a idade, em que valor de experiência a adição de mais um ano efetivamente reduz o salário? Isso faz sentido?
- (iii) Por que, na sua opinião, *anuniv* tem um coeficiente negativo e estatisticamente significativo? (*Sugestão*: Os jogadores da NBA podem ser convocados antes de terminarem a faculdade ou mesmo diretamente quando terminam o curso médio.)
- (iv) Adicione um termo quadrático em *idade* na equação. Ele é necessário? O que isso parece sugerir sobre os efeitos da idade, quando *exper* e *anuniv* são controladas?
- (v) Agora faça uma regressão do  $\log(\text{salário})$  sobre *pontos*, *exper*,  $\text{exper}^2$ , *idade*, e *anuniv*. Descreva os resultados da forma habitual.
- (vi) Verifique se *idade* e *anuniv* são conjuntamente significantes na regressão da parte (v). O que isso implica para saber se *idade* e *anuniv* têm efeitos separados sobre *salário*, quando produtividade e anos de experiência são consideradas?

**6.17** Utilize os dados contidos no arquivo BWGHT2.RAW para fazer este exercício.

- (i) Estime a equação

$$\log(\text{pesonas}) = \beta_0 + \beta_1 \text{conspn} + \beta_2 \text{conspn}^2 + u$$

por MQO, e descreva os resultados da maneira habitual. O termo quadrático é significativo?

- (ii) Mostre que, com base na equação da parte (i) o número estimado de consultas para exame pré-natal que maximiza  $\log(\text{pesonas})$  está em torno de 22. Quantas mulheres fizeram pelo menos 22 consultas para exame pré-natal na amostra?
- (iii) Faz sentido prever que o peso do recém-nascido diminui se a mulher fizer mais de 22 consultas para exame pré-natal? Explique.
- (iv) Adicione a idade da mãe à equação, usando uma forma funcional quadrática. Mantendo a variável *conspn* fixa, com que idade da mãe o peso do recém-nascido é maximizado? Que fração de mulheres na amostra é de mulheres mais velhas que a idade “ótima”?
- (v) Você diria que a idade da mãe e o número de consultas para exame pré-natal explicam muito da variação em  $\log(\text{pesonas})$ ?
- (vi) Usando termos quadráticos tanto para *conspn* como para *idade*, decida qual a melhor forma para prever *pesonas* – o uso do log natural ou em nível da variável *pesonas*.

## CAPÍTULO 7

---

**7.9** Utilize os dados contidos no arquivo GPA1.RAW para fazer este exercício.

- (i) Adicione as variáveis *maesup* e *paisup* à equação estimada em (7.6) e descreva os resultados da forma habitual. O que acontece com o efeito estimado da propriedade de um computador? A variável *PC* ainda é estatisticamente significativa?
- (ii) Teste a significância conjunta de *maesup* e *paisup* na equação da parte (i) e assegure-se de descrever o *p*-valor.
- (iii) Adicione  $nmem^2$  ao modelo da parte (i) e decida se essa generalização é necessária.

**7.10** Utilize os dados contidos no arquivo WAGE2.RAW para fazer este exercício.

- (i) Estime o modelo

$$\log(\text{salário}) = \beta_0 + \beta_1educ + \beta_2exper + \beta_3perm + \beta_4casado + \beta_5negro + \beta_6sul + \beta_7urbano + u$$

e descreva os resultados da forma habitual. Mantendo fixos outros fatores, qual é a diferença aproximada no salário mensal entre negros e não-negros? Essa diferença é estatisticamente significativa?

- (ii) Adicione as variáveis  $exper^2$  e  $perm^2$  à equação e demonstre que elas são conjuntamente não significantes mesmo ao nível de 20%.
- (iii) Estenda o modelo original para permitir que o retorno da educação dependa da raça e teste se o retorno da educação realmente depende da raça.
- (iv) Novamente, comece com o modelo original, mas agora permita que salários difiram entre quatro grupos de pessoas: casado e negro, casado e não-negro, solteiro e negro, e solteiro e não-negro. Qual é o diferencial de salários estimado entre negros casados e não-negros casados?

**7.11** Um modelo que permite que os salários dos jogadores de beisebol da liga principal dos EUA difiram pela posição é

$$\log(\text{salário}) = \beta_0 + \beta_1anos + \beta_2jogosano + \beta_3rebmed + \beta_4hrunano + \beta_5rebrunano + \beta_6runsano + \beta_7perccap + \beta_8porcest + \beta_9pribase + \beta_{10}segbase + \beta_{11}terbase + \beta_{12}interbase + \beta_{13}receptor + u,$$

onde os defensores externos seriam representativos do grupo base.

- (i) Declare a hipótese nula de que, controlando outros fatores, receptores e defensores externos ganhem, em média, a mesma quantia. Teste essa hipótese usando os dados contidos no arquivo MLB1.RAW e comente sobre o tamanho do diferencial salarial estimado.
- (ii) Declare e teste a hipótese nula de que não existe diferença no salário médio entre as posições quando outros fatores são controlados.

- (iii) Os resultados das partes (i) e (ii) são coerentes? Se não forem, explique o que está acontecendo.

**7.12** Utilize os dados contidos no arquivo GPA2.RAW para fazer este exercício.

- (i) Considere a equação

$$nmgrad = \beta_0 + \beta_1 tamclas + \beta_2 tamclas^2 + \beta_3 emperc + \beta_4 sat + \beta_5 feminino + \beta_6 atleta + u,$$

onde  $nmgrad$  é a nota média acumulada no curso superior,  $tamclas$  é o tamanho da classe no ensino médio, em centenas,  $emperc$  é o percentil na turma de formados no ensino médio,  $sat$  é a nota combinada do teste de aptidão acadêmica,  $feminino$  é uma variável binária de gênero e  $atleta$  é uma variável binária igual a um para alunos atletas. Quais são suas expectativas quanto aos coeficientes nessa equação? Para quais deles você está incerto?

- (ii) Estime a equação da parte (i) e descreva os resultados da forma habitual. Qual é o diferencial da nota média de graduação estimado entre atletas e não-atletas? Ele é estatisticamente significativo?
- (iii) Remova  $sat$  do modelo e reestime a equação. Agora, qual é o efeito estimado de ser um atleta? Discuta por que a estimativa é diferente da obtida na parte (ii).
- (iv) No modelo da parte (i), permita que o efeito de ser um atleta difira por gênero e teste a hipótese nula de que não existe diferenças *ceteris paribus* entre mulheres atletas e mulheres não-atletas.
- (v) O efeito de  $sat$  sobre  $nmgrad$  difere por gênero? Justifique sua resposta.

**7.13** No Problema 4.2, adicionamos o retorno da ação da empresa,  $raf$ , a um modelo que explicava o salário dos diretores executivos; constatou-se que  $raf$  não era significativo. Agora, defina uma variável *dummy*,  $rafneg$ , que será igual a um se  $raf < 0$  e igual a zero se  $raf \geq 0$ . Utilize os dados contidos no arquivo CEOSAL1.RAW para estimar o modelo

$$\log(\text{salário}) = \beta_0 + \beta_1 \log(\text{vendas}) + \beta_2 rma + \beta_3 rafneg + u.$$

Discuta a interpretação e a significância estatística de  $\hat{\beta}_3$ .

**7.14** Utilize os dados contidos no arquivo SLEEP75.RAW para fazer este exercício. A equação de interesse é

$$dormir = \beta_0 + \beta_1 trabtot + \beta_2 educ + \beta_3 idade + \beta_4 idade^2 + \beta_5 crianmen + u.$$

- (i) Estime essa equação separadamente para homens e mulheres e descreva os resultados na forma habitual. Existem diferenças notáveis nas estimativas das duas equações?
- (ii) Calcule o teste de Chow quanto à igualdade dos parâmetros na equação de interesse, para homens e mulheres. Use a forma de teste que adiciona  $masculino$  e os termos de interação  $masculino \cdot trabtot$ , ...,  $masculino \cdot crianmen$ , usando o conjunto total de observações. Quais são os  $gl$  relevantes do teste? Você deve rejeitar a hipótese nula ao nível de 5%?

- (iii) Agora, permita um intercepto diferente para homens e mulheres e determine se os termos de interação que envolvem *masculino* são conjuntamente significativos.
- (iv) Dados os resultados das partes (ii) e (iii), qual seria seu modelo final?

**7.15** Utilize os dados contidos no arquivo WAGE1.RAW para fazer este exercício.

- (i) Use a equação (7.18) para estimar o diferencial de gênero quando  $educ = 12,5$ . Compare com o diferencial estimado quando  $educ = 0$ .
- (ii) Compute a regressão usada para obter (7.18), mas com  $feminino \cdot (educ - 12,5)$  substituindo  $feminino \cdot educ$ . Como você interpreta o coeficiente de *feminino* nesse caso?
- (ii) O coeficiente de *feminino* na parte (ii) é estatisticamente significativo? Compare com a equação (7.18) e comente.

**7.16** Utilize os dados contidos no arquivo LOANAPP.RAW para fazer este exercício. A variável binária a ser explicada é *aprovado*, que será igual a um se um empréstimo hipotecário a um indivíduo foi aprovado. A variável explicativa importante é *branco*, uma variável *dummy* igual a um se o pretendente for branco. Os outros pretendentes no conjunto de dados são negros e hispânicos.

Para verificar se existe discriminação no mercado de empréstimos hipotecários, um modelo de probabilidade linear pode ser usado:

$$aprovado = \beta_0 + \beta_1 branco + outros\ fatores.$$

- (i) Se houver discriminação contra as minorias, e os fatores apropriados tiverem sido controlados, qual será o sinal de  $\beta_1$ ?
- (ii) Faça a regressão de *aprovado* sobre *branco* e descreva os resultados na forma habitual. Interprete o coeficiente de *branco*. Ele é estatisticamente significativo? É grande, do ponto de vista prático?
- (iii) Como controles, adicione as variáveis *gastdom*, *outrobr*, *montempr*, *desemp*, *masculino*, *casado*, *dep*, *est*, *aval*, *chist*, *falid*, *inadinp1*, *inadinp2*, e *vr*. O que acontece com o coeficiente de *branco*? Ainda existe evidência de discriminação contra os não-brancos?
- (iv) Agora, permita que o efeito da raça interaja com a variável que registra outras obrigações como uma porcentagem da renda (*outrobr*). O termo de interação é significativo?
- (v) Usando o modelo da parte (iv), qual é o efeito de ser branco sobre a probabilidade de aprovação quando  $outrobr = 32$ , que é aproximadamente o valor da mediana na amostra? Obtenha um intervalo de confiança de 95% desse efeito.

**7.17** Tem havido bastante interesse em saber se os planos de pensão ao alcance de muitos trabalhadores norte-americanos aumentam a poupança líquida. O conjunto de dados do arquivo 401KSUBS.RAW contém informações sobre os ativos financeiros líquidos (*finliq*), renda familiar (*renda*), uma variável binária para a qualificação para um plano de pensão (*e401k*) e várias outras variáveis.

- (i) Que fração das famílias na amostra se classifica para participar de uma conta de um fundo de pensão?

- (ii) Estime um modelo de probabilidade linear explicando a qualificação para ter uma conta nos fundos de pensão, em termos de renda, idade e gênero. Inclua renda e idade na forma quadrática, e descreva os resultados na forma habitual.
- (iii) Você diria que a qualificação para ter uma conta em um fundo de pensão é independente da renda e da idade? E quanto ao gênero? Explique.
- (iv) Obtenha os valores previstos do modelo de probabilidade linear estimado na parte (ii). Algum dos valores estimados é negativo ou maior que um?
- (v) Adicione a variável *plapind* como uma variável explicativa do modelo de probabilidade linear. Supondo todos os outros fatores iguais, se uma família tiver alguém com um plano de aposentadoria individual, o quanto será mais elevada a probabilidade estimada de que a família será qualificada para um plano de pensão? Ela é estatisticamente diferente de zero ao nível de 10%?

**7.18** Utilize os dados contidos no arquivo NBASAL.RAW para fazer este exercício.

- (i) Estime um modelo de regressão linear relacionando pontos por jogo à experiência na liga de basquetebol e à posição (armador, ala, ou pivô). Inclua a experiência na forma quadrática e use pivôs como o grupo base. Descreva os resultados na forma habitual.
- (ii) Por que você não inclui todas as três variáveis simuladas das posições na parte (i)?
- (iii) Mantendo fixa a experiência, um armador marca mais pontos que um pivô? Quantos mais? A diferença é estatisticamente significativa?
- (iv) Agora, adicione o estado civil à equação. Mantendo fixas a posição e a experiência, os jogadores casados são mais produtivos (com base nos pontos por jogo)?
- (v) Adicione interações do estado civil com ambas as variáveis de experiência. Nesse modelo expandido, existe evidência forte de que o estado civil afeta os pontos por jogo?
- (iv) Estime o modelo da parte (iv), mas use assistências por jogo como a variável dependente. Existe alguma diferença notável em relação à parte (iv)? Discuta.

**7.19** Utilize os dados contidos no arquivo 401KSUBS.RAW para fazer este exercício.

- (i) Calcule a média, o desvio-padrão e os valores mínimos e máximos de *finliq* na amostra.
- (ii) Teste a hipótese de que a média de *finliq* não difere pelo *status* da qualificação para ter uma conta em plano de pensão; use uma alternativa bilateral. Qual é o montante em dólares da diferença estimada?
- (iii) Considerando a parte (ii) do Exercício 7.17, fica claro que a variável *e401k* não é exógena em um modelo de regressão simples; no mínimo, ela muda com a renda e a idade. Estime um modelo de regressão linear múltipla de *finliq* que inclua renda, idade, gênero e *e401k* como variáveis explicativas. As variáveis de renda e idade devem aparecer na forma quadrática. Nesse caso, qual é o efeito estimado, em dólares, da qualificação para ter uma conta em um fundo de pensão?
- (iv) Ao modelo estimado na parte (iii), adicione as interações  $e401k \cdot (idade - 41)$  e  $e401k \cdot (idade - 41)^2$ . Observe que a média de idade na amostra está em torno de 41, de forma que no novo modelo o coeficiente de *e401k* é o efeito estimado da qualificação para ter uma conta em um fundo de pensão na idade média. Qual termo de interação é significativo?

- (v) Comparando as estimativas das partes (iii) e (iv), os efeitos estimados da qualificação para ter uma conta em um fundo de pensão na idade de 41 anos diferem muito? Explique.
- (vi) Agora, elimine os termos de interação do modelo, mas defina cinco variáveis *dummy* do tamanho da família:  $tamfam1$ ,  $tamfam2$ ,  $tamfam3$ ,  $tamfam4$  e  $tamfam5$ . A variável  $tamfam5$  é igual a um para famílias compostas de cinco ou mais membros. Inclua as variáveis *dummy* do tamanho da família no modelo estimado na parte (iii); assegure-se de escolher um grupo base. As *dummies* do tamanho das famílias são significativas ao nível de 1%?
- (vii) Agora, faça um teste de Chow do modelo

$$finliq = \beta_0 + \beta_1renda + \beta_2renda^2 + \beta_3idade + \beta_4idade^2 + \beta_5e401k + u$$

entre as cinco categorias de tamanhos de família, permitindo diferenças nos interceptos. A soma dos resíduos quadrados restrita,  $SQR_r$ , é obtida da parte (vi) porque aquela regressão assume que todas as inclinações são as mesmas. A soma dos resíduos quadrados irrestrita é  $SQR_{ir} = SQR_1 + SQR_2 + \dots + SQR_5$ , onde  $SQR_f$  é a soma dos resíduos quadrados da equação estimada usando somente o tamanho da família  $f$ . Você deve se convencer de que existem 30 parâmetros no modelo irrestrito (cinco interceptos mais 25 inclinações) e dez parâmetros no modelo restrito (cinco interceptos mais cinco inclinações). Portanto, o número de restrições testadas é  $q = 20$ , e os  $gl$  do modelo irrestrito são  $9.275 - 30 = 9.245$ .

## CAPÍTULO 8

**8.6** Utilize os dados contidos no arquivo SLEEP75.RAW para estimar a seguinte equação para a variável *dormir*:

$$dormir = \beta_0 + \beta_1trabtot + \beta_2educ + \beta_3idade + \beta_4idade^2 + \beta_5crianmen + \beta_6masculino + u.$$

- (i) Escreva um modelo que permita que a variância de  $u$  difira entre homens e mulheres. A variância não deve depender de outros fatores.
- (ii) Estime os parâmetros do modelo quanto à heteroscedasticidade. (Você terá, primeiro, que estimar a equação *dormir* por MQO, para obter os resíduos MQO.) A variância estimada de  $u$  é maior para homens ou para mulheres?
- (iii) A variância de  $u$  é estatisticamente diferente para homens e para mulheres?
- 8.7** (i) Utilize os dados contidos no arquivo HPRICE1.RAW para obter os erros-padrão robustos em relação à heteroscedasticidade da equação (8.17). Discorra sobre quaisquer diferenças importantes em relação aos erros padrão habituais.
- (ii) Repita a parte (i) para a equação (8.18).
- (iii) O que este exemplo sugere sobre a heteroscedasticidade e a transformação usada na variável dependente?

**8.8** Aplique o teste completo de White de heteroscedasticidade [veja equação (8.19)] na equação (8.18). Usando a forma qui-quadrada da estatística, obtenha o  $p$ -valor. Qual é sua conclusão?



**8.9** Utilize os dados contidos no arquivo VOTE1.RAW para fazer este exercício.

- (i) Estime um modelo com *votoA* como a variável dependente e *forpartA*, *democA*,  $\log(\text{gastoA})$  e  $\log(\text{gastoB})$  como as variáveis independentes. Obtenha os resíduos MQO,  $\hat{u}_i$ , e faça a regressão desses resíduos sobre todas as variáveis independentes. Explique por que você obtém  $R^2 = 0$ .
- (ii) Agora, compute o teste de heteroscedasticidade de Breusch-Pagan. Use a versão da estatística  $F$  e escreva o  $p$ -valor.
- (iii) Compute o caso especial do teste de White de heteroscedasticidade, usando novamente a forma da estatística  $F$ . Qual é a força da evidência de heteroscedasticidade agora?

**8.10** Utilize os dados contidos no arquivo PNTSPRD.RAW para fazer este exercício.

- (i) A variável *sprdcvr* é uma variável binária igual a um se a vantagem que uma casa de apostas der a uma equipe de basquetebol mais fraca for coberta em um determinado jogo. O valor esperado de *sprdcvr*,  $\mu$ , é a probabilidade de que a vantagem seja coberta em um jogo selecionado aleatoriamente. Teste  $H_0: \mu = 0,5$  contra  $H_1: \mu \neq 0,5$  no nível de significância de 10% e discuta os resultados. (*Sugestão*: Isso pode ser feito com facilidade usando um teste  $t$  fazendo a regressão de *sprdcvr* sobre somente um intercepto.)
- (ii) Quantos jogos na amostra de 553 foram jogados em campo neutro?
- (iii) Estime o modelo de probabilidade linear

$$\text{sprdcvr} = \beta_0 + \beta_1 \text{casafav} + \beta_2 \text{neutro} + \beta_3 \text{fav25} + \beta_4 \text{aza25} + u$$

e descreva os resultados da maneira habitual. (Registre os erros-padrão MQO usuais e os erros-padrão robustos em relação à heteroscedasticidade). Quais variáveis são mais significativas, tanto na prática como estatisticamente?

- (iv) Explique a razão, sob a hipótese nula  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ , de não haver heteroscedasticidade no modelo.
- (v) Use a estatística  $F$  habitual para testar a hipótese na parte (iv). Quais suas conclusões?
- (vi) Considerando a análise anterior, você diria ser possível prever sistematicamente se a vantagem concedida pela casa de apostas será coberta usando a informação disponível antes dos jogos?

**8.11** No Exemplo 7.12, estimamos um modelo de probabilidade linear para verificar se um jovem havia sido preso durante o ano de 1986:

$$\text{pris86} = \beta_0 + \beta_1 \text{pcond} + \beta_2 \text{sentmed} + \beta_3 \text{temptot} + \beta_4 \text{temp86} + \beta_5 \text{empr86} + u.$$

- (i) Estime este modelo por MQO e verifique se todos os valores estimados estão estritamente entre zero e um. Quais são o menor e o maior valores estimados?
- (ii) Estime a equação por mínimos quadrados ponderados, como discutido na Seção 8.5.
- (iii) Use as estimativas MQP para determinar se *sentmed* e *temptot* são conjuntamente significantes no nível de 5%.

**8.12** Utilize os dados contidos no arquivo LOANAPP.RAW para fazer este exercício.

- (i) Estime a equação da parte (iii) do Problema 7.16, computando os erros-padrão robustos em relação à heteroscedasticidade. Compare o intervalo de confiança de 95% em  $\beta_{branco}$  com o intervalo de confiança não robusto.
- (ii) Obtenha os valores estimados da regressão na parte (i). Algum deles é menor que zero? Algum deles é maior que um? O que isso significa em termos da aplicação de mínimos quadrados ponderados?

**8.13** Utilize os dados contidos no arquivo GPA1.RAW para fazer este exercício.

- (i) Use o método MQO para estimar um modelo relacionando *nmgrad* com *nmem*, *tac*, *faltas* e *PC*. Obtenha os resíduos MQO.
- (ii) Calcule o caso especial do teste de White de heteroscedasticidade. Na regressão de  $\hat{u}_i^2$  sobre  $nm\hat{grad}$ ,  $nm\hat{grad}^2$ , obtenha os valores estimados, digamos,  $\hat{h}_i$ .
- (iii) Verifique se os valores estimados da parte (ii) são todos estritamente positivos. Em seguida, obtenha as estimativas dos mínimos quadrados ponderados utilizando pesos  $1/\hat{h}_i$ . Compare as estimativas dos mínimos quadrados ponderados para o efeito das faltas às aulas e da propriedade de computadores pessoais com as estimativas MQO correspondentes. O que é possível concluir de suas significâncias estatísticas?
- (iv) Na estimação MQP da parte (iii), obtenha os erros-padrão robustos em relação à heteroscedasticidade. Em outras palavras, leve em conta o fato de que a função da variância estimada na parte (ii) pode ter sido mal especificada. (Veja Questão 8.4.) Os erros-padrão da parte (iii) mudam muito?

**8.14** No Exemplo 8.7, computamos as estimativas MQO e um conjunto de estimativas MQP em uma equação de demanda de cigarros.

- (i) Obtenha as estimativas MQO na equação (8.35).
- (ii) Obtenha  $\hat{h}_i$  usado na estimação MQP da equação (8.36) e reproduza a equação (8.36). Desta equação, obtenha os resíduos e valores estimados não ponderados; chame-os de  $\hat{u}_i$  e  $\hat{y}_i$ , respectivamente. (Por exemplo, no programa econométrico Stata, os resíduos e valores estimados não ponderados são fornecidos por padrão.)
- (iii) Sejam  $\check{u}_i = \hat{u}_i/\sqrt{\hat{h}_i}$  e  $\check{y}_i = \hat{y}_i/\sqrt{\hat{h}_i}$  as quantidades ponderadas. Aplique o caso especial do teste de White de heteroscedasticidade fazendo a regressão de  $\check{u}_i^2$  sobre  $\check{y}_i$ ,  $\check{y}_i^2$ , certificando-se de incluir um intercepto, como sempre. Você encontra heteroscedasticidade nos resíduos ponderados?
- (iv) O que as constatações da parte (iii) implicam sobre a forma de heteroscedasticidade proposta usada na obtenção de (8.36)?
- (v) Obtenha erros-padrão válidos para as estimativas MQP que permitam que a função de variância seja mal especificada.

- 8.15** Utilize os dados contidos no arquivo 401KSUBS.RAW para fazer este exercício.
- (i) Usando o método MQO, estime um modelo de probabilidade linear de *e401k*, utilizando como variáveis explicativas *renda*, *renda*<sup>2</sup>, *idade*, *idade*<sup>2</sup>, e *masculino*. Obtenha ambas as versões de erros-padrão, a habitual por MQO e a robusta em relação à heteroscedasticidade. Existe alguma diferença importante entre elas?
  - (ii) No caso especial do teste de White de heteroscedasticidade, no qual fazemos a regressão dos resíduos quadrados MQO sobre o quadrado dos valores estimados MQO,  $\hat{u}_i^2$  sobre  $\hat{y}_i, \hat{y}_i^2, i = 1, \dots, n$ , argumente que o limite de probabilidade do coeficiente de  $\hat{y}_1$  deve ser um, a probabilidade do coeficiente de  $\hat{y}_i^2$  deve ser  $-1$ , e o limite de probabilidade do intercepto deve ser zero. {*Sugestão*: Lembre-se que  $\text{Var}(y|x_1, x_2, \dots, x_k) = p(\mathbf{x})[1 - p(\mathbf{x})]$ , onde  $p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ .}
  - (iii) Do modelo estimado da parte (i), obtenha o teste de White e veja se as estimativas dos coeficientes correspondem, em linhas gerais, aos valores teóricos descritos na parte (ii).
  - (iv) Após ter verificado que os valores estimados da parte (i) estão todos entre zero e um, obtenha as estimativas dos mínimos quadrados ponderados do modelo de probabilidade linear. Elas diferem, de maneira importante, das estimativas MQO?

## CAPÍTULO 9

---

- 9.6**
- (i) Aplique o teste RESET da equação (9.3) no modelo estimado no Problema 7.13. Existe evidência de má especificação da forma funcional na equação?
  - (ii) Compute uma forma robusta do teste RESET, em relação à heteroscedasticidade. Isso altera a sua conclusão sobre a parte (i)?
- 9.7** Utilize o conjunto de dados do arquivo WAGE2.RAW para este exercício.
- (i) Use a variável *KWW* como uma *proxy* da aptidão em lugar da variável *QI* no Exemplo 9.3. Qual será a estimativa do retorno da educação neste caso?
  - (ii) Agora, use *QI* e *KWW* juntas como variáveis *proxy*. O que acontece com o retorno da educação estimado?
  - (iii) Na parte (ii), *QI* e *KWW* são individualmente significantes? Conjuntamente, elas são significantes?
- 9.8** Utilize os dados do arquivo JTRAIN.RAW para este exercício.
- (i) Considere o modelo de regressão simples

$$\log(\text{rejei}) = \beta_0 + \beta_1 \text{subs} + u,$$

onde *rejei* é o índice de rejeição dos produtos das firmas e *subs* é uma variável *dummy* indicando se uma determinada firma recebeu subsídios para treinamento de pessoal. Você consegue pensar em algumas razões pelas quais os fatores não-observados em *u* possam estar correlacionados com *subs*?

- (ii) Estime o modelo de regressão simples utilizando os dados de 1988. (Você deve ter 54 observações.) O recebimento de subsídios para treinamento de pessoal reduz significativamente a taxa de rejeição das firmas?
- (iii) Agora, adicione como uma variável explicativa  $\log(\text{rejei}_{87})$ . Como isso altera o efeito estimado de *subs*? Interprete o coeficiente de *subs*. Ele é estatisticamente significativo ao nível de 5%, comparado com a alternativa unilateral  $H_1: \beta_{\text{subs}} < 0$ ?
- (iv) Teste a hipótese nula de que o parâmetro de  $\log(\text{rejei}_{87})$  é igual a um, comparado com a alternativa bilateral. Informe o *p*-valor do teste.
- (v) Repita as partes (iii) e (iv), utilizando erros-padrão robustos em relação à heteroscedasticidade, e discuta, de forma resumida, sobre quaisquer diferenças encontradas.

**9.9** Utilize os dados para o ano de 1990 do arquivo INFMRT.RAW para este exercício.

- (i) Reestime a equação (9.37), mas agora inclua uma variável *dummy* para a observação do Distrito de Colúmbia (chamada *DC*). Interprete o coeficiente de *DC* e comente sobre seu tamanho e significância.
- (ii) Compare as estimativas e erros-padrão da parte (i) com as da equação (9.38). O que você conclui sobre a inclusão de uma variável *dummy* para uma única observação?

**9.10** Utilize os dados do arquivo RDCHEM.RAW para melhor examinar os efeitos de observações extremas sobre as estimativas MQO. Em particular, estime o modelo

$$pdintens = \beta_0 + \beta_1 \text{vendas} + \beta_2 \text{vendas}^2 + \beta_3 \text{luclrmarg} + u$$

com e sem a empresa que tem vendas anuais de quase 40 bilhões de dólares e discuta se os resultados diferem em aspectos importantes. As equações serão mais fáceis de serem lidas se você redefinir *vendas* para ser medida em bilhões de dólares antes de prosseguir (veja o Problema 6.3).

**9.11** Refaça o Exemplo 4.10 removendo as escolas em que os benefícios dos professores sejam menores do que 1% do salário.

- (i) Quantas observações são perdidas?
- (ii) A remoção dessas observações produz algum efeito importante na relação de trocas estimada?

**9.12** Utilize os dados do arquivo LOANAPP.RAW para este exercício.

- (i) Quantas observações têm *outrobr* > 40, isto é, outras dívidas maiores que 40% da renda total?
- (ii) Reestime o modelo da parte (iii) do Exercício 7.16, excluindo as observações com *outrobr* > 40. O que acontece com a estimativa e com a estatística *t* da variável *branco*?
- (iii) A estimativa de  $\beta_{\text{branco}}$  aparenta ser demasiadamente afetada pela amostra utilizada?

**9.13** Utilize os dados do arquivo TWOYEAR.RAW para este exercício.

- (i) A variável *stotal* é uma variável de teste padronizada, que pode agir como uma variável *proxy* da aptidão não-observada. Encontre a média amostral e o desvio-padrão de *stotal*.

- (ii) Compute regressões simples de *cp* e de *univ* sobre *stotal*. As duas variáveis sobre educação superior são, estatisticamente, relacionadas com *stotal*? Explique.
- (iii) Adicione *stotal* à equação (4.17) e teste as hipóteses de que os retornos dos cursos profissionalizantes de dois anos e dos cursos de graduação de quatro anos são os mesmos, contra a hipótese alternativa de que o retorno dos cursos de graduação de quatro anos é maior. Como suas constatações podem ser comparadas com as da Seção 4.4?
- (iv) Adicione *stotal*<sup>2</sup> à equação estimada na parte (iii). Parece ser necessário um termo quadrático na variável de teste da pontuação?
- (v) Adicione os termos de interação *stotal*·*cp* e *stotal*·*univ* à equação da parte (iii). Esses termos são conjuntamente significantes?
- (vi) Qual seria seu modelo final que controlaria a aptidão por meio do uso de *stotal*? Justifique sua resposta.

**9.14** Neste exercício você deve comparar as estimativas MQO e MDA dos efeitos da qualificação para os planos de pensão dos EUA sobre os ativos financeiros líquidos. O modelo é

$$finliq = \beta_0 + \beta_1renda + \beta_2renda^2 + \beta_3idade + \beta_4idade^2 + \beta_5masculino + \beta_6e401k + u.$$

- (i) Utilize os dados do arquivo 401KSUBS.RAW para estimar a equação por MQO e descreva os resultados da forma usual. Interprete o coeficiente de *e401k*.
- (ii) Use os resíduos de MQO para testar a heteroscedasticidade utilizando o teste de Breusch-Pagan. A variável *u* é independente das variáveis explicativas?
- (iii) Estime a equação com o método MDA e descreva os resultados da mesma forma como foram informados os resultados do MQO. Interprete a estimativa MDA de  $\beta_6$ .
- (iv) Reconcilie suas constatações das partes (ii) e (iii)

## CAPÍTULO 10

---

**10.7** Em outubro de 1979, o Banco Central norte-americano mudou sua política de metas de oferta monetária e passou a focar diretamente as taxas de juros de curto prazo. Utilizando os dados contidos no arquivo INTDEF.RAW, defina uma variável *dummy* igual a um para os anos após 1979. Inclua essa *dummy* na equação (10.15) para verificar se ocorre uma mudança na equação da taxa de juros após 1979. O que você conclui?

**10.8** Utilize os dados contidos no arquivo BARIUM.RAW para fazer este exercício.

- (i) Adicione uma tendência temporal linear à equação (10.22). Algumas das variáveis, além daquela da tendência, são estatisticamente significantes?
- (ii) Na equação estimada na parte (i) verifique a significância conjunta de todas as variáveis, exceto a de tendência temporal. Qual é sua conclusão?
- (iii) Adicione variáveis *dummy* mensais a essa equação e teste a sazonalidade. A inclusão das *dummies* mensais altera quaisquer outras estimativas ou seus erros-padrão de forma importante?

**10.9** Adicione a variável  $\log(prpnb)$  na equação do salário mínimo (10.38). Essa variável é significativa? Interprete o coeficiente. Como a inclusão de  $\log(prpnb)$  afeta o efeito do salário mínimo estimado?

**10.10** Utilize os dados contidos no arquivo FERTIL3.RAW para verificar que o erro-padrão da PLP na equação (10.19) está em torno de 0,030.

**10.11** Utilize os dados contidos no arquivo EZANDERS.RAW para fazer este exercício. Os dados são de pedidos de auxílio-desemprego mensais na cidade de Anderson, no estado norte-americano de Indiana, de janeiro de 1980 a novembro de 1988. Em 1984, uma zona industrial (ZI) foi instalada em Anderson (como também em outras cidades de Indiana). [Veja Papke (1994) para detalhes.]

- (i) Regrida  $\log(uclms)$  sobre uma tendência temporal linear e sobre 11 variáveis *dummy* mensais. Qual é a tendência geral dos pedidos de auxílio-desemprego ao longo desse período? (Interprete o coeficiente da tendência temporal.) Existe evidência de sazonalidade nos pedidos de auxílio-desemprego?
- (ii) Adicione  $zi$ , uma variável *dummy* igual a um nos meses em que Anderson tinha uma ZI, na regressão da parte (i). O fato de ter uma zona industrial parece reduzir os pedidos de auxílio-desemprego? Em quanto? [Você deve usar a fórmula (7.10) do Capítulo 7].
- (iii) Que hipóteses você precisa fazer para atribuir o efeito da parte (ii) à criação de uma zona industrial?

**10.12** Utilize os dados contidos no arquivo FERTIL3.RAW para fazer este exercício.

- (i) Faça a regressão de  $tgf_t$  sobre  $t$  e  $t^2$  e guarde os resíduos. Isso produzirá um  $tgf_t$  cuja tendência foi removida, digamos  $t\check{g}f_t$ .
- (ii) Regrida  $t\check{g}f_t$  sobre todas as variáveis da equação (10.35), inclusive  $t$  e  $t^2$ . Compare o  $R$ -quadrado com o de (10.35). O que você conclui?
- (iii) Reestime a equação (10.35) mas adicione  $t^3$  a ela. Este termo adicional é estatisticamente significativo?

**10.13** Utilize os dados contidos no arquivo CONSUMP.RAW para fazer este exercício.

- (i) Estime um modelo de regressão simples relacionando o crescimento do consumo real *per capita* (de bens não duráveis e serviços) ao crescimento da renda disponível real *per capita*. Use a mudança nos logaritmos em ambos os casos. Descreva os resultados da forma habitual. Interprete a equação e discuta a significância estatística.
- (ii) Adicione uma defasagem do crescimento da renda disponível real *per capita* da parte (i). Qual sua conclusão sobre as defasagens de ajustamento sobre o crescimento do consumo?
- (iii) Adicione a taxa real de juros à equação da parte (i). Ela afeta o crescimento do consumo?

**10.14** Utilize os dados contidos no arquivo FERTIL3.RAW para fazer este exercício.

- (i) Adicione  $ip_{t-3}$  e  $ip_{t-4}$  à equação (10.19). Faça o teste de significância conjunta dessas defasagens.
- (ii) Encontre a propensão de longo prazo estimada e seus respectivos erros-padrão no modelo da parte (i). Compare-os com os obtidos na equação (10.19).
- (iii) Estime o modelo de defasagens distribuídas polinomial do Problema 10.6. Encontre a PLP e compare-a com o que é obtido do modelo sem restrições.

**10.15** Utilize os dados contidos no arquivo VOLAT.RAW para fazer este exercício. A variável  $rsp500$  é o retorno mensal do índice Standard & Poors do mercado acionário norte-americano, em

termos anuais. (A variável inclui tanto alterações de preços como também dividendos.) A variável  $i3$  é o retorno das letras do Tesouro norte-americano de três meses, e  $pcip$  é a mudança percentual na produção industrial; ambas também estão expressas em termos anuais.

- (i) Considere a equação:

$$rsp500_t = \beta_0 + \beta_1 pcip_t + \beta_2 i3_t + u_t$$

Que sinais você acredita que  $\beta_1$  e  $\beta_2$  deveriam ter?

- (ii) Estime a equação anterior pelo método MQO, descrevendo os resultados na forma padrão. Interprete os sinais e as magnitudes dos coeficientes.  
 (iii) Quais das variáveis são estatisticamente significantes?  
 (iv) Suas descobertas da parte (iii) sugerem que o retorno do S&P 500 é previsível? Explique.

**10.16** Considere o modelo estimado em (10.15); utilize os dados contidos no arquivo INT-DEF.RAW.

- (i) Encontre a correlação entre  $inf$  e  $def$  ao longo do período amostral e comente.  
 (ii) Adicione uma defasagem em  $inf$  e  $def$  na equação e descreva os resultados da forma habitual.  
 (iii) Compare a PLP estimada do efeito da inflação com aquela da equação (10.15). Elas são muito diferentes?  
 (iv) As duas defasagens no modelo são conjuntamente significantes ao nível de 5%?

**10.17** O arquivo TRAFIC2.RAW contém 108 observações mensais sobre acidentes automobilísticos, leis de trânsito e algumas outras variáveis do estado norte-americano da Califórnia, de janeiro de 1981 a dezembro de 1989. Utilize esse conjunto de dados para responder as seguintes questões.

- (i) Em que mês e ano a lei do cinto de segurança entrou em vigência na Califórnia? Quando o limite de velocidade nas rodovias passou para 65 milhas por hora?  
 (ii) Faça a regressão da variável  $\log(totacc)$  em uma tendência temporal linear e 11 variáveis *dummy* mensais, usando janeiro como o mês base. Interprete a estimativa do coeficiente da tendência temporal. Você diria que existe sazonalidade no total de acidentes?  
 (iii) Adicione à regressão da parte (i) as variáveis  $finsem$ ,  $desemp$ ,  $leiveloc$ , e  $leicinto$ . Discuta o coeficiente da variável desemprego. Seu sinal e magnitude têm lógica para você?  
 (iv) Na regressão da parte (iii), interprete os coeficientes de  $leiveloc$  e  $leicinto$ . Os efeitos estimados são os que você esperava? Explique.  
 (v) A variável  $prcfat$  é a porcentagem de acidentes resultando em pelo menos uma fatalidade. Observe que esta variável é uma porcentagem e não uma proporção. Qual é a média da  $prcfat$  ao longo desse período? Sua magnitude parece correta?  
 (vi) Compute a regressão na parte (iii) mas use  $prcfat$  como a variável dependente em lugar de  $\log(totacc)$ . Discuta os efeitos e significância estimados das variáveis referentes às leis da velocidade e do cinto de segurança.

## CAPÍTULO 11

**11.8** Utilize os dados contidos no arquivo HSEINV.RAW para este exercício.

- (i) Encontre a autocorrelação de primeira ordem de  $\log(invpc)$ . Agora, encontre a autocorrelação após a remoção linear da tendência de  $\log(invpc)$ . Faça o mesmo para  $\log(preço)$ . Qual das duas séries pode ter uma raiz unitária?
- (ii) Com base nas suas constatações na parte (i), estime a equação

$$\log(invpc_t) = \beta_0 + \beta_1 \Delta \log(preço_t) + \beta_2 t + u_t$$

e descreva os resultados na forma padrão. Interprete o coeficiente  $\hat{\beta}_1$  e determine se ele é estatisticamente significativo.

- (iii) Retire linearmente a tendência de  $\log(invpc)$  e use a versão sem tendência como variável dependente na regressão da parte (ii) (veja a Seção 10.5). O que acontece com  $R^2$ ?
- (iv) Agora use  $\Delta \log(invpc_t)$  como variável dependente. Como seus resultados da parte (ii) mudam? A tendência temporal ainda é significativa? Por que?

**11.9** No Exemplo 11.7 defina o crescimento no salário por hora e na produção por hora como a mudança no log natural:  $gsalhr = \Delta \log(salhr)$  e  $gprodhr = \Delta \log(prodhr)$ . Considere uma extensão simples do modelo estimado em (11.29):

$$gsalhr_t = \beta_0 + \beta_1 gprodhr_t + \beta_2 gprodhr_{t-1} + u_t$$

Isto permite que um aumento no crescimento da produtividade tenha tanto um efeito corrente como um efeito defasado sobre o crescimento do salário.

- (i) Estime a equação utilizando os dados contidos no arquivo EARN.SRAW e descreva os resultados na forma padrão. O valor defasado de  $gprodhr$  é estatisticamente significativo?
- (ii) Se  $\beta_1 + \beta_2 = 1$ , um aumento permanente no crescimento da produtividade é totalmente transmitido ao crescimento dos salários mais altos após um ano. Teste  $H_0: \beta_1 + \beta_2 = 1$  contra a alternativa bilateral. Lembre-se, a maneira mais fácil de fazer isso é escrever a equação de forma que  $\theta = \beta_1 + \beta_2$  apareça diretamente no modelo, como no Exemplo 10.4 do Capítulo 10.
- (iii) É necessário que  $gprodhr_{t-2}$  esteja no modelo? Explique.

**11.10** (i) No exemplo 11.4, é possível que o valor esperado do retorno no tempo  $t$ , dados os retornos passados, seja uma função quadrática de  $retorno_{t-1}$ . Para verificar essa possibilidade, utilize os dados contidos no arquivo NYSE.RAW para estimar

$$retorno_t = \beta_0 + \beta_1 retorno_{t-1} + \beta_2 retorno_{t-1}^2 + u_t$$

descreva os resultados na forma padrão.

- (ii) Defina e teste a hipótese nula de que  $E(retorno_t | retorno_{t-1})$  não depende de  $retorno_{t-1}$ . (Sugestão: Aqui existem duas restrições a serem testadas.) O que você conclui?
- (iii) Elimine  $retorno_{t-1}^2$  do modelo, mas adicione o termo de interação  $retorno_{t-1} \cdot retorno_{t-2}$ . Agora, teste a hipótese de mercados eficientes.
- (iv) O que você conclui sobre a previsão de retornos semanais de ações com base nos retornos passados?



**11.11** Utilize os dados contidos no arquivo PHILLIPS.RAW para este exercício.

- (i) No exemplo 11.5, assumimos que a taxa natural de desemprego é constante. Uma forma alternativa da curva de Phillips de expectativas aumentadas permite que a taxa natural de desemprego dependa de níveis de desemprego passados. No caso mais simples, a taxa natural no tempo  $t$  é igual a  $desemp_{t-1}$ . Se assumirmos expectativas adaptativas, obtemos uma curva de Phillips na qual inflação e desemprego estão em primeiras diferenças:

$$\Delta inf = \beta_0 + \beta_1 \Delta desemp + u.$$

Estime esse modelo, descreva os resultados da maneira usual e discuta o sinal, tamanho e significância estatística de  $\hat{\beta}_1$ .

- (ii) Que modelo ajusta melhor os dados, (11.19) ou o modelo da parte (i)? Explique.

**11.12** (i) Adicione uma tendência temporal linear na equação (11.27). É necessária uma tendência temporal na equação de primeiras diferenças?

- (ii) Elimine a tendência temporal e adicione as variáveis  $ww2$  e  $pill$  na equação (11.27) (não diferencie essas variáveis *dummy*). Essas variáveis são conjuntamente significantes ao nível de 5%?

- (iii) Usando o modelo da parte (ii), estime a PLP e obtenha seu erro-padrão. Compare com (10.19), onde  $tgf$  e  $ip$  apareceram em níveis, em vez de primeiras diferenças.

**11.13** Seja  $inven_t$  o valor real dos estoques dos Estados Unidos durante o ano  $t$ , seja  $PIB_t$  o produto interno bruto real, e seja  $r3_t$  a taxa de juros real (*ex post*) das letras do Tesouro dos EUA de três meses. A taxa de juros real *ex post* é, aproximadamente,  $r3_t = i3_t - inf_t$ , onde  $i3_t$  é a taxa das letras do Tesouro de três meses e  $inf_t$  é a taxa anual de inflação [veja Mankiw (1994, Seção 6.4)]. A mudança nos estoques,  $\Delta inven_t$ , é o investimento em estoques no ano. O *modelo do acelerador* do investimento em estoques é

$$\Delta inven_t = \beta_0 + \beta_1 \Delta PIB_t + u_t,$$

onde  $\beta_1 > 0$ . [Veja, por exemplo, Mankiw (1994), Capítulo 17.]

- (i) Utilize os dados contidos no arquivo INVEN.RAW para estimar o modelo do acelerador. Descreva os resultados da maneira habitual e interprete a equação.  $\hat{\beta}_1$  é estatisticamente maior que zero?
- (ii) Se a taxa de juros real aumenta, então o custo de oportunidade de manter estoques aumenta, e assim um aumento na taxa de juros real deve fazer decrescer os estoques. Adicione a taxa de juros real ao modelo do acelerador e discuta os resultados.
- (iii) O nível da taxa de juros real funciona melhor que a primeira diferença,  $\Delta r3_t$ ?

**11.14** Utilize os dados contidos no arquivo CONSUMP.RAW para este exercício. Uma versão da *hipótese da renda permanente* (HRP) do consumo é que o *crescimento* do consumo é imprevisível. [Outra versão é que a mudança do consumo em si é imprevisível; veja Mankiw (1994, Capítulo 15) para uma discussão sobre a HRP.] Seja  $gc_t = \log(c_t) - \log(c_{t-1})$  o crescimento no consumo real *per capita* (de bens não-duráveis e serviços). Então, a HRP implica que  $E(gc_t | I_{t-1}) = E(gc_t)$ , onde  $I_{t-1}$  representa a informação conhecida no tempo  $(t - 1)$ ; neste caso,  $t$  representa um determinado ano.

- (i) Teste a HRP estimando  $gc_t = \beta_0 + \beta_1 gc_{t-1} + u_t$ . Especifique com clareza as hipóteses nula e alternativa. O que você conclui?

- (ii) Adicione  $gy_{t-1}$  e  $i3_{t-1}$  à regressão da parte (i), onde  $gy_t$  é o crescimento na renda disponível real *per capita* e  $i3_t$  são as taxas de juros das letras do Tesouro dos EUA de três meses; observe que cada uma delas deve ser defasada na regressão. Essas duas variáveis adicionais são conjuntamente significantes?

**11.15** Utilize os dados contidos no arquivo PHILLIPS.RAW para este exercício.

- (i) Estime um modelo AR(1) da taxa de desemprego. Use essa equação para prever a taxa de desemprego em 1997. Compare-a com a taxa real de desemprego de 1997. (Você pode encontrar essa informação em um volume recente do *Economic Report of the President*, (REP).)
- (ii) Adicione uma defasagem da inflação no modelo AR(1) da parte (i). A variável  $inf_{t-1}$  é estatisticamente significativa?
- (iii) Use a equação da parte (ii) para prever a taxa de desemprego em 1997. O resultado é melhor ou pior do que o do modelo da parte (i)?
- (iv) Use o método da Seção 6.4 para construir um intervalo de previsão de 95% para a taxa de desemprego de 1997. A taxa de desemprego de 1997 está dentro do intervalo?

**11.16** Utilize os dados contidos no arquivo TRAFFIC2.RAW para este exercício. O Exercício 10.17 já solicitou uma análise para estes dados.

- (i) Compute o coeficiente de autocorrelação de primeira ordem para a variável  $prcfat$ . Você está preocupado com a possibilidade de  $prcfat$  conter uma raiz unitária? Faça o mesmo para a taxa de desemprego.
- (ii) Estime um modelo de regressão múltipla relacionando a primeira diferença de  $prcfat$ ,  $\Delta prcfat$ , com as mesmas variáveis da parte (vi) do Exercício 10.17, exceto que você deve primeiro diferenciar a taxa de desemprego. Em seguida, inclua uma tendência temporal linear, variáveis *dummy* mensais, a variável de fim de semana e duas variáveis de política econômica: não as diferencie. Você encontra algum resultado interessante?
- (iii) Comente a seguinte declaração: “Devemos sempre primeiro diferenciar qualquer série temporal que suspeitemos possuir uma raiz unitária antes de computarmos uma regressão múltipla, pois esta é uma estratégia segura e deve produzir resultados semelhantes aos obtidos se usarmos os níveis.” [Para responder a isso, talvez você queira executar a regressão da parte (vi) do Exercício 10.17, se você ainda não o fez.]

**11.17** Usando os dados da Tabela 10.1 adicione os números da inflação e do desemprego dos anos de 1997, 1998, e 1999 aos dados contidos no arquivo PHILLIPS.RAW. Você deve ter agora 52 anos de informação.

- (i) Estime novamente a equação (11.19) e descreva os resultados da forma habitual. As estimativas de intercepto e de inclinação alteram-se de maneira notável quando você adiciona os dados dos três anos recentes?
- (ii) Obtenha uma nova estimativa da taxa natural de desemprego. Compare essa nova estimativa com aquela descrita no Exemplo 11.5.
- (iii) Compute a autocorrelação de primeira ordem de  $desemp$ . Em sua opinião, a raiz unitária está próxima de um?
- (iv) Use  $\Delta desemp$  como a variável explicativa em lugar de  $desemp$ . Qual das variáveis explicativas produz um *R*-quadrado mais alto?

## CAPÍTULO 12

**12.7** No Exemplo 11.6, estimamos um modelo de defasagens distribuídas finitas em primeiras diferenças:

$$\Delta tgf_t = \gamma_0 + \delta_0 \Delta ip_t + \delta_1 \Delta ip_{t-1} + \delta_2 \Delta ip_{t-2} + u_t$$

Utilize os dados contidos no arquivo FERTIL3.RAW para testar se existe correlação serial AR(1) nos erros.

- 12.8**
- (i) Utilizando os dados contidos no arquivo WAGEPRC.RAW, estime o modelo de defasagens distribuídas do Problema 11.5. Use a regressão (12.14) para testar a existência de correlação serial AR(1).
  - (ii) Reestime o modelo usando a estimação iterada de Cochrane-Orcutt. Qual é sua nova estimativa da propensão de longo prazo?
  - (iii) Usando a estimação CO iterada encontre o erro-padrão da PLP. (Isto exige que você estime uma equação modificada.) Determine se a PLP estimada é estatisticamente diferente de um ao nível de 5%.
- 12.9**
- (i) Na parte (i) do Problema 11.13, você foi solicitado a estimar o modelo do acelerador de investimentos em estoques. Teste essa equação quanto à presença de correlação serial AR(1).
  - (ii) Se você encontrar evidência de correlação serial, reestime a equação pelo método de Cochrane-Orcutt e compare os resultados.
- 12.10**
- (i) Utilize o arquivo NYSE.RAW para estimar a equação (12.48). Sejam  $\hat{h}_t$  os valores estimados dessa equação (as estimativas da variância condicional). Quantos  $\hat{h}_t$  são negativos?
  - (ii) Adicione  $\text{retorno}_{t-1}^2$  a (12.48) e, novamente, calcule os valores ajustados,  $\hat{h}_t$ . Algum dos  $\hat{h}_t$  é negativo?
  - (iii) Use os  $\hat{h}_t$  da parte (ii) para estimar (12.47) por mínimos quadrados ponderados (como na Seção 8.4). Compare sua estimativa de  $\beta_1$  com a da equação (11.16). Teste  $H_0: \beta_1 = 0$  e compare o resultado quando o MQO é usado.
  - (iv) Agora, estime (12.47) por MQP, usando o modelo ARCH estimado em (12.51) para obter  $\hat{h}_t$ . Essa mudança altera sus descobertas na parte (iii)?
- 12.11** Considere a versão do modelo de Fair no Exemplo 10.6. Agora, em lugar de prever a proporção dos votos recebida pelos Democratas, estime um modelo de probabilidade linear para verificar se os Democratas vencerão ou não.
- (i) Use a variável binária *demvence* em lugar de *demvoto* em (10.23) e descreva os resultados na forma padrão. Que fatores afetam a probabilidade de vencer? Utilize os dados somente até 1992.
  - (ii) Quantos valores estimados são menores que zero? Quantos são maiores que um?
  - (iii) Use a seguinte regra de previsão: se *demvence* > 0,5 os Democratas vencerão; de outra forma, os Republicanos vencerão. Usando essa regra, determine quantas das 20 eleições foram previstas corretamente pelo modelo.

- (iv) Insira os valores das variáveis explicativas de 1996. Qual é a probabilidade prevista de que Clinton venceria as eleições? Clinton venceu; você obteve a previsão correta?
- (v) Utilize um teste  $t$  robusto em relação à heteroscedasticidade para a correlação serial AR(1) nos erros. O que você encontra?
- (vi) Obtenha os erros-padrão robustos em relação à heteroscedasticidade das estimativas na parte (i). Existe alguma alteração notável em qualquer das estatísticas  $t$ ?

**12.12** (i) No Problema 10.13, você estimou uma relação simples entre o crescimento do consumo e o crescimento da renda disponível. Teste a equação para verificar a existência de correlação serial AR(1) (usando o arquivo CONSUMP.RAW).

- (ii) No Problema 11.14, você testou a hipótese da renda permanente regredindo o crescimento do consumo sobre uma defasagem. Após computar essa regressão, teste a presença de heteroscedasticidade regredindo os resíduos quadrados sobre  $gc_{t-1}$  e  $gc_{t-1}^2$ . Qual é sua conclusão?

**12.13** (i) Para o Exemplo 12.14, utilizando os dados contidos no arquivo BARIUM.RAW, obtenha as estimativas iteradas de Prais-Winsten.

- (ii) As estimativas de Prais-Winsten e de Cochrane-Orcutt são semelhantes? Você esperava que elas fossem?

**12.14** Utilize os dados contidos no arquivo TRAFFIC2.RAW para este exercício.

- (i) Compute uma regressão MQO de *prcfat* sobre uma tendência temporal linear, variáveis *dummy* mensais e sobre as variáveis *finsssem*, *desemp*, *leiveloc*, e *leicinto*. Teste os erros para verificar a presença de correlação serial AR(1), usando a regressão na equação (12.14). Faz sentido usar o teste que assume exogeneidade estrita dos regressores?
- (ii) Obtenha os erros-padrão robustos em relação à correlação serial e à heteroscedasticidade dos coeficientes de *leiveloc* e *leicinto*, usando quatro defasagens no estimador Newey-West. Como isso afeta a significância estatística das duas variáveis de política governamental?
- (iii) Agora, estime o modelo usando o método Prais-Winsten iterativo e compare as estimativas com as do MQO. Existe alguma alteração importante nos coeficientes das variáveis de políticas governamentais ou em suas significâncias estatísticas?

**12.15** O arquivo FISH.RAW contém 97 observações sobre preços e quantidades diárias de peixe no mercado de peixe de Fulton, em Manhattan. Use a variável  $\log(\text{premedio})$  como a variável dependente.

- (i) Regrida  $\log(\text{premedio})$  sobre quatro variáveis *dummy* diárias, tendo a sexta-feira como base. Inclua uma tendência temporal linear. Existe evidência de que o preço varie sistematicamente na semana?
- (ii) Agora, adicione as variáveis *onda2* e *onda3*, que são as medidas da altura das ondas nos últimos dias. Essas variáveis são, individualmente, significantes? Descreva um mecanismo pelo qual mares mais revoltos aumentariam o preço do peixe.
- (iii) O que aconteceu com a tendência temporal quando as variáveis *onda2* e *onda3* foram incluídas na regressão? O que deve estar acontecendo?

- (iv) Explique por que todas as variáveis explicativas na regressão são assumidas, com segurança, como estritamente exógenas.
- (v) Teste os erros para verificar a existência de correlação serial AR(1).
- (vi) Obtenha os erros-padrão Newey-West usando quatro defasagens. O que acontece com as estatísticas  $t$  de *onda2* e *onda3*? Você esperava uma mudança maior ou menor quando comparadas com as estatísticas  $t$  do MQO?
- (vii) Agora, obtenha as estimativas Prais-Winsten do modelo estimado na parte (ii). As variáveis *onda2* e *onda3* são, conjuntamente, estatisticamente significantes?

## CAPÍTULO 13

---

**13.7** Utilize os dados contidos no arquivo FERTIL1.RAW para fazer este exercício.

- (i) Na equação estimada no Exemplo 13.1, teste se o ambiente de vida na idade de 16 anos tem efeito sobre a fertilidade. (O grupo base é cidade grande.) Informe o valor da estatística  $F$  e do  $p$ -valor.
- (ii) Teste se a região do país na idade de 16 anos (sul é o grupo base) tem efeito sobre a fertilidade.
- (iii) Seja  $u$  o termo erro na equação populacional. Suponha que você entenda que a variância de  $u$  muda ao longo do tempo (mas não com *educ*, *idade* etc.). Um modelo que capta isso é

$$u^2 = \gamma_0 + \gamma_1 a74 + \gamma_2 a76 + \dots + \gamma_6 a84 + v.$$

Usando esse modelo, teste a existência de heteroscedasticidade em  $u$ . [Sugestão: seu teste  $F$  deve ter 6 e 1.122 graus de liberdade.]

- (iv) Adicione os termos de interação  $a74 \cdot educ$ ,  $a76 \cdot educ$ , ...,  $a84 \cdot educ$  ao modelo estimado na Tabela 13.1. Explique o que representam esses termos. Eles são conjuntamente significantes?

**13.8** Utilize os dados contidos no arquivo CPS78\_85.RAW para fazer este exercício.

- (i) Como você interpreta o coeficiente de  $a85$  na equação (13.2)? A interpretação sobre ele tem algum interesse? (Cuidado aqui; você deve considerar os termos de interação  $a85 \cdot educ$  e  $y85 \cdot feminino$ .)
- (ii) Mantendo todos os outros fatores fixos, qual é o aumento percentual estimado no salário nominal de um homem com 12 anos de escolaridade? Proponha uma regressão para obter um intervalo de confiança dessa estimativa. [Sugestão: Para obter o intervalo de confiança, substitua  $a85 \cdot educ$  por  $a85 \cdot (educ - 12)$ ; refira-se ao Exemplo 6.3.]
- (iii) Reestime a equação (13.2) mas permita que todos os salários sejam medidos em dólares de 1978. Particularmente, defina o salário real como  $rsalário = salário$  para 1978 e como  $rsalário = salário/1,65$  para 1985. Agora, use  $\log(rsalário)$  em lugar de  $\log(salário)$  para estimar (13.2). Quais coeficientes diferem daqueles da equação (13.2)?

- (iv) Explique a razão de o  $R$ -quadrado da sua regressão na parte (iii) não ser o mesmo da equação (13.2). (*Sugestão*: Os resíduos e, portanto, a soma dos resíduos quadrados das duas equações são idênticos.)
- (v) Descreva como a filiação sindical mudou de 1978 a 1985.
- (vi) Iniciando com a equação (13.2), teste se o diferencial dos salários dos trabalhadores sindicalizados mudou ao longo do tempo. (Isso dever ser um simples teste  $t$ .)
- (vii) Seus resultados na parte (v) são conflitantes com os da parte (vi)? Explique.

**13.9** Utilize os dados contidos no arquivo KIELMC.RAW para fazer este exercício.

- (i) A variável  $dist$  é a distância de cada imóvel do local do incinerador. Considere o modelo

$$\log(preço) = \beta_0 + \delta_0 a81 + \beta_1 \log(dist) + \delta_1 a81 \cdot \log(dist) + u.$$

Se a construção do incinerador reduz o valor dos imóveis mais próximos do local, qual é o sinal de  $\delta_1$ ? O que significa  $\beta_1 > 0$ ?

- (ii) Estime o modelo da parte (i) e descreva os resultados da forma habitual. Interprete o coeficiente de  $a81 \cdot \log(dist)$ . Qual sua conclusão?
- (iii) Adicione  $idade$ ,  $idade^2$ ,  $quartos$ ,  $\log(intst)$ ,  $\log(terreno)$ , e  $\log(área)$  à equação. Agora, qual é a sua conclusão sobre o efeito do incinerador sobre o valor dos imóveis?

**13.10** Utilize os dados contidos no arquivo INJURY.RAW para fazer este exercício.

- (i) Utilizando os dados no estado norte-americano de Kentucky, reestime a equação (13.12), adicionando, como variáveis explicativas,  $masculino$ ,  $casado$ , e um conjunto completo de variáveis  $dummy$  de tipos de empresas e de lesões. Como muda a estimativa de  $apmud \cdot altrend$  quando esses outros fatores são controlados? A estimativa continua estatisticamente significativa?
- (ii) O que você deduz do pequeno  $R$ -quadrado da parte (i)? Isso significa que a equação é inaproveitável?
- (iii) Estime a equação (13.12) utilizando os dados do estado norte-americano de Michigan. Compare as estimativas do termo de interação de Michigan e de Kentucky. A estimativa de Michigan é estatisticamente significativa? O que você deduz disso?

**13.11** Utilize os dados contidos no arquivo RENTAL.RAW para fazer este exercício. Os dados de 1980 e 1990 incluem preços de aluguéis e outras variáveis de cidades universitárias. A idéia é ver se uma forte presença de estudantes afeta os valores dos aluguéis. O modelo de efeitos não observados é

$$\log(alug_{it}) = \beta_0 + \delta_0 a90_t + \beta_1 \log(pop_{it}) + \beta_2 \log(rendfam_{it}) + \beta_3 pctestu_{it} + a_i + u_{it},$$

onde  $pop$  é a população da cidade,  $rendfam$  é a renda média e  $pctestu$  é a população estudantil como uma porcentagem da população da cidade (durante o ano letivo).

- (i) Estime a equação por MQO agrupado e descreva os resultados na forma padrão. O que você deduz da estimativa do coeficiente da variável  $dummy$  de 1990? O que você obtém para  $\hat{\beta}_{pctstu}$ ?

- (ii) Os erros-padrão que você descreveu na parte (i) são válidos? Explique.
- (iii) Agora, faça a diferenciação da equação e a estime por MQO. Compare sua estimativa da  $\beta_{pctstu}$  com a da parte (ii). O tamanho relativo da população estudantil parece afetar os preços dos aluguéis?
- (iv) Obtenha os erros-padrão robustos quanto à heteroscedasticidade da equação de primeiras diferenças na parte (iii). Isso altera suas conclusões?

**13.12** Utilize os dados contidos no arquivo CRIME3.RAW para fazer este exercício.

- (i) No modelo do Exemplo 13.6, teste a hipótese  $H_0: \beta_1 = \beta_2$ . (*Sugestão:* Defina  $\theta_1 = \beta_1 - \beta_2$  e escreva  $\beta_1$  em termos de  $\theta_1$  e  $\beta_2$ . Faça essa substituição na equação e reorganize. Faça um teste  $t$  de  $\theta_1$ .)
- (ii) Se  $\beta_1 = \beta_2$ , mostre que a equação diferenciada pode ser escrita como

$$\Delta \log(\text{crime}_i) = \delta_0 + \delta_1 \Delta \text{medescl}_i + \Delta u_i,$$

onde  $\delta_1 = 2\beta_1$  e  $\text{medescl}_i = (\text{pcescl}_{i,-1} + \text{pcescl}_{i,-2})/2$  é a média percentual de esclarecimentos ao longo dos dois anos anteriores.

- (iii) Estime a equação da parte (ii). Compare o  $R$ -quadrado ajustado com o de (13.22). Qual dos modelos você usaria?

**13.13** Utilize os dados contidos no arquivo GPA3.RAW para fazer este exercício. O conjunto de dados é de 366 estudantes atletas de uma grande universidade dos EUA, para dois semestres. [Uma análise semelhante está em Maloney e McCormick (1993), mas neste caso usamos um conjunto de dados de painel verdadeiro.] Como você tem dois semestres de dados de cada estudante, um modelo de efeitos não observados será apropriado. A questão primordial é esta: os atletas desempenham suas atividades escolares de forma menos efetiva durante a temporada de seu esporte?

- (i) Utilize o MQO agrupado para estimar um modelo com *nsgrad* como a variável dependente. As variáveis explicativas são *semestre1*, *sat*, *emperc*, *feminino*, *negro*, *branco*, *prisem*, *tothrs*, *npgrad*, e *temp*. Interprete o coeficiente de *temp*. Ele é estatisticamente significativo?
- (ii) A maioria dos atletas que praticam seus esportes somente no segundo semestre é de jogadores de futebol. Suponha que o nível de habilidade dos jogadores de futebol difira sistematicamente daquele dos outros atletas. Se a habilidade não for adequadamente capturada pela da pontuação *sat* e pelo percentil da turma de formados do ensino médio (*emperc*), explique por que os estimadores do MQO agrupado serão viesados.
- (iii) Agora, use os dados diferenciados ao longo dos dois semestres. Quais variáveis são eliminadas? Agora, faça um teste do efeito de ser temporada.
- (iv) Você consegue pensar em uma ou mais variáveis com variação temporal, potencialmente importantes, que tenham sido omitidas da análise?

**13.14** O arquivo VOTE2.RAW contém dados de painel das eleições para o Congresso norte-americano em 1988 e 1990. Somente os eleitos em 1988 e que estavam concorrendo à reeleição em 1990 estão na amostra; eles são os que têm mandato. Um modelo de efeitos não observados que explica a participação dos votos dos candidatos que já têm mandato, em termos de gastos por ambos os candidatos é



$$\text{votmand}_{it} = \beta_0 + \delta_0 d90_t + \beta_1 \log(\text{gastmand}_{it}) + \beta_2 \log(\text{gastdes}_{it}) + \beta_3 \text{pgasman}_{it} + a_i + u_{it},$$

onde  $\text{pgasman}_{it}$  é a participação do candidato à reeleição no total de gastos com a campanha (em forma percentual). O efeito não observado  $a_i$  contém características dos candidatos com mandato – tais como “qualidade” – além de informações sobre o distrito, que são constantes. O sexo do candidato e o partido são constantes ao longo do tempo, e portanto são incluídos em  $a_i$ . Estamos interessados no efeito dos gastos de campanha sobre os resultados das eleições.

- (i) Diferencie, ao longo dos dois anos, a equação dada e estime a equação diferenciada por MQO. Quais variáveis são, individualmente, significantes ao nível de 5%, contra uma alternativa bilateral?
- (ii) Na equação da parte (i), teste a significância conjunta de  $\Delta \log(\text{gastmand})$  e  $\Delta \log(\text{gastdes})$ . Informe o  $p$ -valor.
- (iii) Reestime a equação da parte (i) usando  $\Delta \text{pgasman}$  como a única variável independente. Interprete o coeficiente de  $\Delta \text{pgasman}$ . Por exemplo, se a participação dos candidatos à reeleição nos gastos aumentar em dez pontos percentuais, como se espera que isso afete a participação desses candidatos na votação?
- (iv) Refaça a parte (iii), mas agora use somente os pares que tenham concorrentes repetidos. [Isso nos possibilita controlar, também, as características dos concorrentes, o que estaria em  $a_i$ . Levitt (1995) faz uma análise muito mais abrangente do assunto.]

**13.15** Utilize os dados contidos no arquivo CRIME4.RAW para fazer este exercício.

- (i) Adicione os logs de cada variável de salários do conjunto de dados e estime o modelo fazendo uma primeira diferenciação. Como o fato dessas variáveis terem sido incluídas afeta os coeficientes das variáveis da justiça criminal no Exemplo 13.9?
- (ii) Todas as variáveis de salários na parte (i) possuem o sinal esperado? Elas são conjuntamente significantes? Explique.

**13.16** Para fazer este exercício, usamos o arquivo JTRAIN.RAW para determinar o efeito dos subsídios de treinamento de pessoal sobre o número de horas de treinamento por empregado. O modelo básico para os três anos é

$$\text{hrsemp}_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t + \beta_1 \text{subs}_{it} + \beta_2 \text{subs}_{i,t-1} + \beta_3 \log(\text{empreg}_{it}) + a_i + u_{it}.$$

- (i) Estime a equação usando a primeira diferenciação. Quantas empresas são usadas na estimação? Quantas observações totais seriam usadas se cada empresa tivesse dados sobre todas as variáveis (particularmente sobre  $\text{hrsemp}$ ) de todos os três períodos de tempo?
- (ii) Interprete o coeficiente de  $\text{subs}$  e comente sobre sua significância.
- (iii) É surpreendente o fato de o coeficiente de  $\text{subs}_{-1}$  ser não significativo? Explique.
- (iv) As empresas maiores treinam mais, ou menos, seus empregados, em média? O quanto são grandes as diferenças no treinamento?

**13.17** O arquivo MATHPNL.RAW contém dados de painel sobre distritos escolares no estado norte-americano de Michigan nos anos de 1992 a 1998. São dados em nível de distritos análogos aos dados a nível de escolas utilizados por Papke (2001). A variável de resposta de interesse nesta questão é  $\text{mate4}$ , a porcentagem de estudantes de quarta série de um distrito que obtiveram média de aprovação



em um exame padrão de matemática. A variável explicativa principal é  $grpa$  que é o gasto real por aluno no distrito. Os valores estão em dólares de 1997. A variável de gastos aparece na forma logarítmica.

- (i) Considere o modelo estático de efeitos não observados

$$mate4_{it} = \delta_1 a93_t + \dots + \delta_6 a98_t + \beta_1 \log(grpa_{it}) + \beta_2 \log(matricl_{it}) + \beta_3 merenda_{it} + a_i + u_{it},$$

onde  $matricl_{it}$  é o total de matrículas do distrito e  $merenda_{it}$  é a porcentagem de alunos no distrito habilitados a ter acesso ao programa de merenda escolar da escola. (Portanto,  $merenda_{it}$  é uma boa medida da taxa de pobreza em todo o distrito.) Argumente que  $\beta_1/10$  será o ponto percentual de mudança em  $mate4_{it}$ , quando o gasto real por aluno aumentar em aproximadamente 10%.

- (ii) Use a primeira diferenciação para estimar o modelo da parte (i). O método mais simples é admitir um intercepto na equação de primeiras diferenças e incluir variáveis *dummy* para os anos de 1994 a 1998. Interprete o coeficiente da variável de gastos.
- (iii) Agora, adicione uma defasagem da variável de gastos ao modelo e faça a reestimativa usando a primeira diferenciação. Observe que você perde mais um ano de dados, de modo que você está usando mudanças começando em 1994. Discuta os coeficientes e a significância das variáveis de gasto corrente e defasado.
- (iv) Obtenha erros-padrão robustos em relação à heteroscedasticidade para a regressão de primeiras diferenças da parte (iii). Como esses erros-padrão se comparam aos da parte (iii) para as variáveis de gasto?
- (v) Agora, obtenha erros-padrão robustos tanto quanto à heteroscedasticidade como quanto à correlação serial. O que isso faz com a significância da variável de gasto defasada?
- (vi) Verifique que os erros diferenciados  $r_{it} = \Delta u_{it}$  têm correlação serial negativa realizando um teste de correlação serial AR(1).
- (vii) Com base num teste conjunto totalmente robusto, parece ser necessário incluir as variáveis de matrícula e de merenda escolar no modelo?

## CAPÍTULO 14

**14.6** Utilize os dados contidos no arquivo RENTAL.RAW para fazer este exercício. Os dados sobre os preços de aluguéis e outras variáveis em cidades universitárias são dos anos de 1980 e 1990. A idéia é verificar se uma presença mais forte de estudantes afeta os valores dos aluguéis. O modelo de efeitos não observados é

$$\log(alug_{it}) = \beta_0 + \delta_0 a90_t + \beta_1 \log(pop_{it}) + \beta_2 \log(rendfam_{it}) + \beta_3 pctestu_{it} + a_i + u_{it},$$

onde  $pop$  é a população da cidade,  $rendfam$  é a renda média, e  $pctestu$  é a população estudantil como porcentagem da população da cidade (durante o período escolar).

- (i) Estime a equação por MQO agrupado e descreva os resultados na forma padrão. O que você conclui da estimativa da variável *dummy* de 1990? O que você obtém para  $\hat{\beta}_{pctestu}$ ?
- (ii) Os erros-padrão que você descreve na parte (i) são válidos? Explique.

- (iii) Agora, diferencie a equação e a estime por MQO. Compare sua estimativa de  $\hat{\beta}_{pctestu}$  com a da parte (i). O tamanho relativo da população estudantil parece afetar os preços dos aluguéis?
- (iv) Estime o modelo por efeitos fixos para verificar se você obtém estimativas e erros-padrão idênticos aos da parte (iii).

**14.7** Utilize os dados contidos no arquivo CRIME.RAW para fazer este exercício.

- (i) Estime novamente o modelo de efeitos não observados da criminalidade no Exemplo 13.9, mas utilize os efeitos fixos em vez da diferenciação. Existe alguma mudança considerável no sinal ou na magnitude dos coeficientes? O que é possível afirmar sobre a significância estatística?
- (ii) Adicione os logs da variável salários ao conjunto de dados e estime o modelo por efeitos fixos. Como a inclusão dessas variáveis afeta os coeficientes das variáveis de justiça criminal na parte (i)?
- (iii) Todas as variáveis referentes ao salário na parte (ii) têm o sinal esperado? Explique. Elas são conjuntamente significantes?

**14.8** Para fazer este exercício, usamos os dados contidos no arquivo JTRAIN.RAW para determinar o efeito dos subsídios de treinamento de pessoal sobre as horas de treinamento por empregado. O modelo básico para três anos é

$$hrsemp_{it} = \beta_0 + \delta_1 a88_t + \delta_2 a89_t + \beta_1 subs_{it} + \beta_2 subs_{i,t-1} + \beta_3 \log(empreg_{it}) + a_i + u_{it}.$$

- (i) Estime a equação usando efeitos fixos. Quantas empresas são usadas na estimação EF? Quantas observações totais seriam usadas se cada uma das empresas tivesse dados sobre todas as variáveis (particularmente sobre *hrsemp*) para todos os três anos?
- (ii) Interprete o coeficiente de *subs* e comente sobre sua significância.
- (iii) Surpreende o fato de *subs*<sub>-1</sub> ser não significativo? Explique.
- (iv) As empresas maiores oferecem a seus empregados mais, ou menos, treinamento, em média? O quanto são grandes as diferenças? (Por exemplo, se uma empresa tiver 10% mais empregados, qual é a mudança na média de horas de treinamento?)

**14.9** No Exemplo 13.8, usamos os dados de Papke (1994) sobre os pedidos de auxílio-desemprego para estimar o efeito da construção de áreas industriais sobre aqueles pedidos. Papke também usa um modelo que permite que cada cidade tenha sua própria tendência temporal:

$$\log(uclms_{it}) = a_i + c_t + \beta_1 z_{it} + u_{it},$$

onde  $a_i$  e  $c_t$  são, ambas, efeitos não observados. Isso leva em conta maior heterogeneidade entre as cidades.

- (i) Mostre que, quando fazemos a primeira diferenciação da equação anterior obtemos

$$\Delta \log(uclms_{it}) = c_t + \beta_1 \Delta z_{it} + \Delta u_{it}, \quad t = 2, \dots, T.$$

Observe que a equação diferenciada contém um efeito fixo,  $c_t$ .

- (ii) Estime a equação diferenciada por efeitos fixos. Qual é a estimativa de  $\beta_1$ ? Ela é muito diferente da obtida no Exemplo 13.8? O efeito das áreas industriais ainda é estatisticamente significativo?
- (iii) Adicione um conjunto completo de *dummies* anuais à estimação da parte (ii). O que acontece com a estimativa de  $\beta_1$ ?
- 14.10** (i) Na equação de salários do Exemplo 14.4, explique por que as variáveis *dummy* da ocupação podem ser variáveis omitidas importantes para estimarmos o coeficiente de *sindicato*.
- (ii) Se cada pessoa da amostra tivesse ficado na mesma ocupação de 1981 até 1987, seria necessário incluir *dummies* ocupacionais em uma estimação por efeitos fixos? Explique.
- (iii) Utilizando os dados contidos no arquivo WAGEPAN.RAW, inclua oito das variáveis *dummy* ocupacionais na equação e estime-a usando efeitos fixos. O coeficiente de *sindicato* se altera muito? O que você diz sobre sua significância estatística?
- 14.11** Adicione o termo de interação  $\text{sindicato}_{it} \cdot t$  à equação estimada na Tabela 14.2, para verificar se o *crescimento* salarial depende da filiação sindical. Estime a equação por efeitos aleatórios e fixos e compare os resultados.
- 14.12** Use os dados em nível estadual sobre taxas de criminalidade e de execuções contidos no arquivo MURDER.RAW para fazer o seguinte exercício.

- (i) Considere o modelo de efeitos não observados

$$\text{txhomi}_{it} = \theta_t + \beta_1 \text{exec}_{it} + \beta_2 \text{desemp}_{it} + a_i + u_{it},$$

onde  $\theta_t$ , simplesmente representa interceptos de anos diferentes e  $a_i$  é o efeito estadual não observado. Se as execuções passadas de assassinos condenados tiverem um efeito dissuasor, qual será o sinal de  $\beta_1$ ? Que sinal você acha que  $\beta_2$  deveria ter? Explique.

- (ii) Usando apenas os anos de 1990 e 1993, estime a equação da parte (i) por MQO agrupado. Ignore o problema da correlação serial nos erros de composição. Você encontra alguma evidência de um efeito dissuasor?
- (iii) Agora, usando os anos de 1990 e 1993, estime a equação pelos efeitos fixos. Você pode usar a primeira diferenciação já que está usando dados de somente dois anos. E agora, existe alguma evidência de um efeito dissuasor? Se houver, o quanto ele é forte?
- (iv) Compute o erro-padrão robusto em relação à heteroscedasticidade para a estimação na parte (iii). Será mais fácil utilizar a primeira diferenciação.
- (v) Encontre o estado que tenha o número mais alto na variável de execuções em 1993. (A variável *exec* é o total de execuções em 1991, 1992 e 1993.) O quanto esse valor é maior em relação ao segundo maior?
- (vi) Estime a equação usando a primeira diferenciação, eliminando o estado do Texas da análise. Compute os erros-padrão usuais e os robustos em relação à heteroscedasticidade. Agora, o que você constata? O que acontece?

- (vii) Use todos os dados dos três anos e estime o modelo por efeitos fixos. Inclua o estado do Texas na análise. Discuta o tamanho e a significância estatística do efeito dissuasor, em comparação com os resultados obtidos usando somente os anos de 1990 e 1993.

**14.13** Utilize os dados contidos no arquivo MATHPNL.RAW para fazer este exercício. Você fará uma versão com efeitos fixos da primeira diferenciação feita no Exercício 13.17. O modelo de interesse é

$$\begin{aligned} \ln \text{mate}_{it} = & \delta_1 a94_t + \dots + \delta_5 a98_t + \gamma_1 \log(\text{grpa}_{it}) + \gamma_2 \log(\text{grpa}_{i,t-1}) \\ & + \psi_1 \log(\text{matricl}_{it}) + \psi_2 \text{merenda}_{it} + a_i + u_{it}, \end{aligned}$$

onde o primeiro ano disponível (o ano base) é 1993 devido à variável defasada do dispêndio.

- (i) Estime o modelo por MQO agrupado e descreva os erros-padrão habituais. Você deve incluir um intercepto juntamente com as *dummies* anuais para todos os  $a_i$  a fim de obter um valor esperado diferente de zero. Quais são os efeitos estimados das variáveis do dispêndio? Obtenha os resíduos MQO,  $\hat{v}_{it}$ .
- (ii) O sinal do coeficiente de  $\text{merenda}_{it}$  é o que você esperava? Interprete a magnitude do coeficiente. Você diria que a taxa de pobreza da região tem um efeito grande na taxa de aprovação nos testes?
- (iii) Compute um teste da correlação serial AR(1) usando a regressão de  $\hat{v}_{it}$  sobre  $\hat{v}_{i,t-1}$ . Você deve usar os anos de 1994 a 1998 na regressão. Verifique que existe uma forte correlação serial positiva e discuta porque isso ocorre.
- (iv) Agora, estime a equação por efeitos fixos. A variável defasada do dispêndio ainda é significativa?
- (v) Por que você entende, na estimação por efeitos fixos, que as variáveis de matrículas e de merenda são conjuntamente não significantes?
- (vi) Defina o efeito total, ou de longo prazo, do dispêndio como  $\theta_1 = \gamma_1 + \gamma_2$ . Use a substituição  $\gamma_1 = \theta_1 - \gamma_2$  para obter um erro-padrão de  $\hat{\theta}_1$ . [Sugestão: A estimação padrão por efeitos fixos usando  $\log(\text{grpa}_{it})$  e  $z_{it} = \log(\text{grpa}_{i,t-1}) - \log(\text{grpa}_{it})$  como variáveis explicativas deve ajudar a resolver o problema.]

**14.14** O arquivo PENSION.RAW contém informações sobre planos de pensões dirigidos pelos próprios participantes, para os trabalhadores norte-americanos. Algumas das observações são de casais dentro de uma mesma família, de modo que esse conjunto de dados constitui uma pequena amostra de aglomeração (com tamanhos da aglomeração iguais a dois.)

- (i) Ignorando a aglomeração por família, use MQO para estimar o modelo

$$\begin{aligned} \text{pctstck} = & \beta_0 + \beta_1 \text{escolha} + \beta_2 \text{partluc} + \beta_3 \text{feminino} + \beta_4 \text{idade} + \beta_5 \text{educ} \\ & + \beta_6 \text{rendaf25} + \beta_7 \text{rendaf35} + \beta_8 \text{rendaf50} + \beta_9 \text{rendaf75} + \beta_{10} \text{rendaf100} \\ & + \beta_{11} \text{rendaf101} + \beta_{12} \text{riqueza89} + \beta_{13} \text{ações89} + \beta_{14} \text{aposind89} + u, \end{aligned}$$

onde as variáveis estão definidas no conjunto de dados. A variável de maior interesse é *escolha*, que é uma variável *dummy* igual a um se os trabalhadores puderem escolher

como alocar os fundos de pensões entre os diferentes investimentos. Qual é o efeito estimado de *escolha*? Ele é estatisticamente significativo?

- (ii) As variáveis de controle renda, riqueza, posse de ações e de plano de aposentadoria individual são importantes? Explique.
- (iii) Determine quantas famílias diferentes existem no conjunto de dados.
- (iv) Agora, obtenha os erros-padrão do MQO que sejam robustos quanto à correlação de aglomeração dentro de uma família. Eles são muito diferentes dos erros-padrão habituais do MQO? Isso lhe surpreende?
- (v) Estime a equação fazendo a diferenciação somente entre as esposas dentro de uma família. Por que as variáveis explicativas citadas na parte (ii) são eliminadas na estimação da primeira diferenciação?
- (vi) Alguma das variáveis explicativas restantes na parte (v) é significativa? Isso lhe surpreende?

## CAPÍTULO 15

---

**15.12** Utilize os dados contidos no arquivo WAGE2.RAW para fazer este exercício.

- (i) No Exemplo 15.2, usando *irms* como uma instrumental de *educ*, a estimativa VI do retorno da educação é 0,122. Para convencer a si próprio que usar *irms* como uma VI de *educ* não é a mesma coisa que inserir *irms* em *educ* e computar uma regressão por MQO, faça a regressão de  $\log(\text{salário})$  sobre *irms* e explique suas descobertas.
- (ii) A variável *ordnas* é a ordem de nascimento (*ordnas* será um para o primeiro filho, dois para o segundo, e assim por diante). Explique por que *educ* e *ordnas* podem ser negativamente correlacionados. Regrida *educ* sobre *ordnas* para determinar se existe uma correlação negativa estatisticamente significativa.
- (iii) Use *ordnas* como uma VI de *educ* na equação (15.1). Descreva e interprete os resultados.
- (iv) Agora, suponha que incluamos número de irmãos como uma variável explicativa na equação de salários; isso controlará, até certo ponto, o ambiente familiar:

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{irms} + u.$$

Suponha que queiramos usar *ordnas* como uma VI de *educ*, assumindo que *irms* seja exógeno. A forma reduzida de *educ* é

$$\text{educ} = \pi_0 + \pi_1 \text{irms} + \pi_2 \text{irms} + v.$$

Estabeleça e teste a hipótese de identificação.

- (v) Estime a equação da parte (iv) usando *ordnas* como uma VI de *educ* (e *irms* como sua própria VI). Comente sobre os erros-padrão de  $\hat{\beta}_{\text{educ}}$  e  $\hat{\beta}_{\text{irms}}$ .

- (vi) Usando os valores estimados da parte (iv),  $\hat{educ}$ , compute a correlação entre  $\hat{educ}$  e  $irms$ . Use esse resultado para explicar suas descobertas da parte (v).

**15.13** Os dados contidos no arquivo FERTIL2.RAW incluem, para mulheres de Botswana durante o ano de 1988, informações sobre as variáveis número de filhos, anos de educação, idade e condições religiosa e econômica.

- (i) Estime o seguinte modelo por MQO:

$$filhos = \beta_0 + \beta_1 educ + \beta_2 idade + \beta_3 idade^2 + u,$$

interpretando as estimativas. Particularmente, mantendo *idade* fixa, qual será o efeito estimado de um ano a mais de educação sobre a fertilidade? Se 100 mulheres receberem mais um ano de educação, quantos filhos a menos se estima que elas terão?

- (ii) *Prisem* é uma variável *dummy* igual a um se uma mulher tiver nascido durante o primeiro semestre do ano. Assumindo que *prisem* seja não-correlacionado com o termo erro da parte (i), mostre que *prisem* é um candidato razoável a VI de *educ*. (Sugestão: você precisará fazer uma regressão.)
- (iii) Estime o modelo da parte (i) usando *prisem* como uma VI de *educ*. Compare o efeito estimado da educação com o estimado por MQO da parte (i).
- (iv) Adicione as variáveis binárias *eletric*, *tv*, e *bicicleta* ao modelo e assuma que elas sejam exógenas. Estime a equação por MQO e por MQ2E e compare os coeficientes estimados de *educ*. Interprete o coeficiente de *tv* e explique por que o fato de possuir uma televisão tem um efeito negativo sobre a fertilidade.

**15.14** Utilize os dados contidos no arquivo CARD.RAW para fazer este exercício.

- (i) A equação que estimamos no Exemplo 15.4 pode ser escrita da seguinte forma

$$\log(\text{salário}) = \beta_0 + \beta_1 educ + \beta_2 exper + \dots + u,$$

onde as outras variáveis explicativas estão listadas na Tabela 15.1. Para que o método de VI seja consistente, as VIs de *educ* e *proxf4*, devem ser não-correlacionadas com *u*. A variável *proxf4* pode ser correlacionada com outros itens do termo erro, como a aptidão não observada? Explique.

- (ii) Para uma subamostra dos homens no conjunto de dados, existem informações sobre o QI. Faça a regressão de *QI* sobre *proxf4* para verificar se a média de QI varia em função do fato de um homem ter crescido próximo de uma faculdade com cursos de graduação de quatro anos. Quais suas conclusões?
- (iii) Agora, faça a regressão de *QI* sobre *proxf4*, *eprm66* e as variáveis *dummy* regionais *reg662*, ..., *reg669*. As variáveis *QI* e *proxf4* são relacionadas após as variáveis *dummy* geográficas terem sido levadas em conta? Reconcilie isso com suas descobertas da parte (ii).
- (iv) Das partes (ii) e (iii), o que você conclui sobre a importância de controlar *eprm66* e as *dummies* regionais na equação de  $\log(\text{salário})$ ?

**15.15** Utilize os dados contidos no arquivo INTDEF.RAW para fazer este exercício. Uma equação simples relacionando a taxa das letras do Tesouro norte-americano de três meses com a taxa de inflação (construída a partir do índice de preços ao consumidor) é

$$i3_t = \beta_0 + \beta_1 inf_t + u_t.$$

- (i) Estime essa equação por MQO, omitindo o primeiro período de tempo para comparações futuras. Descreva os resultados na forma habitual.
- (ii) Alguns economistas entendem que o índice de preços ao consumidor mede incorretamente a verdadeira taxa de inflação, de forma que o MQO da parte (i) sofre de viés de erro de medida. Reestime a equação da parte (i), usando  $inf_{t-1}$  como uma VI de  $inf_t$ . Como a estimativa de VI de  $\beta_1$  se compara com a do MQO?
- (iii) Agora, faça a primeira diferenciação da equação:

$$\Delta i3_t = \beta_0 + \beta_1 \Delta inf_t + \Delta u_t.$$

Estime essa nova equação por MQO e compare a estimativa de  $\beta_1$  com as estimativas anteriores.

- (iv) Você pode usar  $\Delta inf_{t-1}$  como uma VI de  $\Delta inf_t$  na equação diferenciada na parte (iii)? Explique. (*Sugestão*: serão  $\Delta inf_t$  e  $\Delta inf_{t-1}$  suficientemente correlacionadas?)

**15.16** Utilize os dados contidos no arquivo CARD.RAW para fazer este exercício.

- (i) Na Tabela 15.1, as diferenças entre as estimativas de VI e MQO do retorno da educação são economicamente importantes. Obtenha os resíduos da forma reduzida,  $\hat{v}_2$ , a partir de (15.32). (Veja na Tabela 15.1 as outras variáveis a serem incluídas na regressão.) Use essas informações para verificar se *educ* é exógeno, isto é, determine se a diferença entre o MQO e a VI é *estatisticamente* significativa.
- (ii) Estime a equação por MQ2E, adicionando *proxf2* como uma variável instrumental. O coeficiente de *educ* muda muito?
- (iii) Teste a única restrição sobreidentificadora da parte (ii).

**15.17** Utilize os dados contidos no arquivo MURDER.RAW para fazer este exercício. A variável *txhomi* é a taxa de homicídios, isto é, o número de homicídios por 100.000 habitantes. A variável *exec* é o número total de prisioneiros executados no ano atual e nos dois anos anteriores; *desemp* é a taxa de desemprego no estado.

- (i) Quantos estados executaram pelo menos um prisioneiro em 1991, 1992 ou 1993? Que estado teve o maior número de execuções?
- (ii) Usando os anos de 1990 e 1993, faça uma regressão agrupada de *txhomi* sobre *d93*, *exec* e *desemp*. O que você deduz do coeficiente de *exec*?
- (iii) Usando somente as alterações de 1990 para 1993 (de um total de 51 observações), estime a equação

$$\Delta txhomi = \delta_0 + \beta_1 \Delta exec + \beta_2 \Delta desemp + \Delta u$$

por MQO e descreva os resultados da forma habitual. Agora, a pena capital parece ter um efeito dissuasor?

- (iv) A alteração nas execuções pode ser, pelo menos parcialmente, relacionada às alterações na taxa esperada de homicídios, de forma que  $\Delta exec$  seja correlacionada com  $\Delta u$  na parte (iii). Pode ser razoável assumir que  $\Delta exec_{-1}$  seja não-correlacionada com  $\Delta u$ . (Afinal de contas,  $\Delta exec_{-1}$  depende das execuções que tenham ocorrido há três ou mais anos.) Faça a regressão de  $\Delta exec$  sobre  $\Delta exec_{-1}$  para verificar se elas são suficientemente correlacionadas; interprete o coeficiente de  $\Delta exec_{-1}$ .
- (v) Reestime a equação da parte (iii), usando  $\Delta exec_{-1}$  como uma VI de  $\Delta exec$ . Assuma que  $\Delta desemp$  seja exógena. De que forma mudam suas conclusões em relação às da parte (iii)?

**15.18** Utilize os dados contidos no arquivo PHILLIPS.RAW para fazer este exercício.

- (i) No Exemplo 11.5, estimamos uma curva de Phillips de expectativas aumentadas da forma

$$\Delta inf_t = \beta_0 + \beta_1 desemp_t + e_t,$$

onde  $\Delta inf_t = inf_t - inf_{t-1}$ . Ao estimarmos essa equação por MQO, assumimos que o choque de oferta,  $e_t$ , era não-correlacionado com  $desemp_t$ . Se isso for falso, o que poderá ser dito sobre o estimador MQO de  $\beta_1$ ?

- (ii) Suponha que  $e_t$  não seja previsível, dadas todas as informações passadas:  $E(e_t | inf_{t-1}, desemp_{t-1}, \dots) = 0$ . Explique por que isso faz com que  $desemp_{t-1}$  seja uma boa candidata a VI de  $desemp_t$ .
- (iii) Faça a regressão de  $desemp_t$  sobre  $desemp_{t-1}$ . É possível afirmar que  $desemp_t$  e  $desemp_{t-1}$  são significativamente correlacionadas?
- (iv) Estime a curva de Phillips de expectativas aumentadas por VI. Descreva os resultados da forma habitual e compare-os com as estimativas MQO do Exemplo 11.5.

**15.19** Utilize os dados contidos no arquivo 401KSUBS.RAW para fazer este exercício. A equação de interesse é um modelo de probabilidade linear:

$$plapind = \beta_0 + \beta_1 p401k + \beta_2 renda + \beta_3 renda^2 + \beta_4 idade + \beta_5 idade^2 + u.$$

O objetivo é verificar se existe uma relação de substituição entre ser participante de um plano de pensão e ter um plano de aposentadoria privado. Portanto, queremos estimar  $\beta_1$ .

- (i) Estime a equação por MQO e detalhe o efeito estimado de  $p401k$ .
- (ii) Com o propósito de estimar a relação de substituição *ceteris paribus* entre a participação em dois tipos diferentes de planos de previdência, qual poderia ser o problema com os mínimos quadrados ordinários?
- (iii) A variável  $e401k$  é uma variável binária igual a um se um trabalhador for qualificado para participar do plano de pensão. Explique o que é requerido para que  $e401k$  seja uma VI válida de  $p401k$ . Essas hipóteses parecem razoáveis?



- (iv) Estime a forma reduzida de  $p401k$  e verifique que  $e401k$  tem correlação parcial significativa com  $p401k$ . Como a forma reduzida também é um modelo de probabilidade linear, use um erro-padrão robusto em relação à heteroscedasticidade.
- (v) Agora, estime a equação estrutural por VI e compare a estimativa de  $\beta_1$  com a estimativa MQO. Novamente, você deve obter erros-padrão robustos em relação à heteroscedasticidade.
- (vi) Teste a hipótese nula de que  $p401k$  é de fato exógena, usando um teste robusto em relação à heteroscedasticidade.

**15.20** O propósito deste exercício é comparar as estimativas e erros-padrão obtidos pelo uso correto do MQ2E com os obtidos pelo uso de procedimentos inapropriados. Utilize os dados contidos no arquivo WAGE2.RAW.

- (i) Use uma rotina MQ2E para estimar a equação

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + \beta_4 \text{negro} + u,$$

onde  $irms$  seja uma VI de  $educ$ . Descreva os resultados na forma habitual.

- (ii) Agora, manualmente, execute o MQ2E. Isto é, primeiro faça a regressão de  $educ_i$  sobre  $irms_i$ ,  $exper_i$ ,  $perm_i$ , e  $negro_i$  e obtenha os valores estimados,  $\hat{educ}_i$ ,  $i = 1, \dots, n$ . Depois, execute a segunda etapa da regressão de  $\log(\text{salário}_i)$  sobre  $educ_i$ ,  $exper_i$ ,  $perm_i$ , e  $negro_i$ ,  $i = 1, \dots, n$ . Verifique que os  $\hat{\beta}_j$  são idênticos aos obtidos na parte (i), mas que os erros-padrão são ligeiramente diferentes. Os erros-padrão obtidos do segundo estágio da regressão quando executamos, manualmente o MQ2E são geralmente inapropriados.
- (iii) Agora, use o seguinte procedimento de duas etapas, que geralmente produz estimativas de parâmetros inconsistentes dos  $\beta_j$  e erros-padrão não tão inconsistentes. Na etapa um, faça a regressão de  $educ_i$  somente sobre  $irms_i$ , e obtenha os valores estimados, digamos  $\tilde{educ}_i$ . (Note que essa é uma regressão de primeira etapa incorreta.) Depois, na segunda etapa, execute a regressão de  $\log(\text{salário}_i)$  sobre  $\tilde{educ}_i$ ,  $exper_i$ ,  $perm_i$ , e  $negro_i$ ,  $i = 1, \dots, n$ . Como as estimativas desse procedimento incorreto de duas etapas se comparam com as estimativas corretas de MQ2E do retorno da educação?

**15.21** Utilize os dados contidos no arquivo HTV.RAW para fazer este exercício.

- (i) Execute uma regressão simples por MQO de  $\log(\text{salário})$  sobre  $educ$ . Sem controlar outros fatores, qual é o intervalo de confiança de 95% do retorno de um ano a mais de educação?
- (ii) A variável  $ctuit$  é a mudança do preço pago pelo ensino pelos alunos ao passarem de 17 para 18 anos. Mostre que  $educ$  e  $ctuit$  são essencialmente não-correlacionadas. O que isto diz sobre  $ctuit$  como uma possível VI de  $educ$  em uma análise de regressão simples?
- (iii) Agora, adicione ao modelo de regressão simples na parte (i) um termo quadrático da experiência e um conjunto total de variáveis *dummy* regionais da residência atual e residência na idade de 18 anos. Inclua também os indicadores urbanos das residências atual e na idade de 18 anos. Qual é o retorno estimado de um ano de educação?

- (iv) Novamente usando *ctuit* como uma VI potencial de *educ*, estime a forma reduzida de *educ*. Mostre que *ctuit* é agora estatisticamente significativa na forma reduzida de *educ*.
- (v) Estime o modelo da parte (iii) por VI, usando *ctuit* como uma VI de *educ*. Como se compara o intervalo de confiança do retorno da educação com aquele da parte (iii)?
- (vi) Você acha que o procedimento de VI da parte (v) é convincente?

## CAPÍTULO 16

---

**16.9** Utilize os dados contidos no arquivo SMOKE.RAW para fazer este exercício.

- (i) Um modelo para estimar o efeito do hábito de fumar sobre a renda anual (talvez pelos dias de trabalho perdidos por motivo de doença ou efeitos sobre a produtividade) é

$$\log(\text{renda}) = \beta_0 + \beta_1 \text{cigs} + \beta_2 \text{educ} + \beta_3 \text{idade} + \beta_4 \text{idade}^2 + u_1,$$

onde *cigs* é a quantidade de cigarros fumados por dia, em média. Como você interpreta  $\beta_1$ ?

- (ii) Para refletir o fato de que o consumo de cigarros pode ser determinado conjuntamente com a renda, uma equação da demanda por cigarros é

$$\begin{aligned} \text{cigs} = \gamma_0 + \gamma_1 \log(\text{renda}) + \gamma_2 \text{educ} + \gamma_3 \text{idade} + \gamma_4 \text{idade}^2 \\ + \gamma_5 \log(\text{precig}) + \gamma_6 \text{restaurn} + u_2, \end{aligned}$$

onde *precig* é o preço de um maço de cigarros (em centavos) e *restaurn* é uma variável binária igual a um se a pessoa vive em um estado com restrições sobre fumar em restaurantes. Assumindo que essas variáveis sejam exógenas para o indivíduo, que sinais você esperaria para  $\gamma_5$  e  $\gamma_6$ ?

- (iii) Sob qual hipótese a equação da renda da parte (i) será identificada?
- (iv) Estime a equação da renda por MQO e discuta a estimativa de  $\beta_1$ .
- (v) Estime a forma reduzida de *cigs*. (Lembre que isso acarretará fazer a regressão de *cigs* sobre todas as variáveis exógenas.) É possível afirmar se  $\log(\text{precig})$  e *restaurn* são significantes na forma reduzida?
- (vi) Agora, estime a equação da renda por MQ2E. Detalhe como a estimativa de  $\beta_1$  se compara com aquela estimada por MQO.
- (vii) Você considera que os preços dos cigarros e as restrições ao hábito de fumar em restaurantes são exógenos na equação da renda?

**16.10** Utilize os dados contidos no arquivo MROZ.RAW para fazer este exercício.

- (i) Reestime a função da oferta de mão-de-obra no Exemplo 16.5, usando  $\log(\text{horas})$  como a variável dependente. Compare a elasticidade estimada (que agora é constante) com a estimativa obtida da equação (16.24) na média de horas trabalhadas.
- (ii) Na equação da oferta de mão-de-obra da parte (i), permita que *educ* seja endógena devido à aptidão omitida. Use *educm* e *educp* como VIs de *educ*. Lembre-se, agora você tem duas variáveis endógenas na equação.

- (iii) Teste as restrições sobreidentificadoras na estimação por MQ2E da parte (ii). As VIs passam no teste?

**16.11** Utilize os dados contidos no arquivo OPENNESS.RAW para fazer este exercício.

- (i) Como  $\log(\text{rendpc})$  é não significativa tanto em (16.22) como na forma reduzida de *abertura*, elimine-a da análise. Estime (16.22) por MQO e por VI sem  $\log(\text{rendpc})$ . Alguma das conclusões importantes se altera?
- (ii) Ainda mantendo  $\log(\text{rendpc})$  fora da análise, a variável *área* ou  $\log(\text{área})$  é uma variável instrumental melhor de *abertura*? (Sugestão: Faça a regressão de *abertura* sobre cada uma dessas variáveis separadamente e conjuntamente.)
- (iii) Agora, retorne à equação (16.22). Adicione a variável *dummy petróleo* à equação e trate-a como exógena. Estime a equação por VI. O fato de ser um produtor de petróleo tem um efeito *ceteris paribus* sobre a inflação?

**16.12** Utilize os dados contidos no arquivo CONSUMP.RAW para fazer este exercício.

- (i) No Exemplo 16.7, use o método da Seção 15.5 para testar a restrição sobreidentificadora isolada na estimativa de (16.35). Quais suas conclusões?
- (ii) Campbell e Mankiw (1990) usaram *segundas* defasagens de todas as variáveis como VIs devido a problemas potenciais de mensuração dos dados e de defasagens informativas. Reestime a equação (16.35), usando somente  $gc_{t-2}$ ,  $gy_{t-2}$  e  $r\mathfrak{z}_{t-2}$  como VIs. Como as estimativas se comparam com as de (16.36)?
- (iii) Faça a regressão de  $gy_t$  sobre as VIs da parte (ii) e verifique se  $gy_t$  é suficientemente correlacionada com elas. Por que isso é importante?

**16.13** Utilize o *Economic Report of the President* (2002 ou posterior) para atualizar os dados do arquivo CONSUMP.RAW pelo menos até o ano de 2000. Reestime a equação (16.35). Alguma das conclusões importantes se altera?

**16.14** Utilize os dados contidos no arquivo CEMENT.RAW para fazer este exercício.

- (i) Uma função estática (inversa) de oferta para o crescimento mensal do preço do cimento (*crescprcim*) como uma função do crescimento na quantidade (*crescim*) é

$$\text{crescprcim}_t = \alpha_1 \text{crescim}_t + \beta_0 + \beta_1 \text{crescprpet} + \beta_2 \text{fev}_t + \dots + \beta_{12} \text{dez}_t + u_t^s,$$

onde *crescprpet* (crescimento no preço do petróleo) é assumida como exógena e *fev*, ..., *dez* são variáveis *dummy* mensais. Que sinais você antecipa para  $\alpha_1$  e  $\beta_1$ ? Estime a equação por MQO. A função de oferta se inclina para cima?

- (ii) A variável *crescdef* representa o crescimento mensal nos gastos reais com a defesa nos Estados Unidos. O que você precisa assumir sobre *crescdef* para que ela seja uma boa VI de *crescim*? Teste se *crescim* é parcialmente correlacionado com *crescdef*. (Não se preocupe com a possível correlação serial na forma reduzida.) Você pode usar *crescdef* como uma VI ao estimar a função de oferta?
- (iii) Shea (1993) alega que o crescimento na produção de construção residencial (*crescres*) e não-residencial (*crescnres*) são variáveis instrumentais válidas de *crescim*. A idéia é que elas são deslocadoras da demanda que devem ser em termos gerais não-correlacio-

nadas com o erro da oferta  $u_t^s$ . Teste se *crecscim* é parcialmente correlacionada com *crecres* e *crecscresd*; novamente, não se preocupe com a possível correlação serial na forma reduzida.

- (iv) Estime a função de oferta, usando *crecres* e *crecscres* como VIs de *crecscim*. O que você conclui sobre a função estática da oferta de cimento? [A função dinâmica da oferta tem, aparentemente, uma inclinação para cima; veja Shea (1993).]

**16.15** Refira-se ao Exemplo 13.9 e aos dados contidos no arquivo CRIME4.RAW.

- (i) Suponha que, após ter feito a diferenciação para remover o efeito não-observado, você entenda que  $\Delta \log(polpc)$  seja simultaneamente determinada com  $\Delta \log(txcrim)$ ; em particular, aumentos na criminalidade estão associados com aumentos da força policial. Como isso ajuda a explicar o coeficiente positivo de  $\Delta \log(polpc)$  na equação (13.33)?
- (ii) A variável *imppc* são os impostos coletados por pessoa no município. Parece razoável excluí-la da equação sobre a criminalidade?
- (iii) Estime a forma reduzida de  $\Delta \log(polpc)$  usando o MQO agrupado, inclusive a VI em potencial  $\Delta \log(imppc)$ . Parece que  $\Delta \log(imppc)$  é uma boa candidata a VI? Explique.
- (iv) Suponha que, em vários dos anos, o estado da Carolina do Norte conferiu subsídios a alguns municípios para estes aumentarem o tamanho de suas forças policiais. Como você poderia usar essa informação para estimar o efeito de mais policiais sobre a taxa de criminalidade?

**16.16** Utilize os dados contidos no arquivo FISH.RAW, fornecidos por Graddy (1995), para fazer este exercício. O conjunto de dados também foi usado no Exercício em Computador 12.15. Agora, eles serão utilizados para estimar uma função de demanda por peixe.

- (i) Assuma que a equação de demanda pode ser escrita, em equilíbrio para cada período de tempo, como

$$\log(quantot_t) = \alpha_1 \log(premédio_t) + \beta_{10} + \beta_{11}seg_t + \beta_{12}ter_t + \beta_{13}qua_t + \beta_{14}qui_t + u_{1t},$$

de forma que é permitido que a demanda difira entre os dias da semana. Tratando as variáveis de preço como endógenas, que informação adicional necessitamos para estimar consistentemente os parâmetros da equação de demanda?

- (ii) As variáveis *onda2<sub>t</sub>* e *onda3<sub>t</sub>* representam as medidas da altura das ondas do oceano ao longo dos vários últimos dias. Quais são as duas hipóteses que precisamos fazer para podermos usar *onda2<sub>t</sub>* e *onda3<sub>t</sub>* como VIs de  $\log(premédio_t)$  para estimar a equação da demanda?
- (iii) Faça a regressão de  $\log(premédio_t)$  sobre as *dummies* dos dias da semana e sobre as duas medidas de ondas. As variáveis *onda2<sub>t</sub>* e *onda3<sub>t</sub>* são conjuntamente significantes? Qual é o *p*-valor do teste?
- (iv) Agora, estime a equação da demanda por MQ2E. Qual é o intervalo de confiança de 95% da elasticidade-preço da demanda? A elasticidade estimada é razoável?
- (v) Obtenha os resíduos do MQ2E,  $\hat{u}_{1t}$ . Adicione uma única defasagem,  $\hat{u}_{1,t-1}$  para estimar a equação de demanda por MQ2E. Lembre-se, use  $\hat{u}_{1,t-1}$  como sua própria variável instrumental. Existe evidência de correlação serial AR(1) nos erros da equação de demanda?

- (vi) Considerando que a equação de oferta evidentemente depende das variáveis relativas às ondas, quais duas hipóteses teríamos que fazer para estimar a elasticidade-preço da oferta?
- (vii) Na forma reduzida da equação de  $\log(\text{premed}_t)$ , as *dummies* dos dias da semana são conjuntamente significantes? O que você conclui sobre ter condições de estimar a elasticidade da oferta?

## CAPÍTULO 17

---

**17.8** Utilize os dados contidos no arquivo PNTSPRD.RAW para fazer este exercício.

- (i) A variável *favvence* é uma variável binária que assume o valor um se uma equipe favorecida pela lista de apostas de Las Vegas vencer. Um modelo de probabilidade linear para estimar a probabilidade de a equipe favorecida vencer é

$$P(\text{favvence} = 1 | \text{ltapostas}) = \beta_0 + \beta_1 \text{ltapostas}.$$

Explique por que, se a lista de apostas incorporar todas as informações relevantes, esperamos  $\beta_0 = 0,5$ .

- (ii) Estime o modelo da parte (i) por MQO. Teste  $H_0: \beta_0 = 0,5$  contra uma alternativa bilateral. Utilize tanto os erros-padrão habituais como os robustos em relação à heteroscedasticidade.
- (iii) O coeficiente de *ltapostas* é estatisticamente significativo? Qual é a probabilidade estimada de que a equipe favorecida vença quando *ltapostas* = 10?
- (iv) Agora, estime um modelo probit para  $P(\text{favvence} = 1 | \text{ltapostas})$ . Interprete e teste a hipótese nula de que o intercepto é zero. [Sugestão: Lembre-se que  $\Phi(0) = 0,5$ ].
- (v) Use o modelo probit para estimar a probabilidade de que a equipe favorecida vença quando *ltapostas* = 10. Compare o resultado com a estimativa MPL da parte (iii).
- (vi) Adicione as variáveis *casafav*, *fav25* e *aza25* ao modelo probit e teste a significância conjunta dessas variáveis usando o teste da razão de verossimilhança. (Quanto *gl* estão na distribuição qui-quadrada?) Interprete esse resultado, concentrando-se na questão de se a lista de apostas incorpora todas as informações observáveis antes do jogo.

**17.9** Utilize os dados contidos no arquivo LOANAPP.RAW para fazer este exercício; vide também o Problema 7.16

- (i) Estime um modelo probit de *aprovado* sobre *branco*. Encontre a probabilidade estimada de aprovações de empréstimos tanto para brancos como para não-brancos. Como essas estimativas se comparam com as da probabilidade linear?
- (ii) Agora, adicione as variáveis *gastdom*, *outrobr*, *montempr*, *desemp*, *masculino*, *casado*, *dep*, *est*, *aval*, *chist*, *falid*, *inadimp1*, *inadimp2* e *vr* ao modelo probit. Existe evidência estatisticamente significativa de discriminação contra os não-brancos?
- (iii) Estime o modelo da parte (ii) por logit. Compare o coeficiente de *branco* com a estimativa probit.
- (iv) Como você compararia o tamanho do efeito da discriminação entre o probit e o logit?

**17.10** Utilize os dados contidos no arquivo FRINGE.RAW para fazer este exercício.

- (i) Para que porcentagem dos trabalhadores na amostra *pensão* é igual a zero? Qual é a amplitude de *pensão* para os trabalhadores com benefício de pensão diferente de zero? Por que um modelo Tobit é apropriado para modelar *pensão*?
- (ii) Estime um modelo Tobit explicando *pensão* em termos de *exper*, *idade*, *perm*, *educ*, *deps*, *casado*, *branco*, e *masculino*. É possível afirmar que homens brancos têm valores esperados de benefícios de pensão maiores estatisticamente significantes?
- (iii) Use os resultados da parte (ii) para estimar a diferença nos benefícios de pensão esperados de um homem branco e de uma mulher não-branca, ambas as pessoas com 35 anos de idade, mais de 16 anos de estudo, e 10 anos de experiência.
- (iv) Adicione *sindicato* ao modelo Tobit e comente sobre sua significância.
- (v) Aplique o modelo Tobit da parte (iv), mas com *razpen*, a proporção dos ganhos em relação à pensão, como a variável dependente. (Observe que ela é uma fração entre zero e um, mas, embora muitas vezes ela assuma o valor zero, ela nunca chega perto de ser a unidade. Assim, o modelo Tobit é uma boa aproximação.) O sexo ou a raça têm efeito sobre a proporção dos ganhos sobre a pensão?

**17.11** No Exemplo 9.1, adicionamos os termos quadráticos  $pcond^2$ ,  $ptemp86^2$ , e  $rend86^2$  a um modelo linear de  $npre86$ .

- (i) Utilize os dados contidos no arquivo CRIME.RAW para adicionar esses mesmos termos à regressão de Poisson no Exemplo 17.3.
- (ii) Compute a estimativa de  $\sigma^2$  dada por  $\hat{\sigma}^2 = (n - k - 1)^{-1} \sum_{i=1}^n \hat{u}_i^2 / \hat{y}_i$ . Existe evidência de superdispersão? Como deveriam ser ajustados os erros-padrão da EMV de Poisson?
- (iii) Use os resultados das partes (i) e (ii) e a Tabela 17.3 para computar a estatística quase-razão de verossimilhança para a significância conjunta dos três termos quadráticos. Qual sua conclusão?

**17.12** Refira-se à Tabela 13.1 no Capítulo 13. Ali, usamos os dados contidos no arquivo FERTIL1.RAW para estimarmos um modelo linear de *kids*, o número de filhos que uma mulher já teve.

- (i) Estime um modelo de regressão de Poisson de *kids*, usando as mesmas variáveis da Tabela 13.1. Interprete o coeficiente de  $a82$ .
- (ii) Qual é a diferença percentual estimada na fertilidade entre uma mulher negra e uma mulher não-negra, mantendo fixos todos os outros fatores?
- (iii) Obtenha  $\hat{\sigma}$ . Existe evidência de superdispersão ou subdispersão?
- (iv) Compute os valores estimados da regressão de Poisson e obtenha o  $R$ -quadrado como o quadrado da correlação entre  $kids_i$  e  $\hat{kids}_i$ . Compare o resultado com o  $R$ -quadrado do modelo de regressão linear.

**17.13** Utilize os dados contidos no arquivo RECID.RAW para estimar o modelo do Exemplo 17.4 por MQO, usando somente as 552 durações não-censuradas. Comente, de forma geral, como essas estimativas se comparam com as da Tabela 17.4.

**17.14** Utilize os dados contidos no arquivo MROZ.RAW para fazer este exercício.

- (i) Usando as 428 mulheres que faziam parte da força de trabalho, estime o retorno da educação por MQO, incluindo  $exper$ ,  $exper^2$ ,  $nesprend$ ,  $idade$ ,  $crianmed6$  e  $crianma6$  como variáveis explicativas. Informe a estimativa do coeficiente de  $educ$  e seu erro-padrão.
- (ii) Agora, estime o retorno da educação pelo método Heckit, no qual todas as variáveis exógenas aparecem na segunda etapa da regressão. Em outras palavras, a regressão é  $\log(\text{salário})$  sobre  $educ$ ,  $exper$ ,  $exper^2$ ,  $nesprend$ ,  $idade$ ,  $crianmed6$ ,  $crianma6$  e  $\hat{\lambda}$ . Compare o retorno estimado da variável educação e seu erro-padrão com aqueles da parte (i).
- (iii) Usando somente as 428 observações das mulheres que trabalham, faça a regressão de  $\hat{\lambda}$  sobre  $educ$ ,  $exper$ ,  $exper^2$ ,  $nesprend$ ,  $idade$ ,  $crianmed6$  e  $crianma6$ . Qual o tamanho do  $R$ -quadrado? Como isso ajuda a explicar suas constatações da parte (ii)? (*Sugestão*: Pense na multicolinearidade).

**17.15** O arquivo JTRAIN2.DTA contém dados sobre um programa de treinamento de pessoal para um grupo de homens. Os homens poderiam aderir ao programa a partir de janeiro de 1976 e até meados de 1977. O programa terminou em dezembro de 1977. O objetivo é testar se a participação no programa de treinamento produziu efeito sobre as probabilidades de desemprego e a renda em 1978.

- (i) A variável  $train$  é o indicador do treinamento de pessoal. Quantos homens da amostra participaram do programa de treinamento de pessoal? Qual foi o maior número de meses em que um homem efetivamente participou do programa?
- (ii) Compute uma regressão linear de  $train$  sobre diversas variáveis demográficas e anteriores ao treinamento:  $desemp74$ ,  $desemp75$ ,  $idade$ ,  $educ$ ,  $negro$ ,  $hispan$  e  $casado$ . Essas variáveis são conjuntamente significantes ao nível de 5%?
- (iii) Estime uma versão probit do modelo linear da parte (ii). Compute o teste da razão de verossimilhança para a significância conjunta de todas as variáveis. Qual sua conclusão?
- (iv) Com base em suas respostas para as partes (ii) e (iii), você considera que a participação em programas de treinamento de pessoal pode ser tratada como exógena para explicar a situação de desemprego em 1978? Explique.
- (v) Execute uma regressão simples de  $unem78$  sobre  $train$  e descreva os resultados em forma de equação. Qual é o efeito estimado de participar do programa de treinamento de pessoal sobre a probabilidade de ficar desempregado em 1978? Ele é estatisticamente significativo?
- (vi) Compute uma versão probit de  $unem78$  sobre  $train$ . Faz sentido comparar o coeficiente probit de  $train$  com o coeficiente obtido no modelo linear da parte (v)?
- (vii) Encontre as probabilidades estimadas das partes (v) e (vi). Explique por que elas são idênticas. Qual método você deve usar para medir o efeito e a significância estatística do programa de treinamento de pessoal?
- (viii) Adicione todas as variáveis da parte (ii) como controles adicionais aos modelos das partes (v) e (vi). As probabilidades estimadas são, agora, idênticas? Qual é a correlação entre elas?

**17.16** Utilize os dados contidos no arquivo APPLE.RAW para fazer este exercício. Ele contém dados de uma pesquisa telefônica feita para obter a demanda por maçãs (imaginárias) “ecologi-



camente corretas”. Foi apresentada (aleatoriamente) a cada família uma relação de preços de maçãs normais e maçãs com selo ecológico. Perguntou-se quantas libras de cada tipo de maçã elas comprariam.

- (i) Das 660 famílias da amostra, quantas disseram não querer nenhuma das maçãs com selo ecológico aos preços fornecidos?
- (ii) A variável *ecolbs* parece ter uma distribuição contínua sobre valores estritamente positivos? Que implicações sua resposta tem quanto à adequação de um modelo Tobit para *ecolbs*?
- (iii) Estime um modelo Tobit para *ecolbs* com *ecoprc*, *regprc*, *rendfam*, e *tamfam* como variáveis explicativas. Quais variáveis são significantes ao nível de 1%?
- (iv) Os sinais dos coeficientes das variáveis de preço da parte (iii) são os que você esperava? Explique.
- (v) Seja  $\beta_1$  seja o coeficiente de *ecoprc* e  $\beta_2$  o coeficiente de *regprc*. Teste a hipótese  $H_0: -\beta_1 = \beta_2$  contra uma alternativa bilateral. Informe o *p*-valor do teste. (Você deve rever a Seção 4.4 caso seu programa econométrico não computar com facilidade esse tipo de teste.)
- (vi) Obtenha as estimativas de  $E(ecolbs|\mathbf{x})$  de todas as observações da amostra. [Veja a equação (17.22)]. Chame-as de  $ecôlbs_i$ . Quais são o menor e o maior valores estimados?
- (vii) Compute o quadrado da correlação entre  $ecolbs_i$  e  $ecôlbs_i$ .
- (viii) Agora, estime um modelo linear para *ecolbs* usando as mesmas variáveis explicativas da parte (iii). Por que as estimativas por MQO são tão menores que as do modelo Tobit? Em termos de grau de ajuste, o modelo Tobit é melhor que o modelo linear?
- (ix) Avalie a seguinte afirmação: “Como o *R*-quadrado do modelo Tobit é tão pequeno, os efeitos estimados dos preços provavelmente serão inconsistentes”.

**17.17** Utilize os dados contidos no arquivo SMOKE.RAW para fazer este exercício.

- (i) A variável *cigs* é o número de cigarros fumados por dia. Quantas pessoas na amostra não fumam? Que fração das pessoas declaram fumar 20 cigarros por dia? Por que você acredita haver um acúmulo de pessoas na faixa de 20 cigarros?
- (ii) Com base em sua resposta da parte (i), *cigs* parece ser uma boa candidata para ter uma distribuição de Poisson condicional?
- (iii) Estime um modelo de regressão de Poisson para *cigs*, incluindo  $\log(\text{precig})$ ,  $\log(\text{renda})$ , *branco*, *educ*, *idade*, e  $\text{idade}^2$  como variáveis explicativas. Quais são as elasticidades estimadas em relação ao preço e à renda?
- (iv) Usando os erros-padrão de máxima verossimilhança, as variáveis preço e renda são estatisticamente significantes ao nível de 5%?
- (v) Obtenha a estimativa da  $\sigma^2$  descrita após (17.32). O que é  $\hat{\sigma}$ ? Como você deve ajustar os erros-padrão da parte (iv)?
- (vi) Usando os erros-padrão ajustados da parte (v), as elasticidades em relação ao preço e à renda agora são estatisticamente diferentes de zero? Explique.



- (vii) As variáveis educação e idade são significantes usando os erros-padrão mais robustos? Como você interpreta o coeficiente de *educ*?
- (viii) Obtenha os valores estimados,  $\hat{y}_i$ , do modelo de regressão de Poisson. Encontre os valores mínimo e máximo e detalhe com que eficiência o modelo exponencial prevê o fumante inveterado.
- (ix) Usando os valores estimados da parte (viii), obtenha o coeficiente de correlação elevado ao quadrado entre  $\hat{y}_i$  e  $y_i$ .
- (x) Estime um modelo linear para *cigs* por MQO, usando as variáveis explicativas (e as mesmas formas funcionais) utilizadas na parte (iii). Qual o melhor ajuste – o modelo linear ou o modelo exponencial? O *R*-quadrado dos modelos é muito grande?

## CAPÍTULO 18

---

**18.10** Utilize os dados contidos no arquivo WAGEPRC.RAW para fazer este exercício. O Problema 11.5 forneceu estimativas de um modelo de defasagem distribuída finita de *crpreço* sobre *crsalhr*, onde 12 defasagens de *crsalhr* foram usadas.

- (i) Estime um modelo simples de DD de *crpreço* sobre *crsalhr*. Em particular, estime a equação (18.11) por MQO. Qual será a propensão de impacto estimada e a PLP? Esboce a distribuição de defasagens estimada.
- (ii) Compare PI e PLP estimadas com as obtidas no Problema 11.5. Como se comparam as distribuições de defasagens estimadas?
- (iii) Agora, estime o modelo de defasagem distribuída racional a partir de (18.16). Esboce a distribuição de defasagens e compare as PI e PLP estimadas com as obtidas na parte (ii).

**18.11** Utilize os dados contidos no arquivo HSEINV.RAW para fazer este exercício.

- (i) Teste a existência de uma raiz unitária em  $\log(\text{invpc})$ , incluindo uma tendência temporal linear e duas defasagens de  $\Delta \log(\text{invpc}_i)$ . Use um nível de significância de 5%.
- (ii) Use a abordagem da parte (i) para testar a existência de uma raiz unitária em  $\log(\text{preço})$ .
- (iii) Dados os resultados das partes (i) e (ii), faz sentido testar a existência de co-integração entre  $\log(\text{invpc})$  e  $\log(\text{preço})$ ?

**18.12** Utilize os dados contidos no arquivo VOLAT.RAW para fazer este exercício.

- (i) Estime um modelo AR(3) para *pcip*. Agora, adicione uma quarta defasagem e verifique se essa alteração é muito significativa.
- (ii) Em relação ao modelo AR(3) da parte (i), adicione três defasagens de *pcsp* para testar se *pcsp* Granger-causa *pcip*. Cuidadosamente, descreva suas conclusões.
- (iii) Em relação ao modelo da parte (ii) adicione três defasagens da mudança em *i3*, a taxa das letras do Tesouro de três meses. *Pcsp* Granger-causa *pcip* condicional a  $\Delta i3$  passada?

**18.13** Ao testar a co-integração entre *tgf* e *ip* no Exemplo 18.5, adicione  $t^2$  à equação (18.32) para obter os resíduos MQO. Inclua uma defasagem no teste DF aumentado. O valor crítico a 5% do teste é  $-4,15$ .

**18.14** Utilize os dados contidos no arquivo INTQRT.RAW para fazer este exercício.

- (i) No Exemplo 18.7, estimamos um modelo de correção de erro para os rendimentos das letras do Tesouro de seis meses, onde uma defasagem dos rendimentos das letras do Tesouro de três meses era a variável explicativa. Assumimos que o parâmetro de co-integração era um na equação  $hy6_t = \alpha + \beta hy3_{t-1} + u_t$ . Agora, adicione a mudança de adiantamento,  $\Delta hy3_t$ , a mudança contemporânea,  $\Delta hy3_{t-1}$ , e a mudança defasada,  $\Delta hy3_{t-2}$ , de  $hy3_{t-1}$ . Isto é, estime a equação

$$hy6_t = \alpha + \beta hy3_{t-1} + \phi_0 \Delta hy3_t + \phi_1 \Delta hy3_{t-1} + \rho_1 \Delta hy3_{t-2} + e_t$$

e descreva os resultados em forma de equação. Teste  $H_0: \beta = 1$  contra uma alternativa bilateral. Assuma que o adiantamento e a defasagem são suficientes, de forma que  $\{hy3_{t-1}\}$  seja estritamente exógena nessa equação e não se preocupe com a correlação serial.

- (ii) Em relação ao modelo de correção de erro em (18.39), adicione  $\Delta hy3_{t-2}$  e  $(hy6_{t-2} - hy3_{t-3})$ . Esses termos são conjuntamente significantes? Qual sua conclusão sobre o modelo de correção de erro apropriado?

**18.15** Utilize os dados contidos no arquivo PHILLIPS.RAW, adicionando os valores de 1997 de *desemp* e *inf*: 4,9 e 2,3, respectivamente.

- (i) Estime os modelos em (18.48) e (18.49), utilizando os dados até 1997. As estimativas dos parâmetros alteram-se muito, comparadas às de (18.48) e (18.49)?
- (ii) Use as novas equações para fazer a previsão da *desemp*<sub>1998</sub>: arredonde para duas casas decimais. Use o *Economic Report of the President* (1999 ou mais recente) para obter *desemp*<sub>1998</sub>. Qual das equações produz uma melhor previsão?
- (iii) Como discutimos no texto, a previsão de *desemp*<sub>1998</sub> usando (18.49) é 4,90. Compare esse número com a previsão obtida usando os dados até 1997. O uso do ano extra de dados para obter as estimativas dos parâmetros produz uma melhor previsão?
- (iv) Use o modelo estimado na (18.48) para obter uma previsão com dois passos à frente de *desemp*. Isto é, faça a previsão de *desemp*<sub>1998</sub> usando a equação (18.55) com  $\hat{\alpha} = 1,572$ ,  $\hat{\rho} = 0,732$ , e  $h = 2$ . Essa previsão é melhor ou pior que a obtida com um passo à frente pela inclusão de *desemp*<sub>1997</sub> = 4,9 em (18.48)?

**18.16** Utilize os dados contidos no arquivo BARIUM.RAW para fazer este exercício.

- (i) Estime o modelo de tendência linear  $chnimp_t = \alpha + \beta t + u_t$ , usando as primeiras 119 observações (isso exclui os últimos 12 meses de observações de 1998). Qual é o erro-padrão da regressão?
- (ii) Agora, estime um modelo AR(1) da *chnimp*, novamente usando todos os dados exceto os últimos 12 meses. Compare o erro-padrão da regressão com o da parte (i). Qual dos modelos produz um melhor ajuste dentro da amostra?

- (iii) Use os modelos das partes (i) e (ii) para computar os erros de previsão com um passo à frente para os 12 meses de 1998. (Você deve obter 12 erros de previsão para cada método.) Compute e compare os REQM e os EAM dos dois métodos. Qual método de previsão funciona melhor fora da amostra para previsões com um passo à frente?
- (iv) Adicione variáveis *dummy* mensais à regressão da parte (i). Elas são conjuntamente significantes? (Não se preocupe com a leve correção serial nos erros dessa regressão quando estiver fazendo o teste conjunto).

**18.17** Utilize os dados contidos no arquivo FERTIL3.RAW para fazer este exercício.

- (i) Trace  $tgf$  contra o tempo. Ela contém uma clara tendência de alta ou de baixa ao longo de todo o período da amostra?
- (ii) Usando os dados até 1979, estime um modelo de tendência temporal cúbica de  $tgf$  (isto é, faça a regressão de  $tgf$  sobre  $t$ ,  $t^2$ , e  $t^3$ , juntamente com um intercepto). Comente sobre o  $R$ -quadrado da regressão.
- (iii) Usando o modelo da parte (ii), compute o erro absoluto médio dos erros de previsão com um passo à frente dos anos de 1980 a 1984.
- (iv) Utilizando os dados até 1979, faça a regressão de  $\Delta tgf_t$  somente sobre uma constante. A constante é estatisticamente diferente de zero? Faz sentido assumir que qualquer termo de tendência será zero, se assumirmos que  $tgf_t$  segue um passeio aleatório?
- (v) Agora, faça a previsão de  $tgf$  para os anos de 1980 até 1984, usando um modelo de passeio aleatório: a previsão de  $tgf_{n+1}$  será simplesmente  $tgf_n$ . Encontre o EAM. Como ele se compara com o EAM da parte (iii)? Qual método de previsão você prefere?
- (vi) Agora, estime um modelo AR(2) para  $tgf$ , novamente usando os dados somente até 1979. A segunda defasagem é significativa?
- (vii) Obtenha o EAM de 1980 até 1984, usando o modelo AR(2). Esse modelo mais generalizado funciona melhor fora da amostra do que o modelo de passeio aleatório?

**18.18** Utilize os dados contidos no arquivo CONSUMP.RAW para fazer este exercício.

- (i) Seja  $y_t$  a renda disponível *per capita* real. Use os dados até 1989 para estimar o modelo

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

e descreva os resultados da forma habitual.

- (ii) Use a equação estimada na parte (i) para fazer a previsão de  $y$  em 1990. Qual é o erro de previsão?
- (iii) Compute o erro absoluto médio das previsões com um passo à frente de 1990, usando os parâmetros estimados na parte (i).
- (iv) Agora, compute o EAM ao longo do mesmo período, mas elimine  $y_{t-1}$  da equação. É melhor, ou não, incluir  $y_{t-1}$  na equação?

**18.19** Utilize os dados contidos no arquivo INTQRT.RAW para fazer este exercício.

- (i) Usando os dados de todos os anos, exceto os últimos quatro (16 trimestres), estime um modelo AR(1) de  $\Delta r6_t$ . (Usamos a diferença pois parece que  $r6_t$  tem uma raiz unitária.)

Encontre a REQM das previsões com um passo à frente da  $\Delta r6$ , usando os últimos 16 trimestres.

- (ii) Agora, adicione o termo de correção de erro  $spr_{t-1} = r6_{t-1} - r3_{t-1}$  na equação da parte (i). (Isso assume que o parâmetro de co-integração é um.) Compute o REQM dos últimos 16 trimestres. O termo de correção de erro auxilia na previsão fora da amostra neste caso?
- (iii) Agora, estime o parâmetro de co-integração, em vez de defini-lo como um. Use os 16 últimos trimestres novamente para produzir a REQM fora da amostra. Como isso se compara com as previsões das partes (i) e (ii)?
- (iv) Suas conclusões seriam outras se você quisesse prever  $r6$  em vez de  $\Delta r6$ ? Explique.

**18.20** Utilize os dados contidos no arquivo VOLAT.RAW para fazer este exercício.

- (i) Confirme que  $lsp500 = \log(sp500)$  e  $lip = \log(ip)$  parecem conter raízes unitárias. Use os testes de Dickey-Fuller com quatro mudanças defasadas e faça os testes com e sem uma tendência temporal linear.
- (ii) Compute uma regressão simples de  $lsp500$  sobre  $lip$ . Comente sobre os tamanhos da estatística  $t$  e do  $R$ -quadrado.
- (iii) Use os resíduos da parte (ii) para testar se  $lsp500$  e  $lip$  são co-integrados. Use o teste padrão de Dickey-Fuller e o teste de Dickey-Fuller aumentado (DFA) com duas defasagens. Qual sua conclusão?
- (iv) Adicione uma tendência temporal linear na regressão da parte (ii) e agora faça o teste de co-integração usando os mesmos testes da parte (iii).
- (v) É possível afirmar que os preços das ações e a atividade econômica real têm uma relação de equilíbrio de longo prazo?

**18.21** Este exercício também utiliza os dados do arquivo VOLAT.RAW. O Exercício em Computador 18.20 estuda a relação de longo prazo entre os preços das ações e a produção industrial. Aqui, você estudará a questão da causalidade de Granger usando as mudanças percentuais.

- (i) Estime um modelo AR(4) de  $pcip_t$ , a mudança percentual na produção industrial (descrita como uma taxa anualizada). Mostre que a segunda e terceira defasagens são conjuntamente significantes ao nível de 2,5%.
- (ii) Adicione uma defasagem de  $pcsp_t$  na equação estimada na parte (i). A defasagem é estatisticamente significativa? O que isso lhe diz sobre a causalidade de Granger entre o crescimento da produção industrial e o crescimento dos preços das ações?
- (iii) Refaça a parte (ii) mas obtenha uma estatística  $t$  robusta em relação à heteroscedasticidade. O teste robusto altera suas conclusões da parte (ii)?

**18.22** Utilize os dados contidos no arquivo TRAFFIC2.RAW para fazer este exercício. Esses dados mensais, sobre acidentes de trânsito na Califórnia entre os anos de 1981 a 1989, foram usados no Exercício em Computador 10.17.

- (i) Usando a regressão padrão de Dickey-Fuller, verifique se  $ltotacc_t$  possui uma raiz unitária. Você pode rejeitar uma raiz unitária no nível de 2,5%?

- (ii) Agora, adicione duas mudanças defasadas ao teste da parte (i) e compute o teste Dickey-Fuller aumentado. Qual sua conclusão?
- (iii) Adicione uma tendência temporal linear à regressão DFA da parte (ii). Agora o que acontece?
- (iv) Dadas as constatações das partes (i) a (iii), o que você diria que é a melhor caracterização de  $ltotacc_t$ : um processo I(1) ou um processo I(0) em torno de uma tendência temporal linear?
- (v) Teste a porcentagem de fatalidades,  $prcfat_t$ , para a existência de uma raiz unitária, usando duas defasagens em uma regressão DFA. Neste caso, importa se você incluir uma tendência temporal linear?