

Exercícios em Computador

CAPÍTULO 1

- 1.1 Utilize os dados contidos em WAGE1.RAW para fazer este exercício.
- (i) Encontre o nível médio de escolaridade na amostra. Quais são o maior e o menor número de anos de educação formal?
 - (ii) Encontre o salário médio por hora na amostra. Ele parece ser alto ou baixo?
 - (iii) Os dados sobre salários estão relatados em dólares de 1976. Usando o Economic Report of the President (de 2004 ou posterior) obtenha e relate o Índice de Preços ao Consumidor (IPC) para os anos de 1976 e 2003.
 - (iv) Use os valores do IPC da parte (iii) para encontrar a média de salário médio em dólares de 2003. E agora, o salário médio por hora parece razoável?
 - (v) Quantas mulheres estão na amostra? Quantos homens?
- 1.2 Use os dados de BWGHT.RAW para responder a esta pergunta.
- (i) Quantas mulheres estão na amostra, e quantas informam terem fumado durante a gravidez?
 - (ii) Qual é a média de cigarros fumados por dia? A média é um bom índice da mulher “típica” neste caso? Explique.
 - (iii) Entre as mulheres que fumaram durante a gravidez, qual é a média de cigarros fumados por dia? Como isso se compara com sua resposta da parte (ii), e por quê?
 - (iv) Encontre a média de *educp* na amostra. Por que somente 1.192 observações são usadas para calcular a média?
 - (v) Relate a renda familiar média e seu desvio-padrão em dólares.
- 1.3 Os dados contidos no MEAP01.RAW são do Estado de Michigan do ano de 2001. Use esses dados para responder às seguintes perguntas.
- (i) Encontre o maior e o menor valores de *mate4*. A faixa faz sentido? Explique.
 - (ii) Quantas escolas têm uma taxa de aprovação perfeita do exame de matemática? Qual é a percentagem disso da amostra total?
 - (iii) Quantas escolas têm taxa de aprovação em matemática de exatamente 50 por cento?

- (iv) Compare as médias de aprovação em matemática e da nota de leitura. Qual dos exames é mais difícil de passar?
 - (v) Encontre a correlação entre *mate4* e *read4*. Qual sua conclusão?
 - (vi) A variável *gpa* representa o gasto por aluno. Encontre a média do *gpa* juntamente com seu desvio-padrão. Você diria que existe uma ampla variação no gasto por aluno?
 - (vii) Suponha que a Escola A gaste 6.000 dólares por aluno e a Escola B gaste 5.500 dólares por aluno. Em que percentagem os gastos da Escola A excedem os da Escola B? Compare isto com $100 \cdot [\log(6.000) - \log(5.500)]$, que é a estimativa da diferença percentual baseada na diferença nos logs naturais. (Veja Seção A.4 no Apêndice A).
- 1.4** Os dados contidos no JTRAIN2.RAW são originários de um experimento de treinamento de pessoal conduzido para homens de baixa renda durante 1976-1977; veja Lalonde (1986).
- (i) Use a variável indicadora *train* para determinar a fração de homens que estavam recebendo treinamento de pessoal.
 - (ii) A variável *gr78* representa os rendimentos de 1978 indicados em milhares de dólares de 1982. Encontre as médias *gr78* da amostra de homens que estavam recebendo treinamento de pessoal e da amostra de homens que não estavam recebendo treinamento de pessoal. A diferença é, economicamente, grande?
 - (iii) A variável *desemp78* é um indicador de se um homem estava ou não desempregado em 1978. Que percentagem de homens que receberam treinamento pessoal estava desempregada? E quanto aos homens que não receberam treinamento pessoal? Comente sobre a diferença.
 - (iv) Das partes (ii) e (iii) parece que o programa de treinamento de pessoal foi efetivo? O que tornaria nossas conclusões mais convincentes?

CAPÍTULO 2

2.1 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 percentagem 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

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

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?

2.2 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)? Qual é a permanência mais longa como CEO?
- (iii) Estime o modelo de regressão simples

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

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

2.3 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

$$\text{dormir} = \beta_0 + \beta_1 \text{trabtot} + u,$$

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?

2.4 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?

2.5 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.

2.6 Usamos os dados contidos no MEAP93.RAW para o exemplo 2.12. Agora queremos explorar o relacionamento entre o percentual de alunos aprovados no exame de matemática (*Mate10*) e o gasto por aluno (*gasto*).

- (i) Você acha que cada dólar adicional gasto tem o mesmo efeito na taxa de aprovação, ou um efeito decrescente parece mais apropriado? Explique.
- (ii) No modelo populacional

$$mate10 = \beta_0 + \beta_1 \log(gasto) + u$$

argumente que $\beta_1/10$ é o ponto percentual de alteração em *mate10* dado um aumento de 10% em *gasto*.

- (iii) Utilize os dados no MEAP93.RAW para estimar o modelo da parte (ii). Relate a equação estimada da maneira habitual, inclusive o tamanho da amostra e o *R*-quadrado.
- (iv) Qual a magnitude do efeito estimado dos *gasto*? Ou seja, se os gastos aumentarem em 10%, qual será o ponto percentual de aumento na *mate10*?
- (v) Pode-se reear que a análise de regressão poderá produzir valores ajustados da *mate10* que serão maiores que 100. Por que isso não é muito preocupante neste conjunto de dados?

2.7 Utilize os dados de CHARITY.RAW [obtido de Franses e Paap (2001)] para responder às seguintes questões:

- (i) Qual a média de doação na amostra de 4.268 pessoas (em florins holandeses)? Que porcentagem de pessoas não fez doação?
- (ii) Qual é a média de mala direta por ano? Quais são os valores mínimo e máximo?
- (iii) Estime o modelo:

$$doa = \beta_0 + \beta_1 malaano + u$$

pelos MQO e relate os resultados da maneira habitual, incluindo o tamanho da amostra e o *R*-quadrado.

- (iv) Interprete o coeficiente de inclinação. Se cada mala direta custar um florin, é esperado que a instituição beneficente tenha ganho com cada mala direta? Isto significa que a instituição beneficente tem ganho em todas as malas diretas? Explique.
- (v) Qual é a menor doação predita na amostra? Usando esta análise de regressão simples, é possível predizer zero como doação?

CAPÍTULO 3

3.1 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 fatores. 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

$$pesonas = \beta_0 + \beta_1 cigs + \beta_2 rendfam + u.$$

- (i) Qual é o sinal mais provável de β_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*.

3.2 Use os dados em HPRICE1.RAW para estimar o modelo

$$\text{preço} = \beta_0 + \beta_1 m\text{quad} + \beta_2 \text{banhos} + u,$$

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?

3.3 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 *lucros*. Essas variáveis são altamente correlacionadas? O que isso diz sobre os estimadores MQO?

3.4 Use os dados em ATTEND.RAW para esse exercício.

- (i) Obtenha os valores mínimo, máximo e médio das variáveis *taxafreq*, *supGPA* e *ACT*.
- (ii) Estime o modelo

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

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 *taxafreq* predita se $\text{supGPA} = 3,65$ 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 $\text{supGPA} = 3,1$ e $\text{ACT} = 21$, e Estudante B tem $\text{supGPA} = 2,1$ e $\text{ACT} = 26$, qual é a diferença predita em suas taxas de frequência?

3.5 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 *educ* sobre *exper* e *perm* e salve os resíduos, \hat{r}_1 . Então, regreda $\log(\text{salário})$ sobre \hat{r}_1 . Compare o coeficiente de \hat{r}_1 com o coeficiente de *educ* da regressão de $\log(\text{salário})$ sobre *educ*, *exper* e *perm*.

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

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

3.7 Use os dados de MEAP93.RAW para responder a estas perguntas.

- (i) Estime o modelo

$$\text{mate10} = \beta_0 + \beta_1 + \log(\text{gasto}) + \beta_2 \text{prgalm} + u$$

e relate os resultados da maneira habitual, inclusive o tamanho da amostra e o *R*-quadrado. Os sinais dos coeficientes de inclinação são os que você esperava?

- (ii) O que você conclui do intercepto que você estimou na parte (i)? Particularmente, faz sentido definir as duas variáveis explicativas com valor zero? [*Sugestão*: Lembre-se que $\log(1) = 0$].
- (iii) Agora execute a regressão simples de *mate10* sobre *gasto*, e compare o coeficiente de inclinação com a estimativa obtida na parte (i). O efeito do *gasto* estimado agora é maior ou menor do que na parte (i)?
- (iv) Encontre a correlação entre $\lgasto = \log(\text{gasto})$ e *prgalm*. O seu sinal faz sentido para você?
- (v) Use a parte (iv) para explicar suas constatações na parte (iii).

3.8 Use os dados de DISCRIM.RAW para responder a esta questão. Estes são dados em nível de Código de Endereçamento Postal (CEP) sobre preços de vários itens em restaurantes fast-food juntamente com as características da população do CEP, em Nova Jersey e Pensilvânia. A ideia é verificar se os fast-foods cobram preços maiores em áreas com uma grande concentração de negros.

- (i) Encontre os valores médios de *propnegr* e *renda* na amostra, juntamente com seus desvios-padrão. Quais são as unidades de medida de *propnegr* e de *renda*?
- (ii) Considere um modelo para explicar o preço do refrigerante, *prrefr*, em termos da proporção da população que é negra e com renda média.

$$prref = \beta_0 + \beta_1 propnegr + \beta_2 renda + u$$

Estime este modelo pelos MQO e relate os resultados em forma de equação, inclusive o tamanho da amostra e o *R*-quadrado. (Não use notação científica quando relatar as estimativas). Interprete o coeficiente na *propnegr*. Você acha que ele é economicamente grande?

- (iii) Compare a estimativa da parte (ii) com a estimativa da regressão simples de *prrefr* sobre *propnegr*. O efeito de discriminação é maior ou menor quando você controla a renda?
- (iv) Um modelo com uma elasticidade de preço constante com relação à renda pode ser mais apropriado. Relate as estimativas do modelo

$$\log(prrefr) = \beta_0 + \beta_1 + \log(gasto) + \beta_2 prgalm + u$$

Se *propnegr* aumentar em 0,20 (vinte pontos percentuais), qual será a porcentagem estimada de alteração na *prrefr*? (*Sugestão*: A resposta é 2.xx, em que você preenche a “xx”).

- (v) Agora, adicione a variável *propor* à regressão na parte (iv). O que acontece com $\hat{\beta}_{propnegr}$?
- (vi) Encontre a correlação entre $\log(renda)$ e *propor*. Ela é mais ou menos o que você esperava?
- (vii) Avalie a seguinte declaração: “Como $\log(renda)$ e *propor* são tão altamente correlacionadas, não tem nada a ver estarem na mesma regressão”.

3.9 Utilize os dados contidos em CHARITY.RAW para responder às seguintes questões:

- (i) Estime a equação

$$doa = \beta_0 + \beta_1 malaano + \beta_2 + ultdoa + proprest + u$$

pelos MQO e relate os resultados da maneira habitual, incluindo o tamanho da amostra e o *R*-quadrado. Como o *R*-quadrado se compara com o da regressão simples que omite *ultdoa* e *proprest*?

- (ii) Interprete o coeficiente na *malaano*. Ele é maior ou menor que o coeficiente correspondente na regressão simples?
- (iii) Interprete o coeficiente na *proprest*. Seja metódico em observar as unidades de medição da *proprest*.
- (iv) Agora adicione a variável *meddoa* à equação. O que acontece com o efeito estimado de *malaano*?
- (v) Na equação da parte (iv), o que aconteceu com o coeficiente da *ultdoa*? O que você acha que está acontecendo?

CAPÍTULO 4

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

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

em que $votoA$ é a porcentagem de votos recebidos pelo Candidato A, $gastoA$ e $gastoB$ são os gastos de campanha dos Candidatos A e B, e $forpart$ 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 β_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.)

4.2 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*? Certifique-se de considerar os dados faltantes de *LSAT* e *GPA*.
- (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?

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

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

- (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 θ_1 .
- (ii) Escreva β_2 em termos de θ_1 e β_1 , e coloque isso na equação do $\log(preço)$.
- (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%.

4.4 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.

4.5 Use os dados em `MLB1.RAW` para este exercício.

- (i) Use o modelo estimado na equação (4.31) e retire a variável *rbisyr*. 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*.

4.6 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?

4.7 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 do 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. Qual é o *p*-valor bilateral?

4.8 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 1%?
- (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?

4.9 Use os dados de DISCRIM.RAW para responder a esta questão. (Veja também Exercício em Computador 3.8 no Capítulo 3)

- (i) Use os MQO para estimar o modelo

$$\log(prrefr) = \beta_0 + \beta_1 propneger + \beta_2 \log(renda) + u,$$

e relate os resultados da forma habitual. $\hat{\beta}_1$ é estatisticamente diferente de zero no nível de 5% contra uma alternativa bilateral? E no nível de 1%?

- (ii) Qual é a correlação entre $\log(renda)$ e *propor*? Cada variável é estatisticamente significativa em qualquer caso? Relate os p -valores bilaterais.
- (iii) Adicione a variável $\log(valhab)$ à regressão na parte (i). Interprete seu coeficiente e relate o p -valor bilateral de $\beta_{\log(valhab)} = 0$.
- (iv) Na regressão na parte (iii), o que acontece com a significância estatística individual de $\log(renda)$ e *propor*? Essas variáveis são conjuntamente significantes? (Compute um p -valor). O que você deduz de suas respostas?
- (v) Dado os resultados das regressões anteriores, qual delas você diria ser a mais confiável na determinação de se a maquiagem racial do código de endereçamento posta influência nos preços dos fast-foods locais?

4.10 Utilize os dados de ELEM94_95 para responder a estas questões. Os resultados poderão ser comparados com os da Tabela 4.1. A variável dependente *lsalmed* é o log do salário médio do professor e *bs* é a razão da média de benefícios e o salário médio (por escola).

- (i) Execute a regressão simples de *lsalmed* sobre *bs*. A inclinação estimada é estatisticamente diferente de zero? Ela é estatisticamente diferente de -1 ?
- (ii) Adicione a variável *lmatricl* e *lstaff* à regressão da parte (i). O que acontece com o coeficiente na *bs*? Como a situação se compara com a da Tabela 4.1?
- (iii) Por que o erro-padrão no coeficiente da *bs* é menor na parte (ii) do que na parte (i)? (Sugestão: O que acontece com a variância do erro *versus* a multicolinearidade quando *lmatricl* e *lstaff* são adicionadas?)
- (iv) Por que o coeficiente na *lstaff* é negativo? Ele é grande em magnitude?
- (v) Agora adicione a variável *merenda* à regressão. Mantendo os outros fatores fixos, os professores estão sendo recompensados por darem aulas a estudantes de perfil desfavorável? Explique.
- (vi) No geral, o padrão dos resultados que você encontra no ELEM94_95.RAW é consistente com o padrão na Tabela 4.1?

CAPÍTULO 5

5.1 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?

5.2 Use os dados em GPA2.RAW para este exercício.

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

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

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).

5.3 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).

5.4 Várias estatísticas são comumente usadas para detectar não normalidade em distribuições populacionais subjacentes. Aqui estudaremos uma que avalia a magnitude da simetria em uma distribuição. Recorde que qualquer variável aleatória normalmente distribuída é simétrica em relação à sua média; portanto, se padronizarmos uma variável aleatória simetricamente distribuída, digamos, $z = (y - \mu_y)/\sigma_y$, em que $\mu_y = E(y)$ e $\sigma_y = \text{dp}(y)$, então z terá média zero, variância um, e $E(z^3) = 0$. Dada uma amostra de dados $\{y_i : i = 1, \dots, n\}$, podemos padronizar y_i na amostra com o uso de $z_i = (y_i - \hat{\mu}_y)/\hat{\sigma}_y$, em que $\hat{\mu}_y$ é a média da amostra e $\hat{\sigma}_y$ é o desvio-padrão amostral. (Ignoramos o fato de que estas são estimativas baseadas na amostra.) Uma estatística amostral que avalia a simetria é $n^{-1} \sum_{i=1}^n z_i^3$, ou em que n é substituída por $(n - 1)$ como um ajuste dos graus de liberdade. Se y tiver uma distribuição normal na população, o indicador de simetria na amostra dos valores padronizados não deve diferir significativamente de zero.

- (i) Utilize o conjunto de dados 401KSUBS.RAW, mantendo somente as observações com *tamfam* = 1. Encontre os indicadores de simetria de renda. Faça o mesmo com $\log(\text{renda})$. Qual variável tem mais simetria e, portanto, parece ser menos provável de ser normalmente distribuída?
 (ii) Depois, utilize BWGHT2.RAW. Encontre os indicadores de simetria de *pesonasc* e de $\log(\text{pesonasc})$. Qual a sua conclusão?
 (iii) Avalie a seguinte afirmação: “A transformação logarítmica sempre faz com que uma variável positiva pareça mais normalmente distribuída”.

- (iv) Se estivermos interessados na hipótese de normalidade no contexto da regressão, deveremos avaliar a distribuição incondicional de y e de $\log(y)$. Explique.

CAPÍTULO 6

6.1 Utilize os dados contidos no arquivo KIELMC.RAW, somente do ano de 1981, para responder às 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

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

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

6.2 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 exper^2 é estatisticamente significativa no nível de 1%?
- (iii) Utilizando a aproximação

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

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?

6.3 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 θ_1 representar o retorno da educação (na forma decimal), quando *exper* = 10: $\theta_1 = \beta_1 + 10\beta_3$. Obtenha $\hat{\theta}_1$ e um intervalo de confiança de 95% de θ_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 θ_1 .)

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

- (i) Estime o modelo

$$\text{SAT} = \beta_0 + \beta_1 \text{tamclas} + \beta_2 \text{tamclas}^2 + u,$$

em que *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 $\log(\text{SAT})$ como a variável dependente. Ele é muito diferente do que você obteve na parte (ii)?

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

- (i) Estime o modelo

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

e descreva os resultados no formato MQO habitual.

- (ii) Encontre o valor previsto de $\log(\text{preço})$, 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

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

6.6 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 β_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.

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

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

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

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

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

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

em que $\theta_2 = \beta_2 + 2\beta_4(2,59) + \beta_6(0,82)$. (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 *supGPA(taxafreq - 0,82)*, você use *(supGPA - 2,59) · (taxafreq - 0,82)*. Agora, como você interpreta o coeficiente da *taxafreq* e da *supGPA*?

6.8 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.

6.9 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*, $exper^2$, *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?

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

- (i) Estime a equação

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

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 $\log(pesonas)$ 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 $\log(\textit{pesonas})$?
- (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*.

6.11 Use o APPLE.RAW para verificar algumas das afirmações feitas na Seção 6.3.

- (i) Execute a regressão de *ecolbs* sobre *ecoprc*, *regprc* e relate os resultados da forma normal, inclusive o R-quadrado e o R-quadrado ajustado. Interprete os coeficientes nas variáveis de preço e comente sobre seus sinais e magnitudes.
- (ii) As variáveis de preço são estatisticamente significantes? Relate os *p*-valores dos testes *t* individuais.
- (iii) Qual é a faixa dos valores ajustados da *ecolbs*? Que fração da amostra relata *ecolbs* = 0? Comente.
- (iv) Você acha que as variáveis de preço juntas fazem um bom trabalho em explicar a variação na *ecolbs*? Explique.
- (v) Adicione a variável *rendfam*, *tamfam* (tamanho da família), *educ* e *idade* na regressão da parte (i). Encontre o *p*-valor de sua significância conjunta. Qual sua conclusão?

6.12 Use o subconjunto 401KSUBS.RAW com *tamfam* = 1; isto restringirá a análise em casas com uma só pessoa; veja também Exercício em Computador 4.8.

- (i) Qual é a idade da pessoa mais jovem nesta amostra? Quantas pessoas têm essa idade?
- (ii) No modelo

$$\textit{atfinliq} = \beta_0 + \beta_1 \textit{renda} + \beta_2 \textit{idade} + \beta_3 \textit{idade}^2 + u$$

qual é a interpretação literal de β_2 ? Por si própria, ela é de muito interesse?

- (iii) Estime o modelo da parte (ii) e relate os resultados na forma-padrão. Você se preocupa com o fato de que o coeficiente na *idade* seja negativo? Explique.
- (iv) Como a pessoa mais jovem na amostra tem 25 anos, faz sentido pensar que, em determinado nível de renda, a média mais baixa de ativos financeiros líquidos está na idade de 25 anos. Lembre-se que o efeito parcial de *idade* sobre *atfinliq* é $\beta_2 + 2\beta_3 \textit{idade}$, assim o efeito parcial na idade de 25 anos é $\beta_2 + 2\beta_3(25) = \beta_2 + 50\beta_3$; chame isto de θ_2 . Encontre $\hat{\theta}_2$ e obtenha o *p*-valor bilateral do teste $H_0: \theta_2$. Você deve concluir que $\hat{\theta}_2$ é menor e bastante significativo estatisticamente. [Sugestão: uma maneira de fazer isso é estimar o modelo $\textit{netffa} = \alpha_0 + \beta_1 \textit{renda} + \theta_2 \textit{idade} + \beta_3(\textit{idade} - 25)^2 + u$, em que o intercepto, α_0 , é diferente de β_0 . Existem, também, outras maneiras.]
- (v) Como a evidência contra $H_0: \theta_2 = 0$ é muito fraca, defina-a como zero e estime o modelo

$$\textit{atfinliq} = \alpha_0 + \beta_1 \textit{renda} + \beta_3(\textit{idade} - 25)^2 + u.$$

Em termos de qualidade de ajuste, este modelo é melhor que o da parte (ii)?

- (vi) Na equação na parte (v), defina $renda = 30$ (aproximadamente o valor médio) e trace o relacionamento entre $atfinliq$ e $idade$, mas somente para $idade \geq 25$. Descreva o que você vê.
- (vii) Verifique se a inclusão de um quadrático na renda é necessário.

6.13 Utilize os dados contidos no arquivo MEAP00_01 para responder a estas questões.

- (i) Estime o modelo

$$mate4 = \beta_0 + \beta_2 lgpa + \beta_2 lmatricl + \beta_2 merenda + u$$

pelos MQO, e relate os resultados da forma habitual. É estatisticamente significativo, no nível de 5%, cada uma das variáveis explicativas?

- (ii) Obtenha os valores ajustados da regressão na parte (i). Qual é a faixa dos valores ajustados? Como ela se compara com a faixa dos dados efetivos da $mate4$?
- (iii) Obtenha os resíduos da regressão na parte (i). Qual é o código de edificação da escola que tem o maior resíduo (positivo)? Forneça uma interpretação deste resíduo.
- (iv) Adicione quadráticos de todas as variáveis explicativas na equação, e teste-os quanto à significância conjunta. Você os deixaria no modelo?
- (v) Retornando ao modelo na parte (i), divida a variável dependente e cada uma das variáveis explicativas por seus desvios-padrão amostral, e execute novamente a regressão. (Inclua um intercepto, a menos que você também subtraia a média de cada variável.) Em termos de unidades de desvio-padrão, qual variável explicativa tem o maior efeito na taxa de aprovação em matemática?

CAPÍTULO 7

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

- (i) Adicione as variáveis $maesup$ e $paissup$ 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 $paissup$ na equação da parte (i) e assegure-se de descrever o p -valor.
- (iii) Adicione $EmGPA^2$ ao modelo da parte (i) e decida se essa generalização é necessária.

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

- (i) Estime o modelo

$$\begin{aligned} \log(\text{salário}) = & \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 perm + \beta_4 casado \\ & + \beta_5 negro + \beta_6 sul + \beta_7 urbano + u \end{aligned}$$

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?

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

$$\begin{aligned} \log(\text{salário}) = & \beta_0 + \beta_1 \text{anos} + \beta_2 \text{jogosano} + \beta_3 \text{rebmed} + \beta_4 \text{hrunano} \\ & + \beta_5 \text{rebrunano} + \beta_6 \text{runsano} + \beta_7 \text{perccap} + \beta_8 \text{porcest} \\ & + \beta_9 \text{pribase} + \beta_{10} \text{segbase} + \beta_{11} \text{terbase} + \beta_{12} \text{interbase} + \beta_{13} \text{receptor} + u, \end{aligned}$$

em que 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.

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

- (i) Considere a equação

$$\text{SupGPA} = \beta_0 + \beta_1 \text{tamclas} + \beta_2 \text{tamclas}^2 + \beta_3 \text{emperc} + \beta_4 \text{SAT} + \beta_5 \text{feminino} + \beta_6 \text{atleta} + u,$$

em que SupGPA 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 SupGPA difere por gênero? Justifique sua resposta.

7.5 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 $\hat{\beta}_3$.

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

$$\text{dormir} = \beta_0 + \beta_1 \text{trabt} + \beta_2 \text{educ} + \beta_3 \text{idade} + \beta_4 \text{idade}^2 + \beta_5 \text{crianmen} + u.$$

- (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*·*trabt*, ..., *masculino*·*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?

7.7 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*·(*educ* − 12,5) substituindo *feminino*·*educ*. Como você interpreta o coeficiente de *feminino* nesse caso?
- (iii) O coeficiente de *feminino* na parte (ii) é estatisticamente significativo? Compare com a equação (7.18) e comente.

7.8 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:

$$\text{aprovado} = \beta_0 + \beta_1 \text{branco} + \text{outros fatores}.$$

- (i) Se houver discriminação contra as minorias, e os fatores apropriados tiverem sido controlados, qual será o sinal de β_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.

7.9 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 401(k) (*e401k*) e várias outras variáveis.

- (i) Que fração das famílias na amostra se classifica para participar de uma conta de um fundo de pensão 401(k)?
- (ii) Estime um modelo de probabilidade linear explicando a qualificação para ter uma conta em um plano de pensão 401(k), 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 plano de pensão 401(k) 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) Usando os valores ajustados $\widehat{e401k}_i$ da parte (iv), defina $\widehat{e401k}_i = 1$ se $2\widehat{e401k}_i \geq 0,5$ e $\widehat{e401k}_i = 0$ se $2\widehat{e401k}_i < 0,5$. Das 9.275 famílias, quantas estão preditas como qualificadas para um plano de pensão 401k?
- (vi) Das 5.638 famílias não qualificadas para um plano de pensão 401k, que percentagem está predita que não terá um plano 401k, usando o preditor $\widehat{e401k}_i$? Das 3.637 famílias qualificadas para um plano 401k, que percentagem está predita que terá um? (será útil se seu pacote econométrico tiver o comando “tabular”).
- (vii) A porcentagem global corretamente predita está em torno de 64,9%. Você acha que isto é uma descrição completa da eficiência do modelo, considerando suas respostas na parte (vi)?
- (viii) 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, quanto será mais elevada a probabilidade estimada de que a família será qualificada para um plano de pensão 401(k)? Ela é estatisticamente diferente de zero ao nível de 10%?

7.10 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 *dummies* 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?
- (vi) 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.

7.11 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; 401(k) use uma alternativa bilateral. Qual é o montante em dólares da diferença estimada?
- (iii) Considerando a parte (ii) do Exercício 7.9, 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 401(k)?
- (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 401(k) 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 401(k) 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 quadrados dos resíduos restrita, SQR_r , é obtida da parte (vi) porque aquela regressão assume que todas as inclinações são as mesmas. A soma dos quadrados dos resíduos irrestrita é $SQR_{ir} = SQR_1 + SQR_2 + \dots + SQR_5$, em que SQR_j é a soma dos quadrados

dos resíduos 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 $9.275 - 30 = 9.245$.

7.12 Utilize o conjunto de dados contido no BEAUTY.RAW, que contém um subconjunto das variáveis (mas com observações mais úteis do que nas regressões relatadas por Hamermesh and Biddle (1994).

- (i) Encontre, separadamente, as frações de homens e mulheres que estão classificadas como tendo aparência acima da média. O número maior de pessoas está classificado como tendo a aparência acima ou abaixo da média?
- (ii) Teste a hipótese nula de que as frações da população de homens e mulheres com aparência acima e abaixo da média são as mesmas. Relate o p -valor unilateral de que a fração de mulheres é mais alta. (*Sugestão*: Estimar um modelo de regressão linear simples é mais fácil).
- (iii) Agora estime o modelo

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{abaixmed} + \beta_2 \text{acimmed} + u$$

separadamente para homens e mulheres, e relate os resultados da forma habitual. Em ambos os casos, interprete o coeficiente na *abaixmed*. Explique com suas próprias palavras o que a hipótese $H_0: \beta_1 = 0$ contra $H_1: \beta_1 < 0$ significa, e encontre os p -valores de homens e mulheres.

- (iv) Existe evidência convincente que mulheres com aparência acima da média ganham mais que mulheres com aparência na média? Explique.
- (v) Tanto para homens como para mulheres, adicione as variáveis explicativas *educ*, *exper*, *exper*², *sindicato*, *boasaúde*, *negro*, *casado*, *sul*, *cidadegr*, *cidadepeq* e *serviço*. Os efeitos das variáveis de “aparência” mudam de maneiras importantes?

7.13 Use os dados contidos em APPLE.RAW para responder a esta questão.

- (i) Defina uma variável binária *compreco* = 1 se *ecolbs* > 0 e *compreco* = 0 se *ecolbs* = 0. Em outras palavras, *compreco* indica se, aos preços fornecidos, uma família compraria maçãs com selo ecológico. Que fração de famílias declara que comprariam maçãs com selo ecológico?
- (ii) Estime o modelo de probabilidade linear

$$\text{compreco} = \beta_0 + \beta_1 \text{ecoprc} + \beta_2 \text{regprc} + \beta_3 \text{rendfami} + \beta_4 \text{tamfam} + \beta_5 \text{educ} + \beta_6 \text{idade} + u$$

e relate os resultados da forma habitual. Cuidadosamente interprete os coeficientes nas variáveis de preço

- (iii) As variáveis sem preço são conjuntamente significantes no MPL? (Use a habitual estatística F , embora ela não seja válida quando existe heteroscedasticidade). Qual variável explicativa exceto as variáveis de preço parece ter o efeito mais importante na decisão de comprar maçãs com selo ecológico? Isso faz sentido para você?
- (iv) No modelo da parte (ii), substitua *rendfami* por $\log(\text{rendfami})$. Qual modelo se encaixa melhor nos dados, usando *rendfami* ou $\log(\text{rendfami})$?

- (v) Na estimação na parte (iv), quantas probabilidades estimadas são negativas? Quantas são maiores que um? Você deveria se preocupar com isso?
- (vi) Na estimação na parte (iv), compute a porcentagem corretamente predita de cada resultado, $compreco = 0$ e $compreco = 1$. Qual resultado é mais bem predito pelo modelo?

7.14 Utilize os dados contidos em CHARITY.RAW para responder a esta questão. A variável *resp* é uma variável *dummy* igual a um se uma pessoa deu retorno com uma doação à mais recente mala direta enviada pela organização caritativa. A variável *respultima* uma variável *dummy* igual a um se a pessoa deu retorno à mala direta anterior, *meddoa* é a média das doações anteriores (em Florins holandeses), e *propresp* é a proporção de vezes em que a pessoa deu retorno às malas diretas anteriores.

- (i) Estime um modelo de probabilidade linear relacionando *resp* com *respultima* e *meddoa*. Relate o resultado na forma habitual, e interprete o coeficiente na *respultima*.
- (ii) O valor médio das doações anteriores parece afetar a probabilidade de retorno?
- (iii) Adicione a variável *propresp* ao modelo, e interprete seu coeficiente. (Tenha cuidado aqui: um aumento de um na *propresp* é a maior alteração possível.)
- (iv) O que aconteceu com o coeficiente na *respultima* quando *propresp* foi adicionada à regressão? Isto faz sentido?
- (v) Adicione *malaano*, o número de malas diretas por ano. Quão grande é seu efeito estimado? Por que isto pode não ser uma boa estimativa do efeito causal das malas diretas sobre os retornos?

CAPÍTULO 8

8.1 Utilize o modelo a seguir para explicar o comportamento do sono:

$$dormir = \beta_0 + \beta_1 trabtot + \beta_2 educ + \beta_3 idade + \beta_4 idade^2 + \beta_5 crianmen + \beta_6 masculino + u.$$

- (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) Utilize os dados contidos no arquivo SLEEP.75.RAW para estimar 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?
- 8.2**
- (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?

8.3 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?

8.4 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, calcule o teste de heteroscedasticidade de Breusch-Pagan. Use a versão da estatística F e escreva o p -valor.
- (iii) Calcule 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?

8.5 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*, μ , 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?

8.6 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%.

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

- (i) Estime a equação da parte (iii) do Problema 7.8, computando os erros-padrão robustos em relação à heteroscedasticidade. Compare o intervalo de confiança de 95% em β_{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?

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

- (i) Use o método MQO para estimar um modelo relacionando $SupGPA$ com $emGPA$, ACT , $faltas$ 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 \widehat{SupGPA}_i , \widehat{SupGPA}_i^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?

8.9 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.

8.10 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 $renda$, $renda^2$, $idade$, $idade^2$, 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 quadrados dos resíduos de 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 $\text{Var}(y|x_1, x_2, \dots, x_k) = p(\mathbf{x})[1 - p(\mathbf{x})]$, em que $p(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$.}
- (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 de linear. Elas diferem, de maneira importante, das estimativas MQO?

8.11 Utilize os dados contidos na 401KSUBS.RAW para responder a esta questão, restringindo a amostra a $tamfam = 1$.

- (i) Ao modelo estimado na Tabela 8.1, adicione o termo de interação $e401k \cdot renda$. Estime a equação pelos MQO e obtenha os erros-padrão robustos habituais. Qual sua conclusão sobre a significância estatística do termo de interação?
- (ii) Agora estime o modelo mais geral pelos MQP usando os mesmos pesos, $1/renda$, como na Tabela 8.1. Calcule o erro-padrão habitual e o robusto pelo estimador MQP. O termo de interação é estatisticamente significativo usando o erro-padrão robusto?
- (iii) Examine o coeficiente MQP na $e401k$ no modelo mais geral. Ele é de grande interesse por si só? Explique.
- (iv) Estime novamente o modelo pelos MQP mas use o termo de interação $e401k \cdot (renda - 30)$; a média da renda na amostra está em torno de 29,44. Agora interprete o coeficiente na $e401k$.

8.12 Utilize os dados da MEAP00_1.RAW para responder a esta questão.

- (i) Estime o modelo

$$mat4 = \beta_0 + \beta_1 merenda + \beta_2 \log(matricl) + \beta_3 \log(gpa) + u$$

pelos MQO e obtenha os erros-padrão habituais e os erros-padrão totalmente robustos.

Como eles se comparam, de forma geral?

- (ii) Aplique o caso especial do teste de White de heteroscedasticidade. Qual é o valor do teste F ? Qual sua conclusão?
- (iii) Obtenha \hat{g}_i como os valores ajustados da regressão $\log(\hat{u}_i^2)$ sobre $\widehat{mate4}_i, \widehat{mate4}_i^2$, em que $\widehat{mate4}_i$ são os valores ajustados do MQO e as \hat{u}_i são os resíduos dos MQO. Seja $\hat{h}_i = \exp(\hat{g}_i)$. Use a \hat{h}_i para obter as estimativas MQP. Há grandes diferenças com os coeficientes dos MQO?
- (iv) Obtenha os erros-padrão dos MQP que permitam mal-especificação da função de variância. Eles diferem muito dos habituais erros-padrão dos MQP?
- (v) Para se estimar o efeito dos gastos na $mate4$ os MQO ou MQP parecem ser mais precisos?

CAPÍTULO 9

- 9.1** (i) Aplique o teste RESET da equação (9.3) no modelo estimado no Exercício em Computador 7.5. Existe evidência de má-especificação da forma funcional na equação?
- (ii) Calcule uma forma robusta do teste RESET, em relação à heteroscedasticidade. Isso altera a sua conclusão sobre a parte (i)?

9.2 Utilize o conjunto de dados do arquivo WAGE2.RAW para este exercício.

- (i) Use a variável *KWW* (escore de teste do “Conhecimento do mundo do trabalho”) 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?

9.3 Utilize os dados do arquivo JTRAIN.RAW para este exercício.

- (i) Considere o modelo de regressão simples

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

em que *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 $\log(rejei_{87})$. 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 $H_1: \beta_{subs} < 0$?
- (iv) Teste a hipótese nula de que o parâmetro de $\log(rejei_{87})$ 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.

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

- (i) Reestime a equação (9.43), 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 (9.44). O que você conclui sobre a inclusão de uma variável *dummy* para uma única observação?

9.5 Utilize os dados do arquivo RDCHEM.RAW para melhor examinar os efeitos de observações extremas sobre as estimativas MQO e veja como o MDA é menos sensível às observações extremas. O modelo é:

$$pdintens = \beta_0 + \beta_1 vendas + \beta_2 vendas^2 + \beta_3 lucrmarg + u$$

em que você deve primeiro mudar vendas para bilhões de dólares para tornar a estimativa fácil de interpretar.

- (i) Estime a equação acima pelos MQO, tanto com a firma tendo ou não vendas anuais de 40 bilhões de dólares. Comente sobre quaisquer diferenças notáveis nos coeficientes estimados.
- (ii) Estime a mesma equação pelos MDA, de novo com e sem a maior firma. Comente sobre quaisquer diferenças importantes nos coeficientes estimados.
- (iii) Com base em suas constatações na (i) e na (ii), você diria que os MQO ou os MDA é mais resiliente à atipicidade?

9.6 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?

9.7 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 7.8, excluindo as observações com *outrobr* > 40. O que acontece com a estimativa e com a estatística *t* da variável *branco*?
- (iii) A estimativa de β_{branco} aparenta ser demasiadamente afetada pela amostra utilizada?

9.8 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 de graduação de dois anos e 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*² 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.

9.9 Neste exercício você deve comparar as estimativas MQO e MDA dos efeitos da participação no plano 401(k) 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 β_6 .
- (iv) Reconcilie suas constatações das partes (ii) e (iii)

9.10 Você precisará usar dois conjuntos de dados para fazer este exercício, JTRAIN2.RAW e JTRAIN3.RAW. O primeiro é o resultado de um experimento de treinamento de pessoal. O arquivo JTRAIN3.RAW contém dados observacionais, onde os próprios indivíduos, em grande escala, determinam se eles participam de treinamento de pessoal. Os conjuntos de dados cobrem o mesmo período.

- (i) No conjunto de dados JTRAIN2.RAW, qual é a fração de homens que estão recebendo treinamento de pessoal? Qual é a fração na JTRAIN3.RAW? Por que você acha que existe tamanha diferença?
- (ii) Usando o JTRAIN2.RAW execute uma regressão simples de $gr78$ sobre $trein$. Qual é o efeito estimado da participação num treinamento de pessoal sobre ganhos reais?
- (iii) Agora adicione como controles na regressão na parte(ii) as variáveis $gr74$, $gr75$, $educ$, $idade$, $negro$, e $hispan$. O efeito estimado do treinamento de pessoal sobre a $gr78$ altera muito? De que forma? (Dica: Lembre que são dados experimentais).
- (iv) Faça as regressões nas partes (ii) e (iii) usando os dados de JTRAIN3.RAW, relatando somente os coeficientes estimados na $trein$, juntamente com suas estatísticas t . Qual é o efeito agora por termos controlado os fatores extras e por quê?
- (v) Defina $medgr = (gr74 + gr75)/2$. Encontre as médias amostrais, desvios padrões, e os valores mínimos e máximos nos dois conjuntos de dados. Esses conjuntos de dados são representativos das mesmas populações em 1978?
- (vi) Quase 96% dos homens do conjunto de dados JTRAIN2.RAW tem $medgr$ menor que 10.000 dólares

Usando somente esses homens, execute a regressão

$$gr78 \text{ sobre } trein, gr74, gr75, educ, idade, negro, e hispan.$$

e relate a estimativa de treinamento e sua estatística t . Execute a mesma regressão da JTRAIN3.RAW, usando somente homens com $medgr \leq 10$. Da subamostra de homens com baixa renda, como os efeitos estimados de treinamento se comparam ao longo dos conjuntos de dados experimentais e não experimentais?

- (vii) Agora use cada conjunto de dados para executar uma regressão simples de $gr78$ sobre $trein$, mas somente dos homens que estavam desempregados em 1974 e 1975. Como as estimativas de treinamento se comparam agora?
- (viii) Usando suas constatações das regressões anteriores, comente sobre a importância potencial de ter-se populações comparáveis subjacentes de estimativas experimentais e não experimentais.

9.11 Utilize os dados do ano de 1993 para esta questão, embora você vá precisar, primeiro, obter a taxa defasada de homicídios, digamos $txhomi_{-1}$.

- (i) Execute a regressão de $txhomi$ sobre $exec$, $desemp$. Quais são o coeficiente e estatística t na $exec$? Esta regressão fornece alguma evidência de efeito dissuasor da pena de morte?
- (ii) Quantas execuções são relatadas no Texas durante 1993? (Na verdade, esta é a soma das execuções daquele e dos dois anos anteriores.) Como isto se compara com os outros estados? Adicione uma variável simulada para o Texas na regressão da parte (i). A estatística t é geralmente grande? Disto, o Texas aparenta ser uma observação extrema?
- (iii) Adicione a taxa defasada de homicídios na regressão na parte (i). O que acontece com a $\hat{\beta}_{exec}$ e sua significância estatística?
- (iv) Na regressão na parte (iii), o Texas aparenta ser uma observação extrema? Qual é o efeito na $\hat{\beta}_{exec}$ com a eliminação do Texas da regressão?

9.12 Utilize os dados da ELEM94_95 para responder a esta questão. Veja também o Exercício em Computador 4.10.

- (i) Usando todos os dados, execute a regressão de $lsalmed$ sobre bs , $lmatricl$, $lstaff$, e $merenda$. Relate o coeficiente na bs juntamente com seus erros-padrão habituais e de heteroscedasticidade robusta. O que você conclui sobre a significância econômica e estatística da $\hat{\beta}_{bs}$?
- (ii) Agora elimine as quatro observações com $bs > 0,5$, isto é, em que os benefícios médios são (supostamente) mais de 50% do salário médio. Qual é o coeficiente na bs ? Ele é estatisticamente significativo usando os erros-padrão de heteroscedasticidade robusta?
- (iii) Verifique que as quatro observações com $bs > 0,5$ são 68, 1.127, 1.508. e 1.670. Defina quatro variáveis *dummy* para cada uma dessas observações. (Você pode chamá-las $d68$, $d1127$, $d1508$ e $d1670$.) Adicione-as à regressão da parte (i), e verifique que os coeficientes e erros-padrão dos MQO nas outras variáveis são idênticos aos da parte (ii) Quais das quatro *dummies* têm uma estatística t estatisticamente diferente de zero em um nível de 5%?
- (iv) Verifique que, neste conjunto de dados, o ponto de dado com o maior resíduo estudentizado (maior estatística t na variável *dummy*) na parte (iii) tem uma grande influência nas estimativas MQO. (Isto é, executar os MQO usando todas as observações exceto aquela com o resíduo estudentizado grande.) A eliminação, uma por uma, cada uma das outras observações com $bs > 0,5$ tem efeitos importantes?
- (v) Qual sua conclusão sobre a sensibilidade dos MQO a uma única observação, mesmo com uma amostra de tamanho grande?
- (vi) Verifique que o estimador MDA não é sensível à inclusão da observação identificada na parte (iii).

CAPÍTULO 10

10.1 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?

10.2 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?

10.3 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?

10.4 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.

10.5 Utilize os dados contidos no arquivo EZANDERS.RAW para fazer este exercício. Os dados são de pedidos de seguro-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 seguro-desemprego ao longo desse período? (Interprete o coeficiente da tendência temporal.) Existe evidência de sazonalidade nos pedidos de seguro-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 seguro-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?

10.6 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?

10.7 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?

10.8 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.

10.9 Utilize os dados contidos no arquivo VOLAT.RAW para fazer este exercício. A variável *rsp500* é o retorno mensal do índice Standard & Poors 500 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 β_1 e β_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.

10.10 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%?

10.11 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 às 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 $\log(\text{totacc})$. Discuta os efeitos e significância estimados das variáveis referentes às leis da velocidade e do cinto de segurança.

10.12 (i) Estime a equação (10.2) usando todos os dados do PHILLIPS.RAW e relate os resultados na forma habitual. Quantas observações você tem agora?

- (ii) Compare as estimativas da parte (i) com as da equação (10.14). Em particular, a adição de anos extras ajuda na obtenção de uma estimativa de relação de trocas entre inflação e desemprego? Explique.
- (iii) Agora execute a regressão usando somente os anos de 1997 até 2003. Como essas estimativas diferem das estimativas na equação (10.14)? As estimativas estão usando os mais recentes sete anos de forma suficientemente precisa para extrair quaisquer conclusões das firmas? Explique.
- (iv) Considere uma configuração de regressão simples na qual iniciamos com n observações de séries temporais e depois as dividimos num período de tempo anterior e num período de tempo posterior. No primeiro período de tempo temos n_1 observações e no segundo período n_2 observações. Explore as partes anteriores deste exercício para avaliar a seguinte declaração: “De forma geral, podemos antecipar que a estimativa de inclinação usando todas as n observações será aproximadamente igual a uma média ponderada das estimativas de inclinação nas subamostras anterior e posterior em que os pesos são n_1/n e n_2/n , respectivamente”.

10.13 Utilize os dados do MINWAGE.RAW neste exercício. Em especial, use as séries de emprego e salário do setor 232 (Mobiliário para Homens e Meninos). A variáveis *crescmen232* é o crescimento mensal (mudanças em logs) no salário médio no setor 232, *cresccontr232* é o crescimento na contratação no setor 232, *crescsalmin* é o crescimento no salário mínimo nacional, e *crescipc* é o crescimento no Índice de Preços ao Consumidor (urbano).

- (i) Faça a regressão de *crescmen232* sobre *crescsalmin*, *crescipc*. O sinal e a magnitude de $\hat{\beta}_{\text{crescsalmin}}$ faz sentido para você? Explique. *crescsalmin* é estatisticamente significativa?
- (ii) Adicione as defasagens de 1 a 12 de *crescsalmin* à equação da parte (i). Você acha necessário incluir essas defasagens para estimar o efeito de longo prazo do crescimento do salário mínimo no crescimento do salário no setor 232? Explique.
- (iii) Faça a regressão de *cresccontr232* sobre *crescsalmin*, *crescipc*. O crescimento do salário mínimo aparenta ter um efeito contemporâneo na *cresccontr232*?
- (iv) Adicione as defasagens de 1 a 12 na equação de crescimento da contratação. O crescimento no salário mínimo tem um efeito estatisticamente significativo no crescimento da contratação, no curto ou longo prazo? Explique.

CAPÍTULO 11

11.1 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 $\log(invpc_t)$ 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 quê?

11.2 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.

11.3 (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?

11.4 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.

11.5 (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), em que tgf e ip apareceram em níveis, em vez de primeiras diferenças.

11.6 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, $r3_t = i3_t - inf_t$, em que $i3_t$ 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

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

em que $\beta_1 > 0$. [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, $\Delta r3_t$?

11.7 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 $E(gc_t | I_{t-1}) = E(gc_t)$, em que 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), em que 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?

11.8 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 2004. Compare-a com a taxa real de desemprego de 2004. (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 2004. 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 2004. A taxa de desemprego de 2004 está dentro do intervalo?

11.9 Utilize os dados contidos no arquivo TRAFFIC2.RAW para este exercício. O Exercício 10.11 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 10.11, 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 10.11, se você ainda não o fez.]

11.10 Utilize todos os dados contidos no arquivo PHILLIPS.RAW para responder a esta questão. Você deve ter agora 56 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?

11.11 A Lei de Okun — por exemplo, Mankiw (1994, Capítulo 2) — supõe os seguintes relacionamentos entre a porcentagem anual de mudança no PIB real, $porcpib$, e a mudança na taxa anual de desemprego, $\Delta desemp$:

$$porcpib = 3 - \Delta desemp.$$

Se a taxa de desemprego for estável o PIB real cresce a uma taxa de 3% anualmente. Para cada ponto percentual de aumento na taxa de desemprego o PIB real cresce menos dois pontos percentuais. (Isto não deve ser interpretado em qualquer sentido causal; isto é mais como uma descrição estatística.) Para verificarmos se os dados da economia americana suporta a Lei de Okun, especificamos um modelo que possibilite desvios através de um termo de erro, $porcpib_t = \beta_0 + \beta_1 \Delta desemp_t + u_t$.

- (i) Use os dados do OKUN.RAW para estimar a equação. Você obtém exatamente 3 para o intercepto e -2 para a inclinação? Você esperava que isso acontecesse?
- (ii) Encontre a estatística t do teste $H_0 : \beta_1 = -2$. Você rejeita H_0 contra a alternativa bilateral em qualquer nível de significância?
- (iii) Encontre a estatística t do teste $H_0 : \beta_0 = 3$. Você rejeita no nível de 5% contra a alternativa bilateral? Isso será uma “forte” rejeição?
- (iv) Encontre a estatística F e o p -valor do teste $H_0 : \beta_0 = 3, \beta_1 = -2$. contra a alternativa que H_0 é falsa. No geral, você diria que os dados rejeitam ou tendem a dar respaldo à lei de Okun?

11.12 Utilize os dados do MINWAGE.RAW neste exercício, focando nas séries de salário e emprego do setor 232 (Mobiliário para Homens e Meninos). A variável $crescmen232$ é o crescimento mensal (mudanças em logs) no salário médio no setor 232; $cresccontr232$ é o crescimento na contratação no setor 232; $crescsalmin$ é o crescimento no salário mínimo nacional; e $crescipc$ é o crescimento no Índice de Preços ao Consumidor (urbano).

- (i) Encontre a autocorrelação de primeira ordem na $crescmen232$. Esta série aparenta ser fracamente dependente?
- (ii) Estime o modelo dinâmico

$$crescmen232_t = \beta_0 + \beta_1 crescmen232_{t-1} + \beta_2 crescsalmin_t + \beta_3 crescipc_t + u_t,$$

pelos MQO. Mantendo fixo o crescimento do salário e o crescimento do PIB do último mês, um aumento no salário mínimo nacional resultará num aumento contemporâneo na $crescmen232_t$? Explique.

- (iii) Agora adicione crescimento defasado na contratação, $cresccontr232_{t-1}$, na equação da parte (ii). Ele será estatisticamente significativo?
- (iv) Comparado com o modelo sem $crescmen232_{t-1}$, e $cresccontr232_{t-1}$, a adição das duas variáveis defasadas tem um efeito importante no coeficiente da $crescsalmin$?
- (v) Faça a regressão de $crescsalmin_t$ sobre $crescmen232_{t-1}$ e $cresccontr232_{t-1}$, e informe o R -quadrado. Comente sobre como o valor do R -quadrado ajuda a explicar sua resposta da parte (iv).

CAPÍTULO 12

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

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

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

12.2 (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%.

12.3 (i) Na parte (i) do Problema 11.6, 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.

12.4 (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 $retorno_{t-1}^2$ 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 β_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 seus resultados na parte (iii)?

12.5 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 $\widehat{demvence} > 0,5$ 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 ?
- 12.6**
- (i) No Exercício em Computador 10.7, 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 Exercício em Computador 11.7, 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 quadrados dos resíduos sobre gc_{t-1} e gc_{t-1}^2 . Qual é sua conclusão?
- 12.7**
- (i) Para o Exemplo 12.14, utilizando os dados contidos no arquivo BARIUM.RAW, obtenha as estimativas iteradas de Cochrane-Orcutt.
 - (ii) As estimativas de Prais-Winsten e de Cochrane-Orcutt são semelhantes? Você esperava que elas fossem?
- 12.8** 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 *finsem*, *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?
- 12.9** 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?

12.10 Use os dados do PHILLIPS.RAW para responder estas questões.

- (i) Usando a totalidade do conjunto de dados, estime a equação da curva estática de Phillips $inf_t = \beta_0 + \beta_1 desemp + u_t$ pelos MQO e relate os resultados na forma habitual.
- (ii) Obtenha os resíduos dos MQO da parte (i), \hat{u}_t , e obtenha ρ da regressão de \hat{u}_t sobre \hat{u}_{t-1} . (É bom incluir um intercepto nesta regressão). Existe evidência forte de correlação serial?
- (iii) Agora estime o modelo da curva estática de Phillips pelo Prais-Winsten iterativo. Compare a estimativa de β_1 com a obtida na Tabela 12.2. Existe muita diferença na estimativa quando os anos posteriores β_1 são adicionados?
- (iv) Em vez de usar Prais-Winsten use o Cochrane-Orcutt iterativo. Quão similares são as estimativas finais de ρ ? Quão similares são as estimativas PW e CO da β_1 ?

12.11 Use os dados da NYSE.RAW para responder a estas questões.

- (i) Estime o modelo na equação (12.47) e obtenha os resíduos quadrados dos MQO. Encontre os valores médio, mínimo e máximo de \hat{u}_t^2 em toda a amostra.
- (ii) Use os resíduos quadrados dos MQO para estimar o seguinte modelo de heteroscedasticidade:

$$\text{Var}(u_t | \text{retorno}_{t-1}, \text{retorno}_{t-2}, \dots) = \text{Var}(u_t | \text{retorno}_{t-1}) = \delta_0 + \delta_1 \text{retorno}_{t-1} + \delta_2 \text{retorno}_{t-1}^2.$$

Relate os coeficientes estimados, os erros padrões, o R -quadrado, e o R -quadrado ajustado.

- (iii) Esboce a variância condicional como uma função de retorno_{-1} defasada. Para que valor de retorno_{-1} a variância é a menor, e qual é a variância?
- (iv) Na predição da variância dinâmica, o modelo na parte (ii) produz qualquer estimativa de variância negativa?
- (v) O modelo na parte (ii) parece se adaptar melhor ou pior que o modelo HCAR(1) no Exemplo 12.9? Explique.
- (vi) À regressão HCAR(1) na equação (12.51), adicione a segunda defasagem \hat{u}_{t-2} . Esta defasagem parece importante? O modelo HCAR(2) se adapta melhor que o modelo na parte (ii)?

12.12 Use os dados de INVEN.RAW para este exercício, veja também Exercício em Computador 11.6.

- (i) Obtenha os resíduos dos MQO do modelo acelerador $\Delta inven = \beta_0 + \beta_1 \Delta PIB_t + u_t$ e use a regressão de \hat{u}_t sobre \hat{u}_{t-1} para verificar se existe correlação serial. Qual é a estimativa de ρ ? Quão problemática a correlação serial parece ser?
- (ii) Estime o modelo acelerador pelo método de PW e compare a estimativa de ρ com a estimativa pelos MQO. Por que você espera que elas sejam similares?

12.13 Use os dados da OKUN.RAW para responder esta questão; veja, também, o Exercício de Computador 11.11.

- (i) Estime a equação $porcpib_t = \beta_0 + \beta_1 \Delta desemp_t + u_t$ e teste os erros quanto a correlação serial AR(1), sem assumir que $\{\Delta desemp_t : t = 1, 2, \dots\}$ é estritamente exógena. Qual sua conclusão?
- (ii) Regresse os resíduos quadrados, \hat{u}_t^2 , sobre $\Delta desemp_t$ (este é teste de Breusch-Pagan de heteroscedasticidade no caso de regressão simples). Qual sua conclusão?
- (iii) Obtenha o erro padrão de heteroscedasticidade-robusta da estimativa pelos MQO $\hat{\beta}_1$. É ele substancialmente diferente do erro-padrão dos MQO?

12.14 Use os dados da MINWAGE.RAW para este exercício, focando no setor 232.

- (i) Estime a equação

$$crescmen232_t = \beta_0 + \beta_1 crescsamin_t + \beta_2 crescipc_t + u_t$$

e teste os erros em busca de correlação serial AR(1). Importa se você assumir que $crescsalmin_t$ e $crescipc_t$ são estritamente exógenas. No geral, qual sua conclusão?

- (ii) Obtenha o erro-padrão de Newey-West das estimativas MQO na parte(i), usando uma defasagem de 12. Como os erros-padrão de Newey-West se comparam com os erros-padrão costumeiros dos MQO?
- (iii) Agora obtenha os erros-padrão de heteroscedasticidade-robusta dos MQO, e compare-os com os costumeiros erros padrões e com os erros padrões de Newey-West. A correlação serial ou heteroscedasticidade aparentam ser mais um problema nesta aplicação?
- (iv) Use o teste de Breusch-Pagan na equação original para verificar que os erros exibem forte heteroscedasticidade.
- (v) Adicione defasagens 1 a 12 de $crescsalmin$ na equação na parte (i). Obtenha o p -valor do teste F conjunto das defasagens 1 a 12, e compare-o com o p -valor do teste de heteroscedasticidade-robusta. Como o ajuste da heteroscedasticidade afeta a significância das defasagens?
- (vi) Obtenha o p -valor do teste de significância conjunta na parte (v) usando o método de Newey-West. Qual é a sua conclusão, agora?
- (vii) Se você deixar de fora as defasagens da $crescsalmin$, a estimativa de propensão de longo prazo será muito diferente?

CAPÍTULO 13

13.1 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 de 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?

13.2 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 quadrados dos resíduos 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.

13.3 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 δ_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*, *banhos*, $\log(intst)$, $\log(terreno)$, e $\log(área)$ à equação. Agora, qual é a sua conclusão sobre o efeito do incinerador sobre o valor dos imóveis?
- (iv) Como pode o coeficiente da $\log(dist)$ ser positivo e estatisticamente significativo na parte (ii) mas não na parte (iii)? O que isto indica sobre os controles usados na parte (iii)?

13.4 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?

13.5 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

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

em que *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 β_{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?

13.6 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 β_1 em termos de θ_1 e β_2 . Faça essa substituição na equação e reorganize. Faça um teste t de θ_1 .)
- (ii) Se $\beta_1 = \beta_2$, mostre que a equação diferenciada pode ser escrita como

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

em que $\delta_1 = 2\beta_1$ 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?

13.7 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 em 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*, *npGPA*, 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?

13.8 O arquivo VOTE2.RAW contém dados em 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

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

em que *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 $\Delta \log(gastmand)$ e $\Delta \log(gastdes)$. Informe o p -valor.
- (iii) Reestime a equação da parte (i) usando $\Delta pgasman$ como a única variável independente. Interprete o coeficiente de $\Delta pgasman$. 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.]

13.9 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.

13.10 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

$$hrsemp_{it} = \beta_0 + \delta_1 d88_t + \delta_2 d89_t + \beta_1 subs_{it} + \beta_2 subs_{i,t-1} + \beta_3 \log(empreg_{it}) + a_i + u_{it}.$$

- (i) Estime a equação usando 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 $hrsemp$) de todos os três períodos de tempo?
- (ii) Interprete o coeficiente de $subs$ e comente sobre sua significância.
- (iii) É surpreendente o fato de o coeficiente de $subs_{-1}$ ser não significativo? Explique.
- (iv) As empresas maiores treinam mais, ou menos, seus empregados, em média? O quanto são grandes as diferenças no treinamento?

13.11 O arquivo MATHPNL.RAW contém dados em 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 em nível de escolas utilizados por Papke (2005). A variável de resposta de interesse nesta questão é $mate4$, 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

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

em que $matricl_{it}$ 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 $mate_{it}$, 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?

13.12 Use os dados de MURDER.RAW para este exercício.

- (i) Usando os anos de 1990 e 1993, estime a equação

$$txhomi_{it} = \delta_0 + \delta_1 d93 + \beta_1 exec_{it} + \beta_2 desemp + a_i + u_{it}, t = 1, 2$$

pelos MQO agrupados e relate os resultados na forma habitual. Não se preocupe que os erros-padrão habituais dos MQO sejam impróprios devido à presença de a_i . Você estima um efeito dissuasor da pena de morte?

- (ii) Compute as estimativas PD (use somente as diferenças de 1990 para 1993; você deve ter 51 observações na regressão PD). Agora qual sua conclusão sobre um efeito dissuasor?
- (iii) Na regressão PD na parte (ii), obtenha os resíduos, digamos, \hat{e}_i . Execute a regressão de Breusch-Pagan \hat{e}_i^2 sobre $\Delta exec_i$, $\Delta desemp_i$ e compute o teste F de heteroscedasticidade. Faça o PD mesmo para o caso especial do teste de White [isto é, \hat{e}_i^2 sobre \hat{y}_i , \hat{y}_i^2 em que os valores ajustados são da parte (ii)]. Qual sua conclusão sobre a heteroscedasticidade na equação PD?
- (iv) Execute a mesma regressão da parte (ii), mas obtenha as estatísticas t de heteroscedasticidade robusta. O que acontece?
- (v) Com qual estatística t na $\Delta exec_i$ você se sente mais confortável em confiar, a habitual ou a de heteroscedasticidade robusta? Por quê?

13.13 Utilize os dados do WAGEPAN.RAW para este exercício.

- (i) Considere o modelo de efeitos não observados

$$lsalário_{it} = \beta_0 + \delta_1 d81t + \dots + \delta_7 d87t + \beta_1 educ_i + \gamma_1 \delta 81, educ_i + \dots + \delta_7 d87, educ_i + \beta_2 sindicato_{it} + a_i + u_{it}$$

em que a_i é permitida ser correlacionada com $educ_i$ e $sindicato_{it}$. Quais parâmetros você pode estimar usando a primeira diferenciação?

- (ii) Estime a equação da parte (i) pelas PD, e teste a hipótese nula de que o retorno da educação não mudou ao longo do tempo.
- (iii) Teste a hipótese da parte (ii) usando um teste plenamente robusto, isto é, um que permita heteroscedasticidade arbitrária e correlação serial nos erros PD, Δu_{it} . A sua conclusão muda?
- (iv) Agora permita que o diferencial sindical mude ao longo do tempo (juntamente com a educação) e estime a equação pelas PD. Qual será o diferencial sindical estimado em 1980? E em 1987? A diferença é estatisticamente significativa?
- (v) Teste a hipótese nula de que o diferencial sindical não mudou ao longo do tempo, e comente seus resultados à luz da sua resposta para a parte (iv).

13.14 Utilize os dados do JTRAIN3.RAW para esta questão.

- (i) Estime o modelo de regressão simples $gr78 = \beta_0 + \beta_1 trein + u$, e relate os resultados na forma habitual. Baseado nesta regressão, o treinamento de pessoal, que ocorreu em 1976 e 1977, parece ter tido um efeito positivo nos ganhos reais da força de trabalho em 1978?
- (ii) Agora, use a mudança nos ganhos reais da força de trabalho, $cre = gr78 - gr75$, como as variáveis dependentes. (Não precisamos diferenciar $trein$ pois assumimos que não havia treinamento de pessoal antes de 1975. Isto é, se definirmos $ctrain = gr78 - gr75$ então $ctrain = train78$ pois $train75 = 0$.) Agora qual é o efeito estimado do treinamento? Detalhe como ele se compara com a estimativa na parte (i).
- (iii) Encontre o intervalo de confiança de 95% do efeito do treinamento usando os habituais erros-padrão MQO e os erros-padrão de heteroscedasticidade robusta, e descreva seus resultados.

CAPÍTULO 14

14.1 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

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

em que 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 alugueis?
- (iv) Estime o modelo por efeitos fixos para verificar se você obtém estimativas e erros-padrão idênticos aos da parte (iii).

14.2 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?

14.3 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*₋₁ 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?)

14.4 No Exemplo 13.8, usamos os dados de Papke (1994) sobre os pedidos de seguro-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:

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

em que 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

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

Observe que a equação diferenciada contém um efeito fixo, c_i .

- (ii) Estime a equação diferenciada por efeitos fixos. Qual é a estimativa de β_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 β_1 ?
- 14.5**
- (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?
- 14.6** Adicione o termo de interação $sindicato_{it} \cdot t$ 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.
- 14.7** 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

$$txhomi_{it} = \theta_t + \beta_1 exec_{it} + \beta_2 desemp_{it} + a_i + u_{it},$$

em que θ_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 β_1 ? Que sinal você acha que β_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 compostos. 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.

14.8 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 em Computador 13.11. O modelo de interesse é

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

em que 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 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 por que 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.]

14.9 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 por agrupamento (com tamanhos do agrupamento iguais a dois.)

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

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

em que 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 agrupamento 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?

14.10 Use os dados do AIRFARE.RAW para este exercício. Estamos interessados em estimar o modelo

$$\log(\text{passag}_{it}) = \theta_t + \beta_1 \text{concen}_{it} + \beta_2 \log(\text{dist}_i) + \beta_3 [\log(\text{dist}_i)]^2 + a_i + u_{it}, t = 1, \dots, 4,$$

em que θ_t significa que permitimos diferentes interceptos anuais.

- (i) Estime a equação acima pelos MQO agrupados, assegurando-se de incluir variáveis simuladas dos anos. Se $\Delta \text{concen} = 0,10$, qual é a porcentagem de aumento estimada *passag*?
- (ii) Qual é o intervalo de confiança de 95% dos MQO da β_1 ? Por que ele provavelmente não será confiável? Se você tiver acesso a um pacote estatístico que compute erros-padrão totalmente robustos, encontre o IC de 95% totalmente robusto da β_1 . Compare-o com o IC habitual e comente.
- (iii) Descreva o que está acontecendo com o quadrático na $\log(\text{dist})$. Em particular, para que valor de *dist* o relacionamento entre $\log(\text{passag})$ e *dist* se torna positivo? [Dica: Primeiro descubra o valor do ponto de inflexão de $\log(\text{dist})$, e depois exponencie]. O ponto de inflexão está fora da faixa dos dados?
- (iv) Agora estime a equação usando os efeitos aleatórios. Como a estimativa de β_1 é alterada?
- (v) Agora estime a equação usando os efeitos ajustados. Qual é a estimativa EF da β_1 ? Por que ela é bastante semelhante à estimativa EA? (Dica: O que é $\hat{\lambda}$ na estimação EA?)
- (vi) Cite duas características de uma rota (exceto a distância entre paradas) que são capturadas pela a_i . Elas podem estar correlacionadas com ?
- (vii) Você está convencido que concentração mais alta numa rota aumenta a tarifa aérea? Qual é sua melhor estimativa?

14.11 Esta questão parte do princípio que você tem acesso a pacotes de programas estatísticos que computam erros-padrão robustos de correlação serial arbitrária e heteroscedasticidade de métodos de dados em painel.

- (i) Das estimativas agrupada de MQO na Tabela 14.1, obtenha os erros-padrão que permitem correlação serial arbitrária (nos erros compostos, $v_{it} = a_i + u_{it}$) e heterosce-

dasticidade. Como os erros-padrão robustos das *educ*, *casado*, e *sindicato* se comparam com os não robustos?

- (ii) Agora obtenha erros-padrão robustos das estimativas de efeitos ajustados que permitem correlação serial arbitrária e heteroscedasticidade no erro idiossincrático, u_{it} . Como eles se comparam com os erros-padrão não robustos dos EF?
- (iii) Para qual dos métodos, MQO ou EF o ajustamento dos erros-padrão é mais importante? Por quê?

14.12 Utilize os dados do ELEM94_95 para responder a esta questão. Os dados são de escolas primárias em Michigan. Neste exercício, vemos os dados como uma amostra por agrupamento, em que cada escola é parte de um grupo distrital.

- (i) Quais são o menor e o maior número de escolas num distrito? Qual o número médio de escolas por distrito?
- (ii) Usando os MQO agrupados, (isto é, fazendo o agrupamento em todas as 1.848 escolas), estime um modelo relacionando *lsalmed*, com *bs*, *lmatricl*, *lstaff*, e *merenda*; veja também Exercício em Computador 9.11. Quais são o coeficiente e erro-padrão da *bs*?
- (iii) Obtenha os erros-padrão que sejam robustos quanto à correlação de agrupamento no interior do distrito (e também heteroscedasticidade). O que acontece com a estatística *t* da *bs*?
- (iv) Ainda usando os MQO agrupados, elimine as quatro observações com $bs > 0,5$ e obtenha $\hat{\beta}_{bs}$ e seu erro-padrão de agrupamento robusto. Existe agora muita evidência de relação de trocas entre salário e benefícios?
- (v) Estime a equação pelos efeitos ajustados permitindo um efeito distrital comum de escolas pertencentes a um distrito. Novamente, elimine as observações com $bs > 0,5$. Agora qual sua conclusão sobre a relação de trocas entre salário e benefícios?
- (vi) À luz das suas estimativas das partes (iv) e (v), comente sobre a importância de se permitir que a remuneração dos professores varie sistematicamente entre os distritos via um efeito ajustado de um distrito.

CAPÍTULO 15

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

- (i) No Exemplo 15.2, usando *irms* como um instrumento 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 $\hat{\beta}_{educ}$ e $\hat{\beta}_{irms}$.
- (vi) Usando os valores estimados da parte (iv), \widehat{educ} , compute a correlação entre \widehat{educ} e *irms*. Use esse resultado para explicar suas descobertas da parte (v).

15.2 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:

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

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.

15.3 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

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

em que 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 de 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 $\log(\text{salário})$?

15.4 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 β_1 se compara com a do MQO?
- (iii) Agora, faça a primeira diferença 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 β_1 com as estimativas anteriores.

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

15.5 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).

15.6 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 de morte 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 Δu na parte (iii). Pode ser razoável assumir que $\Delta exec_{-1}$ seja não correlacionada com Δu . (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)?

15.7 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

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

em que $\Delta inf_t = inf_t - inf_{t-1}$. 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 β_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.

15.8 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 401(k) e ter um plano de aposentadoria privado. Portanto, queremos estimar β_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 401(k). 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 β_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.

15.9 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

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

em que $irms$ 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, \widehat{educ}_i , $i = 1, \dots, n$. Depois, execute a segunda etapa da regressão de $\log(\text{salário}_i)$ sobre \widehat{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 β_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 \widetilde{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 \widetilde{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?

15.10 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*. [Naturalmente, a forma reduzida de *educ* agora inclui a variável explicativa na parte III]. 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

16.1 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

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

em que *cigs* é a quantidade de cigarros fumados por dia, em média. Como você interpreta β_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

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

em que *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 γ_5 e γ_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 β_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 β_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?

16.2 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?

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

- (i) Como $\log(\text{rendpc})$ é não significativo 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?

16.4 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^3_{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?

16.5 Utilize o *Economic Report of the President* (2005 ou posterior) para atualizar os dados do arquivo CONSUMP.RAW pelo menos até o ano de 2003. Reestime a equação (16.35). Alguma das conclusões importantes se altera?

16.6 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

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

em que *crescprpet* (crescimento no preço do petróleo) é assumida como exógeno e *fev*, ..., *dez* são variáveis *dummy* mensais. Que sinais você antecipa para α_1 e β_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 ideia é que elas são deslocadoras da demanda que devem ser em termos gerais não correlacionadas com o erro da oferta u_t^s . Teste se *crescim* é parcialmente correlacionada com *crescres* e *crescnres*; novamente, não se preocupe com a possível correlação serial na forma reduzida.
- (iv) Estime a função de oferta, usando *crescres* e *crescnres* como VIs de *crescim*. 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).]

16.7 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?

16.8 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 12.9. 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_t* e *onda3_t* 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_t* e *onda3_t* como VIs de $\log(premédio_t)$ para estimar a equação da demanda?

- (iii) Faça a regressão de $\log(\text{premed}_t)$ sobre as *dummies* dos dias da semana e sobre as duas medidas de ondas. As variáveis onda2_t e onda3_t 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}_{t1} . Adicione uma única defasagem, $\hat{u}_{t-1,1}$ para estimar a equação de demanda por MQ2E. Lembre-se, use $\hat{u}_{t-1,1}$ 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?

16.9 Para este exercício, use os dados de AIRFARE.RAW, mas somente do ano de 1997.

- (i) Uma função simples de demanda por lugares em companhias aérea em rotas nos Estados Unidos é

$$\log(\text{passageiro}) = \beta_{10} + \alpha_1 \text{passag} + \beta_{11} \log(\text{dist}) + \beta_{12} [\log(\text{dist})]^2 + u_1,$$

em que

passageiro = média de passageiros por dia

passag = tarifa aérea média

dist = a distância da rota (em milhas).

Se esta for uma verdadeira função de demanda, qual deverá ser o sinal de α_1 ?

- (ii) Estime a equação da parte (i) pelos MQO. Qual é a elasticidade de preço estimada?
- (iii) Considere a variável *concen*, que é um indicador da concentração do mercado. (Especificamente, ela é a participação no mercado registrada pela maior transportadora aérea). Explique em suas palavras o que devemos assumir para tratarmos *concen* como exógena na equação de demanda.
- (iv) Agora assuma que *concen* seja exógena na equação de demanda. Estime a forma reduzida de $\log(\text{passag})$ e confirme que *concen* tem um efeito (parcial) positivo na $\log(\text{passag})$.
- (v) Estime a função de demanda usando VI. Agora qual é a estimativa da elasticidade de preço da demanda? Como ela se compara com a estimativa pelos MQO?
- (vi) Usando as estimativas VI, descreva a demanda por assentos depende da distância da rota.

16.10 Use a totalidade do conjunto de dados de painel do AIRFARE.RAW para este exercício. A equação de demanda num modelo de efeitos não observados de equações simultâneas é

$$\log(\text{passageiro}_{it}) = \theta_{it} + \alpha_1 \log(\text{passag}_{it}) + a_{i1} + u_{it1},$$

Em que absorveremos as variáveis de distância na a_{i1} .

- (i) Estime a função de demanda usando efeitos ajustados, tendo a certeza de incluir variáveis simuladas anuais para registrar os diferentes interceptos. Qual é a elasticidade estimada?
- (ii) Use efeitos ajustados para estimar a forma reduzida

$$\log(\text{passageiro}_{it}) = \theta_{i2} + \pi_{21}\text{concen}_{it} + a_{i2} + v_{it2},$$

Faça o teste apropriado para garantir que concen_{it} pode ser usada como uma VI de $\log(\text{passag}_{it})$.

- (iii) Agora estime a função de demanda usando as transformações dos efeitos ajustados juntamente com VI, como na equação (16.42). Agora qual é a elasticidade estimada? Ela é estatisticamente significativa?

CAPÍTULO 17

17.1 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.

17.2 Utilize os dados contidos no arquivo LOANAPP.RAW para fazer este exercício; vide também o Exercício em Computador 7.8.

- (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) Use a equação (17.7) para estimar o tamanho do efeito da discriminação entre o probit e o logit?

17.3 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 não possuem dependentes, 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?

17.4 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 σ^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 da quase-razão de verossimilhança para a significância conjunta dos três termos quadráticos. Qual sua conclusão?

17.5 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 \widehat{kids}_i . Compare o resultado com o R -quadrado do modelo de regressão linear.

17.6 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.

17.7 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).

17.8 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?
- (ix) Usando o modelo da parte (viii), estime o efeito parcial médio de *trein* na probabilidade de desemprego de 1978. Use a (17.17) com $c_k = 0$. Como a estimativa se compara com a feita pelos MQO da parte (viii)

17.9 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) “ecologicamente 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 no nível de 1%?
- (iv) *rendfam* e *tamfam* são conjuntamente significantes?
- (v) Os sinais dos coeficientes das variáveis de preço da parte (iii) são os que você esperava? Explique.
- (vi) Seja β_1 seja o coeficiente de *ecoprc* e β_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.)
- (vii) Obtenha as estimativas de $E(\text{ecolbs}|\mathbf{x})$ de todas as observações da amostra. [Veja a equação (17.25)]. Chame-as de ecolbs_i . Quais são o menor e o maior valores estimados?
- (viii) Compute o quadrado da correlação entre ecolbs_i e $\widehat{\text{ecolbs}}_i$.
- (ix) 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 qualidade de ajuste, o modelo Tobit é melhor que o modelo linear?
- (x) Avalie a seguinte afirmação: “Como o R -quadrado do modelo Tobit é tão pequeno, os efeitos estimados dos preços provavelmente serão inconsistentes.”

17.10 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 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 no nível de 5%?
- (v) Obtenha a estimativa da σ^2 descrita após (17.35). 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?

17.11 Use os dados de CPS91.RAW para fazer este exercício. Esses dados são sobre mulheres casadas, em que também temos informações sobre a renda de cada marido e dados populacionais.

- (i) Que fração das mulheres informaram estarem em uma força de trabalho?
- (ii) Usando apenas os dados de mulheres que trabalham—você não tem escolha—estime a equação salarial

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + \beta_4 \text{negro} + \beta_5 \text{hispan} + u$$

pelos mínimos quadrados ordinários. Relate os resultados na forma habitual. Parece haver diferenças significantes por raça e etnia?

- (iii) Estime um modelo probit de *naft* que inclua as variáveis explicativas na equação salarial da parte (ii) como também *nesprend* e *filhm6*. Estas últimas duas variáveis têm coeficientes do sinal esperado? São elas estatisticamente significantes?
- (iv) Explique por que, com o propósito de testar e, possivelmente, corrigir a equação salarial da seleção na força de trabalho, é importante que a *nesprend* e a *filhm6* ajudem a explicar a *naft*. O que você deve assumir sobre a *nesprend* e a *filhm6* na equação salarial?
- (v) Compute o inverso da razão de Mills (de cada observação) e o adicione como um regressor adicional na equação salarial da parte (ii). Qual é seu *p*-valor bilateral? Você acha que isso é particularmente pequeno com 3.286 observações?
- (vi) A adição do inverso da razão de Mills altera os coeficientes na regressão salarial de maneiras importantes? Explique.

17.12 Utilize os dados Do CHARITY.RAW para responder a estas questões.

- (i) A variável *resp* é uma variável binária igual a um se uma pessoa deu retorno com uma doação a mais recente mala direta. O banco de dados consiste somente de pessoas que

deram retorno pelo menos uma vez no passado. Que fração de pessoas deu retorno mais recentemente?

- (ii) Estime um modelo probit de *resp*, usando *respúlti*, *semanasúlti*, *propresp*, *malaano*, e *meddoa* como as variáveis explicativas. Qual das variáveis explicativas é estatisticamente significativa?
- (iii) Encontre o efeito parcial médio de *malaano* e compare-o com o coeficiente de um modelo de probabilidade linear.
- (iv) Usando as mesmas variáveis explicativas, estime um modelo Tobit de *doa*, o montante da mais recente *doa* (em florins holandeses). Agora qual variável explicativa é estatisticamente significativa?
- (v) Compare o EPM Tobit de *malaano* com o da regressão linear. Eles são similares?
- (vi) As estimativas das partes (ii) e (iv) são inteiramente compatíveis no modelo Tobit? Explique.

17.13 Use os dados do HTV.RAW para responder a esta questão.

- (i) Usando os MQO na amostra completa estime um modelo de $\log(\text{salário})$ usando as variáveis explicativas *educ*, *hábil*, *exper*, *cn*, *oeste*, *sul*, e *urban*. Relate o retorno da educação estimado e seu erro-padrão.
- (ii) Agora estime a equação da parte (i) usando somente pessoas com *educ* < 16. Que percentagem da amostra é perdida? Agora qual é o retorno estimado de um ano de educação formal? Como isso se compara com a parte (i)?
- (iii) Agora elimine todas as observações com *salário* ≥ 20 , de forma que todos os que permanecerem na amostra ganhe menos que 20 dólares por hora. Faça a regressão da parte (i) e comente sobre o coeficiente na *educ*. (Pela razão de o modelo de regressão truncada normal assumir que *y* é contínua, não importa em teoria se eliminamos observações com *salário* ≥ 20 ou com *salário* > 20. Na prática, a inclusão nesta aplicação, pode importar um pouco, pois existem pessoas que ganham exatamente 20 dólares por hora.)
- (iv) Usando a amostra da parte (iii), aplique regressão truncada [com o ponto de truncamento superior sendo $\log(20)$]. A regressão truncada aparenta recuperar o retorno da educação na população total, assumindo que a estimativa da parte (i) é consistente? Explique.

CAPÍTULO 18

18.1 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*, em que 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).

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

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

18.3 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-cause $pcip$. 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?

18.4 Ao testar a cointegração 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 $-4,15$.

18.5 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, em que uma defasagem dos rendimentos das letras do Tesouro de três meses era a variável explicativa. Assumimos que o parâmetro de cointegração 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?

18.6 Utilize os dados contidos no arquivo PHILLIPS.RAW, para responder a estas questões.

- (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 $desemp_{1998}$; arredonde para duas casas decimais. Qual das equações produz uma melhor previsão?

- (iii) Como discutimos no texto, a previsão de $desemp_{1998}$ 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 dois passos à frente de $desemp$. Isto é, faça a previsão de $desemp_{1998}$ 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 $desemp_{1997} = 4,9$ em (18.48)?

18.7 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).

18.8 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 Δtgf_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?

18.9 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?

18.10 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 cointegração 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 cointegração, 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.

18.11 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 cointegrados. 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?

18.12 Este exercício também utiliza os dados do arquivo VOLAT.RAW. O Exercício em Computador 18.11 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 no 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)?

18.13 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 10.11.

- (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?

18.14 Use os dados do MINWAGE.DTA do setor 232 para responder às seguintes questões.

- (i) Confirme que $lsalário232$ e $lemp232$ são melhores caracterizadas como processos I(1). Use o teste de Dickey-Fuller aumentado com uma defasagem de $crescmen232$ e $cresccontr232$, respectivamente e uma tendência temporal linear. Existe alguma dúvida que estas séries devam ser consideradas como tendo raízes unitárias?
- (ii) Faça a regressão de $lemp232$ sobre $lsalário232$ e verifique se há cointegração, tanto com como sem uma tendência temporal, levando em conta duas defasagens no teste de Engle-Granger aumentado. Qual sua conclusão?
- (iii) Agora faça a regressão de $lsalário232$ sobre o log da taxa real de salário, $lrsalário232_t = lsalário232_t - lipc_t$ e uma tendência temporal. Você encontra cointegração? Elas estão “próximas” de serem cointegradas quando você usa salário real em vez de salários nominais?
- (iv) Quais são alguns fatores que podem estar faltando da regressão cointegrante na parte (iii)?