Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Wooldridge, Jeffrey M.
Introdução à econometria : uma abordagem moderna / Jeffrey M. Wooldridge ; tradução José Antônio Ferreira ; revisão técnica Galo Carlos Lopez Noriega. -- São Paulo : Cengage Learning, 2010.

Título original: Introductory econometrics : a modern approach
4. ed. norte-americana
Bibliografia.
ISBN 978-85-221-0446-8

1. Econometria II. Título.

10-11298 CDD-330.015195

Índices para catálogo sistemático:

1. Econometria 330.015195

Introdução à Econometria

Uma Abordagem Moderna

Jeffrey M. Wooldridge
Michigan State University

Tradução da quarta edição norte-americana

Tradução
José Antônio Ferreira

Revisão Técnica
Galo Carlos Lopez Noriega, MSc.
Docente de métodos quantitativos no MBA do Insper Ibmec São Paulo
e coordenador acadêmico de Educação Executiva do Insper Ibmec São Paulo

CENGAGE Learning

Austrália Brasil Canadá Cingapura Espanha Estados Unidos México Reino Unido

13

O Agrupamento de Cortes Transversais ao Longo do Tempo. Métodos Simples de Dados em Painel

Até agora, estudamos a análise de regressão múltipla usando dados puros de corte transversal ou de séries temporais. Embora esses dois casos surjam com frequência no trabalho aplicado, conjuntos de dados que possuem as duas dimensões, corte transversal e séries temporais, estão sendo usados cada vez mais na pesquisa empírica. Métodos de regressão múltipla podem ainda ser usados em tais conjuntos de dados. Aliás, dados com aspectos de corte transversal e de séries temporais podem, muitas vezes, esclarecer questões importantes de política econômica. Usaremos vários exemplos, neste capítulo.

Analisaremos, ainda, dois tipos de conjuntos de dados. Um **agrupamento independente de cortes transversais** é obtido fazendo-se uma amostragem aleatória de dados de uma população grande, em diferentes períodos de tempo (geralmente, mas não necessariamente, em anos diferentes). Por exemplo, de cada ano, podemos extrair uma amostra aleatória de salários por hora, educação, experiência etc., da população de trabalhadores nos Estados Unidos. Ou, de cada ano, podemos extrair uma amostra aleatória de preços de venda, área construída, número de banheiros etc., das casas vendidas em determinada área metropolitana. Do ponto de vista estatístico, esses conjuntos de dados possuem uma importante característica: eles consistem de observações amostrais coletadas *independentemente*. Esse também foi um aspecto fundamental em nossas análises de dados de corte transversal: entre outras coisas, ele elimina a correlação nos erros entre as diferentes observações.

Um agrupamento independente de cortes transversais difere de uma amostra aleatória única pelo fato de que fazer amostragem de uma população em períodos de tempo diferentes provavelmente levará a observações que não são identicamente distribuídas. Por exemplo, distribuições de salários e educação vêm mudando ao longo do tempo, na maioria dos países. Como veremos, é fácil de lidar com isso na prática, permitindo que em um modelo de regressão múltipla o intercepto e, em alguns casos, a inclinação, mudem ao longo do tempo. Tratamos de tais modelos na Seção 13.1. Na Seção 13.2, discutiremos como o agrupamento de cortes transversais ao longo do tempo pode ser usado para avaliar alterações de política econômica.

Um conjunto de **dados em painel**, embora tenha dimensões tanto de corte transversal como de série temporal, difere em alguns aspectos importantes de um agrupamento independente de cortes transversais. Para coletar dados em painel — algumas vezes chamados de **dados longitudinais** —, nós acompanhamos (ou tentamos acompanhar) os *mesmos* indivíduos, famílias, empresas, cidades, estados, ou o que seja, ao longo do tempo. Por exemplo, um conjunto de dados em painel sobre salários individuais, horas, educação e outros fatores, é coletado fazendo-se uma seleção aleatória de pessoas de uma população em determinado momento. Depois, essas *mesmas* pessoas são entrevistadas em vários períodos de tempo subsequentes. Isso nos fornecerá dados sobre salários, horas, educação etc. do mesmo grupo de pessoas em anos diferentes.

Conjuntos de dados em painel relativos a distritos escolares, cidades, municípios, estados e países são muito fáceis de serem coletados, e a análise da política governamental é muito mais aprimorada com o uso de conjuntos de dados em painel; veremos alguns exemplos nas discussões a seguir. Na análise econométrica de dados em painel, não podemos supor que as observações sejam independentemente distribuídas ao longo do tempo. Por exemplo, fatores não observados (como a aptidão) que afetaram o salário-hora de um indivíduo em 1990 também afetarão o salário dessa pessoa em 1991; fatores não observados que afetaram a taxa de criminalidade de uma cidade em 1985 também afetarão a taxa de criminalidade dessa cidade em 1990. Por essa razão, modelos e métodos especiais foram desenvolvidos para analisar dados em painel. Nas Seções 13.3, 13.4 e 13.5, descreveremos o método objetivo da diferenciação para remover atributos constantes no tempo não observados das unidades de estudo. Como os métodos de dados em painel são de certa forma mais avançados, vamos nos valer muito da intuição na descrição das propriedades estatísticas dos procedimentos de estimação, deixando o detalhamento das hipóteses para o apêndice do capítulo. Seguiremos a mesma estratégia no Capítulo 14, que tratará de métodos de dados em painel mais complexos.

13.1 O AGRUPAMENTO INDEPENDENTE DE CORTES TRANSVERSAIS AO LONGO DO TEMPO

Pesquisas sobre pessoas, famílias e empresas são repetidas a intervalos regulares, muitas vezes a cada ano. Um exemplo é o *Current Population Survey* (Contagem da População Atual) — ou CPS —, que de forma aleatória faz pesquisa domiciliar a cada ano. Se uma amostra aleatória for extraída a cada período de tempo, o agrupamento das amostras aleatórias resultantes produz um agrupamento independente de cortes transversais.

Uma razão para usar agrupamentos independentes de cortes transversais é aumentar o tamanho da amostra. Ao agrupar amostras aleatórias extraídas da mesma população, mas em períodos de tempo diferentes, podemos obter estimadores mais precisos e estatísticas de testes mais poderosas. O agrupamento é útil em relação à isso somente se a relação entre a variável dependente e pelo menos uma das variáveis independentes permanecer constante ao longo do tempo.

Como mencionado na introdução, o uso de cortes transversais agrupados provoca apenas pequenas complicações estatísticas. Em geral, para refletir o fato de que a população pode ter distribuições diferentes em períodos de tempo diferentes, permitimos que o intercepto difira ao longo dos períodos, normalmente anos. Isso é facilmente conseguido com a inclusão de variáveis *dummy* para todos os anos menos um, em que o primeiro ano da amostra é habitualmente escolhido como o ano-base. Também é possível que a variância do erro mude ao longo do tempo, assunto que discutiremos mais tarde.

Algumas vezes, o padrão dos coeficientes das variáveis *dummy* anuais é de interesse particular. Por exemplo, um demógrafo pode estar interessado na seguinte questão: *após* ter controlado a variável educação, o padrão de fertilidade entre mulheres com mais de 35 anos mudou entre 1972 e 1984? O seguinte exemplo ilustra como essa questão pode ser respondida de maneira simples, com o uso da análise de regressão múltipla com **variáveis *dummy* anuais**.

EXEMPLO 13.1

(Fertilidade Feminina ao Longo do Tempo)

O conjunto de dados do arquivo FERTIL1.RAW, que é semelhante ao usado por Sander (1992), provém do *General Social Survey* (Pesquisa Social Geral) do *National Opinion Research Center* (Centro de Pesquisa de Opinião Nacional) para os anos de 1972 a 1984, inclusive. Usamos esses dados para estimar um modelo que explique o número total de nascimentos por mulheres (*kids*).

Uma questão de interesse é a seguinte: após termos controlado todos os outros fatores observáveis, o que aconteceu com as taxas de fertilidade ao longo do tempo? Os fatores que controlamos são anos de educação, idade, raça, região do país onde as mulheres residiam quando tinham 16 anos, e ambiente em que viviam quando tinham essa mesma idade. As estimativas estão na Tabela 13.1.

O ano-base é 1972. Os coeficientes das variáveis *dummy* anuais mostram uma nítida queda da fertilidade no início dos anos 1980. Por exemplo, o coeficiente de *a82* indica que, mantendo fixos educação, idade e outros fatores, uma mulher teve, em média, 0,52 menos filhos em 1982 do que em 1972. Isso é uma queda bastante grande: mantendo fixos *educ*, *idade* e os outros fatores, prevê-se que 100 mulheres em 1982 teriam 52 crianças a menos se comparadas com 100 mulheres em 1972. Como estamos controlando a variável educação, essa queda é separada do declínio da fertilidade em razão do aumento nos níveis de educação. (A média de anos de escolaridade é 12,2 em 1972 e 13,3 em 1984.) Os coeficientes de *a82* e *a84* representam queda na fertilidade por razões que não estão captadas nas variáveis explicativas.

Considerando que as variáveis *dummy* anuais de 1982 e 1984 são individualmente bastante significantes, não é surpreendente que, como um grupo, as variáveis simuladas anuais sejam, conjuntamente, bastante significantes: o *R*-quadrado da regressão sem as *dummies* anuais é 0,1019, e isso leva a $F_{6,1111} = 5,87$ e *p*-valor ≈ 0.

Mulheres com mais anos de escolaridade têm menor número de filhos, e a estimativa é, estatisticamente, bastante significante. Com todos os outros fatores permanecendo iguais, 100 mulheres com curso superior terão, em média, 51 filhos a menos do que 100 mulheres com apenas ensino médio: $0,128(4) = 0,512$. A idade tem um efeito redutor sobre a fertilidade. (O ponto de inflexão do termo quadrático está próximo da *idade* = 46, quando a maior parte das mulheres parou de ter filhos.)

O modelo estimado na Tabela 13.1 presume que o efeito de cada variável explicativa, particularmente a da educação, permaneceu constante. Isso pode ou não ser verdade; solicitaremos que você explore esse assunto em Exercício em Computador 13.1, no site da Cengage.

Finalmente, pode haver heteroscedasticidade no termo de erro adjacente da equação estimada. Podemos tratar desse assunto usando os métodos do Capítulo 8. Existe, aqui, uma diferença interessante: agora, a variância do erro pode mudar ao longo do tempo mesmo que ela não mude com os valores de *educ*, *idade*, *negro* etc. Os erros-padrão e as estatísticas de testes robustos em relação à heteroscedasticidade, contudo, serão válidos. O teste de Breusch-Pagan será obtido fazendo-se a regressão do quadrado dos resíduos de MQO sobre *todas* as variáveis independentes da Tabela 13.1, inclusive as *dummies* anuais. (No caso especial da estatística de White, os valores estimados $\widehat{kids}$ e os quadrados dos valores estimados são usados como variáveis independentes, como sempre). Um procedimento de mínimos quadrados ponderados deve explicar as variâncias que possivelmente mudem ao longo do tempo. No procedimento discutido na Seção 8.4, *dummies* anuais seriam incluídas na equação (8.32).

QUESTÃO 13.1

Tendo em vista a Tabela 13.1, a seguir, alega-se que, se todos os demais fatores permanecerem iguais, espera-se que uma mulher negra tenha um filho a mais que uma mulher não negra. Você concorda com essa alegação?

Podemos também interagir uma variável *dummy* anual com variáveis explicativas básicas para verificar se o efeito dessa variável mudou ao longo de certo período de tempo. O próximo exemplo examina como o retorno da educação e a diferença salarial por gênero mudaram de 1978 a 1985.

Tabela 13.1

Determinantes da fertilidade feminina.

Variável dependente: ***kids***		
Variáveis independentes	**Coeficientes**	**Erros-padrão**
educ	−0,128	0,018
idade	0,532	0,138
idade²	−0,0058	0,0016
negra	1,076	0,174
leste	0,217	0,133
centnorte	0,363	0,121
oeste	0,198	0,167
fazenda	−0,053	0,147
outrural	−0,163	0,175
cidade	0,084	0,124
cidpeq	0,212	0,160
a74	0,268	0,173
a76	−0,097	0,179
a78	−0,069	0,182
a80	−0,071	0,183
a82	−0,522	0,172
a84	−0,545	0,175
constante	−7,742	3,052
$n = 1.129$ $R^2 = 0,1295$ $\bar{R}^2 = 0,1162$		

EXEMPLO 13.2

(Mudanças no Retorno da Educação e a Diferença Salarial por Gênero)

Uma equação log(*salárioh*) (na qual *salárioh* representa o salário por hora) agrupada ao longo dos anos de 1978 (o ano-base) e 1985 é

$$\log(salárioh) = \beta_0 + \delta_0 a85 + \beta_1 educ + \delta_1 a85 \cdot educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 sindicato + \beta_5 feminino + \delta_5 a85 \cdot feminino + u, \quad \textbf{(13.1)}$$

em que já devemos estar familiarizados com a maioria das variáveis explicativas. A variável *sindicato* é uma variável *dummy* igual a um se a pessoa for sindicalizada e igual a zero, caso contrário. A variável *a85* é uma variável *dummy* igual a um se a observação corresponde ao ano de 1985 e zero se for de 1978. Existem 550 pessoas na amostra em 1978 e conjunto diferente de 534 pessoas em 1985.

O intercepto de 1978 é β_0, e o intercepto de 1985 é $\beta_0 + \delta_0$. O retorno da educação em 1978 é β_1, e o retorno da educação em 1985 é $\beta_1 + \delta_1$. Portanto, δ_1 mede como o retorno de mais um ano de estudo mudou ao longo do período de sete anos. Finalmente, em 1978 o diferencial log(*salárioh*) entre homens e mulheres é β_5; o diferencial em 1985 é $\beta_5 + \delta_5$. Assim, podemos testar a hipótese nula de que nada aconteceu com o diferencial por gênero ao longo desse período de sete anos, fazendo o teste H_0: $\delta_5 = 0$. A hipótese alternativa de que o diferencial por gênero tenha sido *reduzido* é H_1: $\delta_5 > 0$. Para simplificar, presumimos que a experiência e a filiação sindical têm o mesmo efeito sobre os salários em ambos os períodos de tempo.

Antes de apresentarmos as estimativas, existe um outro problema do qual precisamos tratar, ou seja, o salário por hora, aqui, está expresso em dólares nominais (correntes). Como salários nominais aumentam em razão da inflação, nosso verdadeiro interesse está no efeito de cada variável explicativa sobre os salários reais. Suponha que nos concentremos em medir os salários em dólares de 1978. Isso exigirá que deflacionemos os salários de 1985 para valores em dólares de 1978. [Usando o índice de preços ao consumidor do *Economic Report of the President* (Relatório Econômico do Presidente) de 1997, o fator de deflação é 107,6/65,2 ≈ 1,65.] Embora possamos, com facilidade, dividir cada salário-hora de 1985 por 1,65, isso acaba não sendo necessário, dado que uma *dummy* anual de 1985 foi incluída na regressão *e* log(*salárioh*) (no lugar de *salárioh*) foi usado como variável dependente. O uso do salário real ou nominal em uma forma funcional logarítmica somente afeta o coeficiente da *dummy* anual, *a85*. Para verificar isso, seja *P85* o fator de deflação para os salários-hora de 1985 (1,65, se usarmos o IPC). Então, o log do salário por hora real de cada pessoa *i* na amostra de 1985 é

$$\log(salárioh_i/P85) = \log(salárioh_i) - \log(P85).$$

Agora, embora $salárioh_i$ difira entre as pessoas, *P85* não difere. Portanto, log(*P85*) será absorvido pelo intercepto de 1985. (Essa conclusão seria modificada se, por exemplo, usássemos índices de preços diferentes para pessoas de diferentes regiões do país.) O ponto principal é que, para examinarmos como o retorno da educação ou o diferencial por gênero mudaram, não precisamos converter salários nominais em salários reais na equação (13.1). O Exercício em Computador 13.2, disponível no site da Cengage, solicitará que você verifique isso neste exemplo.

Se nos esquecermos de admitir diferentes interceptos em 1978 e 1985, o uso de salários nominais pode produzir resultados seriamente equivocados. Se usarmos *salárioh* em lugar de log(*salárioh*) como variável dependente, é importante usar o salário real e incluir uma *dummy* anual.

EXEMPLO 13.2 (continuação)

A discussão anterior geralmente é válida quando usamos valores monetários na variável dependente ou nas variáveis independentes. Desde que os valores monetários apareçam na forma logarítmica e que sejam usadas variáveis *dummy* para todos os períodos de tempo (exceto, é claro, para o período-base), o uso de deflatores de preços agregados afetará somente os interceptos; nenhuma das estimativas de inclinação será alterada.

Agora, utilizamos os dados contidos no arquivo CPS78_85.RAW para estimar a equação:

$$\begin{aligned}\widehat{\log(salárioh)} = {} & \underset{(0{,}093)}{0{,}459} + \underset{(0{,}124)}{0{,}118}\, a85 + \underset{(0{,}0067)}{0{,}0747}\, educ + \underset{(0{,}0094)}{0{,}0185}\, a85 \cdot educ \\ & + \underset{(0{,}0036)}{0{,}0296}\, exper - \underset{(0{,}00008)}{0{,}00040}\, exper^2 + \underset{(0{,}030)}{0{,}202}\, sindicato \\ & - \underset{(0{,}037)}{0{,}317}\, feminino + \underset{(0{,}051)}{0{,}085}\, a85 \cdot feminino \\ & n = 1.084,\ R^2 = 0{,}426,\ \bar{R}^2 = 0{,}422.\end{aligned} \quad \textbf{(13.2)}$$

O retorno da educação em 1978 é estimado em torno de 7,5%; o retorno da educação em 1985 é cerca de 1,85 pontos percentuais *mais alto*, ou cerca de 9,35%. Como a estatística *t* no termo de interação é 0,0185/0,0094 ≈ 1,97, a diferença no retorno da educação é estatisticamente significante no nível de 5% contra uma alternativa bilateral.

E a diferença salarial por gênero? Em 1978, outros fatores que permaneceram iguais, uma mulher ganhava cerca de 31,7% menos que um homem (27,2% é uma estimativa mais precisa). Em 1985, a diferença em log(*salárioh*) é − 0,317 + 0,085 = − 0,232. Portanto, a diferença salarial por gênero parece ter caído de 1978 para 1985 em cerca de 8,5 pontos percentuais. A estatística *t* do termo de interação é de cerca de 1,67, o que significa que ela é significante no nível de 5% contra a alternativa unilateral positiva.

O que acontece se fizermos a interação de *todas* as variáveis independentes com *a85* na equação (13.2)? Seria o mesmo que estimarmos duas equações separadas, uma para 1978 e outra para 1985. Algumas vezes, isso é preferível. Por exemplo, no Capítulo 7, discutimos um estudo feito por Krueger (1993), no qual ele estimou o retorno do uso de computador no trabalho. Krueger estima duas equações separadas, uma usando o CPS de 1984 e a outra usando o de 1989. Comparando como o retorno da educação muda ao longo do tempo e se o uso de computadores está, ou não, controlado, ele estima que de um terço à metade do aumento observado no retorno da educação ao longo desse período de cinco anos pode ser atribuído ao aumento do uso de computadores. [Veja Tabelas VIII e IX em Krueger (1993).]

O Teste de Chow de Mudança Estrutural ao Longo do Tempo

No Capítulo 7, discutimos como o teste de Chow — que é, simplesmente, um teste *F* — pode ser usado para determinar se uma função de regressão múltipla difere entre dois grupos. Também podemos aplicar esse teste em dois períodos de tempo diferentes. Uma forma do teste obtém a soma dos quadrados dos resíduos da estimação agrupada como a SQR restrita. A SQR sem restrições é a soma das SQR dos dois períodos de tempo estimados separadamente. A mecânica do cálculo da estatística é exatamente a mesma da Seção 7.4. Também há uma versão da estatística robusta em relação à heteroscedasticidade (veja Seção 8.2).

O Exemplo 13.2 sugere outra maneira de calcular o teste de Chow para dois períodos de tempo, fazendo a interação de cada variável com uma *dummy* anual de um dos dois anos e testando a significância conjunta da *dummy* anual e todos os termos de interação. Como o intercepto em um modelo de regressão muitas vezes muda ao longo do tempo (em razão da inflação, digamos, no exemplo dos preços dos imóveis), esse teste de Chow completo pode verificar tais mudanças. Em geral, é mais interessante considerar uma diferença de interceptos e depois testar se determinados coeficientes de inclinação mudam ao longo do tempo (como fizemos no Exemplo 13.2).

Um teste de Chow também pode ser calculado para mais de dois períodos de tempo. Assim como no caso de dois períodos, em geral é mais interessante permitir que os interceptos mudem ao longo do tempo e depois testar se os coeficientes de inclinação mudaram também ao longo do tempo. Podemos testar a constância dos coeficientes de inclinação fazendo, geralmente, a interação de todas as *dummies* anuais (exceto da que define o grupo-base) com uma, várias ou todas as variáveis explicativas, e verificar a significância conjunta dos termos de interação. Os Exercícios em Computador 13.1 e 13.2, disponíveis no site da Cengage, são exemplos. Quando temos muitos períodos de tempo e muitas variáveis explicativas, construir um conjunto completo de interações pode ser cansativo. Como alternativa, pode-se adaptar o método descrito na parte (vi) do Exercício em Computador 7.11, disponível no site da Cengage, Capítulo 7. Primeiro, estime o modelo restrito fazendo uma regressão agrupada admitindo diferentes interceptos de tempo; isso produz SQR_r. Depois, execute uma regressão para cada um dos, digamos, T períodos de tempo e obtenha a soma dos quadrados dos resíduos para cada período de tempo. A soma dos quadrados dos resíduos sem restrições é obtida como $SQR_{ur} = SQR_1 + SQR_2 + \ldots + SQQ_T$. Se houver k variáveis explicativas (sem incluir os interceptos das *dummies* temporais) com T períodos de tempo, estaremos, então, testando $(T - 1)k$ restrições, e haverá $T + Tk$ parâmetros estimados no modelo sem restrições. Assim, se $n = n_1 + n_2 + \ldots + n_T$ for o número total de observações, então, os *gl* do teste F serão $(T - 1)k$ e $n - T - Tk$. Calculamos a estatística F da maneira habitual: $[(SQR_r - SQR_{ur})/SQR_{ur}][(n - T - Tk)/(T - 1)k]$. Infelizmente, como em qualquer teste F baseado nas somas dos quadrados dos resíduos ou em R-quadrados, esse teste não é robusto quanto à heteroscedasticidade (inclusive quanto às mudanças nas variâncias ao longo do tempo). Para obter um teste robusto em relação à heteroscedasticidade, devemos construir os termos de interações e fazer uma regressão agrupada.

13.2 ANÁLISE DE DECISÃO DE POLÍTICAS COM AGRUPAMENTOS DE CORTES TRANSVERSAIS

Cortes transversais agrupados podem ser muito úteis para a avaliação do impacto de determinado evento ou decisão política. O exemplo seguinte de um estudo de evento mostra como dois conjuntos de dados de cortes transversais, coletados antes e depois da ocorrência de um evento, podem ser usados para determinar seu efeito sobre resultados econômicos.

EXEMPLO 13.3

(Efeito da Localização de um Incinerador de Lixo sobre os Preços de Imóveis)

Kiel e McClain (1995) estudaram o efeito que um novo incinerador de lixo teve sobre os valores dos imóveis em North Andover, Massachusetts. Eles utilizaram dados de muitos anos e uma análise econométrica um tanto complicada. Utilizaremos dados de dois anos e alguns modelos simplificados, mas nossa análise é semelhante à deles.

O rumor de que um novo incinerador seria construído em North Andover começou após 1978 e a construção começou em 1981. Esperava-se que o incinerador entrasse em operação em pouco tempo com o início das obras; na realidade, ele entrou em operação em 1985. Utilizaremos dados de preços dos imóveis vendidos em 1978 e outra amostra dos vendidos em 1981. A hipótese é que os preços dos imóveis localizados próximos do incinerador cairiam em relação aos preços dos imóveis mais distantes.

EXEMPLO 13.3 (continuação)

Como ilustração, estabelecemos que uma casa está próxima do incinerador se estiver localizada a menos de três milhas (4,8 km). [No Exercício em Computador 13.3, disponível no site da Cengage, você terá que utilizar as distâncias reais das casas ao incinerador, como em Kiel e McClain (1995).] Começaremos verificando o efeito monetário sobre os preços dos imóveis. Isso requer que façamos a mensuração dos preços em moeda constante. Medimos todos os imóveis aos preços de 1978, utilizando o índice de preços de imóveis de Boston. Suponhamos que *rpreço* represente o preço dos imóveis em termos reais.

Um analista ingênuo usaria somente os dados de 1981 e estimaria um modelo muito simples:

$$rpreço = \gamma_0 + \gamma_1 proxincin + u, \qquad \textbf{(13.3)}$$

em que *proxincin* é uma variável binária igual a um se o imóvel estiver localizado próximo ao incinerador e zero, caso contrário. Estimando a equação utilizando os dados contidos no arquivo KIELMC.RAW, resulta

$$\begin{aligned} \widehat{rpreço} &= 101.307{,}5 - 30.688{,}27\, proxincin \\ &\quad\;\; (3.093{,}0) \quad\;\; (5.827{,}71) \\ n &= 142,\ R^2 = 0{,}165. \end{aligned} \qquad \textbf{(13.4)}$$

Como essa é uma regressão simples sobre uma única variável *dummy*, o intercepto é a média dos preços de venda dos imóveis afastados do incinerador, enquanto o coeficiente de *proxincin* é a diferença no preço médio de venda entre os imóveis situados próximos ao incinerador e os distantes dele. A estimativa mostra que o preço médio de venda dos imóveis para o primeiro grupo era de 30.688,27 dólares a menos que o do segundo grupo. A estatística *t* é maior que cinco, em valor absoluto, de modo que podemos rejeitar com certeza a hipótese de que os preços médios de venda dos imóveis situados próximos do incinerador e daqueles distantes dele sejam os mesmos.

Infelizmente, a equação (13.4) *não* implica a localização do incinerador como a causa dos menores valores dos imóveis. Aliás, se computarmos a mesma regressão para o ano de 1978 (antes de sequer haver rumores sobre o incinerador), obteremos

$$\begin{aligned} \widehat{rpreço} &= 82.517{,}23 - 18.824{,}37\, proxincin \\ &\quad\;\; (2.653{,}79) \quad (4.744{,}59) \\ n &= 179,\ R^2 = 0{,}082. \end{aligned} \qquad \textbf{(13.5)}$$

Portanto, mesmo *antes* de haver qualquer comentário sobre o incinerador, o valor médio de um imóvel próximo do local era 18.824,37 dólares menor que o de outro distante do local (82.517,23 dólares); a diferença também é estatisticamente significante. Isso é coerente com a percepção de que o incinerador foi construído em uma área de imóveis de menor valor.

Como, então, podemos dizer se a construção de um novo incinerador reduz os valores dos imóveis? O segredo está em verificar como o coeficiente de *proxincin* mudou entre 1978 e 1981. A diferença na média dos valores dos imóveis era muito maior em 1981 do que em 1978 (30.688,27 dólares contra 18.824,37 dólares), mesmo como uma percentagem do valor médio dos imóveis distantes do local do incinerador. A diferença nos dois coeficientes de *proxincin* é

EXEMPLO 13.3 (continuação)

$$\hat{\delta}_1 = -30.688,27 - (-18.824,37) = -11.863,9.$$

Essa é nossa estimativa do efeito do incinerador sobre os valores dos imóveis próximos de sua localização. Em economia empírica, $\hat{\delta}_1$ tornou-se conhecido como **estimador de diferença em diferenças** porque ele pode ser expresso como

$$\hat{\delta}_1 = (\overline{rpreço}_{81,pr} - \overline{rpreço}_{81,af}) - (\overline{rpreço}_{78,pr} - \overline{rpreço}_{78,af}), \tag{13.6}$$

em que "pr" significa "próximo ao local do incinerador" e "af" significa "afastado do local do incinerador". Em outras palavras, $\hat{\delta}_1$ é a diferença, ao longo do tempo, das diferenças das médias dos preços dos imóveis nas duas localizações.

Para testar se $\hat{\delta}_1$ é estatisticamente diferente de zero, precisamos encontrar seu erro-padrão utilizando uma análise de regressão. De fato, $\hat{\delta}_1$ pode ser obtido ao estimarmos

$$rpreço = \beta_0 + \delta_0 a81 + \beta_1 proxincin + \delta_1 a81 \cdot proxincin + u, \tag{13.7}$$

utilizando os dados agrupados de ambos os anos. O intercepto, β_0, é o preço médio de um imóvel distante do incinerador em 1978. O parâmetro δ_0 indica as alterações em *todos* os valores dos imóveis em North Andover de 1978 a 1981. [Uma comparação das equações (13.4) e (13.5) mostra que os valores dos imóveis em North Andover, com relação ao índice de preços de imóveis de Boston, aumentou nitidamente ao longo desse período.] O coeficiente de *proxincin*, β_1, mede o efeito da localização, que *não* é em razão da presença do incinerador: como vimos na equação (13.5), mesmo em 1978, os imóveis próximos do local onde seria construído o incinerador eram vendidos por preços mais baixos do que os de outras áreas afastadas do local.

O parâmetro de interesse está no termo de interação *a81·proxincin*: δ_1 mede o declínio nos valores dos imóveis em razão do novo incinerador, desde que presumamos que tanto os imóveis próximos quanto os distantes do local do incinerador não tenham sido, por outras razões, valorizados a taxas diferentes.

As estimativas das equações (13.7) são apresentadas na coluna (1) da Tabela 13.2.

Tabela 13.2

Efeito da localização de incinerador sobre os preços de imóveis.

Variável dependente: *rpreço*			
Variável independente	**(1)**	**(2)**	**(3)**
constante	82.517,23 (2.726,91)	89.116,54 (2.406,05)	13.807,67 (11.166,59)
a81	18.790,29 (4.050,07)	21.321,04 (3.443,63)	13.928,48 (2.798,75)
proxincin	−18.824,37 (4.875,32)	9.397,94 (4.812,22)	3.780,34 (4.453,42)

(*cont.*)

EXEMPLO 13.3 (continuação)

Tabela 13.2 (*cont.*)

Variável independente	(1)	(2)	(3)
a81 · proxincin	−11.863,90 (7.456,65)	−21.920,27 (6.359,75)	−14.177,93 (4.987,27)
Outros controles	Não	*idade*, *idade*^{2}	Conjunto total
Observações *R*-quadrado	321 0,174	321 0,414	321 0,660

O único número que não conseguimos obter das equações (13.4) e (13.5) é o erro-padrão de $\hat{\delta}_1$. A estatística *t* de $\hat{\delta}_1$ está em torno de −1,59, que é marginalmente significante contra uma alternativa unilateral (*p*-valor ≈ 0,057).

Kiel e McClain (1995) incluíram várias características dos imóveis em suas análises da localização do incinerador. Existem duas boas razões para fazer isso. A primeira é que os tipos de imóveis à venda em 1981 podem ter sido sistematicamente diferentes dos vendidos em 1978; e assim, é importante controlar as características que possam ter sido diferentes. Entretanto, tão importante quanto, mesmo que as características médias dos imóveis tenham sido as mesmas em ambos os anos, sua inclusão pode reduzir bastante a variância do erro, que por sua vez pode diminuir o erro-padrão de $\hat{\delta}_1$. (Veja Seção 6.3 do Capítulo 6 para discussão desse assunto.) Na coluna (2), controlamos as idades dos imóveis, utilizando um termo quadrático. Isso aumenta substancialmente o *R*-quadrado (ao reduzir a variância do resíduo). O coeficiente de *a81 · proxincin* é agora muito maior em magnitude, e seu erro-padrão é menor.

Além das variáveis de idade na coluna (2), a coluna (3) controla a distância até a rodovia interestadual (*intst*), a área do terreno (*terreno*), a área construída (*área*), o número de quartos (*quartos*) e o número de banheiros (*banheiros*). Isso produz uma estimativa de *a81 · proxincin* mais próxima daquela sem nenhum controle, mas produz um erro-padrão muito menor: a estatística *t* de $\hat{\delta}_1$ é de cerca de −2,84. Portanto, encontramos um efeito muito mais significante na coluna (3) do que na coluna (1). As estimativas da coluna (3) são preferidas, pois controlam a maioria dos fatores e possuem os menores erros-padrão (exceto na constante, que, nesse caso, não é importante). O fato de que *proxincin* tem um coeficiente muito menor e é não significante na coluna (3) indica que as características incluídas na coluna (3) indicam em grande parte as características dos imóveis que são mais importantes para a determinação dos preços dos imóveis.

Com o propósito de introdução do método, usamos o nível de preços reais na Tabela 13.2. Faz mais sentido usar log(*preço*) [ou log(*rpreço*)] na análise, para obter um efeito percentual aproximado. O modelo básico torna-se

$$\log(preço) = \beta_0 + \delta_0 a81 + \beta_1 proxicin + \delta_1 a81 \cdot proxincin + u. \tag{13.8}$$

Agora, $100 \cdot \delta_1$ é a redução percentual aproximada nos valores dos imóveis em razão do incinerador. [Assim como no Exemplo 13.2, o uso de log(*preço*) *versus* log(*rpreço*) apenas afeta o coeficiente de *a81*.] O uso das mesmas 321 observações agrupadas produz

EXEMPLO 13.3 (continuação)

$$\widehat{\log(preço)} = 11{,}29 + 0{,}457a81 - 0{,}340\,proxincin - 0{,}063\,a81 \cdot proxincin \quad (13.9)$$
$$(0{,}31)\ (0{,}045)\ (0{,}055)\ (0{,}083)$$
$$n = 321, R^2 = 0{,}409.$$

O coeficiente do termo de interação resulta em que em razão do novo incinerador, os imóveis próximos dele perderam cerca de 6,3% em valor. Porém, essa estimativa não é estatisticamente diferente de zero. Entretanto, quando usamos um conjunto completo de controles, como na coluna (3) da Tabela 13.2 (mas com *intst, terreno* e *área* aparecendo na forma logarítmica), o coeficiente de *a81·proxincin* passa a ser −0,132, com uma estatística *t* em torno de −2,53. Novamente, o controle dos demais fatores resulta de grande importância. Usando a forma logarítmica, estimamos que os imóveis próximos do incinerador desvalorizaram-se em cerca de 13,2%.

A metodologia aplicada no exemplo anterior tem inúmeras aplicações, especialmente quando os dados são provenientes de um **experimento natural** (ou **quase-experimento**). Um experimento natural ocorre quando algum evento exógeno — frequentemente uma mudança na política governamental — altera o ambiente no qual indivíduos, famílias, empresas ou cidades operam. Um experimento natural sempre tem um grupo de controle, que não é afetado pela mudança na política, e um grupo de tratamento, que é afetado pela mudança na política. Diferentemente de um experimento verdadeiro, no qual os grupos de tratamento e de controle são escolhidos aleatória e explicitamente, esses grupos, nos experimentos naturais, surgem da mudança específica na política governamental. Para controlar diferenças sistemáticas entre os grupos de controle e de tratamento, necessitamos de dois anos de dados, um anterior à mudança na política e outro após a mudança. Assim, nossa amostra será convenientemente dividida em quatro grupos: o grupo de controle antes da mudança, o grupo de controle após a mudança, o grupo de tratamento antes da mudança e o grupo de tratamento após a mudança.

Chamemos *A* o grupo de controle e *B* o grupo de tratamento, definindo *dB* igual à unidade para os do grupo *B* de tratamento e zero, caso contrário. Então, definindo *d2* como uma variável *dummy* para o segundo período de tempo (após a mudança na política), a equação de interesse é

$$y = \beta_0 + \delta_0 d2 + \beta_1 dB + \delta_1 d2 \cdot dB + \text{outros fatores}, \quad (13.10)$$

em que y é a variável de interesse resultante. Como no Exemplo 13.3, δ_1 mede o efeito da decisão da política do governo. Sem outros fatores na regressão, $\hat{\delta}_1$ será o estimador de diferenciamento:

$$\hat{\delta}_1 = (\bar{y}_{2,B} - \bar{y}_{2,A}) - (\bar{y}_{1,B} - \bar{y}_{1,A}), \quad (13.11)$$

em que a barra significa média, o primeiro subscrito representa o ano e o segundo subscrito representa o grupo.

TABELA 13.3

Ilustração do estimador de diferença em diferenças.

	Antes	Após	Após – Antes
Controle	β_0	$\beta_0 + \delta_0$	δ_0
Tratamento	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Tratamento — Controle	β_1	$\beta_1 + \delta_1$	δ_1

A configuração geral do estimador de diferença em diferenças é mostrada na Tabela 13.3. A Tabela 13.3 sugere que o parâmetro δ_1, algumas vezes chamado de **efeito médio de tratamento** (pois ele mede o efeito do "tratamento" ou critério no resultado médio de y), pode ser estimado de duas maneiras: (1) calculando as diferenças nas médias entre os grupos de tratamento e de controle em cada período de tempo e depois tirando a primeira diferença dos resultados ao longo do tempo; da mesma forma que na equação (13.11); (2) calculando a alteração nas médias ao longo do tempo de cada um dos grupos de tratamento e controle, e então tirando a primeira diferença dessas alterações, o que significa que simplesmente escrevemos $\hat{\delta}_1 = (\bar{y}_{2,B} - \bar{y}_{1,B}) - (\bar{y}_{2,A} - \bar{y}_{1,A})$. Naturalmente, a $\hat{\delta}_1$ estimada não depende da maneira como tiramos a primeira diferença, como pode ser visto pela simples reorganização.

Quando são adicionadas variáveis explicativas na equação (13.10) (para controlar o fato de que as populações das quais foram extraídas as amostras podem diferir sistematicamente ao longo dos dois períodos), a estimativa MQO de δ_1 não mais tem a forma simples de (13.11), mas sua interpretação é semelhante.

EXEMPLO 13.4

(Efeitos das Leis de Indenizações Trabalhistas sobre os Prazos de Afastamento dos Trabalhadores)

Meyer, Viscusi & Durbin (1995) (a partir daqui, MVD) estudaram a extensão do tempo (em semanas) em que um trabalhador acidentado recebe remuneração por conta de indenização trabalhista. Em 15 de julho de 1980, o estado norte-americano de Kentucky aumentou o limite dos ganhos semanais que eram cobertos por essa remuneração. Um aumento no limite não tem efeito sobre os benefícios para os trabalhadores de baixa renda, mas torna menos oneroso para um trabalhador de alta renda permanecer afastado recebendo indenização trabalhista. Portanto, o grupo de controle é o dos trabalhadores de baixa renda, e o grupo de tratamento é o dos trabalhadores de alta renda; trabalhadores de alta renda são definidos como os que estavam posicionados no teto antes da mudança da política do governo. Usando amostras aleatórias, tanto do período anterior como do período posterior à mudança, os MVD puderam testar se uma remuneração mais generosa faria com que os trabalhadores ficassem mais tempo sem trabalhar (tudo mais mantido inalterado). Eles iniciaram com uma análise de diferença em diferenças, usando log(*duração*) como a variável dependente. Façamos *apmud* representar uma variável *dummy* das observações após a mudança da política e *altrend*, a variável dos trabalhadores de altos rendimentos. Utilizando os dados contidos no arquivo INJURY.RAW, a equação estimada, com os erros-padrão entre parênteses, é

EXEMPLO 13.4 (continuação)

$$\widehat{\log(duração)} = \underset{(0{,}031)}{1{,}126} + \underset{(0{,}0447)}{0{,}0077}\ apmud + \underset{(0{,}047)}{0{,}256}\ altrend + \underset{(0{,}069)}{0{,}191}\ apmud \cdot altrend \qquad (13.12)$$

$$n = 5.626,\ R^2 = 0{,}021.$$

Portanto, $\hat{\delta}_1 = 0{,}191$ ($t = 2{,}77$), o que implica aumento no tempo médio em que os trabalhadores de alta renda permaneceram sem trabalhar recebendo indenização trabalhista em cerca de 19% em razão do aumento do limite dos ganhos. O coeficiente de *apmud* é pequeno e estatisticamente não significante: como esperado, o aumento do limite dos ganhos não tem efeito sobre a duração do período de afastamento dos trabalhadores de baixa renda.

Esse é um bom exemplo de como podemos obter uma estimativa razoavelmente precisa do efeito de uma mudança da política governamental, mesmo que não possamos explicar muito bem a variação na variável dependente. As variáveis *dummy* em (13.12) explicam somente 2,1% da variação em log(*duração*). Isso faz sentido: existem claramente muitos fatores, inclusive a gravidade da lesão, que afetam a duração do tempo em que um trabalhador receberá indenização. Felizmente, temos uma amostra bastante grande e isso nos possibilita obter uma estatística *t* significante.

MVD também adicionaram uma variedade de controles para gênero, estado civil, tipo de atividade e tipo do ferimento. Isso leva em conta o fato de que as características das pessoas e os tipos de ferimentos podem diferir sistematicamente nos dois anos. O controle desses fatores acaba tendo pouco efeito sobre a estimativa de δ_1. (Veja o Exercício em Computador 13.4, no site da Cengage.)

QUESTÃO 13.2

O que você conclui do coeficiente e da estatística *t* de *altrend* na equação (13.12)?

Algumas vezes, os dois grupos são constituídos por pessoas que moram em dois estados norte-americanos vizinhos. Por exemplo, para avaliar o impacto da mudança dos impostos sobre o consumo dos cigarros, podemos obter amostras aleatórias dos dois estados para dois anos. No Estado A, o grupo de controle não sofreu alterações nos impostos. No Estado B, o imposto aumentou (ou foi reduzido) entre os dois anos. A variável resultante seria uma medição do consumo de cigarros, e a equação (13.10) pode ser estimada para determinar o efeito dos impostos sobre o consumo de cigarros.

Para um interessante levantamento da metodologia sobre experimentos naturais e vários exemplos adicionais, veja Meyer (1995).

13.3 ANÁLISE DE DADOS EM PAINEL DE DOIS PERÍODOS

Retornamos agora à análise do tipo mais simples de dados em painel: para um corte transversal de indivíduos, escolas, empresas, cidades, ou o que seja, temos dados de dois anos; vamos chamá-los $t = 1$ e $t = 2$. Esses anos não precisam ser adjacentes, mas $t = 1$ corresponde ao ano mais antigo. Por exemplo, o arquivo CRIME2.RAW contém dados sobre (entre outras coisas) taxas de criminalidade e de desemprego de 46 cidades em 1982 e 1987. Portanto, $t = 1$ corresponde a 1982 e $t = 2$, a 1987.

O que acontece se usarmos o corte transversal de 1987 e executarmos uma regressão simples de *txcrim* sobre *desemp*? Obteremos

$$\widehat{txcrim} = \underset{(20{,}76)}{128{,}38} - \underset{(3{,}42)}{4{,}16}\ desemp$$

$$n = 46,\ R^2 = 0{,}033.$$

Se interpretarmos a equação estimada de forma causal, ela implica um aumento na taxa de desemprego que *reduz* a taxa de criminalidade. Com certeza, isso não é o que esperávamos. O coeficiente de *desemp* não é estatisticamente significante aos níveis-padrão de significância: na melhor das hipóteses, não encontramos ligação entre as taxas de criminalidade e desemprego.

Como temos enfatizado ao longo deste texto, essa equação de regressão simples possivelmente sofre do problema de variáveis omitidas. Uma possível solução é tentar controlar mais fatores, como a distribuição por idade, a distribuição por sexo, níveis de educação, esforços para a imposição da lei etc., em uma análise de regressão múltipla. Porém, pode ser difícil controlar muitos desses fatores. No Capítulo 9, mostramos como a inclusão de *txcrim* de um ano anterior — neste caso, 1982 — pode auxiliar a controlar o fato de que cidades distintas têm taxas de criminalidade historicamente diferentes. Essa é uma maneira de usar dados de dois anos na estimativa de um efeito causal.

Um modo alternativo de usar dados em painel é separar os fatores não observados que afetam a variável dependente em dois tipos: os que são constantes e os que variam ao longo do tempo. Fazendo i representar a unidade de corte transversal e t o período de tempo, podemos escrever um modelo com uma única variável explicativa observada como

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + a_i + u_{it},\ t = 1{,}2. \qquad (13.13)$$

Na notação y_{it}, i é a pessoa, empresa, cidade etc., e t é o período de tempo. A variável $d2_t$ é uma variável *dummy* igual a zero quando $t = 1$ e um quando $t = 2$; ela não muda ao longo de i, razão pela qual ela não tem o subscrito i. Portanto, o intercepto de $t = 1$ é β_0, e o intercepto de $t = 2$ é $\beta_0 + \delta_0$. Da mesma forma, quando usamos agrupamentos independentes de cortes transversais, permitir que o intercepto mude ao longo do tempo é importante na maioria das aplicações. No exemplo sobre criminalidade, tendências seculares nos Estados Unidos farão com que as taxas de criminalidade em todas as cidades do país mudem, talvez de forma considerável, ao longo de um período de cinco anos.

A variável a_i capta todos os fatores não observados, constantes no tempo, que afetam y_{it}. (O fato de a_i não ter um subscrito t nos diz que ele não muda ao longo do tempo.) De forma genérica, a_i é chamado de **efeito não observado**. Também é comum no trabalho aplicado encontrar a_i referido como **efeito fixo**, o que nos ajuda a lembrar que a_i é fixo ao longo do tempo. O modelo em (13.13) é chamado de **modelo de efeitos não observados** ou **modelo de efeitos fixos**. Em aplicações, pode-se encontrar também a_i referido como **heterogeneidade não observada** (ou *heterogeneidade do indivíduo*, *heterogeneidade da empresa*, *heterogeneidade da cidade* etc.).

O erro u_{it} muitas vezes é chamado de **erro idiossincrático** ou erro de variação temporal, porque ele representa fatores não observados que mudam ao longo do tempo e afetam y_{it}. Eles são muito parecidos com os erros em uma equação de regressão de série temporal.

Um modelo simples de efeitos não observados da taxa de criminalidade de uma cidade em 1982 e 1987 é

$$txcrim_{it} = \beta_0 + \delta_0 d87_t + \beta_1 desemp_{it} + a_i + u_{it}, \quad \textbf{(13.14)}$$

em que *d87* é uma variável *dummy* para 1987. Como *i* representa cidades diferentes, chamamos a_i de *efeito não observado da cidade* ou *efeito fixo da cidade*: ele representa todos os fatores que afetam a taxa de criminalidade da cidade que não mudam ao longo do tempo. Detalhes geográficos, como a localização da cidade, estão incluídos em a_i. Muitos outros fatores podem não ser exatamente constantes, mas podem ser aproximadamente constantes ao longo de um período de cinco anos. Entre eles é possível encontrar características demográficas da população (idade, raça e educação). Cada cidade pode ter seus próprios métodos de registrar a criminalidade, e os habitantes dessas cidades podem ter atitudes diferentes ante a criminalidade; em geral, a mudança desses aspectos é lenta. Por razões históricas, as cidades podem ter taxas de criminalidade bastante diferentes, e os fatores históricos são efetivamente capturados pelo efeito não observado a_i.

Como devemos estimar o parâmetro de interesse, β_1, a partir de dois anos de dados em painel? Uma possibilidade é agrupar os dois anos e usar o MQO, essencialmente como na Seção 13.1. Esse método tem duas inconveniências. A mais importante delas é que, para o MQO agrupado produzir um estimador consistente de β_1, teremos que presumir o efeito não observado, a_i, como não correlacionado com x_{it}. Podemos ver isso facilmente escrevendo (13.13) como

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + v_{it}, \quad t = 1,2, \quad \textbf{(13.15)}$$

em que $v_{it} = a_i + u_{it}$ é muitas vezes chamado de **erro composto**. Pelo que conhecemos do MQO, temos que considerar ser v_{it} não correlacionado com x_{it}, em que $t = 1$ ou 2, para que o MQO estime β_1 (e os outros parâmetros) consistentemente. Isso é verdade independentemente de usarmos um único corte transversal ou agruparmos os dois cortes transversais. Portanto, mesmo presumindo que o erro idiossincrático u_{it} seja não correlacionado com x_{it}, o MQO agrupado será viesado e inconsistente se a_i e x_{it} forem correlacionados. O viés resultante no MQO agrupado algumas vezes é chamado de **viés de heterogeneidade**, mas na realidade é apenas um viés causado pela omissão de uma variável constante no tempo.

QUESTÃO 13.3

Suponha que a_i, u_{i1} e u_{i2} tenham média zero e que sejam não correlacionados dois a dois. Mostre que $\mathrm{Cov}(v_{i1},v_{i2}) = \mathrm{Var}(a_i)$, de forma que os erros compostos sejam positiva e serialmente correlacionados ao longo do tempo, a menos que $a_i = 0$. O que isso sugere sobre os erros-padrão MQO habituais da estimação MQO agrupada?

Para ilustrar o que acontece, utilizamos os dados contidos no arquivo CRIME2.RAW para estimar (13.14) pelo MQO agrupado. Como existem 46 cidades e dois anos para cada cidade, há um total de 92 observações:

$$\begin{aligned} \widehat{txcrim} &= \underset{(12,74)}{93{,}42} + \underset{(7,98)}{7{,}94}\, d87 + \underset{(1,188)}{0{,}427}\, desemp \\ n &= 92,\ R^2 = 0{,}012. \end{aligned} \quad \textbf{(13.16)}$$

(Quando descrevemos a equação estimada, normalmente abandonamos os subscritos *i* e *t*). O coeficiente de *desemp*, embora positivo em (13.16), tem uma estatística *t* muito pequena. Assim, o uso do MQO agrupado dos dois anos não mudou nada, substancialmente, em relação ao uso de um único corte transversal. Isso não surpreende, já que o uso do MQO agrupado não resolve o problema de variáveis omitidas. (Os erros-padrão nessa equação estão incorretos em razão da correlação serial descrita na Questão 13.3, mas vamos ignorar isso, já que nosso foco aqui não é o MQO agrupado.)

Na maioria das aplicações, a principal razão para coletar dados em painel é considerar que o efeito não observado, a_i, é correlacionado com as variáveis explicativas. Por exemplo, na equação sobre a criminalidade, queremos que os fatores da cidade não mensurados em a_i que afetam a taxa de criminalidade também sejam correlacionados com a taxa de desemprego. Isso acaba sendo fácil de fazer: como a_i é constante ao longo do tempo, podemos diferenciar os dados ao longo dos dois anos. Mais precisamente, para uma observação *i* de corte transversal, escreva os dois anos como

$$y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2} \ (t = 2)$$
$$y_{i1} = \beta_0 + \beta_1 x_{i1} + a_i + u_{i1} \ (t = 1).$$

Se subtrairmos a *segunda* equação da *primeira*, obtemos

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1(x_{i2} - x_{i1}) + (u_{i2} - u_{i1}),$$

ou

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i, \quad \textbf{(13.17)}$$

em que Δ representa a mudança de $t = 1$ para $t = 2$. O efeito não observado, a_i, não aparece em (13.17): ele foi "descartado pela diferenciação". Além disso, o intercepto em (13.17) é, na realidade, a *mudança* no intercepto de $t = 1$ para $t = 2$.

A equação (13.17), que chamamos de **equação de primeiras diferenças**, é muito simples. Ela é apenas uma equação única de corte transversal, mas cada variável é diferenciada ao longo do tempo. Podemos analisar (13.17) utilizando os métodos que desenvolvemos na Parte 1, desde que as hipóteses fundamentais sejam satisfeitas. A mais importante delas é que Δu_i seja não correlacionado com Δx_i. Essa hipótese será mantida se o erro idiossincrático em cada tempo *t*, u_{it}, for não correlacionado com a variável explicativa em *ambos* os períodos de tempo. Essa é outra versão da hipótese de **exogeneidade estrita** que encontramos no Capítulo 10 sobre modelos de séries temporais. Especificamente, essa hipótese exclui o caso em que x_{it} é a variável dependente defasada, $y_{i,t-1}$. Diferentemente do Capítulo 10, permitimos que x_{it} seja correlacionado com fatores não observáveis que sejam constantes ao longo do tempo. Quando obtemos o estimador MQO de β_1 de (13.17), chamamos o estimador resultante de **estimador de primeiras diferenças**.

No exemplo da criminalidade, presumir que Δu_i e $\Delta desemp_i$ sejam não correlacionados pode ser razoável, mas também pode não ser possível. Por exemplo, suponha que o empenho na imposição da lei (que está no erro idiossincrático) aumente mais nas cidades onde a taxa de desemprego diminui. Isso pode causar correlação negativa entre Δu_i e $\Delta desemp_i$, o que levaria a um viés no estimador MQO. Naturalmente, esse problema pode, até certo ponto, ser contornado pela inclusão de mais fatores na equação, assunto que veremos mais tarde. Em geral, é sempre possível que não tenhamos levado em conta suficientes fatores variáveis ao longo do tempo.

Outra condição crucial é que Δx_i deve ter alguma variação ao longo de i. Essa qualificação não se sustenta se a variável explicativa não mudar ao longo do tempo, para qualquer observação do corte transversal, ou se ela mudar pela mesma magnitude, em cada observação. Isso não é um problema no exemplo da taxa de criminalidade, pois a taxa de desemprego muda ao longo do tempo em quase todas as cidades. Mas, se i representar um indivíduo e x_{it} for uma variável *dummy* de gênero, $\Delta x_i = 0$ para todo i; claramente, não podemos estimar (13.17) por MQO nesse caso. Isso, na verdade, faz muito sentido: como permitimos que a_i seja correlacionado com x_{it}, não podemos ter esperança de separar o efeito de a_i sobre y_{it} do efeito de qualquer variável que não mude ao longo do tempo.

A única outra hipótese que necessitamos aplicar às estatísticas habituais do MQO é que (13.17) satisfaça a hipótese de homoscedasticidade. Isso é razoável em muitos casos e, se ela não se sustentar, sabemos como testar e corrigir a heteroscedasticidade utilizando os métodos do Capítulo 8. Algumas vezes é sensato presumir que (13.17) satisfaz todas as hipóteses do modelo linear clássico. Os estimadores MQO são não viesados e todas as inferências estatísticas são exatas em tais casos.

Quando estimamos a (13.17) para o exemplo da taxa de criminalidade, obtemos

$$\begin{aligned}\Delta\widehat{txcrim} &= \underset{(4,70)}{15,40} + \underset{(0,88)}{2,22}\,\Delta desemp \\ n &= 46,\ R^2 = 0,127,\end{aligned} \tag{13.18}$$

que agora fornece uma relação entre as taxas de criminalidade e desemprego positiva e estatisticamente significante. Assim, a diferenciação para eliminar os efeitos constantes no tempo faz uma grande diferença nesse exemplo. O intercepto em (13.18) também revela algo interessante. Mesmo com $\Delta desemp = 0$, espera-se um aumento na taxa de criminalidade (crimes por 1.000 pessoas) de 15,40. Isso reflete um aumento duradouro das taxas de criminalidade, por todos os Estados Unidos, de 1982 a 1987.

Mesmo que não iniciemos com o modelo de efeitos não observados (13.13), o uso da diferenciação ao longo do tempo é, intuitivamente, lógico. Em vez de estimar uma relação-padrão de corte transversal — que pode sofrer o problema de variáveis omitidas, consequentemente tornando difíceis as conclusões *ceteris paribus* — a equação (13.17) explicitamente considera como as alterações na variável explicativa ao longo do tempo afetam a alteração em y ao longo do mesmo período de tempo. Mesmo assim, ainda é muito útil ter (13.13) em mente: ela mostra explicitamente que podemos estimar o efeito de x_{it} sobre y_{it}, mantendo a_i fixo.

Embora a diferenciação de dados em painel de dois anos seja um meio poderoso de controlar efeitos não observados, ele tem um custo. Primeiro, os conjuntos de dados em painel são mais difíceis de coletar do que um corte transversal, especialmente de indivíduos. Precisamos usar uma pesquisa e acompanhar o indivíduo para uma pesquisa complementar. Muitas vezes é difícil localizar o mesmo indivíduo para uma segunda pesquisa. Em unidades como empresas, algumas delas podem falir ou passar por uma fusão com outras empresas. Dados em painel são mais fáceis de serem obtidos de escolas, cidades, municípios, estados e países.

Mesmo que tenhamos coletado um conjunto de dados em painel, a diferenciação utilizada para eliminar a_i pode reduzir bastante a variação nas variáveis explicativas. Embora x_{it} frequentemente tenha variação substancial no corte transversal para cada t, Δx_i pode não ter muita variação. Sabemos, do Capítulo 3, que pequenas variações em Δx_i podem levar a grandes erros-padrão de $\hat{\beta}_1$ quando estimamos por (13.17) MQO. Podemos combater esse fato usando um corte transversal grande, mas isso nem sempre é possível. Além disso, o uso de diferenciações maiores ao longo do tempo algumas vezes é melhor que o uso de mudanças ano a ano.

Como exemplo, considere o problema de estimar o retorno da educação, desta vez usando dados em painel de indivíduos, de dois anos. O modelo por pessoa i é

$$\log(salárioh_{it}) = \beta_0 + \delta_0 d2_t + \beta_1 educ_{it} + a_i + u_{it},\ t = 1, 2,$$

em que a_i contém aptidão não observada — que provavelmente é correlacionada com $educ_{it}$. Novamente, consideramos interceptos diferentes ao longo do tempo, para levar em conta ganhos de produtividade agregados (e inflação, se $salárioh_{it}$ estiver em termos nominais). Como, por definição, a aptidão inata não muda ao longo do tempo, os métodos de dados em painel parecem idealmente apropriados para estimar o retorno da educação. A equação de primeiras diferenças é

$$\Delta\log(salárioh_i) = \delta_0 + \beta_1\Delta educ_i + \Delta u_i, \tag{13.19}$$

e podemos fazer essa estimativa por MQO. O problema é que estamos interessados nos adultos que trabalham, e para a maioria dos indivíduos empregados, a educação não muda ao longo do tempo. Se apenas uma pequena fração de nossa amostra tiver $\Delta educ_i$ diferente de zero, será difícil obter um estimador preciso de β_1 de (13.19), a menos que tenhamos uma amostra de tamanho bastante grande. Em teoria, o uso de uma equação de primeira diferença para estimar o retorno da educação é uma boa ideia, mas ela não funciona muito bem com a maioria dos dados em painel correntemente disponíveis.

A adição de muitas variáveis explicativas não causa dificuldades. Iniciamos com o modelo de efeitos não observados

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it1} + \beta_2 x_{it2} + \ldots + \beta_k x_{itk} + a_i + u_{it}, \tag{13.20}$$

para $t = 1$ e 2. Essa equação parece mais complicada do que é, pois cada variável explicativa tem três subscritos. O primeiro representa o número da observação do corte transversal, o segundo refere-se ao período de tempo, e o terceiro é apenas um rótulo de variável.

EXEMPLO 13.5

(Dormir *Versus* Trabalhar)

Utilizamos os dados em painel de dois anos do arquivo SLP75_81.RAW, de Biddle e Hamermesh (1990), para estimar a relação de substituição entre o tempo gasto dormindo e trabalhando. No Problema 3.3, localizado no final do Capítulo 3, usamos apenas o corte transversal de 1975. O conjunto de dados em painel de 1975 e 1981 tem 239 pessoas, muito menor que o corte transversal de 1975, que inclui mais de 700 pessoas. Um modelo de efeitos não observados do total de minutos dormidos por semana é

$$\begin{aligned} dormtot_{it} = {} & \beta_0 + \delta_0 d81_t + \beta_1 trabtot_{it} + \beta_2 educ_{it} + \beta_3 casado_{it} \\ & + \beta_4 crianmen_{it} + \beta_5 boasaúde_{it} + a_i + u_{it},\quad t = 1, 2. \end{aligned}$$

O efeito não observado, a_i, será chamado de *efeito individual não observado* ou *efeito fixo não observado*. Ele é potencialmente importante para possibilitar que a_i seja correlacionada com $trabtot_{it}$: os mesmos fatores (alguns biológicos) que fazem com que as pessoas durmam mais ou menos (capturados em a_i) possivelmente são correlacionados com o tempo gasto trabalhando. Algumas pessoas têm mais energia e isso faz com

EXEMPLO 13.5 (continuação)

que elas durmam menos e trabalhem mais. A variável *educ* representa anos de escolaridade, *casado* é uma variável *dummy* indicando o estado civil, *crianmen* é uma variável *dummy* indicando a presença de criança pequena e *boasaúde* é uma variável *dummy* indicando se a pessoa goza de boa saúde. Observe que não incluímos gênero ou raça (como fizemos na análise do corte transversal), já que esses fatores não mudam ao longo do tempo; eles fazem parte de a_i. Nosso principal interesse está em β_1.

A diferenciação ao longo dos dois anos produz a equação estimável

$$\Delta dormtot_i = \delta_0 + \beta_1 \Delta trabtot_i + \beta_2 \Delta educ_i + \beta_3 \Delta casado_i + \beta_4 \Delta crianmen_i + \beta_5 \Delta boasaúde_i + \Delta u_i.$$

Presumindo que a mudança no erro idiossincrático, Δu_i, seja não correlacionada com as mudanças em todas as variáveis explicativas, podemos obter estimadores consistentes de MQO. Isso produz

$$\begin{aligned} \widehat{\Delta dormtot} = &-92{,}63 - 0{,}227\ \Delta trabtot - 0{,}024\ \Delta educ \\ &(45{,}87)\ (0{,}036) \qquad\qquad (48{,}759) \\ &+ 104{,}21\ \Delta casado + 94{,}67\ \Delta crianmen + 87{,}58\ \Delta boasaúde \\ &(92{,}86) \qquad\qquad (87{,}65) \qquad\qquad (76{,}60) \\ &n = 239,\ R^2 = 0{,}150. \end{aligned} \tag{13.21}$$

O coeficiente de $\Delta trabtot$ indica uma relação de substituição entre dormir e trabalhar: mantendo os outros fatores fixos, uma hora a mais de trabalho está associada com 0,227(60) = 13,62 minutos a menos dormindo. A estatística t (−6,31) é bastante significante. Nenhuma outra estimativa, exceto o intercepto, é estatisticamente diferente de zero. O teste F de significância conjunta de todas as variáveis, exceto $\Delta trabtot$, fornece p-valor = 0,49, e significa que elas são conjuntamente não significantes a qualquer nível razoável de significância e poderiam ser eliminadas da equação.

O erro-padrão de $\Delta educ$ é especialmente grande em relação à estimativa. Esse é o fenômeno descrito anteriormente para a equação do salário-hora. Na amostra de 239 pessoas, 183 (76,6%) não apresentam mudança na educação ao longo do período de seis anos; 90% das pessoas apresentam alteração no grau de escolaridade de, no máximo, um ano. Como está refletido pelo extremamente grande erro-padrão de $\hat{\beta}_2$, quase não existe variação suficiente na educação para estimar $\hat{\beta}_2$ com alguma precisão. De qualquer forma, $\hat{\beta}_2$ é concretamente muito pequeno.

Dados em painel também podem ser usados para estimar modelos de defasagens distribuídas finitas. Mesmo que especifiquemos a equação para somente dois anos, precisamos coletar dados de maior número de anos para obter as variáveis explicativas defasadas. O que segue é um exemplo simples.

EXEMPLO 13.6

(Defasagens Distribuídas da Taxa de Criminalidade sobre a Taxa de Esclarecimento de Crimes)

Eide (1994) utiliza dados em painel de distritos policiais da Noruega para estimar um modelo de defasagens distribuídas de taxas de criminalidade. A única variável explicativa é o "percentual de esclarecimento" de crimes (*pcescl*) — a percentagem de crimes que levaram a uma condenação. Os dados sobre a taxa de criminalidade são dos anos de 1972 e 1978. Seguindo os passos de Eide, defasamos *pcescl* para um e dois anos: é possível que as taxas de esclarecimento de crimes do passado tenham um efeito dissuasor sobre a criminalidade atual. Isso leva ao seguinte modelo de efeitos não observados dos dois anos:

EXEMPLO 13.6 (continuação)

$$\log(crime_{it}) = \beta_0 + \delta_0 d78_t + \beta_1 pcescl_{i,t-1} + \beta_2 pcescl_{i,t-2} + a_i + u_{it}.$$

Quando fazemos a diferenciação da equação e a estimamos utilizando os dados de CRIME3.RAW, obtemos

$$\begin{aligned} \widehat{\Delta\log(crime)} = &\ 0{,}086 - 0{,}0040\ \Delta pcescl_{-1} - 0{,}132\ \Delta pcescl_{-2} \\ &(0{,}064)\ (0{,}0047) \qquad\qquad (0{,}0052) \\ &n = 53,\ R^2 = 0{,}193,\ \bar{R}^2 = 0{,}161. \end{aligned} \tag{13.22}$$

A segunda defasagem é negativa e estatisticamente significante, o que sugere que uma taxa de esclarecimento de crimes mais alta dois anos atrás desencorajaria a criminalidade neste ano. Especificamente, um aumento de 10 pontos percentuais em *pcescl* dois anos atrás levaria a uma estimativa de redução na taxa de criminalidade deste ano de 13,2%. Isso sugere que o uso de mais recursos na solução de crimes e na obtenção de condenações dos criminosos pode reduzir a criminalidade no futuro.

A Organização dos Dados em Painel

Quando se usa dados em painel em estudos econométricos, é importante saber como os dados devem ser armazenados. Devemos ser cuidadosos ao arrumar os dados, de forma que os diferentes períodos de tempo da mesma unidade de corte transversal (pessoa, firma, cidade etc.) sejam facilmente encadeados. Concretamente, suponha que o conjunto de dados seja de cidades, de dois anos. Para a maioria dos propósitos, a melhor maneira de armazenar os dados é ter *dois* registros para cada cidade, um para cada ano: o primeiro registro de cada cidade corresponde ao ano mais antigo, e o segundo ao ano mais recente. Esses dois registros devem ser adjacentes. Portanto, o conjunto de dados de 100 cidades e dois anos conterá 200 registros. Os dois primeiros registros são da primeira cidade na amostra, os próximos dois são da segunda cidade, e assim sucessivamente. (Para um exemplo, veja Tabela 1.5 no Capítulo 1.) Isso facilitará a construção das diferenças para armazená-las no segundo registro de cada cidade, e fazer uma análise de agrupamento de cortes transversais, que pode ser comparada com a estimação por diferenciação.

A maioria dos conjuntos de dados em painel de dois períodos que são citados neste texto foram armazenados dessa maneira (por exemplo, CRIME2.RAW, CRIME3.RAW, GPA3.RAW, LOWBRTH.RAW e RENTAL.RAW). Usamos uma extensão direta desse esquema para conjuntos de dados em painel com mais de dois períodos de tempo.

Uma segunda maneira de organizar dois períodos de dados em painel é ter apenas um registro por unidade de corte transversal. Isso exige duas entradas para cada variável, uma para cada período de tempo. Os dados em painel do arquivo SLP75_81.RAW estão organizados dessa maneira. Cada indivíduo tem dados das variáveis *dormtot75*, *dormtot81*, *trabtot75*, *trabtot81*, e assim por diante. Criar as diferenças de 1975 a 1981 é fácil. Outros conjuntos de dados em painel com essa estrutura são TRAFFIC1.RAW e VOTE2.RAW. A inconveniência de colocar os dados em um registro é que isso impossibilita uma análise agrupada de MQO utilizando os dois períodos de tempo dos dados originais.

Além disso, esse método organizacional não funciona para conjuntos de dados em painel com mais de dois períodos de tempo, um caso que consideraremos na Seção 13.5.

13.4 ANÁLISE DE DECISÕES DE POLÍTICAS COM DADOS EM PAINEL DE DOIS PERÍODOS

Conjuntos de dados em painel são muito úteis para a análise de decisões de políticas, particularmente na avaliação de programas. Na estrutura mais simples de avaliação de programas, uma amostra de indivíduos, firmas, cidades etc. é obtida no primeiro período de tempo. Algumas dessas unidades, as pertencentes ao grupo de tratamento, farão parte de um programa específico em um período de tempo posterior; as que não farão parte estão no grupo de controle. Isso é semelhante à literatura sobre experimentos naturais discutida anteriormente, com uma importante diferença: as *mesmas* unidades do corte transversal aparecem em cada período de tempo.

Como exemplo, suponha que queiramos avaliar o efeito de um programa de treinamento de pessoal de Michigan sobre a produtividade dos trabalhadores de firmas manufatureiras (veja também o Exercício em Computador 9.3, no site da Cengage). Façamos ref_{it} representar a taxa de refugo dos produtos da firma i durante o ano t (o número de itens, em cada 100, que devem ser rejeitados devido a defeitos). Seja $subs_{it}$ um indicador binário igual a um se a firma i no ano t recebeu subsídio de treinamento de pessoal. Para os anos de 1987 e 1988, o modelo é

$$ref_{it} = \beta_0 + \delta_0 a88_t + \beta_1 subs_{it} + a_i + u_{it}, \quad t = 1, 2, \tag{13.23}$$

em que $a88_t$ é uma variável *dummy* para 1988 e a_i é o *efeito não observado da firma* ou o *efeito fixo da firma*. O efeito não observado contém elementos como a aptidão média dos empregados, capital e capacidade gerencial; esses fatores são, em linhas gerais, constantes ao longo de um período de dois anos. Estamos preocupados em saber se a_i está sistematicamente relacionado ao fato de uma firma receber subsídio. Por exemplo, os administradores do programa poderiam dar prioridade às firmas cujos trabalhadores tenham menos especialização. Ou o problema oposto poderia ocorrer: para que o treinamento de pessoal pareça eficiente, os administradores do programa poderiam conceder subsídios a empregadores com trabalhadores mais produtivos. Na realidade, neste programa específico, os subsídios foram conferidos na ordem de entrada das solicitações. Porém, e se o fato de uma firma entrar com a solicitação na frente estiver relacionado com a produtividade do trabalhador? Nesse caso, uma análise usando um único corte transversal ou somente agrupando os cortes transversais produzirá estimadores viesados e não consistentes.

Fazendo a diferenciação para remover a_i, obtemos

$$\Delta ref_i = \delta_0 + \beta_1 \Delta subs_i + \Delta u_i. \tag{13.24}$$

Portanto, simplesmente regredimos as mudanças na taxa de refugo sobre as mudanças do indicador de subsídio. Como nenhuma firma recebeu subsídio em 1987, $subs_{i1} = 0$ para todo i, e assim $\Delta subs_i = subs_{i2} - subs_{i1} = subs_{i2}$, que simplesmente indica se a firma recebeu algum subsídio em 1988. Porém, geralmente é importante diferenciar todas as variáveis (inclusive as variáveis *dummy*), pois isso é necessário para remover a_i no modelo de efeitos não observados (13.23).

A estimação da equação de primeiras diferenças utilizando os dados contidos no arquivo JTRAIN.RAW produz

$$\widehat{\Delta ref} = -0{,}564 - 0{,}739\ \Delta subs$$
$$(0{,}405)\quad (0{,}683)$$
$$n = 54,\ R^2 = 0{,}022.$$

Portanto, estimamos que o fato de haver um subsídio de treinamento reduziu a taxa de refugo, na média, em $-0{,}739$. Mas a estimativa não é estatisticamente diferente de zero.

Obtemos resultados mais consistentes se usarmos $\log(ref)$ e estimarmos o efeito percentual:

$$\widehat{\Delta \log(ref)} = -0{,}057 - 0{,}317\ \Delta subs$$
$$(0{,}097)\quad (0{,}164)$$
$$n = 54,\ R^2 = 0{,}067.$$

A taxa de refugo estimada reduziu-se em cerca de 27,2% devido ao subsídio de treinamento de pessoal. [Obtivemos essa estimativa da equação (7.10): $\exp(-0{,}317) - 1 \approx -0{,}272$.] A estatística t é de cerca de $-1{,}93$, que é marginalmente significante. Em contraposição, usando o MQO agrupado de $\log(ref)$ sobre $a88$ e $subs$ produz $\hat{\beta}_1 = 0{,}057$ (erro-padrão $= 0{,}431$). Assim, não encontramos qualquer relação significante entre a taxa de refugo e o subsídio de treinamento de pessoal. Como essa conclusão difere muito da estimativa de primeiras diferenças, ela sugere que as firmas com empregados menos especializados estão mais propensas a receber subsídio de treinamento.

É útil estudar o modelo de avaliação do programa de forma mais generalizada. Façamos y_{it} representar uma variável de resultado e $prog_{it}$ uma variável *dummy* de participação no programa. O modelo mais simples de efeitos não observados é

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 prog_{it} + a_i + u_{it}. \tag{13.25}$$

Se a participação no programa somente ocorreu no segundo período, então, o estimador MQO de β_1 na equação diferenciada terá uma representação muito simples:

$$\hat{\beta}_1 = \overline{\Delta y}_{trat} - \overline{\Delta y}_{control}. \tag{13.26}$$

Ou seja, calculamos a média da mudança em y ao longo dos dois períodos de tempo para os grupos de tratamento e de controle. Então, $\hat{\beta}_1$ será a diferença entre eles. Essa é a versão de dados em painel do estimador de diferença em diferenças na equação (13.11) de dois cortes transversais agrupados. Com dados em painel, temos uma vantagem potencialmente importante: podemos diferenciar y ao longo do tempo para as *mesmas* unidades de corte transversal. Isso nos possibilita controlar efeitos específicos de pessoas, firmas ou cidades, como deixa claro o modelo em (13.25).

Se a participação no programa ocorrer nos dois períodos, $\hat{\beta}_1$ não pode ser escrito como em (13.26), mas o interpretamos da mesma maneira: ele é a mudança no valor médio de y em razão da participação no programa.

O controle de fatores que variam ao longo do tempo não altera nada que tenha importância. Simplesmente diferenciamos tais variáveis e as incluímos com $\Delta prog$. Isso nos possibilita controlar as variáveis com variação temporal que possam estar correlacionadas com a especificação do programa.

O mesmo método de diferenciação funciona na análise dos efeitos de qualquer política que varie entre cidades ou estados. O que segue é um exemplo simples.

EXEMPLO 13.7

(Efeitos da Legislação a Respeito da Condução de Veículos sob Embriaguez sobre as Fatalidades no Trânsito)

Muitos estados dos Estados Unidos vêm adotando diferentes políticas, na tentativa de coibir a condução de veículos sob embriaguez. Dois tipos de leis que estudaremos aqui são as *leis de recipientes abertos* — que consideram ilegal os passageiros de um veículo ter em seu poder recipientes abertos de bebidas alcoólicas — e as *leis administrativas propriamente ditas* — que autorizam a Justiça a suspender a carteira de habilitação do motorista preso por dirigir embriagado, mesmo antes de sua condenação. Uma análise possível é usar um único corte transversal de estados para regredir fatalidades no trânsito (relacionadas com o ato de dirigir embriagado) sobre variáveis *dummy* indicadoras da presença de cada lei. Há pouca possibilidade de isso funcionar de forma satisfatória, pois os estados decidem, por meio de processos legislativos, se eles precisam de tais leis. Portanto, a presença de leis possivelmente estará relacionada com a média das fatalidades no trânsito ocorridas nos anos recentes. Uma análise mais convincente utiliza dados em painel de um período de tempo em que alguns estados tenham adotado novas leis (e alguns estados que tenham revogado as leis até então existentes). O arquivo TRAFFIC1.RAW contém dados de 1985 e 1990 de todos os 50 estados norte-americanos e do Distrito de Colúmbia. A variável dependente é o número de mortes no trânsito por 100 milhões de milhas (*txmort*). Em 1985, 19 estados tinham leis de recipientes abertos, enquanto 22 estados tinham tais leis em 1990. Em 1985, 21 estados tinham leis administrativas propriamente ditas; esse número subiu para 29 em 1990.

O uso de MQO após a primeira diferença produz

$$\Delta \widehat{txmort} = -0{,}497 - 0{,}420\ \Delta abertos - 0{,}151\ \Delta admin \quad (13.27)$$
$$(0{,}052)\quad (0{,}206) \qquad\qquad (0{,}117)$$
$$n = 51,\ R^2 = 0{,}119.$$

As estimativas sugerem que a adoção de uma lei de recipientes abertos reduziu a taxa de fatalidades no trânsito em 0,42, um efeito nada desprezível considerando que a taxa média de mortalidade em 1985 era de 2,7 com um desvio-padrão de cerca de 0,6. A estimativa é estatisticamente significante ao nível de 5% contra uma alternativa bilateral. A lei administrativa propriamente dita tem um efeito menor, e sua estatística *t* é de somente −1,29; mas as estimativas dão os sinais que esperávamos. O intercepto nessa equação mostra que as fatalidades no trânsito caíram substancialmente em todos os estados ao longo do período de cinco anos, tenha ou não havido mudanças de leis. Os estados que adotaram uma lei de recipientes abertos durante esse período observaram uma redução adicional, em média, nas taxas de mortalidade.

Outras leis podem também afetar as fatalidades no trânsito, como as leis sobre o uso de cinto de segurança, leis sobre o uso de capacetes por motociclistas e sobre limites máximos de velocidade. Além disso, podemos querer controlar as distribuições por idade e gênero, como também avaliar a influência que organizações como a Mothers Against Drunk Driving (Mães contra a Condução de Veículos sob Embriaguez) têm em cada estado.

QUESTÃO 13.4

No exemplo 13.7, $\Delta admin = -1$ para o Estado de Washington. Explique o que isso significa.

13.5 A DIFERENCIAÇÃO COM MAIS DE DOIS PERÍODOS DE TEMPO

Também podemos usar a diferenciação com mais de dois períodos de tempo. Como ilustração, suponha que temos N indivíduos e $T = 3$ períodos de tempo para cada indivíduo. Um modelo genérico de efeitos fixos é

$$y_{it} = \delta_1 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}, \quad (13.28)$$

para $t = 1$, 2 e 3. (O número total de observações é, portanto, $3N$.) Observe que agora incluímos duas *dummies* de período de tempo em adição ao intercepto. É uma boa ideia deixar um intercepto separado para cada período de tempo, especialmente quando temos um pequeno número deles. O período-base, como sempre, é $t = 1$. O intercepto do segundo período é $\delta_1 + \delta_2$, e assim por diante. Nosso primeiro interesse está em $\beta_1, \beta_2, \ldots, \beta_k$. Se o efeito não observado a_i for correlacionado com qualquer das variáveis explicativas, o uso do MQO agrupado nos três anos de dados resultará em estimativas viesadas e inconsistentes.

A hipótese crucial é a de que os erros idiossincráticos são não correlacionados com as variáveis explicativas em cada período de tempo:

$$\text{Cov}(x_{itj}, u_{is}) = 0, \text{ para todo } t, s \text{ e } j. \quad (13.29)$$

Ou seja, as variáveis explicativas são *estritamente exógenas* após retirarmos o efeito não observado a_i. (A hipótese de exogeneidade estrita, especificada em termos de uma expectativa condicional zero, é explicada no Apêndice deste capítulo.) A Hipótese (13.29) impede casos em que variáveis explicativas futuras reagem às mudanças correntes nos erros idiossincráticos, como deve ser o caso se x_{itj} for uma variável dependente defasada. Se omitirmos uma variável com variação temporal importante, então, (13.29) geralmente será violada. Erros de medida em uma ou mais das variáveis explicativas podem fazer com que (13.29) seja falsa, como no Capítulo 9. Nos Capítulos 15 e 16, discutiremos o que pode ser feito em tais casos.

Se a_i for correlacionado com x_{itj}, então, x_{itj} será correlacionado com o erro *composto*, $v_{it} = a_i + u_{it}$, sob (13.29). Podemos eliminar a_i fazendo a diferenciação dos períodos adjacentes. No caso de $T = 3$, subtraímos o período de tempo um do período de tempo dois e o período de tempo dois do período de tempo três. Isso produz

$$\Delta y_{it} = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \beta_1 \Delta x_{it1} + \ldots + \beta_k \Delta x_{itk} + \Delta u_{it}, \quad (13.30)$$

para $t = 2$ e 3. Não temos uma equação diferenciada para $t = 1$ por não haver nada a ser subtraído da equação $t = 1$. Agora, (13.30) representa *dois* períodos de tempo para cada indivíduo da amostra. Se essa equação satisfizer as hipóteses do modelo linear clássico, o MQO agrupado produzirá estimadores não viesados e as estatísticas t e F usuais serão válidas para a hipótese. Podemos também recorrer a resultados assimptóticos. O requisito importante para que o MQO seja consistente é Δu_{it} ser não correlacionado com Δx_{itj} para todo j e $t = 2$ e 3. Essa é a extensão natural do caso com dois períodos de tempo.

Observe como (13.30) contém as diferenças das *dummies* anuais, $d2_t$ e $d3_t$. Para $t = 2$, $\Delta d2_t = 1$ e $\Delta d3_t = 0$; para $t = 3$, $\Delta d2_t = -1$ e $\Delta d3_t = 1$. Portanto, (13.30) não contém um intercepto. Isso é inconveniente para certos propósitos, inclusive no cálculo do R-quadrado. A menos que os interceptos de tempo no modelo original (13.28) sejam de interesse direto — raramente eles são — é melhor estimar a equação de primeiras diferenças com um intercepto e um único período *dummy* de tempo, normalmente para o terceiro período. Em outras palavras, a equação passa a ser

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + \beta_1 \Delta x_{it1} + \ldots + \beta_k \Delta x_{itk} + \Delta u_{it}, \quad \text{para } t = 2 \text{ e } 3.$$

As estimativas de β_j são idênticas em qualquer das formulações.

Com mais de três períodos de tempo, a ideia é semelhante. Se tivermos os mesmos T períodos de tempo para cada N unidades de corte transversal, dizemos que esse conjunto de dados é um **painel equilibrado**: temos os mesmos períodos de tempo para todos os indivíduos, firmas, cidades etc. Quando T for pequeno em relação a N, devemos incluir uma variável *dummy* para cada período de tempo e levar em conta as mudanças duradouras que não estejam sendo modeladas. Portanto, após fazermos a primeira diferença, a equação parecerá com

$$\Delta y_{it} = \alpha_0 + \alpha_3 d3_t + \alpha_4 d4_t + \ldots + \alpha_T dT_t + \beta_1 \Delta x_{it1} + \ldots + \beta_k \Delta x_{itk} + \Delta u_{it}, \quad t = 2, 3, \ldots, T, \qquad \textbf{(13.31)}$$

em que temos $T - 1$ períodos de tempo em cada unidade i para a equação de primeiras diferenças. O número total de observações é $N(T - 1)$.

É simples estimar (13.31) pelo MQO agrupado, desde que as observações tenham sido adequadamente organizadas e a diferenciação tenha sido feita cuidadosamente. Para facilitar a primeira diferença, o arquivo de dados deve consistir de NT registros. Os primeiros registros T são das observações do primeiro corte transversal, ordenados cronologicamente; os segundos registros T são das observações do segundo corte transversal, ordenados cronologicamente, e assim por diante. Depois, calculamos as diferenças, armazenando a mudança de $t - 1$ para t no registro de tempo t. Portanto, as diferenças de $t = 1$ devem ser os valores ausentes de todas as N observações do corte transversal. Se não fizermos isso, corremos o risco de obter observações fictícias na análise de regressão. Uma observação inválida é criada quando a última observação, digamos, da pessoa $i - 1$ é subtraída da primeira observação da pessoa i. Se fizermos a regressão sobre os dados diferenciados, e NT ou $NT - 1$ observações forem registradas, é porque esquecemos de definir como ausentes as observações $t = 1$.

Quando usamos mais de dois períodos de tempo, devemos supor que Δu_{it} é não correlacionado ao longo do tempo, para que os habituais erros-padrão e estatísticas de testes sejam válidos. Essa hipótese é razoável algumas vezes, mas ela não se sustenta se presumirmos que os erros idiossincráticos originais, u_{it}, são não correlacionados ao longo do tempo (uma hipótese que usaremos no Capítulo 14). De fato, se presumirmos que os u_{it} são serialmente não correlacionados com variância constante, é possível mostrar que a correlação entre Δu_{it} e $\Delta u_{i,t+1}$ é igual a $-0{,}5$. Se u_{it} seguir um modelo AR(1) estável, Δu_{it} será serialmente correlacionado. Somente quando u_{it} seguir um passeio aleatório Δu_{it} será serialmente não correlacionado.

É fácil testar a existência de correlação serial na equação de primeiras diferenças. Seja $r_{it} = \Delta u_{it}$ a primeira diferença do erro original. Se r_{it} seguir o modelo AR(1) $r_{it} = \rho r_{i,t-1} + e_{it}$, podemos testar com facilidade H_0: $\rho = 0$. Primeiro, estimamos (13.31) pelo MQO agrupado e obtemos os resíduos $\hat{r}_{it}$.

Depois, executamos uma regressão simples por MQO agrupado de $\hat{r}_{it}$ sobre $\hat{r}_{i,t-1}$, $t = 3, \ldots, T$, $i = 1, \ldots, N$ e calculamos o teste t padrão do coeficiente de $\hat{r}_{i,t-1}$. (Ou podemos tornar a estatística t robusta quanto à heteroscedasticidade.) O coeficiente $\hat{\rho}$ de $\hat{r}_{i,t-1}$ é um estimador consistente de ρ. Como estamos usando o resíduo defasado, perdemos outro período de tempo. Por exemplo, se iniciamos com $T = 3$, a equação diferenciada tem dois períodos de tempo, e o teste para verificar a existência de correlação serial é simplesmente uma regressão de corte transversal dos resíduos do terceiro período de tempo sobre os resíduos do segundo período de tempo. Daremos Exemplo mais tarde.

Podemos corrigir a presença da correlação serial AR(1) em r_{it} com o uso do MQG factível. Fundamentalmente, em cada observação do corte transversal, utilizamos a transformação de Prais-Winsten com base em $\hat{\rho}$ descrita no parágrafo anterior. (Sem dúvida, preferimos, nesse caso, o método de Prais-Winsten ao de Cochrane-Orcutt, pois a eliminação do primeiro período de tempo agora significará perder N observações do corte transversal.) Infelizmente, programas-padrão que efetuam correções AR(1) em regressões de séries temporais não funcionam nesse caso. Os métodos-padrão de Prais-Winsten tratarão as observações como se elas tivessem seguido um processo AR(1) ao longo de i e t; isso não faz sentido, pois estamos presumindo que as observações são independentes ao longo de i. Correções nos erros-padrão de MQO que permitam formas arbitrárias de correlação serial (e heteroscedasticidade) podem ser calculadas quando N é grande (e N deve ser consideravelmente maior que T). Um tratamento detalhado desses tópicos está além do escopo deste texto [veja Wooldridge (2002, Capítulo 10)], mas eles são fáceis de serem rodados em alguns programas de regressão.

QUESTÃO 13.5

A correlação serial em Δu_{it} faz com que o estimador de primeiras diferenças seja viesado e inconsistente? Por que a correlação serial é uma preocupação?

Se não houver correlação serial nos erros, os métodos habituais de tratar a heteroscedasticidade são válidos. Podemos usar os testes de heteroscedasticidade de Breusch-Pagan e de White do Capítulo 8, e também podemos calcular erros-padrão robustos.

Diferenciar mais de dois anos de dados em painel é muito útil para a análise de decisão de políticas, como mostrado no exemplo seguinte.

EXEMPLO 13.8

(Efeitos das Zonas Industriais sobre os Pedidos de Seguro-Desemprego)

Papke (1994) estudou o efeito do programa de instalação de zonas industriais (ZI) no Estado norte-americano de Indiana sobre os pedidos de seguro-desemprego. Ela analisou 22 cidades de Indiana ao longo do período de 1980 a 1988. Seis zonas industriais foram criadas em 1984 e mais quatro em 1985. Doze das cidades da amostra não criaram zonas industriais nesse período; elas serviram como grupo de controle.

Um modelo simples de avaliação de política

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

em que $uclms_{it}$ é o número de pedidos de seguro-desemprego registrados durante o ano t na cidade i. O parâmetro θ_t apenas representa um intercepto diferente para cada período de tempo. De forma geral, os pedidos

EXEMPLO 13.8 (continuação)

de seguro-desemprego estavam caindo em âmbito estadual ao longo desse período, e isso deveria estar refletido nos interceptos dos diferentes anos. A variável binária zi_{it} é igual a um se a cidade i no tempo t possuía uma zona industrial; estamos interessados em β_1. O efeito não observado a_i representa fatores fixos que afetam o meio ambiente econômico na cidade i. Como o estabelecimento de zonas industriais não foi feito de maneira aleatória — zonas industriais normalmente são estabelecidas em áreas economicamente debilitadas — é provável que zi_{it} e a_i sejam positivamente correlacionados (a_i elevado significa maior número de pedidos de seguro-desemprego, o que leva a uma maior probabilidade de ser criada uma ZI). Assim, devemos diferenciar a equação para eliminar a_i:

$$\Delta\log(uclms_{it}) = \alpha_0 + \alpha_1 d82_t + \ldots + \alpha_7 d88_t + \beta_1 \Delta zi_{it} + \Delta u_{it}. \quad \textbf{(13.32)}$$

A variável dependente nessa equação, a mudança em log($uclms_{it}$), é a taxa aproximada de crescimento anual nos pedidos de seguro-desemprego do ano $t - 1$ para t. Podemos estimar essa equação para os anos de 1981 a 1988, utilizando os dados contidos no arquivo EZUNEM.RAW; o tamanho total da amostra é $22 \cdot 8 = 176$. A estimativa de β_1 é $\hat{\beta}_1 = -0{,}182$ (erro-padrão $= 0{,}078$). Portanto, parece que a presença de uma ZI provoca uma queda aproximada de 16,6% [$\exp(-0{,}182) - 1 \approx -0{,}166$] nos pedidos de seguro-desemprego. Esse é um efeito economicamente grande e estatisticamente significante.

Não há comprovação de heteroscedasticidade na equação: o teste F de Breusch-Pagan produz $F = 0{,}85$, p-valor $= 0{,}557$. Porém, quando adicionamos os resíduos defasados do MQO à equação diferenciada (e perdemos o ano de 1981), obtemos $\hat{\rho} = -0{,}197$ ($t = -2{,}44$), de modo que há evidência de uma mínima correlação serial negativa nos erros em primeiras diferenças. Ao contrário da correlação serial positiva, os erros-padrão usuais do MQO podem não subestimar muito os erros-padrão corretos quando os erros são negativamente correlacionados (veja Seção 12.1 no Capítulo 12). Assim, a significância da variável *dummy* da zona industrial provavelmente não será afetada.

EXEMPLO 13.9

(Taxas de Criminalidade Municipais na Carolina do Norte)

Cornwell e Trumbull (1994) usaram dados de 90 municípios da Carolina do Norte, dos anos de 1981 a 1987, para estimar um modelo de efeitos não observados da criminalidade; os dados estão contidos no arquivo CRIME4.RAW. Aqui, estimamos uma versão mais simples do modelo deles, e fazemos a diferenciação da equação ao longo do tempo para eliminar a_i, o efeito não observado. (Cornwell e Trumbull utilizam uma transformação diferente, sobre a qual discorreremos no Capítulo 14.) Vários fatores, inclusive localização geográfica, atitudes diante da criminalidade, registros históricos e convenções sobre os registros podem estar contidos em a_i. A taxa de criminalidade é o número de crimes por pessoa, *prpris* é a probabilidade estimada de prisão, *prcond* é a probabilidade de condenação (tendo havido uma prisão), *prspris* é a probabilidade de cumprir pena prisional (tendo havido uma condenação), *sentmed* é a duração média da sentença cumprida, e *polpc* é o número de policiais *per capita*. Como é padrão em estudos criminométricos, usamos os logs de todas as variáveis para estimar a elasticidade. Também incluímos um conjunto completo de *dummies* anuais para controlar as tendências estaduais das taxas de criminalidade. Podemos usar os anos de 1982 a 1987 para estimar a equação diferenciada. As quantidades entre parênteses são os erros-padrão habituais do MQO; as quantidades entre colchetes são erros-padrão robustos tanto quanto à correlação serial como quanto à heteroscedasticidade:

EXEMPLO 13.9 (continuação)

$$\begin{aligned}
\Delta\widehat{\log(txcrim)} = {} & \underset{\substack{(0,017)\\[0,014]}}{0{,}008} - \underset{\substack{(0,024)\\[0,022]}}{0{,}100}\, d83 - \underset{\substack{(0,024)\\[0,020]}}{0{,}048}\, d84 - \underset{\substack{(0,023)\\[0,025]}}{0{,}005}\, d85 \\
& + \underset{\substack{(0,024)\\[0,021]}}{0{,}028}\, d86 + \underset{\substack{(0,024)\\[0,024]}}{0{,}041}\, d87 - \underset{\substack{(0,030)\\[0,056]}}{0{,}327}\, \Delta\log(prpris) \\
& - \underset{\substack{(0,018)\\[0,040]}}{0{,}238}\, \Delta\log(prcond) - \underset{\substack{(0,026)\\[0,046]}}{0{,}165}\, \Delta\log(prspris) \\
& - \underset{\substack{(0,022)\\[0,026]}}{0{,}022}\, \Delta\log(sentmed) + \underset{\substack{(0,027)\\[0,103]}}{0{,}398}\, \Delta\log(polpc) \\
& n = 540,\ R^2 = 0{,}433,\ \bar{R}^2 = 0{,}422.
\end{aligned} \quad \textbf{(13.33)}$$

As três variáveis de probabilidade — de prisão, de condenação e de cumprir pena prisional — têm o indício esperado, e todas são estatisticamente significantes. Por exemplo, estima-se que um aumento de 1% na probabilidade de prisão reduza a taxa de criminalidade em torno de 0,33%. A variável de duração média da sentença mostra um modesto efeito dissuasor, mas não estatisticamente significante.

O coeficiente da variável número de policiais *per capita* é algo surpreendente e é uma característica da maioria dos estudos que buscam explicar a taxa de criminalidade. Interpretado de maneira causal, ele diz que o aumento de 1% no número de policiais *per capita aumenta* a taxa de criminalidade em cerca de 0,4%. (A estatística t habitual é muito grande, quase 15.) É difícil acreditar que um número maior de policiais fará com que ocorra um maior número de crimes. O que estará acontecendo aqui? Existem pelo menos duas possibilidades. Primeiro, a variável da taxa de criminalidade é calculada com base nos crimes *denunciados*. Pode ser que, quando há mais policiais, mais crimes são registrados. Segundo, a variável do número de policiais pode ser endógena na equação por outras razões: os municípios podem aumentar a força policial quando estão prevendo um aumento na criminalidade. Nesse caso, (13.33) não pode ser interpretada de forma causal. Nos Capítulos 15 e 16, discorreremos sobre modelos e métodos de estimação que podem levar em conta essa forma adicional de endogeneidade.

O caso especial do teste de White quanto à heteroscedasticidade na Seção 8.3 do Capítulo 8, produz $F = 75{,}48$ e p-valor $= 0{,}0000$, de modo que haja forte evidência de heteroscedasticidade. (Tecnicamente, esse teste não será válido se também houver correlação serial, mas ele é bastante sugestivo.) O teste da existência de correlação serial AR(1) produz $\hat{\rho} = -0{,}233$, $t = -4{,}77$, o que significa que existe correlação serial negativa. Os erros-padrão entre colchetes fazem os ajustes da correlação serial e heteroscedasticidade. [Não daremos os detalhes disso; os cálculos são semelhantes aos descritos na Seção 12.5 do capítulo 12 e são executados por muitos dos programas econométricos. Veja Wooldridge (2002, Capítulo 10) para mais detalhes.] Nenhuma variável perdeu significância estatística, mas as estatísticas t nas variáveis dissuasórias significantes foram notavelmente menores. Por exemplo, a estatística t da variável de probabilidade de condenação vai de $-13{,}22$ com o uso do erro-padrão usual do MQO para $-6{,}10$ com o uso do erro-padrão totalmente robusto. De forma equivalente, os intervalos de confiança construídos com o uso dos erros-padrão robustos serão, de forma apropriada, muito mais amplos do que os baseados nos erros-padrão habituais do MQO.

Naturalmente, podemos aplicar o teste de Chow em modelos de dados em painel estimados pela primeira diferença. Como no caso das seções transversais agrupadas, raramente queremos testar se os interceptos são constantes ao longo do tempo; por razões diversas, esperamos que os interceptos sejam diferentes. Muito mais interessante é testar se os coeficientes de inclinação mudaram ao longo do tempo, e podemos facilmente realizar tais testes pela interação das variáveis explicativas de interesse com as variáveis *dummy* de período de tempo. Curiosamente, embora não possamos estimar as inclinações de variáveis que não mudam ao longo do tempo, podemos testar se os efeitos parciais das variáveis constantes no tempo sofreram alterações ao longo do tempo. Como uma ilustração, suponha que observamos três anos de dados numa amostra aleatória de pessoas que estavam trabalhando nos anos de 2000, 2002 e 2004, e especifiquemos o modelo (para o log de salário-hora, *lsalárioh*),

$$\begin{aligned} lsalárioh_{it} = {} & \beta_0 + \delta_1 d02_t + \delta_2 d04_t + \beta_1 feminino_i + \gamma_1 d02_t feminino_i \\ & + \gamma_2 d04_t feminino_i + \mathbf{z}_{it}\boldsymbol{\lambda} + a_i + u_{it}, \end{aligned}$$

em que $\mathbf{z}_{it}\boldsymbol{\lambda}$ é uma forma abreviada das outras variáveis explicativas incluídas no modelo e seus coeficientes. Quando fazemos a primeira diferença, eliminamos o intercepto do ano 2000, β_0, e também a diferença salarial entre os sexos de 2000, β_1. Porém, a alteração na $d01_t feminino_i$ é $(\Delta d01_t) feminino_i$, que não é eliminada. Consequentemente podemos estimar como a diferença salarial se alterou em 2002 e 2004 em relação a 2000, e podemos testar se $\gamma_1 = 0$, ou $\gamma_2 = 0$, ou ambos. Também devemos perguntar se o prêmio nos salários dos sindicalizados mudou ao longo do tempo, caso em que incluiríamos no modelo $sindicato_{it}$, $d02_t sindicato_{it}$, e $d04_t sindicato_{it}$. Os coeficientes de todas estas variáveis explicativas podem ser estimados, pois a $sindicato_{it}$, presumivelmente teria alguma variação de tempo.

Se alguém tentar estimar um modelo que contenha interações por diferenciamento feitas manualmente, poderá ser um pouco complicado. Por exemplo, na equação anterior com o *status* de sindicato, devemos simplesmente diferenciar os termos de interação $d02_t sindicato_{it}$ e $d04_t sindicato_{it}$. *Não podemos* calcular as diferenças exatas, digamos, $d02_t \Delta sindicato_{it}$ e $d04_t \Delta sindicato_{it}$, ou mesmo substituir $d02_t$ e $d04_t$ por suas primeiras diferenças.

Como uma observação de caráter geral, é importante retornarmos ao modelo original e lembrarmos que o diferenciamento é usado para eliminarmos a_i. É mais fácil usar um comando embutido que possibilite primeira diferença como uma opção na análise de dados em painel. (Veremos algumas das outras opções no Capítulo 14.)

Armadilhas Potenciais na Primeira Diferença de Dados em Painel

Nesta e nas seções anteriores, temos argumentado que tirar a diferença de dados em painel ao longo do tempo, para eliminar um efeito não observado constante no tempo, é um método valioso para obter-se efeito causal. Não obstante, a diferença não é isenta de dificuldades. Já discutimos problemas potenciais com o método quando as principais variáveis explicativas não têm grandes variações ao longo do tempo (e o método é inútil com variáveis explicativas que nunca variam ao longo do tempo). Infelizmente, mesmo quando temos suficiente variação temporal na x_{itj} a estimação pelas primeiras diferenças (PD) pode estar sujeita a sérios vieses. Já mencionamos que a exogeneidade estrita dos regressores é uma hipótese crítica. Infelizmente (como discutido em Wooldridge (2002, Seção 11.1), tendo-se mais períodos temporais em geral não reduz a inconsistência no estimador PD quando os regressores não são estritamente exógenos (digamos, se $y_{i,t-1}$ estiver incluída entre as x_{itj}).

Outra importante inconveniência do estimador PD é que ele pode ser pior que o MQO agrupado se uma ou mais das variáveis explicativas estiver sujeita a erro de medição, especialmente o modelo de erros clássicos nas variáveis discutido na Seção 9.3. Tirar a diferença de um regressor deficientemente medido reduz sua variação em relação a sua correlação com o erro diferenciado causado pelo erro de medição clássico, resultando num viés potencialmente considerável. Resolver tais problemas pode ser muito difícil. Veja a Seção 15.8 e Wooldridge (2002, Capítulo 11).

RESUMO

Estudamos métodos para analisar conjuntos de cortes transversais agrupados independentemente de dados em painel. Cortes transversais independentes surgem quando diferentes amostras aleatórias são obtidas em diferentes períodos de tempo (geralmente anos). O MQO usando dados agrupados é o principal método de estimação, e os procedimentos habituais de inferência são eficazes, inclusive a correção da heteroscedasticidade. (A correlação serial não é um problema, pois as amostras são independentes ao longo do tempo.) Em razão da dimensão das séries temporais, frequentemente admitimos diferentes interceptos temporais. Podemos também interagir *dummies* temporais com determinadas variáveis cruciais para verificar como elas mudaram ao longo do tempo. Isso é especialmente importante na literatura de avaliação das políticas em experimentos naturais.

Os conjuntos de dados em painel estão sendo cada vez mais usados no trabalho aplicado, especialmente na análise de políticas. Eles são conjuntos de dados nos quais as mesmas unidades de corte transversal são acompanhadas ao longo do tempo. Os conjuntos de dados em painel são muito úteis quando se quer controlar características não observadas constantes no tempo — de pessoas, firmas, cidades etc. — que pensamos poderem estar correlacionadas com as variáveis explicativas de nosso modelo. Uma maneira de remover o efeito não observado é diferenciar os dados nos períodos de tempo adjacentes. Assim, uma análise padrão MQO das diferenças pode ser usada. O uso de dados de dois períodos resulta em uma regressão de corte transversal dos dados diferenciados. Os procedimentos habituais de inferência são assimptoticamente válidos sob homoscedasticidade; a inferência exata é acessível em condições de normalidade.

Para mais de dois períodos de tempo, podemos usar o MQO agrupado sobre os dados diferenciados: perdemos o primeiro período de tempo em razão da diferenciação. Além da homoscedasticidade, devemos presumir que os erros *diferenciados* são serialmente não correlacionados, para podermos aplicar as estatísticas *t* e *F* usuais. (O apêndice deste capítulo contém uma lista meticulosa das hipóteses.) Naturalmente, qualquer variável que seja constante ao longo do tempo é eliminada da análise.

PROBLEMAS

13.1 No exemplo 13.1, suponha que as médias de todos os fatores, exceto *educ*, tenha permanecido constante ao longo do tempo e que o nível médio de educação seja 12,2 na amostra de 1972 e 13,3 na amostra de 1984. Utilizando as estimativas da Tabela 13.1, encontre a mudança estimada na média da fertilidade entre 1972 e 1984. (Certifique-se de levar em conta a mudança no intercepto e a mudança na média da escolaridade.)

13.2 Utilizando os dados contidos no arquivo KIELMC.RAW, as seguintes equações foram estimadas usando os anos de 1978 e 1981:

$$\begin{aligned} \widehat{\log(preço)} &= \underset{(0,26)}{11,49} - \underset{(0,058)}{0,547}\, proxincin + \underset{(0,080)}{0,394}\, a81 \cdot proxincin \\ n &= 321,\ R^2 = 0,220 \end{aligned}$$

e

$$\widehat{\log(preço)} = 11{,}18 - 0{,}563\ a81 - 0{,}403\ a81 \cdot proxincin$$
$$(0{,}27) \quad (0{,}044) \qquad (0{,}067)$$
$$n = 321,\ R^2 = 0{,}337.$$

Compare as estimativas do termo de interação $a81 \cdot proxincin$ com as da equação (13.9). Por que as estimativas são diferentes?

13.3 Por que não podemos usar as primeiras diferenças quando temos cortes transversais independentes em dois anos (ao contrário dos dados em painel)?

13.4 Se pensarmos que β_1 é positivo em (13.14) e que Δu_i e $\Delta desemp_i$ são negativamente correlacionados, qual será o viés no estimador MQO de β_1 na equação de primeiras diferenças? (*Sugestão*: Reveja a equação (5.4).

13.5 Suponha que queremos estimar o efeito de diversas variáveis sobre a poupança anual e que temos um conjunto de dados em painel sobre indivíduos coletado em 31 de janeiro de 1990 e 31 de janeiro de 1992. Se incluirmos uma *dummy* anual para o ano de 1992 e usarmos a primeira diferença, poderemos também incluir a idade no modelo original? Explique.

13.6 Em 1985, nem a Flórida nem a Geórgia tinham leis banindo recipientes abertos de bebidas alcoólicas nos compartimentos de veículos de passageiros. Em 1990, a Flórida sancionou tal lei, mas a Geórgia não.

(i) Suponha que você colete amostras aleatórias da população com idade para dirigir de ambos os estados, de 1985 e 1990. Defina *prisão* como uma variável binária igual à unidade se uma pessoa foi presa por dirigir embriagada durante o ano. Sem controlar quaisquer outros fatores, escreva um modelo de probabilidade linear que possibilite verificar se a lei de recipientes abertos reduziu a probabilidade de alguém ser preso por dirigir embriagado. Que coeficiente em seu modelo mede o efeito da lei?

(ii) Por que você pode querer controlar outros fatores nesse modelo? Quais poderiam ser esses fatores?

(iii) Agora, suponha que somente seja possível coletar dados de 1985 e 1990 em nível de municípios dos dois estados. A variável dependente seria a fração dos motoristas habilitados presos por dirigirem embriagados durante o ano. Como essa estrutura de dados difere dos dados em nível individual descritos na parte (i)? Que método econométrico você usaria?

13.7 Leia os itens abaixo e responda às questões.

(i) Utilizando os dados contidos no arquivo INJURY.RAW para Kentucky, a equação estimada quando *apmud* é retirada da (13.12) é

$$\widehat{\log(duração)} = 1{,}129 + 0{,}253 altrend + 0{,}198 \cdot altrend$$
$$(0{,}022) \quad (0{,}042) \qquad (0{,}052)$$
$$n = 5.626,\ R^2 = 0{,}021.$$

É surpreendente que a estimativa na interação seja razoavelmente próxima da obtida na (13.12)? Explique.

(ii) Quando *apmud* é incluída, mas altrend é excluída, o resultado é

$$\widehat{\log(duração)} = 1{,}233 - 0{,}100 apmud + 0{,}447 apmud \cdot altrend$$
$$(0{,}023) \quad (0{,}040) \qquad (0{,}050)$$
$$n = 5.626,\ \mathrm{R}^2 = 0{,}016.$$

Por que o coeficiente no termo de interação agora é tão maior do que na (13.12)? [*Sugestão*: Na equação (13.10), qual é a hipótese que se está fazendo sobre os grupos de tratamento e controle se $\beta_1 = 0$?]

APÊNDICE 13A

Hipóteses do MQO Agrupado Usando Primeiras Diferenças

Neste apêndice, apresentamos observações cuidadosas sobre as hipóteses do estimador de primeiras diferenças. Um pouco da verificação dessas afirmações aparece aqui, mas elas podem ser encontradas de forma mais completa em Wooldridge (2002, Capítulo 10).

Hipótese PD.1

Para cada *i*, o modelo é

$$y_{it} = \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}, \quad t = 1, \ldots, T,$$

em que os β_j são os parâmetros a estimar e a_i é o efeito não observado.

Hipótese PD.2

Temos uma amostra aleatória do corte transversal.

Hipótese PD.3

Cada variável explicativa muda ao longo do tempo (para pelo menos algum *i*), e não existem relações lineares perfeitas entre as variáveis explicativas.

Na próxima hipótese é conveniente que X_i represente as variáveis explicativas de todos os períodos da observação *i* do corte transversal; assim, X_i conterá x_{itj}, t 5 1, ...,T, j 5 1, ..., k.

Hipótese PD.4

Para cada *t*, o valor esperado do erro idiossincrático, dadas as variáveis explicativas em todos os períodos de tempo e os efeitos não observados, é zero: $\mathrm{E}(u_{it}|\mathbf{X}_i, a_i) = 0$.

Quando a hipótese PD.4 se mantém, algumas vezes dizemos que os x_{itj} são *estritamente exógenos condicionais ao efeito não observado*. A ideia é que, assim que tenhamos o controle de a_i, não haverá correlação entre x_{isj} e o erro idiossincrático remanescente, u_{it}, para todo s e t.

Como declarado, a Hipótese PD.4 é mais forte que o necessário. Usamos esta forma da hipótese porque ela enfatiza que estamos interessados na equação

$$E(y_{it} \mid \mathbf{X}_i, a_i) = E(y_{it} \mid \mathbf{x}_{it}, a_i) = \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i,$$

de forma que β_j indique os efeitos parciais das variáveis explicativas observadas mantendo-se fixo, ou "controlando-se," o efeito não observado, a_i. No entanto, uma implicação importante da PD.4, e que é suficiente para a ausência de viés no estimador PD, é $E(\Delta u_{it} \mid \mathbf{X}_i) = 0$, t $= 2, \ldots, T$. De fato, por consistência podemos simplesmente presumir que Δx_{itj} é não correlacionado com Δu_{it} para todos $t = 2, \ldots$T e $j = 1, \ldots, k$. Veja Wooldridge (2002, Capítulo 10) para obter mais informação.

Sob essas quatro primeiras hipóteses, os estimadores de primeiras diferenças são não viesados. A hipótese crucial é PD.4, que é a exogeneidade estrita das variáveis explicativas. Sob essas mesmas hipóteses, também podemos mostrar que o estimador PD é consistente com um T fixo e quando $N \to \infty$ (e talvez de forma mais genérica).

As próximas duas hipóteses asseguram que os erros-padrão e testes estatísticos resultantes do MQO agrupado em primeiras diferenças são (assimptoticamente) válidos.

Hipótese PD.5

A variância dos erros diferenciados, condicional a todas as variáveis explicativas, é constante: $\text{Var}(\Delta u_{it} \mid \mathbf{X}_i) = \sigma^2$, $t = 2, \ldots, T$.

Hipótese PD.6

Para todo $t \neq s$, as *diferenças* nos erros idiossincráticos são não correlacionadas (condicionais a todas as variáveis explicativas): $\text{Cov}(\Delta u_{it}, \Delta u_{is} \mid \boldsymbol{X}_i) = 0$, $t \neq s$.

A Hipótese PD.5 garante que os erros diferenciados, Δu_{it}, são homoscedásticos. A Hipótese PD.6 estabelece que os erros diferenciados são serialmente não correlacionados, o que significa que u_{it} segue um passeio aleatório ao longo do tempo (veja Capítulo 11). Sob as Hipóteses PD.1 a PD.6, o estimador PD de β_j é o melhor estimador linear não viesado (condicional às variáveis explicativas).

Hipótese PD.7

Condicional a $\mathbf{X}_i$, os Δu_{it} são variáveis aleatórias normais independentes e identicamente distribuídas.

Quando adicionamos a Hipótese PD.7, os estimadores PD são normalmente distribuídos, e as estatísticas t e F do MQO agrupado das diferenças têm distribuições t e F exatas. Sem a Hipótese PD.7, podemos recorrer às aproximações assimptóticas habituais.