

INTRODUÇÃO À ECONOMETRIA

UMA ABORDAGEM MODERNA

JEFFREY M. WOOLDRIDGE

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

Wooldridge, Jeffrey M., 1960-
Introdução à econometria : uma abordagem moderna /
Jeffrey M. Wooldridge ; tradução Rogério Cézar de Souza,
José Antônio Ferreira ; revisão técnica Nelson
Carvalheiro. -- São Paulo : Pioneira Thomson Learning,
2006.

Titulo original: *Introductory econometrics :
a modern approach*
Bibliografia
ISBN 85-221-0421-2

I. Econometria I. Carvalheiro, Nelson.
II. Título.

05-8002

CDD-330.015195

índices para catálogo sistemático:

1. Econometria 330.015195

INTRODUÇÃO À ECONOMETRIA

Uma Abordagem Moderna

Jeffrey M. Wooldridge

Michigan State University

Tradução

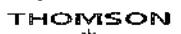
Rogério César de Souza
José Antônio Ferreira

Revisão Técnica

Nelson Carvalheiro

Doutor em Economia pela USP e Professor Titular
do Departamento de Economia da FEA/PUC-SP

THOMSON



Gerente Editorial:
Dulcy Grisolia

**Editora de
Desenvolvimento:**
Tatiana Pavanelli Valsi

**Supervisora de Produção
Editorial:**
Patrícia La Rosa

Produtor Editorial:
Fábio Gonçalves

Produtora Gráfica:
Fabiana Alencar Albuquerque

Título original:
*Introductory Econometrics:
A Modern Approach –
2nd edition*
(ISBN: 0-324-11364-1)

Tradutores:
Rogério César de Souza
(Capítulos 1-5),
José Antônio Ferreira

Revisão Técnica:
Nelson Carvalheiro

Copidesque:
Peterso Roberto Rissatti

Revisão:
Andréa Medeiros
Ana Paula Ribeiro,
Silvana Gouveia

Editoração Eletrônica:
ERJ – Composição Editorial
e Artes Gráficas Ltda.

Capa:
Eduardo Bertolini

COPYRIGHT © 2002 de
Thomson Learning, Inc.
COPYRIGHT © 2006
para a Língua Portuguesa
adquirido por Pioneira
Thomson Learning Ltda., uma
divisão da Thomson Learning,
Inc. Thomson Learning™ é
uma marca registrada aqui
utilizada sob licença.

Impresso no Brasil.
Printed in Brazil.
1 2 3 4 08 07 06

Rua Traipu, 114 – 3^º andar
Perdizes – CEP 01235-000
São Paulo – SP
Tel.: (11) 3665-9900
Fax: (11) 3665-9901
sac@thomsonlearning.com.br
www.thomsonlearning.com.br

Todos os direitos reservados.
Nenhuma parte deste livro
poderá ser reproduzida, sejam
quais forem os meios empregados,
sem a permissão, por
escrito, da Editora. Aos infratores
aplicam-se as sanções previstas
nos artigos 102, 104, 106 e 107
da Lei
nº 9.610, de 19 de fevereiro de
1998.

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

Wooldridge, Jeffrey M., 1960-
Introdução à econometria : uma
abordagem moderna / Jeffrey M.
Wooldridge ; tradução Rogério
Cézar de Souza, José Antônio
Ferreira ; revisão técnica Nelson
Carvalheiro. — São Paulo :
Pioneira Thomson Learning, 2006.
Título original: *Introductory
econometrics : a modern approach*
Bibliografia
ISBN 85-221-0421-2
I. Econometria I. Carvalheiro,
Nelson.
II. Título.
05-8002
CDD-330.015195
Índices para catálogo sistemático:
1. Econometria 330.015195

Prefácio

Minha motivação para escrever *Introdução à Econometria — Uma Abordagem Moderna* vem de uma lacuna, razoavelmente ampla, que identifiquei entre como a econometria é ensinada nos cursos de graduação e o que os pesquisadores empíricos pensam sobre os métodos econométricos e suas aplicações. Com igual importância, convenci-me de que ensinar econometria introdutória da perspectiva dos usuários profissionais da econometria simplificaria, de fato, a apresentação, além de tornar o assunto mais interessante.

Baseado em numerosas reações positivas à primeira edição norte-americana, parece que minha filosofia sobre como ensinar a econometria introdutória é partilhada por muitas pessoas. É gratificante que professores de formações e interesses variados — incluindo a microeconomia aplicada, a macroeconomia aplicada, a análise de política econômica, a ciência política e a econometria teórica —, ensinando estudantes com níveis de preparação muito diferentes, tenham abraçado a abordagem moderna da econometria adotada neste livro. Conseqüentemente, a estrutura desta primeira edição brasileira é muito parecida com a da primeira edição norte-americana. A ênfase ainda está em aplicar a econometria aos problemas do mundo real. Todo método econométrico é motivado por uma questão particular com a qual o pesquisador se defronta ao analisar dados não-experimentais. O foco principal da obra está em entender e interpretar as hipóteses à luz das aplicações empíricas reais: a matemática requerida não vai além da álgebra dos cursos de graduação e da probabilidade e estatística básicas.

ORGANIZADO PARA O ECONOMETRISTA DE HOJE

A característica mais perceptível desta edição, que distingue este texto de outros, é a separação dos tópicos por tipo de dados analisados. Essa é uma diferença clara em relação à abordagem tradicional, que apresenta um modelo linear, lista todas as hipóteses que podem ser necessárias em algum ponto futuro da análise e, então, prova ou afirma resultados sem conectar-los claramente às hipóteses. Minha abordagem é, em primeiro lugar, tratar, na Parte Um, da análise de regressão múltipla com dados de corte transversal, sob a hipótese de amostragem aleatória. Essa estrutura é completamente natural para os estudantes, pois eles estão familiarizados com ela desde os cursos de estatística introdutória. O mais importante é que ela nos permite distinguir hipóteses feitas sobre o modelo de regressão da população subjacente — hipóteses que podem ter um conteúdo econômico determinado ou um conteúdo comportamental geral — de hipóteses sobre como os dados foram extraídos para compor uma amostra.

As discussões sobre as consequências da amostragem não-aleatória podem ser tratadas de um modo intuitivo, após os estudantes terem um bom domínio do modelo de regressão múltipla aplicado a amostras aleatórias.

Uma característica importante de uma abordagem moderna é que as variáveis explicativas — com a variável dependente — são tratadas como resultados de variáveis aleatórias. Para as ciências sociais, admitir variáveis explicativas aleatórias é muito mais realista do que a hipótese tradicional de variáveis explicativas não-aleatórias. Um benefício importante é que a abordagem modelo populacional/amostragem aleatória que utilize reduz bastante o número de hipóteses que os estudantes devem absorver e entender. Ironicamente, a abordagem clássica da análise de regressão, que trata as variáveis explicativas como valores fixos em amostras repetidas e está difundida nos livros introdutórios, aplica-se literalmente a dados coletados em uma estrutura experimental. Além disso, as contorções exigidas para formular e explicar as hipóteses podem ser confusas para os estudantes.

Meu foco sobre o modelo populacional enfatiza que as hipóteses fundamentais subjacentes à análise de regressão, tal como a hipótese de média zero dos fatores não-observados, estão apropriadamente formuladas, condicionadas às variáveis explicativas. Isso leva a um entendimento claro dos tipos de problemas, tal como a heteroscedasticidade (variância não-constante), que podem invalidar os procedimentos padrões da inferência. Adicionalmente, pude eliminar várias interpretações erradas que surgem nos textos de econometria em todos os níveis. Apenas para citar alguns exemplos, explico a razão de o *R*-quadrado usual ainda ser válido como uma medida do grau de ajuste na presença de heteroscedasticidade (e, mais adiante, na presença dos erros serialmente correlacionados nas equações de séries de tempo); discuto, em um nível bastante intuitivo, porque os testes para a forma funcional não devem ser vistos como testes gerais de variáveis omitidas não observadas; e posso facilmente explicar porque sempre se deve incluir, em um modelo de regressão, variáveis extras de controle que são não-correlacionadas com a variável explicativa de interesse (tal como uma variável de decisão).

Como as hipóteses da análise de corte transversal são relativamente diretas e realistas, os estudantes ficam envolvidos mais cedo com aplicações sérias de corte transversal, sem ter de se preocupar com as questões espinhosas de tendência, sazonalidade, correlação serial, alta persistência e regressão espúria que aparecem em abundância nos modelos de regressão de séries temporais. Inicialmente, imaginei que meu tratamento da regressão com dados de corte transversal, seguida pela regressão com dados de séries de tempo, cairia nas boas graças dos professores cujos interesses de pesquisa estão na microeconomia aplicada, e parece que esse é o caso. Tem sido gratificante que aqueles que adotaram este livro e que têm uma inclinação para as séries de tempo ficaram igualmente entusiasmados com a estrutura da obra. Ao postergar a análise econométrica de séries temporais, pude colocar o foco apropriado sobre as armadilhas potenciais da análise de dados de séries de tempo que não surgem com dados de corte transversal. Com efeito, a econometria de séries temporais obteve, por fim, o tratamento sério que ela merece em um livro introdutório.

Para esta edição, escolhi conscientemente os tópicos que são importantes para ler artigos de revistas e para realizar pesquisas empíricas básicas. Dentro de cada tópico, omiti deliberadamente muitos testes e procedimentos de estimação que, embora tradicionalmente incluídos nos livros-texto, não têm resistido ao teste empírico do tempo. Da mesma forma, enfatizei os tópicos mais recentes que têm se mostrado claramente úteis, tal como obter estatísticas de teste que são robustas em relação à heteroscedasticidade (ou à correlação serial) cuja forma é desconhecida, usar dados de vários anos para a análise de decisão, ou resolver o problema de variáveis omitidas pelo método de variáveis instrumentais. Parece que fiz as escolhas corretas, pois me lembro de bem poucas sugestões para acrescentar ou excluir material, especialmente nos capítulos da regressão básica das Partes Um e Dois. Um tópico que, de fato, expandi foi a estimação dos mínimos desvios absolutos (MDA) do Capítulo 9. O MDA está se tornando cada vez mais popular nos trabalhos empíricos, especialmente quando a distribuição condicional da variável dependente é assimétrica ou tem caudas largas. Ao ler pesquisas empíricas da área da economia do trabalho, da economia do setor público e outros campos, é provável que os estudantes encontrem cada vez mais modelos estimados por MDA.

Ao reescrever alguns trechos desta edição, tentei aperfeiçoar mais a abordagem sistemática. O termo sistemática significa que cada tópico está fundamentado, de um modo lógico, no material anterior, e as hipóteses são introduzidas somente se forem necessárias para obter uma conclusão. Por exemplo, os pesquisadores aplicados, bem como os teóricos, sabem que nem todas as hipóteses de Gauss-Markov são necessárias para mostrar que os estimadores de mínimos quadrados ordinários (MQO) são não-viesados. Contudo, quase todos os livros de econometria introduzem o conjunto completo de hipóteses (muitas das quais são redundantes ou, em alguns casos, logicamente conflitantes) antes de provar a inexistência de viés do MQO. De forma semelhante, a hipótese de normalidade é incluída entre as hipóteses que são necessárias para o Teorema de Gauss-Markov, quando é razoavelmente bem conhecido que a normalidade não desempenha nenhum papel para mostrar que os estimadores de MQO são os melhores estimadores lineares não-viesados.

Minha abordagem sistemática estende-se para o estudo das propriedades de amostras grandes, em que as hipóteses de consistência são introduzidas somente quando necessárias. Isso torna relativamente fácil cobrir tópicos mais avançados, como usar cortes transversais agrupados, explorar as estruturas de dados de painel e aplicar os métodos de variáveis instrumentais. Trabalhei para fornecer uma visão unificada da econometria, de acordo com a qual todos os estimadores e as estatísticas de testes são obtidas usando um pouco de princípios, intuitivamente racionais, de estimação e testes (os quais, evidentemente, também têm justificativas rigorosas). Por exemplo, os testes básicos de regressão para a heteroscedasticidade e a correlação serial são fáceis de ser compreendidos pelos estudantes porque eles já têm um conhecimento sólido de regressão. Isso contrasta com os tratamentos que fornecem um conjunto de receitas desconexas para procedimentos econométricos ultrapassados.

Como, ao longo deste livro, enfatizo as relações *ceteris paribus*, vou diretamente para a análise de regressão múltipla após abordar somente um capítulo do modelo de regressão simples. Isso motiva os estudantes a pensar sobre aplicações sérias mais cedo. Também atribuo muito mais destaque à análise de decisão utilizando todos os tipos de estruturas. Os tópicos práticos, como usar variáveis *proxy* para obter efeitos *ceteris paribus* e obter erros-padrão dos efeitos parciais nos modelos com termos de interação, são discutidos de modo simples.

NOVO NESTA EDIÇÃO

No Capítulo 3, há uma discussão completa do viés de variáveis omitidas no modelo de regressão múltipla; acaba não sendo especialmente difícil caracterizar o viés no caso geral. O apêndice do Capítulo 3 contém uma derivação do viés que requer somente álgebra e estatística básicas.

O Capítulo 6 contém uma discussão mais detalhada de modelos com termos de interação. Percebi que os estudantes podem ter dificuldades ao interpretar os parâmetros de tais modelos, de modo que tentei dar melhores explicações. Os Capítulos 7 e 13 apresentam mais detalhes sobre como computar os testes de Chow para os modelos de regressão entre grupos de diferentes unidades de corte transversal e entre diferentes períodos de tempo. Nos Capítulos 8 e 12, forneço um argumento explícito e simples sobre porque o *R*-quadrado ainda é válido como uma medida do grau de ajuste quando o modelo sofre de heteroscedasticidade ou de correlação serial. O Capítulo 9 inclui uma discussão simples de amostragem estratificada básica, um tópico que nasceu de um curso de segundo semestre no qual ministrei aulas.

O Capítulo 17 contém material complexo, e expandi a discussão sobre como interpretar modelos não-lineares de variável dependente limitada, inclusive como eles podem ser comparados com as estimativas de modelos lineares padrão. Novos gráficos foram usados para ilustrar as comparações.

O livro traz mais de 80 bancos de dados disponíveis para *download* pela senha 4212 na página deste livro no site da Thomson (www.thomsonlearning.com.br), dos quais muitos permitem comparar

as estimativas dos retornos de freqüentar cursos profissionalizantes e cursos de graduação; testar se a poupança dos planos de pensão norte-americanos substituem outras formas de poupança ou representa poupança nova; testar se os restaurantes *fast-food* praticam discriminação de preços contra minorias; testar se o casamento afeta a produtividade ou a remuneração dos jogadores profissionais de basquete-bol; estudar o efeito de mais escolhas sobre os investimentos em planos de pensão; estimar os efeitos dos gastos públicos em escolas sobre o desempenho dos estudantes; testar se as leis de uso do cinto de segurança e de limite de velocidade afetam as taxas de acidentes e de mortes; e estimar funções de demanda para maçãs e peixe ecologicamente produzidos. Além disso, os exercícios para computador — disponíveis no site da Thomson — foram expandidos para explorar os novos bancos de dados. Alguns dos conjuntos de dados não são utilizados no livro. Em vez disso, eles podem ser usados em problemas, em exames ou para servir de base para um trabalho de final de curso.

PROJETADO PARA ESTUDANTES DE CURSOS DE GRADUAÇÃO E ADAPTADO PARA ESTUDANTES DE CURSOS DE PÓS-GRADUAÇÃO

O livro é direcionado para estudantes de cursos de graduação em economia que estudaram álgebra e um semestre de probabilidade e estatística introdutórias. (Os Apêndices A, B, C, D, E e F — disponíveis no site da Thomson — contêm o material de pré-requisito.) Não se espera que um curso de econometria de um semestre ou de um trimestre abranja tudo, ou mesmo alguma parte, dos tópicos mais avançados da Parte Três. Um curso de introdução típico incluiria os Capítulos 1 a 8, os quais compreendem as bases das regressões simples e múltipla para dados de corte transversal. Dado que a ênfase está na intuição e na interpretação dos exemplos empíricos, o material dos oito primeiros capítulos deveria ser colocado à disposição dos estudantes dos cursos de graduação na maioria dos departamentos de economia. A maioria dos professores também vai querer cobrir pelo menos partes dos capítulos sobre análise de regressão com dados de séries de tempo (Capítulos 10, 11 e 12), com graus variados de profundidade. No curso de um semestre em que leciono, no Estado de Michigan, trabalho o Capítulo 10 cuidadosamente, dou uma visão geral do material do Capítulo 11 e abordo o material sobre correlação serial do Capítulo 12. Acredito que esse curso básico de um semestre fornece ao aluno fundamentos sólidos para escrever trabalhos empíricos, como uma monografia de final de curso ou um texto para um seminário. O Capítulo 9 contém tópicos mais especializados que surgem ao analisar dados de corte transversal, incluindo problemas de dados tais como *outliers* e amostragem não aleatória. Para um curso de um semestre, esse capítulo pode ser abandonado sem perda de continuidade.

A estrutura do livro é ideal para um curso com foco em corte transversal/análise de decisão: os capítulos de séries de tempo podem ser abandonados, dando lugar a tópicos dos Capítulos 9, 13, 14 ou 15. O Capítulo 13 é “avançado” somente no sentido de que ele trata de duas novas estruturas de dados: cortes transversais independentemente agrupados e análise de dados de painel para dois períodos. Tais estruturas de dados são especialmente úteis para análise de decisão, e esse capítulo fornece vários exemplos a esse respeito. Os estudantes com um bom domínio dos Capítulos 1 a 8 terão pouca dificuldade com o Capítulo 13. O Capítulo 14 trata dos métodos de dados de painel mais avançados e provavelmente será coberto somente em um segundo curso. Uma boa maneira de finalizar um curso sobre métodos de corte transversal é compreender os rudimentos da estimação de variáveis instrumentais do Capítulo 15.

Tenho utilizado material selecionado da Parte Três, incluindo os Capítulos 13, 15 e 17, nos seminários do curso de graduação direcionados para produzir um trabalho de pesquisa sério. Além do curso básico de um semestre, os estudantes que foram expostos à análise básica de dados de painel, de estimação de variáveis instrumentais e de modelos de variável dependente limitada estão preparados para

ler boa parte da literatura aplicada das ciências sociais. O Capítulo 17 apresenta uma introdução aos modelos mais comuns de variável dependente limitada.

O livro também é adequado para um curso introdutório de pós-graduação, no qual a ênfase está mais nas aplicações do que nas derivações que usam álgebra matricial. Além disso, para os professores que querem apresentar a matéria na forma matricial, os apêndices D e E abordam, de modo auto-suficiente, a álgebra matricial e o modelo de regressão múltipla na forma matricial. O Apêndice E abrange a análise assimptótica em profundidade maior para estudantes avançados.

No Estado de Michigan, os estudantes dos cursos de doutorado das muitas áreas que requerem análise de dados — incluindo contabilidade, economia agrícola, economia do desenvolvimento, finanças, economia internacional, economia do trabalho, macroeconomia, ciência política e finanças públicas — descobriram que o livro é uma ponte útil entre o trabalho empírico que eles leem e a economia mais teórica que eles aprendem no nível de doutoramento.

CARACTERÍSTICAS BÁSICAS

Os professores e estudantes parecem apreciar as questões formuladas no texto, cujas respostas encontram-se no Apêndice F. Essas questões têm a intenção de dar ao estudante um retorno imediato sobre seu desempenho. Cada capítulo contém muitos exemplos. Vários deles são estudos de caso retirados de artigos publicados recentemente, levemente modificados para simplificar a análise, sem sacrificar seus principais pontos.

Os problemas de final de capítulo e exercícios para computador — disponíveis no site da Thomson — são totalmente orientados para o trabalho empírico, em vez das derivações complicadas. Os estudantes são solicitados a fundamentar cuidadosamente suas respostas, com base no que aprenderam. Os exercícios para computador expandem, em geral, os exemplos do texto. Vários exercícios usam bancos de dados de trabalhos publicados ou conjuntos de dados similares que são motivados por pesquisas publicadas em economia e em outros campos.

Uma característica única deste livro é o extensivo glossário. As definições e descrições curtas serão um lembrete de grande auxílio para os alunos que estejam estudando para exames ou lendo pesquisas empíricas que usam métodos econometríticos.

CONJUNTOS DE DADOS

Mais de 80 bancos de dados estão disponíveis em Excel e em “.DES”, arquivos de texto que descrevem as variáveis e que podem ser abertos no programa Excel. A maioria dos conjuntos de dados é proveniente de pesquisas reais, de modo que alguns são bastante grandes. Exceto quando for o caso ilustrar as várias estruturas de dados, os bancos de dados não são descritos no texto. Este livro está direcionado para um curso em que o trabalho com o computador desempenha papel importante. Os conjuntos de dados podem ser encontrados na página deste livro no site da Thomson.

Um banco de dados em Access com os nomes e as informações dos dados estatísticos também está disponível no site e permite ao leitor uma busca por nome do arquivo. Os nomes de diversas variáveis foram traduzidos e adaptados ao longo do texto e estão grafados com acentos (como *salário*, por exemplo) para facilitar a memorização do leitor brasileiro. Entretanto, como os dados estatísticos em Excel estão em inglês, a busca pelo banco de dados permite ao leitor encontrar o nome da variável em inglês e sua busca no respectivo arquivo em Excel.

SUGESTÕES PARA MONTAR SEU CURSO

Já comentei sobre o conteúdo da maioria dos capítulos e possíveis estruturas de cursos. Aqui, farei alguns comentários mais específicos sobre o material dos capítulos que podem ser abordados ou postergados.

O Capítulo 9 tem exemplos interessantes (tal como uma regressão que inclui a pontuação do QI como uma variável explicativa). Os nomes das variáveis *proxy* não devem ser formalmente apresentados para descrever esses tipos de exemplos, e costumo apresentá-los quando termino a análise de corte transversal. No Capítulo 12, para um curso de um semestre, não apresento o material sobre inferência robusta na presença de correlação serial quando estou tratando da análise de mínimos quadrados ordinários, bem como de modelos dinâmicos de heteroscedasticidade.

Mesmo em um segundo curso, prefiro despender pouco tempo no Capítulo 16, que trata da análise de equações simultâneas. Se há uma questão sobre a qual as pessoas divergem é a importância das equações simultâneas. Alguns consideram que esse material é fundamental; outros pensam que é raramente aplicável. Minha visão é que os modelos de equações simultâneas são demasiadamente utilizados (veja o Capítulo 16 para uma discussão). Se lermos os trabalhos aplicados cuidadosamente, variáveis omitidas e erros de medida são provavelmente uma das maiores razões para adotar a estimativa de variáveis instrumentais, e é por isso que uso variáveis omitidas para motivar a estimativa de variáveis instrumentais no Capítulo 15. Além disso, os modelos de equações simultâneas são indispensáveis para estimar funções de demanda e oferta, e eles também são aplicáveis em alguns outros casos importantes.

O Capítulo 17 é o único que considera modelos inherentemente não-lineares em seus parâmetros, e isso impõe uma carga adicional para o estudante. O primeiro material que deveria ser tratado nesse capítulo são os modelos de resposta binária probit e logit. Minha apresentação dos modelos Tobit e de regressão censurada ainda parecem originais: reconheço explicitamente que o modelo Tobit é aplicável a resultados de solução de canto em amostras aleatórias, ao passo que a regressão censurada é aplicável quando o processo de coleta de dados censura a variável dependente.

O Capítulo 18 trata de alguns tópicos mais recentes da econometria de séries de tempo, inclusive o teste de raízes unitárias e a cointegração. Abordo esse material somente no segundo semestre de um curso, seja no nível de graduação ou de pós-graduação. Uma introdução razoavelmente detalhada para a previsão também está incluída no Capítulo 18.

O Capítulo 19, que deveria ser acrescentado ao programa de cursos que exigem um trabalho de conclusão, é muito mais extensivo que capítulos semelhantes de outros livros. Ele resume alguns métodos apropriados para vários tipos de problemas e estruturas de dados, aponta dificuldades potenciais, explica com algum detalhe como escrever um trabalho de conclusão de curso em economia empírica e inclui sugestões de possíveis projetos.

AGRADECIMENTOS

Gostaria de agradecer aqueles que revisaram a primeira ou a segunda edição norte-americana, fizeram extensos comentários ou influenciaram de alguma maneira o livro. São eles:

Richard Agnello
University of Delaware

Esfandiar Maasoumi
Southern Methodist University

Scott Baier
Clemson University

Kristin McCue
Texas A &M University

Eli Berman
Boston University

Philip Meguire
University of Canterbury

James Cardon
Brigham Young University

John Mullaly
University of Wisconsin

Rogério César de Souza
Universidade Paulista

William Neils
Texas A &M University

Amitabh Chandra
Dartmouth University

David Neumark
Michigan State University

Christopher Cornwell
University of Georgia

Leslie Papke
Michigan State University

Edward Coulson
Pennsylvania State University

Soo-Bin Park
Carleton University

William Even
Miami University of Ohio

Jeffrey Pliskin
Hamilton College

Adrian Fleissig
St. Louis University

Joseph Quinn
Boston College

Arthur Goldberger
University of Wisconsin

Nagesh Revankar
SUNY, Buffalo

Daniel Hamermesh
University of Texas

Louise Russell
Rutgers University

Bruce Hansen
University of Wisconsin

Shinichi Sakata
University of Michigan

KyungSo In
Wichita State University

Mark Showalter
Brigham Young University

Datelina Ivañovo
SUNY, Albany

Jeffrey Smith
University of Maryland

Heejoon Kang
Indiana University

John Spitzer
SUNY, Brockport

Manfred Keil
Claremont McKenna College

Leanna Stiefel
New York University

Neha Khanna
SUNY, Binghamton

Wendy Stock
Montana State University

Christopher Tuber
Northwestern University

George Tavlas
Bank of Greece

Larry Taylor
Lehigh University

Pravin Trivedi
Indiana University

Robert Trost
George Washington University

Hiroki Tsurumi
Rutgers University

Robert Turner
Colgate University

Timothy Vogelsang
Cornell University

Melvyn Weeks
University of Cambridge

Diana Whistler
University of British Columbia

Paul Wilson
University of Texas

Keith Womar
University of Mississippi

Jeffrey Zabel
Tufts University

Considerei muitos dos seus comentários sobre a primeira edição norte-americana, mesmo que não tenha alterado o material de acordo com alguma preferência específica. Em alguns casos, as opiniões dos comentaristas e dos usuários conflitaram, e decidi deixar a organização do livro como está. Em outros casos, continuarei a cogitar sugestões específicas feitas por um ou mais comentaristas. Naturalmente, estou aberto a sugestões sobre possíveis melhorias futuras.

Muitos estudantes e assistentes de professores encontraram erros na primeira edição norte-americana ou sugeriram a reformulação de alguns parágrafos. O número de tais colaboradores é tão grande que não há espaço aqui para listá-los. Gostaria de agradecer os esforços de Chirok Han, que leu cuidadosamente as provas da primeira edição, e a Ali Becker, que verificou esta edição.

Uma vez mais, adorei trabalhar com o pessoal da South-Western/Thomson Learning. Coletivamente, eles conduziram a realização desta edição de modo gentil, mas com mão firme. Meu editor de desenvolvimento, Andy McGuire, ajudou-me a refinar muitos comentários de colaboradores e leitores sobre a primeira edição, e foi de grande auxílio ao construir uma estratégia de revisão do livro. Starratt Alexander encarregou-se da tarefa onerosa de editor de produção e fez um trabalho excepcional. Peggy Buskey e Vicky True cuidaram, com grande habilidade, do site do livro. Uma vez mais, Malvine Litten e seu grupo do LEAP, e em particular Rachel Morris, fizeram um trabalho magnífico de digitação e edição do manuscrito.

Esta edição ainda é dedicada à minha mulher, Leslie, e aos nossos filhos, Edmund e Gwenyth. Leslie fez comentários valiosos sobre o livro e identificou alguns erros de impressão. Além disso, ela continua me encorajando muito, apesar de ter ficado bastante cética quando eu disse, após a publicação da primeira edição norte-americana, o quanto estava feliz por ter finalizado o “projeto”.

Sobre o Autor

Jeffrey M. Wooldridge é professor emérito de Economia na Universidade Estadual de Michigan, onde leciona desde 1991. De 1986 a 1991, foi professor-assistente de Economia no Massachusetts Institute of Technology. Obteve seu bacharelado, com especialização em Ciência da Computação e Economia, na Universidade da Califórnia, Berkeley, em 1982, e seu doutorado em Economia na Universidade da Califórnia, San Diego, em 1986. Publicou mais de duas dezenas de artigos em revistas internacionalmente reconhecidas e muitos capítulos de livros. É autor de *Econometric Analysis of Cross Section and Panel Data*. Seus prêmios incluem: um Alfred P. Sloan Research Fellowship, o Plura Scripsit da *Econometric Theory*; o Sir Richard Stone do *Journal of Applied Econometrics*; e três prêmios de professor do ano da pós-graduação do MIT. Além de ser membro do *Journal of Econometrics*, é também editor do *Journal of Business and Economic Statistics*, co-editor de econometria do *Economics Letters* e participa do corpo editorial do *Journal of Econometrics* e da *Review of Economics and Statistics*. Ocionalmente, também atua como consultor de econometria para a Arthur Andersen, de Chicago, e para a Charles River Associates, de Boston.

Sumário

Capítulo 1 A Natureza da Econometria e dos Dados Econômicos	1
1.1 O que é Econometria?	1
1.2 Passos na Análise Econômica Empírica	2
1.3 A Estrutura dos Dados Econômicos	5
<i>Dados de Corte Transversal</i>	5
<i>Dados de Séries de Tempo</i>	8
<i>Cortes Transversais Agrupados</i>	9
<i>Dados de Painel ou Longitudinais</i>	10
<i>Um Comentário sobre Estruturas de Dados</i>	12
1.4 A Causalidade e a Noção de <i>Ceteris Paribus</i> na Análise Econométrica	12
Resumo	17
 PARTE 1	
ANÁLISE DE REGRESSÃO COM DADOS DE CORTE TRANSVERSAL	19
 Capítulo 2 O Modelo de Regressão Simples	20
2.1 Definição do Modelo de Regressão Simples	20
2.2 Derivação das Estimativas de Mínimos Quadrados Ordinários	25
<i>Uma Nota sobre Terminologia</i>	34
2.3 Mecânica do Método MQO	34
<i>Valores Estimados e Resíduos</i>	34
<i>Propriedades Algébricas das Estatísticas de MQO</i>	36
<i>Grau de ajuste</i>	38
2.4 Unidades de Medida e Forma Funcional	39
<i>Os Efeitos de Mudanças das Unidades de Medida sobre as Estatísticas de MQO</i>	40
<i>Incorporação de Não-Linearidades na Regressão Simples</i>	41
<i>O Significado da Regressão "Linear"</i>	44
2.5 Valores Esperados e Variâncias dos Estimadores de MQO	45
<i>Inexistência de Viés em MQO</i>	45
<i>Variâncias dos Estimadores de MQO</i>	51
<i>Estimação da Variância do Erro</i>	55

2.6	Regressão através da Origem	58
	Resumo	59
	Problemas	59
	Apêndice 2A	62
Capítulo 3	Análise de Regressão Múltipla: Estimação	64
3.1	Funcionabilidade da Regressão Múltipla	64
	<i>Modelo com duas Variáveis Independentes</i>	64
	<i>Modelo com k Variáveis Independentes</i>	67
3.2	Mecânica e Interpretação dos Mínimos Quadrados Ordinários	69
	<i>Obtenção das Estimativas de MQO</i>	69
	<i>Interpretação da Equação de Regressão de MQO</i>	70
	<i>Sobre o Significado de "Manter Outros Fatores Fixos"</i>	
	<i>na Regressão Múltipla</i>	73
	<i>Variação de mais de uma Variável Independente Simultaneamente</i>	73
	<i>Valores Estimados e Resíduos de MQO</i>	74
	<i>Interpretação de "Parcialidade" da Regressão Múltipla</i>	75
	<i>Comparação das Estimativas das Regressões Simples e Múltipla</i>	75
	<i>Grau de Ajuste</i>	77
	<i>Regressão através da Origem</i>	79
3.3	O Valor Esperado dos Estimadores de MQO	80
	<i>Inclusão de Variáveis Irrelevantes em um Modelo de Regressão</i>	85
	<i>Viés de Variável Omitida: O Caso Simples</i>	86
	<i>Viés de Variável Omitida: Casos mais Gerais</i>	90
3.4	A Variância dos Estimadores de MQO	91
	<i>Os Componentes das Variâncias de MQO: Multicolinearidade</i>	92
	<i>Variâncias em Modelos Mal Especificados</i>	96
	<i>Estimação de σ^2: Os Erros-Padrão dos Estimadores de MQO</i>	97
3.5	Eficiência de MQO: O Teorema de Gauss-Markov	99
	Resumo	100
	Problemas	101
	Apêndice 3A	105
Capítulo 4	Análise de Regressão Múltipla: Inferência	110
4.1	Distribuições Amostrais dos Estimadores de MQO	110
4.2	Testes de Hipóteses sobre um único Parâmetro Populacional: O Teste t	113
	<i>Teste contra Hipóteses Alternativas Unilaterais</i>	116
	<i>Teste contra Hipóteses Alternativas Bilaterais</i>	121
	<i>Testes de outras Hipóteses sobre β_j</i>	123
	<i>Cálculos dos p-Valores dos Testes t</i>	126
	<i>Lembrete sobre a Linguagem do Teste de Hipóteses Clássico</i>	129
	<i>Significância Econômica ou Prática versus Significância Estatística</i>	129
4.3	Intervalos de Confiança	131
4.4	Testes de Hipóteses sobre uma Combinação Linear dos Parâmetros	134
4.5	Testes de Restrições Lineares Múltiplas: O Teste F	137
	<i>Teste de Restrições de Exclusão</i>	137
	<i>Relação entre as Estatísticas F e t</i>	143

	<i>A Forma R-quadrado da Estatística F</i>	144
	<i>Cálculo dos p-Valores para Testes F</i>	146
	<i>A Estatística F para a Significância Geral de uma Regressão</i>	147
	<i>Teste de Restrições Lineares Gerais</i>	148
4.6	Descrição dos Resultados da Regressão	149
	Resumo	152
	Problemas	152
Capítulo 5	Análise de Regressão Múltipla: MQO Assimptótico	158
5.1	Consistência	158
	<i>A Derivação da Inconsistência no Método MQO</i>	161
5.2	Normalidade Assimptótica e Inferência de Amostras Grandes	163
	<i>Outros Testes de Amostras Grandes:</i>	
	<i>A Estatística Multiplicador de Lagrange</i>	166
5.3	Eficiência Assimptótica de MQO	169
	Resumo	171
	Problemas	171
	Apêndice 5A	172
Capítulo 6	Análise de Regressão Múltipla: Problemas Adicionais	174
6.1	Efeitos da Dimensão dos Dados nas Estatísticas MQO	174
	<i>Os Coeficientes Beta</i>	177
6.2	Um pouco mais sobre a Forma Funcional	179
	<i>Um pouco mais sobre o Uso de Formas Funcionais Logarítmicas</i>	179
	<i>Modelos com Funções Quadráticas</i>	182
	<i>Modelos com Termos de Interação</i>	187
6.3	Um pouco mais sobre o Grau de Ajuste e a Seleção de Regressores	189
	<i>O R-Quadrado Ajustado</i>	189
	<i>O Uso do R-quadrado Ajustado para a Escolha entre Modelos Não-Aninhados</i>	191
	<i>O Controle de muitos Fatores na Análise de Regressão</i>	193
	<i>A Adição de Regressores para Reduzir a Variância do Erro</i>	194
6.4	Previsão e Análise de Resíduos	195
	<i>Intervalos de Confiança de Previsões</i>	195
	<i>Análise de Resíduos</i>	199
	<i>Previsão de y quando a Variável Dependente é $\log(y)$</i>	201
	Resumo	203
	Problemas	204
Capítulo 7	Análise de Regressão Múltipla com Informações Qualitativas: Variáveis Binárias (ou Dummy)	207
7.1	A Descrição das Informações Qualitativas	207
7.2	Uma Única Variável Dummy Independente	209
	<i>A Interpretação dos Coeficientes de Variáveis Dummy Explicativas quando a Variável Dependente é Expressa como $\log(y)$</i>	214
7.3	O Uso de Variáveis Dummy para Categorias Múltiplas	216
	<i>Incorporação de Informações Ordinais com o Uso de Variáveis Dummy</i>	218

7.4	Interações Envolvendo Variáveis <i>Dummy</i>	221
	<i>Interações entre Variáveis Dummy</i>	221
	<i>Consideração de Inclinações Diferentes</i>	223
	<i>Verificação de Diferenças nas Funções de Regressão entre Grupos</i>	227
7.5	Uma Variável Dependente Binária: O Modelo de Probabilidade Linear	230
7.6	Um Pouco mais sobre Análise e Avaliação de Políticas e Programas Governamentais	236
	Resumo	238
	Problemas	239
Capítulo 8	Heteroscedasticidade	243
8.1	Conseqüências da Heteroscedasticidade para o Método MQO	243
8.2	Inferência Robusta em Relação à Heteroscedasticidade após e Estimação MQO	244
	<i>Computando Testes LM Robustos em Relação à Heteroscedasticidade</i>	249
8.3	O Teste da Existência de Heteroscedasticidade	251
	<i>O Teste de White para a Heteroscedasticidade</i>	254
8.4	Estimação de Mínimos Quadrados Ponderados	256
	<i>A Heteroscedasticidade É Percebida como uma Constante Multiplicativa</i>	256
	<i>A Necessidade de Estimar a Função de Heteroscedasticidade:</i>	
	<i>O MQG Factível</i>	262
8.5	O Modelo de Probabilidade Linear Revisitado	266
	Resumo	269
	Problemas	270
Capítulo 9	Problemas Adicionais de Especificação e de Dados	272
9.1	Má Especificação da Forma Funcional	272
	<i>O Teste RESET como um Teste Geral da Má Especificação da Forma Funcional</i>	275
	<i>Testes Contra Alternativas Não-Aninhadas</i>	277
9.2	Utilizando Variáveis Proxy para Variáveis Explicativas Não-Observadas	278
	<i>O Uso de Variáveis Dependentes Defasadas como Variáveis Proxy</i>	283
9.3	Propriedades do Método MQO quando há Erros de Medida	285
	<i>Erro de Medida na Variável Dependente</i>	285
	<i>Erro de Medida em uma Variável Explicativa</i>	287
9.4	Ausência de Dados, Amostras Não-Aleatórias e Observações Extremas	292
	<i>Ausência de Dados</i>	292
	<i>Amostras Não-Aleatórias</i>	293
	<i>Observações Extremas ou Atípicas</i>	295
	Resumo	300
	Problemas	301
PARTE 2		
ANÁLISE DE REGRESSÃO COM DADOS DE SÉRIES TEMPORAIS		305
Capítulo 10	O Básico da Análise de Regressão com Dados de Séries Temporais	306
10.1	A Natureza dos Dados das Séries Temporais	306

10.2	Exemplos de Modelos de Regressão de Séries Temporais <i>Modelos Estáticos</i>	307 308
	<i>Modelos de Desasagens Distributivas Finitas</i>	308
	<i>Convenção sobre o Índice Temporal</i>	311
10.3	Propriedades de Amostra Finita do MQO sob as Hipóteses Clássicas <i>Inexistência de Viés do MQO</i>	311 311
	<i>As Variâncias dos Estimadores MQO e o Teorema de Gauss-Markov</i>	315
	<i>Inferência sob as Hipóteses do Modelo Linear Clássico</i>	318
10.4	Forma Funcional, Variáveis Dummy e Números-Índices	320
10.5	Tendência e Sazonalidade <i>Caracterização de Séries Temporais com Tendência</i>	327 327
	<i>Uso de Variáveis com Tendência na Análise de Regressão</i>	330
	<i>Interpretação sobre a Retirada da Tendência de Regressões com a Inclusão de uma Tendência Temporal</i>	333
	<i>Cálculo do R-Quadrado quando a Variável Dependente Apresenta Tendência</i>	334
	<i>Sazonalidade</i>	336
	Resumo	338
	Problemas	338
Capítulo 11	Questões Adicionais quanto ao Uso do MQO com Dados de Séries Temporais	340
11.1	Séries Temporais Estacionárias e Fracamente Dependentes <i>Séries Temporais Estacionárias e Não-Estacionárias</i>	340 341
	<i>Séries Temporais Fracamente Dependentes</i>	342
11.2	Propriedades Assimptóticas do MQO	345
11.3	O Uso de Séries Temporais Altamente Persistentes na Análise de Regressão <i>Séries Temporais Altamente Persistentes</i>	352 353
	<i>Transformações de Séries Temporais Altamente Persistentes</i>	357
	<i>A Decisão sobre uma Série de Tempo Ser I(I)</i>	358
11.4	Modelos Dinamicamente Completos e a Ausência de Correlação Serial	360
11.5	A Hipótese de Homoscedasticidade para Modelos de Séries Temporais Resumo	363 364
	Problemas	365
Capítulo 12	Correlação Serial e Heteroscedasticidade em Regressões de Séries Temporais	368
12.1	As Propriedades do MQO com Erros Serialmente Correlacionados <i>Inexistência de Viés e Consistência</i>	368 368
	<i>Eficiência e Inferência</i>	369
	<i>O Grau de Ajuste</i>	370
	<i>A Correlação Serial na Presença da Variável Dependente Defasada</i>	371
12.2	O Teste da Correlação Serial <i>O Teste t de Correlação Serial AR(1) com Regressores</i>	372
	<i>Estritamente Exógenos</i>	373
	<i>O Teste de Durbin-Watson sob as Hipóteses Clássicas</i>	375

<i>O Teste da Correlação Serial AR(1) sem Regressores</i>	
<i>Estritamente Exógenos</i>	376
<i>O Teste da Correlação Serial de Ordem mais Elevada</i>	378
12.3 A Correção da Correlação Serial com Regressores Estritamente Exógenos	380
<i>A Obtenção do Melhor Estimador Linear Não-Viesado no Modelo AR(1)</i>	380
<i>A Estimação MQG Factivel com Erros AR(1)</i>	382
<i>Comparação entre MQO e MQGF</i>	384
<i>A Correção da Correlação Serial para Ordens mais Elevadas</i>	386
12.4 Diferenciação e Correlação Serial	387
12.5 Inferência Robusta em Relação à Correlação Serial após o MQO	388
12.6 Heteroscedasticidade em Regressões de Séries Temporais	392
<i>Estatísticas Robustas em Relação à Heteroscedasticidade</i>	392
<i>O Teste da Heteroscedasticidade</i>	393
<i>A Heteroscedasticidade Condicional Auto-Regressiva</i>	394
<i>Heteroscedasticidade e Correlação Serial em Modelos de Regressão</i>	396
Resumo	397
Problemas	398
PARTE 3	
TÓPICOS AVANÇADOS	401
Capítulo 13 O Agrupamento de Cortes Transversais ao Longo do Tempo.	
Métodos Simples de Dados de Painel	402
13.1 O Agrupamento Independente de Cortes Transversais ao Longo do Tempo	403
<i>O Teste de Chow de Mudança Estrutural ao Longo do Tempo</i>	407
13.2 Análise de Decisões Governamentais com Agrupamentos de Cortes Transversais	408
13.3 Análise de Dados de Painel de dois Períodos	414
<i>A Organização dos Dados de Painel</i>	420
13.4 Análise de Decisões Governamentais com Dados de Painel de dois Períodos	421
13.5 A Diferenciação com mais de dois Períodos de Tempo	424
Resumo	429
Problemas	430
Apêndice 13A	431
Capítulo 14 Métodos Avançados de Dados de Painel	433
14.1 Estimação de Efeitos Fixos	433
<i>A Regressão das Variáveis Dummy</i>	437
<i>Efeitos Fixos ou Primeira Diferenciação?</i>	439
<i>Efeitos Fixos com Painéis Não Equilibrados</i>	440
14.2 Modelos de Efeitos Aleatórios	441
<i>Efeitos Aleatórios ou Efeitos Fixos?</i>	445
14.3 A Aplicação de Métodos de Dados de Painel a outras Estruturas de Dados	445
Resumo	447
Problemas	448
Apêndice 14A	449

Capítulo 15 Estimação de Variáveis Instrumentais e Mínimos Quadrados de dois Estágios	453
15.1 Motivação: Variáveis Omitidas em um Modelo de Regressão Simples <i>Inferência Estatística com o Estimador de VI</i>	454
<i>Propriedades da VI com uma Variável Instrumental Pobre</i>	457
<i>O Cálculo do R-Quadrado após a Estimação de VI</i>	462
15.2 Estimação de VI do Modelo de Regressão Múltipla	464
15.3 Mínimos Quadrados de dois Estágios <i>Uma Única Variável Explicativa Endógena</i>	468
<i>Multicolinearidade e MQ2E</i>	471
<i>Variáveis Explicativas Endógenas Múltiplas</i>	472
<i>O Teste de Hipóteses Múltiplas após a Estimação por MQ2E</i>	473
15.4 Soluções de VI de Problemas de Erros nas Variáveis	473
15.5 O Teste de Endogeneidade e o Teste de Restrições <i>Sobreidentificadoras</i>	475
<i>O Teste de Endogeneidade</i>	475
<i>O Teste de Restrições Sobreidentificadoras</i>	477
15.6 O MQ2E com Heteroscedasticidade	478
15.7 A Aplicação do MQ2E a Equações de Séries Temporais	479
15.8 A Aplicação do MQ2E em Cortes Transversais Agrupados e em Dados de Painel	481
Resumo	484
Problemas	484
Apêndice 15A	488
Capítulo 16 Modelos de Equações Simultâneas	491
16.1 A Natureza dos Modelos de Equações Simultâneas	491
16.2 Viés de Simultaneidade no MQO	496
16.3 A Identificação e a Estimação de uma Equação Estrutural <i>A Identificação em um Sistema de Duas Equações</i>	498
<i>Estimação por MQ2E</i>	498
16.4 Sistemas com mais de duas Equações <i>Identificação em Sistemas com três ou mais Equações</i>	503
<i>Estimação</i>	505
16.5 Modelos de Equações Simultâneas com Séries Temporais	506
16.6 Modelos de Equações Simultâneas com Dados de Painel Resumo	510
Problemas	513
	514
Capítulo 17 Modelos com Variáveis Dependentes Limitadas e Correções da Seleção Amostral	517
17.1 Modelos Logit e Probit de Resposta Binária <i>A Especificação de Modelos Logit e Probit</i>	518
<i>Estimação de Máxima Verossimilhança de Modelos Logit e Probit</i>	518
<i>Testes de Hipóteses Múltiplas</i>	521
<i>A Interpretação das Estimativas Logit e Probit</i>	522
17.2 O Modelo Tobit para Resposta de Solução de Canto	523
	529

<i>A Interpretação das Estimativas Tobit</i>	531
<i>Problemas de Especificação nos Modelos Tobit</i>	536
17.3 O Modelo de Regressão de Poisson	537
17.4 Modelos de Regressão Censurada e Truncada	542
<i>Modelos de Regressão Censurada</i>	543
<i>Modelos de Regressão Truncada</i>	547
17.5 Correções da Seleção Amostral	549
<i>Quando o MQO é Consistente na Amostra Selecionada?</i>	549
<i>Truncamento Ocasional</i>	551
Resumo	555
Problemas	556
Apêndice 17A	558
Capítulo 18 Tópicos Avançados sobre Séries Temporais	559
18.1 Modelos de Defasagem Distribuída Infinita	560
<i>A Defasagem Distribuída Geométrica (ou de Koyck)</i>	562
<i>Modelos de Defasagem Distribuída Racional</i>	564
18.2 O Teste de Raízes Unitárias	567
18.3 Regressão Espúria	572
18.4 Co-Integração e Modelos de Correção de Erro	574
<i>Co-Integração</i>	574
<i>Modelos de Correção de Erro</i>	579
18.5 Previsão	581
<i>Tipos de Modelos de Regressão Utilizados na Previsão</i>	583
<i>Previsão um Passo à Frente</i>	584
<i>A Comparação de Previsões um Passo à Frente</i>	588
<i>Previsão com Múltiplos Passos à Frente</i>	589
<i>A Previsão de Processos com Tendência, Sazonais e Integrados</i>	592
Resumo	597
Problemas	599
Capítulo 19 A Montagem de um Projeto na Prática	602
19.1 A Formulação de uma Pergunta	602
19.2 A Revisão da Literatura	604
19.3 A Compilação dos Dados	605
<i>A Decisão sobre o Conjunto de Dados Apropriado</i>	605
<i>A Entrada e o Armazenamento de Seus Dados</i>	606
<i>Inspeção, Limpeza e Sumário de Seus Dados</i>	608
19.4 A Análise Econométrica	609
19.5 A Redação de um Ensaio Empírico	613
<i>Introdução</i>	613
<i>Estrutura Conceitual (ou Teórica)</i>	613
<i>Modelos Econométricos e Métodos de Estimação</i>	614
<i>Os Dados</i>	616
<i>Resultados</i>	617

<i>Conclusões</i>	618
<i>Sugestões de Estilo</i>	618
Resumo	621
Amostra de Projetos Empíricos	621
Lista de Periódicos	626
Fontes de Dados	627
Apêndice G Tabelas Estatísticas	629
Referências Bibliográficas	637
Glossário	645
Índice Remissivo	667



A Natureza da Econometria e dos Dados Econômicos

Capítulo 1 examina o escopo da econometria e propõe questões gerais que resultam da aplicação dos métodos econométricos. A Seção 1.3 examina os tipos de dados usados em negócios, economia e outras ciências sociais. A Seção 1.4 faz uma discussão intuitiva das dificuldades associadas com a inferência da causalidade nas ciências sociais.

1.1 O QUE É ECONOMETRIA?

Imagine que você seja contratado pelo governo de seu Estado para avaliar a eficácia de um programa de treinamento financiado com recursos públicos. Suponha que esse programa ensine aos trabalhadores várias maneiras de como usar computadores no processo produtivo. O programa, com duração de 20 semanas, oferece cursos fora do horário do expediente. Qualquer trabalhador horista da produção pode participar, e a matrícula em todo o programa, ou em parte dele, é voluntária. Você deve determinar qual o efeito, se houver, do programa de treinamento sobre o salário-hora de cada trabalhador.

Suponha, agora, que você trabalhe para um banco de investimentos. Você deve estudar os retornos de diferentes estratégias de investimento que envolvem títulos do Tesouro dos Estados Unidos para decidir se elas estão de acordo com as teorias econômicas a elas associadas.

A tarefa de responder a tais questões pode parecer desanimadora à primeira vista. Nesse ponto, você deve ter somente uma vaga idéia de qual tipo de dados coletar. Até o fim deste curso de princípios de econometria, você provavelmente saberá como usar os métodos econométricos para avaliar, formalmente, um programa de treinamento ou testar uma simples teoria econômica.

A econometria é baseada no desenvolvimento de métodos estatísticos para estimar relações econômicas, testar teorias, avaliar e implementar políticas de governo e de negócios. A aplicação mais comum da econometria é a previsão de importantes variáveis macroeconômicas, tais como taxas de juros, taxas de inflação e produto interno bruto (PIB). Ainda que as previsões de indicadores econômicos sejam bastante visíveis e, muitas vezes, extensamente publicadas, os métodos econométricos podem ser usados em áreas econômicas que não têm nada a ver com previsões macroeconômicas. Por exemplo, estudaremos os efeitos de gastos em campanhas políticas sobre os resultados de eleições. No campo da educação, consideraremos o efeito de gastos públicos com escolas sobre o desempenho de estudantes. Além disso, aprenderemos como usar métodos econométricos para prever séries de tempo econômicas.

A econometria evoluiu como uma disciplina separada da estatística matemática, porque enfoca problemas inerentes à coleta e à análise de dados econômicos não-experimentais. **Dados não-experimentais** não são acumulados por meio de experimentos controlados de indivíduos, firmas ou seg-

mentos da economia. (Dados não-experimentais são, às vezes, chamados de **dados observacionais** para enfatizar o fato de que o pesquisador é um coletor passivo de dados.) **Dados experimentais** são freqüentemente coletados em ambientes de laboratório nas ciências naturais, mas são muito mais difíceis de serem obtidos nas ciências sociais. Embora seja possível realizar alguns experimentos sociais, geralmente é impossível conduzir os tipos de experimentos controlados necessários para avaliar questões econômicas, seja porque eles são proibitivamente dispendiosos ou moralmente repugnantes. Na Seção 1.4, apresentaremos alguns exemplos específicos das diferenças entre dados experimentais e não-experimentais.

Naturalmente, os econometristas, sempre que possível, valem-se dos estatísticos matemáticos. O método de análise da regressão múltipla é o esteio de ambos os campos, mas seu foco e sua interpretação podem diferir de forma marcante. Além disso, os economistas criaram novas técnicas para lidar com as complexidades dos dados econômicos e para testar as previsões das teorias econômicas.

1.2 PASSOS NA ANÁLISE ECONÔMICA EMPÍRICA

Os métodos econométricos são relevantes em, virtualmente, todos os ramos da economia aplicada. Eles entram em cena quando temos uma teoria econômica para testar ou quando temos em mente uma relação que apresenta alguma importância para decisões de negócios ou análises de políticas públicas. Uma análise empírica usa dados para testar uma teoria ou estimar uma relação.

Como se estrutura uma análise econômica empírica? Pode parecer óbvio, mas é importante enfatizar que o primeiro passo em qualquer análise empírica é a formulação cuidadosa da questão de interesse. Essa questão pode ser a de testar certo aspecto de uma teoria ou os efeitos de uma política governamental. Em princípio, métodos econométricos podem ser usados para responder a uma gama de questões.

Em alguns casos, especialmente aqueles que envolvem o teste de teorias econômicas, constrói-se um **modelo econômico** formal. Um modelo econômico consiste em equações matemáticas que descrevem várias relações. Os economistas são conhecidos por suas construções de modelos os quais descrevem um amplo leque de comportamentos. Por exemplo, em microeconomia intermediária, as decisões de consumo individual, sujeitas a uma restrição orçamentária, são descritas por modelos matemáticos. A premissa básica que fundamenta esses modelos é a *maximização da utilidade*. A hipótese de que os indivíduos fazem escolhas para maximizar seu bem-estar, sujeitas às restrições de recursos, oferece-nos um arcabouço muito poderoso para criar modelos econômicos tratáveis e fazer previsões bem definidas. No contexto das decisões de consumo, a maximização da utilidade leva a um conjunto de *equações de demanda*. Em uma equação de demanda, a quantidade demandada de cada produto depende do seu próprio preço, do preço dos bens substitutos e complementares, da renda do consumidor e das características individuais que influem no gosto. Essas equações podem formar a base de uma análise econométrica da demanda do consumidor.

Os economistas têm usado ferramentas econômicas básicas, tais como o arcabouço da maximização da utilidade, para explicar comportamentos que, à primeira vista, podem parecer de natureza não econômica. Um exemplo clássico é o modelo econômico de Becker (1968) sobre o comportamento criminoso.

EXEMPLO 1.1**(Modelo Econômico do Crime)**

Em um artigo inspirador, o prêmio Nobel Gary Becker postulou um arcabouço da maximização da utilidade para descrever a participação de um indivíduo no crime. Certos crimes têm recompensas econômicas evidentes, mas muitos comportamentos criminosos têm custos. O custo de oportunidade do crime impede o criminoso de participar de outras atividades, como um emprego legal. Além disso, há custos associados com a possibilidade de ser capturado, e, se condenado, há os custos associados com o cumprimento de pena. Da perspectiva de Becker, a decisão de empreender a atividade ilegal é uma decisão de alocação de recursos com os benefícios e custos das atividades concorrentes sendo considerados.

Sob hipóteses gerais podemos derivar uma equação que descreve a quantidade de tempo gasto na atividade criminosa como uma função de vários fatores. Podemos representar tal função como

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7), \quad (1.1)$$

em que

- y = horas gastas em atividades criminosas,
- x_1 = “salário” por hora ocupada em atividade criminosa,
- x_2 = salário-hora em emprego legal,
- x_3 = renda de outras atividades que não o crime ou um emprego legal,
- x_4 = probabilidade de ser capturado,
- x_5 = probabilidade de ser condenado se capturado,
- x_6 = sentença esperada se condenado, e
- x_7 = idade.

Outros fatores geralmente afetam a decisão de uma pessoa de participar de atividades criminosas, mas a lista acima representa o que poderia resultar de uma análise econômica formal. Como é comum na teoria econômica, não fomos específicos sobre a função $f(\cdot)$ em (1.1). Essa função depende de uma função utilidade subjacente, raramente conhecida. Entretanto, podemos usar a teoria econômica – ou a introspecção – para prever o efeito que cada variável teria sobre as atividades criminosas. Essa é a base para uma análise econométrica das atividades criminosas individuais.

A modelagem econômica formal é, às vezes, o ponto de partida da análise empírica, porém é mais comum usar a teoria econômica de modo menos formal, ou mesmo contar inteiramente com a intuição. Você pode concordar quanto aos determinantes do comportamento criminoso que aparecem na equação (1.1) serem razoavelmente baseados no senso comum; poderíamos chegar a tal equação diretamente, sem partir da maximização da utilidade. Essa visão tem algum mérito, embora haja casos em que derivações formais geram idéias que a intuição pode ignorar.

Vejamos o exemplo de uma equação que foi derivada por meio de um raciocínio um tanto informal.

EXEMPLO 1.2**(Treinamento e Produtividade do Trabalhador)**

Considere o problema proposto no início da Seção 1.1. Um economista especializado em trabalho gostaria de examinar os efeitos do treinamento sobre a produtividade do trabalhador. Nesse caso, há pouca necessidade de teoria econômica formal. Um entendimento econômico básico é suficiente para perceber que fatores tais como educação, experiência e treinamento influenciam a produtividade do trabalhador. Os economistas também estão bem cientes de que os trabalhadores são pagos de acordo com sua produtividade. Esse raciocínio simples leva a um modelo tal que

$$\text{salárioh} = f(\text{educ}, \text{exper}, \text{treina}), \quad (1.2)$$

em que salárioh é o salário-hora, educ representa os anos de educação formal, exper refere-se aos anos de experiência no mercado de trabalho e treina corresponde a semanas ocupadas em treinamento. Novamente, outros fatores geralmente influenciam a taxa de salário, mas (1.2) captura a essência do problema.

Após especificarmos um modelo econômico, precisamos voltar ao que chamamos de modelo econométrico. Visto que trabalharemos com modelos econométricos ao longo deste texto, é importante saber como eles se relacionam com os modelos econômicos. Considere a equação (1.1) como exemplo. A forma da função $f(\cdot)$ deve ser especificada antes de podermos empreender uma análise econométrica. Uma segunda questão concernente a (1.1) é como lidar com variáveis que não podem ser razoavelmente observadas. Por exemplo, considere o “salário” que uma pessoa pode receber na atividade criminosa. Em princípio, tal quantidade é bem-definida, mas poderia ser difícil, se não impossível, observar o “salário” para um determinado indivíduo. Mesmo variáveis como a probabilidade de ser preso não podem ser obtidas de modo realista para um determinado indivíduo, mas pelo menos podemos observar estatísticas de detenção relevantes e derivar uma variável que se aproxime da probabilidade de prisão. Muitos outros fatores que não podemos listar, nem mesmo observar, afetam o comportamento criminoso, mas devemos de algum modo considerá-los.

As ambigüidades inerentes ao modelo econômico do crime são resolvidas ao se especificar um modelo econométrico particular, tal como:

$$\begin{aligned} \text{crime} = & \beta_0 + \beta_1 \text{saláriom} + \beta_2 \text{outrenda} + \beta_3 \text{freqpris} + \beta_4 \text{freqcond} \\ & + \beta_5 \text{sentmed} + \beta_6 \text{idade} + u, \end{aligned} \quad (1.3)$$

em que crime é alguma medida de freqüência da atividade criminosa, saláriom é o salário que poderia ser ganho em um emprego legal, outrenda é a renda de outras fontes (ativos, herança etc.), freqpris é a freqüência de prisões por infrações anteriores (para aproximar a probabilidade de detenção), freqcond é a freqüência de condenações e sentmed é a duração média da sentença após as condenações. A escolha dessas variáveis é determinada pela teoria econômica e por considerações sobre os dados. O termo u contém fatores não observados, tais como o “salário” da atividade criminosa, o caráter moral, a formação da família e erros na mensuração de coisas como a atividade criminosa e a probabilidade de detenção. Podemos adicionar variáveis de formação da família ao modelo, tais como o número de irmãos, a educação dos pais, e assim por diante, mas nunca poderemos eliminar u inteiramente. De fato, lidar com esse *termo de erro* ou *termo de扰动* é, talvez, o componente mais importante de qualquer análise econométrica.

As constantes $\beta_0, \beta_1, \dots, \beta_6$ são os *parâmetros* do modelo econométrico e descrevem as direções e as influências da relação entre *crime* e os fatores usados para determinar *crime* no modelo.

Um modelo econométrico completo para o Exemplo 1.2 poderia ser

$$\text{salário} = \beta_0 + \beta_1 \text{edu} + \beta_2 \text{exper} + \beta_3 \text{treina} + u, \quad (1.4)$$

em que o termo u contém fatores tais como “aptidão inata”, qualidade da educação, formação da família e uma miríade de outros fatores que podem influenciar o salário de uma pessoa. Se estivermos especialmente interessados nos efeitos do treinamento de trabalho, então β_3 é o parâmetro de interesse.

Na maioria dos casos, a análise econométrica começa pela especificação de um modelo econométrico, sem consideração de detalhes da criação do modelo. Geralmente seguimos essa abordagem, pois, em grande parte, a derivação cuidadosa de algo como o modelo econômico do crime toma muito tempo e pode nos levar para algumas áreas especializadas e freqüentemente difíceis da teoria econômica. O raciocínio econômico desempenhará um papel importante em nossos exemplos, e incorporaremos toda teoria econômica subjacente na especificação do modelo econométrico. No modelo econômico do exemplo do crime, começaríamos com um modelo econométrico tal como (1.3) e usaríamos o raciocínio econômico e o senso comum como guias para escolher as variáveis. Embora essa abordagem perca algumas das profusões da análise econômica, ela é comum e efetivamente aplicada por pesquisadores cautelosos.

Visto que um modelo econométrico tal como (1.3) ou (1.4) tenha sido especificado, várias *hipóteses* de interesse podem ser formuladas em termos dos parâmetros desconhecidos. Por exemplo, na equação (1.3), poderíamos levantar a hipótese de que salário_{it} , o salário que poderia ser ganho no emprego legal, não tem efeito sobre o comportamento criminoso. No contexto desse modelo econométrico específico, a hipótese é equivalente a $\beta_1 = 0$.

Uma análise empírica, por definição, requer dados. Após os dados sobre as variáveis relevantes terem sido coletados, os métodos econométricos são usados para estimar os parâmetros do modelo econométrico e para, formalmente, testar as hipóteses de interesse. Em alguns casos, o modelo econométrico é usado para fazer previsões com a finalidade de testar de uma teoria a estudo do impacto de uma política.

Como a coleta de dados é muito importante em trabalhos empíricos, a Seção 1.3 descreverá os tipos de dados com os quais, provavelmente, nos defrontaremos.

1.3 A ESTRUTURA DOS DADOS ECONÔMICOS

Os dados econômicos apresentam-se em uma variedade de tipos. Embora alguns métodos econométricos possam ser aplicados com pouca ou nenhuma modificação para muitos tipos diferentes de informações, as características especiais de alguns dados devem ser consideradas ou deveriam ser exploradas. Descreveremos a seguir as estruturas de dados mais importantes encontradas nos trabalhos aplicados.

Dados de Corte Transversal

Um **conjunto de dados de corte transversal** consiste em uma amostra de indivíduos, consumidores, empresas, cidades, estados, países ou uma variedade de outras unidades, tomada em um determinado ponto no tempo. Às vezes, os dados de todas as unidades não correspondem precisamente ao mesmo período. Por exemplo, muitas famílias podem ser pesquisadas durante diferentes semanas de um ano. Em uma an-

lise pura de dados de corte transversal, ignorariamós, na coleta de dados, quaisquer diferenças de tempo não importantes. Se o conjunto de famílias fosse pesquisado durante diferentes semanas do mesmo ano, ainda veríamos isso como um conjunto de dados de corte transversal.

Uma importante característica dos dados de corte transversal é que não podemos, freqüentemente, assumir que eles foram obtidos por amostragem aleatória da população subjacente. Por exemplo, se obtemos informações sobre salários, educação, experiência e outras características ao extrair aleatoriamente 500 pessoas de uma população de trabalhadores, teremos uma amostra aleatória da população de todas as pessoas que trabalham. A amostragem aleatória, matéria aprendida nos cursos introdutórios de estatística, simplifica a análise de dados de corte transversal. Uma revisão sobre amostragem aleatória aparece no Apêndice C disponível em www.thomsonlearning.com.br, na página deste livro.

Algumas vezes, a amostragem aleatória não é apropriada como uma hipótese para analisar dados de corte transversal. Por exemplo, suponha que estejamos interessados em estudar fatores que influenciam na acumulação de riqueza das famílias. Podemos estudar uma amostra aleatória de famílias, mas algumas talvez se recusem a relatar suas riquezas. Se, por exemplo, for menos provável que famílias mais ricas revelem sua riqueza, a amostra resultante sobre a riqueza não é uma amostra aleatória extraída da população de todas as famílias. Este é um exemplo de um problema de seleção amostral, um tópico avançado que discutiremos no Capítulo 17.

Outra violação da amostragem aleatória ocorre quando construímos uma amostra a partir de unidades grandes relacionadas à população, em especial a unidades geográficas. O problema provável em tais casos é que a população não é suficientemente grande para se supor, de maneira razoável, que as observações são extrações independentes. Por exemplo, se queremos explicar novas atividades de negócios entre estados, como função de taxas de salários, preços de energia, alíquotas de impostos, serviços prestados, qualidade da força de trabalho e outras características estaduais, é improvável que as atividades de negócios em um Estado próximo a outro sejam independentes. Isso revela que os métodos econômétricos que discutimos funcionam, de fato, em tais situações, mas algumas vezes necessitam ser refinados. Na maioria dos casos, ignoraremos as complexidades que surgem ao analisar tais situações e trataremos esses problemas dentro do arcabouço da amostragem aleatória, mesmo quando não for tecnicamente correto fazê-lo.

Os dados de corte transversal são amplamente usados em economia e em outras ciências sociais. Em economia, a análise de dados de corte transversal está intimamente alinhada com campos da microeconomia aplicada, tais como economia do trabalho, finanças públicas estaduais e locais, organização industrial, economia urbana, demografia e economia da saúde. Dados sobre indivíduos, famílias, empresas e cidades em um determinado ponto do tempo são importantes para testar hipóteses microeconômicas e avaliar políticas governamentais.

Para a análise econômica, os dados de corte transversal usados podem ser representados e armazenados em computadores. A Tabela 1.1 contém, de forma abreviada, um conjunto de dados de corte transversal para o ano de 1976, de 526 trabalhadores. (Esse é um subconjunto dos dados do arquivo WAGE1.RAW*.) As variáveis incluem *salárioh* (salário por hora), *educ* (anos de educação formal), *exper* (anos de experiência no mercado de trabalho), *feminino* (indicador de gênero) e *casado* (estado civil). Estas duas últimas variáveis são binárias (zero-um) por natureza, e servem para indicar características qualitativas dos indivíduos. (A pessoa é do sexo feminino ou não; a pessoa é casada ou não.) Falaremos mais sobre variáveis binárias no Capítulo 7 e seguintes.

* NRT: Todos os arquivos mencionados no texto têm a designação “*.RAW”, mas no banco de dados os arquivos são planilhas em Excel (“.XLS”), e os arquivos de trabalho de programas econômétricos (como “*.WFI”, do Eviews[®]). Portanto, a designação “*.RAW” é genérica e dá um significado de “* matéria-prima” para aplicações e exercícios.

Tabela 1.1

Conjunto de Dados de Corte Transversal sobre Salários e outras Características Individuais

<i>nobsi</i>	<i>salárioh</i>	<i>educ</i>	<i>exper</i>	<i>feminino</i>	<i>casado</i>
1	3,10	11	2	1	0
2	3,24	12	22	1	1
3	3,00	11	2	0	0
4	6,00	8	44	0	1
5	5,30	12	7	0	1
.
.
.
525	11,56	16	5	0	1
526	3,50	14	5	1	0

A variável *nobsi* na Tabela 1.1 é o número da observação atribuído a cada indivíduo na amostra. Diferentemente das outras variáveis, ela não é uma característica do indivíduo. Todos os programas econômétricos e estatísticos atribuem a cada unidade um número de observação. A intuição deveria dizer-lhe que, para dados como os da Tabela 1.1, não importa qual pessoa é classificada como observação um, qual pessoa é designada pela observação dois, e assim por diante. O fato de que a ordenação dos dados não importa para a análise econômética é uma característica fundamental dos conjuntos de dados de corte transversal obtidos a partir da amostragem aleatória.

Às vezes, variáveis diferentes correspondem a diferentes períodos nos conjuntos de dados de corte transversal. Por exemplo, a fim de determinar os efeitos de políticas governamentais sobre o crescimento econômico de longo prazo, os economistas têm estudado a relação entre crescimento do PIB *per capita* real ao longo de certo período (digamos, 1960 a 1985) e variáveis determinadas, em parte, pela política governamental em 1960 (consumo do governo como percentagem do PIB e taxas de ensino médio de adultos). Tais conjuntos de dados poderiam ser representados como na Tabela 1.2, a qual constitui parte do conjunto de dados usados no estudo de De Long e Summers (1991) sobre as taxas de crescimento entre países.

A variável *cpibpcr* representa o crescimento médio do PIB *per capita* real ao longo do período 1960 a 1985. O fato de que *consgov60* (consumo do governo como percentagem do PIB) e *second60* (percentagem da população adulta com ensino médio) correspondem ao ano de 1960, enquanto *cpibpcr* é o crescimento médio ao longo do período 1960 a 1985, não leva a quaisquer problemas especiais ao tratar essas informações como um conjunto de dados de corte transversal. As observações estão ordenadas alfabeticamente por país, mas essa ordenação não afeta em nada qualquer análise subsequente.

Tabela 1.2

Conjunto de Dados sobre Taxas de Crescimento Econômico e Características de Países

<i>nobsp</i>	<i>país</i>	<i>cpiibpcr</i>	<i>consgov60</i>	<i>second60</i>
1	Argentina	0,89	9	32
2	Áustria	3,32	16	50
3	Bélgica	2,56	13	69
4	Bolívia	1,24	18	12
.
.
.
61	Zimbábue	2,30	17	6

Dados de Séries de Tempo

Um **conjunto de dados de séries de tempo** consiste em observações sobre uma variável ou muitas variáveis ao longo do tempo. Exemplos de dados de séries temporais incluem preços de ações, oferta de moeda, índice de preços ao consumidor, produto interno bruto, taxas anuais de homicídios e números de vendas de automóveis. Como eventos passados podem influenciar eventos futuros, e como, nas ciências sociais, as defasagens do comportamento são prevalentes, o tempo é uma dimensão importante em um conjunto de dados de séries de tempo. Diferentemente do arranjo dos dados de corte transversal, a ordenação cronológica das observações em uma série de tempo transmite informações potencialmente importantes.

Uma característica essencial dos dados de séries de tempo que torna mais difícil a análise do que os dados de corte transversal é o fato de que raramente é possível assumir (se é que é possível) que as observações econômicas são independentes ao longo do tempo. A maioria das séries de tempo econômicas, bem como de outras séries de tempo, está relacionada – muitas vezes fortemente relacionada – com seus históricos recentes. Por exemplo, saber algo sobre o produto interno bruto do último trimestre nos diz muito sobre a provável variação do PIB durante este trimestre, visto que o PIB tende a permanecer razoavelmente estável de um trimestre para o próximo. Embora muitos procedimentos econométricos possam ser usados tanto com dados de corte transversal como com dados de séries de tempo, outros pontos podem ser considerados para especificar, apropriadamente, os modelos econométricos que usam dados de séries de tempo. Além disso, as modificações e embelezamentos das técnicas econométricas comuns foram desenvolvidas com a finalidade de considerar e explorar a natureza dependente das séries de tempo e para tratar de outras questões, tal como o fato de que algumas variáveis econômicas tendem a exibir claras tendências ao longo do tempo.

Outra característica dos dados de séries de tempo que pode requerer atenção especial é a **frequência dos dados**, na qual eles são coletados. Em economia, as frequências mais comuns são: diária, semanal, mensal, trimestral e anual. Os preços de ações são registrados em intervalos diários (excluindo sábados e domingos). A oferta de moeda na economia dos Estados Unidos é informada semanalmente. Muitas séries macroeconômicas são tabuladas mensalmente, incluindo as taxas de inflação e desemprego. Outras séries macroeconômicas são registradas menos freqüentemente, como a cada três meses (todo trimestre). O produto interno bruto é um exemplo importante de uma série trimestral. Outras séries de tempo, como as taxas de mortalidade infantil dos estados norte-americanos, estão disponíveis somente em bases anuais.

Muitas séries de tempo econômicas, sejam semanais, mensais ou trimestrais, exibem um forte padrão sazonal, o qual pode ser um importante fator na análise de séries de tempo. Por exemplo, dados mensais sobre o início da construção de moradias se diferenciam entre os meses simplesmente devido a mudanças das condições climáticas. Aprenderemos como trabalhar com séries de tempo no Capítulo 10.

A Tabela 1.3 contém um conjunto de dados de séries de tempo, obtido de um artigo de Castillo-Freeman e Freeman (1992), sobre os efeitos do salário mínimo em Porto Rico. O ano mais antigo no conjunto de dados é a primeira observação, e o ano mais recente disponível é a última observação. Quando os métodos econométricos são utilizados para analisar dados de séries de tempo, os dados devem ser armazenados em ordem cronológica.

Tabela 1.3

Salário Mínimo, Desemprego e Dados Relacionados para Porto Rico

<i>nobsa</i>	<i>ano</i>	<i>minmed</i>	<i>cobmed</i>	<i>desemp</i>	<i>pnb</i>
1	1950	0,20	20,1	15,4	878,7
2	1951	0,21	20,7	16,0	925,0
3	1952	0,23	22,6	14,8	1.015,9
.
.
.
37	1986	3,35	58,1	18,9	4.281,6
38	1987	3,35	58,2	16,8	4.496,7

A variável *minmed* se refere ao salário mínimo médio no ano, *cobmed* é a taxa de cobertura média (o percentual de trabalhadores cobertos pela lei de salário mínimo), *desemp* é a taxa de desemprego e *pnb* é o produto nacional bruto. Usaremos esses dados mais adiante em uma análise de séries de tempo do efeito do salário mínimo sobre o emprego.

Cortes Transversais Agrupados

Alguns conjuntos de dados têm tanto características de corte transversal quanto de séries de tempo. Por exemplo, suponha que dois estudos sobre famílias sejam realizados nos Estados Unidos com dados de corte transversal, um em 1985 e outro em 1990. Em 1985, uma amostra aleatória de famílias é pesquisada para variáveis tais como renda, poupança, tamanho da família, e assim por diante. Em 1990, uma nova amostra aleatória de famílias é extraída usando as mesmas questões da pesquisa. A fim de aumentar nosso tamanho de amostra, podemos formar um **corte transversal agrupado** ao combinar os dois anos.

Agrupar cortes transversais de diferentes anos é, frequentemente, um modo eficaz de analisar os efeitos de uma nova política de governo. A idéia é coletar dados de anos anteriores e posteriores a uma importante mudança de política governamental. Como exemplo, considere o seguinte conjunto de dados sobre os preços da moradia coletados em 1993 e 1995 nos Estados Unidos, quando houve uma redução nos impostos sobre a propriedade em 1994. Suponha que tenhamos dados sobre 250 residências para 1993 e sobre 270 para 1995. Um modo de armazenar tais dados é apresentado na Tabela 1.4.

Tabela 1.4

Cortes-Transversais Agrupados: Dois Anos de Preços de Moradias

<i>nobsm</i>	<i>ano</i>	<i>preçoc</i>	<i>imppro</i>	<i>arquad</i>	<i>qdorm</i>	<i>banhos</i>
1	1993	85.500	42	1.600	3	2,0
2	1993	67.300	36	1.440	3	2,5
3	1993	134.000	38	2.000	4	2,5
.
.
.
250	1993	243.600	41	2.600	4	3,0
251	1995	65.000	16	1.250	2	1,0
252	1995	182.400	20	2.200	4	2,0
253	1995	97.500	15	1.540	3	2,0
.
.
.
520	1995	57.200	16	1.100	2	1,5

As observações 1 a 250 correspondem às residências vendidas em 1993, e as observações 251 a 520 correspondem às 270 residências vendidas em 1995. Embora a ordem na qual armazenamos os dados não se revele crucial, não se esqueça de que o ano de cada observação é, geralmente, muito importante. Essa é a razão de introduzirmos *ano* como uma variável separada.

A análise de um corte transversal agrupado é muito parecida com a de um corte transversal padrão, exceto pelo fato de que precisamos, freqüentemente, considerar diferenças periódicas das variáveis ao longo do tempo. De fato, além de aumentar o tamanho da amostra, a característica de uma análise de corte transversal agrupada é, freqüentemente, ver como uma relação fundamental mudou ao longo do tempo.

Dados de Painel ou Longitudinais

Um conjunto de **dados de painel** (ou dados longitudinais) consiste em uma série de tempo para *cada* membro do corte transversal do conjunto de dados. Como exemplo, suponha que tenhamos o histórico de salário, educação e emprego para um conjunto de indivíduos ao longo de um período de dez anos, ou que possamos coletar informações, tais como dados de investimento e financeiros, sobre o mesmo conjunto de empresas ao longo de um período de cinco anos. Dados de painel também podem ser coletados para unidades geográficas. Por exemplo, podemos coletar dados para o mesmo conjunto de municípios dos Estados Unidos sobre fluxos de imigração, impostos, taxas de salários, gastos governamentais etc., para os anos de 1980, 1985 e 1990.

A característica essencial dos dados de painel que os distingue dos dados de corte transversal agrupado é o fato de que as *mesmas* unidades do corte transversal (indivíduos, empresas ou municípios nos exemplos anteriores) são acompanhadas ao longo de um determinado período. Os dados na Tabela 1.4 não são considerados um conjunto de dados de painel porque as residências vendidas são provavelmente diferentes em 1993 e 1995; se houver quaisquer repetições, o número é provavelmente bem pequeno para ser significante. Em contraste, a Tabela 1.5 contém um conjunto de dados de painel de dois anos sobre o crime e as estatísticas relacionadas para 150 cidades nos Estados Unidos.

Há várias características interessantes na Tabela 1.5. Primeiro, a cada cidade foi dado um número de 1 a 150. Qual cidade decidimos chamar de cidade 1, cidade 2, e assim por diante, é irrelevante. Assim como em um corte transversal puro, a ordenação no corte transversal de um conjunto de dados de painel não é importante. Poderíamos usar o nome da cidade em lugar de um número, mas é freqüentemente útil ter ambos.

Tabela 1.5

Conjunto de Dados de Painel sobre Estatísticas de Crime nas Cidades para Dois Anos

<i>nobsc</i>	<i>cidade</i>	<i>ano</i>	<i>homicds</i>	<i>população</i>	<i>desemp</i>	<i>pólicia</i>
1	1	1986	5	350.000	8,7	440
2	1	1990	8	359.200	7,2	471
3	2	1986	2	64.300	5,4	75
4	2	1990	1	65.100	5,5	75
.
.
.
297	149	1986	10	260.700	9,6	286
298	149	1990	6	245.000	9,8	334
299	150	1986	25	543.000	4,3	520
300	150	1990	32	546.200	5,2	493

Um segundo ponto é que os dois anos dos dados para a cidade 1 preenchem as duas primeiras linhas ou observações. As observações 3 e 4 correspondem à cidade 2, e assim por diante. Como cada uma das 150 cidades tem duas linhas de dados, qualquer pacote econométrico verá isso como 300 observações. Esse conjunto de dados pode ser tratado como um corte transversal agrupado, em que as mesmas cidades aparecem em cada ano. Porém, como veremos nos Capítulos 13 e 14, podemos também usar a estrutura de painel para responder a questões que não podem ser respondidas simplesmente vendo isso como um corte transversal agrupado.

Ao organizar as observações na Tabela 1.5, colocamos os dois anos dos dados de cada cidade um ao lado do outro, com o primeiro ano antecedendo o segundo em todos os casos. Apenas por questões práticas, esse é o modo preferido de se ordenar conjuntos de dados de painel. Essa organização contrasta com o modo pelo qual os cortes transversais agrupados são armazenados na Tabela 1.4. Em resu-

mo, a razão para ordenar os dados de painel como na Tabela 1.5 é que precisaremos fazer transformações dos dados para cada cidade nos dois anos.

Como os dados de painel requerem a repetição das mesmas unidades ao longo do tempo, os conjuntos de dados de painel, especialmente aqueles sobre indivíduos, famílias e empresas, são mais difíceis de se obter que os cortes transversais agrupados. Não surpreendentemente, observar as mesmas unidades ao longo do tempo traz várias vantagens sobre os dados de corte transversal ou mesmo sobre os de dados de cortes transversais agrupados. O benefício que salientaremos neste livro é que ter múltiplas observações sobre as mesmas unidades nos permite controlar certas características não observáveis dos indivíduos, firmas etc. Como veremos, o uso de mais de uma observação pode facilitar a inferência causal em situações em que inferir causalidade seria muito difícil se somente um único corte transversal estivesse disponível. Uma segunda vantagem dos dados de painel é que eles, frequentemente, nos permitem estudar a importância das defasagens do comportamento ou o resultado de tomar decisões. Essa informação pode ser importante, visto que se pode esperar o impacto em muitas políticas públicas somente após algum tempo.

A maior parte dos livros para cursos de nível superior não contém uma discussão de métodos econométricos para dados de painel. Entretanto, os economistas agora reconhecem que algumas questões são difíceis, se não impossíveis, de serem respondidas satisfatoriamente sem dados de painel. Como você verá, podemos fazer consideráveis progressos com análises simples de dados de painel, um método que não é muito mais difícil do que trabalhar com um conjunto de dados de corte transversal padrão.

Um Comentário sobre Estruturas de Dados

A Parte 1 deste livro cobre a análise de dados de corte transversal, já que ela propõe menos conceitos e dificuldades técnicas. Ao mesmo tempo, ela ilustra muitos dos temas essenciais da análise econométrica. Usaremos os métodos e as idéias da análise de corte transversal no restante do texto.

Embora a análise econométrica de séries de tempo use muitas das mesmas ferramentas que a análise de corte transversal, ela é mais complicada devido à existência de tendência que traduz a natureza altamente persistente de muitas séries de tempo econômicas. Acredita-se que agora são considerados falhos muitos exemplos que têm sido tradicionalmente usados para ilustrar a maneira pela qual os métodos econométricos podem ser aplicados a dados de séries de tempo. Faz pouco sentido usar tais exemplos inicialmente, visto que esse hábito somente reforça uma prática econométrica insatisfatória. Portanto, postergaremos o tratamento da econometria de séries de tempo até a Parte 2, quando questões importantes concernentes à tendência, persistência, dinâmica e sazonalidade serão introduzidas.

Na Parte 3, trataremos explicitamente de cortes transversais agrupados e dados de painel. A análise de cortes transversais independentemente agrupados e a análise simples de dados de painel são ambas independentemente, extensões claras e diretas da análise pura de corte transversal. Entretanto, vamos esperar até o Capítulo 13 para tratar desses tópicos.

1.4 A CAUSALIDADE E A NOÇÃO DE *CETERIS PARIBUS* NA ANÁLISE ECONOMÉTRICA

Em muitos testes de teoria econômica, e certamente para avaliar políticas públicas, o objetivo do economista é inferir que uma variável (tal como a educação) tem um efeito causal sobre outra variável (tal como a produtividade do trabalhador). Encontrar simplesmente uma associação entre duas ou mais variáveis pode ser sugestivo, mas, a não ser que se possa estabelecer uma causalidade, raramente ela é convincente.

A noção de *ceteris paribus* – que significa “outros fatores (relevantes) permanecendo iguais” – desempenha um papel importante na análise causal. Essa idéia esteve implícita em algumas de nossas discussões anteriores, particularmente nos Exemplos 1.1 e 1.2, mas até agora não a mencionamos explicitamente.

Você provavelmente se lembra de que na economia introdutória muitas questões econômicas são *ceteris paribus* por natureza. Por exemplo, na análise da demanda do consumidor, estamos interessados em conhecer o efeito da variação do preço de um bem sobre sua quantidade demandada, enquanto todos os outros fatores – tais como renda, preços de outros bens e gostos individuais – permanecem fixos. Se outros fatores não forem mantidos fixos, não poderemos conhecer o efeito causal de uma variação do preço sobre a quantidade demandada.

Manter fixos os outros fatores também é crucial para a análise da política governamental. No exemplo do treinamento de trabalho (Exemplo 1.2), poderíamos nos interessar pelo efeito de outra semana de treinamento sobre os salários, com todos os outros componentes permanecendo iguais (em particular, educação e experiência). Se conseguirmos manter fixos todos os outros fatores relevantes e, em seguida, acharmos uma ligação entre treinamento e salários, poderemos concluir que o treinamento tem um efeito causal sobre a produtividade do trabalhador. Embora isso possa parecer simples, mesmo nesse estágio inicial deve ficar claro que, exceto em casos muito especiais, não será possível, literalmente, manter tudo o mais igual. A questão fundamental na maioria dos estudos empíricos é: foram mantidos fixos em número suficiente outros fatores, para que se possa inferir a causalidade? Raramente avalia-se um estudo econométrico sem levantar essa questão.

Em muitas aplicações sérias, o número de fatores que podem afetar a variável de interesse – tal como a atividade criminosa ou os salários – é imenso, e isolar qualquer variável particular pode parecer um esforço inútil. Entretanto, veremos no final que, quando cuidadosamente aplicados, os métodos econométricos podem simular um experimento *ceteris paribus*.

Neste ponto, não podemos ainda explicar como os métodos econométricos são usados para estimar efeitos *ceteris paribus*; desse modo, consideraremos alguns problemas que podem surgir ao se tentar inferir causalidade em economia. Não vamos usar nenhuma equação nessa discussão. Para cada exemplo, o problema de inferir causalidade desaparece se um experimento apropriado puder ser conduzido. Assim, é útil descrever como tal experimento poderia ser estruturado e observar que, em muitos casos, obter dados experimentais é impraticável. Também é de grande auxílio pensar por que os dados disponíveis às vezes não têm as principais características de um conjunto de dados experimentais.

Daqui em diante, contaremos com a compreensão intuitiva dos termos *aleatório*, *independência* e *correlação*, que devem ser familiares para quem estudou probabilidade e estatística. (Esses conceitos são revistos no Apêndice B, disponível na página do livro, no site www.thomsonlearning.com.br.) Vamos começar com um exemplo que ilustra algumas dessas questões importantes.

EXEMPLO 1.3

(Efeitos dos Fertilizantes sobre a Produção Agrícola)

Alguns dos primeiros estudos econométricos [por exemplo, Griliches (1957)] consideraram os efeitos de novos fertilizantes sobre a produção agrícola. Suponha a soja como o produto em consideração. Como a quantidade de fertilizantes é somente um fator que afeta a produção – outros fatores incluem chuva, qualidade da terra e presença de parasitas –, essa questão deve ser levantada como uma questão *ceteris paribus*. Uma maneira de determinar o efeito causal da quantidade de fertilizantes sobre a produção de soja é conduzir um experimento, que poderia incluir os seguintes passos. Escolha vários lotes de terra de um acre. Aplique diferentes quantidades de fertilizante em cada lote e, subseqüentemente, mensure a produção; isso nos dá

EXEMPLO 1.3 (continuação)

um conjunto de dados de corte transversal. Em seguida, use os métodos estatísticos (a serem introduzidos no Capítulo 2) para medir a associação entre produção de soja e quantidades de fertilizantes.

Como descrito anteriormente, isso pode não parecer um experimento muito bom, pois não dissemos nada sobre escolher lotes de terra que sejam idênticos em todos os aspectos, com exceção da quantidade de fertilizantes. De fato, escolher lotes de terra com essa característica não é exequível: alguns dos fatores, como a qualidade da terra, não podem ser, de fato, observados. Como sabemos que os resultados desse experimento podem ser usados para mensurar o efeito *ceteris paribus* dos fertilizantes? A resposta depende das especificidades de como as quantidades de fertilizantes são escolhidas. Se os níveis de fertilizantes são atribuídos aos lotes independentemente de outras características do lote que afetam a produção – isto é, outras características dos lotes são completamente ignoradas quando se decide sobre as quantidades de fertilizantes –, então podemos começar a fazer o que planejamos. Justificaremos essa afirmação no Capítulo 2.

O próximo exemplo é mais representativo das dificuldades que surgem ao se inferir causalidade em economia aplicada.

EXEMPLO 1.4**(Medindo o Retorno da Educação)**

Os economistas especializados em trabalho e os *formuladores de políticas públicas* há muito se interessam pelo “retorno da educação”. De modo um tanto informal, a questão é colocada da seguinte maneira: se uma pessoa é escolhida de uma população, e recebe um ano a mais de educação, em quanto aumentará seu salário? Assim como nos exemplos anteriores, essa é uma questão *ceteris paribus*, que implica que todos os outros fatores são mantidos fixos enquanto a pessoa recebe um ano a mais de educação.

Podemos imaginar um planejador social esquematizando um experimento para estudar essa questão, da mesma maneira que o pesquisador agrícola pode projetar um experimento para estimar os efeitos dos fertilizantes. Uma abordagem é seguir o exemplo dos fertilizantes no Exemplo 1.3: escolha um grupo de pessoas, dê aleatoriamente a cada pessoa uma quantidade de educação (algumas pessoas recebem alguns anos de estudo que equivalem ao ensino fundamental, a outras é dado uma educação que equivale ao ensino médio etc.), e, em seguida, mensure seus salários (assumindo que cada uma delas no momento trabalha). Aqui, as pessoas são como os lotes no exemplo dos fertilizantes, em que a educação desempenha o papel dos fertilizantes, e o salário, o da produção da soja. Como no Exemplo 1.3, se níveis de educação forem atribuídos independentemente de outras características que afetam a produtividade (tal como experiência e aptidão inata), uma análise que ignorar esses outros fatores produzirá resultados úteis. Uma vez mais, no Capítulo 2, faremos algum esforço para justificar essa afirmação; por ora, ela é formulada sem sustentação.

Diferentemente do exemplo fertilizante-produção, o experimento descrito no Exemplo 1.4 é inexistível. As questões morais – sem mencionar os custos econômicos – associadas à determinação aleatória dos níveis de educação para um grupo de indivíduos são óbvias. Além disso, não há lógica em simplesmente atribuir a alguém alguns anos de educação se tal pessoa já completou o curso superior.

Embora dados experimentais não possam ser obtidos para medir o retorno da educação, podemos certamente coletar dados não-experimentais sobre níveis de educação e salários para um grupo grande, fazendo amostras aleatórias da população de trabalhadores. Tais dados estão disponíveis em uma variedade de pesquisas usadas em economia do trabalho, mas esses conjuntos de dados têm uma característica que torna difícil estimar o retorno *ceteris paribus* da educação.

As pessoas *escolhem* seus próprios níveis de educação; portanto, os níveis de educação não são, provavelmente, determinados independentemente de todos os outros fatores que afetam os salários. Esse problema é uma característica compartilhada de muitos conjuntos de dados não-experimentais.

Um fator que afeta o salário é a experiência da força de trabalho. Visto que possuir mais educação requer, geralmente, um adiamento da entrada na força de trabalho, aqueles com mais educação têm, muitas vezes, menos experiência. Assim, em um conjunto de dados não-experimentais sobre salários e educação, provavelmente a educação está negativamente associada com uma variável fundamental que também afeta o salário. Acredita-se também que pessoas com mais aptidão inata escolhem, freqüentemente, níveis de educação mais altos. Como aptidão maior leva a salários maiores, temos novamente uma correlação entre educação e um fator crucial que afeta o salário.

Os fatores omitidos no exemplo dos salários, experiência e aptidão, têm semelhança no exemplo dos fertilizantes. A experiência é, em geral, fácil de mensurar e, portanto, similar a uma variável como a chuva. A aptidão, no entanto, é algo vago e difícil de quantificar; ela é similar à qualidade da terra no exemplo dos fertilizantes. Como veremos ao longo deste livro, considerar outros fatores, como a experiência, ao estimar o efeito *ceteris paribus* de outra variável, como a educação, é algo relativamente direto e simples. Também descobriremos que considerar fatores inherentemente não observáveis, como a aptidão, é muito mais problemático. Pode-se dizer que muitos dos avanços nos métodos econométricos têm tentado lidar com fatores não observados nos modelos econométricos.

Podemos fazer um último paralelo entre os exemplos 1.3 e 1.4. Suponha que, no caso dos fertilizantes, as quantidades de fertilizantes não sejam completamente determinadas de modo aleatório. Em vez disso, o assistente que escolheu os níveis de fertilizante pensou que seria melhor colocar mais fertilizante nas áreas de terra de maior qualidade. (Os pesquisadores agrícolas devem ter uma idéia aproximada sobre quais áreas de terra têm melhor qualidade, ainda que eles não possam ser capazes de quantificar totalmente as diferenças.) Essa situação é completamente análoga à da relação estabelecida entre o nível de escolaridade e a aptidão não observada no Exemplo 1.4. Como terras melhores levam a safras maiores, e mais fertilizantes foram usados nas melhores áreas, qualquer relação observada entre produção de soja e quantidade de fertilizantes poderia ser espúria.

EXEMPLO 1.5

(O Efeito do Cumprimento da Lei sobre os Níveis de Criminalidade das Cidades)

A questão de como impedir a criminalidade está – e provavelmente continuará – entre nós há um bom tempo. Uma indagação especialmente importante sobre esse aspecto é: a presença de mais policiais nas ruas detém a criminalidade?

A questão *ceteris paribus* é fácil de formular. Se uma cidade fosse escolhida aleatoriamente e recebesse, por exemplo, dez policiais a mais, em quanto suas taxas de criminalidade cairiam? Outra maneira de formular a questão é: se duas cidades fossem, em todos os aspectos, iguais, exceto que a cidade A tivesse dez policiais a mais que a cidade B, em quanto difeririam as taxas de criminalidade das duas cidades?

Seria virtualmente impossível encontrar pares de comunidades idênticas em todos os aspectos, exceto no que respeita ao tamanho de suas forças policiais. Felizmente, a análise econométrica não requer isso. O que, de fato, precisamos saber é se os dados que podemos coletar sobre os níveis de criminalidade de uma comunidade e o tamanho de sua força policial podem ser vistos como experimentais. Podemos, certamente, imaginar um experimento verdadeiro, envolvendo um grande número de cidades, em que decidimos quantos policiais cada uma delas usará no ano seguinte.

EXEMPLO 1.5 (continuação)

Embora os policiais possam ser usados para produzir um efeito sobre o tamanho das forças policiais, certamente não podemos dizer a cada cidade quantos policiais ela deve empregar. Se, como é provável, a decisão de uma cidade sobre quantos policiais empregar estiver correlacionada com outros fatores relativos às cidades que afetam a criminalidade, os dados deverão ser vistos como não experimentais. De fato, um modo de ver esse problema é observar que as escolhas de uma cidade relativamente ao tamanho da força policial e a quantidade de crimes são *simultaneamente determinadas*. Vamos tratar explicitamente desse problema no Capítulo 16.

Os três primeiros exemplos que discutimos utilizaram com dados de corte transversal em vários níveis de agregação (por exemplo, do indivíduo ou da cidade). Os mesmos obstáculos surgem ao se inferir causalidade em problemas de séries de tempo.

EXEMPLO 1.6**(O Efeito do Salário Mínimo sobre o Desemprego)**

Uma questão de política governamental importante, e talvez controversa, diz respeito ao efeito do salário mínimo sobre as taxas de desemprego para vários grupos de trabalhadores. Embora esse problema possa ser estudado dentro de uma variedade de estruturas de dados (dados de corte transversal, de séries de tempo ou de painel), os dados de séries de tempo são, freqüentemente, usados para observar efeitos agregados. Um exemplo de um conjunto de dados de séries de tempo relativo a taxas de desemprego e salários mínimos foi dado na Tabela 1.3.

A análise-padrão de oferta e demanda implica que, quando o salário mínimo cresce acima do salário de equilíbrio de mercado, há um movimento para cima ao longo da curva de demanda por trabalho e o emprego total diminui. (A oferta de trabalho excede a demanda por trabalho.) Para quantificar esse efeito, podemos estudar a relação entre emprego e salário mínimo ao longo do tempo. Além de algumas dificuldades especiais que podem surgir ao se lidar com dados de séries de tempo, há possíveis problemas com a inferência de causalidade. O salário mínimo nos Estados Unidos não é determinado individualmente. Várias forças econômicas e políticas exercem forte influência sobre o salário mínimo de qualquer ano. (O salário mínimo, uma vez determinado, fica geralmente congelado por muitos anos, a não ser que esteja indexado à inflação.) Assim, é provável que o salário mínimo esteja relacionado com outros fatores que têm efeito sobre os níveis de emprego.

Podemos imaginar o governo dos Estados Unidos conduzindo um experimento para determinar os efeitos do salário mínimo sobre o emprego (em vez de se preocupar com o bem-estar dos trabalhadores que ganham salários baixos). O salário mínimo poderia ser estabelecido aleatoriamente pelo governo a cada ano, enquanto os resultados do emprego poderiam ser tabulados. Os dados experimentais de séries de tempo resultantes poderiam, em seguida, ser analisados usando métodos econométricos razoavelmente simples. Mas esse cenário dificilmente descreveria como os salários mínimos são determinados.

Se pudéssemos controlar suficientemente outros fatores relacionados com o emprego, ainda poderíamos esperar estimar o efeito *ceteris paribus* do salário mínimo sobre o emprego. Nesse sentido, o problema seria muito similar aos exemplos anteriores de corte transversal.

Mesmo quando as teorias econômicas não são mais naturalmente descritas em termos de causalidade, elas geralmente têm previsões que podem ser testadas por meio de métodos econométricos. O exemplo seguinte demonstra essa abordagem.

EXEMPLO 1.7**(A Hipótese das Expectativas)**

A *hipótese das expectativas* da economia financeira afirma que, dadas todas as informações disponíveis ao investidor no momento de investir, o retorno *esperado* de quaisquer dois investimentos é o mesmo. Por exemplo, considere dois possíveis investimentos, com um horizonte de investimento de três meses, adquiridos no mesmo momento. (1) Comprar um título do Tesouro norte-americano de três meses, com valor de face de \$ 10.000, por um preço abaixo de \$ 10.000; em três meses, você recebe \$ 10.000. (2) Comprar um título de seis meses (a um preço abaixo de \$ 10.000) e, em três meses, vendê-lo como um título de três meses. Cada investimento requer, mais ou menos, a mesma quantidade de capital inicial, mas há uma diferença importante. Para o primeiro investimento, você sabe exatamente qual é o retorno no momento da compra, porque você sabe o preço inicial do título de três meses e de seu valor de face. Isso não é verdade para o segundo investimento: embora saiba o preço do título de três meses quando o compra, você não conhece o preço pelo qual o venderá em três meses. Portanto, há incerteza com relação a esse investimento para aqueles que têm um horizonte de investimento de três meses.

Os retornos reais desses dois investimentos serão, em geral, diferentes. De acordo com a hipótese das expectativas, o retorno esperado do segundo investimento, dadas todas as informações no momento do investimento, deve-se igualar ao retorno de se adquirir um título de três meses. Essa teoria é razoavelmente fácil de testar, como veremos no Capítulo 11.

Neste capítulo introdutório, discutimos o propósito e o escopo da análise econométrica. A econometria é utilizada em todos os campos da economia aplicada para testar teorias econômicas, para informar o governo, principalmente os formuladores de políticas públicas e o setor privado, e para prever séries de tempo econômicas. Às vezes, um modelo econométrico é derivado de um modelo econômico formal, mas, em outros casos, os modelos econométricos são baseados em raciocínios econômicos informais e na intuição. O objetivo da análise econométrica é estimar os parâmetros do modelo e testar as hipóteses sobre esses parâmetros; os valores e os sinais dos parâmetros determinam a validade de uma teoria econômica e os efeitos de determinadas políticas públicas.

Dados de corte transversal, de séries de tempo, de cortes transversais agrupados e de painel são os tipos mais comuns de estruturas de dados usadas na econometria aplicada. Conjuntos de dados que envolvem uma dimensão temporal, como os dados de séries de tempo e de painel, requerem tratamento especial devido à correlação através do tempo de muitas séries econômicas. Outras questões, tais como tendências e sazonalidade, surgem na análise de séries de tempo, mas não na análise de dados de corte transversal.

Na Seção 1.4, discutimos as noções de *ceteris paribus* e de inferência causal. Em muitos casos, as hipóteses das ciências sociais são, por natureza, *ceteris paribus*: todos os outros fatores relevantes devem estar fixos ao se estudar a relação entre duas variáveis. Por causa da natureza não-experimental de muitos dados coletados nas ciências sociais, descobrir relações causais é muito desafiador.

Análise de Regressão com Dados de Corte Transversal

parte 1 do texto aborda a análise de regressão com dados de corte transversal. Ela se apóia na álgebra estudada nos cursos superiores e em conceitos básicos de probabilidade e estatística. Os Apêndices A, B e C contêm revisões completas sobre esses tópicos.

O Capítulo 2 tem início com o modelo de regressão linear simples, no qual explicamos uma variável em termos de outra. Embora a regressão simples não seja amplamente usada em econometria aplicada, ela é utilizada ocasionalmente e serve como um ponto de partida natural, pois sua álgebra e suas interpretações são relativamente simples.

Os capítulos 3 e 4 cobrem os fundamentos da análise de regressão múltipla, em que permitimos que mais variáveis afetem a variável que estamos tentando explicar. A regressão múltipla é ainda o método mais geralmente usado na pesquisa empírica, de modo que esses capítulos merecem atenção cuidadosa. O capítulo 3 enfatiza a álgebra do método de mínimos quadrados ordinários (MQO), estabelecendo ainda as condições necessárias para que os estimadores MQO sejam não-viesados e também os melhores estimadores lineares não-viesados. O Capítulo 4 trata do importante tópico da inferência estatística.

O Capítulo 5 discute as propriedades referentes às amostras grandes, ou assintóticas, dos estimadores MQO. Essa discussão oferece a justificativa para os procedimentos de inferência do Capítulo 4 quando os erros em um modelo de regressão não são normalmente distribuídos. O capítulo 6 cobre alguns tópicos adicionais da análise de regressão, incluindo questões avançadas sobre forma funcional, transformação dos dados, previsão e grau de ajuste da estimação. O Capítulo 7 explica como a informação qualitativa pode ser incorporada em modelos de regressão múltipla.

O Capítulo 8 ilustra como testar e corrigir o problema da heteroscedasticidade, ou variância não-constante, no termo erro. Mostramos como as estatísticas MQO usuais podem ser ajustadas e também apresentamos uma extensão do método MQO, conhecida como método dos *mínimos quadrados ponderados*, que explica diretamente as diferentes variâncias dos erros. O Capítulo 9 explora o importante problema da correlação entre o termo erro e uma ou mais das variáveis explicativas. Demonstramos como a utilização de uma variável *proxy* pode resolver o problema de variáveis omitidas. Adicionalmente, determinamos o viés e a inconsistência dos estimadores MQO na presença de certos tipos de erros de medida nas variáveis. Diversos problemas de tratamento dos dados são também discutidos, incluindo o problema dos *outliers*.

O Modelo de Regressão Simples

modelo de regressão simples pode ser usado para estudar a relação entre duas variáveis. Por razões que veremos adiante, o modelo de regressão simples tem limitações enquanto ferramenta geral para a análise empírica. No entanto, às vezes ele é apropriado como ferramenta empírica. Aprender como interpretar o modelo de regressão simples é uma boa prática para estudar a regressão múltipla, o que faremos nos capítulos subsequentes.

2.1 DEFINIÇÃO DO MODELO DE REGRESSÃO SIMPLES

Grande parte da análise econometrística começa com a seguinte premissa: y e x são duas variáveis, representando alguma população, e estamos interessados em “explicar y em termos de x ”, ou em “estudar como y varia com variações em x ”. Discutimos alguns exemplos no Capítulo 1, incluindo: y é a produção de soja, e x , a quantidade de fertilizantes; y é o salário-hora, e x , anos de educação; e y é uma taxa de criminalidade em uma comunidade, e x , o número de policiais.

Ao escrever um modelo que “explicará y em termos de x ”, defrontamo-nos com três questões. Primeira, como nunca há uma relação exata entre duas variáveis, como consideramos outros fatores que afetam y ? Segunda, qual é a relação funcional entre y e x ? E terceira, como podemos estar certos de que estamos capturando uma relação *ceteris paribus* entre y e x (se esse for um objetivo desejado)?

Podemos resolver essas ambigüidades escrevendo uma equação que relaciona y a x . Uma equação simples é

$$y = \beta_0 + \beta_1 x + u. \quad (2.1)$$

A equação (2.1), que supostamente é válida para a população de interesse, define o **modelo de regressão linear simples**. Ela também é chamada *modelo de regressão linear de duas variáveis* ou *modelo de regressão linear bivariada*, pois relaciona as duas variáveis x e y . Vamos discutir, agora, o significado de cada uma das quantidades em (2.1). (A propósito, o termo “regressão” tem origens que não são especialmente importantes para muitas das aplicações econometrísticas modernas, de modo que não o explicaremos aqui. Veja Stigler (1986) para uma história interessante da análise de regressão.)

Quando relacionadas por (2.1), as variáveis y e x têm vários nomes diferentes, os quais são intercambiáveis, como explicado em seguida. y é chamada a **variável dependente**, a **variável explicada**, a **variável de resposta**, a **variável prevista**, ou o **regressando**. x é chamada a **variável independente**, a **variável explicativa**, a **variável de controle**, a **variável previsora**, ou o **regressor**. (O termo

covariável também é usado para x .) Os termos “variável dependente” e “variável independente” são usados com freqüência em econometria. Mas esteja consciente de que o nome “independente” não se refere aqui à noção estatística de independência entre variáveis aleatórias (veja Apêndice B, disponível na página do livro, no site www.thomsonlearning.com.br).

Os termos variáveis “explicada” e “explicativa” são, provavelmente, os mais descriptivos. “Resposta” e “controle” são muito usados nas ciências experimentais, em que a variável x está sob o controle do pesquisador. Não usaremos os termos “variável prevista” e “previsora”, embora algumas vezes você os veja no texto. Nossa terminologia para a regressão simples está resumida na Tabela 2.1.

Tabela 2.1

Terminologia para a Regressão Simples

y	x
Variável Dependente	Variável Independente
Variável Explicada	Variável Explicativa
Variável de Resposta	Variável de Controle
Variável Prevista	Variável Previsora
Regressando	Regressor

A variável u , chamada de **termo erro ou perturbação** da relação, representa outros fatores, além de x , que afetam y . Uma análise de regressão simples trata, efetivamente, todos os fatores, além de x , que afetam y como não-observados. Você pode pensar em u , convenientemente, como representando o “não-observado”.

A equação (2.1) também trata da questão da relação funcional entre y e x . Se os outros fatores em u são mantidos fixos, de modo que a variação em u é zero, $\Delta u = 0$, então x tem um efeito linear sobre y :

$$\Delta y = \beta_1 \Delta x \text{ se } \Delta u = 0. \quad (2.2)$$

Assim, a variação em y é, simplesmente, β_1 multiplicado pela variação em x . Isso significa que β_1 é o **parâmetro de inclinação** da relação entre y e x , mantendo fixos os outros fatores em u ; ele é de interesse fundamental em economia aplicada. O **parâmetro de intercepto** β_0 também tem seus usos, embora ele raramente seja central para uma análise.

EXEMPLO 2.1**(Produção de Soja e Fertilizantes)**

Suponha que a produção de soja seja determinada pelo modelo

$$\text{produção} = \beta_0 + \beta_1 \text{fertilizante} + u, \quad (2.3)$$

EXEMPLO 2.1 (continuação)

de modo que $y = \text{produção}$ e $x = \text{fertilizantes}$. O pesquisador agrícola está interessado no efeito dos fertilizantes sobre a produção, mantendo outros fatores fixos. Esse efeito é dado por β_1 . O termo erro u contém fatores como qualidade da terra, chuva etc. O coeficiente β_1 mede o efeito dos fertilizantes sobre a produção, mantendo outros fatores fixos: $\Delta\text{produção} = \beta_1\Delta\text{fertilizante}$.

EXEMPLO 2.2**(Uma Equação Simples do Salário)**

Um modelo que relaciona o salário de uma pessoa à educação observada e outros fatores não-observados é

$$\text{salárioh} = \beta_0 + \beta_1\text{educ} + u. \quad (2.4)$$

Se salárioh é medido em dólares por hora e educ corresponde a anos de educação formal, β_1 mede a variação no salário-hora dado um ano a mais de educação, mantendo todos os outros fatores fixos. Alguns desses fatores incluem experiência da força de trabalho, aptidão inata, permanência com o empregador atual, ética no trabalho e inumeráveis outras coisas.

A linearidade de (2.1) implica que uma variação de uma unidade em x tem o mesmo efeito sobre y , independentemente do valor inicial de x . Isso é irrealista para muitas aplicações econômicas. Por exemplo, no salário-educação, poderíamos querer considerar retornos crescentes: o próximo ano de educação teria, em relação ao anterior, um efeito *maior* sobre os salários. Veremos como considerar tais possibilidades na Seção 2.4.

A questão mais difícil é saber se o modelo (2.1) realmente nos permite tirar conclusões *ceteris paribus* sobre como x afeta y . Acabamos de ver, na equação (2.2), que β_1 mede, *de fato*, o efeito de x sobre y , mantendo todos os outros fatores (em u) fixos. Encerra-se com isso a questão da causalidade? Infelizmente, não. Como podemos esperar aprender algo, em geral, sobre o efeito *ceteris paribus* de x sobre y , mantendo outros fatores fixos, quando estamos ignorando todos aqueles outros fatores?

A Seção 2.5 mostrará que somos capazes de obter estimadores confiáveis de β_0 e β_1 de uma amostra aleatória de dados somente quando fazemos uma hipótese que restrinja a maneira de como o termo não-observável u está relacionado à variável explicativa x . Sem tal restrição, não seremos capazes de estimar o efeito *ceteris paribus*, β_1 . Como u e x são variáveis aleatórias, precisamos de um conceito baseado em probabilidade.

Antes de expormos a hipótese crucial de como x e u são relacionados, podemos sempre fazer uma hipótese sobre u . Se o intercepto β_0 está incluído na equação, nada se perde ao assumir que o valor médio de u na população é zero.

Matematicamente,

$$E(u) = 0. \quad (2.5)$$

A hipótese (2.5) não diz nada sobre a relação entre u e x ; ela simplesmente faz uma afirmação sobre a distribuição dos fatores não-observáveis na população. Usando os exemplos anteriores como ilustração, podemos ver que a hipótese (2.5) não é muito restritiva. No Exemplo 2.1, não perdemos nada ao normalizar os fatores não-observáveis que afetam a produção de soja, tal como a qualidade da terra, para ter uma média zero na população de todos os lotes cultivados. O mesmo é verdadeiro para os fatores não-observáveis do Exemplo 2.2. Sem perda de generalidade, podemos assumir que coisas como a média da aptidão são zero na população de todas as pessoas que trabalham. Se você não está convencido, pode trabalhar com o Problema 2.2 para ver que podemos sempre redefinir o intercepto na equação (2.1) para tornar (2.5) verdadeiro.

Agora, vamos voltar à hipótese crucial concernente à u e x como são relacionados. Uma medida natural de associação entre duas variáveis aleatórias é o *coeficiente de correlação*. (Veja Apêndice B, disponível no site da Thomson, para definição e propriedades.) Se u e x são *não-correlacionados*, logo, enquanto variáveis aleatórias, não são *linearmente* relacionados. Assumir que u e x são não-relacionados requer um caminho longo para definir o sentido em que u e x deveriam ser não-correlacionados na equação (2.1). Mas isso não vai longe o suficiente, pois a correlação mede somente a dependência linear entre u e x . A correlação tem uma característica algo contra-intuitiva: é possível que u seja não-correlacionado com x e seja correlacionado com funções de x , tal como x^2 . (Veja Seção B.4 para uma discussão adicional.) Essa possibilidade não é aceitável para muitos propósitos da regressão, visto que causa problemas para interpretar o modelo e para derivar propriedades estatísticas. Uma hipótese melhor envolve o *valor esperado de u , dado x* .

Como u e x são variáveis aleatórias, podemos definir a distribuição condicional de u , dado qualquer valor de x . Em particular, para qualquer x , podemos obter o valor esperado (ou médio) de u para aquela fatia da população descrita pelo valor de x . A hipótese crucial é que o valor médio de u não depende do valor de x . Podemos escrever isso como

$$E(u|x) = E(u) = 0, \quad (2.6)$$

em que a segunda igualdade resulta de (2.5). A primeira igualdade na equação (2.6) é a hipótese nova. Ela diz que, para qualquer valor de x , a média dos fatores não-observáveis é a mesma e, portanto, deve igualar-se ao valor médio de u na população. Quando combinamos a primeira igualdade da equação (2.6) com a hipótese (2.5), obtemos a **hipótese de média condicional zero**.

Vamos ver o que (2.6) acarreta ao exemplo do salário. Para simplificar a discussão, assuma que u seja o mesmo que aptidão inata. Então, (2.6) requer que o nível médio de aptidão seja o mesmo, independentemente dos anos de educação formal. Por exemplo, se $E(\text{aptidão}|8)$ representa a aptidão média para o grupo de todas as pessoas com oito anos de educação formal, e $E(\text{aptidão}|16)$ representa a aptidão média entre pessoas na população com 16 anos de educação formal, portanto (2.6) implica que essas médias devem ser as mesmas. De fato, o nível de aptidão média deve ser o mesmo para *todos* os níveis de educação. Se, por exemplo, entendemos que a aptidão média aumenta com os anos de educação formal, então (2.6) é falsa. (Isso aconteceria se, em média, pessoas com maior aptidão escolhessem tornar-se mais educadas.) Como não podemos observar aptidão inata, não temos um modo de saber se a aptidão média é ou não a mesma para todos os níveis de educação. Essa é uma questão que devemos resolver antes de aplicar a análise de regressão simples.

No exemplo dos fertilizantes, se as quantidades de fertilizantes são escolhidas independentemente de outras características dos lotes, então (2.6) se sustentará: a qualidade média da terra não dependerá da quantidade de fertilizantes. Entretanto, se mais fertilizantes forem usados em lotes de terra de melhor qualidade, então o valor esperado de u varia com o nível de fertilizantes, e (2.6) não se sustenta.

Suponha que a nota de um exame final (*nota*) dependa da freqüência às aulas (*freq*) e de fatores não-observados que afetam o desempenho dos estudantes (tal como a aptidão). Então:

$$\text{nota} = \beta_0 + \beta_1 \text{freq} + u. \quad (2.7)$$

Em que situação você esperaria que esse modelo satisfaça (2.6)?

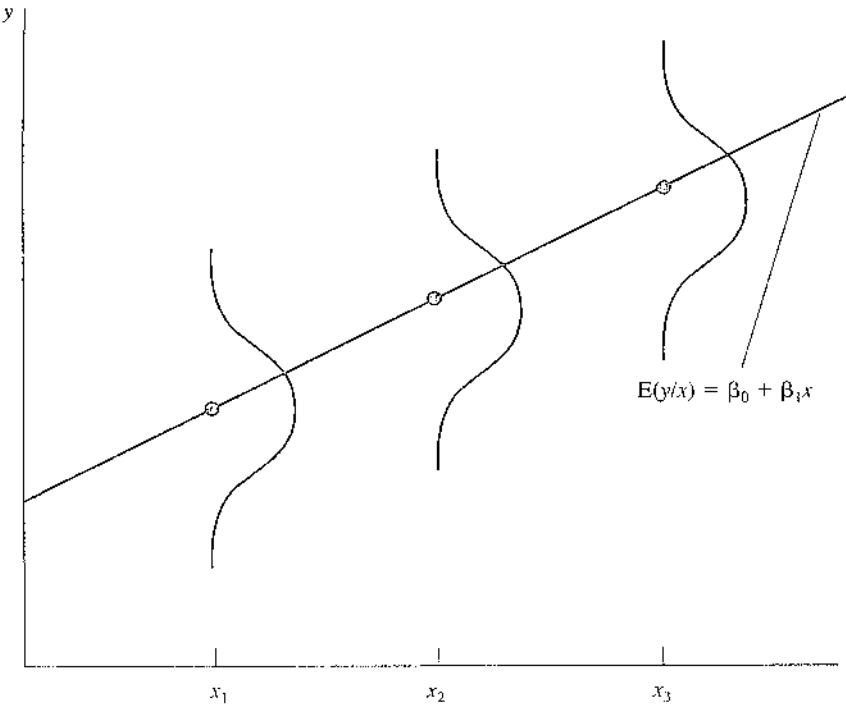
A hipótese (2.6) dá a β_1 outra interpretação que é, freqüentemente, útil. Considerando o valor esperado de (2.1) condicionado a x e usando $E(u|x) = 0$, obtém-se

$$E(y|x) = \beta_0 + \beta_1 x. \quad (2.8)$$

A equação (2.8) mostra que a **função de regressão populacional** (FRP), $E(y|x)$, é uma função linear de x . A linearidade significa que o aumento de uma unidade em x faz com que o *valor esperado* de y varie segundo a magnitude de β_1 . Para qualquer valor dado de x , a distribuição de y está centrada ao redor de $E(y|x)$, como ilustrado na Figura 2.1.

Figura 2.1

$E(y|x)$ como função linear de x .



Quando (2.6) é verdadeira, é útil dividir y em dois componentes. A parte $\beta_0 + \beta_1x$ é algumas vezes chamada a *parte sistemática* de y — isto é, a parte de y explicada por x —, e u é chamado a *parte não-sistemática*, ou a parte de y não explicada por x . Usaremos a hipótese (2.6) na próxima seção para encontrar as estimativas de β_0 e β_1 . Essa hipótese também é crucial para a análise estatística na Seção 2.5.

2.2 DERIVAÇÃO DAS ESTIMATIVAS DE MÍNIMOS QUADRADOS ORDINÁRIOS

Agora que discutimos os ingredientes básicos do modelo de regressão simples, trataremos da importante questão de como estimar os parâmetros β_0 e β_1 da equação (2.1). Para tanto, necessitamos de uma amostra da população. Vamos considerar $\{(x_i, y_i): i=1, \dots, n\}$ como uma amostra aleatória de tamanho n da população. Visto que esses dados vêm de (2.1), podemos escrever

$$y_i = \beta_0 + \beta_1x_i + u_i, \quad (2.9)$$

para cada i . Aqui, u_i é o termo erro para a observação i , uma vez que ele contém todos os fatores, além de x_i , que afetam y_i .

Como um exemplo, x_i poderia ser a renda anual e y_i , a poupança anual para a família i durante um determinado ano. Se coletarmos dados de 15 famílias, então $n = 15$. Um gráfico de tal conjunto de dados é dado pela Figura 2.2, juntamente com a função de regressão populacional (necessariamente fictícia).

Devemos decidir como usar esses dados, a fim de obter estimativas do intercepto e da inclinação na regressão populacional da poupança sobre a renda.

Há muitas maneiras de colocar em prática o seguinte procedimento de estimação. Usaremos (2.5) e uma importante implicação da hipótese (2.6): na população, u tem média zero e é não-correlacionado com x . Portanto, vemos que u tem valor esperado zero e que a covariância entre x e u é zero:

$$E(u) = 0 \quad (2.10)$$

e

$$\text{Cov}(x, u) = E(xu) = 0, \quad (2.11)$$

onde a primeira igualdade em (2.11) resulta de (2.10). (Veja Seção B.4 do Apêndice B, disponível no site da Thomson, para definição e propriedades da covariância.) Em termos das variáveis observáveis x e y e dos parâmetros desconhecidos β_0 e β_1 , as equações (2.10) e (2.11) podem ser escritas como

$$E(y - \beta_0 - \beta_1x) = 0 \quad (2.12)$$

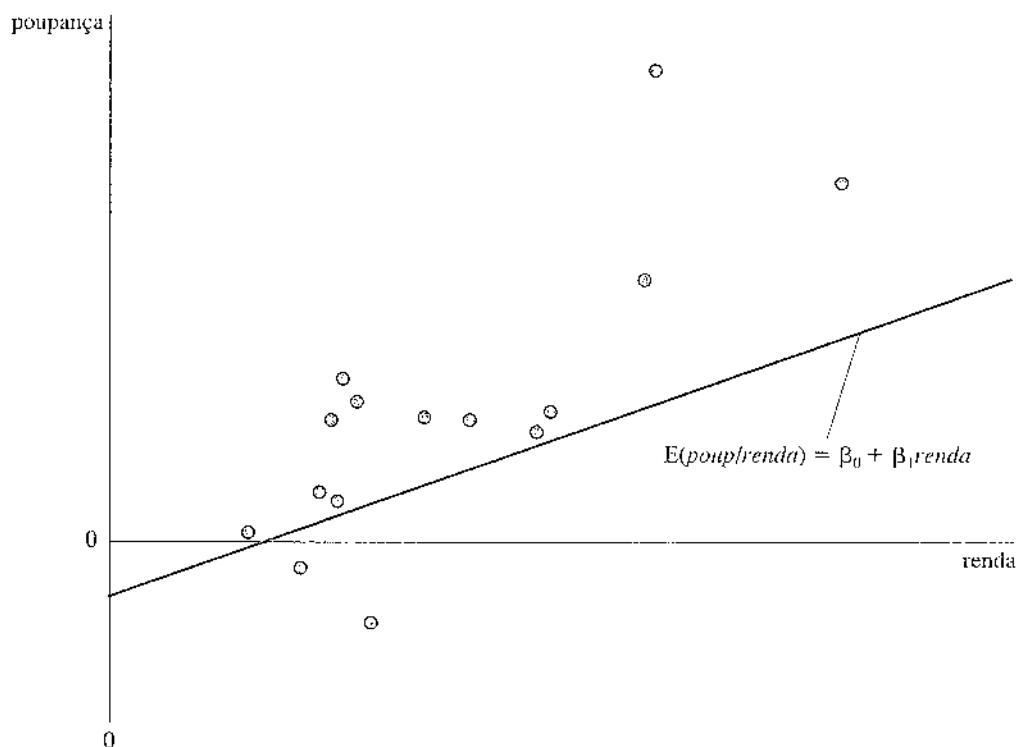
e

$$E[x(y - \beta_0 - \beta_1x)] = 0, \quad (2.13)$$

respectivamente. As equações (2.12) e (2.13) implicam duas restrições sobre a distribuição de probabilidade conjunta de (x, y) na população. Como há dois parâmetros desconhecidos para estimar, poderíamos esperar que as equações (2.12) e (2.13) pudessem ser usadas para obter bons estimadores de β_0 e β_1 . De fato, elas podem ser usadas. Dada uma amostra de dados, escolhemos as estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ para resolver as equivalências *amostrais* de (2.12) e (2.13):

Figura 2.2

Gráfico da dispersão de poupança e renda de 15 famílias e a regressão populacional $E(poup|renda) = \beta_0 + \beta_1 renda$.



$$n^{-1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad (2.14)$$

e

$$n^{-1} \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0. \quad (2.15)$$

Esse é um exemplo da abordagem do *método dos momentos* para a estimação. (Veja a Seção C.4 do Apêndice C, disponível no site da Thomson, para uma discussão das diferentes abordagens de estimação.) Essas equações podem ser resolvidas para $\hat{\beta}_0$ e $\hat{\beta}_1$.

Usando as propriedades básicas do operador somatório a partir do Apêndice A (disponível no site da Thomson), a equação (2.14) pode ser escrita como

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}, \quad (2.16)$$

em que $\bar{y} = n^{-1} \sum_{i=1}^n y_i$ é a média amostral de y_i , e igualmente para \bar{x} . Essa equação nos permite escrever $\hat{\beta}_0$ em termos de $\hat{\beta}_1$, \bar{y} e \bar{x} :

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}. \quad (2.17)$$

Portanto, uma vez que temos a estimativa de inclinação $\hat{\beta}_1$, obtém-se diretamente a estimativa de intercepto $\hat{\beta}_0$, dados \bar{y} e \bar{x} .

Suprimindo o n^{-1} em (2.15) (já que ele não afeta a solução) e inserindo (2.17) em (2.15), obtemos

$$\sum_{i=1}^n x_i [y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i] = 0$$

a qual, após rearranjo, pode ser escrita

$$\sum_{i=1}^n x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^n x_i (x_i - \bar{x}).$$

Das propriedades básicas do operador somatório [veja (A.7) e (A.8) disponível no site da Thomson],

$$\sum_{i=1}^n x_i (x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2 \text{ e } \sum_{i=1}^n x_i (y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

Portanto, desde que

$$\sum_{i=1}^n (x_i - \bar{x})^2 > 0, \quad (2.18)$$

a inclinação estimada é

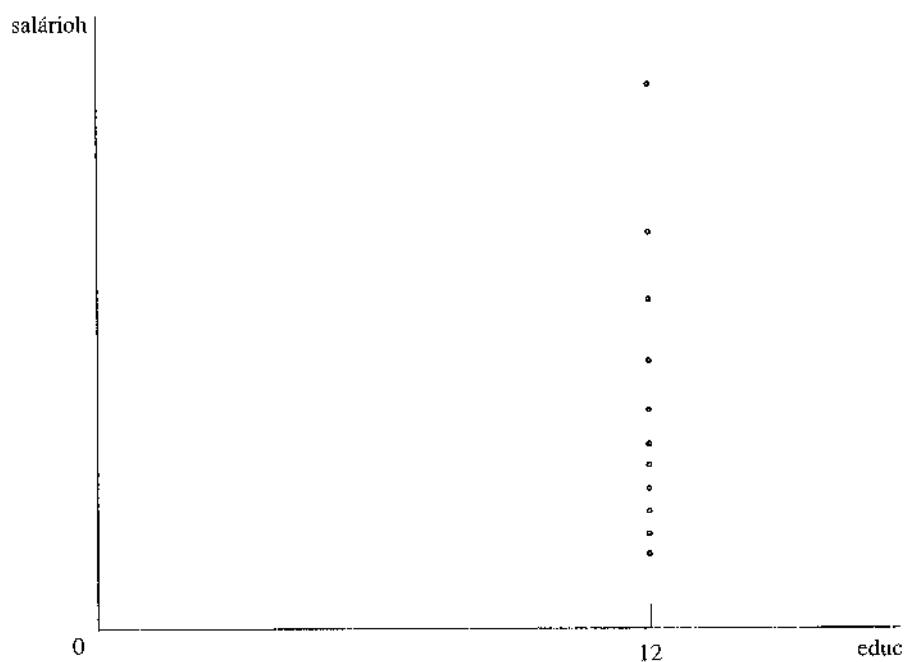
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.19)$$

A equação (2.19) é simplesmente a covariância amostral entre x e y , dividida pela variância amostral de x . (Veja Apêndice C, disponível no site da Thomson. Dividir tanto o numerador como o denominador por $n - 1$ não altera o resultado.) Isso faz sentido, pois β_1 é igual à covariância populacional dividida pela variância de x quando $E(u) = 0$ e $Cov(x, u) = 0$. Uma implicação imediata é que se x e y são positivamente correlacionados na amostra, então $\hat{\beta}_1$ é positivo; se x e y são negativamente correlacionados, então $\hat{\beta}_1$ é negativo.

Embora o método para obter (2.17) e (2.19) decorra de (2.6), a única hipótese necessária para se calcular as estimativas para uma amostra particular é (2.18). Mas essa raramente é uma hipótese: (2.18) é verdadeira sempre que os x_i na amostra não são todos iguais a um mesmo valor. Se (2.18) não se sustentar, então fomos infelizes em obter nossa amostra da população, ou não especificamos um problema interessante (x não varia na população). Por exemplo, se $y = \text{salárioh}$ e $x = \text{educ}$, então (2.18) não se mantém se todos na amostra têm a mesma quantidade de anos de educação formal. (Por exemplo, se todos têm o equivalente ao ensino médio concluído. Veja a Figura 2.3.) Se apenas uma pessoa tem uma quantidade diferente de anos de educação formal, então (2.18) se sustenta, e as estimativas de MQO podem ser calculadas.

Figura 2.3

Gráfico da dispersão de salários e educação, quando $\text{educ}_i = 12$ para todo i .



As estimativas dadas em (2.17) e (2.19) são chamadas de estimativas de **mínimos quadrados ordinários (MQO)** de β_0 e β_1 . Para justificar esse nome, defina, para qualquer $\hat{\beta}_0$ e $\hat{\beta}_1$, um valor estimado para y quando $x = x_i$, tal como

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i, \quad (2.20)$$

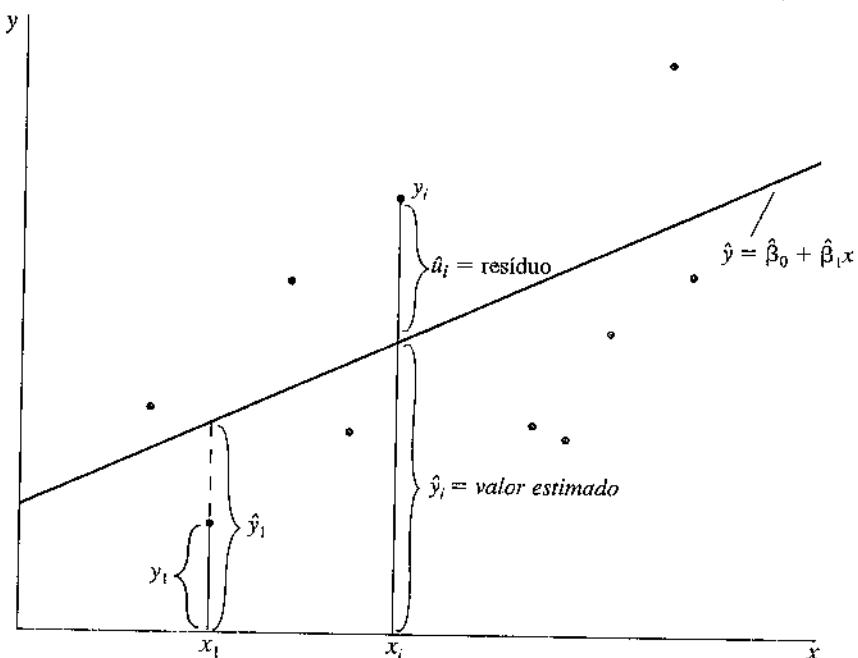
para o intercepto e a inclinação dados. Esse é o valor que prevemos para y quando $x = x_i$. Há um valor estimado para cada observação na amostra. O **resíduo** para a observação i é a diferença entre o valor verdadeiro de y_i e seu valor estimado:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i. \quad (2.21)$$

Novamente, há n desses resíduos. [Eles *não* são iguais aos erros em (2.9), um ponto ao qual retornaremos na Seção 2.5.] Os valores estimados e os resíduos estão indicados na Figura 2.4.

Figura 2.4

Valores estimados e resíduos.



Agora, suponha que escolhemos $\hat{\beta}_0$ e $\hat{\beta}_1$ com a finalidade de tornar a **soma dos resíduos quadrados**,

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2, \quad (2.22)$$

tão pequena quanto possível. O apêndice deste capítulo mostra que as condições necessárias para $(\hat{\beta}_0, \hat{\beta}_1)$ minimizarem (2.22) são dadas exatamente pelas equações (2.14) e (2.15), sem n^{-1} . As equações (2.14) e (2.15) são freqüentemente chamadas de **condições de primeira ordem** para as estimativas de MQO, um termo que vem da otimização utilizada em cálculo (veja o Apêndice A, disponível no site da Thomson). De nossos cálculos anteriores, sabemos que as soluções para as condições de primeira

ordem de MQO são dadas por (2.17) e (2.19). O nome “mínimos quadrados ordinários” vem do fato de que essas estimativas minimizam a soma dos resíduos quadrados.

Quando vemos o método de mínimos quadrados ordinários como um método que minimiza a soma dos resíduos quadrados, é natural perguntar: por que não minimizar alguma outra função dos resíduos, como o valor absoluto dos resíduos? De fato, como discutiremos brevemente na Seção 9.4, minimizar a soma dos valores absolutos dos resíduos é, algumas vezes, muito útil. Mas esse procedimento também tem suas desvantagens. Primeiro, não podemos obter fórmulas para os estimadores resultantes; dado um conjunto de dados, as estimativas devem ser obtidas por rotinas de otimização numérica. Em consequência, a teoria estatística para estimadores que minimizam a soma dos resíduos absolutos é muito complicada. Minimizar outras funções dos resíduos, como a soma de cada resíduo elevado à quarta potência, tem desvantagens similares. (Nunca deveríamos escolher nossos estimadores para minimizar, por exemplo, a soma dos próprios resíduos, pois resíduos grandes em magnitude e com sinais opostos tendem a se cancelar.) Com o método MQO, seremos capazes de derivar, de modo relativamente fácil, inexistência de viés, consistência e outras importantes propriedades estatísticas. E mais, como as equações (2.13) e (2.14) sugerem, e como veremos na Seção 2.5, o método MQO é adequado para estimar os parâmetros que aparecem na função de média condicional (2.8).

Uma vez determinados os estimadores de intercepto e inclinação de MQO, construímos a **reta de regressão de MQO**:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x, \quad (2.23)$$

em que $\hat{\beta}_0$ e $\hat{\beta}_1$ foram obtidos ao usar as equações (2.17) e (2.19). A notação \hat{y} — leia-se “y chapéu” — enfatiza que os valores previstos da equação (2.23) são estimativas. O intercepto $\hat{\beta}_0$ é o valor previsto de y quando $x = 0$, embora, em alguns casos, não faça sentido considerar $x = 0$. Nestas situações, $\hat{\beta}_0$ não é, por si mesmo, muito interessante. Ao usar (2.23) para calcular os valores previstos de y para vários valores de x , devemos considerar o intercepto nos cálculos. A equação (2.23) é também chamada **função de regressão amostral (FRA)**, pois ela é a versão estimada da função de regressão populacional $E(y|x) = \beta_0 + \beta_1 x$. É importante lembrar que a FRA é algo fixo, porém desconhecido, na população. Como a FRA é obtida para uma dada amostra de dados, uma amostra nova gerará uma inclinação e um intercepto diferentes na equação (2.23).

Em muitos casos, a estimativa do coeficiente de inclinação, que podemos escrever como

$$\hat{\beta}_1 = \Delta \hat{y} / \Delta x, \quad (2.24)$$

é de interesse fundamental. Ela nos diz o quanto varia \hat{y} quando x aumenta em uma unidade. Equivalente,

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x, \quad (2.25)$$

de modo que, dada qualquer variação em x (seja positiva ou negativa), podemos calcular a variação prevista em y .

Agora, vamos apresentar vários exemplos de regressões simples obtidas de dados reais. Em outras palavras, vamos encontrar as estimativas de intercepto e de inclinação a partir das equações (2.17) e (2.19). Como esses exemplos compreendem muitas observações, os cálculos foram feitos usando pro-

gramas econôméticos. Neste ponto, não se preocupe muito em interpretar as regressões; elas não estão, necessariamente, revelando uma relação causal. Até aqui, não dissemos nada sobre as propriedades estatísticas do método MQO. Na Seção 2.5, consideraremos as propriedades depois de impormos explicitamente hipóteses sobre a equação do modelo populacional (2.1).

EXEMPLO 2.3

{Salários de Diretores Executivos e Retornos de Ações}

Para a população de diretores executivos, seja y o salário anual (*salário*) em milhares de dólares. Assim, $y = 856,3$ indica um salário anual de \$ 856.300, e $y = 1.452,6$ indica um salário de \$ 1.452.600. Seja x o retorno médio da ação sobre o patrimônio (*rma*), dos três anos anteriores, da empresa do diretor executivo. (O retorno da ação sobre o patrimônio é definido em termos de renda líquida, como uma percentagem do patrimônio comum.) Por exemplo, se $rma = 10$, então o retorno médio da ação sobre o patrimônio é de 10%.

Para estudar a relação entre essa medida do desempenho das empresas e a remuneração dos seus diretores executivos, postulamos o modelo simples

$$\text{salário} = \beta_0 + \beta_1 rma + u.$$

O parâmetro de inclinação β_1 mede a variação no salário anual, em milhares de dólares, quando o retorno da ação aumenta em um ponto percentual. Como um *rma* mais elevado é melhor para a empresa, esperamos que $\beta_1 > 0$.

O conjunto de dados do arquivo CEOSAL1.RAW contém informações sobre 209 diretores executivos para o ano de 1990; esses dados foram obtidos da revista *Business Week* (6.5.91). Na amostra, o salário médio anual é \$ 1.281.120; sendo que o menor e o maior são \$ 223.000 e \$ 14.822.000, respectivamente. O retorno médio das ações para os anos 1988, 1989 e 1990 é de 17,18%, sendo que os valores menor e maior são de 0,5 e 56,3%, respectivamente.

Usando os dados do arquivo CEOSAL1.RAW, a reta de regressão de MQO que relaciona *salário* a *rma* é

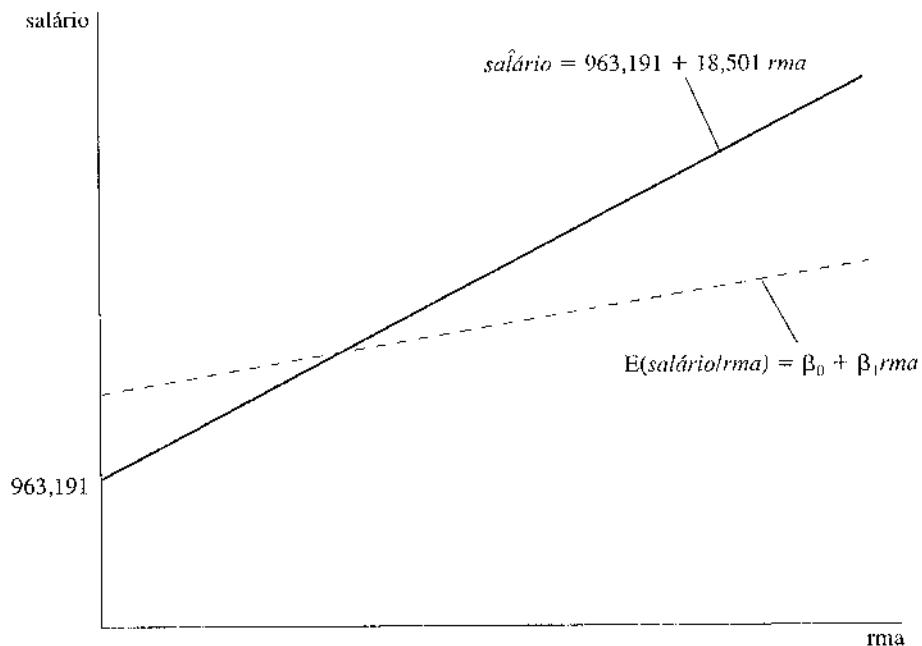
$$\text{salário} = 963,191 + 18,501 rma, \quad (2.26)$$

na qual as estimativas de intercepto e de inclinação foram arredondadas em três casas decimais; usamos "salário chapéu" para indicar que essa é uma equação estimada. Como interpretamos a equação? Primeiro, se o retorno da ação é zero, $rma = 0$, então o salário previsto é o intercepto, 963,191, que é igual a \$ 963.191, visto que *salário* é mensurado em milhares. Em seguida, podemos escrever a variação prevista no salário como uma função da variação em *rma*: $\Delta \text{salário} = 18,501(\Delta rma)$. Isso significa que se o retorno da ação aumenta um ponto percentual, $\Delta rma = 1$, então espera-se que *salário* variará cerca de 18,5, ou \$ 18.500. Como (2.26) é uma equação linear, esse valor é a variação estimada independentemente do salário inicial.

Podemos facilmente usar (2.26) para comparar salários previstos para valores diferentes de *rma*. Suponha $rma = 30$. Portanto, $\text{salário} = 963,191 + 18,501(30) = 1.518,221$, o que está pouco acima de \$ 1,5 milhão. Entretanto, isso não significa que um determinado diretor executivo, cuja empresa tenha um *rma* = 30, ganhe \$ 1.518.221. Muitos outros fatores afetam o salário. Essa é somente a nossa previsão a partir da reta de regressão de MQO de (2.26). A reta estimada está representada na Figura 2.5, juntamente com a função de regressão populacional $E(\text{salário}|rma)$. Nunca conheceremos a FRP; assim, não podemos dizer o quanto próxima a FRA está da FRP. Outra amostra de dados levará a uma reta de regressão diferente, que pode estar, ou não, mais próxima da reta de regressão populacional.

Figura 2.5

A reta de regressão de MQO $\hat{salário} = 963,191 + 18,501 rma$ e a função de regressão populacional (desconhecida).

**EXEMPLO 2.4****(Salários e Educação)**

Para a população de pessoas na força de trabalho em 1976, seja $y = salário_{ih}$, em que $salário_{ih}$ é mensurado em dólares por hora. Assim, para uma determinada pessoa, se $salário_{ih} = 6,75$, o salário-hora é \$ 6,75. Vamos chamar anos de escolaridade formal de $x = educ$; por exemplo, $educ = 12$ corresponde ao ensino médio completo (nos Estados Unidos). O salário horário médio na amostra é \$ 5,90, o que equivale, de acordo com o índice de preços ao consumidor dos Estados Unidos, a \$ 16,64 em dólares de 1997.

Usando os dados do arquivo WAGE1.RAW, em que $n = 526$ indivíduos, obtemos a seguinte reta de regressão de MQO (ou função de regressão amostral):

$$\hat{salário} = -0,90 + 0,54 educ. \quad \{2.27\}$$

Devemos interpretar essa equação com cautela. O intercepto de $-0,90$ significa, literalmente, que uma pessoa sem nenhuma educação formal tem um salário-hora previsto de -90 centavos de dólar por hora. Isso, evidentemente, é tolice. Ocorre que apenas 18 pessoas na amostra de 526 têm menos que oito anos de educação formal. Conseqüentemente, não é surpreendente que a reta de regressão não faça boas previsões para

EXEMPLO 2.4 (continuação)

níveis de educação formal muito baixos. Para uma pessoa com oito anos de educação formal, o salário previsto é $\text{salário} = -0,90 + 0,54(8) = 3,42$, ou \$ 3,42 por hora (em dólares de 1976).

A inclinação estimada em (2.27) implica que um ano a mais de educação formal aumenta o salário horário em 54 centavos de dólar por hora. Portanto, quatro anos a mais de educação formal aumentam o salário horário previsto em $4(0,54) = 2,16$, ou \$ 2,16 por hora. Esses efeitos são razoavelmente grandes. Devido à natureza linear de (2.27), outro ano de educação formal aumenta o salário na mesma quantidade, independentemente do nível inicial de educação. Na Seção 2.4, discutiremos alguns métodos que levam em consideração efeitos marginais não-constantes de nossas variáveis explicativas.

O salário horário estimado em (2.27), quando $\text{educ} = 8$, é \$ 3,42, em dólares de 1976. Qual é esse valor em dólares de 1997? (Sugestão: você tem informação suficiente, no Exemplo 2.4, para responder a essa questão.)

EXEMPLO 2.5**(Resultados Eleitorais e Gastos de Campanha)**

O arquivo VOTE1.RAW contém dados sobre resultados eleitorais e gastos de campanha de 173 disputas entre dois partidos, para a *House of Representatives* dos Estados Unidos (equivalente a uma câmara federal). Há dois candidatos em cada disputa: A e B. Seja $votoA$ a percentagem de votos recebida pelo Candidato A e $partA$ a percentagem dos gastos totais de campanha que cabem ao Candidato A. Muitos outros fatores além de $partA$ afetam o resultado eleitoral (incluindo a qualidade dos candidatos e os valores absolutos dos gastos de A e B). No entanto, podemos estimar um modelo de regressão simples para descobrir se gastar mais do que o concorrente implica uma percentagem maior de votos.

A equação estimada usando as 173 observações é

$$votoA = 26,81 + 0,464 partA. \quad (2.28)$$

Isso significa que, se a parte dos gastos do Candidato A aumenta em um ponto percentual, o Candidato A recebe quase meio ponto percentual (0,464) a mais da votação total. Não fica claro se isso revela ou não um efeito causal, mas isso é crível. Se $partA = 50$, prevê-se que $votoA$ será cerca de 50, ou metade da votação.

Em alguns casos, a análise de regressão não é usada para determinar a causalidade, mas para simplesmente observar se duas variáveis são positiva ou negativamente relacionadas, de modo muito parecido com uma análise padrão de correlação. Um exemplo disso ocorre no Problema 2.12, que pede que você use os dados de Biddle e Hamermesh (1990) referentes ao tempo que se gasta dormindo e trabalhando a fim de investigar a relação entre esses dois fatores.

No Exemplo 2.5, qual é a votação prevista para o Candidato A se $partA = 60$ (que significa 60%)? A resposta parece razoável?

Uma Nota sobre Terminologia

Em muitos casos, indicaremos a estimativa de uma relação através de MQO ao escrever uma equação como (2.26), (2.27) ou (2.28). Algumas vezes, por motivo de brevidade, é útil indicar que uma regressão de MQO foi estimada sem realmente escrever a equação. Freqüentemente, indicaremos que a equação (2.23) foi obtida por MQO ao dizer que nós *rodamos* a regressão de

$$y \text{ sobre } x, \quad (2.29)$$

ou simplesmente que *regredimos* y sobre x . As posições de y e x em (2.29) indicam qual é a variável dependente e qual é a variável independente: sempre regredimos a variável dependente sobre a variável independente. Para aplicações específicas, substituiremos y e x por seus nomes. Assim, para obter (2.26), regredimos *salário* sobre *rma*, ou, para obter (2.28), regredimos *votoA* sobre *partA*.

Ao usarmos essa terminologia em (2.29), sempre estaremos dizendo que planejamos estimar o intercepto, $\hat{\beta}_0$, juntamente com o coeficiente de inclinação, $\hat{\beta}_1$. Esse caso é apropriado para a maioria das aplicações. Ocasionalmente, podemos querer estimar a relação entre y e x *assumindo* que o intercepto é zero (de modo que $x = 0$ implica $\hat{y} = 0$); cobriremos esse caso, brevemente, na Seção 2.6. A não ser que seja explicitamente dito de outro modo, sempre estimaremos um intercepto juntamente com uma inclinação.

2.3 MECÂNICA DO MÉTODO MQO

Nesta seção, cobriremos algumas propriedades algébricas da reta de regressão de MQO estimada. Talvez, a melhor maneira de pensar nessas propriedades é perceber que elas são características de MQO para uma determinada amostra de dados. Elas podem ser contrastadas com as propriedades *estatísticas* de MQO, o que requer a derivação das características das distribuições amostrais dos estimadores. Discutiremos as propriedades estatísticas na Seção 2.5.

Muitas das propriedades algébricas que derivaremos parecerão triviais. No entanto, ter uma compreensão dessas propriedades ajuda-nos a entender o que acontece com as estimativas de MQO e estatísticas relacionadas quando os dados são manipulados de determinadas maneiras, como quando variam as unidades de medida das variáveis dependentes e independentes.

Valores Estimados e Resíduos

Assumimos que as estimativas de intercepto e de inclinação, $\hat{\beta}_0$ e $\hat{\beta}_1$, foram obtidas de uma dada amostra de dados. Dados $\hat{\beta}_0$ e $\hat{\beta}_1$, podemos obter o valor estimado \hat{y}_i para cada observação. [Isso é dado pela equação (2.20).] Por definição, cada valor estimado de \hat{y}_i está sobre a reta de regressão de MQO. O resíduo de MQO associado a cada observação i , \hat{u}_i , é a diferença entre y_i e seu valor estimado, como dado na equação (2.21). Se \hat{u}_i é positivo, a reta subestima y_i ; se \hat{u}_i é negativo, a reta superestima y_i . O caso ideal para a observação i é quando $\hat{u}_i = 0$, mas na maior parte dos casos *todos* os resíduos são

diferentes de zero. Em outras palavras, nenhum dos pontos dos dados deve, realmente, estar sobre a reta de MQO.

EXEMPLO 2.6

(Salário de Diretores Executivos e Retornos de Ações)

A Tabela 2.2 contém uma lista das 15 primeiras observações do conjunto de dados dos salários dos diretores executivos, juntamente com os valores estimados, chamados de *salchapéu*, e os resíduos, estimados de *uchapéu*.

Tabela 2.2

Valores Estimados e Resíduos dos 15 Primeiros Diretores Executivos

<i>nobsd</i>	<i>rma</i>	<i>salário</i>	<i>salchapéu</i>	<i>uchapéu</i>
1	14,1	1.095	1.224,058	-129,0581
2	10,9	1.001	1.164,854	-163,8542
3	23,5	1.122	1.397,969	-275,9692
4	5,9	578	1.072,348	-494,3484
5	13,8	1.368	1.218,508	149,4923
6	20,0	1.145	1.333,215	-188,2151
7	16,4	1.078	1.266,611	-188,6108
8	16,3	1.094	1.264,761	-170,7606
9	10,5	1.237	1.157,454	79,54626
10	26,3	833	1.449,773	-616,7726
11	25,9	567	1.442,372	-875,3721
12	26,8	933	1.459,023	-526,0231
13	14,8	1.339	1.237,009	101,9911
14	22,3	937	1.375,768	-438,7678
15	56,3	2.011	2.004,808	6,191895

Os quatro primeiros diretores executivos têm salários menores do que os previstos a partir da reta de regressão de MQO (2.26); em outras palavras, dado somente o *rma* da empresa, esses diretores executivos ganham menos do que prevemos. Como pode ser visto dos *uchapéus* positivos, o quinto diretor executivo ganha mais do que prevemos a partir da reta de regressão de MQO.

Propriedades Algébricas das Estatísticas de MQO

Há várias propriedades algébricas úteis das estimativas de MQO e das estatísticas a elas associadas. Vamos discutir as três mais importantes.

(1) A soma, e portanto a média amostral, dos resíduos de MQO, é zero. Matematicamente,

$$\sum_{i=1}^n \hat{u}_i = 0. \quad (2.30)$$

Essa propriedade não precisa de prova; ela resulta, imediatamente, da condição de primeira ordem de MQO (2.14), quando lembramos que os resíduos são definidos por $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$. Em outras palavras, as estimativas de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ são *escolhidas* para fazer com que a soma dos resíduos seja zero (para qualquer conjunto de dados). Isso não diz nada sobre o resíduo de qualquer observação i em particular.

(2) A covariância amostral entre os regressores e os resíduos de MQO é zero. Isso resulta da condição de primeira ordem (2.15), que pode ser escrita em termos dos resíduos, como

$$\sum_{i=1}^n x_i \hat{u}_i = 0. \quad (2.31)$$

A média amostral dos resíduos de MQO é zero, de modo que o lado esquerdo de (2.31) é proporcional à covariância amostral entre x_i e \hat{u}_i .

(3) O ponto (\bar{x}, \bar{y}) sempre está sobre a reta de regressão de MQO. Em outras palavras, se considerarmos a equação (2.23) e inserirmos \bar{x} no lugar de x , então o valor estimado é \bar{y} . Isso é exatamente o que a equação (2.16) nos mostrou.

EXEMPLO 2.7

(Salários e Educação)

Para os dados do arquivo em WAGE1.RAW, o salário-hora médio da amostra é 5,90, arredondado para duas casas decimais, e a educação formal média (medida em anos) é 12,56. Se inserirmos $educ = 12,56$ na reta de regressão de MQO (2.27), obtemos $\hat{salário}_{ih} = -0,90 + 0,54(12,56) = 5,8824$, igual a 5,9 quando arredondamos para uma casa decimal. A razão de esses números não serem exatamente os mesmos é que nós arredondamos o salário-hora e os anos de educação formal médios, assim como as estimativas de intercepto e de inclinação. Se, inicialmente, não tivéssemos arredondado nenhum dos valores, obteríamos respostas mais aproximadas, mas essa prática teria pouco efeito útil.

Ao escrever cada y_i como o seu valor estimado mais seu resíduo, temos outro modo de interpretar uma regressão de MQO. Para cada i , escreva

$$y_i = \hat{y}_i + \hat{u}_i \quad (2.32)$$

Da propriedade (1), a média dos resíduos é zero; equivalentemente, a média amostral dos valores estimados, \hat{y}_i , é a mesma que a média amostral de y_i , ou $\bar{\hat{y}} = \bar{y}$. Além disso, as propriedades (1) e (2)

podem ser usadas para mostrar que a covariância amostral entre \hat{y}_i e \hat{u}_i é zero. Portanto, podemos ver MQO como um método que decompõe cada y_i em duas partes: um valor estimado e um resíduo. Os valores estimados e os resíduos são não-correlacionados na amostra.

Defina a **soma dos quadrados total (SQT)**, a **soma dos quadrados explicada (SQE)** e a **soma dos quadrados dos resíduos (SQR)** (também conhecida como a soma dos resíduos quadrados), como a seguir:

$$\text{SQT} = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (2.33)$$

$$\text{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \quad (2.34)$$

$$\text{SQR} = \sum_{i=1}^n \hat{u}_i^2. \quad (2.35)$$

SQT é uma medida da variação amostral total em y_i ; isto é, ela mede o quanto dispersos estão os y_i na amostra. Se dividirmos SQT por $n - 1$, obteremos a variância amostral de y , como discutido no Apêndice C; no site da Thomson. Semelhantemente, SQE mede a variação amostral em \hat{y}_i (em que usamos o fato de que $\bar{\hat{y}} = \bar{y}$), e SQR mede a variação amostral em \hat{u}_i . A variação total em y pode sempre ser expressa como a soma da variação explicada e da variação não-explicada SQR. Assim,

$$\text{SQT} = \text{SQE} + \text{SQR}. \quad (2.36)$$

Provar (2.36) não é difícil; mas requer o uso de todas as propriedades do operador somatório apresentadas no Apêndice A, no site da Thomson. Escreva

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n [(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^n [\hat{u}_i + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 + 2 \sum_{i=1}^n \hat{u}_i(\hat{y}_i - \bar{y}) + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ &= \text{SQR} + 2 \sum_{i=1}^n \hat{u}_i(\hat{y}_i - \bar{y}) + \text{SQE}. \end{aligned}$$

Agora, (2.36) é válida se mostrarmos que

$$\sum_{i=1}^n \hat{u}_i(\hat{y}_i - \bar{y}) = 0. \quad (2.37)$$

Mas já dissemos que a covariância amostral entre os resíduos e os valores estimados é zero, e essa covariância é justamente a equação (2.37) dividida por $n - 1$. Conseqüentemente, confirmamos (2.36).

Algumas palavras de precaução sobre SQT, SQE e SQR devem ser mencionadas. Não há concordância uniforme sobre os nomes e abreviações das três quantidades definidas nas equações (2.33), (2.34) e (2.35). A soma dos quadrados total é chamada SQT ou STQ, de modo que aqui não há grandes confusões. Infelizmente, a soma dos quadrados explicada é, às vezes, chamada de “soma dos quadrados da regressão”. Se a esse termo é dado sua abreviação natural, ele pode ser facilmente confundido com o termo “soma dos quadrados dos resíduos”. Alguns programas econométricos referem-se à soma dos quadrados explicada como “soma dos quadrados do modelo”.

Para tornar as coisas ainda piores, a soma dos quadrados dos resíduos é freqüentemente chamada de “soma dos quadrados dos erros”. Esse termo é um tanto inadeguado, pois, como veremos na Seção 2.5, os erros e os resíduos são quantidades diferentes. Assim, sempre chamaremos (2.35) de soma dos quadrados dos resíduos ou soma dos resíduos quadrados. Preferimos usar a abreviação SQR para representar a soma dos resíduos quadrados, pois ela é mais comum nos programas econométricos.

Grau de Ajuste

Até aqui, não apresentamos uma maneira de mensurar o quanto bem a variável explicativa ou independente, x , explica a variável dependente, y . Muitas vezes, é útil calcular um número que resume o quanto bem a reta de regressão de MQO se ajusta aos dados. Na discussão seguinte, lembre-se de que assumimos estimar o intercepto com a inclinação.

Ao assumirmos que a soma dos quadrados total, SQT, não é igual a zero — o que é verdadeiro, a não ser no evento muito improvável de todos os y_i serem iguais a um mesmo valor —, podemos dividir (2.36) por SQT para obter $1 = \text{SQE}/\text{SQT} + \text{SQR}/\text{SQT}$. O ***R*-quadrado** da regressão, algumas vezes chamado coeficiente de determinação, é definido como

$$R^2 = \text{SQE}/\text{SQT} = 1 - \text{SQR}/\text{SQT}. \quad (2.38)$$

R^2 é a razão entre a variação explicada e a variação total; assim, ele é interpretado como a *fração da variação amostral em y que é explicada por x*. A segunda equação em (2.38) fornece outra maneira de calcular R^2 .

De (2.36), o valor de R^2 está sempre entre zero e um, visto que SQE não pode ser maior que SQT. Quando interpretamos R^2 , usualmente o multiplicamos por 100 para transformá-lo em percentual: $100 \cdot R^2$ é a *percentagem da variação amostral em y que é explicada por x*.

Se todos os pontos dos dados estiverem sobre a mesma reta, MQO fornece um ajuste perfeito aos dados. Nesse caso, $R^2 = 1$. Um valor de R^2 quase igual a zero indica um ajuste ruim da reta de MQO: muito pouco da variação em y_i é capturado pela variação em \hat{y}_i (que está sobre a reta de regressão de MQO). De fato, pode ser mostrado que R^2 é igual ao *quadrado* do coeficiente de correlação amostral entre y_i e \hat{y}_i . É daí que vem o termo “*R-quadrado*”. (A letra *R* era, tradicionalmente, usada para denominar uma estimativa do coeficiente de correlação populacional, e seu uso sobreviveu na análise de regressão.)

EXEMPLO 2.8**(Salário de Diretores Executivos e Retornos de Ações)**

Na regressão de salários de diretores executivos, estimamos a seguinte equação:

$$\hat{\text{salário}} = 963,191 + 18,501 \text{ rma} \quad (2.39)$$

$$n = 209, R^2 = 0,0132.$$

Por motivos de clareza, reproduzimos a reta de regressão de MQO e o número de observações. Usando o R -quadrado (arredondado para quatro casas decimais) apresentado para essa equação, podemos ver quanto da variação no salário é, realmente, explicada pelo retorno da ação. A resposta é: não muito. O retorno da ação da empresa explica somente 1,3% da variação nos salários dessa amostra de 209 diretores executivos. Isso significa que 98,7% das variações salariais desses diretores executivos são deixadas sem explicação. Essa falta de poder explicativo não deve ser surpreendente demais, já que muitas outras características, tanto da empresa como do diretor executivo individual, devem influenciar o salário; esses fatores estão, necessariamente, incluídos nos erros de uma análise de regressão simples.

Nas ciências sociais não são incomuns R -quadrados baixos nas equações de regressão, especialmente na análise de corte transversal. Discutiremos essa questão, de modo mais geral, sob a análise de regressão múltipla, mas vale a pena enfatizar agora que um R -quadrado aparentemente baixo não significa, necessariamente, que uma equação de regressão de MQO é inútil. Ainda, é possível que (2.39) seja uma boa estimativa da relação *ceteris paribus* entre *salário* e *rma*; se isso for verdade ou não, não depende diretamente da magnitude do R -quadrado. Os estudantes que estão se defrontando com econometria pela primeira vez tendem, ao avaliar equações de regressão, a pôr muito peso na magnitude do R -quadrado. Por enquanto, esteja ciente de que usar o R -quadrado como o principal padrão de medida de sucesso de uma análise econométrica pode levar a confusões.

Algumas vezes, a variável explicativa elucida uma parte substancial da variação amostral na variável dependente.

EXEMPLO 2.9**(Resultados Eleitorais e Gastos de Campanha)**

Na equação de resultados eleitorais (2.28), $R^2 = 0,856$. Assim, a participação dos candidatos nos gastos de campanha explica mais de 85% da variação nos resultados eleitorais nessa amostra. Essa é uma explicação considerável.

2.4 UNIDADES DE MEDIDA E FORMA FUNCIONAL

Duas questões importantes em economia aplicada são: (1) entender como, ao mudar as unidades de medida das variáveis dependentes e/ou independentes, são afetadas as estimativas de MQO e (2) saber como incorporar, à análise de regressão, formas funcionais populares usadas em economia.

A matemática necessária para uma compreensão completa das questões sobre a forma funcional está revista no Apêndice A, disponível no site da Thomson.

Os Efeitos de Mudanças das Unidades de Medida sobre as Estatísticas de MQO

No Exemplo 2.3, escolhemos mensurar o salário anual em milhares de dólares, e o retorno das ações foi medido como uma percentagem (em vez de um decimal). É crucial saber como *salário* e *rma* são medidos nesse exemplo, a fim de dar sentido às estimativas da equação (2.39).

Devemos também saber que as estimativas de MQO mudam de maneira completamente esperada, quando as unidades de medida das variáveis dependente e independente mudam. No Exemplo 2.3, suponhamos que, em vez de medir o salário em milhares de dólares, nós o medimos em dólares. Seja *salardol* o salário em dólares (*salardol* = 845.761 seria interpretado como \$ 845.761). Evidentemente, *salardol* tem uma relação simples com o salário medido em milhares de dólares: *salardol* = 1.000. *salário*. Não precisamos, realmente, computar a regressão *salardol* sobre *rma* para saber que a equação estimada é:

$$\text{salardol} = 963.191 + 18.501 \text{ rma}. \quad (2.40)$$

Obtemos o intercepto e a inclinação em (2.40) ao, simplesmente, multiplicarmos o intercepto e o coeficiente de inclinação em (2.39) por 1.000. Isso dá às equações (2.39) e (2.40) a mesma interpretação. Olhando para (2.40), se *rma* = 0, então *salardol* = 963.191, de modo que o salário previsto é \$ 963.191 [o mesmo valor que obtivemos da equação (2.39)]. Além disso, se *rma* aumenta em um, então o salário previsto aumenta em \$ 18.501; novamente, isso é o que concluímos de nossa análise anterior da equação (2.39).

Em geral, é fácil fazer uma idéia do que acontece às estimativas de intercepto e de inclinação quando se altera a unidade de medida da variável dependente. Se a variável dependente é multiplicada pela constante *c* — o que significa dizer que cada valor na amostra é multiplicado por *c* —, então as estimativas de MQO de intercepto e de inclinação também são multiplicadas por *c*. (Isso assume que nada foi alterado com respeito à variável independente.) No exemplo do salário dos diretores executivos, *c* = 1.000 ao passarmos de *salário* para *salardol*.

Suponha que o salário seja mensurado em centenas de dólares, em vez de milhares de dólares, e o chamemos *salarcent*. Quais serão as estimativas de intercepto e de inclinação na regressão de *salarcent* sobre *rma*?

Também podemos usar o exemplo do salário dos diretores executivos para ver o que acontece quando as unidades de medida da variável independente são mudadas. Defina *rmadec* = *rma*/100 como sendo o equivalente decimal de *rma*; assim, *rmadec* = 0,23 significa um retorno da ação de 23%. A fim de centrarmos o foco na mudança das unidades de medida da variável independente, retornaremos à nossa variável dependente original, *salário*, mensurada em milhares de dólares. Quando regresemos *salário* sobre *rmadec*, obtemos

$$\text{salário} = 963,191 + 1.850,1 \text{ rmadec}. \quad (2.41)$$

O coeficiente de *rmadec* é 100 vezes o coeficiente de *rma* em (2.39). Isso é o que deveria ser. Variar *rma* em um ponto percentual é equivalente a $\Delta \text{rmadec} = 0,01$. De (2.41), se $\Delta \text{rmadec} = 0,01$, então

$\Delta\hat{\text{salário}} = 1.850,1(0,01) = 18,501$, que é igual ao obtido ao se usar (2.39). Observe que, ao passarmos de (2.39) para (2.41), a variável independente foi dividida por 100, e assim a estimativa de inclinação de MQO foi multiplicada por 100, preservando a interpretação da equação. Em geral, se a variável independente é dividida ou multiplicada por alguma constante diferente de zero, c , então o coeficiente de inclinação de MQO é multiplicado ou dividido por c , respectivamente.

O intercepto não mudou em (2.41), pois $r_{\text{madec}} = 0$ ainda corresponde a um retorno zero da ação. Em geral, mudar somente as unidades de medida da variável independente não afeta o intercepto.

Na seção anterior, definimos R -quadrado como uma medida de grau de ajuste para a regressão de MQO. Podemos também questionar o que acontece ao R^2 quando é mudada a unidade de medida da variável independente ou da variável dependente. Sem fazer nenhuma álgebra, deveríamos saber o resultado: o grau de ajuste do modelo não depende das unidades de medida de nossas variáveis. Por exemplo, a quantidade de variação no salário, explicada pelo retorno da ação, não deve depender de o salário ser medido em dólares ou em milhares de dólares, ou de o retorno da ação ser uma percentagem ou um decimal. Essa intuição pode ser verificada matematicamente: usando a definição de R^2 , pode ser mostrado que R^2 é, de fato, invariante a mudanças nas unidades de y ou x .

Incorporação de Não-Linearidades na Regressão Simples

Até aqui, enfatizamos as relações *lineares* entre as variáveis dependente e independente. Como mencionamos no Capítulo 1, relações lineares não são, em geral, suficientes para todas as aplicações econômicas. Felizmente, é bastante fácil incorporar muitas não-linearidades na análise de regressão simples ao definir apropriadamente as variáveis dependente e independente. Vamos tratar aqui de duas possibilidades que freqüentemente aparecem em trabalhos aplicados.

Ao ler trabalhos aplicados nas ciências sociais, com freqüência você encontrará equações de regressão em que a variável dependente aparece na forma logarítmica. Por que isso é feito? Lembre-se do exemplo salários-educação, em que regredimos o salário-hora sobre os anos de educação formal. Obtivemos uma estimativa da inclinação de 0,54 [veja a equação (2.27)], o que significa dizer que, para cada ano adicional de educação, é previsto um aumento de 54 centavos de dólar no salário-hora. Devido à natureza linear de (2.27), 54 centavos de dólar é o aumento tanto para o primeiro ano de educação quanto para o vigésimo ano; isso pode não ser razoável.

Suponha, em vez disso, que o aumento percentual no salário é o mesmo, dado um ano a mais de educação formal. O modelo (2.27) não implica um aumento percentual constante: o aumento depende do salário inicial. Um modelo que gera (aproximadamente) um efeito percentual constante é

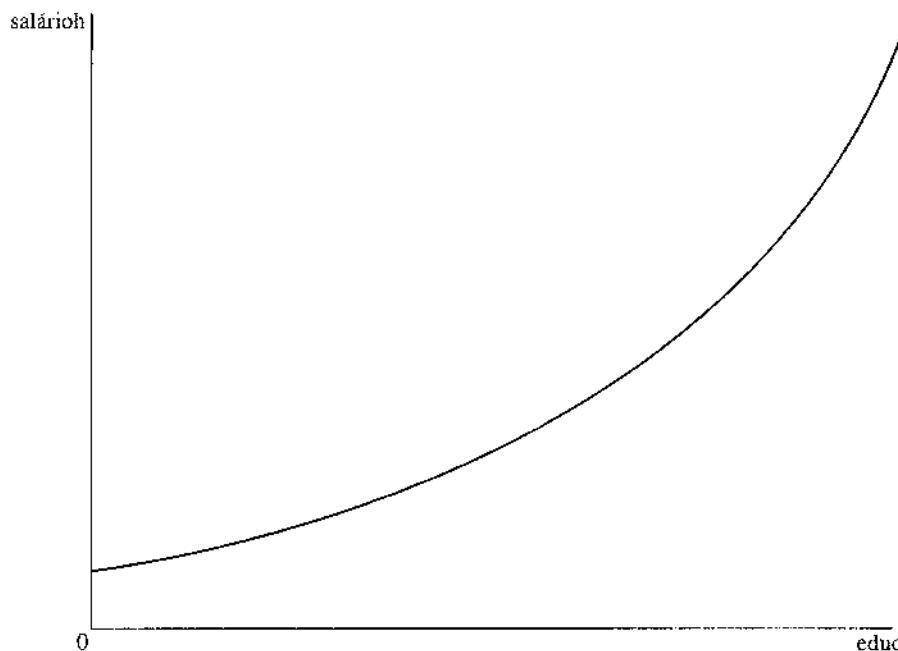
$$\log(\text{salário}_h) = \beta_0 + \beta_1 \text{educ} + u, \quad (2.42)$$

em que $\log(\cdot)$ é o logaritmo *natural*. (Veja Apêndice A, disponível no site da Thomson, para uma revisão sobre logaritmos.) Em particular, se $\Delta u = 0$, então

$$\% \Delta \text{salário}_h \approx (100 \cdot \beta_1) \Delta \text{educ}. \quad (2.43)$$

Figura 2.6

$\text{salárioh} = \exp(\beta_0 + \beta_1 \text{educ})$, com $\beta_1 > 0$.



Observe como multiplicamos β_1 por 100 para obter a variação percentual em salárioh dado um ano adicional de educação formal. Como a variação percentual em salárioh é a mesma para cada ano adicional de educação, a variação em salárioh , para um ano extra de educação formal, *aumenta* quando a educação formal aumenta; em outras palavras, (2.42) implica um retorno *crescente* da educação formal. Com a exponenciação de (2.42), podemos escrever $\text{salárioh} = \exp(\beta_0 + \beta_1 \text{educ} + u)$. O gráfico dessa equação aparece na Figura 2.6, com $u = 0$.

Quando se usa a regressão simples, a estimativa de um modelo como (2.42) é imediata. Apenas defina a variável dependente, y , como $y = \log(\text{salárioh})$. A variável independente é representada por $x = \text{educ}$. A mecânica de MQO é a mesma de antes: as estimativas de intercepto e de inclinação são dadas pelas fórmulas (2.17) e (2.19). Em outras palavras, obtemos $\hat{\beta}_0$ e $\hat{\beta}_1$ da regressão de MQO de $\log(\text{salárioh})$ sobre educ .

EXEMPLO 2.10

(Uma Equação do Logaritmo dos Salários-Hora)

Utilizando os mesmos dados do Exemplo 2.4, mas usando $\log(\text{salárioh})$ como a variável dependente, obtemos a seguinte relação:

$$\hat{\log(\text{salário})} = 0,584 + 0,083 \text{ educ} \quad (2.44)$$

EXEMPLO 2.10 (continuação)

$$n = 526, R^2 = 0,186.$$

O coeficiente de *educ* tem uma interpretação percentual quando ele é multiplicado por 100: para cada ano adicional de educação formal, *salárioh* aumenta 8,3%. Isso é o que os economistas querem dizer quando se referem ao “retorno de um ano adicional de educação formal”.

É importante lembrar que a principal razão para usar o log de *salárioh* em (2.42) é impor um efeito percentual constante da educação formal sobre *salárioh*. Uma vez obtida a equação (2.42), o log natural de *salárioh* é raramente mencionado. Em particular, não é correto dizer que um ano adicional de educação formal aumenta $\log(\text{salárioh})$ em 8,3%.

O intercepto em (2.42) não tem muito significado, visto que ele é o $\log(\text{salárioh})$ previsto quando *educ* = 0. O *R*-quadrado mostra que *educ* explica cerca de 18,6% da variação em $\log(\text{salárioh})$ (não em *salário*). Finalmente, a equação (2.44) pode não capturar toda a não-linearidade da relação entre salário-hora e escolaridade formal. Se houver “efeitos-diploma”, o décimo segundo ano de educação — formatura do ensino médio nos Estados Unidos — deve ser muito mais valioso que o décimo primeiro ano. No Capítulo 7 aprenderemos como lidar com esse tipo de não-linearidade.

Outro uso importante do log natural está em obter um **modelo de elasticidade constante**.

EXEMPLO 2.11**(Salários de Diretores Executivos e Vendas das Empresas)**

Podemos estimar um modelo de elasticidade constante que relaciona o salário dos diretores executivos às vendas das empresas. O conjunto de dados é o mesmo utilizado no Exemplo 2.3, exceto que agora relacionamos *salário* a *vendas*. Seja *vendas* as vendas anuais das empresas, mensurada em milhões de dólares. Um modelo de elasticidade constante é

$$\hat{\log(\text{salário})} = \beta_0 + \beta_1 \log(\text{vendas}) + u, \quad (2.45)$$

em que β_1 é a elasticidade de *salário* com respeito a *vendas*. Esse modelo está compreendido no modelo de regressão simples ao se definir a variável dependente como $y = \log(\text{salário})$, e a variável independente como $x = \log(\text{vendas})$. Ao estimar essa equação por MQO, temos

$$\hat{\log(\text{salário})} = 4,822 + 0,257\log(\text{vendas}) \quad (2.46)$$

$$n = 209, R^2 = 0,211.$$

O coeficiente de $\log(\text{vendas})$ é a elasticidade estimada de *salário* em relação a *vendas*. Ela implica que um aumento de 1% nas vendas das empresas aumenta o salário dos diretores executivos em cerca de 0,257% — a interpretação usual de uma elasticidade.

As duas formas consideradas nesta seção surgirão no restante deste texto. Tratamos aqui de modelos que contêm logaritmos naturais porque eles aparecem muito freqüentemente em trabalhos aplicados. A interpretação desses modelos não será muito diferente no caso da regressão múltipla.

É também útil observar o que acontece às estimativas de intercepto e de inclinação se mudarmos as unidades de medida da variável dependente quando ela aparece na forma logarítmica. Pelo fato de a variação da forma logarítmica aproximar-se de uma variação proporcional, faz sentido que *nada* aconteça com a inclinação. Podemos ver isso ao escrever a variável em uma nova escala como c_1y_i para cada observação i . A equação original é $\log(y_i) = \beta_0 + \beta_1x_i + u_i$. Se adicionarmos $\log(c_1)$ a ambos os lados da equação, obtemos $\log(c_1) + \log(y_i) = [\log(c_1) + \beta_0] + \beta_1x_i + u_i$; ou $\log(c_1y_i) = [\log(c_1) + \beta_0] + \beta_1x_i + u_i$. (Lembre-se de que a soma dos logs é igual ao log de seus produtos, como mostrado no Apêndice A, disponível no site da Thomson.) Portanto, a inclinação ainda é β_1 , mas o intercepto agora é $\log(c_1) + \beta_0$. Semelhantemente, se a variável independente for $\log(x)$, e mudarmos as unidades de medida de x antes de considerarmos o log, a inclinação permanece a mesma, mas o intercepto muda. Pediremos que você verifique essas assertões no Problema 2.9.

Finalizamos esta subseção resumindo quatro combinações de formas funcionais construídas a partir da variável original ou de seu logaritmo natural. Na Tabela 2.3, x e y representam as variáveis em suas formas originais. O modelo com y como a variável dependente e x como a variável independente é chamado modelo *nível-nível*, pois cada variável aparece em sua forma de nível. O modelo com $\log(y)$ como a variável dependente e x como a variável independente é chamado modelo *log-nível*. Não discutiremos aqui, explicitamente, o modelo *nível-log*, pois ele aparece menos freqüentemente na prática. De qualquer forma, veremos exemplos desse modelo em outros capítulos.

A última coluna na Tabela 2.3 mostra a interpretação de β_1 . No modelo log-nível, $100 \cdot \beta_1$ é algumas vezes chamado **semi-elasticidade** de y em relação a x . Como mencionamos no Exemplo 2.11, no modelo log-log β_1 é a **elasticidade** de y em relação a x . A Tabela 2.3 requer um estudo cuidadoso, já que vamos, com freqüência, nos referir a ela no restante do texto.

O Significado da Regressão “Linear”

O modelo de regressão simples que estudamos neste capítulo também é chamado modelo de regressão *linear simples*. No entanto, como acabamos de ver, o modelo geral também permite certas relações *não-lineares*. Portanto, o que significa “linear” aqui? Você pode observar, ao olhar a equação (2.1), que $y = \beta_0 + \beta_1x + u$. O importante é que essa equação é linear nos *parâmetros*, β_0 e β_1 . Não há restrições de como y e x se relacionam com as variáveis explicada e explicativa originais de interesse. Como vimos nos Exemplos 2.7 e 2.8, y e x podem ser os logaritmos naturais de variáveis, e isso é muito comum em aplicações. Mas não precisamos parar aqui. Por exemplo, *nada* nos impede de usar a regressão simples para estimar um modelo tal como $cons = \beta_0 + \beta_1\sqrt{rend} + u$, em que $cons$ é o consumo anual e $rend$ é a renda anual.

Tabela 2.3

Resumo das Formas Funcionais Envolvendo Logaritmos

Modelo	Variável Dependente	Variável Independente	Interpretação de β_1
nível-nível	y	x	$\Delta y = \beta_1 \Delta x$
nível-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\% \Delta x$
log-nível	$\log(y)$	x	$\% \Delta y = (100\beta_1) \Delta x$
log-log	$\log(y)$	$\log(x)$	$\% \Delta y = \beta_1 \% \Delta x$

Embora a mecânica da regressão simples não dependa de como y e x são definidos, a interpretação dos coeficientes depende, realmente, de suas definições. Para que os trabalhos empíricos sejam bem-sucedidos, é muito mais importante tornar-se proficiente em interpretar coeficientes do que eficiente no cálculo de fórmulas como (2.19). Obteremos muito mais prática em como interpretar as estimativas da reta de regressão de MQO quando estudarmos a regressão múltipla.

Muitos modelos *não podem* ser considerados modelos de regressão linear, porque eles não são lineares em seus parâmetros; um exemplo é $\text{cons} = 1/(\beta_0 + \beta_1 \text{rend}) + u$. A estimativa desses modelos leva-nos ao campo âmbito do *modelo de regressão não-linear*, o qual está além do escopo deste texto. Para muitas aplicações, é suficiente escolher um modelo que possa ser expresso dentro do arcabouço da regressão linear.

2.5 VALORES ESPERADOS E VARIÂNCIAS DOS ESTIMADORES DE MQO

Na Seção 2.1 definimos o modelo populacional $y = \beta_0 + \beta_1 x + u$ e afirmamos que a hipótese fundamental para que a análise de regressão simples seja útil é que o valor esperado de u , dado qualquer valor de x , seja zero. Nas seções 2.2, 2.3 e 2.4 discutimos as propriedades algébricas da estimação de MQO. Retornamos agora ao modelo populacional e estudaremos as propriedades *estatísticas* da estimativa de MQO. Em outras palavras, veremos agora $\hat{\beta}_0$ e $\hat{\beta}_1$ como *estimadores* dos parâmetros β_0 e β_1 que aparecem no modelo populacional. Isso significa que estudaremos as propriedades das distribuições de $\hat{\beta}_0$ e $\hat{\beta}_1$ de diferentes amostras aleatórias da população. (O Apêndice C, disponível no site da Thomson, contém as definições de estimadores e revisões de algumas de suas principais propriedades.)

Inexistência de Viés em MQO

Vamos iniciar estabelecendo a inexistência de viés do método MQO sob um conjunto simples de hipóteses. Para referências futuras, é útil numerar essas hipóteses usando o prefixo "RLS" para regressão linear simples. A primeira hipótese define o modelo populacional.

H I P Ó T E S E R L S . 1 (LINEAR NOS PARÂMETROS)

No modelo populacional, a variável dependente y está relacionada à variável independente x e ao erro (ou perturbação) u como

$$y = \beta_0 + \beta_1 x + u, \quad (2.47)$$

em que β_0 e β_1 são os parâmetros de intercepto e de inclinação populacionais, respectivamente.

Ao especificar o modelo populacional — e para ser realista —, y , x e u são todos vistos como variáveis aleatórias. Discutimos, em alguma extensão, a interpretação desse modelo na Seção 2.1 e demos vários exemplos. Na seção anterior aprendemos que a equação (2.47) não é tão restritiva quanto inicialmente parecia; escolhendo y e x apropriadamente, podemos obter relações não-lineares interessantes (como os modelos de elasticidade constante).

Estamos interessados em usar os dados de y e x para estimar os parâmetros β_0 e, especialmente, β_1 . Assumimos que nossos dados foram obtidos de uma amostra aleatória. (Veja o Apêndice C, disponível no site da Thomson, para uma revisão sobre amostragem aleatória.)

H I P Ó T E S E R L S . 2 (AMOSTRAGEM ALEATÓRIA)

Podemos usar uma amostra aleatória de tamanho n , $\{(x_i, y_i); i = 1, 2, \dots, n\}$, proveniente de um modelo populacional.

Em capítulos posteriores que abordam a análise de séries de tempo e problemas de seleção de amostra, teremos de dar um tratamento ao fato de a hipótese de amostragem aleatória não ser mais válida. Nem todas as amostras de corte transversal podem ser vistas como resultados de amostras aleatórias, mas muitas podem ser assim entendidas.

Podemos escrever (2.47), em termos da amostra aleatória como

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \dots, n, \quad (2.48)$$

em que u_i é o erro ou perturbação da observação i (por exemplo, pessoa i , empresa i , cidade i etc.). Assim, u_i contém os fatores não-observáveis da observação i que afetam y_i . Os u_i não devem ser confundidos com os resíduos, \hat{u}_i , definidos na Seção 2.3. Mais adiante, exploraremos a relação entre os erros e os resíduos. Para interpretar β_0 e β_1 em uma aplicação particular, (2.47) é mais instrutiva, mas (2.48) também é necessária para algumas derivações estatísticas.

A relação (2.48) pode ser colocada em um gráfico para um registro particular dos dados, como mostrado na Figura 2.7.

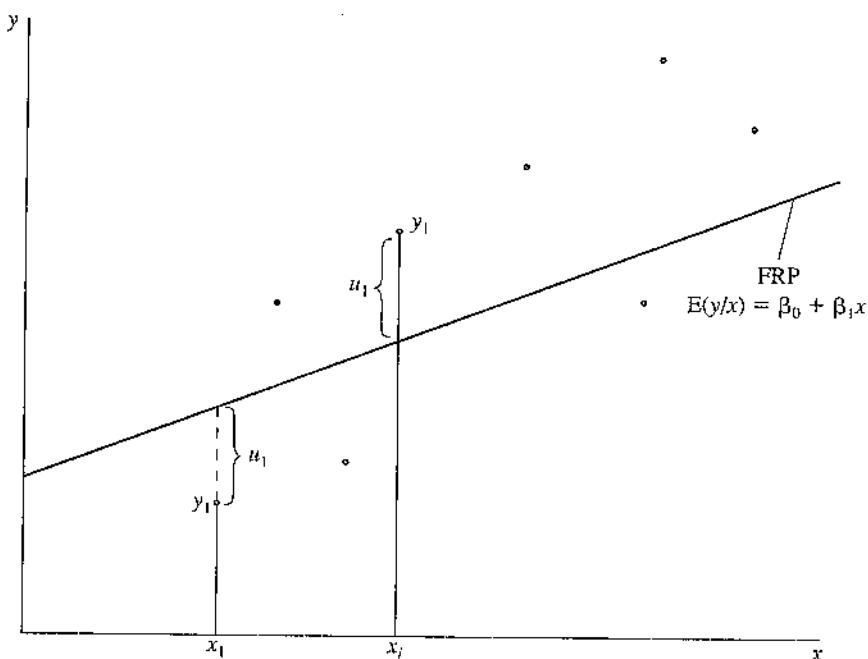
A fim de obter estimadores não-viesados de β_0 e β_1 , precisamos impor a hipótese de média condicional zero que discutimos, com algum detalhe, na Seção 2.1. Agora, vamos adicioná-la explicitamente à nossa lista de hipóteses.

H I P Ó T E S E R L S . 3 (MÉDIA CONDICIONAL ZERO)

$$E(u|x) = 0.$$

Para uma amostra aleatória, essa hipótese implica que $E(u_i|x_i) = 0$, para todo $i = 1, 2, \dots, n$.

Além de restringir a relação entre u e x na população, a hipótese de média condicional zero — juntamente com a hipótese de amostra aleatória — permite uma simplificação técnica conveniente. Em particular, podemos derivar as propriedades estatísticas dos estimadores de MQO como *condicionais* aos valores de x_i em nossa amostra. Tecnicamente, em derivações estatísticas, condicionar aos valores amostrais da variável independente é o mesmo que tratar x_i como *fixo em amostras repetidas*. Esse processo envolve vários passos. Primeiro, escolhemos n valores amostrais para x_1, x_2, \dots, x_n . (Esses valores podem ser repetidos.) Dados esses valores, obtemos uma amostra de y (efetivamente, obtendo uma amostra aleatória de u_i). Em seguida, obtém-se outra amostra de y , usando novamente os mesmos x_1, x_2, \dots, x_n . É assim por diante.

Figura 2.7Gráfico de $y_i = \beta_0 + \beta_1 x_i + u_i$ 

O fixo no cenário de amostras repetidas não é muito realista no contexto não-experimental. Por exemplo, na amostragem de indivíduos do exemplo salários-educação, faz pouco sentido pensar em escolher os valores de *educ* antecipadamente e, em seguida, fazer uma amostra de indivíduos com aqueles níveis particulares de educação formal. A amostragem aleatória, na qual os indivíduos são escolhidos aleatoriamente e seus salários e anos de educação formal são registrados, é um processo representativo de como muitos conjuntos de dados são obtidos para a análise empírica nas ciências sociais. Já que assumimos $E(u|x) = 0$, e temos amostragem aleatória, nada se perde nas derivações ao tratar os x_i como não-aleatórios. O perigo é que o fixo na hipótese de amostras repetidas *sempre* implica que u_i e x_i são independentes. Ao decidir quando a análise de regressão simples produzirá estimadores não-viesados, é crucial pensar em termos da hipótese RLS.3.

Visto que concordamos em condicionar as derivações estatísticas aos valores de x_i , precisamos de uma última hipótese para a inexistência de viés.

H I P Ó T E S E R L S . 4 (VARIAÇÃO AMOSTRAL NA VARIÁVEL INDEPENDENTE)

Na amostra, as variáveis independentes x_i , $i = 1, 2, \dots, n$, não são todas iguais a uma mesma constante. Isso exige alguma variação em x na população.

Encontramos a hipótese RLS.4 quando derivamos as fórmulas dos estimadores de MQO; ela é equivalente a $\sum_{i=1}^n (x_i - \bar{x})^2 > 0$. Das quatro hipóteses feitas, esta é a menos importante, pois ela essen-

cialmente nunca falha em aplicações interessantes. Se a hipótese RLS.4 não se sustentar, não podemos calcular os estimadores de MQO, o que significa que a análise estatística é irrelevante.

Usando o fato de que $\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})y_i$ (veja o Apêndice A, no site da Thomson), podemos escrever o estimador de inclinação de MQO na equação (2.19) como

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.49)$$

Como agora estamos interessados no comportamento de $\hat{\beta}_1$ ao longo de todas as amostras possíveis, $\hat{\beta}_1$ é apropriadamente visto como uma variável aleatória.

Podemos escrever $\hat{\beta}_1$ em termos dos coeficientes populacionais e dos erros ao substituir o lado direito de (2.48) em (2.49). Temos

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\text{SQT}_x} = \frac{\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\text{SQT}_x}, \quad (2.50)$$

onde definimos a variação total em x_i como $\text{SQT}_x = \sum_{i=1}^n (x_i - \bar{x})^2$, a fim de simplificar a notação. (Essa expressão não é exatamente a variância amostral de x_i , pois não a dividimos por $n - 1$.) Usando a álgebra do operador somatório, vamos escrever o numerador de $\hat{\beta}_1$ como

$$\begin{aligned} & \sum_{i=1}^n (x_i - \bar{x})\beta_0 + \sum_{i=1}^n (x_i - \bar{x})\beta_1 x_i + \sum_{i=1}^n (x_i - \bar{x})u_i \\ &= \beta_0 \sum_{i=1}^n (x_i - \bar{x}) + \beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i + \sum_{i=1}^n (x_i - \bar{x})u_i. \end{aligned} \quad (2.51)$$

Como mostrado no Apêndice A, $\sum_{i=1}^n (x_i - \bar{x}) = 0$ e $\sum_{i=1}^n (x_i - \bar{x})x_i = \sum_{i=1}^n (x_i - \bar{x})^2 = s_x^2$. Portanto, podemos escrever o numerador de $\hat{\beta}_1$ como $\beta_1 \text{SQT}_x + \sum_{i=1}^n (x_i - \bar{x})u_i$. Escrevendo isso no denominador resulta

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\text{SQT}_x} = \beta_1 + (1/\text{SQT}_x) \sum_{i=1}^n d_i u_i, \quad (2.52)$$

onde $d_i = x_i - \bar{x}$. Vemos agora que o estimador $\hat{\beta}_1$ é igual à inclinação populacional β_1 mais um termo que é uma combinação linear dos erros $\{u_1, u_2, \dots, u_n\}$. Condicionada aos valores de x_i , a aleatoriedade em $\hat{\beta}_1$ deve-se inteiramente aos erros na amostra. O fato de que esses erros sejam, em geral, diferentes de zero é o que faz com que $\hat{\beta}_1$ seja diferente de β_1 .

Ao usar a representação em (2.52), podemos provar a primeira importante propriedade estatística do método MQO.

T E O R E M A 2 . 1 (INEXISTÊNCIA DE VIÉS EM MQO)

Usando as hipóteses RLS.1 a RLS.4,

$$E(\hat{\beta}_0) = \beta_0 \text{ e } E(\hat{\beta}_1) = \beta_1, \quad (2.53)$$

para quaisquer valores de β_0 e β_1 . Em outras palavras, $\hat{\beta}_0$ é não-viesado para β_0 e $\hat{\beta}_1$ é não-viesado para β_1 .

PROVA: Nesta prova, os valores esperados estão condicionados aos valores amostrais da variável independente. Visto que SQT_x e d_i são funções somente de x_i , eles são não-aleatórios quando condicionais. Portanto, de (2.52), e mantendo o condicionamento a $\{x_1, x_2, \dots, x_n\}$ implícito, temos

$$\begin{aligned} E(\hat{\beta}_1) &= \beta_1 + E[(1/SQT_x) \sum_{i=1}^n d_i u_i] = \beta_1 + (1/SQT_x) \sum_{i=1}^n E(d_i u_i) \\ &= \beta_1 + (1/SQT_x) \sum_{i=1}^n d_i E(u_i) = \beta_1 + (1/SQT_x) \sum_{i=1}^n d_i \cdot 0 = \beta_1, \end{aligned}$$

onde usamos o fato de que o valor esperado de cada u_i (condicional a $\{x_1, x_2, \dots, x_n\}$) é zero sob as hipóteses RLS.2 e RLS.3. Como a inexistência de viés se mantém para qualquer resultado condicionado a $\{x_1, x_2, \dots, x_n\}$, a inexistência de viés também se mantém sem se condicionar a $\{x_1, x_2, \dots, x_n\}$.

A prova para $\hat{\beta}_0$ é agora direta. Obtenha a média de (2.48) através de i para obter $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}$, e insira essa equação na fórmula de $\hat{\beta}_0$:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} + \bar{u} - \hat{\beta}_1 \bar{x} = \beta_0 + (\beta_1 - \hat{\beta}_1) \bar{x} + \bar{u}.$$

Então, condicional aos valores de x_i ,

$$E(\hat{\beta}_0) = \beta_0 + E[(\beta_1 - \hat{\beta}_1) \bar{x}] + E(\bar{u}) = \beta_0 + E[(\beta_1 - \hat{\beta}_1)] \bar{x},$$

já que, pelas hipóteses RLS.2 e RLS.3, $E(\bar{u}) = 0$. No entanto, mostramos que $E(\hat{\beta}_1) = \beta_1$, o que implica que $E[(\hat{\beta}_1 - \beta_1)] = 0$. Assim, $E(\hat{\beta}_0) = \beta_0$. Ambos os argumentos são válidos para quaisquer valores de β_0 e β_1 , e assim estabelecemos a inexistência de viés.

Lembre-se de que a inexistência de viés é uma característica das distribuições amostrais de $\hat{\beta}_1$, e $\hat{\beta}_0$, o que não nos diz nada sobre a estimativa que obtemos de uma dada amostra. Esperamos que, se a amostra que obtemos é de algum modo “típica”, então nossa estimativa deve estar “próxima” do valor populacional. Infelizmente, é sempre possível obter uma amostra ruim que nos dê uma estimativa pontual distante de β_1 , e nós *nunca* saberemos, com certeza, se esse é o caso. Neste ponto, você pode querer revisar o material sobre estimadores não-viesados no Apêndice C (disponível no site da Thomson), especialmente o exercício de simulação da Tabela C.1, o qual ilustra o conceito de inexistência de viés.

Em geral, a inexistência de viés não é válida se qualquer uma das nossas quatro hipóteses não for válida. Isso significa que é importante pensar na veracidade de cada hipótese em uma aplicação particular. Como já discutimos, se a hipótese RLS.4 não for válida, então não seremos capazes de obter as estimativas de MQO. A hipótese RLS.1 requer que y e x sejam linearmente relacionados com uma perturbação adicionada. Certamente, isso pode não ser válido. Mas sabemos também que y e x podem ser escolhidos de forma que possam produzir relações não-lineares interessantes. Estudar com a não validade de (2.47) requer métodos mais avançados, que estão além do escopo deste texto.

Posteriormente, veremos como relaxar a hipótese RLS.2, a hipótese de amostragem aleatória, na análise de séries de tempo. Porém, o que dizer de seu uso na análise de corte transversal? A amostragem aleatória pode não ser válida em um corte transversal quando as amostras não são representativas da população subjacente; de fato, alguns conjuntos de dados são construídos fazendo-se, intencionalmente, amostras de partes diferentes da população. Discutiremos os problemas de amostragem não-aleatória nos Capítulos 9 e 17.

A hipótese na qual devemos nos concentrar agora é a RLS.3. Se RLS.3 se mantém, as estimativas de MQO são não-viesadas. Do mesmo modo, se RLS.3 não se mantém, os estimadores de MQO serão, em geral, viesados. Há maneiras de determinar a direção e o tamanho prováveis do viés, algo que estudaremos no Capítulo 3.

A possibilidade de que x seja correlacionado com u é quase sempre uma preocupação na análise de regressão simples com dados não-experimentais, como indicamos por meio de vários exemplos na Seção 2.1. Usar a regressão simples quando u contém fatores que afetam y e que também estão correlacionados com x pode resultar em *correlação espúria*: isto é, achamos uma relação entre y e x que se deve, em verdade, a outros fatores que afetam y e que também estão correlacionados com x .

EXEMPLO 2.12

[Desempenho em Matemática de Estudantes e o Programa de Merenda Escolar]

Seja $mate10$ a percentagem de alunos do primeiro ano do ensino médio aprovados em um exame de matemática. Suponha que desejamos estimar o efeito do programa de merenda escolar financiado pelo governo sobre o desempenho dos alunos. Esperamos que o programa de merenda tenha um efeito *ceteris paribus* positivo sobre o desempenho: todos os outros fatores permanecendo iguais, se um estudante, bastante pobre para ter regularmente refeições, torna-se qualificado para o programa de merenda escolar, seu desempenho deveria melhorar. Seja $prgalm$ a percentagem de estudantes que estão aptos para participar do programa de merenda escolar. Portanto, o modelo de regressão simples é

$$mate10 = \beta_0 + \beta_1 prgalm + u, \quad (2.54)$$

em que u contém características da escola e do estudante que afetam o desempenho escolar total. Usando os dados do arquivo de MEAP93.RAW de 408 escolas de Michigan no ano escolar 1992-1993, obtemos

EXEMPLO 2.12 (continuação)

$$\hat{ma\grave{e}te10} = 32,14 - 0,319 \text{ prgalm}$$

$$n = 408, R^2 = 0,171.$$

Essa equação prevê que se a participação dos estudantes no programa de merenda escolar aumenta em dez pontos percentuais, a percentagem de estudantes que passa no exame de matemática cai cerca de 3,2 pontos percentuais. Realmente devemos acreditar que a participação maior no programa de merenda escolar causa, de fato, um desempenho pior? Muito provavelmente não. Uma explicação melhor é que o termo erro u na Equação (2.54) está correlacionado com prgalm . De fato, u contém fatores como a taxa de pobreza das crianças que freqüentam a escola, que afeta o desempenho dos estudantes e está altamente correlacionada com a qualificação no programa de merenda. Variáveis como qualidade e recursos da escola também estão contidas em u e, provavelmente, estão correlacionados com prgalm . É importante lembrar que a estimativa $-0,319$ é somente para essa amostra particular, mas seu sinal e magnitude nos fazem suspeitar de que u e x sejam correlacionadas, de modo que a regressão linear é viésada.

Além de variáveis omitidas, há outras razões para que x esteja correlacionado com u no modelo de regressão simples. Como essas mesmas questões surgirão na análise de regressão múltipla, postergaremos até lá um tratamento sistemático do problema.

Variâncias dos Estimadores de MQO

Além de saber que a distribuição amostral de $\hat{\beta}_1$ está centrada em torno de β_1 ($\hat{\beta}_1$ é não-viesado), é importante saber o quanto distante, em média, podemos esperar que $\hat{\beta}_1$ esteja de β_1 . Entre outras coisas, isso nos permite escolher o melhor estimador entre todos os estimadores não-viesados — ou pelo menos entre uma ampla classe deles. A medida de dispersão da distribuição de $\hat{\beta}_1$ (e $\hat{\beta}_0$) com a qual é mais fácil trabalhar é a variância, ou sua raiz quadrada, o desvio-padrão. (Veja o Apêndice C, disponível no site da Thomson, para uma discussão mais detalhada.)

A variância dos estimadores de MQO pode ser calculada sob as hipóteses RLS.1 a RLS.4. Entretanto, as expressões dessas variâncias são complexas. Em vez disso, vamos adicionar uma hipótese que é tradicional na análise de corte transversal. Essa hipótese afirma que a variância do termo não-observável, u , condicionado a x , é constante. Ela é conhecida como a hipótese de **homoscedasticidade** ou de “variância constante”.

H I P Ó T E S E R L S . 5 (HOMOSCEDASTICIDADE)
 $\text{Var}(u|x) = \sigma^2$.

Devemos enfatizar que a hipótese de homoscedasticidade é completamente distinta da hipótese de média condicional zero, $E(u|x) = 0$. A hipótese RLS.3 compreende o *valor esperado* de u , enquanto a hipótese RLS.5 diz respeito à *variância* de u (ambos condicionados a x). Lembre-se de que nós estabelecemos a inexistência de viés de MQO sem a hipótese RLS.5: a hipótese de homoscedasticidade não desempenha *qualquer* papel para mostrar que $\hat{\beta}_0$ e $\hat{\beta}_1$ são não-viesados. Adicionamos a hipótese

RLS.5 pois ela simplifica os cálculos da variância de $\hat{\beta}_0$ e $\hat{\beta}_1$, e porque ela implica que o método de mínimos quadrados ordinários tenha certas propriedades de eficiência, algo que veremos no Capítulo 3. Se assumíssemos que u e x são *independentes*, a distribuição de u , dado x , não dependeria de x , e assim $E(u|x) = E(u) = 0$ e $\text{Var}(u|x) = \sigma^2$. No entanto, algumas vezes independência é, uma hipótese forte demais.

Como $\text{Var}(u|x) = E(u^2|x) - [E(u|x)]^2$ e $E(u|x) = 0$, $\sigma^2 = E(u^2|x)$, o que significa que σ^2 também é a esperança *não-condicional* de u^2 . Portanto, $\sigma^2 = E(u^2) = \text{Var}(u)$, pois $E(u) = 0$. Em outras palavras, σ^2 é a variância *não-condicional* de u , e por isso σ^2 é freqüentemente chamado de **variância do erro** ou variância da perturbação. A raiz quadrada de σ^2 , σ , é o desvio-padrão do erro. Um σ grande significa que a distribuição dos fatores não-observáveis que afetam y é mais dispersa.

Freqüentemente, é útil escrever as hipóteses RLS.3 e RLS.5 em termos da média condicional e da variância condicional de y :

$$E(y|x) = \beta_0 + \beta_1 x. \quad (2.55)$$

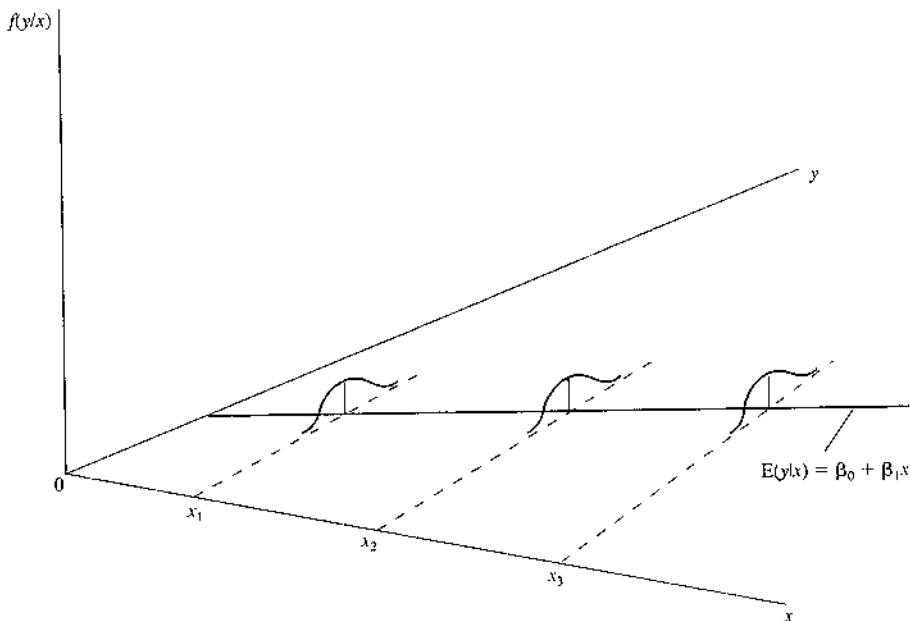
$$\text{Var}(y|x) = \sigma^2. \quad (2.56)$$

Em outras palavras, a esperança condicional de y , dado x , é linear em x , mas a variância de y , dado x , é constante. Essa situação está ilustrada na Figura 2.8, em que $\beta_0 > 0$ e $\beta_1 > 0$.

Quando $\text{Var}(u|x)$ depende de x , diz-se que o termo erro apresenta **heteroscedasticidade** (ou variância não-constante). Como $\text{Var}(u|x) = \text{Var}(y|x)$, a heteroscedasticidade está presente sempre que $\text{Var}(y|x)$ é uma função de x .

Figura 2.8

O modelo de regressão simples sob homoscedasticidade.

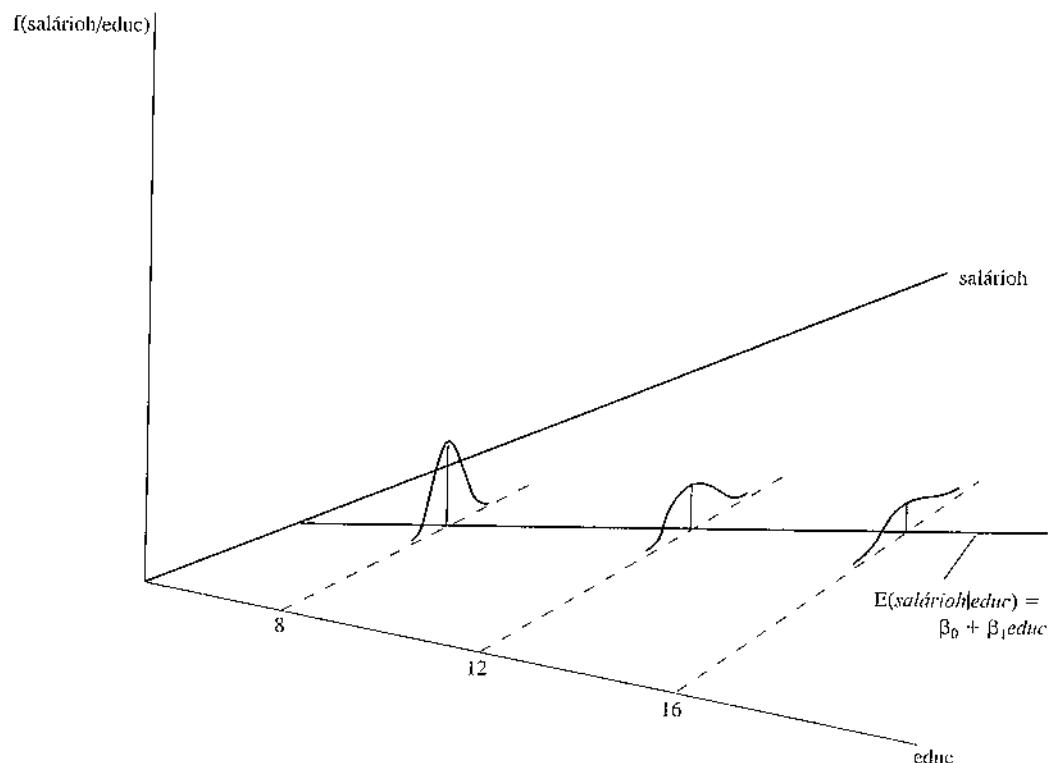


EXEMPLO 2.13**(Heteroscedasticidade em uma Equação de Salários)**

A fim de obter um estimador não-viesado do efeito *ceteris paribus* de *educ* sobre *salárioh*, devemos assumir que $E(u|educ) = 0$, e isso implica $E(salárioh|educ) = \beta_0 + \beta_1 educ$. Se também usarmos a hipótese de homoscedasticidade, então $\text{Var}(u|educ) = \sigma^2$ não depende do nível de educação formal, que é o mesmo que assumir que $\text{Var}(salárioh|educ) = \sigma^2$. Assim, enquanto se deixa o salário-hora médio aumentar com o nível de educação formal – é essa taxa de crescimento que estamos interessados em descrever – assume-se que a variabilidade no salário horário em torno de sua média é constante através de todos os níveis de educação formal: isso pode não ser realista. É provável que pessoas com mais tempo de educação formal tenham uma variedade maior de interesses e de oportunidades de trabalho, o que poderia levar a uma variabilidade maior nos níveis de educação formal mais elevados. Pessoas com níveis de educação formal bastante baixos têm muito poucas oportunidades e, freqüentemente, precisam trabalhar recebendo salário mínimo; isso tem o efeito de reduzir a variabilidade salarial nos níveis baixos de educação formal. Essa situação é mostrada na Figura 2.9. Em última análise, se a hipótese RLS.5 se mantém ou não é uma questão empírica. No Capítulo 8 mostraremos como testar a hipótese RLS.5.

Figura 2.9

$\text{Var}(\text{salárioh}/\text{educ})$ crescendo com educ .



Embora o valor esperado de $\hat{\beta}_0$ iguale-se a β_0 , e similarmente para $\hat{\beta}_1$, \hat{u}_i não é o mesmo que u_i . A diferença entre eles tem, de fato, um *valor esperado* igual a zero.

Agora que entendemos a diferença entre os erros e os resíduos, podemos retornar para estimar σ^2 . Primeiro, $\sigma^2 = E(u^2)$, de modo que um “estimador” não-viesado de σ^2 é $n^{-1} \sum_{i=1}^n u_i^2$. Infelizmente, esse não é um estimador verdadeiro, pois não observamos os erros u_i . Mas, temos, de fato, as estimativas de u_i , a saber, os resíduos \hat{u}_i de MQO. Se substituíssemos os erros pelos resíduos de MQO, temos $n^{-1} \sum_{i=1}^n \hat{u}_i^2 = SQR/n$. Esse é um estimador verdadeiro, porque ele fornece uma regra computável para qualquer amostra de dados sobre x e y . Uma ligeira desvantagem desse estimador é que ele resulta viesado (embora o viés seja pequeno para n grande). Como é fácil calcular um estimador não-viesado, usamos esse como substituto.

O estimador SQR/n é viesado, essencialmente, porque ele não explica a razão de duas restrições que devem ser satisfeitas pelos resíduos de MQO. Essas restrições são dadas pelas duas condições de primeira ordem de MQO:

$$\sum_{i=1}^n \hat{u}_i = 0, \quad \sum_{i=1}^n x_i \cdot \hat{u}_i = 0. \quad (2.60)$$

Uma maneira de ver essas restrições é esta: se nós conhecemos $n - 2$ dos resíduos, podemos sempre obter os outros dois resíduos usando as restrições implicadas pelas condições de primeira ordem em (2.60). Assim, há somente $n - 2$ graus de liberdade nos resíduos de MQO, em oposição a n graus de liberdade nos erros. Se substituíssemos \hat{u}_i por u_i em (2.60), as restrições não mais se manteriam. O estimador não-viesado de σ^2 que utilizaremos faz um ajustamento dos graus de liberdade:

$$\hat{\sigma}^2 = \frac{1}{(n - 2)} \sum_{i=1}^n \hat{u}_i^2 = SQR/(n - 2). \quad (2.61)$$

(Esse estimador é, às vezes, denominado por s^2 , mas continuaremos a usar a convenção de colocar “chapéus” sobre os estimadores.)

T E O R E M A 2 . 3 (ESTIMAÇÃO NÃO-VIESADA DE σ^2)

Sob as hipóteses RLS.1 a RLS.5,

$$E(\hat{\sigma}^2) = \sigma^2.$$

PROVA: Se construirmos a média da equação (2.59) para todos os i e usarmos o fato de que os resíduos de MQO têm média igual a zero, temos $0 = \bar{u} - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)\bar{x}$; subtraindo essa equação de (2.59), resulta $\hat{u}_i = (u_i - \bar{u}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x})$. Portanto, $\hat{u}_i^2 = (u_i - \bar{u})^2 + (\hat{\beta}_1 - \beta_1)^2(x_i - \bar{x})^2 - 2(u_i - \bar{u})(\hat{\beta}_1 - \beta_1)(x_i - \bar{x})$. A soma ao longo de todos os i resulta na equação $\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n (u_i - \bar{u})^2 + (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 - 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^n u_i (x_i - \bar{x})$. Agora, o valor esperado do primeiro termo é $(n - 1)\sigma^2$, algo que está mostrado no Apêndice C, no site da Thomson.

T E O R E M A 2 . 3 (ESTIMAÇÃO NÃO-VIESADA DE σ^2) (continuação)

O valor esperado do segundo termo é, simplesmente, σ^2 , porque $E[(\hat{\beta}_1 - \beta_1)^2] = \text{Var}(\hat{\beta}_1) = \sigma^2/s_v^2$.

Finalmente, o terceiro termo pode ser escrito como: $2(\hat{\beta}_1 - \beta_1)^2 s_v^2$; aplicando as esperanças, resulta em $2\sigma^2$.

Colocando esses três termos juntos, obtemos $E\left(\sum_{i=1}^n \hat{u}_i^2\right) = (n - 1)\sigma^2 + \sigma^2 - 2\sigma^2 = (n - 2)\sigma^2$, de modo que $E[\text{SQR}/(n - 2)] = \sigma^2$.

Se $\hat{\sigma}^2$ for inserido nas fórmulas da variância (2.57) e (2.58), então teremos estimadores não-viesados de $\text{Var}(\hat{\beta}_1)$ e $\text{Var}(\hat{\beta}_0)$. Posteriormente, necessitaremos de estimadores dos desvios padrão de $\hat{\beta}_1$ e $\hat{\beta}_0$, e isso requer estimar σ . O estimador natural de σ é

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} \quad (2.62)$$

e é chamado **erro-padrão da regressão** (EPR). (Outros nomes para $\hat{\sigma}$ são *erro-padrão da estimativa* e *raiz do erro quadrado médio*, mas não os usaremos.) Ainda que $\hat{\sigma}$ não seja um estimador não-viesado de σ , podemos mostrar que ele é um estimador *consistente* de σ (veja Apêndice C, disponível no site da Thomson), e que ele servirá muito bem para nossos propósitos.

A estimativa $\hat{\sigma}$ é interessante, já que ela é uma estimativa do desvio-padrão dos fatores não observáveis que afetam y ; equivalentemente, ela estima o desvio-padrão em y após os efeitos de x terem sido retirados. A maior parte dos programas econometrícios informa o valor de $\hat{\sigma}$ juntamente com o R -quadrado, o intercepto, a inclinação e outras estatísticas de MQO (sob um dos vários nomes listados anteriormente). Por enquanto, nosso principal interesse está em usar $\hat{\sigma}$ para estimar os desvios-padrão de $\hat{\beta}_0$ e $\hat{\beta}_1$. Como $d\text{p}(\hat{\beta}_1) = \sigma/s_v$, o estimador natural de $d\text{p}(\hat{\beta}_1)$ é

$$\text{ep}(\hat{\beta}_1) = \hat{\sigma}/s_v = \hat{\sigma}/\left(\sum_{i=1}^n (x_i - \bar{x})^2\right)^{1/2},$$

que é chamado de **erro-padrão de $\hat{\beta}_1$** . Observe que $\text{ep}(\hat{\beta}_1)$ é visto como uma variável aleatória quando pensamos em processar o método MQO usando diferentes amostras de y ; isso ocorre porque σ^2 varia com diferentes amostras. Para uma dada amostra, $\text{ep}(\hat{\beta}_1)$ é um número, exatamente como $\hat{\beta}_1$ é simplesmente um número quando nós o calculamos a partir de dados conhecidos.

Semelhantemente, $\text{ep}(\hat{\beta}_0)$ é obtido de $d\text{p}(\hat{\beta}_0)$ ao substituir σ por $\hat{\sigma}$. O erro-padrão de qualquer estimativa nos dá uma idéia de qual preciso é o estimador. Os erros-padrão desempeham um papel central em todo este texto; nós os usaremos para construir estatísticas de testes e intervalos de confiança para todos os procedimentos econometrícios que cobriremos, a partir do Capítulo 4.

2.6 REGRESSÃO ATRAVÉS DA ORIGEM

Em raros casos, desejamos impor a restrição de que, quando $x = 0$, o valor esperado de y é zero. Há certas relações para as quais isso é razoável. Por exemplo, se a renda (x) for zero, então os gastos com o imposto de renda (y) devem ser zero. Além disso, há problemas quando um modelo que originalmente tem um intercepto diferente de zero é transformado em um modelo sem um intercepto.

Formalmente, nós escolhemos agora um estimador da inclinação, que chamaremos de $\tilde{\beta}_1$, e uma reta da forma

$$\tilde{y} = \tilde{\beta}_1 x, \quad (2.63)$$

em que os sinais gráficos do til sobre $\tilde{\beta}_1$ e \tilde{y} são usados para distinguir esse problema do problema muito mais comum de estimar um intercepto juntamente com uma inclinação. Costuma-se chamar (2.63) de **regressão através da origem**, pois a reta (2.63) passa pelo ponto $x = 0, \tilde{y} = 0$. Para obter a estimativa de inclinação em (2.63), nós ainda contamos com o método de mínimos quadrados ordinários, que minimiza, nesse caso, a soma dos resíduos quadrados:

$$\sum_{i=1}^n (y_i - \tilde{\beta}_1 x_i)^2. \quad (2.64)$$

Usando cálculo, pode-se mostrar que $\tilde{\beta}_1$ deve resolver a condição de primeira ordem:

$$\sum_{i=1}^n x_i (y_i - \tilde{\beta}_1 x_i) = 0. \quad (2.65)$$

Daí, podemos resolver para $\tilde{\beta}_1$:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}, \quad (2.66)$$

desde que nem todos os x_i sejam zero — um caso que excluímos.

Observe como $\tilde{\beta}_1$ se compara com a estimativa de inclinação quando também estimamos o intercepto (em vez de determiná-lo igual a zero). Essas duas estimativas são as mesmas se, e somente se, $\bar{x} = 0$. [Veja a equação (2.49) para $\hat{\beta}_1$.] Obter uma estimativa de β_1 usando a regressão através da origem não é freqüente em trabalhos aplicados, e por boas razões: se o intercepto $\beta_0 \neq 0$, logo $\tilde{\beta}_1$ é um estimador viesado de β_1 . Você será solicitado a provar isso no Problema 2.8.

Neste capítulo introduzimos o modelo de regressão simples e cobrimos suas propriedades básicas. Dada uma amostra aleatória, o método de mínimos quadrados ordinários é usado para estimar os parâmetros de inclinação e de intercepto no modelo populacional. Demonstramos a álgebra da reta de regressão de MQO, incluindo os cálculos dos valores estimados e dos resíduos, e a obtenção das variações previstas na variável dependente a partir de uma dada variação na variável independente. Na Seção 2.4 discutimos duas questões de importância prática: (1) o comportamento das estimativas de MQO quando mudamos as unidades de medida da variável dependente ou da variável independente e (2) o uso do log natural na elaboração de modelos de elasticidade constante e de semi-elasticidade constante.

Na Seção 2.5, mostramos que, sob as quatro hipóteses RLS.1 a RLS.4, os estimadores de MQO são não-viesados. A hipótese fundamental é que o termo erro u tem média zero, dado qualquer valor da variável independente x . Infelizmente, há razões para pensar que isso é falso em muitas aplicações da regressão simples nas ciências sociais, em que os fatores omitidos em u estão freqüentemente correlacionados com x . Quando adicionamos a hipótese de que a variância do erro, dado x , é constante, obtemos fórmulas simples das variâncias amostrais dos estimadores de MQO. Como vimos, a variância do estimador de inclinação $\hat{\beta}_1$ cresce quando a variância do erro cresce, e ela decresce quando há mais variação amostral na variável independente. Também derivamos um estimador não-viesado para $\sigma^2 = \text{Var}(u)$.

Na Seção 2.6, discutimos brevemente a regressão através da origem, cujo estimador de inclinação é obtido sob a hipótese de que o intercepto é zero. Às vezes, tal regressão é útil, mas ela não aparece com freqüência em trabalhos aplicados.

Temos ainda muito trabalho por fazer. Por exemplo, ainda não sabemos como testar hipóteses sobre os parâmetros populacionais, β_0 e β_1 . Assim, embora saibamos que o método MQO é, sob as hipóteses RLS.1 a RLS.4, não-viesado para os parâmetros populacionais, não temos um modo de fazer inferências sobre a população. Outros tópicos, tais como a eficiência de MQO relativa a outros procedimentos possíveis, também foram omitidos.

As questões de intervalos de confiança, testes de hipóteses e eficiência também são centrais para a análise de regressão múltipla. Como a maneira que construímos os intervalos de confiança e as estatísticas de testes é muito similar para a regressão múltipla — e porque a regressão simples é um caso especial da regressão múltipla —, nosso tempo será mais bem gasto se nos movermos para a regressão múltipla, que é muito mais aplicável que a regressão simples. Nossa propósito, no Capítulo 2, foi fazer você pensar nas questões que surgem na análise econometrística dentro de uma estrutura clara e simples.

2.1 Seja $filhos$ o número de filhos de uma mulher e $educ$ os anos de educação da mulher. Um modelo simples que relaciona a fertilidade a anos de educação é

$$filhos = \beta_0 + \beta_1 educ + u,$$

em que u é um erro não-observável.

- (i) Que tipos de fatores estão contidos em u ? É provável que eles estejam correlacionados com o nível de educação?

- (ii) Uma análise de regressão simples mostrará o efeito *ceteris paribus* da educação sobre a fertilidade? Explique.

2.2 No modelo de regressão linear simples $y = \beta_0 + \beta_1 x + u$, suponha que $E(u) \neq 0$. Fazendo $\alpha_0 = E(u)$, mostre que o modelo pode sempre ser reescrito com a mesma inclinação, mas com um novo intercepto e erro, em que o novo erro tem um valor esperado zero.

2.3 A tabela seguinte contém as variáveis *nmgrad* (nota média em curso superior nos Estados Unidos) e *tac* (nota do teste de avaliação de conhecimentos para ingresso em curso superior nos Estados Unidos) com as notas hipotéticas de oito estudantes de curso superior. A nota *nmgrad* está baseada em uma escala de quatro pontos e foi arredondada para um dígito após o ponto decimal. A nota *tac* baseia-se em uma escala de 36 pontos e foi arredondada para um número inteiro.

Estudante	<i>nmgrad</i>	<i>tac</i>
1	2,8	21
2	3,4	24
3	3,0	26
4	3,5	27
5	3,6	29
6	3,0	25
7	2,7	25
8	3,7	30

- (i) Estime a relação entre *nmgrad* e *tac* usando MQO; isto é, obtenha as estimativas de intercepto e de inclinação da equação

$$\hat{nmgrad} = \hat{\beta}_0 + \hat{\beta}_1 tac.$$

Comente a direção da relação. O intercepto tem uma interpretação útil aqui? Explique. Qual deveria ser o valor previsto de *nmgrad* se a nota *tac* aumentasse em cinco pontos?

- (ii) Calcule os valores estimados e os resíduos de cada observação e verifique que a soma dos resíduos é (aproximadamente) zero.
 (iii) Qual é o valor previsto de *nmgrad* quando *tac* = 20?
 (iv) Quanto da variação de *nmgrad* dos 8 estudantes é explicada por *tac*? Explique.

2.4 Os dados do arquivo BWGHT.RAW contém dados de nascimentos por mulheres nos Estados Unidos. As duas variáveis de interesse são: a variável dependente, peso dos recém-nascidos em onças* (*pesonas*), e a variável explicativa, número médio de cigarros que a mãe fumou por dia durante a gravidez (*cigs*). A seguinte regressão simples foi estimada usando dados de $n = 1.388$ nascimentos:

$$\hat{pesonas} = 119,77 - 0,514 cigs$$

* NT: 1 onça = 31,10 g.

- (i) Qual é o peso de nascimento previsto quando $cigs = 0$? E quando $cigs = 20$ (um maço por dia)? Comente a diferença.
- (ii) O modelo de regressão simples necessariamente captura uma relação causal entre o peso de nascimento da criança e os hábitos de fumar da mãe? Explique.
- (iii) Para prever um peso de nascimento de 125 onças, qual deveria ser a magnitude de $cigs$? Comente.
- (iv) Qual a fração de mulheres na amostra que não fumaram enquanto estiveram grávidas? Isso ajuda a reconciliar sua conclusão da parte (iii)?

2.5 Na função de consumo linear

$$\hat{cons} = \hat{\beta}_0 + \hat{\beta}_1 rend,$$

a propensão marginal a consumir PMgC (estimada) é simplesmente a inclinação $\hat{\beta}_1$, enquanto a propensão média a consumir PmeC é $\hat{\beta}_0/rend + \hat{\beta}_1$. Usando as observações de renda e consumo anuais de 100 famílias (ambas medidas em dólares), obteve-se a seguinte equação:

$$\hat{cons} = -124,84 + 0,853 rend$$

$$n = 100, R^2 = 0,692.$$

- (i) Interprete o intercepto dessa equação e comente seu sinal e magnitude.
- (ii) Qual é o consumo previsto quando a renda familiar é \$ 30.000?
- (iii) Com $rend$ sobre o eixo de x , desenhe um gráfico da PMgC e da PmeC estimadas.

2.6 Usando dados de casas vendidas em 1988 em Andover, Massachusetts [Kiel e McClain (1995)], a equação seguinte relaciona os preços das casas ($preço$) à distância de um incinerador de lixo recentemente construído ($dist$):

$$\log(préço) = 9,40 + 0,312 \log(dist)$$

$$n = 135, R^2 = 0,162.$$

- (i) Interprete o coeficiente de $\log(dist)$. O sinal dessa estimativa é o que você esperava?
- (ii) Você considera que a regressão simples oferece um estimador não-viesado da elasticidade *ceteris paribus* de $preço$ em relação a $dist$? (Pense na decisão da cidade sobre onde colocar o incinerador.)
- (iii) Quais outros fatores relativos a casas afetam seu preço? Eles poderiam estar correlacionados com a distância do incinerador?

2.7 Considere a função de poupança

$$poup = \beta_0 + \beta_1 rend + u, \quad u = \sqrt{rend} \cdot e,$$

onde e é uma variável aleatória com $E(e) = 0$ e $\text{Var}(e) = \sigma_e^2$. Assuma que e é independente de $rend$.

- (i) Mostre que $E(u|rend) = 0$, de modo que a hipótese de média condicional zero (hipótese RLS.3) é satisfeita. [Sugestão: se e é independente de $rend$, então $E(e|rend) = E(e)$.]
- (ii) Mostre que $\text{Var}(u|rend) = \sigma_e^2 rend$, de modo que a hipótese de homoscedasticidade RLS.5 é violada. Em particular, a variância de $poup$ aumenta com $rend$. [Sugestão: $\text{Var}(e|rend) = \text{Var}(e)$, se e e $rend$ são independentes.]
- (iii) Faça uma discussão que sustente a hipótese de que a variância da poupança aumenta com a renda da família.

2.8 Considere o modelo de regressão simples padrão $y = \beta_0 + \beta_1 x + u$, sob as hipóteses RLS.1 a RLS.4. Os estimadores usuais $\hat{\beta}_0$ e $\hat{\beta}_1$ são não-viesados para seus respectivos parâmetros populacionais. Seja $\tilde{\beta}_1$ o estimador de β_1 obtido ao assumir que o intercepto é zero (veja a Seção 2.6).

- (i) Encontre $E(\tilde{\beta}_1)$ em termos de x_i , β_0 e β_1 . Verifique que $\tilde{\beta}_1$ é não-viesado para β_1 quando o intercepto populacional (β_0) é zero. Há outros casos em que $\tilde{\beta}_1$ é não-viesado?
 - (ii) Encontre a variância de $\tilde{\beta}_1$. [Sugestão: a variância não depende de β_0 .]
 - (iii) Mostre que $\text{Var}(\tilde{\beta}_1) \leq \text{Var}(\hat{\beta}_1)$. [Sugestão: para qualquer amostra de dados, $\sum_{i=1}^n x_i^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$, com a desigualdade estrita preponderando, a não ser que $\bar{x} = 0$.]
 - (iv) Comente a relação entre viés e variância, ao escolher entre $\hat{\beta}_1$ e $\tilde{\beta}_1$.
- 2.9** (i) Sejam $\hat{\beta}_0$ e $\hat{\beta}_1$ o intercepto e a inclinação da regressão de y_i sobre x_i , usando n observações. Sejam c_1 e c_2 constantes, com $c_2 \neq 0$. Sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ o intercepto e a inclinação da regressão de $c_1 y_i$ sobre $c_2 x_i$. Mostre que $\tilde{\beta}_1 = (c_1/c_2) \hat{\beta}_1$ e $\tilde{\beta}_0 = c_1 \hat{\beta}_0$, verificando as observações sobre as unidades de medida da Seção 2.4. [Sugestão: para obter $\tilde{\beta}_1$, insira as transformações de x e y em (2.19). Em seguida, use (2.17) para $\tilde{\beta}_0$, estando seguro de usar as transformações de x e y e a inclinação correta.]
- (ii) Agora, sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ os parâmetros estimados da regressão de $(c_1 + y_i)$ sobre $(c_2 + x_i)$ (sem qualquer restrição sobre c_1 ou c_2). Mostre que $\tilde{\beta}_1 = \hat{\beta}_1$ e $\tilde{\beta}_0 = \hat{\beta}_0 + c_1 - c_2 \hat{\beta}_1$.
- (iii) Agora, sejam $\hat{\beta}_0$ e $\hat{\beta}_1$ as estimativas de MQO da regressão $\log(y_i)$ sobre x_i , para a qual devemos assumir $y_i > 0$ para todo i . Para $c_1 > 0$, sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ o intercepto e a inclinação da regressão de $\log(c_1 y_i)$ sobre x_i . Mostre que $\tilde{\beta}_1 = \hat{\beta}_1$ e $\tilde{\beta}_0 = \log(c_1) + \hat{\beta}_0$.
- (iv) Agora, assumindo que $x_i > 0$ para todo i , sejam $\tilde{\beta}_0$ e $\tilde{\beta}_1$ o intercepto e a inclinação da regressão de y_i sobre $\log(c_2 x_i)$. Como $\tilde{\beta}_0$ e $\tilde{\beta}_1$ comparam-se com o intercepto e a inclinação da regressão de y_i sobre $\log(x_i)$?

Minimizando a Soma dos Resíduos Quadrados

Mostramos aqui que as estimativas de MQO $\hat{\beta}_0$ e $\hat{\beta}_1$ minimizam a soma dos resíduos quadrados, como afirmado na Seção 2.2. Formalmente, o problema é caracterizar as soluções $\hat{\beta}_0$ e $\hat{\beta}_1$ para o problema de minimização

$$\min_{b_0, b_1} \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2,$$

onde b_0 e b_1 são argumentos *dummies* para o problema de otimização; por simplicidade, chame essa função $Q(b_0, b_1)$. De um resultado fundamental do cálculo multivariado (veja Apêndice A, disponível no site de Thomson), uma condição necessária para $\hat{\beta}_0$ e $\hat{\beta}_1$ resolverem o problema de minimização é que as derivadas parciais de $Q(b_0, b_1)$ em relação a b_0 e b_1 devem ser zero quando avaliadas com $\hat{\beta}_0$ e $\hat{\beta}_1$: $\partial Q(\beta_0, \hat{\beta}_1)/\partial b_0 = 0$ e $\partial Q(\hat{\beta}_0, \hat{\beta}_1)/\partial b_1 = 0$. Usando a regra da cadeia do cálculo, essas duas equações tornam-se

$$\begin{aligned} -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\ -2 \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0. \end{aligned}$$

Essas duas equações são exatamente (2.14) e (2.15) multiplicadas por $-2n$ e, portanto, são solucionadas pelos mesmos $\hat{\beta}_0$ e $\hat{\beta}_1$.

Como sabemos que, realmente, minimizamos a soma dos resíduos quadrados? As condições de primeira ordem são necessárias, mas não são suficientes. Uma maneira de verificar que minimizamos a soma dos resíduos quadrados é escrever, para qualquer b_0 e b_1 ,

$$\begin{aligned} Q(b_0, b_1) &= \sum_{i=1}^n [y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i + (\hat{\beta}_0 - b_0) + (\hat{\beta}_1 - b_1) x_i]^2 \\ &= \sum_{i=1}^n [\hat{u}_i + (\hat{\beta}_0 - b_0) + (\hat{\beta}_1 - b_1) x_i]^2 \\ &= \sum_{i=1}^n \hat{u}_i^2 + n(\hat{\beta}_0 - b_0)^2 + (\hat{\beta}_1 - b_1)^2 \sum_{i=1}^n x_i^2 + 2(\hat{\beta}_0 - b_0)(\hat{\beta}_1 - b_1) \sum_{i=1}^n x_i, \end{aligned}$$

onde usamos as equações (2.30) e (2.31). A soma dos resíduos quadrados não depende de b_0 e b_1 , enquanto a soma dos últimos três termos pode ser escrita como

$$\sum_{i=1}^n [(\hat{\beta}_0 - b_0) + (\hat{\beta}_1 - b_1) x_i]^2,$$

como pode ser verificado diretamente por álgebra. Visto que essa expressão é uma soma de termos quadrados, ela é maior que zero. Portanto, seu menor valor ocorre quando $b_0 = \hat{\beta}_0$ e $b_1 = \hat{\beta}_1$.

Análise de Regressão Múltipla: Estimação

No Capítulo 2 aprendemos a usar a análise de regressão simples para explicar uma variável dependente y como função de uma única variável independente x . A desvantagem principal de usar a análise de regressão simples em trabalhos empíricos é o fato de ser muito difícil obter conclusões *ceteris paribus* sobre como x afeta y : a hipótese fundamental, RLS.3 — todos os outros fatores que afetam y são não-correlacionados com x —, é freqüentemente irreal.

A análise de regressão múltipla é mais receptiva à análise *ceteris paribus*, pois ela nos permite controlar *explicitamente* muitos outros fatores que, de maneira simultânea, afetam a variável dependente. Isso é importante tanto para testar teorias econômicas quanto para avaliar efeitos da política governamental quando devemos nos basear em dados não-experimentais. Como os modelos de regressão múltipla podem acomodar muitas variáveis explicativas que podem estar correlacionadas, podemos esperar inferir causalidade nos casos em que a análise de regressão simples seria enganosa.

Naturalmente, se adicionarmos ao nosso modelo mais fatores que são úteis para explicar y , então mais da variação de y poderá ser explicada. Assim, a análise de regressão múltipla pode ser usada para construir modelos melhores para prever a variável dependente.

Uma vantagem adicional da análise de regressão múltipla é que ela pode incorporar, completamente, relações de formas funcionais gerais. No modelo de regressão simples, somente a função de uma variável explicativa pode aparecer na equação. Como veremos, o modelo de regressão múltipla permite muito mais flexibilidade.

A Seção 3.1 introduz, formalmente, o modelo de regressão múltipla e, mais adiante, discute as vantagens da regressão múltipla sobre a regressão simples. Na Seção 3.2, demonstramos como estimar os parâmetros do modelo de regressão múltipla por meio do método de mínimos quadrados ordinários. Nas seções 3.3, 3.4 e 3.5 descrevemos várias propriedades dos estimadores de MQO, incluindo a inexistência de viés e a eficiência.

O modelo de regressão múltipla ainda é o veículo mais extensamente usado da análise empírica em economia e em outras ciências sociais. Igualmente, o método de mínimos quadrados ordinários é popularmente usado para estimar os parâmetros do modelo de regressão múltipla.

3.1 FUNCIONALIDADE DA REGRESSÃO MÚLTIPLA

Modelo com Duas Variáveis Independentes

Iniciaremos com alguns exemplos simples para mostrar como a análise de regressão múltipla pode ser usada para resolver problemas que não podem ser resolvidos pela regressão simples.

O primeiro exemplo é uma variação simples da equação do salário introduzida no Capítulo 2 para obter o efeito da educação sobre o salário-hora:

$$\text{salárioh} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u, \quad (3.1)$$

em que exper representa anos de experiência no mercado de trabalho. Assim, salárioh é determinado por duas variáveis explicativas ou independentes, educação e experiência, e por outros fatores não-observados, contidos em u . Basicamente, ainda estamos interessados no efeito de educ sobre salárioh , mantendo fixos todos os outros fatores que afetam salárioh ; isto é, estamos interessados no parâmetro β_1 .

Comparada com uma análise de regressão simples que relaciona salárioh a educ , a equação (3.1) remove, efetivamente, exper do termo erro e a coloca explicitamente na equação. Como exper aparece na equação, seu coeficiente, β_2 , mede o efeito *ceteris paribus* de exper sobre salárioh , que também é de algum interesse.

Não surpreendentemente, assim como na regressão simples, teremos de fazer hipóteses sobre como u , em (3.1), está relacionado às variáveis independentes, educ e exper . Entretanto, como vemos na Seção 3.2, há uma coisa da qual podemos estar seguros: visto que (3.1) contém a experiência de modo explícito, seremos capazes de mensurar o efeito da educação sobre o salário horário, mantendo a experiência fixa. Na análise de regressão simples — que coloca exper no termo erro —, teríamos de assumir que experiência é não-correlacionada com educação, uma hipótese tenua.

Como segundo exemplo, considere o problema de explicar o efeito do gasto público por estudante (gasto) sobre a nota média padronizada (notmed) do ensino médio. Suponha que a nota média dependa do gasto público, da renda familiar média (rendfam) e de outros fatores não-observáveis:

$$\text{notmed} = \beta_0 + \beta_1 \text{gasto} + \beta_2 \text{rendfam} + u. \quad (3.2)$$

Para o propósito de análise da política governamental, o coeficiente de interesse é β_1 , o efeito *ceteris paribus* de gasto sobre notmed . Ao incluir rendfam explicitamente no modelo, somos capazes de controlar seu efeito sobre notmed . Isso é provavelmente importante, pois a renda familiar média tende a estar correlacionada com o gasto público por estudante: os níveis de gasto público são, freqüentemente, determinados tanto por impostos locais sobre a propriedade como sobre a renda. Na análise de regressão simples, rendfam estaria incluída no termo erro, que estaria provavelmente correlacionado com gasto , fazendo com que o estimador de β_1 de MQO fosse viesado no modelo de duas variáveis.

Nesses dois exemplos similares, mostramos que outros fatores observáveis, além da variável de interesse primordial [educ na equação (3.1) e gasto na equação (3.2)] podem ser incluídos em um modelo de regressão. Em geral, podemos escrever um modelo com duas variáveis independentes como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad (3.3)$$

em que β_0 é o intercepto, β_1 mede a variação em y com relação a x_1 , mantendo fixos outros fatores, e β_2 mede a variação em y com relação a x_2 , mantendo outros fatores fixos.

A análise de regressão múltipla também é útil para generalizar relações funcionais entre variáveis. Como exemplo, suponha que o consumo da família (cons) é uma função quadrática da renda familiar (rend):

$$cons = \beta_0 + \beta_1 rend + \beta_2 rend^2 + u, \quad (3.4)$$

em que u contém outros fatores que afetam o consumo. Nesse modelo, o consumo depende somente de um fator observado, a renda; desse modo, pareceria que ele pode ser tratado dentro do arcabouço da regressão simples. No entanto, esse modelo está fora do padrão da regressão simples, porque ele contém duas funções da renda, $rend$ e $rend^2$ (e, portanto, três parâmetros: β_0 , β_1 e β_2). Portanto, a função consumo é facilmente escrita como modelo de regressão com duas variáveis, e fazendo $x_1 = rend$ e $x_2 = rend^2$.

Mecanicamente, não há *nenhuma* diferença em usar o método de mínimos quadrados ordinários (introduzido na Seção 3.2) para estimar equações tão diferentes como (3.1) e (3.4). Cada uma dessas equações pode ser escrita como (3.3), que é tudo o que importa para os cálculos. Há, entretanto, uma diferença importante em como interpretar os parâmetros. Na equação (3.1), β_1 é o efeito *coeteris paribus* de *educ* sobre *salárioh*. O parâmetro β_1 não tem tal interpretação em (3.4). Em outras palavras, não faz sentido medir o efeito de *rend* sobre *cons* mantendo, ao mesmo tempo, *rend*² fixo, porque se *rend* varia, então *rend*² deve variar! Em vez disso, a variação no consumo com respeito à variação na renda — a propensão marginal a consumir — é aproximada por

$$\frac{\Delta cons}{\Delta rend} \approx \beta_1 + 2\beta_2 rend$$

Para os cálculos necessários quanto a derivação dessa equação, veja o Apêndice A, disponível na página do livro, no site www.thomsonlearning.com.br. Em outras palavras, o efeito marginal da renda sobre o consumo depende tanto de β_2 como de β_1 e do nível de renda. Esse exemplo mostra que, em qualquer aplicação particular, as definições das variáveis independentes são cruciais. Mas, para o desenvolvimento teórico da regressão múltipla, podemos ser vagos com relação a tais detalhes. No Capítulo 6 estudaremos exemplos como esse de forma mais completa.

No modelo com duas variáveis independentes, a hipótese fundamental sobre como u está relacionado a x_1 e x_2 é

$$E(u|x_1, x_2) = 0. \quad (3.5)$$

A interpretação da condição (3.5) é similar à interpretação da hipótese RLS.3 da análise de regressão simples. Ela significa que, para qualquer valor de x_1 e x_2 na população, o fator não-observável médio é igual a zero. Assim como na regressão simples, a parte importante da hipótese é que o valor esperado de u é o mesmo para todas as combinações de x_1 e x_2 ; dizer que esse valor comum é zero está longe de ser uma hipótese, desde que o intercepto β_0 esteja incluído no modelo (veja a Seção 2.1).

Como podemos interpretar a hipótese de média condicional zero no exemplo anterior? Na equação (3.1), a hipótese é $E(u|educ, exper) = 0$. Isso implica que outros fatores que afetam *salárioh* não estão, em média, relacionados a *educ* e *exper*. Portanto, se entendermos que aptidão inata é parte de u , então precisaremos que os níveis médios de aptidão sejam os mesmos em todas as combinações de educação e experiência na população que trabalha. Isso pode ou não ser verdadeiro, mas, como vemos na Seção 3.3, essa é a questão que precisamos fazer a fim de determinar se o método de mínimos quadrados ordinários produz estimadores não-viesados.

O exemplo que mede o desempenho dos estudantes [equação (3.2)] é similar à equação do salário. A hipótese de média condicional zero é $E(u|gasto, rendfam) = 0$, o que significa que os outros fato-

res que afetam as notas — características das escolas e dos estudantes — são, em média, não-relacionados aos gastos públicos por estudante e à renda familiar média.

Um modelo simples para explicar as taxas de homicídio nas cidades (*taxahom*) em termos da probabilidade de condenação (*prcond*) e da duração média da sentença (*sentmed*) é

$$taxahom = \beta_0 + \beta_1 prcond + \beta_2 sentmed + u.$$

Que fatores estão contidos em u ? Você entende ser provável que a hipótese (3.5) se mantenha?

Quando aplicada à função quadrática do consumo em (3.4), a hipótese da média condicional zero tem uma interpretação ligeiramente diferente. A equação (3.5), escrita literalmente, é $E(u|rend, rend^2) = 0$. Como $rend^2$ é conhecido quando se conhece $rend$, é redundante incluir $rend^2$ na esperança: $E(u|rend, rend^2) = 0$ é o mesmo que $E(u|rend) = 0$. Não há problema em colocar $rend^2$ junto com $rend$ na esperança quando expressamos a hipótese, mas $E(u|rend) = 0$ é mais conciso.

Modelo com k Variáveis Independentes

Como estamos no contexto da regressão múltipla, não há necessidade de ficarmos com duas variáveis independentes. A análise de regressão múltipla permite que muitos fatores observados afetem y . No exemplo do salário, poderíamos também incluir semanas de treinamento de trabalho, anos de permanência com o empregador atual, medidas de aptidão e mesmo variáveis demográficas, como o número de irmãos ou a educação da mãe. No exemplo do gasto público por estudante, poderiam ser incluídos variáveis adicionais que medissem a qualidade dos professores e o tamanho das escolas.

O **modelo de regressão linear múltipla** geral (também chamado modelo de regressão múltipla) pode ser escrito, na população, como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u, \quad (3.6)$$

onde β_0 é o **intercepto**, β_1 é o parâmetro associado a x_1 , β_2 é o parâmetro associado a x_2 , e assim por diante. Como há k variáveis independentes e um intercepto, a equação (3.6) contém $k + 1$ parâmetros (desconhecidos) populacionais. Para simplificar, algumas vezes vamos nos referir aos outros parâmetros que não o intercepto como **parâmetros de inclinação**, ainda que, literalmente, nem sempre eles tenham esse significado. [Veja a equação (3.4), em que nenhum dos parâmetros, β_1 e β_2 , é, por si mesmo, uma inclinação, mas juntos determinam a inclinação da relação entre o consumo e a renda.]

A terminologia da regressão múltipla é similar àquela da regressão simples e é apresentada na Tabela 3.1. Exatamente como na regressão simples, a variável u é o **termo erro** ou **perturbação**. Ele contém outros fatores, além de x_1, x_2, \dots, x_k , que afetam y . Não importa quantas variáveis explicativas incluímos em nosso modelo, pois sempre haverá fatores que não podemos incluir, e eles estão contidos, coletivamente, em u .

Tabela 3.1

Terminologia para a Regressão Múltipla

y	x_1, x_2, \dots, x_k
Variável Dependente	Variáveis Independentes
Variável Explicada	Variáveis Explicativas
Variável de Resposta	Variáveis de Controle
Variável Prevista	Variáveis Previsoras
Regressando	Regressores

Ao aplicar o modelo de regressão múltipla geral, devemos saber como interpretar os parâmetros. Agora e nos capítulos subsequentes vamos adquirir bastante prática, mas é útil, neste ponto, relembrarmos algumas coisas que já sabemos. Suponha que os salários (*salário*) dos diretores executivos estejam relacionados às vendas das empresas (*vendas*) e à permanência dos diretores executivos nas empresas (*permceo*) pela equação

$$\log(\text{salário}) = \beta_0 + \beta_1 \log(\text{vendas}) + \beta_2 \text{permceo} + \beta_3 \text{permceo}^2 + u. \quad (3.7)$$

Essa equação enquadra-se no modelo de regressão múltipla (com $k = 3$) ao definirmos $y = \log(\text{salário})$, $x_1 = \log(\text{vendas})$, $x_2 = \text{permceo}$ e $x_3 = \text{permceo}^2$. Como sabemos do Capítulo 2, o parâmetro β_1 é a elasticidade (*ceteris paribus*) de *salário* em relação a *vendas*. Se $\beta_3 = 0$, então 100 β_2 é, aproximadamente, o aumento percentual *ceteris paribus* em *salário* quando *permceo* aumenta em um ano. Quando $\beta_3 \neq 0$, o efeito de *permceo* sobre *salário* é mais complicado. Postergaremos até o Capítulo 6 um tratamento detalhado de modelos com termos quadráticos.

A equação (3.7) fornece um lembrete importante sobre a análise de regressão múltipla. O termo “linear” na expressão “modelo de regressão linear múltipla” significa que a equação (3.6) é linear nos parâmetros, β_j . A equação (3.7) é um exemplo de modelo de regressão múltipla que, ao mesmo tempo, é linear nos β_j e é uma relação não-linear entre *salário* e as variáveis *vendas* e *permceo*. Muitas aplicações da regressão múltipla envolvem relações não-lineares entre as variáveis subjacentes.

É fácil expressar a hipótese essencial para o modelo de regressão múltipla geral em termos de uma esperança condicional:

$$E(u|x_1, x_2, \dots, x_k) = 0. \quad (3.8)$$

No mínimo, a equação (3.8) requer que todos os fatores no termo erro não-observado sejam não-correlacionados com as variáveis explicativas. Ela também significa que consideramos corretamente a relação funcional entre as variáveis explicada e as explicativas. Qualquer problema que faça com que u seja correlacionado com qualquer variável independente fazem com que (3.8) não seja válida. Na Seção 3.3 mostraremos que a hipótese (3.8) implica que o método MQO é não-viesado e derivaremos o viés que surge quando uma variável-chave for omitida da equação. Nos capítulos 15 e 16 estudaremos outras razões que podem fazer com que (3.8) não seja válida e mostraremos o que pode ser feito nesses casos.

3.2 MECÂNICA E INTERPRETAÇÃO DOS MÍNIMOS QUADRADOS ORDINÁRIOS

Vamos resumir, agora, algumas características computacionais e algébricas do método de mínimos quadrados ordinários, quando ele se aplica a um conjunto particular de dados. Discutiremos também como interpretar a equação estimada.

Obtenção das Estimativas de MQO

Vamos considerar, primeiramente, a estimação do modelo com duas variáveis independentes. A equação de MQO estimada é escrita de forma similar ao caso da regressão simples:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2, \quad (3.9)$$

onde $\hat{\beta}_0$ é a estimativa de β_0 , $\hat{\beta}_1$ é a estimativa de β_1 , e $\hat{\beta}_2$ é a estimativa de β_2 . Porém, como obtemos $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$? O método de **mínimos quadrados ordinários** escolhe as estimativas que minimizam a soma dos resíduos quadrados. Isto é, dadas n observações de y , x_1 e x_2 , $\{(x_{i1}, x_{i2}, y_i) : i = 1, 2, \dots, n\}$, as estimativas $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$ são escolhidas, simultaneamente, para fazer com que a expressão

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2 \quad (3.10)$$

tenha o menor tamanho possível.

A fim de entender o que o método MQO está fazendo, é importante dominar o significado da indexação das variáveis independentes em (3.10). As variáveis independentes têm, aqui, dois subscritos: i seguido de 1 ou 2. O subscrito i refere-se ao número da observação. Assim, a soma em (3.10) contempla todas as observações de $i = 1$ a n . O segundo índice é simplesmente um método para distinguir as diferentes variáveis independentes. No exemplo que relaciona *salário_i* a *educ* e *exper*, $x_{i1} = \text{educ}_i$ é a educação formal da pessoa i na amostra, e $x_{i2} = \text{exper}_i$ é a experiência da pessoa i . A soma dos resíduos quadra-

dos na equação (3.10) é $\sum_{i=1}^n (\text{salário}_i - \hat{\beta}_0 - \hat{\beta}_1 \text{educ}_i - \hat{\beta}_2 \text{exper}_i)^2$. No que vem a seguir, o subscrito i é reservado para indexar o número da observação. Se escrevermos x_{ij} , então isso significa a i -ésima observação da j -ésima variável independente. (Alguns autores preferem mudar a ordem do número da observação e do número da variável, de modo que x_{1i} é a observação i da variável um. Mas isso é apenas um problema de gosto notacional.)

No caso geral com k variáveis independentes, procuramos estimar $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ na equação

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (3.11)$$

As $k + 1$ estimativas de MQO delas foram escolhidas para minimizar a soma dos resíduos quadrados:

$$\sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik})^2. \quad (3.12)$$

Esse problema de minimização pode ser resolvido usando cálculo multivariado (veja o Apêndice 3A, disponível no site da Thomson). Isso leva a $k + 1$ equações lineares com $k + 1$ estimadores $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ desconhecidos:

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ \sum_{i=1}^n x_{i2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ &\vdots \\ &\vdots \\ &\vdots \\ \sum_{i=1}^n x_{ik} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0. \end{aligned} \quad (3.13)$$

Essas equações são,, chamadas freqüentemente de **condições de primeira ordem** de MQO. Assim, como no modelo de regressão simples da Seção 2.2, as condições de primeira ordem de MQO podem ser obtidas pelo método dos momentos: sob a hipótese (3.8), $E(u) = 0$ e $E(x_j u) = 0$, onde $j = 1, 2, \dots, k$. As equações em (3.13) são contrapartidas amostrais desses momentos da população, embora tenhamos omitido a divisão pelo tamanho da amostra n .

Mesmo para n e k de tamanhos moderados, resolver as equações em (3.13) fazendo os cálculos manualmente é tedioso. Não obstante, computadores modernos que processam programas padrões de estatística e econometria podem resolver essas equações com n e k grandes muito rapidamente.

Há somente um pequeno aviso: devemos assumir que as equações em (3.13) podem ser resolvidas *unicamente* para os $\hat{\beta}_j$. Por enquanto, assumimos apenas isso, como é usualmente o caso em modelos bem definidos. Na Seção 3.3 formulamos a hipótese necessária para a existência de estimativas de MQO únicas (veja a hipótese RLM.4).

Como na análise de regressão simples, a equação (3.11) é chamada **reta de regressão de MQO** ou a **função de regressão amostral (FRA)**. Chamaremos $\hat{\beta}_0$ a **estimativa de intercepto de MQO** e $\hat{\beta}_1, \dots, \hat{\beta}_k$ de **estimativas de inclinação de MQO** (correspondentes às variáveis independentes x_1, x_2, \dots, x_k).

A fim de indicar que uma regressão de MQO foi computada, escreveremos a equação (3.11) com y e x_1, \dots, x_k substituídos pelos seus nomes de variável (tais como *salário*, *educ* e *exper*), ou diremos que “rodamos uma regressão de MQO de y sobre x_1, x_2, \dots, x_k ou que regredimos y sobre x_1, x_2, \dots, x_k ”. Essas expressões são modos de dizer que o método de mínimos quadrados ordinários foi usado para obter a equação de MQO (3.11). A não ser que afirmemos explicitamente, sempre estaremos um intercepto juntamente com as inclinações.

Interpretação da Equação de Regressão de MQO

Mais importante que os detalhes subjacentes à computação dos $\hat{\beta}_j$ é a *interpretação* da equação estimada. Iniciaremos com o caso de duas variáveis independentes:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \quad (3.14)$$

O intercepto $\hat{\beta}_0$ na equação (3.14) é o valor previsto de y quando $x_1 = 0$ e $x_2 = 0$. Às vezes, colocar x_1 e x_2 iguais a zero é um cenário interessante; em outros casos, isso não fará sentido. Não obstante, para obter uma previsão de y a partir da reta de regressão de MQO, o intercepto sempre é necessário, como (3.14) deixa claro.

As estimativas $\hat{\beta}_1$ e $\hat{\beta}_2$ têm interpretações de **efeito parcial**, ou *ceteris paribus*. Da equação (3.14), temos

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1 + \hat{\beta}_2\Delta x_2,$$

de modo que podemos obter a variação prevista em y dadas as variações em x_1 e x_2 . (Observe que o intercepto não tem nada a ver com as variações em y .) Em particular, quando x_2 é mantido fixo, de modo que $\Delta x_2 = 0$, então

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1,$$

mantendo x_2 fixo. O ponto fundamental é que, ao incluir x_2 no nosso modelo, obtemos um coeficiente de x_1 com uma interpretação *ceteris paribus*. Essa é a razão de a análise de regressão múltipla ser tão útil. Semelhantemente,

$$\Delta\hat{y} = \hat{\beta}_2\Delta x_2,$$

mantendo x_1 fixo.

EXEMPLO 3.1

(Determinantes da Nota Média em Curso Superior nos Estados Unidos)

As variáveis do arquivo GPA1.RAW incluem a nota média em um curso superior (*nmgrad*), a nota média do ensino médio (*nmem*) e a nota do teste de avaliação de conhecimentos para ingresso em curso superior (*tac*) para uma amostra de 141 estudantes de uma grande universidade dos Estados Unidos; tanto *nmgrad* como *nmem* estão baseados em uma escala de quatro pontos. Obtemos a seguinte reta de regressão de MQO para estimar *nmgrad* a partir de *nmem* e *tac*:

$$nmgrad = 1,29 + 0,453 nmem + 0,0094 tac. \quad (3.15)$$

Como interpretamos essa equação? Primeiro, o intercepto de 1,29 é o valor previsto de *nmgrad* se tanto *nmem* como *tac* forem iguais a zero. Como ninguém que freqüenta um curso superior teve nota média no ensino médio igual a zero ou uma nota no teste de ingresso no curso superior igual a zero, o intercepto nessa equação não é, por si mesmo, significativo.

As estimativas mais interessantes são os coeficientes de inclinação de *nmem* e *tac*. Como esperado, há uma relação parcial positiva entre *nmgrad* e *nmem*: mantendo *tac* fixo, um ponto adicional em *nmem* está associado a 0,453 de um ponto em *nmgrad*, ou quase meio ponto. Em outras palavras, se escolhermos dois estudantes, A e B, e esses estudantes tiverem a mesma nota *tac*, mas *nmem* do estudante A é um ponto maior que a *nmem* do estudante B, prevemos que o estudante A tem *nmgrad*

EXEMPLO 3.1 (continuação)

0,453 maior que *nmgrad* do estudante B. (Isso não diz nada sobre quaisquer duas pessoas reais, mas é a nossa melhor previsão.)

O sinal *tac* implica que, mantendo *nmem* fixo, uma variação de 10 pontos na nota em *tac* — uma variação muito grande, visto que a nota média na amostra é de cerca de 24, com um desvio-padrão menor que três — afeta *nmgrad* em menos de um décimo de um ponto. Esse é um efeito pequeno e sugere que, uma vez considerada o *nmem*, a nota do *tac* não é um forte previsor de *nmgrad*. (Naturalmente, há muitos outros fatores que contribuem para *nmgrad*, mas aqui estamos enfatizamos as estatísticas disponíveis de estudantes do ensino médio.) Posteriormente, após discutirmos a inferência estatística, mostraremos que o coeficiente de *tac* não é somente pequeno na prática, mas ele também é estatisticamente não significante.

Se colocarmos o foco na análise de regressão simples relacionando somente *nmgrad* e *tac*, obtemos

$$nmgrad = 2,40 + 0,0271 tac;$$

assim, o coeficiente *tac* é quase três vezes maior que a estimativa em (3.15). No entanto, essa equação *não* nos permite comparar duas pessoas com o mesmo *nmem*; ela corresponde a um experimento diferente. Mais adiante, falaremos mais sobre as diferenças entre as regressões múltipla e simples.

O caso com mais de duas variáveis independentes é similar. A reta de regressão de MQO é

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (3.16)$$

Escrita em termos de variações,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + \dots + \hat{\beta}_k \Delta x_k. \quad (3.17)$$

O coeficiente de x_1 mede a variação em \hat{y} devido a um aumento de uma unidade em x_1 , mantendo todas as outras variáveis independentes fixas. Isto é,

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1, \quad (3.18)$$

mantendo x_2, x_3, \dots, x_k fixos. Assim, controlamos as variáveis x_2, x_3, \dots, x_k ao estimar o efeito de x_1 sobre y . Os outros coeficientes têm uma interpretação similar.

O exemplo seguinte contém três variáveis independentes.

EXEMPLO 3.2**(Equação do Salário Horário)**

Usando as 526 observações de trabalhadores do arquivo WAGE1.RAW, incluímos *educ* (anos de educação formal), *exper* (anos de experiência no mercado de trabalho) e *perm* (anos com o empregador atual) na equação que explica *log(salárioh)*. A equação estimada é

EXEMPLO 3.2 (continuação)

$$\log(\text{salárioh}) = 0,284 + 0,092 \text{ educ} + 0,0041 \text{ exper} + 0,022 \text{ perm.} \quad (3.19)$$

Como no caso da regressão simples, os coeficientes têm uma interpretação de percentagem. A única diferença é que eles também têm uma interpretação *ceteris paribus*. O coeficiente 0,092 significa que, mantendo *exper* e *perm* fixos, um ano a mais de educação formal aumenta o valor esperado de $\log(\text{salárioh})$ em 0,092, o que se traduz em um aumento aproximado de 9,2% [100(0,092)] em *salárioh*. Alternativamente, se considerarmos duas pessoas com os mesmos níveis de experiência e permanência no trabalho, o coeficiente de *educ* é a diferença proporcional no salário horário previsto quando seus níveis de educação diferem em um ano. Essa medida de retorno da educação mantém fixos ao menos dois importantes fatores de produtividade; saber se ela é uma boa estimativa do retorno *ceteris paribus* de mais um ano de educação formal requer que estudemos as propriedades estatísticas de MQO (veja a Seção 3.3).

Sobre o Significado de “Manter Outros Fatores Fixos” na Regressão Múltipla

Como a interpretação de efeito parcial dos coeficientes de inclinação na análise de regressão múltipla pode causar alguma confusão; assim, vamos tentar impedir o surgimento desse problema agora.

No Exemplo 3.1, observamos que o coeficiente *tac* mede a diferença prevista em *nmgrad*, mantendo *nmem* fixo. O poder da análise de regressão múltipla é que ela proporciona uma interpretação *ceteris paribus* mesmo que os dados não sejam coletados de uma maneira *ceteris paribus*. Ao dar ao coeficiente de *tac* uma interpretação de efeito parcial, pode parecer que, realmente, saímos a campo e extraímos amostras compostas de pessoas com a mesma *nmem* e, possivelmente, com diferentes notas do *tac*. Isso não é verdade. Os dados são uma amostra aleatória de uma universidade grande: não há restrições colocadas sobre os valores amostrais de *nmem* ou *tac* na obtenção dos dados. De fato, raramente temos o luxo de manter certas variáveis fixas na obtenção de nossa amostra. Se pudéssemos coletar uma amostra de indivíduos com a mesma *nmem*, então poderíamos realizar uma análise de regressão simples relacionando *nmgrad* a *tac*. A regressão múltipla nos permite, efetivamente, simular essa situação sem restringir os valores de quaisquer variáveis independentes.

O poder que a análise de regressão múltipla tem é que ela nos permite fazer, em ambientes não-experimentais, o que os cientistas naturais são capazes de fazer em um ambiente controlado de laboratório: manter outros fatores fixos.

Variação de mais de uma Variável Independente Simultaneamente

Às vezes, queremos variar mais que uma variável independente ao mesmo tempo para encontrar o efeito resultante sobre a variável dependente. Isso é facilmente feito usando a equação (3.17). Por exemplo, na equação (3.19), podemos obter o efeito estimado sobre *salárioh* quando um indivíduo permanece na mesma empresa por mais um ano: ambos *exper* (experiência geral da força de trabalho) e *perm* aumentam em um ano. O efeito total (mantendo *educ* fixo) é

$$\Delta \log(\text{salárioh}) = 0,0041 \Delta \text{exper} + 0,022 \Delta \text{perm} = 0,0041 + 0,022 = 0,0261,$$

ou cerca de 2,6%. Como *exper* e *perm* aumentam, cada um, em um ano, apenas somamos os coeficientes de *exper* e *perm* e multiplicamos por 100 para converter o efeito em uma percentagem.

Valores Estimados e Resíduos de MQO

Após obter a reta de regressão de MQO (3.11), podemos obter um *valor ajustado* ou *preditivo* para cada observação. Para a observação i , o valor ajustado é simplesmente

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}, \quad (3.20)$$

que é exatamente o valor previsto obtido inserindo os valores das variáveis independentes da observação i na equação (3.11). Ao obter os valores estimados, não devemos esquecer do intercepto; de outro modo, a resposta pode ser muito equivocada. Como exemplo, se em (3.15) $nmem_i = 3,5$ e $tac_i = 24$, $nmgrad_i = 1,29 + 0,453(3,5) + 0,0094(24) = 3,101$ (arredondado em três casas após o decimal).

Normalmente, para qualquer observação i , o valor real y_i não se iguala ao valor previsto, \hat{y}_i . O método de MQO minimiza o erro quadrado médio de previsão, que não diz nada sobre o erro de previsão de qualquer observação específica. O **resíduo** da observação i é definido exatamente como no caso da regressão simples,

$$\hat{u}_i = y_i - \hat{y}_i. \quad (3.21)$$

Há um resíduo para cada observação. Se $\hat{u}_i > 0$, então \hat{y}_i está abaixo de y_i , o que significa que, para essa observação, y_i é subestimado. Se $\hat{u}_i < 0$, então $y_i < \hat{y}_i$, e y_i é superestimado.

Os valores estimados de MQO e os resíduos têm algumas propriedades importantes que são extensões imediatas do caso da variável única:

1. A média amostral dos resíduos é zero.
2. A covariância amostral entre cada variável independente e os resíduos de MQO é zero. Conseqüentemente, a covariância amostral entre os valores estimados de MQO e os resíduos de MQO é zero.
3. O ponto $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k, \bar{y})$ está sempre sobre a reta de regressão de MQO: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \dots + \hat{\beta}_k \bar{x}_k$.

As duas primeiras propriedades são consequências imediatas do conjunto de equações usadas para obter as estimativas de MQO. A primeira equação em (3.13) diz que a soma dos resíduos é zero. As equações restantes são da forma $\sum_{i=1}^n x_{ij} \hat{u}_i = 0$, implicando que cada variável independente tem covariância amostral zero com \hat{u}_i . A Propriedade 3 decorre imediatamente da Propriedade 1.

No Exemplo 3.1, a reta estimada de MQO que explica *nmgrad* em termos de *nmem* é

$$nmgrad = 1,29 + 0,453 nmem + 0,0094 tac.$$

Se *nmem* médio é de cerca de 3,4 e a nota média do *tac* está em torno de 24,2, qual é o *nmgrad* médio na amostra?

Interpretação de “Parcialidade” da Regressão Múltipla

Ao aplicar o método MQO, não precisamos saber as fórmulas explícitas dos $\hat{\beta}_j$ que solucionam o sistema de equações em (3.13). Entretanto, para certas derivações, precisamos de fórmulas explícitas dos $\hat{\beta}_j$. Essas fórmulas também ajudam a esclarecer o funcionamento de MQO.

Considere, uma vez mais, o caso com $k = 2$ variáveis independentes, em que $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$. Para uma ideia mais concreta, vamos enfatizar $\hat{\beta}_1$. Um modo de expressar $\hat{\beta}_1$ é

$$\hat{\beta}_1 = \left(\sum_{i=1}^n \hat{r}_{i1} y_i \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right), \quad (3.22)$$

onde os \hat{r}_{ij} são os resíduos de MQO de uma regressão simples de x_1 sobre x_2 , usando a amostra à mão. Regredimos nossa primeira variável independente, x_1 , sobre nossa segunda variável independente, x_2 , e, em seguida, obtemos os resíduos (y não desempenha nenhum papel aqui). A equação (3.2) mostra que podemos, portanto, fazer uma regressão simples de y sobre \hat{r}_{i1} para obter $\hat{\beta}_1$. (Observe que os resíduos \hat{r}_{ij} têm uma média amostral zero, e assim $\hat{\beta}_1$ é a estimativa de inclinação usual da regressão simples.)

A representação da equação (3.22) dá outra demonstração da interpretação do efeito parcial de $\hat{\beta}_1$. Os resíduos \hat{r}_{i1} são a parte de x_{i1} que é não-correlacionada com x_{i2} . Outro modo de dizer isso é que \hat{r}_{i1} é x_{i1} após o efeito de x_{i2} ter sido *isolado*, ou *deduzido*. Assim, $\hat{\beta}_1$ mede a relação amostral entre y e x_1 após x_2 ter sido imparcializado.

Na análise de regressão simples, não há parcialização de outras variáveis, porque outras variáveis não estão incluídas na regressão. O Problema 3.17 o conduzirá, passo a passo, pelo processo de parcialização usando os dados de salários do Exemplo 3.2. Para propósitos práticos, o importante é que $\hat{\beta}_1$, na equação $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, mede a variação em y , dado um aumento de uma unidade em x_1 , mantendo x_2 fixo.

No modelo geral com k variáveis explicativas, $\hat{\beta}_1$ pode ainda ser escrito como na equação (3.22), mas os resíduos \hat{r}_{i1} vêm da regressão de x_1 sobre x_2, \dots, x_k . Assim, $\hat{\beta}_1$ mede o efeito de x_1 sobre y após x_2, \dots, x_k terem sido isolados ou deduzidos.

Comparação das Estimativas das Regressões Simples e Múltipla

Há dois casos especiais em que tanto a regressão simples de y sobre x_1 como a regressão de y sobre x_1 e x_2 produzirão a mesma estimativa de MQO de x_1 . Para maior precisão, escreva a regressão simples de y sobre x_1 como $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$, e escreva a regressão múltipla como $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$. Sabemos que o coeficiente da regressão simples $\tilde{\beta}_1$ não se iguala, geralmente, ao coeficiente da regressão múltipla $\hat{\beta}_1$. Acontece que há uma relação simples entre $\tilde{\beta}_1$ e $\hat{\beta}_1$, que permite comparações interessantes entre as regressões simples e múltipla:

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1,$$

em que $\tilde{\delta}_1$ é o coeficiente de inclinação da regressão simples de x_{i2} sobre x_{i1} , $i = 1, \dots, n$. Essa equação mostra como $\tilde{\beta}_1$ difere do efeito parcial de x_1 sobre \hat{y} . O termo que pode causar confusão é o efeito parcial de x_2 sobre \hat{y} multiplicado pela inclinação da regressão amostral de x_2 sobre x_1 . (Veja a Seção 3A.4, no apêndice do capítulo, disponível no site da Thomson, para uma verificação mais geral.)

A relação entre $\tilde{\beta}_1$ e $\hat{\beta}_1$ também mostra que há dois casos distintos em que eles são iguais:

1. O efeito parcial de x_2 sobre \hat{y} é zero na amostra, isto é, $\hat{\beta}_2 = 0$.
2. x_1 e x_2 são não-correlacionados na amostra, isto é, $\tilde{\delta}_1 = 0$.

Ainda que as estimativas das regressões múltipla e simples quase nunca sejam idênticas, podemos usar a fórmula anterior para caracterizar o motivo pelo qual elas deveriam ser muito diferentes ou bastante similares. Por exemplo, se $\hat{\beta}_2$ é pequeno, deveríamos esperar que as estimativas das regressões múltipla e simples de β_1 sejam semelhantes. No Exemplo 3.1, a correlação amostral entre *nmem* e *tac* é de cerca de 0,346, o que não é uma correlação trivial. Porém, o coeficiente de *tac* é razoavelmente pequeno. Não é surpreendente descobrir que a regressão de *nmgrad* sobre *nmem* produz uma estimativa de inclinação de 0,482, não é muito diferente da estimativa de 0,453 em (3.15).

EXEMPLO 3.3

[Participação nos Planos de Pensão 401(k)]

Vamos usar os dados do arquivo em 401K.RAW para estimar o efeito de uma taxa de contribuição para um plano (*taxcont*) sobre a taxa de participação (*taxap*) dos trabalhadores nos planos de pensão de contribuição definida existentes nos Estados Unidos. A taxa de contribuição é a quantidade com a qual a firma contribui para um fundo de trabalhadores, para cada dólar de contribuição do trabalhador (até certo limite); assim, *taxcont* = 0,75 significa que a firma contribui com 75 centavos de dólar para cada dólar contribuído pelo trabalhador. A taxa de participação é a percentagem de trabalhadores habilitados a ter uma conta no plano de pensão. A variável *idade* é a idade do plano de pensão. Há 1.534 planos no banco de dados, a *taxap* média é 87,36, a *taxcont* média é 0,732 e a *idade* média é 13,2.

Regredindo *taxap* sobre *taxcont* e *idade* resulta na equação

$$\hat{taxap} = 80,12 + 5,52 taxcont + 0,243 idade. \quad (3.23)$$

Assim, *taxcont* e *idade* têm os efeitos esperados. O que aconteceria se não controlássemos a variável *idade*? O efeito estimado de *idade* não é trivial, e portanto poderíamos esperar uma variação grande no efeito estimado de *taxcont* se *idade* fosse excluída da regressão. Entretanto, a regressão simples de *taxap* sobre *taxcont* produz $\hat{taxap} = 83,08 + 5,86 taxcont$. A estimativa de regressão simples do efeito de *taxcont* sobre *taxap* é, claramente, diferente da estimativa de regressão múltipla, mas a diferença não é muito grande. (A *taxap* é, claramente, diferente da estimativa de regressão múltipla, mas a diferença não é muito grande.) Isso pode ser explicado pelo fato de a correlação amostral entre *taxcont* e *idade* ser somente de 0,12.

No caso com k variáveis independentes, a regressão simples de y sobre x_1 , e a regressão múltipla de y sobre x_1, x_2, \dots, x_k produzem uma estimativa idêntica de x_1 , somente se: (1) os coeficientes de MQO de x_2 até x_k forem todos zero ou (2) x_1 for não-correlacionado com *cada um* dos x_2, \dots, x_k . Na prática, nenhuma dessas possibilidades é muito provável. Porém, se os coeficientes de x_2 até x_k forem pequenos, ou as correlações amostrais entre x_1 e as outras variáveis independentes forem pouco substanciais, então as estimativas das regressões simples e múltiplas do efeito de x_1 sobre y podem ser similares.

Grau de Ajuste

Assim como na regressão simples, podemos definir a **Soma dos Quadrados Total (SQT)**, a **Soma dos Quadrados Explicada (SQE)** e a **Soma dos Quadrados dos Resíduos ou Soma dos Resíduos Quadrados (SQR)** como

$$\text{SQT} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (3.24)$$

$$\text{SQE} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (3.25)$$

$$\text{SQR} = \sum_{i=1}^n \hat{u}_i^2. \quad (3.26)$$

Usando o mesmo argumento utilizado no caso da regressão simples, podemos mostrar que

$$\text{SQT} = \text{SQE} + \text{SQR}. \quad (3.27)$$

Em outras palavras, a variação total em $\{y_i\}$ é a soma das variações totais em $\{\hat{y}_i\}$ e em $\{\hat{u}_i\}$.

Assumindo que a variação total em y não é zero — como é o caso, a não ser que y_i seja constante na amostra —, podemos dividir (3.27) por SQT para obter

$$\text{SQR/SQT} + \text{SQE/SQT} = 1.$$

Exatamente como no caso da regressão simples, o R -quadrado é definido como

$$R^2 = \text{SQE/SQT} = 1 - \text{SQR/SQT}, \quad (3.28)$$

e é interpretado como a proporção da variação amostral em y_i que é explicada pela reta de regressão de MQO. Por definição, R^2 é um número entre zero e um.

Pode-se também mostrar que R^2 é igual ao quadrado do coeficiente de correlação entre os valores reais y_i real e os valores estimados \hat{y}_i ajustado. Isto é,

$$R^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \right)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \right)}. \quad (3.29)$$

[Inserimos a média de \hat{y}_i em (3.29) por coerência com a fórmula do coeficiente de correlação; sabemos que essa média é igual a \bar{y} , porque a média amostral dos resíduos é zero e $y_i = \hat{y}_i + \hat{u}_i$.]

Um fato importante sobre R^2 é que ele nunca diminui, e geralmente aumenta, quando outra variável independente é adicionada à regressão. Esse fato algébrico ocorre por definição, pois a

soma dos resíduos quadrados nunca aumenta quando regressores adicionais são acrescentados ao modelo.

O fato de que R^2 nunca diminui quando *qualquer* variável for adicionada a uma regressão faz dele um instrumento fraco para decidir se uma variável ou diversas variáveis deveriam ser adicionadas ao modelo. O fator que deve determinar se uma variável explicativa pertence a um modelo é se a variável explicativa tem, na *população*, um efeito parcial sobre y diferente de zero. No Capítulo 4, quando cobrirmos a inferência estatística, mostraremos como testar essa hipótese. Veremos também que, quando usado apropriadamente, R^2 permite-nos *testar* um grupo de variáveis com a finalidade de ver se ele é importante para explicar y . Por enquanto, usaremos R^2 como uma medida do grau de ajuste para um dado modelo.

EXEMPLO 3.4

(Determinantes de *nmgrad*)

Da regressão de *nmgrad* que fizemos anteriormente, a equação com R^2 é

$$\begin{aligned} nmgrad &= 1,29 + 0,453 nmem + 0,0094 tac \\ n &= 141, R^2 = 0,176. \end{aligned}$$

Isso significa que *nmem* e *tac* explicam, juntos, cerca de 17,6% da variação em *nmgrad* da amostra de estudantes. Isso pode não parecer uma percentagem alta, mas devemos nos lembrar de que há muitos outros fatores — incluindo formação da família, personalidade, qualidade da educação do ensino médio, afinidade com o curso escolhido — que contribuem para o desempenho dos estudantes. Se *nmem* e *tac* explicassem quase toda a variação em *nmgrad*, então o desempenho no curso superior seria predeterminado pelo desempenho no ensino médio!

EXEMPLO 3.5

(Explicando os Registros de Prisões)

O arquivo CRIME1.RAW contém dados de prisões durante o ano de 1986 e outras informações sobre 2.725 homens nascidos em 1960 ou 1961 na Califórnia. Cada homem na amostra foi preso pelo menos uma vez antes de 1986. A variável *npre86* é o número de vezes que determinado homem foi preso em 1986; ela é zero para muitos homens da amostra (72,29%), e varia de 0 a 12. (A percentagem de homens presos uma vez em 1986 foi de 20,51%.) A variável *pcond* é a proporção (não a percentagem) de prisões anteriores a 1986 que levaram à condenação, *sentmed* é a duração média da sentença cumprida por condenação prévia (zero para muitas pessoas), *ptemp86* são os meses passados na prisão em 1986 e *empr86* é o número de trimestres durante o qual determinado homem ficou empregado em 1986 (de zero a quatro).

Um modelo linear explicando as prisões é

$$npre86 = \beta_0 + \beta_1 pcond + \beta_2 sentmed + \beta_3 ptemp86 + \beta_4 empr86 + u,$$

em que *pcond* é uma variável (*proxy*) da probabilidade de um homem ser condenado por um crime, e *sentmed* é uma medida do rigor esperado da pena, em caso de condenação. A variável *ptemp86* capture o efeito de confinamento do crime: se um indivíduo está na prisão, ele não pode ser preso por um crime fora da prisão. As oportunidades no mercado de trabalho são capturadas grosseiramente por *empr86*.

Primeiro, estimamos o modelo sem a variável *sentmed*. Obtemos

EXEMPLO 3.5 (continuação)

$$\begin{aligned} \hat{np_{re86}} &= 0,712 - 0,150 pcond - 0,034 ptemp86 - 0,104 empr86 \\ n &= 2.725, R^2 = 0,0413. \end{aligned}$$

Essa equação diz que, como um grupo, as três variáveis *pcond*, *ptemp86* e *empr86* explicam cerca de 4,1% da variação em *npre86*.

Cada um dos coeficientes de inclinação de MQO tem o sinal esperado. Um aumento na proporção de condenações diminui o número previsto de prisões. Se aumentarmos *pcond* em 0,50 (um aumento grande na probabilidade de condenação), então, mantendo os outros fatores fixos, $\Delta\hat{np_{re86}} = -0,150(0,50) = -0,075$. Isso pode parecer pouco usual, porque uma prisão não pode ser uma fração. No entanto, podemos usar esse valor para obter a variação prevista das prisões esperadas de um grande grupo de homens. Por exemplo, entre cem homens, a queda esperada de prisões quando *pcond* aumenta em 0,50 é $-7,5$.

Semelhantemente, um período de prisão mais longo leva a um número previsto menor de prisões. De fato, se *ptemp86* aumenta de 0 para 12, as prisões previstas para um determinado homem diminuem em $0,034(12) = 0,408$. Um trimestre a mais no qual o emprego legal é informado diminui as prisões esperadas em 0,104, o que significaria 10,4 prisões entre cem homens.

Se *sentmed* for adicionado ao modelo, sabemos que R^2 aumentará. A equação estimada é

$$\begin{aligned} \hat{np_{re86}} &= 0,707 - 0,151 pcond + 0,0074 sentmed - 0,037 ptemp86 - 0,103 empr86 \\ n &= 2.725, R^2 = 0,0422. \end{aligned}$$

Assim, ao adicionar a variável sentença média, R^2 aumenta de 0,0413 para 0,0422, um efeito praticamente insignificante. O sinal do coeficiente de *sentmed* também é inesperado: ele diz que uma duração mais longa da sentença média aumenta a atividade criminal.

O Exemplo 3.5 merece uma palavra final de cautela. O fato de as quatro variáveis explicativas incluídas na segunda regressão explicarem somente 4,2% da variação em *npre86* não necessariamente significa que a equação é inútil. Ainda que, coletivamente, essas variáveis não expliquem muito da variação nas prisões, é possível que as estimativas de MQO sejam estimativas confiáveis dos efeitos *ceteris paribus* de cada variável independente sobre *npre86*. Como veremos, se esse for o caso, isso não depende, diretamente, do tamanho do R^2 . Em geral, um R^2 baixo indica que é difícil prever resultados individuais sobre *y* com muita precisão, algo que estudaremos com mais detalhes no Capítulo 6. No exemplo da prisão, o R^2 pequeno reflete algo sobre o qual suspeitamos nas ciências sociais: geralmente, é muito difícil prever o comportamento individual.

Regressão através da Origem

Algumas vezes, uma teoria econômica, ou o senso comum, sugere que β_0 deveria ser zero, e por isso devemos mencionar, brevemente, a estimativa de MQO quando o intercepto é zero. Especificamente, vamos agora buscar uma equação da forma

$$\bar{y} = \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2 + \dots + \tilde{\beta}_k x_k, \quad (3.30)$$

em que o símbolo gráfico “~” sobre as estimativas é utilizado para distingui-las das estimativas de MQO obtidas juntamente com o intercepto [como em (3.11)]. Em (3.30), quando $x_1 = 0, x_2 = 0, \dots, x_k = 0$, o

valor previsto é zero. Nesse caso, diz-se que $\tilde{\beta}_1, \dots, \tilde{\beta}_k$ são as estimativas de MQO da regressão de y sobre x_1, x_2, \dots, x_k através da origem.

As estimativas de MQO em (3.30), como sempre, minimizam a soma dos resíduos quadrados, mas com o intercepto igualado a zero. Você deve estar prevenido de que as propriedades de MQO que derivamos anteriormente não se mantêm mais para a regressão através da origem. Em particular, os resíduos de MQO não têm mais uma média amostral zero. Além disso, se R^2 for definido como $1 - SQR/SQT$, em que SQT está dado em (3.24) e SQR é agora $\sum_{i=1}^n (y_i - \tilde{\beta}_1 x_{i1} - \dots - \tilde{\beta}_k x_{ik})^2$, então R^2 pode ser, de fato, negativo. Isso significa que a média amostral, \bar{y} , “explica” mais da variação em y_i do que as variáveis explicativas.

Devemos incluir um intercepto na regressão ou devemos concluir que as variáveis explicativas explicam fracamente y . A fim de sempre ter um R -quadrado não-negativo, alguns economistas preferem calcular R^2 como o quadrado do coeficiente de correlação entre os valores reais e estimados de y , como em (3.29). (Nesse caso, o valor estimado médio deve ser calculado diretamente, já que ele não é mais igual a \bar{y} .) Entretanto, não há um conjunto de regras sobre como calcular o R -quadrado para a regressão através da origem.

Uma desvantagem séria com a regressão através da origem é que, se o intercepto β_0 for diferente de zero no modelo populacional, então os estimadores dos parâmetros de inclinação serão viesados. O viés pode ser severo em alguns casos. O custo de estimar um intercepto quando β_0 é realmente zero é que as variâncias dos estimadores de inclinação de MQO são maiores.

3.3 O VALOR ESPERADO DOS ESTIMADORES DE MQO

Vamos nos voltar, agora, para as propriedades estatísticas do método de MQO, para estimar os parâmetros de um modelo da população subjacente. Nesta seção, derivamos o valor esperado dos estimadores de MQO. Em particular, formulamos e discutimos quatro hipóteses, que são extensões diretas das hipóteses do modelo de regressão simples, sob as quais os estimadores de MQO são estimadores não-viesados dos parâmetros da população. Também obtemos explicitamente o viés em MQO, quando uma variável importante for omitida da regressão.

Você deve lembrar que propriedades estatísticas não têm nada a ver com uma amostra particular, mas sim, mais precisamente, com a propriedade dos estimadores quando a amostragem aleatória é feita repetidamente. Assim, as seções 3.3, 3.4 e 3.5 são um pouco abstratas. Apesar de darmos exemplos de derivação do viés de modelos específicos, não é significativo falar sobre as propriedades estatísticas de um conjunto de estimativas de uma única amostra.

A primeira hipótese que vamos fazer define, simplesmente, o modelo de regressão linear múltipla (RLM).

H I P Ó T E S E R L M . 1 (LINEAR NOS PARÂMETROS)

O modelo na população pode ser escrito como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u, \quad (3.31)$$

em que $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros desconhecidos (constantes) de interesse, e u é um erro aleatório não-observável ou um termo de perturbação aleatória.

A equação (3.31) especifica, formalmente, o **modelo populacional**, algumas vezes chamado **modelo verdadeiro**, para considerar a possibilidade de podermos estimar um modelo diferente de (3.31). A característica fundamental é que o modelo é linear nos parâmetros $\beta_0, \beta_1, \dots, \beta_k$. Como sabemos, (3.31) é bastante flexível, pois y e as variáveis independentes podem ser funções arbitrárias de variáveis subjacentes de interesse, como os logaritmos naturais e os quadrados [veja, por exemplo, a equação (3.7)].

H I P Ó T E S E R L M . 2 (AMOSTRAGEM ALEATÓRIA)

Temos uma amostra aleatória de n observações, $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i); i = 1, 2, \dots, n\}$, do modelo populacional descrito por (3.31).

Às vezes, precisamos escrever a equação de uma observação particular i : para uma observação extraída aleatoriamente da população, temos

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i \quad (3.32)$$

Lembre-se de que i refere-se à observação, enquanto o segundo subscrito em x é o número da variável. Por exemplo, podemos escrever uma equação do salário de diretores executivos para um diretor executivo específico particular i como

$$\log(\text{salário}) = \beta_0 + \beta_1 \log(\text{vendas}_i) + \beta_2 \text{permceo}_i + \beta_3 \text{perceo}_i^2 + u_i \quad (3.33)$$

O termo u_i contém os fatores não-observáveis para o diretor executivo i que afetam seu salário. Nas aplicações, é usualmente mais fácil escrever o modelo na forma populacional, como em (3.31). Ela é menos desordenada e enfatiza que estamos interessados em estimar a relação populacional.

À luz do modelo (3.31), os estimadores de MQO $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ da regressão de y sobre x_1, \dots, x_k são agora considerados estimadores de $\beta_0, \beta_1, \dots, \beta_k$. Vimos, na Seção 3.2, que MQO escolhe as estimativas de uma amostra particular, de modo que os resíduos sejam, em média, iguais a zero e a correlação amostral entre cada variável independente e os resíduos seja zero. Para que MQO seja não-viesado, é preciso que a versão *populacional* dessa condição seja verdadeira.

H I P Ó T E S E R L M . 3 (MÉDIA CONDICIONAL ZERO)

O erro u tem um valor esperado igual a zero, dados quaisquer valores das variáveis independentes. Em outras palavras,

$$E(u|x_1, x_2, \dots, x_k) = 0. \quad (3.34)$$

Uma maneira como a hipótese RLM.3 pode ser violada é quando a relação funcional entre as variáveis explicadas e explicativas está mal-especificada na equação (3.31): por exemplo, se esquece-

mos de incluir o termo quadrático $rend^2$ na função consumo $consumo = \beta_0 + \beta_1 rend + \beta_2 rend^2 + u$ quando estimamos o modelo. Outra forma funcional mal-especificada ocorre quando usamos o nível da variável e, de fato, é o log da variável que aparece no modelo populacional, ou vice-versa. Por exemplo, se o modelo verdadeiro tiver $\log(salário)$ como variável dependente, mas usarmos $salário$ como variável dependente em nossa análise de regressão, então os estimadores serão viesados. Intuitivamente, isso deveria ser muito claro. No Capítulo 9 discutiremos maneiras de detectar a má especificação da forma funcional.

Omitir um fator importante que está correlacionado com qualquer uma das variáveis x_1, x_2, \dots, x_k faz com que a hipótese RLM.3 também não se sustente. Com a análise de regressão múltipla, somos capazes de incluir muitos fatores entre as variáveis explicativas e, por isso, variáveis omitidas são menos prováveis de serem um problema na análise de regressão múltipla do que na análise de regressão simples. Não obstante, em qualquer aplicação, há sempre muitos fatores que, devido à limitação de dados ou à ignorância deles, não somos capazes de incluir. Se acharmos que esses fatores devem ser controlados e que eles estão correlacionados com uma ou mais variáveis independentes, então a hipótese RLM.3 será violada. Posteriormente, derivaremos esse viés.

Há outros modos pelos quais u pode estar correlacionado com uma variável explicativa. No Capítulo 15 discutiremos o problema do erro de medida em uma variável explicativa. No Capítulo 16 cobriremos o problema, conceitualmente mais difícil, em que uma ou mais variáveis explicativas é determinada conjuntamente com y . Vamos postergar nosso estudo desses problemas até que tenhamos um domínio firme da análise de regressão múltipla sob um conjunto ideal de hipóteses.

Quando a hipótese RLM.3 se mantém, dizemos freqüentemente que temos **variáveis explicativas exógenas**. Se x_j for correlacionado com u por alguma razão, então se diz que x_j é uma **variável explicativa endógena**. Os termos “exógena” e “endógena” originaram-se da análise de equações simultâneas (veja Capítulo 16), mas o significado do termo “variável explicativa endógena” evoluiu e passou a incluir qualquer caso em que uma variável explicativa pode estar correlacionada com o termo erro.

A última hipótese de que precisamos para mostrar que MQO é não-viesado assegura que os estimadores de MQO são, realmente, bem definidos. Para a regressão simples, precisamos assumir que a única variável independente não era constante na amostra. A hipótese correspondente para a análise de regressão múltipla é mais complicada.

H I P Ó T E S E R L M . 4 (COLINEARIDADE NÃO PERFEITA)

Na amostra (e, portanto, na população), nenhuma das variáveis independentes é constante, e não há relações lineares exatas entre as variáveis independentes.

A hipótese de colinearidade não perfeita somente diz respeito às variáveis independentes. Estudantes de econometria iniciantes tendem a confundir as hipóteses RLM.4 e RLM.3, de modo que enfatizamos aqui que RLM.4 não diz *nada* sobre a relação entre u e as variáveis explicativas.

A hipótese RLM.4 é mais complicada que sua contrapartida na regressão simples, pois agora devemos examinar as relações entre todas as variáveis independentes. Se uma variável independente em (3.31) é uma combinação linear exata de outras variáveis independentes, dizemos que o modelo sofre de **colinearidade perfeita**, e ele não pode ser estimado por MQO.

É importante observar que a hipótese RLM.4 permite, *de fato*, que as variáveis independentes sejam correlacionadas; elas apenas não podem ser correlacionadas *perfeitamente*. Se não permitíssemos qualquer correlação entre as variáveis independentes, então a regressão múltipla não seria muito útil para a análise econométrica. Por exemplo, no modelo que relaciona notas de estudantes aos gastos com educação e à renda familiar,

$$notmed = \beta_0 + \beta_1 gasto + \beta_2 rendfam + u,$$

esperamos, com certeza, que *gasto* e *rendfam* sejam correlacionados: distritos escolares com rendas familiares médias altas tendem a gastar mais em educação por estudante. De fato, a principal motivação para incluir *rendfam* na equação é que suspeitamos que ela seja correlacionada com *gasto*, e, desse modo, gostaríamos de mantê-la fixa na análise. A hipótese RLM.4 somente exclui a correlação perfeita entre *gasto* e *rendfam* em nossa amostra. Teríamos muito azar se obtivéssemos uma amostra em que os gastos por estudante fossem perfeitamente correlacionados com a renda familiar média. Porém, espera-se que haja alguma correlação — talvez uma quantidade substancial — e certamente ela é permitida.

A maneira mais simples como duas variáveis independentes podem ser perfeitamente correlacionadas é quando uma variável é um múltiplo constante da outra. Isso pode acontecer quando um pesquisador, inadvertidamente, coloca a mesma variável medida em unidades diferentes dentro da equação de regressão. Por exemplo, ao estimar a relação entre consumo e renda, não faz sentido incluir como variáveis independentes a renda mensurada em dólares e a renda mensurada em milhares de dólares. Uma delas é redundante. Que sentido faria manter a renda mensurada em dólares fixa, enquanto a renda mensurada em milhares de dólares varia?

Já sabemos que diferentes funções não-lineares da mesma variável *podem* aparecer entre os regressores. Por exemplo, o modelo $cons = \beta_0 + \beta_1 rend + \beta_2 rend^2 + u$ não viola a hipótese RLM.4: ainda que $x_2 = rend^2$ seja uma função exata de $x_1 = rend$, $rend^2$ não é uma função *linear* de *rend*. Incluir $rend^2$ no modelo é uma maneira útil de generalizar a forma funcional, diferentemente de incluir a renda mensurada em dólares e em milhares de dólares.

O senso comum nos diz para não incluir a mesma variável explicativa medida em diferentes unidades na mesma equação de regressão. Há também maneiras mais sutis de uma variável independente poder ser um múltiplo de outra. Suponha que gostaríamos de estimar uma extensão da função de consumo de elasticidade constante. Poderia parecer natural especificar um modelo tal como

$$\log(cons) = \beta_0 + \beta_1 \log(rend) + \beta_2 \log(rend^2) + u, \quad (3.35)$$

em que $x_1 = \log(rend)$ e $x_2 = \log(rend^2)$. Usando as propriedades básicas do log natural (veja o Apêndice A, disponível no site de Thomson), $\log(rend^2) = 2\log(rend)$. Isto é, $x_2 = 2x_1$, e naturalmente isso é válido para todas as observações na amostra. Isso viola a hipótese RLM.4. Em vez disso, deve-ríamos incluir $[\log(rend)]^2$, e não $\log(rend^2)$, juntamente com $\log(rend)$. Essa é uma extensão prudente do modelo de elasticidade constante, e veremos como interpretar tais modelos no Capítulo 6.

Outra maneira de as variáveis independentes serem perfeitamente colineares ocorre quando uma variável independente pode ser expressa como uma função linear exata de duas ou mais das outras variáveis independentes. Por exemplo, suponha que queremos estimar o efeito dos gastos de campanha sobre os resultados da campanha. Por simplicidade, assuma que cada eleição tem dois candidatos. Seja *votoA* a percentagem de votos do Candidato A; seja *gastoA* os gastos de campanha do Candidato A; seja *gastoB* os gastos de campanha do Candidato B; e seja *totalgasto* os gastos totais de campanha; todas as últimas três variáveis são medidas em dólares. Pode parecer natural especificar o modelo como

$$votoA = \beta_0 + \beta_1 gastoA + \beta_2 gastoB + \beta_3 totalgasto + u, \quad (3.36)$$

a fim de isolar os efeitos dos gastos de cada candidato e da quantidade total de gastos. No entanto, esse modelo viola a hipótese RLM.4, porque $x_3 = x_1 + x_2$ por definição. Tentar interpretar essa equação ao estilo *ceteris paribus* revela o problema. Supõe-se que o parâmetro de β_1 na equação (3.36) mede o efeito de um aumento de um dólar nos gastos do Candidato A sobre os votos do Candidato A, mantendo os gastos do Candidato B e os gastos totais fixos. Isso é uma tolice, pois, se *gastoB* e *totalgasto* forem mantidos fixos, não podemos aumentar *gastoA*.

A solução para a colinearidade perfeita em (3.36) é simples: retire qualquer uma das três variáveis do modelo. Provavelmente, tiraríamos *totalgasto*, e consequentemente o coeficiente de *gastoA* mensuraria o efeito de aumentar os gastos de A sobre a percentagem de votos recebidos por A, mantendo os gastos de B fixos.

O exemplo anterior mostra que a hipótese RLM.4 pode ser violada se não formos cuidadosos ao especificar nosso modelo. Essa hipótese também não se mantém se o tamanho da amostra, n , é muito pequeno em relação ao número de parâmetros que são estimados. No modelo de regressão geral da equação (3.31), há $k + 1$ parâmetros, e RLM.4 não se mantém se $n < k + 1$. Intuitivamente, isso faz sentido: para estimar $k + 1$ parâmetros, necessitamos de pelo menos $k + 1$ observações. Não surpreendentemente, é melhor ter tantas observações quanto possível, algo que veremos em nossos cálculos da variância na Seção 3.4.

No exemplo anterior, se usarmos como variáveis explicativas *gastoA*, *gastoB* e *partA*, em que $partA = 100 \cdot (gastoA/totalgasto)$ é a participação percentual dos gastos totais de campanha feitos pelo Candidato A, isso viola a hipótese RLM.4?

Se o modelo for cuidadosamente especificado e $n \geq k + 1$, a hipótese RLM.4 pode não se manter em casos raros devido a um azar ao coletar a amostra. Por exemplo, em uma equação de salários, com educação e experiência como variáveis, poderíamos obter uma amostra aleatória em que cada indivíduo tivesse exatamente duas vezes mais educação que anos de experiência. Esse cenário faria com que a hipótese RLM.4 falhasse, mas isso pode ser considerado muito improvável, a não ser que tenhamos um tamanho de amostra extremamente pequeno.

Agora, estamos prontos para mostrar que, sob essas quatro hipóteses da regressão múltipla, os estimadores de MQO são não-viesados. Como no caso da regressão simples, as esperanças estão condicionadas aos valores das variáveis independentes da amostra, mas não mostraremos esse condicionamento explicitamente.

TEOREMA 3.1 (INEXISTÊNCIA DE VIÉS DE MQO)

Sob as hipóteses RLM.1 a RLM.4,

$$E(\hat{\beta}_j) = \beta_j, j = 0, 1, \dots, k, \quad (3.37)$$

para qualquer valor do parâmetro populacional β_j . Em outras palavras, os estimadores de MQO são estimadores não-viesados dos parâmetros da população.

Em nossos exemplos empíricos anteriores, a hipótese RLM.4 foi satisfeita (visto que fomos capazes de calcular as estimativas de MQO). Além disso, em sua maior parte, as amostras são aleatoriamente escolhidas de uma população bem-definida. Se acreditamos que os modelos especificados estão corretos sob a hipótese fundamental RLM.3, então podemos concluir que MQO é não-viesado nesses exemplos.

Como estamos nos aproximando do ponto em que podemos usar a regressão múltipla no trabalho empírico é útil lembrar o significado de inexistência de viés. É tentador, nos exemplos como o da equação do salário em (3.19), dizer algo como “9,2% é uma estimativa não-viesada do retorno da educação”. Como sabemos, uma estimativa não pode ser viesada: uma estimativa é um número fixo, obtido de uma amostra particular, usualmente diferente do parâmetro populacional. Quando dizemos que MQO é não-viesado sob as hipóteses RLM.1 a RLM.4, estamos dizendo que o *procedimento* pelo qual as estimativas de MQO foram obtidas é não-viesado, e tal procedimento é visto como algo aplicado em todas as amostras aleatórias possíveis. Esperamos que tenhamos obtido uma amostra que nos dê uma estimativa próxima do valor da população, mas, infelizmente, isso não pode ser garantido.

Inclusão de Variáveis Irrelevantes em um Modelo de Regressão

Uma questão que podemos dispensar com rapidez razoável é a **inclusão de uma variável irrelevante ou a superespecificação do modelo** na análise de regressão múltipla. Isso significa que uma (ou mais) das variáveis independentes está incluída no modelo, embora ela não tenha efeito parcial sobre y na população. (Isto é, seu coeficiente populacional é zero.)

Para ilustrar a questão, suponha que especificarmos o modelo como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u, \quad (3.38)$$

e esse modelo satisfaz as hipóteses RLM.1 a RLM.4. Entretanto, x_3 não tem efeito sobre y após x_1 e x_2 terem sido controlados, o que significa que $\beta_3 = 0$. A variável x_3 pode ou não ser correlacionada com x_1 e x_2 ; o que importa é que, uma vez que x_1 e x_2 estejam controlados, x_3 não tem efeito sobre y . Em termos de esperanças condicionais, $E(y|x_1, x_2, x_3) = E(y|x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

Como não sabemos que $\beta_3 = 0$, somos inclinados a estimar a equação com x_3 :

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3. \quad (3.39)$$

Incluímos a variável irrelevante, x_3 , em nossa regressão. Qual é o efeito de incluir x_3 em (3.39), quando seu coeficiente no modelo populacional (3.38) é zero? Em termos da inexistência de viés de $\hat{\beta}_1$ e $\hat{\beta}_2$, não há *nenhum efeito*. Essa conclusão não requer nenhuma derivação especial, já que ela decorre imediatamente do Teorema 3.1. Lembrase, a inexistência de viés significa $E(\hat{\beta}_j) = \beta_j$ para qualquer valor de β_j , incluindo $\beta_j = 0$. Assim, podemos concluir que $E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_1$, $E(\hat{\beta}_2) = \beta_2$ e $E(\hat{\beta}_3) = 0$ (para quaisquer valores de β_0 , β_1 e β_2). Mesmo que $\hat{\beta}_3$, por si mesmo, nunca seja exatamente zero, seu valor médio obtido de muitas amostras aleatórias será zero.

A conclusão do exemplo anterior é muito mais geral: incluir uma ou mais variáveis irrelevantes no modelo de regressão múltipla, ou superespecificar o modelo, não afeta a inexistência de viés dos estimadores de MQO. Isso significa que incluir variáveis irrelevantes é inócuo? Não. Como veremos na Seção 3.4, incluir variáveis irrelevantes pode ter efeitos indesejáveis sobre as *variâncias* dos estimadores de MQO.

Viés de Variável Omitida: O Caso Simples

Suponha agora que, em vez de incluir uma variável irrelevante, omitimos uma variável que, realmente, pertence ao modelo verdadeiro (ou populacional). Isso é freqüentemente chamado problema de **excluir uma variável relevante** ou de **subespecificar o modelo**. No Capítulo 2, e anteriormente neste capítulo, afirmamos que esse problema geralmente faz com que os estimadores de MQO sejam viesados. Agora é o momento de mostrar isso explicitamente e, não menos importante, derivar a direção e o tamanho do viés.

Derivar o viés causado ao omitir uma variável importante é um exemplo de **análise de má-especificação**. Iniciaremos com o caso em que o modelo populacional verdadeiro tem duas variáveis explicativas e um termo erro:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad (3.40)$$

e assumimos que esse modelo satisfaz as hipóteses RLM.1 a RLM.4.

Suponha que nosso interesse primordial esteja em β_1 , o efeito parcial de x_1 sobre y . Por exemplo, y é o salário horário (ou log do salário horário), x_1 é educação e x_2 é uma medida de aptidão inata. A fim de obter um estimador não-viesado de β_1 , deveríamos computar a regressão de y sobre x_1 e x_2 (o que resulta em estimadores não-viesados de β_0 , β_1 e β_2). Entretanto, devido à nossa ignorância ou indisponibilidade de dados, estimamos o modelo *excluindo* x_2 . Em outras palavras, executamos somente uma regressão simples de y sobre x_1 , obtendo a equação

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1. \quad (3.41)$$

Usamos o símbolo gráfico “~” em vez de “^” para enfatizar que $\tilde{\beta}_1$ vem de um modelo subespecificado.

Ao aprender, pela primeira vez, o problema de variável omitida, pode ser difícil para o estudante distinguir entre o modelo verdadeiro subjacente, (3.40) nesse caso, e o modelo que realmente estimamos, capturado pela regressão em (3.41). Pode parecer bobagem omitir a variável x_2 se ela pertence ao modelo, mas freqüentemente não temos escolha. Por exemplo, suponha que *salárioh* seja determinado pela equação

$$salárioh = \beta_0 + \beta_1 educ + \beta_2 aptid + u. \quad (3.42)$$

Como a aptidão não é observada, estimamos, em vez disso, o modelo

$$salárioh = \beta_0 + \beta_1 educ + v,$$

onde $v = \beta_2 aptid + u$. O estimador de β_1 da regressão simples de *salárioh* sobre *educ* é o que estamos chamando $\tilde{\beta}_1$.

Vamos derivar o valor esperado de $\tilde{\beta}_1$ condicionado aos valores amostrais de x_1 e x_2 . Derivar essa esperança não é difícil, pois $\tilde{\beta}_1$ é exatamente o estimador de inclinação de MQO de uma regressão simples, e já estudamos esse estimador extensivamente no Capítulo 2. A diferença aqui é que devemos analisar suas propriedades quando o modelo de regressão simples é mal-especificado devido a uma variável omitida.

Da equação (2.49), podemos expressar $\tilde{\beta}_1$ como

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)y_i}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}. \quad (3.43)$$

O próximo passo é o mais importante. Visto que (3.40) é o modelo verdadeiro, escrevemos y para cada observação i como

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i \quad (3.44)$$

(e *não* $y_i = \beta_0 + \beta_1 x_{i1} + u_i$, já que o modelo verdadeiro contém x_2). Seja SQT₁ o denominador em (3.43). Se inserirmos em (3.43) o y_i de (3.44), o numerador em (3.43) passa a ser

$$\begin{aligned} & \sum_{i=1}^n (x_{i1} - \bar{x}_1)(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i) \\ &= \beta_1 \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i2} + \sum_{i=1}^n (x_{i1} - \bar{x}_1)u_i \\ &= \beta_1 \text{SQT}_1 + \beta_2 \sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i2} + \sum_{i=1}^n (x_{i1} - \bar{x}_1)u_i, \end{aligned} \quad (3.45)$$

Se dividirmos (3.45) por SQT₁, considerarmos a esperança condicionada aos valores das variáveis independentes e usarmos $E(u_i) = 0$, obteremos

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i2}}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2} \quad (3.46)$$

Assim, $E(\tilde{\beta}_1)$ não é, geralmente, igual a β_1 : $\tilde{\beta}_1$ é viesado para β_1 .

A razão que multiplica β_2 em (3.46) tem uma interpretação simples: ela é exatamente o coeficiente de inclinação da regressão de x_2 sobre x_1 , utilizando nossa amostra de variáveis independentes, que pode ser escrita como

$$\tilde{x}_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1. \quad (3.47)$$

Como estamos condicionados aos valores amostrais de ambas as variáveis independentes, $\tilde{\delta}_1$ não é aleatório aqui. Portanto, podemos escrever (3.46) como

$$E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1, \quad (3.48)$$

o que implica que o viés em $\tilde{\beta}_1$ é $E(\tilde{\beta}_1) - \beta_1 = \beta_2 \tilde{\delta}_1$. Essa expressão é chamada freqüentemente de **viés de variável omitida**.

Da equação (3.48), vemos que há dois casos em que $\tilde{\beta}_1$ é não-viesado. O primeiro é bastante óbvio: se $\beta_2 = 0$ — de modo que x_2 não aparece no modelo verdadeiro (3.40) —, então $\tilde{\beta}_1$ é não-viesado. Já sabemos isso da análise de regressão simples do Capítulo 2. O segundo caso é mais interessante. Se $\tilde{\delta}_1 = 0$, então $\tilde{\beta}_1$ é não-viesado para β_1 , mesmo se $\beta_2 \neq 0$.

Como $\tilde{\delta}_1$ é a covariância amostral entre x_1 e x_2 sobre a variância amostral de x_1 , $\tilde{\delta}_1 = 0$ se, e somente se, x_1 e x_2 forem não-correlacionados na amostra. Assim, temos a importante conclusão de que, se x_1 e x_2 forem não-correlacionados na amostra, então $\tilde{\beta}_1$ é não-viesado. Isso não é surpreendente: na Seção 3.2, mostramos que o estimador da regressão simples $\tilde{\beta}_1$ e o estimador da regressão múltipla $\hat{\beta}_1$ são iguais quando x_1 e x_2 forem não-correlacionados na amostra. [Podemos também mostrar que $\tilde{\beta}_1$ é não-viesado sem condicionar a x_{i2} se $E(x_2|x_1) = E(x_2)$; então, para a estimativa de β_1 , deixar x_2 no termo erro não viola a hipótese de média condicional zero do erro, uma vez que ajustamos o intercepto.]

Quando x_1 e x_2 forem correlacionados, $\tilde{\delta}_1$ tem o mesmo sinal da correlação entre x_1 e x_2 : $\tilde{\delta}_1 > 0$ se x_1 e x_2 forem positivamente correlacionados, e $\tilde{\delta}_1 < 0$ se x_1 e x_2 forem negativamente correlacionados. O sinal do viés em $\tilde{\beta}_1$ depende tanto do sinal de β_2 como de $\tilde{\delta}_1$ e está resumido na Tabela 3.2 para os quatro casos possíveis quando há viés. A Tabela 3.2 justifica um estudo cuidadoso. Por exemplo, o viés em $\tilde{\beta}_1$ é positivo se $\beta_2 > 0$ (x_2 tem um efeito positivo sobre y) e x_1 e x_2 são positivamente correlacionados; o viés é negativo se $\beta_2 > 0$ e x_1 e x_2 são negativamente correlacionados, e assim por diante.

Tabela 3.2

Sumário do Viés em $\tilde{\beta}_1$, quando x_2 é Omitida na Estimação da Equação (3.40)

	Corr(x_1, x_2) > 0	Corr(x_1, x_2) < 0
$\beta_2 > 0$	viés positivo	viés negativo
$\beta_2 < 0$	viés negativo	viés positivo

A Tabela 3.2 resume a direção do viés, mas o tamanho do viés também é muito importante. Um viés pequeno de qualquer dos dois sinais não precisa ser causa de preocupação. Por exemplo, se o retorno da educação formal é de 8,6% na população e o viés do estimador de MQO é de 0,1% (um décimo de um ponto percentual), então não precisaríamos ficar muito preocupados. De outro lado, um viés da ordem de três pontos percentuais seria muito mais sério. O tamanho do viés é determinado pelos tamanhos de β_2 e $\tilde{\delta}_1$.

Na prática, como β_2 é um parâmetro populacional desconhecido, não podemos estar certos se β_2 é positivo ou negativo. Entretanto, temos geralmente uma boa idéia sobre a direção do efeito parcial de x_2 sobre y . Além disso, ainda que o sinal da correlação entre x_1 e x_2 não possa ser conhecido se x_2 não é observado, em muitos casos, podemos fazer uma suposição criteriosa sobre se x_1 e x_2 são positiva ou negativamente correlacionados.

Na equação do salário (3.42), por definição, mais aptidão conduz a uma produtividade maior e, portanto, a salários maiores: $\beta_2 > 0$. Há também razões para acreditar que *educ* e *aptid* sejam positivamente correlacionados: em média, indivíduos com mais aptidão inata escolhem níveis maiores de educação formal. Assim, as estimativas de MQO da equação de regressão simples do *salárioh* = $\beta_0 + \beta_1 \text{educ} + \nu$ são, *em média*, muito grandes. Isso não significa que a estimativa obtida de nossa amostra seja enorme. Somente podemos dizer que, se coletarmos muitas amostras aleatoriamente e obtivermos as estimativas da regressão simples a cada vez, a média dessas estimativas será maior que β_1 .

EXEMPLO 3.6

(Equação do Salário Horário)

Suponha que o modelo $\log(\text{salárioh}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{aptid} + u$ satisfaça as hipóteses RLM.1 a RLM.4. O conjunto de dados no arquivo WAGE1.RAW não contém dados sobre aptidão, de modo que estimamos β_1 a partir da regressão simples

$$\begin{aligned}\log(\tilde{\text{salárioh}}) &= 0,584 + 0,083 \text{ educ} \\ n &= 526, R^2 = 0,186.\end{aligned}$$

Esse é somente o resultado de uma única amostra, de modo que não podemos dizer que 0,083 é maior que β_1 ; o retorno verdadeiro da educação poderia ser menor ou maior que 8,3% (nunca saberemos com certeza). Entretanto, sabemos que a média dos estimadores de todas as amostras aleatórias seria bastante grande.

Como um segundo exemplo, suponha que, no nível fundamental do ensino, a nota média dos estudantes de um exame padronizado seja determinado por

$$\text{notmed} = \beta_0 + \beta_1 \text{gasto} + \beta_2 \text{taxpob} + u,$$

em que *gasto* é o gasto público por estudante, e *taxpob* é a taxa de pobreza das crianças da escola. Usando dados do distrito da escola, temos somente observações da percentagem de estudantes com uma nota de aprovação e gastos públicos por estudante; não temos informações sobre taxas de pobreza. Assim, estimamos β_1 a partir da regressão simples de *notmed* sobre *gasto*.

Podemos obter, uma vez mais, o viés provável em $\tilde{\beta}_1$. Primeiro, β_2 é provavelmente negativo: há ampla evidência de que crianças que vivem na pobreza têm, em média, notas mais baixas em testes padronizados. Segundo, o gasto público médio por estudante é, provavelmente, negativamente correlacionado com a taxa de pobreza: quanto maior a taxa de pobreza menor o gasto público médio por estudante, de modo que $\text{Corr}(x_1, x_2) < 0$. De acordo com a Tabela 3.2, $\tilde{\beta}_1$ terá um viés positivo. Essa observação tem implicações importantes. Pode ser que o efeito verdadeiro do gasto público fosse zero; isto é, $\beta_1 = 0$. Entretanto, a estimativa de β_1 da regressão simples será, geralmente, maior que zero, e isso poderia nos levar a concluir que os gastos públicos são importantes quando eles não são.

Ao ler e ao fazer trabalhos empíricos em economia, é importante dominar a terminologia associada aos estimadores viesados. No contexto de omissão de uma variável do modelo (3.40), se $E(\tilde{\beta}_1) > \beta_1$, então dizemos que $\tilde{\beta}_1$ tem um **viés para cima**. Quando $E(\tilde{\beta}_1) < \beta_1$, $\tilde{\beta}_1$ tem um **viés para baixo**. Essas definições são as mesmas, seja β_1 positivo ou negativo. A expressão **viesado para zero** refere-se aos casos em que $E(\tilde{\beta}_1)$ está mais próxima de zero do que de β_1 . Portanto, se β_1 for positivo, $\tilde{\beta}_1$ será viesado para zero se ele tiver um viés para baixo. De outro lado, se $\beta_1 < 0$, $\tilde{\beta}_1$ será viesado para zero se ele tiver um viés para cima.

Viés de Variável Omitida: Casos mais Gerais

É mais difícil derivar o sinal do viés de variável omitida quando há múltiplos regressores no modelo estimado. Devemos lembrar que a correlação entre uma única variável explicativa e o erro resulta, geralmente, em *todos* os estimadores de MQO serem viesados. Por exemplo, suponha que o modelo populacional

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \quad (3.49)$$

satisfaca as hipóteses RLM.1 a RLM.4. No entanto, omitimos x_3 e estimamos o modelo como

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2. \quad (3.50)$$

Agora suponha que x_2 e x_3 sejam não-correlacionados, mas que x_1 é correlacionado com x_3 . Em outras palavras, x_1 é correlacionado com a variável omitida, mas x_2 não é. É tentador pensar que, embora provavelmente $\tilde{\beta}_1$ seja viesado com base na derivação da subseção anterior, $\tilde{\beta}_2$ seja não-viesado, pois x_2 é não-correlacionado com x_3 . Infelizmente, esse *não* é, geralmente, o caso: normalmente, tanto $\tilde{\beta}_1$ como $\tilde{\beta}_2$ serão viesados. A única exceção a isso ocorre quando x_1 e x_2 também são não-correlacionados.

Mesmo em um modelo razoavelmente simples como o apresentado, pode ser difícil obter a direção do viés em $\tilde{\beta}_1$ e $\tilde{\beta}_2$. Isso se deve ao fato de que x_1 , x_2 e x_3 podem estar correlacionados aos pares. Entretanto, uma aproximação é, freqüentemente, útil na prática. Se assumirmos que x_1 e x_2 são não-correlacionados, podemos estudar o viés em $\tilde{\beta}_1$ como se x_2 estivesse ausente dos modelos populacional e estimado. De fato, quando x_1 e x_2 são não-correlacionados, pode-se mostrar que

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)x_{i3}}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$$

Essa equação é exatamente igual a (3.46), mas β_3 substitui β_2 , e x_3 substitui x_2 . Portanto, o viés em $\tilde{\beta}_1$ é obtido ao se substituir β_2 por β_3 e x_2 por x_3 na Tabela 3.2. Se $\beta_3 > 0$ e $\text{Corr}(x_1, x_3) > 0$, o viés em $\tilde{\beta}_1$ é positivo, e assim por diante.

Como um exemplo, suponha que acrescentamos *exper* ao modelo do salário:

$$\text{salárioh} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{aptid} + u.$$

Se *aptid* for omitida do modelo, ambos os estimadores de β_1 e β_2 serão viesados, mesmo se assumirmos que *exper* é não-correlacionado com *aptid*. Estamos principalmente interessados no retorno da educação formal, de modo que seria bom se pudéssemos concluir que $\tilde{\beta}_1$ tem um viés para cima ou para baixo devido à omissão da aptidão. Essa conclusão não é possível sem hipóteses adicionais. Como uma *aproximação*, suponhamos que, além de *exper* e *aptid* serem não-correlacionadas, *educ* e *exper* também sejam não-correlacionadas. (Na realidade, elas são negativamente correlacionadas.) Como $\beta_3 > 0$ e *educ* e *aptid* são positivamente correlacionadas, $\tilde{\beta}_1$ teria um viés para cima, exatamente como se *exper* não estivesse no modelo.

O raciocínio usado no exemplo anterior é, muitas vezes, compreendido como um guia aproximado para obter o viés provável dos estimadores em modelos mais complicados. Geralmente, o foco está

na relação entre uma variável explicativa particular, por exemplo x_1 , e o fator omitido fundamental. Estritamente falando, ignorar todas as outras variáveis explicativas é uma prática válida somente quando cada uma delas é não-correlacionada com x_1 , mas essa ainda é uma orientação útil. O Apêndice 3A (disponível no site da Thomson) contém uma análise mais cuidadosa do viés de variável omitida com múltiplas variáveis explicativas.

3.4 A VARIÂNCIA DOS ESTIMADORES DE MQO

Obteremos, agora, a variância dos estimadores de MQO, de modo que, além de conhecermos as tendências centrais dos $\hat{\beta}_j$, também teremos uma medida da dispersão de sua distribuição amostral. Antes de encontrarmos as variâncias, vamos adicionar uma hipótese de homoscedasticidade, como no Capítulo 2. Fazemos isso por duas razões. Primeira, ao impor a hipótese de variância constante do erro, as fórmulas são simplificadas. Segunda, veremos na Seção 3.5 que MQO tem uma propriedade importante de eficiência se acrescentarmos a hipótese de homoscedasticidade.

No arcabouço da regressão múltipla, a homoscedasticidade é expressa como a seguir:

$$\text{H I P Ó T E S E R L M . 5 (HOMOSCEDASTICIDADE)} \\ \text{Var}(u|x_1, \dots, x_k) = \sigma^2.$$

A hipótese RLM.5 significa que a variância do termo erro, u , condicionada às variáveis explicativas, é a mesma para todas as combinações de resultados das variáveis explicativas. Se essa hipótese é violada, o modelo exibe heteroscedasticidade, exatamente como no caso de duas variáveis.

Na equação

$$\text{salárioh} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + u,$$

a homoscedasticidade requer que a variância do erro não-observado u não dependa dos níveis de educação, experiência ou permanência. Isto é,

$$\text{Var}(u|\text{educ}, \text{exper}, \text{perm}) = \sigma^2.$$

Se a variância varia com qualquer uma das três variáveis explicativas, então a heteroscedasticidade está presente.

As hipóteses RLM.1 a RLM.5 são, em conjunto, conhecidas como as **hipóteses de Gauss-Markov** (para a regressão de corte transversal). Até agora, nossas asserções sobre as hipóteses são adequadas somente quando aplicadas à análise de corte transversal com amostragem aleatória. Como veremos, as hipóteses de Gauss-Markov para a análise de séries de tempo — e para outras situações, como a análise de dados de painel — são mais difíceis de se manterem, embora haja muitas similaridades.

Na discussão a seguir, usaremos o símbolo x para representar o conjunto de todas as variáveis independentes, (x_1, \dots, x_k) . Assim, na regressão do salário horário com educ , exper e perm como variáveis independentes, $x = (\text{educ}, \text{exper}, \text{perm})$. Consequentemente, podemos escrever as hipóteses RLM.1 e RLM.3 como

$$E(y|x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_3 x_3,$$

e a hipótese RLM.5 é a mesma que $\text{Var}(y|x) = \sigma^2$. Expressar as hipóteses desse modo ilustra como a hipótese RLM.5 difere muito da hipótese RLM.3. Esta diz que o valor esperado de y , dado x , é linear nos parâmetros, mas ele certamente depende de x_1, x_2, \dots, x_k . A hipótese RLM.5 diz que a variância de y , dado x , não depende dos valores das variáveis independentes.

Podemos obter, agora, as variâncias dos $\hat{\beta}_j$, que uma vez mais, estão condicionadas aos valores amostrais das variáveis independentes. A prova está no apêndice deste capítulo.

TEOREMA 3.2 (VARIÂNCIAS AMOSTRAIS DOS ESTIMADORES DE INCLINAÇÃO DE MQO)

Sob as hipóteses RLM.1 a RLM.5, condicionadas aos valores amostrais das variáveis independentes,

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\text{SQT}_j(1 - R_j^2)}, \quad (3.51)$$

para $j = 1, 2, \dots, k$, em que $\text{SQT}_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ é a variação amostral total em x_j , e R_j^2 é o R -quadrado da regressão x_j sobre todas as outras variáveis independentes (incluindo um intercepto).

Antes de estudarmos a equação (3.51) mais detalhadamente, é importante saber que todas as hipóteses de Gauss-Markov são usadas na obtenção dessa fórmula. Embora não necessitemos da hipótese de homoscedasticidade para concluir que MQO é não-viesado, precisamos dela para validar a equação (3.51).

O tamanho de $\text{Var}(\hat{\beta}_j)$ é importante na prática. Uma variância maior significa um estimador menos preciso, e isso se traduz em intervalos de confiança maiores e testes de hipóteses menos acurados (como veremos no Capítulo 4). Na próxima subseção, discutiremos os elementos que compreendem (3.51).

Os Componentes das Variâncias de MQO: Multicolinearidade

A equação (3.51) mostra que a variância de $\hat{\beta}_j$ depende de três fatores: σ^2 , SQT_j e R_j^2 . Lembre-se de que o índice j representa simplesmente qualquer uma das variáveis independentes (como a educação ou a taxa de pobreza). Agora, vamos considerar cada um dos fatores que afetam $\text{Var}(\hat{\beta}_j)$.

A VARIÂNCIA DO ERRO, σ^2 . Da equação (3.51), um σ^2 maior significa variâncias maiores dos estimadores de MQO. Isso não é totalmente surpreendente: mais “ruído” na equação (um σ^2 maior) torna mais difícil estimar o efeito parcial de qualquer uma das variáveis independentes sobre y , e isso é refletido nas variâncias maiores dos estimadores de inclinação de MQO. Visto que σ^2 é uma característica da população, ele não tem nada a ver com o tamanho da amostra. Ele é o componente de (3.51) que é desconhecido. Veremos mais adiante como obter um estimador não-viesado de σ^2 .

Para uma dada variável dependente y , há de fato somente uma maneira de reduzir a variância do erro, que é adicionar mais variáveis explicativas à equação (retirar alguns fatores do termo erro). Isso nem sempre é possível, nem sempre é desejável, por razões discutidas posteriormente neste capítulo.

A VARIAÇÃO AMOSTRAL TOTAL EM x_j , SQT_j . Da equação (3.51), vemos que quanto maior a variação total em x_j , menor é $\text{Var}(\hat{\beta}_j)$. Assim, tudo o mais sendo igual para estimar $\hat{\beta}_j$, preferimos ter tanta variação amostral em x_j quanto possível. Já descobrimos isso no caso da regressão simples, no Capítulo 2. Embora raramente seja possível escolher os valores amostrais das variáveis independentes, há uma maneira de aumentar a variação amostral em cada uma das variáveis independentes: aumentar o tamanho da amostra. De fato, na amostragem aleatória de uma população, SQT_j aumenta sem limite quando o tamanho da amostra torna-se maior. Esse é o componente da variância que depende sistematicamente do tamanho da amostra.

Quando SQT_j é pequeno, $\text{Var}(\hat{\beta}_j)$ pode ficar muito grande, mas um SQT_j pequeno não é uma violação da hipótese RLM.4. Tecnicamente, quando SQT_j tende a zero, $\text{Var}(\hat{\beta}_j)$ aproxima-se do infinito. O caso extremo de nenhuma variação amostral em x_j , $SQT_j = 0$, não é permitido pela hipótese RLM.4.

AS RELAÇÕES LINEARES ENTRE AS VARIÁVEIS INDEPENDENTES, R_j^2 . O termo R_j^2 na equação (3.51) é o mais difícil dos três componentes de se entender. Esse termo não aparece na análise de regressão simples porque há somente uma variável independente em tal caso. É importante compreender que esse R -quadrado é distinto do R -quadrado da regressão de y sobre x_1, x_2, \dots, x_k : R_j^2 é obtido de uma regressão que envolve somente as variáveis independentes do modelo original, em que x_j desempenha o papel de uma variável dependente.

Considere, primeiro, o caso $k = 2$: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Então, $\text{Var}(\hat{\beta}_1) = \sigma^2/[SQT_1(1 - R_1^2)]$, em que R_1^2 é o R -quadrado da regressão simples de x_1 sobre x_2 (e um intercepto, como sempre). Como o R -quadrado mede o grau de ajuste, um valor de R_1^2 próximo de um indica que x_2 explica bastante da variação de x_1 na amostra. Isso significa que x_1 e x_2 são altamente correlacionados.

Quando R_1^2 cresce em direção a um, $\text{Var}(\hat{\beta}_1)$ torna-se maior. Assim, um grau elevado de relação linear entre x_1 e x_2 pode levar a variâncias grandes dos estimadores de inclinação de MQO. (Um argumento similar se aplica a $\hat{\beta}_2$.) Veja a Figura 3.1 para a relação entre $\text{Var}(\hat{\beta}_j)$ e o R -quadrado da regressão de x_1 sobre x_2 .

No caso geral, R_j^2 é a proporção da variação total de x_j que pode ser explicada pelas outras variáveis independentes que aparecem na equação. Para dados σ^2 e SQT_j , a menor $\text{Var}(\hat{\beta}_j)$ é obtida quando $R_j^2 = 0$, que ocorre se, e somente se, x_j tem correlação amostral zero com *cada uma das outras* variáveis independentes. Esse é o melhor caso para estimar β_j , mas é raramente encontrado.

O outro caso extremo, $R_j^2 = 1$, é excluído pela hipótese RLM.4, pois $R_j^2 = 1$ significa que, na amostra, x_j é uma combinação linear *perfeita* de algumas das outras variáveis independentes da regressão. Um caso mais relevante é quando R_j^2 está “próximo” de um. Da equação (3.51) e da Figura 3.1, vemos que

isso pode fazer com que $\text{Var}(\hat{\beta}_j)$ seja grande: $\text{Var}(\hat{\beta}_j) \rightarrow \infty$ quando $R_j^2 \rightarrow 1$. Correlação alta (mas não perfeita) entre duas ou mais variáveis independentes é chamada **multicolinearidade**.

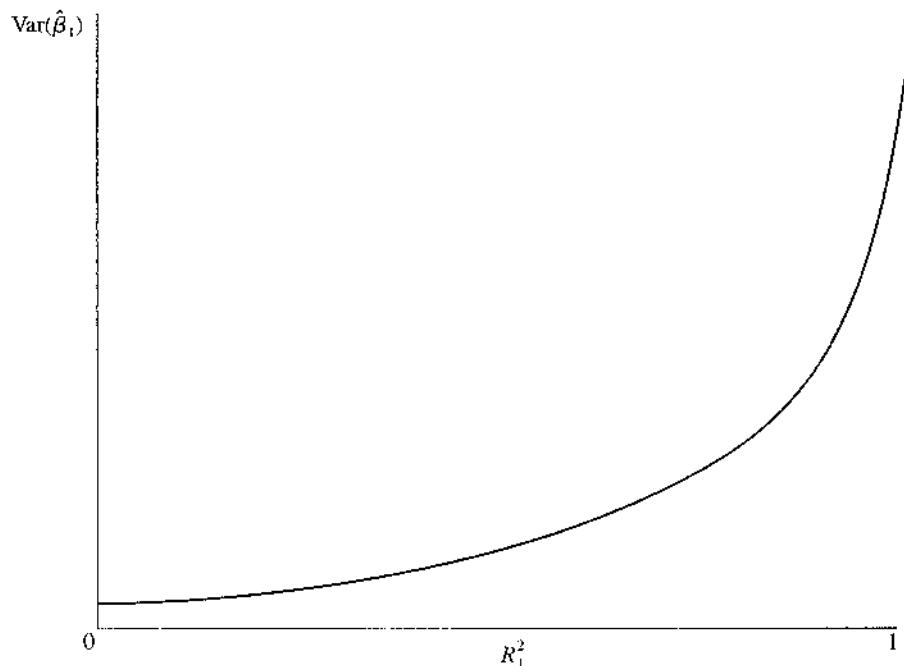
Antes de discutirmos mais a questão da multicolinearidade, é importante que uma coisa esteja bem clara: um caso em que R_j^2 está próximo de um *não* é uma violação da hipótese RLM.4.

Como a multicolinearidade não viola nenhuma de nossas hipóteses, o “problema” da multicolinearidade não é, de fato, bem definido. Ao dizer que a multicolinearidade surge ao estimarmos β_j , quando R_j^2 está “próximo” de um, colocamos “próximo” dentro de aspas porque não há um número absoluto que podemos citar para concluir que a multicolinearidade é um problema. Por exemplo, $R_j^2 = 0,9$ significa que 90% da variação amostral em x_j pode ser explicada pelas outras variáveis independentes do modelo de regressão. Inquestionavelmente, isso significa que x_j tem uma forte relação linear com as outras variáveis independentes. No entanto, se isso se traduz em uma $\text{Var}(\hat{\beta}_j)$ que é grande demais para ser útil, depende dos tamanhos de σ^2 e SQT_j . Como veremos no Capítulo

4, para a inferência estatística, o que essencialmente importa é quanto $\hat{\beta}_j$ é grande com relação a seu desvio-padrão.

Figura 3.1

Var($\hat{\beta}_j$) como uma função de R_j^2 .



Assim como um valor grande de R_j^2 pode causar uma $\text{Var}(\hat{\beta}_j)$ grande, um valor pequeno de SQT_j também pode fazer com que $\text{Var}(\hat{\beta}_j)$ seja grande. Portanto, um tamanho pequeno da amostra pode também levar a variâncias amostrais grandes. Preocupar-se com graus elevados de correlação entre variáveis independentes da amostra não é, de fato, diferente de se preocupar com um tamanho pequeno da amostra: ambos funcionam para aumentar $\text{Var}(\hat{\beta}_j)$.

O famoso econometrista da Universidade de Wisconsin Arthur Goldberger, reagindo à obsessão dos econometristas pela multicolinearidade, criou (jocosamente) o termo **micronumerosidade**, que ele define como o “problema do tamanho pequeno da amostra”. [Para uma discussão interessante sobre multicolinearidade e micronumerosidade, veja Goldberger (1991).]

Embora o problema da multicolinearidade não possa ser claramente definido, uma coisa é clara: tudo mais sendo igual, para estimar β_j , é melhor ter menos correlação entre x_j e as outras variáveis independentes. Essa observação muitas vezes leva a uma discussão de como “resolver” o problema da multicolinearidade. Nas ciências sociais, em que somos geralmente coletores passivos de dados, não há uma boa maneira de reduzir as variâncias dos estimadores não-viesados que não seja coletar mais dados. Para um determinado conjunto de dados, podemos tentar, num esforço para reduzir a multicolinearidade, suprimir outras variáveis independentes do modelo. Infelizmente, suprimir uma variável que pertence ao modelo populacional pode levar viés, como vimos na Seção 3.3.

Neste ponto, talvez um exemplo ajude a esclarecer algumas das questões aqui levantadas relativas à multicolinearidade. Suponha que estamos interessados em estimar o efeito de várias categorias de

despesas de escolas sobre o desempenho de estudantes. É provável que as despesas com salários de professores, materiais institucionais, atletismo etc. estejam altamente correlacionadas: escolas mais ricas tendem a gastar mais com tudo, e escolas mais pobres gastam menos com tudo. Não surpreendentemente, pode ser difícil estimar o efeito de qualquer categoria de despesa específica sobre o desempenho dos estudantes quando há pouca variação em uma categoria que não pode ser, em grande medida, explicada por variações das outras categorias de despesas (isso leva a um R^2 alto para cada uma das variáveis de despesas). Esses problemas de multicolinearidade podem ser mitigados ao coletar mais dados mas assim, em certo sentido, nós mesmos nos impusemos o problema: estamos formulando questões que podem ser sutis demais para que os dados disponíveis as respondam com alguma precisão. Provavelmente, podemos fazer algo muito melhor mudando o escopo da análise e agrupando todas as categorias de despesa em uma única categoria, desde que não mais estivéssemos tentando estimar o efeito parcial de cada categoria separadamente.

Outro ponto importante é que um elevado grau de correlação entre certas variáveis independentes pode ser irrelevante no que diz respeito a quanto bem podemos estimar outros parâmetros do modelo. Por exemplo, considere um modelo com três variáveis independentes:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,$$

em que x_2 e x_3 são altamente correlacionados. Então, $\text{Var}(\hat{\beta}_2)$ e $\text{Var}(\hat{\beta}_3)$ podem ser grandes. Mas o valor da correlação entre x_2 e x_3 não tem efeito direto sobre $\text{Var}(\hat{\beta}_1)$. De fato, se x_1 é não-correlacionado com x_2 e x_3 , então $R^2_1 = 0$ e $\text{Var}(\hat{\beta}_1) = \sigma^2/\text{SQT}_1$, independentemente da quantia de correlação existir entre x_2 e x_3 . Se β_1 é o parâmetro de interesse, realmente não devemos nos preocupar com o valor da correlação entre x_2 e x_3 .

Suponha que você postula um modelo que explica a nota do exame final em termos da freqüência às aulas. Assim, a variável dependente é a nota do exame final, e a principal variável explicativa é o número de aulas freqüentadas. A fim de controlar as aptidões dos estudantes e pelos esforços fora da sala de aula, você inclui entre as variáveis explicativas a nota acumulada durante todo o curso, a nota do teste de avaliação de conhecimentos para ingresso em curso superior e as medidas do desempenho do estudante no ensino médio. Alguém diz: "Você não pode esperar aprender nada com esse exercício, pois todas essas variáveis são, provavelmente, altamente colineares". Qual seria sua resposta?

A observação anterior é importante porque os economistas freqüentemente incluem muitas variáveis de controle a fim de isolar o efeito causal de uma variável particular. Por exemplo, ao olhar para a relação entre as taxas de aprovação de empréstimos e a percentagem de minorias em uma região, poderíamos incluir variáveis como renda média, valor médio das moradias, medidas de inadimplência, e assim por diante, pois esses fatores precisam ser considerados a fim de se extrair conclusões causais sobre a discriminação. Renda, preços de moradia e inadimplência são, geralmente, altamente correlacionados entre si. No entanto, correlações altas entre essas variáveis não tornam mais difícil determinar os efeitos da discriminação.

Variâncias em Modelos Mal Especificados

A escolha de incluir ou não uma variável particular em um modelo de regressão pode ser feita ao analisar o dilema entre viés e variância. Na Seção 3.3 derivamos o viés produzido pela omissão de uma variável relevante quando o modelo verdadeiro contém duas variáveis explicativas. Vamos continuar a análise desse modelo comparando as variâncias dos estimadores de MQO.

Escreva o modelo populacional verdadeiro que satisfaz as hipóteses de Gauss-Markov, como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

Consideremos dois estimadores de β_1 . O estimador $\hat{\beta}_1$ é proveniente da regressão múltipla

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2. \quad (3.52)$$

Em outras palavras, incluímos x_2 , juntamente com x_1 , no modelo de regressão. O estimador $\tilde{\beta}_1$ é obtido ao omitir x_2 do modelo e computando uma regressão simples de y sobre x_1 :

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1. \quad (3.53)$$

Quando $\beta_2 \neq 0$, a equação (3.53) exclui uma variável relevante do modelo e, como vimos na Seção 3.3, isso induz um viés em $\tilde{\beta}_1$, a não ser que x_1 e x_2 sejam não-correlacionados. De outro lado, $\tilde{\beta}_1$ é não-viesado para β_1 , para qualquer valor de β_2 , incluindo $\beta_2 = 0$. Segue-se que, se o viés for usado como único critério, $\tilde{\beta}_1$ é preferível a $\hat{\beta}_1$.

A conclusão de que $\tilde{\beta}_1$ é sempre preferível a $\hat{\beta}_1$ não se sustenta quando trazemos a variância para dentro da análise. Condicionando aos valores de x_1 e x_2 na amostra, temos, de (3.51),

$$\text{Var}(\hat{\beta}_1) = \sigma^2 / [\text{SQT}_1(1 - R_1^2)], \quad (3.54)$$

em que SQT_1 é a variação total em x_1 , e R_1^2 é o R -quadrado da regressão de x_1 sobre x_2 . Além disso, uma simples modificação da prova para a regressão de duas variáveis do Capítulo 2 mostra que

$$\text{Var}(\tilde{\beta}_1) = \sigma^2 / \text{SQT}_1. \quad (3.55)$$

Comparar (3.55) a (3.54) mostra que $\text{Var}(\tilde{\beta}_1)$ é sempre *menor* que $\text{Var}(\hat{\beta}_1)$, a menos que x_1 e x_2 sejam não-correlacionados na amostra, caso em que os dois estimadores $\tilde{\beta}_1$ e $\hat{\beta}_1$ são os mesmos. Ao assumir que x_1 e x_2 são não-correlacionados, podemos ter as seguintes conclusões:

1. Quando $\beta_2 \neq 0$, $\tilde{\beta}_1$ é viesado, $\hat{\beta}_1$ é não-viesado e $\text{Var}(\tilde{\beta}_1) < \text{Var}(\hat{\beta}_1)$.
2. Quando $\beta_2 = 0$, $\tilde{\beta}_1$ e $\hat{\beta}_1$ são ambos não-viesados e $\text{Var}(\tilde{\beta}_1) < \text{Var}(\hat{\beta}_1)$.

Da segunda conclusão, é claro que $\tilde{\beta}_1$ é preferido se $\beta_2 = 0$. Intuitivamente, se x_2 não tem um efeito parcial sobre y , incluí-lo no modelo pode somente exacerbar o problema da multicolinearidade, o que

leva a um estimador menos eficiente de β_1 . O custo de incluir uma variável irrelevante no modelo é uma variância maior do estimador de β_1 .

O caso em que $\beta_2 \neq 0$ é mais difícil. Omitindo x_2 do modelo leva a um estimador viesado de β_1 . Tradicionalmente, econometristas têm sugerido comparar o tamanho provável do viés devido à omissão de x_2 com a redução na variância — resumida no tamanho de R^2 — para decidir se x_2 deve ser incluído. Entretanto, quando $\beta_2 \neq 0$, há duas razões favoráveis para incluir x_2 no modelo. A mais importante delas é que qualquer viés em $\tilde{\beta}_1$ não diminui quando o tamanho da amostra cresce; de fato, o viés não segue, necessariamente, qualquer padrão. Portanto, podemos em geral pensar o viés como mais ou menos o mesmo para qualquer tamanho de amostra. De outro lado, $\text{Var}(\tilde{\beta}_1)$ e $\text{Var}(\tilde{\beta}_2)$ tendem a zero quando n torna-se grande, o que significa que a multicolinearidade induzida pela adição de x_2 torna-se menos importante quando o tamanho da amostra cresce. Em amostras grandes, preferiríamos $\tilde{\beta}_1$.

A outra razão para preferir $\hat{\beta}_1$ é mais sutil. A fórmula da variância em (3.55) está condicionada aos valores de x_1 e x_2 na amostra, o que oferece o melhor cenário para $\tilde{\beta}_1$. Quando $\beta_2 \neq 0$, a variância de $\tilde{\beta}_1$ condicionada somente a x_1 é maior que aquela apresentada em (3.55). Intuitivamente, quando $\beta_2 \neq 0$ e x_2 é excluído do modelo, a variância do erro aumenta porque o erro efetivamente contém parte de x_2 . Mas, (3.55) ignora o aumento da variância do erro porque ela trata ambos os regressores como não-aleatórios. Uma discussão completa de quais variáveis independentes deveriam ser condicionadas nos desviaria demais de nosso cairinho. É suficiente dizer que (3.55) é bastante generosa quando ela aparece para medir a precisão de $\tilde{\beta}_1$.

Estimação de σ^2 : Os Erros-Padrão dos Estimadores de MQO

Vamos, agora, mostrar como escolher um estimador não-viesado de σ^2 , que nos permitirá obter estimadores não-viesados de $\text{Var}(\tilde{\beta}_j)$.

Como $\sigma^2 = E(u^2)$, um “estimador” não-viesado de σ^2 é a média amostral dos erros quadrados: $n^{-1} \sum_{i=1}^n u_i^2$. Infelizmente, esse não é um estimador verdadeiro, pois não observamos os u_i . Não obstante, lembre-se de que os erros podem ser escritos como $u_i = y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \dots - \beta_k x_{ik}$, e a razão real de não observarmos os u_i é que não conhecemos os β_j . Quando substituímos cada β_j por seu estimador de MQO, obtemos os resíduos de MQO:

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik}.$$

Parece natural estimar σ^2 ao substituir u_i por \hat{u}_i . No caso da regressão simples, vimos que isso leva a um estimador viesado. O estimador não-viesado de σ^2 no caso geral da regressão múltipla é

$$\hat{\sigma}^2 = \left(\sum_{i=1}^n \hat{u}_i^2 \right) / (n - k - 1) = \text{SQR}/(n - k - 1). \quad (3.56)$$

Já encontramos esse estimador no caso $k = 1$ da regressão simples.

O termo $n - k - 1$ em (3.56) representa os **graus de liberdade** (*gl*) do problema geral de MQO com n observações e k variáveis independentes. Como há $k + 1$ parâmetros em um modelo de regressão com k variáveis independentes e um intercepto, podemos escrever

$$gl = n - (k + 1) \quad (3.57)$$

= (número de observações) - (número de parâmetros estimados)

Essa é a maneira mais fácil de calcular os graus de liberdade em uma aplicação particular: conte o número de parâmetros, incluindo o intercepto, e subtraia esse valor do número de observações. (No caso raro em que o intercepto não é estimado, o número de parâmetros diminui em um.)

Tecnicamente, a divisão por $n - k - 1$ em (3.56) é proveniente do fato de que o valor esperado da soma dos resíduos quadrados é $E(SQR) = (n - k - 1) \sigma^2$. Intuitivamente, podemos entender porque o ajustamento de graus de liberdade é necessário ao retornarmos às condições de primeira ordem dos estimadores de MQO. Elas podem ser escritas como $\sum_{i=1}^n u_i = 0$ e $\sum_{i=1}^n x_{ij} \hat{u}_i = 0$, onde $j = 1, 2, \dots, k$. Assim, na obtenção dos estimadores de MQO, $k + 1$ restrições são impostas sobre os resíduos de MQO. Isso significa que, dados $n - (k + 1)$ dos resíduos, os $k + 1$ resíduos restantes são conhecidos: há somente $n - (k + 1)$ graus de liberdade nos resíduos. (Isso pode ser contrastado com os erros u_i , os quais têm n graus de liberdade na amostra.)

Para referência, vamos resumir essa discussão com o Teorema 3.3. Provamos esse teorema para o caso da análise de regressão simples no Capítulo 2 (veja o Teorema 2.3). (No Apêndice E, disponível no site de Thomson, é dada uma prova geral, que requer álgebra matricial.)

TEOREMA 3 . 3 (ESTIMADOR NÃO-VIESADO DE σ^2)

Sob as hipóteses de Gauss-Markov RLM.1 a RLM.5, $E(\hat{\sigma}^2) = \sigma^2$.

A raiz quadrada positiva de $\hat{\sigma}^2$, denominada $\hat{\sigma}$, é chamada **Erro-Padrão da Regressão (EPR)**. O EPR é um estimador do desvio-padrão do termo erro. Essa estimativa é usualmente informada pelos programas de regressão, embora ela seja chamada de nomes diferentes pelos diferentes programas. (Além de EPR, $\hat{\sigma}$ também é chamado *erro-padrão da estimativa* e a *raiz do erro quadrado médio*.)

Observe que $\hat{\sigma}$ pode diminuir ou aumentar quando outra variável independente é acrescentada a uma regressão (para uma dada amostra). Isso ocorre pois, embora SQR deva cair quando outra variável explicativa é adicionada, os graus de liberdade também diminuem em um. Como SQR está no numerador e gl está no denominador, não podemos dizer, de antemão, qual efeito prevalecerá.

Para construir intervalos de confiança e conduzir testes no Capítulo 4, precisaremos estimar o **desvio-padrão de $\hat{\beta}_j$** , que é exatamente a raiz quadrada da variância:

$$dp(\hat{\beta}_j) = \sigma / [SQT_j(1 - R_j^2)]^{1/2}.$$

Como σ é desconhecido, ele é substituído pelo seu estimador, $\hat{\sigma}$. Isso nos dá o **erro-padrão de $\hat{\beta}_j$** :

$$ep(\hat{\beta}_j) = \hat{\sigma} / [SQT_j(1 - R_j^2)]^{1/2}. \quad (3.58)$$

Assim como as estimativas de MQO podem ser obtidas para qualquer amostra dada, os erros-padrão também podem. Como $ep(\hat{\beta}_j)$ depende de $\hat{\sigma}$, o erro-padrão tem uma distribuição amostral, que desempenhará um papel importante no Capítulo 4.

Devemos enfatizar algo sobre os erros-padrão. Como (3.58) é obtido diretamente da fórmula da variância em (3.51), e como (3.51) se apóia na hipótese de homoscedasticidade RLM.5, a fórmula do erro-padrão em (3.58) *não* é um estimador válido de $d\bar{p}(\hat{\beta}_j)$ se os erros exibem heteroscedasticidade. Assim, enquanto a presença de heteroscedasticidade não causa viés em $\hat{\beta}_j$, ela leva viés da fórmula usual da $\text{Var}(\hat{\beta}_j)$, o que invalida, portanto, os erros-padrão. Isso é importante porque qualquer programa de regressão calcula (3.58) como o erro-padrão básico de cada coeficiente (com uma interpretação um pouco diferente para o intercepto). Se suspeitarmos de heteroscedasticidade, então os erros-padrão de MQO “habitualis” não são válidos, e alguma ação corretiva deve ser tomada. No Capítulo 8, veremos quais métodos estão disponíveis para trabalhar com a heteroscedasticidade.

3.5 EFICIÊNCIA DE MQO: O TEOREMA DE GAUSS-MARKOV

Nesta seção, apresentaremos e discutiremos o importante **Teorema de Gauss-Markov**, que justifica o uso do método de MQO em vez de usar uma variedade de estimadores concorrentes. Já conhecemos uma justificativa para MQO: sob as hipóteses RLM.1 a RLM.4, MQO é não-viesado. Entretanto, há *muitos* estimadores não-viesados de β_j sob essas hipóteses (por exemplo, veja o Problema 3.12). Poderia haver outros estimadores não-viesados com variâncias menores que as dos estimadores de MQO?

Se limitarmos apropriadamente a classe de estimadores concorrentes, podemos mostrar que MQO é o melhor dentro de sua classe. Especificamente, argumentamos que, sob as hipóteses RLM.1 a RLM.5, o estimador de MQO $\hat{\beta}_j$ para β_j é o **melhor estimador linear não-viesado** (*Best Linear Unbiased Estimator — BLUE*). A fim de formular o teorema, precisamos entender cada componente da sigla “BLUE”. Primeiro, sabemos o que é um estimador: ele é uma regra que pode ser aplicada a qualquer amostra de dados para produzir uma estimativa. Também sabemos o que é um estimador não-viesado: no contexto correto, um estimador, por exemplo $\tilde{\beta}_j$, de β_j é um estimador não-viesado de β_j se $E(\tilde{\beta}_j) = \beta_j$ para qualquer $\beta_0, \beta_1, \dots, \beta_k$.

E o que dizer sobre o significado do termo “linear”? No contexto atual, um estimador $\tilde{\beta}_j$ de β_j é linear se, e somente se, ele puder ser expresso como uma função linear dos dados da variável dependente:

$$\tilde{\beta}_j = \sum_{i=1}^n w_{ij} y_i \quad (3.59)$$

em que cada w_{ij} pode ser uma função dos valores amostrais de todas as variáveis independentes. Os estimadores de MQO são lineares, como pode ser visto na equação (3.22).

Finalmente, como definir “o melhor”? Para o teorema correto, o melhor é definido como a *variância menor*. Dados dois estimadores não-viesados, é lógico preferir aquele com a variância menor (veja o Apêndice C, disponível no site de Thomson).

Agora, vamos chamar de $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ os estimadores de MQO do modelo (3.31) sob as hipóteses RLM.1 a RLM.5. O Teorema de Gauss-Markov diz que, para qualquer estimador $\tilde{\beta}_j$ que é *linear e não-viesado*, $\text{Var}(\hat{\beta}_j) \leq \text{Var}(\tilde{\beta}_j)$, e a desigualdade é geralmente estrita. Em outras palavras, na classe dos estimadores lineares não-viesados, MQO tem a menor variância (sob as cinco hipóteses de Gauss-Markov). De fato, o teorema diz mais do que isso. Se desejarmos estimar qualquer função linear de β_j , a combinação linear correspondente dos estimadores de MQO alcança a menor variância entre todos os estimadores não-viesados. Vamos concluir com um teorema, o provado no Apêndice 3A, disponível no site da Thomson.

T E O R E M A 3 . 4 (TEOREMA DE GAUSS-MARKOV)

Sob as hipóteses RLM.1 a RLM.5, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ são os melhores estimadores lineares não-viesados (BLUEs) de $\beta_0, \beta_1, \dots, \beta_k$, respectivamente.

É por causa desse teorema que as hipóteses RLM.1 a RLM.5 são conhecidas como as hipóteses de Gauss-Markov (para a análise de corte transversal).

A importância do Teorema de Gauss-Markov é que, quando um conjunto padrão de hipóteses se mantém, não necessitamos procurar por estimadores não-viesados alternativos da forma expressa em (3.59): nenhum será melhor que MQO. Equivalentemente, se somos apresentados a um estimador que é tanto linear como não-viesado, então sabemos que a variância desse estimador é pelo menos tão grande quanto a variância de MQO; nenhum cálculo adicional é necessário para mostrar isso.

Para nossos propósitos, o Teorema 3.4 justifica o uso de MQO para estimar modelos de regressão múltipla. Se qualquer uma das hipóteses de Gauss-Markov for violada, o teorema não é mais válido. Já sabemos que a falha da hipótese de média condicional zero (hipótese RLM.3) faz com que MQO seja viesado, de modo que o Teorema 3.4 também deixa de ser válido. Também sabemos que a heteroscedasticidade (falha da hipótese RLM.5) não faz com que MQO seja viesado. Entretanto, MQO não tem mais a menor variância entre os estimadores lineares não-viesados na presença da heteroscedasticidade. No Capítulo 8, analisaremos um estimador que melhora MQO quando conhecemos o tipo da heteroscedasticidade.

1. O modelo de regressão múltipla nos permite, efetivamente, manter os outros fatores fixos ao examinarmos os efeitos de uma variável independente particular sobre a variável dependente. Ele permite, explicitamente, que as variáveis sejam correlacionadas.
2. Embora o modelo seja linear em seus *parâmetros*, ele pode ser usado para modelar relações não-lineares ao se escolher, apropriadamente, as variáveis dependente e independente.
3. O método de mínimos quadrados ordinários é facilmente aplicado para estimar o modelo de regressão múltipla. Cada estimativa de inclinação mede o efeito parcial da variável independente correspondente sobre a variável dependente, mantendo todas as outras variáveis independentes fixas.
4. R^2 é a proporção da variação amostral da variável dependente explicada pelas variáveis independentes, e é utilizado como uma medida do grau de ajuste. É importante não dar importância demais ao valor do R^2 na avaliação de modelos econômicos.
5. Sob as primeiras quatro hipóteses de Gauss-Markov (RLM.1 a RLM.4), os estimadores de MQO são não-viesados. Isso implica que incluir uma variável irrelevante em um modelo não tem nenhum efeito sobre a inexistência de viés dos estimadores de intercepto e de inclinação. De outro lado, omitir uma variável importante faz com que MQO seja viesado. Em muitas circunstâncias, a direção do viés pode ser determinada.
6. Sob as cinco hipóteses de Gauss-Markov, a variância de um estimador de inclinação de MQO é dada por $\text{Var}(\hat{\beta}_j) = \sigma^2 / [\text{SQT}_j(1 - R_j^2)]$. Quando a variância do erro σ^2 cresce, o mesmo ocorre com $\text{Var}(\hat{\beta}_j)$, enquanto $\text{Var}(\hat{\beta}_j)$ diminui quando a variação amostral em x_j , SQT_j , aumenta. O termo R_j^2 mede a magnitude da colinearidade entre x_j e as outras variáveis explicativas. Quando R_j^2 aproxima-se de um, $\text{Var}(\hat{\beta}_j)$ é ilimitada.

7. Adicionar uma variável irrelevante a uma equação geralmente aumenta as variâncias dos demais estimadores de MQO, por causa da multicolinearidade.
8. Sob as hipóteses de Gauss-Markov (RLM.1 a RLM.5), os estimadores de MQO são os melhores estimadores lineares não-viesados (BLUE).
-

3.1 Usando os dados do arquivo GPA2.RAW sobre 4.137 estudantes de curso superior nos Estados Unidos, estimou-se a seguinte equação por MQO:

$$\hat{nmgrad} = 1,392 - 0,0135 \text{ emperc} + 0,00148 \text{ sat}$$

$$n = 4.137, R^2 = 0,273,$$

em que $nmgrad$ é mensurada em uma escala de quatro pontos, $emperc$ é o percentil da turma de formados do ensino médio (definido de modo que, por exemplo, $emperc = 5$ significa os cinco por cento melhores da sala), e sat é uma nota média ponderada de matemática e habilidade verbal do estudante para ingresso em curso superior.

- (i) Por que faz sentido que o coeficiente de $emperc$ seja negativo?
- (ii) Qual é o valor previsto de $nmgrad$ quando $emperc = 20$ e $sat = 1.050$?
- (iii) Suponha que dois alunos do ensino médio, A e B, estejam no mesmo percentil no ensino médio, mas a nota sat do Estudante A foi 140 pontos maior (cerca de um desvio-padrão na amostra). Qual é a diferença prevista em $nmgrad$ para esses dois estudantes? A diferença é grande?
- (iv) Mantendo $emperc$ fixo, que diferença na nota sat levaria a uma diferença prevista de $nmgrad$ de 0,50? Comente sua resposta.

3.2 Os dados do arquivo WAGE2.RAW, sobre homens que trabalham, foram utilizados para estimar a seguinte equação:

$$\hat{educ} = 10,36 - 0,094 \text{ irms} + 0,131 \text{ educm} + 0,210 \text{ educp}$$

$$n = 722, R^2 = 0,214,$$

em que $educ$ é anos de escolaridade formal, $irms$ é o número de irmãos, $educm$ é anos de escolaridade formal da mãe e $educp$ é anos de escolaridade formal do pai.

- (i) $irms$ tem o efeito esperado? Explique. Mantendo $educm$ e $educp$ fixos, em quanto deveria $irms$ aumentar para reduzir os anos previstos da educação formal em um ano? (Uma resposta incompleta é aceitável aqui.)
- (ii) Discuta a interpretação do coeficiente de $educm$.
- (iii) Suponha que o Homem A não tenha irmãos, e sua mãe e seu pai tenham, cada um, 12 anos de educação formal. Suponha também que o Homem B não tenha irmãos, e sua mãe e seu pai tenham, cada um, 16 anos de educação formal. Qual é a diferença prevista em anos de educação formal entre B e A?

3.3 O modelo seguinte é uma versão simplificada do modelo de regressão múltipla usado por Biddle e Hamermesh (1990) para estudar a escolha entre o tempo gasto dormindo e trabalhando e para observar outros fatores que afetam o sono:

$$dormir = \beta_0 + \beta_1 trabtot + \beta_2 educ + \beta_3 idade + u,$$

em que *dormir* e *trabtot* (trabalho total) são mensurados em minutos por semana e *educ* e *idade* são mensurados em anos. (Veja também o Problema 2.12.)

- (i) Se os adultos escolhem entre dormir e trabalhar, qual é sinal de β_1 ?
- (ii) Que sinais você espera que β_2 e β_3 terão?
- (iii) Usando os dados do arquivo SLEEP75.RAW, a equação estimada é

$$\hat{dormir} = 3.638,25 - 0,148 trabtot - 11,13 educ + 2,20 idade \\ n = 706, R^2 = 0,113.$$

Se alguém trabalha 5 horas a mais por semana, qual é a queda, em minutos, no valor esperado de dormir? Esse valor representa uma escolha grande?

- (iv) Discuta o sinal e a magnitude do coeficiente de *educ*.
- (v) Você diria que *trabtot*, *educ* e *idade* explicam muito da variação de *dormir*? Quais outros fatores poderiam afetar o tempo gasto dormindo? É provável que eles sejam correlacionados com *trabtot*?

3.4 O salário inicial (mediano) para recém-formados em direito é determinado pela equação

$$\log(salárioim) = \beta_0 + \beta_1 lsat + \beta_2 nmgrad + \beta_3 \log(volbib) + \beta_4 \log(custo) + \beta_5 rank + u,$$

em que *lsat* é a nota mediana do *lsat* (nota de ingresso no curso de direito) dos recém-formados, *nmgrad* é a nota mediana dos recém-formados nas disciplinas do curso de direito, *volbib* é o número de volumes da biblioteca da escola de direito, *custo* é o custo anual da escola de direito e *rank* é a classificação da escola de direito (com *rank* = 1 sendo o melhor posto de classificação).

- (i) Explique a razão de esperarmos $\beta_5 \leq 0$.
- (ii) Quais são os sinais que você espera para os outros parâmetros de inclinação? Justifique sua resposta.
- (iii) Utilizando os dados do arquivo LAWSCH85.RAW, a equação estimada é

$$\begin{aligned} \hat{\log}(salárioim) &= 8,34 + 0,0047 lsat + 0,248 nmgrad + 0,095 \log(volbib) \\ &\quad + 0,038 \log(custo) - 0,0033 rank \\ &\quad n = 136, R^2 = 0,842. \end{aligned}$$

Qual é a diferença *ceteris paribus* prevista no salário para as escolas com um *nmgrad* médio diferente em um ponto? (Descreva sua resposta em percentual.)

- (iv) Interprete o coeficiente da variável $\log(volbib)$.
- (v) Você diria que é melhor freqüentar uma escola de direito que tem uma classificação melhor? Qual é a diferença no salário inicial esperado para uma escola que tem uma classificação igual a 20?

3.5 Em um estudo que relaciona a nota média em curso superior (*nmgrad*) ao tempo gasto em várias atividades, você distribui uma pesquisa para vários estudantes. Os estudantes devem responder quan-

tas horas eles despendem, em cada semana, em quatro atividades: estudo, sono, trabalho e lazer. Toda atividade é colocada em uma das quatro categorias, de modo que, para cada estudante, a soma das horas nas quatro atividades deve ser igual a 168.

- (i) No modelo

$$nungrad = \beta_0 + \beta_1 estudar + \beta_2 dormir + \beta_3 trabalhar + \beta_4 lazer + u,$$

faz sentido manter *dormir*, *trabalhar* e *lazer* fixos, enquanto *estudar* varia?

- (ii) Explique a razão de esse modelo violar a hipótese RLM.4.
- (iii) Como você poderia reformular o modelo, de modo que seus parâmetros tivessem uma interpretação útil e ele satisfizesse a hipótese RLM.4?

3.6 Considere o modelo de regressão múltipla contendo três variáveis independentes, sob as hipóteses RLM.1 a RLM.4:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

Você está interessado em estimar a soma dos parâmetros de x_1 e x_2 ; chame-a de $\theta_1 = \beta_1 + \beta_2$. Mostre que $\hat{\theta}_1 = \hat{\beta}_1 + \hat{\beta}_2$ é um estimador não-viesado de θ_1 .

3.7 Quais dos seguintes itens podem fazer com que os estimadores de MQO sejam viesados?

- (i) Heteroscedasticidade.
- (ii) Omitir uma variável importante.
- (iii) Um coeficiente de correlação amostral de 0,95 entre duas variáveis independentes incluídas no modelo.

3.8 Suponha que a produtividade média do trabalhador da indústria (*prodmed*) dependa de dois fatores — horas médias de treinamento do trabalhador (*treinmed*) e aptidão média do trabalhador (*aptidmed*):

$$prodmed = \beta_0 + \beta_1 treinmed + \beta_2 aptidmed + u.$$

Assuma que essa equação satisfaça as hipóteses de Gauss-Markov. Se um subsídio foi dado às empresas cujos trabalhadores têm uma aptidão menor do que a média, de modo que *treinmed* e *aptidmed* sejam negativamente correlacionados, qual é o provável viés em $\hat{\beta}_1$ obtido da regressão simples de *prodmed* sobre *treinmed*?

3.9 A equação seguinte descreve o preço mediano das residências de uma comunidade em termos da quantidade de poluição (*oxn*, de óxido nitroso) e do número médio de cômodos nas residências da comunidade (*comods*):

$$\log(preço) = \beta_0 + \beta_1 \log(oxn) + \beta_2 comods + u.$$

- (i) Quais são os prováveis sinais de β_1 e β_2 ? Qual é a interpretação de β_1 ? Explique.
- (ii) Por que *oxn* [ou, mais precisamente, $\log(oxn)$] e *comods* deveriam ser negativamente correlacionados? Se esse é o caso, a regressão simples de $\log(preço)$ sobre $\log(oxn)$ produz um estimador viesado para cima ou para baixo de β_1 ?

(iii) Utilizando os dados do arquivo HPRICE2.RAW foram estimadas as seguintes equações:

$$\log(\hat{p}\hat{re}ço) = 11,71 - 1,043 \log(oxn), n = 506, R^2 = 0,264.$$

$$\log(\hat{p}\hat{re}ço) = 9,23 - 0,718 \log(oxn) + 0,306 comods, n = 506, R^2 = 0,514.$$

A relação entre as estimativas da elasticidade do *preço* das regressões simples e múltipla é a que você previu, tomando como base suas respostas na parte (ii)? Pode-se dizer que $-0,718$ está claramente mais próximo da elasticidade verdadeira que $-1,043$?

3.10 Suponha que o modelo populacional que determina y seja:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u,$$

e esse modelo satisfaz as hipóteses de Gauss-Markov. Entretanto, estimamos o modelo que omite x_3 . Sejam $\tilde{\beta}_0$, $\tilde{\beta}_1$ e $\tilde{\beta}_2$ os estimadores de MQO da regressão de y sobre x_1 e x_2 . Mostre que o valor esperado de $\tilde{\beta}_1$ (dados os valores das variáveis independentes da amostra) é

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3 \frac{\sum_{i=1}^n \hat{r}_{i1} x_{i3}}{\sum_{i=1}^n \hat{r}_{i1}^2},$$

em que os \hat{r}_{i1} são os resíduos de MQO da regressão de x_1 sobre x_2 . [Sugestão: a fórmula de $\tilde{\beta}_1$ é proveniente da equação (3.22). Coloque $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i$ nessa equação. Após alguma álgebra, aplique o operador esperança, tratando x_{i3} e \hat{r}_{i1} como não-aleatórios.]

3.11 A seguinte equação representa os efeitos das receitas totais de impostos sobre o crescimento subsequente do emprego para a população de municípios dos Estados Unidos:

$$cresc = \beta_0 + \beta_1 parc_p + \beta_2 parc_r + \beta_3 parc_v + outros fatores,$$

em que $cresc$ é a variação percentual do emprego de 1980 a 1990, enquanto o total das receitas de impostos tem a seguinte distribuição: $parc_p$ é a parcela dos impostos sobre a propriedade, $parc_r$ é a parcela das receitas de impostos sobre a renda e $parc_v$ é a parcela das receitas de impostos sobre as vendas. Todas essas variáveis estão mensuradas em 1980. A parcela omitida, $parc_v$, inclui taxas e impostos variados. Por definição, as quatro parcelas somam um. Outros fatores incluiriam despesas com educação, infra-estrutura, e assim por diante (todos mensurados em 1980).

- (i) Por que devemos omitir uma das variáveis de parcela de impostos da equação?
 - (ii) Dê uma interpretação cuidadosa de β_1 .
- 3.12** (i) Considere o modelo de regressão simples $y = \beta_0 + \beta_1 x + u$, sob as primeiras quatro hipóteses de Gauss-Markov. Para alguma função $g(x)$, por exemplo, $g(x) = x^2$ ou $g(x) = \log(1 + x^2)$, defina $z_i = g(x_i)$. Defina um estimador de inclinação como

$$\tilde{\beta}_1 = \left(\sum_{i=1}^n (z_i - \bar{z}) y_i \right) / \left(\sum_{i=1}^n (z_i - \bar{z}) x_i \right).$$

Mostre que $\hat{\beta}_1$ é linear e não-viesado. Lembre-se: como $E(u|x) = 0$, você pode tratar tanto x_i como z_i como não-aleatórios em sua derivação.

- (ii) Acrescente a hipótese de homoscedasticidade, RLM.5. Mostre que

$$\text{Var}(\hat{\beta}_1) = \sigma^2 \left(\sum_{i=1}^n (z_i - \bar{z})^2 \right) / \left(\sum_{i=1}^n (z_i - \bar{z}) x_i \right)^2.$$

- (iii) Mostre diretamente que, sob as hipóteses de Gauss-Markov, $\text{Var}(\hat{\beta}_1) \leq \text{Var}(\tilde{\beta}_1)$, em que $\hat{\beta}_1$ é o estimador de MQO. [Sugestão: a desigualdade de Cauchy-Schwartz do Apêndice B (disponível no site da Thomson) implica que

$$\left(n^{-1} \sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x}) \right)^2 \leq \left(n^{-1} \sum_{i=1}^n (z_i - \bar{z})^2 \right) \left(n^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right);$$

Observe que podemos retirar \bar{x} da covariância amostral.]

3A.1 Derivação das Condições de Primeira Ordem da Equação (3.13)

A análise é muito similar à do caso da regressão simples. Devemos caracterizar as soluções para o problema

$$\min_{b_0, b_1, \dots, b_k} \sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - \dots - b_k x_{ik})^2.$$

Considerando as derivadas parciais em relação a cada um dos b_j (veja o Apêndice A, disponível no site da Thomson), avaliando-as nas soluções e igualando-as a zero resulta

$$\begin{aligned} -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0 \\ -2 \sum_{i=1}^n x_{ij} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) &= 0, \text{ para todo } j = 1, \dots, k. \end{aligned}$$

Cancelando -2 obtemos as condições de primeira ordem em (3.13).

3A.2 Derivação da Equação (3.22)

Para derivar (3.22), escreva x_{ij} em termos de seus valores estimados e seus resíduos a partir da regressão de x_1 sobre x_2, \dots, x_k : $x_{ij} = \hat{x}_{ij} + \hat{r}_{ij}$, para todo $i = 1, \dots, n$. Agora, insira essa expressão na segunda equação de (3.13):

$$\sum_{i=1}^n (\hat{x}_{ij} + \hat{r}_{ij})(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0 \quad (3.60)$$

Pela definição do resíduo de MQO \hat{u}_i , como \hat{x}_{i1} é exatamente uma função linear das variáveis explicativas x_{i2}, \dots, x_{ik} , segue que $\sum_{i=1}^n \hat{x}_{i1} \hat{u}_i = 0$. Portanto, a equação (3.60) pode ser expressa como

$$\sum_{i=1}^n \hat{r}_{i1} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0. \quad (3.61)$$

Como os \hat{r}_{i1} são os resíduos da regressão de x_1 sobre x_2, \dots, x_k , $\sum_{i=1}^n x_{ij} \hat{r}_{i1} = 0$, para todo $j = 2, \dots, k$. Portanto, (3.61) é equivalente a $\sum_{i=1}^n \hat{r}_{i1} (y_i - \hat{\beta}_1 x_{i1}) = 0$. Finalmente, usamos o fato de que $\sum_{i=1}^n \hat{x}_{i1} \hat{r}_{i1} = 0$, o que significa que $\hat{\beta}_1$ soluciona

$$\sum_{i=1}^n \hat{r}_{i1} (y_i - \hat{\beta}_1 \hat{r}_{i1}) = 0.$$

Agora, com um pouco de álgebra chegamos a (3.22), desde que, evidentemente, $\sum_{i=1}^n \hat{r}_{i1}^2 > 0$; isso é garantido pela hipótese RLM.4.

3A.3 Prova do Teorema 3.1

Vamos provar o Teorema 3.1 para $\hat{\beta}_1$; a prova para os outros parâmetros de inclinação é, virtualmente, idêntica. (Veja o Apêndice E, disponível no site da Thomson para uma prova mais sucinta utilizando matrizes.) Sob a hipótese RLM.4, os estimadores de MQO existem, e podemos escrever $\hat{\beta}_1$ como em (3.22). Sob a hipótese RLM.1, podemos escrever y_i como em (3.32); substitua-o pelo y_i de (3.22). Então, usando $\sum_{i=1}^n \hat{r}_{i1} = 0$, $\sum_{i=1}^n x_{ij} \hat{r}_{i1} = 0$, para todo $j = 2, \dots, k$ e $\sum_{i=1}^n x_{j1} \hat{r}_{i1} = \sum_{i=1}^n \hat{r}_{i1}^2$, temos

$$\hat{\beta}_1 = \beta_1 + \left(\sum_{i=1}^n \hat{r}_{i1} u_i \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right). \quad (3.62)$$

Agora, sob as hipóteses RLM.2 e RLM.3, o valor esperado de cada u_i , dadas todas as variáveis independentes na amostra, é zero. Como os \hat{r}_{ij} são justamente funções das variáveis independentes da amostra, segue-se que

$$E(\hat{\beta}_1 | X) = \beta_1 + \left(\sum_{i=1}^n \hat{r}_{i1} E(u_i | X) \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right)$$

$$= \beta_1 + \left(\sum_{i=1}^n \hat{r}_{i1} \cdot 0 \right) / \left(\sum_{i=1}^n \hat{r}_{i1}^2 \right) = \beta_1,$$

em que X representa os dados de todas as variáveis independentes, e $E(\hat{\beta}_1 | X)$ é o valor esperado de $\hat{\beta}_1$, dado x_{i1}, \dots, x_{ik} , para todo $i = 1, \dots, n$. Isso completa a prova.

3A.4 Viés de Variável Omitida no Modelo Geral

Podemos derivar o viés de variável omitida no modelo geral da equação (3.31) sob as quatro primeiras hipóteses de Gauss-Markov. Em particular, sejam $\hat{\beta}_j, j = 0, 1, \dots, k$ os estimadores de MQO da regressão ao se usar o conjunto completo de variáveis explicativas. Sejam $\tilde{\beta}_j, j = 0, 1, \dots, k-1$ os estimadores de MQO da regressão que omite x_k . Sejam $\tilde{\delta}_j, j = 1, \dots, k-1$ os coeficientes de inclinação de x_i da regressão auxiliar de x_{ik} sobre $x_{i1}, x_{i2}, \dots, x_{i,k-1}$, $i = 1, \dots, n$. Um fato útil é que

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j \quad (3.63)$$

Isso mostra explicitamente que, quando não controlamos x_k na regressão, o efeito parcial estimado de x_j é igual ao efeito parcial quando incluímos x_k mais o efeito parcial de x_k sobre \hat{y} vezes a relação parcial entre a variável omitida, x_k , e $x_j, j < k$. Condicionado ao conjunto inteiro de variáveis explicativas, X , sabemos que os $\hat{\beta}_j$ são todos não-viesados para os correspondentes $\beta_j, j = 1, \dots, k$. Além disso, como $\tilde{\delta}_j$ é exatamente uma função de X , temos

$$\begin{aligned} E(\tilde{\beta}_j | X) &= E(\hat{\beta}_j | X) + E(\hat{\beta}_k | X) \tilde{\delta}_j \\ &= \beta_j + \beta_k \tilde{\delta}_j \end{aligned} \quad (3.64)$$

A equação (3.64) mostra que $\tilde{\beta}_j$ é viesado para β_j , a menos que $\beta_k = 0$ — caso em que x_k não tem efeito parcial na população —, ou $\tilde{\delta}_j$ é igual a zero, o que significa que x_{ik} e x_{ij} são parcialmente não-correlacionados na amostra. A chave para obter a equação (3.64) é a equação (3.63). Para mostrar a equação (3.63), podemos usar a equação (3.22) várias vezes. Por simplicidade, vamos olhar para $j = 1$. Agora, $\tilde{\beta}_1$ é o coeficiente de inclinação da regressão simples de y_i sobre \tilde{r}_{i1} , $i = 1, \dots, n$, em que os \tilde{r}_{i1} são os resíduos de MQO da regressão de x_{i1} sobre $x_{i2}, x_{i3}, \dots, x_{i,k-1}$. Considere o numerador da expressão de $\tilde{\beta}_1$: $\sum_{i=1}^n \tilde{r}_{i1} y_i$. Para cada i , podemos escrever $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik} + \hat{u}_i$ e colocar em y_i . Agora, pelas propriedades dos resíduos de MQO, os \tilde{r}_{i1} têm média amostral zero e são não-correlacionados com $x_{i2}, x_{i3}, \dots, x_{i,k-1}$ na amostra. Semelhantemente, os \hat{u}_i têm média amostral zero e correlação amostral zero com $x_{i1}, x_{i2}, \dots, x_{ik}$. Segue-se que os \tilde{r}_{i1} e \hat{u}_i são não-correlacionados na amostra (visto que os \tilde{r}_{i1} são exatamente uma combinação linear de $x_{i1}, x_{i2}, \dots, x_{i,k-1}$). Assim

$$\sum_{i=1}^n \tilde{r}_{i1} y_i = \hat{\beta}_1 \left(\sum_{i=1}^n \tilde{r}_{i1} x_{i1} \right) + \hat{\beta}_k \left(\sum_{i=1}^n \tilde{r}_{i1} x_{ik} \right). \quad (3.65)$$

Agora, $\sum_{i=1}^n \tilde{r}_{i1} x_{i1} = \sum_{i=1}^n \tilde{r}_{i1}^2$, que é também o denominador de $\tilde{\beta}_1$. Portanto, mostramos que

$$\begin{aligned} \tilde{\beta}_1 &= \hat{\beta}_1 + \hat{\beta}_k \left(\sum_{i=1}^n \tilde{r}_{i1} x_{ik} \right) / \left(\sum_{i=1}^n \tilde{r}_{i1}^2 \right) \\ &= \hat{\beta}_1 + \hat{\beta}_k \tilde{\delta}_1. \end{aligned}$$

Essa é a relação que queríamos mostrar.

3A.5 Prova do Teorema 3.2

Novamente, vamos provar o teorema para $j = 1$. Escreva $\hat{\beta}_1$ como na equação (3.62). Agora, sob RLM.5, $\text{Var}(u_i|X) = \sigma^2$, para todo $i = 1, \dots, n$. Sob amostragem aleatória, os u_i são independentes, mesmo condicionados a X , e os \hat{r}_{ii} são não-aleatórios condicionados a X . Portanto,

$$\begin{aligned}\text{Var}(\hat{\beta}_1|X) &= \left(\sum_{i=1}^n \hat{r}_{ii}^2 \text{Var}(u_i|X) \right) / \left(\sum_{i=1}^n \hat{r}_{ii}^2 \right)^2 \\ &= \left(\sum_{i=1}^n \hat{r}_{ii}^2 \sigma^2 \right) / \left(\sum_{i=1}^n \hat{r}_{ii}^2 \right)^2 = \sigma^2 / \left(\sum_{i=1}^n \hat{r}_{ii}^2 \right).\end{aligned}$$

Agora, visto que $\sum_{i=1}^n \hat{r}_{ii}^2$ é a soma dos quadrados dos resíduos da regressão de x_1 sobre x_2, \dots, x_k , $\sum_{i=1}^n \hat{r}_{ii}^2 = \text{SQT}_1(1 - R_1^2)$. Isso completa a prova.

3A.6 Prova do Teorema 3.4

Mostramos que, para qualquer outro estimador linear não-viesado $\tilde{\beta}_1$ de β_1 , $\text{Var}(\tilde{\beta}_1) \geq \text{Var}(\hat{\beta}_1)$, em que $\hat{\beta}_1$ é o estimador de MQO. Não se perde generalidade ao jogarmos o foco em $j = 1$.

Para $\tilde{\beta}_1$ como na equação (3.59), podemos inserir em y_i para obter

$$\tilde{\beta}_1 = \beta_0 \sum_{i=1}^n w_{i1} + \beta_1 \sum_{i=1}^n w_{i1} x_{i1} + \beta_2 \sum_{i=1}^n w_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n w_{i1} x_{ik} + \sum_{i=1}^n w_{i1} u_i.$$

Agora, visto que os w_{ij} são funções de x_{ij} ,

$$\begin{aligned}\text{E}(\tilde{\beta}_1|X) &= \beta_0 \sum_{i=1}^n w_{i1} + \beta_1 \sum_{i=1}^n w_{i1} x_{i1} + \beta_2 \sum_{i=1}^n w_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n w_{i1} x_{ik} + \sum_{i=1}^n w_{i1} \text{E}(u_i|X) \\ &= \beta_0 \sum_{i=1}^n w_{i1} + \beta_1 \sum_{i=1}^n w_{i1} x_{i1} + \beta_2 \sum_{i=1}^n w_{i1} x_{i2} + \dots + \beta_k \sum_{i=1}^n w_{i1} x_{ik}\end{aligned}$$

porque $\text{E}(u_i|X) = 0$, para todo $i = 1, \dots, n$, sob RLM.2 e RLM.3. Portanto, para $\text{E}(\tilde{\beta}_1|X)$ igualar-se a β_1 para quaisquer valores dos parâmetros, devemos ter

$$\sum_{i=1}^n w_{i1} = 0, \quad \sum_{i=1}^n w_{i1} x_{i1} = 1, \quad \sum_{i=1}^n w_{i1} x_{ij} = 0, \quad j = 2, \dots, k. \quad (3.66)$$

Agora, sejam \hat{r}_{ii} os resíduos da regressão de x_{i1} sobre x_{i2}, \dots, x_{ik} . Então, de (3.66), segue-se que

$$\sum_{i=1}^n w_{i1} \hat{r}_{ii} = 1 \quad (3.67)$$

visto que $x_{i1} = \hat{x}_{i1} + \hat{r}_{i1}$ e $\sum_{i=1}^n w_{i1} \hat{x}_{i1} = 0$. Agora, considere a diferença entre $\text{Var}(\tilde{\beta}_1|X)$ e $\text{Var}(\hat{\beta}_1|X)$ sob RLM.1 a RLM.5:

$$\sigma^2 \sum_{i=1}^n w_{ii}^2 = \sigma^2 \left(\sum_{i=1}^n \hat{r}_{ii}^2 \right). \quad (3.68)$$

Por causa de (3.67), podemos escrever a diferença em (3.68), sem σ^2 , como

$$\sum_{i=1}^n w_{ii}^2 - \left(\sum_{i=1}^n w_{ii} \hat{r}_{ii} \right)^2 \left/ \left(\sum_{i=1}^n \hat{r}_{ii}^2 \right) \right.. \quad (3.69)$$

No entanto, (3.69) é simplesmente

$$\sum_{i=1}^n (w_{ii} - \hat{\gamma}_1 \hat{r}_{ii})^2, \quad (3.70)$$

em que $\hat{\gamma}_1 = \left(\sum_{i=1}^n w_{ii} \hat{r}_{ii} \right) \left/ \left(\sum_{i=1}^n \hat{r}_{ii}^2 \right) \right.$, como pode ser visto ao se elevar cada termo em (3.70) ao quadrado, somando e cancelando os termos. Como (3.70) é exatamente a soma dos resíduos quadrados da regressão simples de w_{ii} sobre \hat{r}_{ii} — lembre-se de que a média amostral de \hat{r}_{ii} é zero —, (3.70) deve ser não-negativo. Isso completa a prova.

Análise de Regressão Múltipla: Inferência

ste capítulo continua o estudo da análise de regressão múltipla. Vamos nos voltar agora para o problema de testar hipóteses sobre os parâmetros do modelo da regressão populacional.

Iniciaremos encontrando as distribuições dos estimadores de MQO sob a hipótese adicional de que o erro populacional é normalmente distribuído. As seções 4.2 e 4.3 cobrem os testes de hipóteses sobre os parâmetros individuais, enquanto a Seção 4.4 discute como testar uma única hipótese que envolve mais de um parâmetro. Na Seção 4.5, vamos focalizar o teste de restrições múltiplas, bem como dedicar atenção especial em determinar se um grupo de variáveis independentes pode ser omitido do modelo.

4.1 DISTRIBUIÇÕES AMOSTRAIS DOS ESTIMADORES DE MQO

Até este ponto, construímos um conjunto de hipóteses sob as quais o método MQO é não-viesado; também derivamos e discutimos o viés causado por variáveis omitidas. Na Seção 3.4, obtivemos as variâncias dos estimadores de MQO sob as hipóteses de Gauss-Markov. Na seção 3.5, mostramos que essa variância é a menor entre os estimadores lineares não-viesados.

Conhecer o valor esperado e a variância dos estimadores de MQO é útil para descrever sua precisão. Entretanto, para a inferência estatística necessitamos conhecer mais do que apenas os dois primeiros momentos de $\hat{\beta}_j$; precisamos conhecer a distribuição amostral completa de $\hat{\beta}_j$. Mesmo sob as hipóteses de Gauss-Markov, a distribuição de $\hat{\beta}_j$ pode ter, virtualmente, qualquer forma.

Quando estabelecemos um condicionamento aos valores das variáveis independentes de nossa amostra, fica claro que as distribuições amostrais dos estimadores de MQO dependem da distribuição subjacente dos erros. Para tornar as distribuições amostrais de $\hat{\beta}_j$ passíveis de tratamento, vamos assumir agora que o erro não-observado é *normalmente distribuído* na população. Chamamos isso de **hipótese da normalidade**.

H I P Ó T E S E R L M . 6 (NORMALIDADE)

O erro populacional u é independente das variáveis explicativas x_1, x_2, \dots, x_k e é normalmente distribuído, com média zero e variância σ^2 : $u \sim \text{Normal}(0, \sigma^2)$.

A hipótese RLM.6 é muito mais forte que qualquer uma das nossas hipóteses anteriores. De fato, como u é independente de x_j sob RLM.6, $E(u|x_1, \dots, x_k) = E(u) = 0$ e $\text{Var}(u|x_1, \dots, x_k) = \text{Var}(u) = \sigma^2$. Assim, ao fazermos a hipótese RLM.6, necessariamente estamos assumindo RLM.3 e RLM.5. Para enfatizar que estamos assumindo mais do que antes, vamos nos referir ao conjunto completo de hipóteses RLM.1 a RLM.6.

Nas aplicações da regressão de corte transversal, as hipóteses RLM.1 a RLM.6 são chamadas **hipóteses do modelo linear clássico (MLC)**. Assim, vamos nos referir ao modelo sob essas seis hipóteses como o **modelo linear clássico**. É melhor pensar as hipóteses do MLC como contendo todas as hipóteses de Gauss-Markov *mais* a hipótese de um termo erro normalmente distribuído.

Sob as hipóteses do MLC, os estimadores de MQO $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ têm uma propriedade de eficiência mais forte do que teriam sob as hipóteses de Gauss-Markov. Pode-se mostrar que os estimadores de MQO são os **estimadores não-viesados de variância mínima**, o que significa que MQO tem a menor variância entre os estimadores não-viesados; não temos mais de restringir nossa comparação com os estimadores que são lineares em y_i . Essa propriedade de MQO sob as hipóteses do MLC é discutida mais adiante, no Apêndice E.

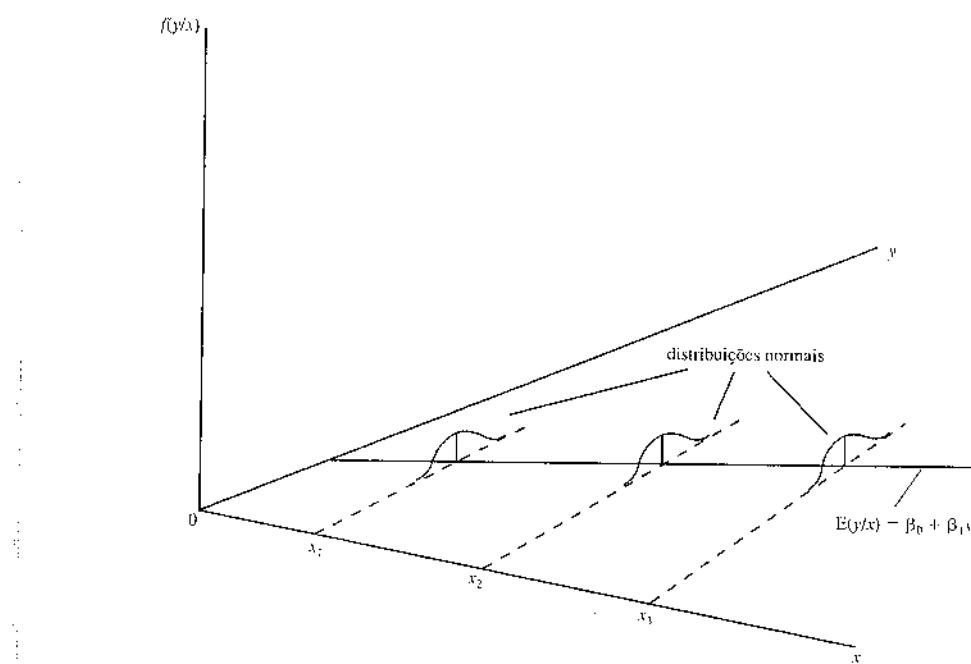
Uma maneira sucinta de resumir as hipóteses do MLC na população é

$$y|x \sim \text{Normal}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \sigma^2),$$

em que x é, novamente, uma maneira de escrever (x_1, \dots, x_k) . Assim, condicionado a x , y tem uma distribuição normal com média linear em x_1, \dots, x_k e uma variância constante. Para uma única variável independente x , essa situação está ilustrada na Figura 4.1.

Figura 4.1

A distribuição normal homoscedástica com uma única variável explicativa.



O argumento para justificar a distribuição normal dos erros é usualmente este: como u é a soma de muitos fatores diferentes não-observados que afetam y , podemos invocar o teorema do limite central (veja Apêndice C, disponível na página do livro, no site www.thomsonlearning.com.br) para concluir que u tem uma distribuição normal aproximada. Esse argumento tem algum mérito, mas não sem debilidades. Primeiro, os fatores em u podem ter distribuições muito diferentes na população (por exemplo, a aptidão e a qualidade da educação no erro de uma equação de salário). Embora o teorema do limite central (TLC) possa ainda ser válido em tais casos, a aproximação à normal pode ser insatisfatória, dependendo de quantos fatores aparecem em u e de quão diferentes são suas distribuições.

Um problema mais sério com o argumento do TLC é que ele assume que todos os fatores não-observados afetam y de um modo separado e aditivo. Nada garante que isso seja assim. Se u é uma função complicada dos fatores não-observados, então o argumento do TLC realmente não se aplica.

Em qualquer aplicação, saber se a normalidade de u pode ser assumida é uma questão empírica. Por exemplo, não há teorema dizendo que salárioh condicionado a educ , exper e perm é normalmente distribuído. De qualquer modo, o simples raciocínio sugere que o oposto é verdadeiro: visto que o salário por hora nunca pode ser menor que zero, ele não pode, estritamente falando, ter uma distribuição normal. Além disso, visto que há leis de salário mínimo, alguma fração da população ganha exatamente o salário mínimo, o que viola a hipótese de normalidade. Confuso, como uma questão prática, podemos perguntar se a distribuição condicional do salário está “próxima” de ser normal. A evidência empírica passada sugere que normalidade não é uma boa hipótese para os salários.

Freqüentemente, fazer uma transformação, especialmente tomando o log, produz uma distribuição que está mais próxima da normal. Por exemplo, algo como $\log(\text{preço})$ tende a ter uma distribuição que parece mais normal do que a distribuição de preço . Uma vez mais, essa é uma questão empírica. Discutiremos as consequências da normalidade para a inferência estatística no Capítulo 5.

Há alguns exemplos em que RLM.6 é claramente falsa. Sempre que y assume apenas alguns valores, ela não pode ter uma distribuição próxima de uma distribuição normal. A variável dependente do Exemplo 3.5 dá um bom exemplo. A variável $npre86$, o número de vezes que um homem jovem foi preso em 1986, assume um conjunto pequeno de valores inteiros e é igual a zero para a maioria dos homens. Assim, $npre86$ está longe de ser normalmente distribuída. O que pode ser feito nesses casos? Como veremos no Capítulo 5 – e isso é importante –, a normalidade dos erros não é um problema sério com tamanhos grandes de amostra. Por ora, vamos apenas fazer a hipótese da normalidade.

A normalidade do termo erro traduz-se nas distribuições normais amostrais dos estimadores de MQO:

T E O R E M A 4 . 1 (DISTRIBUIÇÕES AMOSTRAIS NORMAIS)

Sob as hipóteses do MLC, RLM.1 a RLM.6, condicional aos valores amostrais das variáveis independentes,

$$\hat{\beta}_j \sim \text{Normal}[\beta_j, \text{Var}(\hat{\beta}_j)], \quad [4.1]$$

onde $\text{Var}(\hat{\beta}_j)$ foi estudada no Capítulo 3 [equação (3.51)]. Portanto,

$$(\hat{\beta}_j - \beta_j)/\text{dp}(\hat{\beta}_j) \sim \text{Normal}(0,1).$$

A prova de (4.1) não é tão difícil, dadas as propriedades das variáveis aleatórias normalmente distribuídas descritas no Apêndice B, no site da Thomson. Cada $\hat{\beta}_j$ pode ser escrito como $\hat{\beta}_j = \beta_j + \sum_{i=1}^n w_{ij} u_i$, onde $w_{ij} = \hat{r}_{ij}/\text{SQR}_j$, \hat{r}_{ij} é o i -ésimo resíduo da regressão de x_j sobre todas as outras variáveis independentes, e SQR_j é a soma dos resíduos quadrados dessa regressão [veja a equação (3.62)]. Como os w_{ij} dependem somente das variáveis independentes, eles podem ser tratados com não-aleatórios. Assim, $\hat{\beta}_j$ é exatamente uma combinação linear dos erros na amostra, $\{u_i; i = 1, 2, \dots, n\}$. Sob a hipótese RLM.6 (e a hipótese de amostragem aleatória RLM.2), os erros são variáveis aleatórias independentes e identicamente distribuídas com distribuição Normal $(0, \sigma^2)$. Um fato importante sobre variáveis aleatórias normais independentes é que uma combinação linear de tais variáveis é normalmente distribuída (veja Apêndice B, disponível no site da Thomson). Isso basicamente completa a prova. Na Seção 3.3, vimos que $E(\hat{\beta}_j) = \beta_j$, e derivamos a $\text{Var}(\hat{\beta}_j)$ na Seção 3.4; não há necessidade de derivarmos novamente essas expressões.

Suponha que u é independente das variáveis explicativas, e que assume os valores $-2, -1, 0, 1$ e 2 com probabilidade igual a $1/5$. Isto infringe as hipóteses de Gauss-Markov? Isto infringe as hipóteses do MLC?

A segunda parte deste teorema segue imediatamente do fato de que, quando padronizamos uma variável aleatória normal ao subtrair dela sua média e dividi-la pelo seu desvio-padrão, obtemos uma variável aleatória normal padronizada.

As conclusões do Teorema 4.1 podem ser fortalecidas. Além de (4.1), qualquer combinação linear dos $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ também é normalmente distribuída, e qualquer subconjunto dos $\hat{\beta}_j$ tem uma distribuição normal conjunta. Esses fatos estão na base dos resultados dos testes no restante deste capítulo. No Capítulo 5 mostraremos que a normalidade dos estimadores de MQO ainda é *aproximadamente* verdadeira em amostras grandes, mesmo sem normalidade dos erros.

4.2 TESTES DE HIPÓTESES SOBRE UM ÚNICO PARÂMETRO POPULACIONAL: O TESTE t

Esta Seção cobre o importante tópico de testar hipóteses sobre um único parâmetro da função de regressão populacional. O modelo populacional pode ser escrito como

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u, \quad (4.2)$$

e assumimos que ele satisfaz as hipóteses do MLC. Sabemos que MQO produz estimadores não-viesados de β_j . Nesta seção, estudaremos como testar hipóteses sobre um particular β_j . Para um entendimento completo dos testes de hipóteses, devemos recordar que os β_j são características desconhecidas da população, e nunca os conheceremos com certeza. No entanto, podemos fazer *hipóteses* sobre o valor de β_j , e, em seguida, utilizar inferência estatística para testar nossa hipótese.

A fim de construir os testes de hipóteses, precisamos do seguinte resultado:

TEOREMA 4 . 2 (A DISTRIBUIÇÃO t PARA OS ESTIMADORES PADRONIZADOS)

Sob as hipóteses do MLC, RLM.1 a RLM.6,

$$(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) \sim t_{n-k-1}, \quad (4.3)$$

em que $k + 1$ é o número de parâmetros desconhecidos do modelo populacional $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$ (k parâmetros de inclinação mais o intercepto β_0).

Esse resultado difere do Teorema 4.1 em alguns aspectos notáveis. O Teorema 4.1 mostrou que, sob as hipóteses do MLC, $(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) \sim \text{Normal}(0, 1)$. A distribuição t em (4.3) é proveniente do fato de que a constante σ em $\text{ep}(\hat{\beta}_j)$ foi substituída pela variável aleatória $\hat{\sigma}$. A prova de que isso leva a uma distribuição t com $n - k - 1$ graus de liberdade não é particularmente perceptida. Essencialmente, a prova mostra que (4.3) pode ser escrita como a razão da variável aleatória normal padronizada $(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j)$ sobre a raiz quadrada de $\hat{\sigma}^2/\sigma^2$. Pode-se mostrar que essas variáveis aleatórias são independentes e $(n - k - 1) \hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-k-1}$. O resultado decorre da definição de uma variável aleatória t (veja Seção B.5 disponível no site da Thomson).

O Teorema 4.2 é importante porque ele nos permite testar hipóteses que envolvem os β_j . Na maioria das aplicações, nosso principal interesse é testar a **hipótese nula**

$$H_0: \beta_j = 0, \quad (4.4)$$

em que j corresponde a qualquer uma das k variáveis independentes. É importante entender o que (4.4) significa e ser capaz de descrever essas hipóteses em uma linguagem simples em uma determinada aplicação. Como β_j mede o efeito parcial de x_j sobre (o valor esperado de) y , após controlar todas as outras variáveis independentes, (4.4) significa que, uma vez que $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_k$ foram considerados, x_j não tem *nenhum efeito* sobre o valor esperado de y . Não podemos expressar a hipótese nula como “ x_j tem realmente um efeito parcial sobre y ” porque isso é verdadeiro para qualquer outro valor de β_j que não zero. O teste clássico é apropriado para testar *hipóteses simples* como (4.4).

Como um exemplo, considere a equação do salário

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

A hipótese nula $H_0: \beta_2 = 0$ significa que, uma vez que a educação formal e a permanência foram consideradas, o número de anos no mercado de trabalho (*exper*) não tem nenhum efeito sobre o salário horário. Essa é uma hipótese economicamente interessante. Se ela é verdadeira, ela implica que o histórico de trabalho de uma pessoa, anterior ao emprego atual, não afeta o salário. Se $\beta_2 > 0$, então a experiência prévia de trabalho contribui para a produtividade e, portanto, para o salário.

Você provavelmente se lembra, de seus cursos de estatística, dos rudimentos do teste de hipótese para a média de uma população normal. (Há uma revisão no Apêndice C, no site da Thomson.) Os mecanismos do teste (4.4) no contexto da regressão múltipla são muitos semelhantes. A parte difícil está em obter as estimativas dos coeficientes, os erros-padrão e os valores críticos, mas a maior parte desse trabalho é feita automaticamente por programas econométricos. Nossa trabalho é aprender como o resultado da regressão pode ser usado para testar as hipóteses de interesse.

A estatística que usamos para testar (4.4) (contra qualquer alternativa) é chamada “a” **estatística t** ou “a” **razão t** de $\hat{\beta}_j$, e é definida como

$$t_{\hat{\beta}_j} = \hat{\beta}_j / \text{ep}(\hat{\beta}_j). \quad (4.5)$$

Colocamos “a” entre aspas porque, como veremos em breve, para testar outras hipóteses sobre é necessária uma forma mais geral da estatística t . Por ora, é importante saber que (4.5) é apropriada somente para testar (4.4). Quando não houver possibilidade de causar nenhuma confusão, algumas vezes escreveremos t no lugar de $t_{\hat{\beta}_j}$.

A estatística t de $\hat{\beta}_j$ é simples de calcular, dados $\hat{\beta}_j$ e seu erro-padrão. De fato, a maioria dos programas de regressão faz a divisão automaticamente e informa a estatística t juntamente com cada coeficiente e seu erro-padrão.

Antes de discutir como usar (4.5) para testar, formalmente, $H_0: \beta_j = 0$, é útil ver porque $t_{\hat{\beta}_j}$ tem características que o tornam razoável enquanto uma estatística de teste para detectar $\beta_j \neq 0$. Em primeiro lugar, como $\text{ep}(\hat{\beta}_j)$ é sempre positivo, $t_{\hat{\beta}_j}$ tem o mesmo sinal de $\hat{\beta}_j$: se $\hat{\beta}_j$ é positivo, então do mesmo modo é $t_{\hat{\beta}_j}$, e se $\hat{\beta}_j$ é negativo, igualmente negativo é $t_{\hat{\beta}_j}$. Segundo, para um dado valor de $\text{ep}(\hat{\beta}_j)$, um valor maior de $\hat{\beta}_j$ leva a valores maiores de $t_{\hat{\beta}_j}$. Se $\hat{\beta}_j$ fica mais negativo, do mesmo modo fica $t_{\hat{\beta}_j}$.

Como estamos testando $H_0: \beta_j = 0$, é natural olhar nosso estimador não-viesado de β_j , $\hat{\beta}_j$, como um guia. Em qualquer aplicação interessante, a estimativa pontual $\hat{\beta}_j$ nunca será exatamente zero, seja H_0 verdadeira ou não. A questão é: Quão distante está $\hat{\beta}_j$ de zero? Um valor amostral de $\hat{\beta}_j$ muito distante de zero fornece evidência contra $H_0: \hat{\beta}_j = 0$. Entretanto, devemos reconhecer que há um erro amostral em nossa estimativa $\hat{\beta}_j$, de modo que o tamanho de $\hat{\beta}_j$ deve ser ponderado pelo seu erro amostral. Como o erro-padrão de $\hat{\beta}_j$ é uma estimativa do desvio-padrão de $\hat{\beta}_j$, $t_{\hat{\beta}_j}$ mede quantos desvios-padrão estimados $\hat{\beta}_j$ estão afastados de zero. Isso é precisamente o que fazemos ao testar se a média de uma população é zero usando a estatística t padrão da estatística introdutória. Valores de $t_{\hat{\beta}_j}$ suficientemente distantes de zero resultarão em uma rejeição de H_0 . A regra de rejeição depende da hipótese alternativa e do nível de significância escolhido do teste.

Determinar uma regra para rejeitar (4.4) a um dado nível de significância — isto é, a probabilidade de rejeitar H_0 quando ela é verdadeira — requer conhecer a distribuição amostral de $t_{\hat{\beta}_j}$ quando H_0 é verdadeira. Do Teorema 4.2, sabemos que ela é t_{n-k-1} . Esse é o resultado teórico essencial necessário para testar (4.4).

Antes de continuarmos, é importante lembrar que estamos testando hipóteses sobre parâmetros *populacionais*. Não estamos testando hipóteses sobre estimativas de uma amostra particular. Assim, nunca fará sentido formular a hipótese nula como “ $H_0: \hat{\beta}_1 = 0$ ” ou, ainda pior, como “ $H_0: 0,237 = 0$ ” quando a estimativa do parâmetro for 0,237 na amostra. Estamos testando se o valor populacional desconhecido, β_1 , é zero.

Alguns tratamentos da análise de regressão definem a estatística t como o *valor absoluto* de (4.5), de modo que a estatística t sempre é positiva. Essa prática tem a desvantagem de tornar um pouco confuso o teste contra hipóteses alternativas unilaterais. Ao longo deste livro, a estatística t sempre tem o mesmo sinal da estimativa do coeficiente de MQO correspondente.

Teste contra Hipóteses Alternativas Unilaterais

A fim de determinar uma regra para rejeitar H_0 , precisamos decidir sobre a **hipótese alternativa** relevante. Primeiro, considere uma **hipótese alternativa unilateral** do tipo

$$H_1: \beta_j > 0. \quad (4.6)$$

Isso significa que não nos preocupamos com alternativas de H_0 do tipo $H_1: \beta_j < 0$; por alguma razão, talvez tornando como base a introspecção ou a teoria econômica, estamos excluindo os valores populacionais de β_j menores que zero. (Outra maneira de pensar a respeito é que a hipótese nula é realmente $H_0: \beta_j \leq 0$; em qualquer caso, a estatística $t_{\hat{\beta}_j}$ é usada como a estatística de teste.)

Como devemos escolher uma regra de rejeição? Em primeiro lugar, devemos decidir sobre um **nível de significância** ou uma probabilidade de rejeitar H_0 quando ela é, de fato, verdadeira. Em termos mais concretos, suponha que decidimos por um nível de significância de 5%, já que esta é a escolha mais comum. Assim, estamos dispostos a rejeitar erroneamente H_0 , quando ela é verdadeira 5% das vezes. Agora, embora $t_{\hat{\beta}_j}$ tenha uma distribuição t sob H_0 — de modo que ele tem média igual a zero —, sob a hipótese alternativa $\beta_j > 0$, o valor esperado de $t_{\hat{\beta}_j}$ é positivo. Assim, estamos procurando um valor positivo “suficientemente grande” de $t_{\hat{\beta}_j}$ a fim de rejeitar $H_0: \beta_j = 0$ em favor de $H_1: \beta_j > 0$. Valores negativos de $t_{\hat{\beta}_j}$ não fornecem evidência em favor de H_1 .

A definição de “suficientemente grande”, com um nível de significância de 5%, é o 95º percentil de uma distribuição t com $n - k - 1$ graus de liberdade; denominemos esse ponto de c . Em outras palavras, a **regra de rejeição** é que H_0 é rejeitada em favor de H_1 , ao nível de significância de 5%, se

$$t_{\hat{\beta}_j} > c. \quad (4.7)$$

De acordo com nossa escolha do **valor crítico** c , a rejeição de H_0 , quando ela é verdadeira, ocorrerá em 5% de todas as amostras aleatórias.

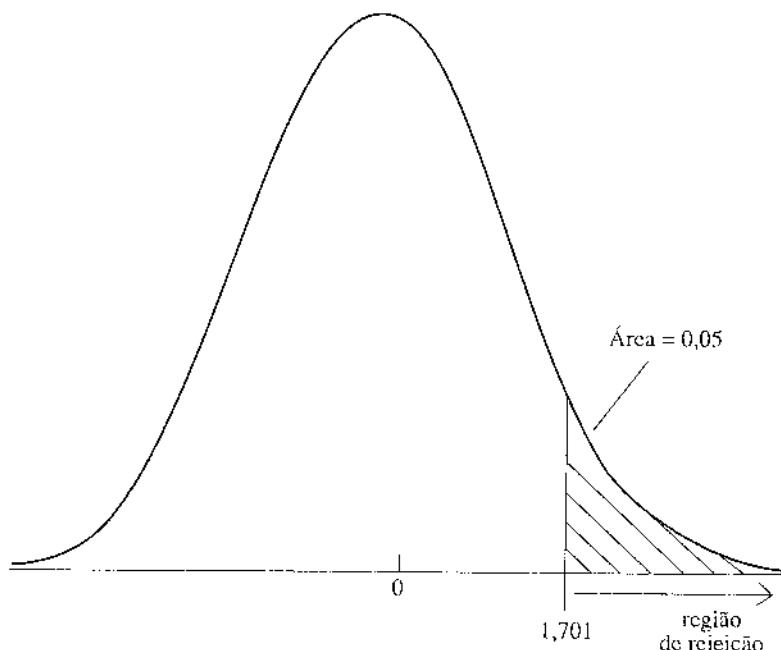
A regra de rejeição em (4.7) é um exemplo de um **teste monocaudal**. Para obter c , necessitamos somente do nível de significância e dos graus de liberdade. Por exemplo, para um teste ao nível de 5% e com $n - k - 1 = 28$ graus de liberdade, o valor crítico é $c = 1,701$. Se $t_{\hat{\beta}_j} < 1,701$, não rejeitamos H_0 em favor de (4.6) ao nível de 5%. Observe que um valor negativo de $t_{\hat{\beta}_j}$, não importando o tamanho desse valor em termos absolutos, leva a uma negativa em rejeitar H_0 em favor de (4.6). (Veja a Figura 4.2.)

O mesmo procedimento pode ser usado com outros níveis de significância. Para um teste ao nível de 10% e se $gl = 21$, o valor crítico é $c = 1,323$. Para um nível de significância de 1% e se $gl = 21$, $c = 2,518$. Todos esses valores críticos são obtidos diretamente da Tabela G.2, do Apêndice G. Você deve observar um padrão nos valores críticos: quando o nível de significância cai, o valor crítico aumenta, de modo que, para rejeitar H_0 , exigimos um valor cada vez maior de $t_{\hat{\beta}_j}$. Assim, se H_0 é rejeitada, por exemplo, ao nível de 5%, então ela também é, automaticamente, rejeitada ao nível de 10%. Não faz sentido rejeitar a hipótese nula, por exemplo, ao nível de 5% e, em seguida, refazer o teste para nos certificarmos do resultado ao nível de 10%.

Quando os graus de liberdade da distribuição t ficam maiores, a distribuição t aproxima-se da distribuição normal padronizada. Por exemplo, quando $n - k - 1 = 120$, o valor crítico de 5% para a hipótese alternativa unilateral (4.7) é 1,658, comparável ao valor normal padronizado de 1,645. Esses valores, para objetivos práticos, são suficientemente próximos; para graus de liberdade maiores que 120, pode-se usar os valores críticos da distribuição normal padronizada.

Figura 4.2

Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j > 0$, com 28 gl.

**EXEMPLO 4.1****(Equação do Salário Horário)**

Usando os dados do arquivo WAGE1.RAW, obtemos a equação estimada

$$\begin{aligned} \log(\text{salárioh}) &= 0,284 + 0,092 \text{ educ} + 0,041 \text{ exper} + 0,022 \text{ perm} \\ (0,104) &\quad (0,007) \quad (0,0017) \quad (0,003) \\ n &= 526, R^2 = 0,316, \end{aligned}$$

em que os erros-padrão aparecem em parênteses abaixo dos coeficientes estimados. Seguiremos essa convenção ao longo do livro. Essa equação pode ser usada para testar se o retorno de *exper*, controlando *educ* e *perm*, é zero na população, contra a hipótese alternativa de que ele é positivo. Para tanto, escreva $H_0: \beta_{\text{exper}} = 0$ versus $H_1: \beta_{\text{exper}} > 0$. (Nas aplicações, indexar um parâmetro pelo nome da variável ao qual está

EXEMPLO 4.1 (continuação)

associado é uma maneira hábil de caracterizar os parâmetros, visto que os índices numéricos que usamos são, em geral, arbitrários e podem causar confusão.) Lembre-se de que β_{exper} representa o parâmetro populacional desconhecido. Não faz sentido algum escrever " $H_0: 0,0041 = 0$ " ou " $H_0: \hat{\beta}_{exper} = 0$ ".

Como temos 522 graus de liberdade, podemos usar os valores críticos da distribuição normal padronizada. O valor crítico a 5% é 1,645, e o valor crítico a 1% é 2,326. A estatística t para β_{exper} é

$$t_{\hat{\beta}_{exper}} = 0,0041/0,0017 \approx 2,41,$$

e portanto $\hat{\beta}_{exper}$, ou $exper$, é estatisticamente significante mesmo ao nível de 1%. Também dizemos que " $\hat{\beta}_{exper}$ é estatisticamente maior que zero ao nível de significância de 1%".

O retorno estimado para um ano a mais de experiência, mantendo fixas a permanência e a educação formal, não é muito grande. Por exemplo, acrescentar três anos a mais aumenta $\log(salárioh)$ em $3(0,0041) = 0,0123$, de modo que o salário é somente cerca de 1,2% maior. No entanto, mostramos, de modo convincente, que o efeito parcial da experiência é positivo na população.

A hipótese alternativa unilateral cujo parâmetro é menor que zero,

$$H_1: \beta_j < 0, \quad (4.8)$$

também aparece nas aplicações. A regra de rejeição para a hipótese alternativa (4.8) é exatamente a imagem espelhada do caso anterior. Agora, o valor crítico vem da cauda esquerda da distribuição t . Na prática, é mais fácil pensar a regra de rejeição como

$$t_{\hat{\beta}_j} < -c, \quad (4.9)$$

em que c é o valor crítico da hipótese alternativa $H_1: \beta_j > 0$. Para simplificar, sempre assumimos que c é positivo, visto que é assim que os valores críticos são apresentados nas tabelas t , e, portanto, o valor crítico $-c$ é um número negativo.

Sejam as taxas de aprovação de empréstimos de uma cidade determinadas por

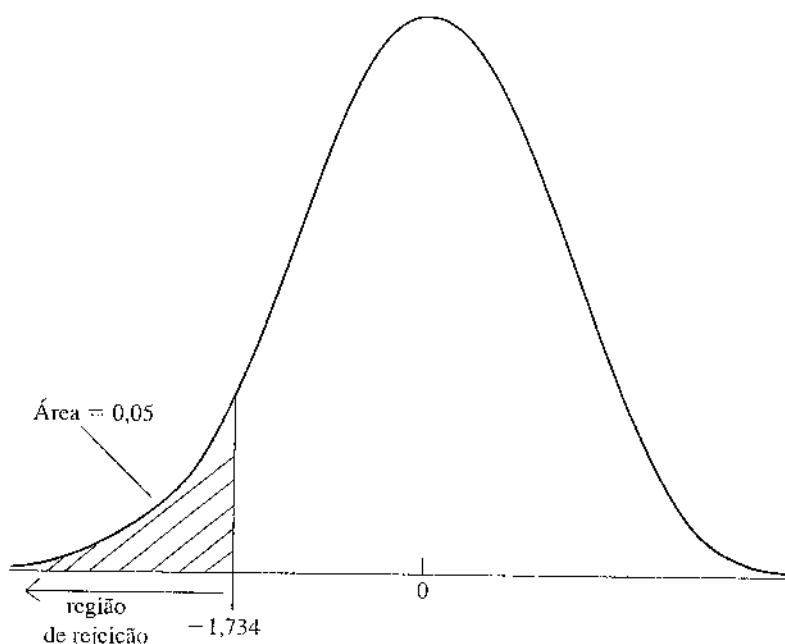
$$\begin{aligned} taxaprov = & \beta_0 + \beta_1 porcmin + \beta_2 rendmed + \\ & \beta_3 riquemed + \beta_4 dívidamed + u, \end{aligned}$$

em que $porcmin$ é a percentagem de minorias na cidade, $rendmed$ é a renda média, $riquemed$ é a riqueza média e $dívidamed$ é alguma medida de dívidas médias. Como você formularia a hipótese nula de que não há diferença nas taxas de empréstimos entre os bairros devido à composição racial e étnica quando a renda média, a riqueza média e a dívida média foram controladas? Como você formularia a hipótese alternativa de que há discriminação contra as minorias nas taxas de aprovação de empréstimos?

Por exemplo, se o nível de significância é 5% e o número de graus de liberdade é 18, então $c = 1,734$, e assim $H_0: \beta_j = 0$ é rejeitada em favor de $H_1: \beta_j < 0$ ao nível de 5% se $t_{\hat{\beta}_j} < -1,734$. É importante lembrar que, para rejeitar H_0 contra a hipótese alternativa negativa (4.8), devemos obter uma estatística t negativa. Uma razão t positiva, não importa o quão grande ela seja, não fornece evidência em favor de (4.8). A regra de rejeição está ilustrada na Figura 4.3.

Figura 4.3

Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j < 0$, com 18 gl.



EXEMPLO 4.2

(Desempenho de Estudantes e Tamanho de Escolas)

Há muito interesse no efeito do tamanho das escolas sobre o desempenho dos estudantes. (Veja, por exemplo, *The New York Times Magazine*, 28.5.95.) Afirma-se que, tudo o mais sendo igual, os estudantes de escolas menores saem-se melhor do que aqueles de escolas maiores. Assume-se que essa hipótese é verdadeira, mesmo após considerar as diferenças nos tamanhos das salas entre as escolas.

O arquivo MEAP93.RAW contém dados sobre 408 escolas de ensino médio em Michigan para o ano de 1993. Podemos usar esses dados para testar a hipótese nula de que o tamanho da escola não tem efeito sobre as notas de testes padronizados, contra a hipótese alternativa de que o tamanho tem um efeito negativo. O desempenho é medido pela percentagem de estudantes que recebem uma nota de aprovação no teste de matemática, *mate10*. O tamanho da escola é medido pelo número de estudantes matriculados (*matrici*). A hipótese nula é $H_0: \beta_{matrici} = 0$, e a hipótese alternativa é $H_1: \beta_{matrici} < 0$. Por ora, vamos controlar outros

EXEMPLO 4.2 (continuação)

dois fatores: o salário anual médio dos professores (*totsal*) e o número de funcionários por mil estudantes (*staff*). O salário do professor é uma medida da qualidade do professor, e o tamanho de *staff* é uma medida aproximada de quanta atenção os estudantes recebem.

A equação estimada, com os erros-padrão entre parênteses, é

$$\hat{ma\acute{e}10} = 2,274 + 0,00046 \text{ } totsal + 0,048 \text{ } staff - 0,00020 \text{ } matricl$$

$$(6,113) \quad (0,00010) \quad (0,040) \quad (0,00022)$$

$$n = 408, R^2 = 0,0541.$$

O coeficiente de *matricl*, $-0,00020$, está de acordo com a conjectura de que escolas maiores tolhem o desempenho: números maiores de matrículas levam a uma percentagem menor de estudantes com uma nota de aprovação. (Os coeficientes de *totsal* e *staff* também têm os sinais que esperamos.) O fato de *matricl* ter um coeficiente estimado diferente de zero pode ser devido, justamente, ao erro de amostragem; para nos convencermos de um efeito, precisamos conduzir um teste *t*.

Como $n - k - 1 = 408 - 4 = 404$, usamos o valor crítico normal padronizado. Ao nível de 5%, o valor crítico é $-1,65$; para rejeitar H_0 ao nível de 5%, a estatística *t* de *matricl* deve ser menor que $-1,65$.

A estatística de *matricl* é $-0,00020/0,00022 \approx -0,91$, que é maior que $-1,65$: não podemos rejeitar H_0 em favor de H_1 ao nível de 5%. De fato, o valor crítico ao nível de 15% é $-1,04$, e, como $-0,91 > -1,04$, não é possível rejeitar H_0 mesmo ao nível de 15%. Concluímos que *matricl* não é estatisticamente significante ao nível de 15%.

A variável *totsal* é estatisticamente significante mesmo ao nível de significância de 1% porque sua estatística *t* é 4,6. Do outro lado, a estatística *t* de *staff* é 1,2, portanto não podemos rejeitar $H_0: \beta_{staff} = 0$ contra $H_1: \beta_{staff} > 0$, mesmo ao nível de significância de 10%. (O valor crítico, calculado a partir da distribuição normal padronizada, é $c = 1,28$.)

Para ilustrar como a mudança da forma funcional pode afetar nossas conclusões, vamos estimar também o modelo com todas as variáveis independentes na forma logarítmica. Isso permite, por exemplo, que o efeito do tamanho da escola diminua quando o tamanho da escola aumenta. A equação estimada é

$$\hat{ma\acute{e}10} = -207,66 + 21,16 \log(totsal) + 3,98 \log(staff) - 1,29 \log(matricl)$$

$$(48,70) \quad (4,06) \quad (4,19) \quad (0,69)$$

$$n = 408, R^2 = 0,0654$$

A estatística *t* de $\log(matricl)$ é cerca de $-1,87$; como esse valor está abaixo do valor crítico ao nível de 5%, $-1,65$, rejeitamos $H_0: \beta_{\log(matricl)} = 0$ em favor de $H_1: \beta_{\log(matricl)} < 0$ ao nível de 5%.

No Capítulo 2, encontramos um modelo em que a variável dependente aparecia em sua forma original (chamada forma em *nível*), enquanto a variável independente aparecia na forma log (chamado modelo *nível-log*). A interpretação dos parâmetros é a mesma no contexto da regressão múltipla, exceto, evidentemente, que podemos dar uma interpretação *ceteris paribus* aos parâmetros. Mantendo fixos *totsal* e *staff*, temos $\Delta\hat{ma\acute{e}10} = -1,29[\Delta\log(matricl)]$, de modo que

$$\Delta\hat{ma\acute{e}10} \approx - (1,29/100)(\% \Delta matricl) \approx -0,013(\% \Delta matricl).$$

EXEMPLO 4.2 (continuação)

Uma vez mais, usamos o fato de que a variação em $\log(\text{matrcl})$, quando multiplicada por 100, é aproximadamente a variação percentual em matrcl . Assim, se o número de matrículas é 10% maior em uma escola, prevê-se que $\text{mate10} \approx 0,013(10) = 0,13$ ponto percentual menor (mate10 é mensurado como uma percentagem).

Que modelo preferimos: aquele que usa o nível de matrcl ou aquele que usa $\log(\text{matrcl})$? No modelo nível-nível, o número de matrículas não tem um efeito estatisticamente significante, mas no modelo nível-log ele tem. Isso se traduz em um R^2 -quadrado maior para o modelo nível-log, o que significa dizer que explicamos mais da variação em mate10 ao usar matrcl na forma logarítmica (6,5% contra 5,4%). O modelo nível-log é preferível, pois ele captura, de modo mais próximo, a relação entre mate10 e matrcl . No Capítulo 6 falaremos mais sobre como usar o R^2 -quadrado para escolher a forma funcional.

Teste contra Hipóteses Alternativas Bilaterais

Nas aplicações, é comum testar a hipótese nula $H_0: \beta_j = 0$ contra uma hipótese alternativa bilateral, ou seja,

$$H_1: \beta_j \neq 0. \quad (4.10)$$

Sob essa hipótese alternativa, x_j tem um efeito *ceteris paribus* sobre y sem especificar se o efeito é positivo ou negativo. Ela é a hipótese alternativa relevante quando o sinal de β_j não é bem-determinado pela teoria (ou pelo senso comum). Mesmo quando sabemos se β_j é positivo ou negativo sob a hipótese alternativa, um teste bilateral é muitas vezes prudente. No mínimo, usar uma alternativa bilateral nos impede de olhar a equação estimada e, então, buscar a hipótese alternativa em se $\hat{\beta}_j$ é positivo ou negativo. Usar as estimativas da regressão para nos ajudar a formular as hipóteses nula e alternativa não é permitido, porque a inferência estatística clássica pressupõe que formulamos as hipóteses nula e alternativa sobre a população antes de olhar os dados. Por exemplo, não devemos estimar em primeiro lugar a equação que relaciona o desempenho em matemática ao número de matrículas, observar que o efeito estimado é negativo e, em seguida, decidir que a hipótese alternativa relevante é $H_1: \beta_{\text{matrcl}} < 0$.

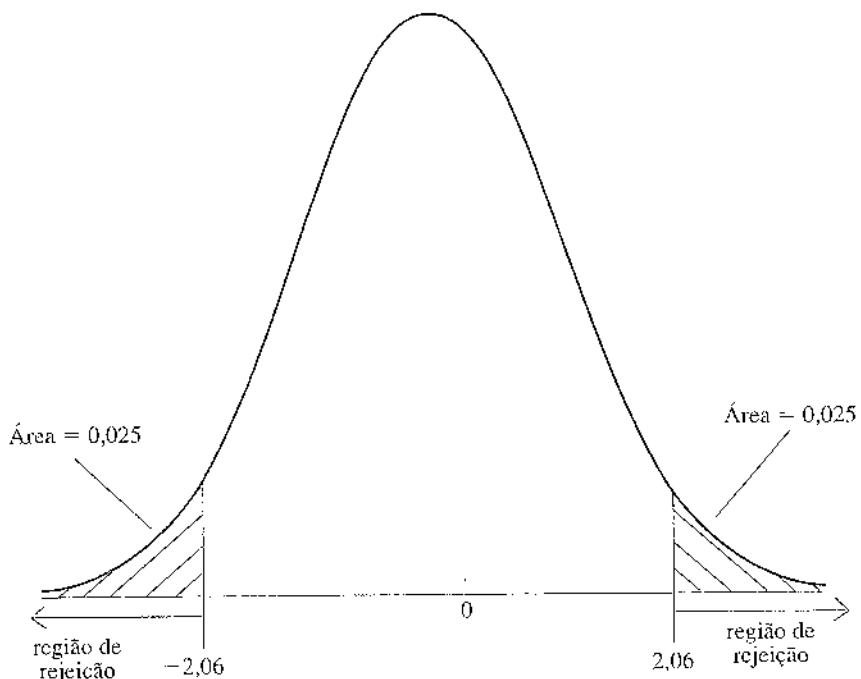
Quando a alternativa é bilateral, estamos interessados no valor absoluto da estatística t . A regra de rejeição para $H_0: \beta_j = 0$ contra (4.10) é

$$|t_{\hat{\beta}_j}| > c, \quad (4.11)$$

em que $|\cdot|$ representa o valor absoluto e c é um valor crítico apropriadamente escolhido. Para achar c , vamos especificar novamente um nível de significância, por exemplo, de 5%. Para um teste bi-caudal, c é escolhido de tal forma a fazer com que a área em cada cauda da distribuição t seja igual a 2,5%. Em outras palavras, c é o 97,5º percentil da distribuição t com $n - k - 1$ graus de liberdade. Quando $n - k - 1 = 25$, o valor crítico de 5% para um teste bilateral é $c = 2,060$. A Figura 4.4 ilustra essa distribuição.

Figura 4.4

Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j \neq 0$ com 25 gl.



Quando uma hipótese alternativa específica não é formulada, considera-se geralmente que ela é bilateral. No restante deste livro, o padrão será uma hipótese alternativa bilateral, e 5% será o nível de significância padrão. Ao conduzir uma análise econométrica empírica, é sempre uma boa idéia explicitar qual é a hipótese alternativa e o nível de significância. Se H_0 é rejeitada em favor de (4.10) ao nível de 5%, em geral dizemos que “ x_j é estatisticamente significante, ou estatisticamente diferente de zero, ao nível de 5%”. Se H_0 não é rejeitada, dizemos que “ x_j é estatisticamente não significante ao nível de 5%”.

EXEMPLO 4.3

{Determinantes de nmgrad}

Usamos os dados do arquivo GPA1.RAW para estimar um modelo que explique a nota média em curso superior (*nmgrad*), utilizando o número de faltas às aulas por semana (*faltas*) como uma variável explicativa adicional. O modelo estimado é

$$\begin{aligned} \hat{nmgrad} &= 1,39 + 0,412 nmem + 0,015 tac + 0,083 faltas \\ (0,33) &\quad (0,094) \quad (0,011) \quad (0,026) \\ n &= 141, R^2 = 0,234. \end{aligned}$$

EXEMPLO 4.3 (continuação)

Podemos facilmente calcular as estatísticas t para verificar quais variáveis são estatisticamente significantes ao usar uma hipótese alternativa bilateral em cada caso. O valor crítico de 5% é cerca de 1,96, visto que os graus de liberdade ($141 - 4 = 137$) são suficientemente grandes para usar a aproximação normal padronizada. O valor crítico de 1% é cerca de 2,58.

A estatística t de $nmem$ é 4,38, significante a níveis de significância muito pequenos. Assim, dizemos que " $nmem$ é estatisticamente significante a qualquer nível de significância convencional". A estatística t de tac é 1,36, que não é estatisticamente significante ao nível de 10% contra uma alternativa bilateral. O coeficiente de tac também é, na prática, pequeno: um aumento de 10 pontos em tac , que é grande, faz com que o valor previsto de $nmgrad$ cresça somente 0,15 ponto. Assim, a variável tac é, na prática, bem como estatisticamente, não significante.

O coeficiente de $faltas$ tem uma estatística t de $-0,083/0,026 = -3,19$, de modo que $faltas$ é estatisticamente significante ao nível de significância de 1% ($3,19 > 2,58$). Isso significa que uma falta a mais por semana diminui o $nmgrad$ previsto em cerca de 0,083. Assim, mantendo $nmem$ e tac fixos, a diferença prevista em $nmgrad$ entre um estudante que não falta a nenhuma aula por semana e um estudante que falta a cinco aulas por semana é de 0,42. Lembre-se de que isso não diz nada sobre estudantes específicos; referindo-se apenas aos estudantes médios dentro da população.

No exemplo, para cada variável do modelo, poderíamos argumentar que uma hipótese alternativa unilateral é apropriada. As variáveis $nmem$ e $faltas$ são muito significantes ao se usar um teste bicaudal e têm os sinais que esperamos, de modo que não há razão para fazer um teste monocaudal. Do outro lado, contra uma hipótese alternativa unilateral ($\beta_3 > 0$), tac é significante ao nível de 10% mas não ao nível de 5%. Isso não muda o fato de o coeficiente de tac ser muito pequeno.

Testes de outras Hipóteses sobre β_j

Embora $H_0: \beta_j = 0$ seja a hipótese mais comum, algumas vezes queremos testar se β_j é igual a alguma outra constante dada. Dois exemplos comuns são $\beta_j = 1$ e $\beta_j = -1$. Em geral, se a hipótese nula é expressa como

$$H_0: \beta_j = a_j, \quad \{4.12\}$$

em que a_j é o nosso valor hipotético de β_j , então a estatística t apropriada é

$$t = (\hat{\beta}_j - a_j)/\text{ep}(\hat{\beta}_j).$$

Assim como antes, t mede quantos desvios-padrão estimados $\hat{\beta}_j$ está distante do valor hipotético de β_j . A estatística t geral é usualmente escrita como

$$t = \frac{(\text{estimativa} - \text{valor hipotético})}{\text{erro-padrão}}. \quad \{4.13\}$$

Sob (4.12), essa estatística t é distribuída como t_{n-k-1} , de acordo com o Teorema 4.2. A estatística t usual é obtida quando $a_j = 0$.

Podemos usar a estatística t geral para fazer o teste contra hipóteses alternativas unilaterais ou bilaterais. Por exemplo, se as hipóteses nula e alternativa são $H_0: \beta_j = 1$ e $H_1: \beta_j > 1$, então encontramos o valor crítico para uma alternativa unilateral *exatamente* como antes: a diferença está em como calculamos a estatística t , não em como obtemos o c apropriado. Rejeitamos H_0 em favor de H_1 se $t > c$. Nesse caso, diríamos que “ $\hat{\beta}_j$ é estatisticamente maior que um” ao nível de significância apropriado.

EXEMPLO 4.4

(Crimes no Campus e Matrículas)

Considere um modelo simples que relaciona o número anual de crimes no *campus* de uma universidade (*crime*) ao número de estudantes matriculados na universidade (*matricl*):

$$\log(crime) = \beta_0 + \beta_1 \log(matricl) + u.$$

Esse é um modelo de elasticidade constante, em que β_1 é a elasticidade do crime em relação às matrículas. Não é muito útil testar $H_0: \beta_1 = 0$, se esperamos que o número total de crimes aumente quando o tamanho do *campus* aumenta. Uma hipótese mais interessante seria supor que a elasticidade do crime em relação a matrículas é igual a um: $H_0: \beta_1 = 1$. Isso significa que um aumento de 1% nas matrículas leva, em média, a um aumento de 1% nos crimes. Uma hipótese alternativa digna de nota é $H_1: \beta_1 > 1$, implicando que um aumento de 1% nas matrículas aumenta o crime no campus em *mais* de 1%. Se $\beta_1 > 1$, então, em um sentido relativo — não exatamente um sentido absoluto —, o crime é mais um problema de *campi* maiores. Uma maneira de ver isso é considerar o exponencial da equação:

$$crime = \exp(\beta_0) matricl^{\beta_1} \exp(u).$$

(Veja o Apêndice A, disponível no site da Thomson, para as propriedades do logaritmo natural e das funções exponenciais.) Para $\beta_0 = 0$ e $u = 0$, essa equação está representada na Figura 4.5, com $\beta_1 < 1$, $\beta_1 = 1$ e $\beta_1 > 1$.

Vamos testar $\beta_1 = 1$ contra $\beta_1 > 1$, usando os dados de 97 faculdades e universidades dos Estados Unidos no ano de 1992, os quais estão contidos no arquivo CAMPUS.RAW. Os dados são provenientes da publicação *FBI's Uniform Crime Reports*, e o número médio de crimes no *campus* é cerca de 394 na amostra, enquanto o número médio de matrículas é aproximadamente 16.076. A equação estimada (com as estimativas e os erros-padrão arredondados em duas casas decimais) é

$$\begin{aligned} \log(crime) &= -6,63 + 1,27 \log(matricl) \\ &\quad (1,03) \quad (0,11) \\ n &= 97, R^2 = 0,585. \end{aligned} \tag{4.14}$$

A elasticidade estimada de crime em relação a *matricl*, 1,27, está na direção da hipótese alternativa $\beta_1 > 1$. Porém, há evidência suficiente para concluir que $\beta_1 > 1$? Precisamos tomar cuidado ao testar essa hipótese, especialmente porque os resultados estatísticos dos programas padrão de regressão são muito mais complexos do que o resultado simplificado informado pela equação (4.14). Nossa primeira instinto deveria ser construir “a” estatística t , tornando o coeficiente de $\log(matricl)$ e dividindo-o pelo seu erro-padrão, que é a estatística t descrita por um programa de regressão. No entanto, essa é a estatística errada

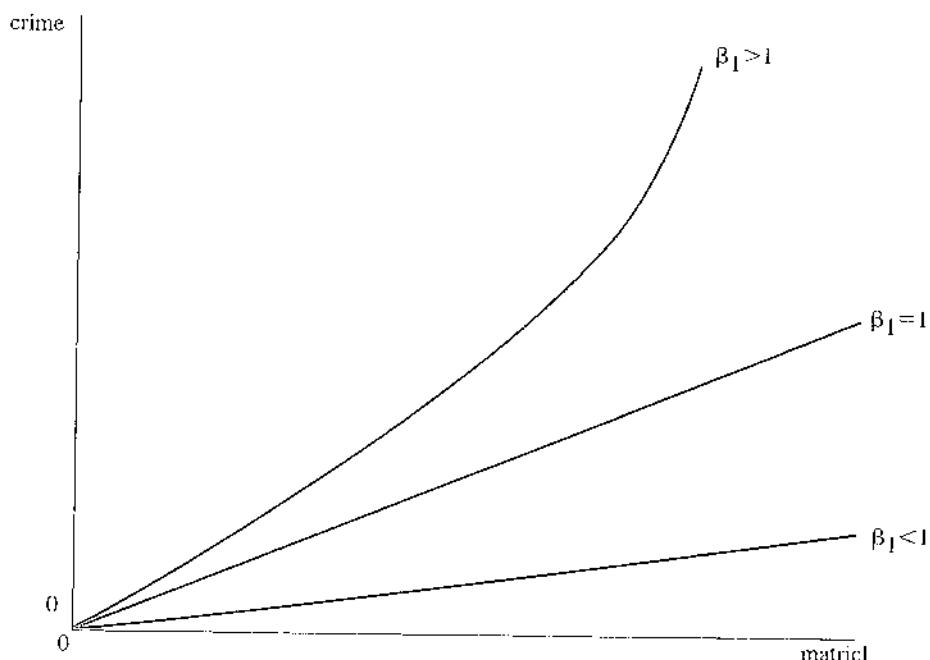
EXEMPLO 4.4 (continuação)

para testar $H_0: \beta_1 = 1$. A estatística t correta é obtida de (4.13): subtraímos o valor hipotético, um, da estimativa e dividimos o resultado pelo seu erro-padrão de $\hat{\beta}_1$: $t = (1,27 - 1)/0,11 = 0,27/0,11 \approx 2,45$. O valor crítico unilateral de 5% para uma distribuição t com $97 - 2 = 95$ g.l é cerca de 1,66 (usando $g.l = 120$), de modo que, claramente, rejeitamos $\beta_1 = 1$ em favor de $\beta_1 > 1$ ao nível de 5%. De fato, o valor crítico de 1% é cerca de 2,37, e portanto rejeitamos a hipótese nula em favor da hipótese alternativa, ao nível de 1%.

Devemos ter em mente que essa análise não mantém os outros fatores constantes e, portanto, a elasticidade de 1,27 não é necessariamente uma boa estimativa do efeito *ceteris paribus*. É possível que um número maior de matrículas esteja correlacionado com outros fatores que tornam o crime maior: escolas maiores podem estar localizadas em áreas de incidência maior de crimes. Poderíamos controlar isso ao coletar dados sobre taxas de crimes em cada cidade.

Figura 4.5

Gráfico de $crime = matricl^{\beta_1}$ para $\beta_1 < 1$, $\beta_1 = 1$ e $\beta_1 > 1$.



Para um teste alternativo bilateral, por exemplo, $H_0: \beta_j = -1$, $H_1: \beta_j \neq -1$, ainda calculamos a estatística t como em (4.13): $t = (\hat{\beta}_j + 1)/\text{ep}(\hat{\beta}_j)$ (observe que subtrair -1 significa adicionar 1). A regra de rejeição para o teste bicaudal é a usual: rejeitar H_0 se $|t| > c$, em que c é o valor crítico bicaudal. Se H_0 é rejeitada, dizemos que “ $\hat{\beta}_j$ é estatisticamente diferente do valor negativo um” ao nível de significância apropriado.

EXEMPLO 4.5**(Preços de Casas e Poluição do Ar)**

Para uma amostra de 506 comunidades na área de Boston, estimamos um modelo que relaciona o preço mediano das casas (*preço*) nas comunidades a várias características das comunidades: *oxn* é a quantidade de óxido nitroso no ar, em partes por milhão; *dist* é uma distância ponderada da comunidade em relação a cinco centros de emprego, em milhas; *comods* é o número médio de cômodos nas casas da comunidade; e *razestud* é a razão média estudante-professor nas escolas da comunidade. O modelo populacional é

$$\log(\text{preço}) = \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 \log(\text{dist}) + \beta_3 \text{comods} + \beta_4 \text{razestud} + u.$$

Assim, β_1 é a elasticidade do preço em relação a *oxn*. Queremos testar $H_0: \beta_1 = -1$ contra a hipótese alternativa $H_1: \beta_1 \neq -1$. A estatística t para fazer esse teste é $t = (\hat{\beta}_1 + 1)/\text{ep}(\hat{\beta}_1)$.

Usando os dados do arquivo HPRICE2.RAW, o modelo estimado é

$$\begin{aligned} \log(\text{preço}) &= 11,08 - 0,954 \log(\text{oxn}) - 0,134 \log(\text{dist}) + 0,255 \text{comods} - 0,052 \text{razestud} \\ &\quad (0,32) \quad (0,117) \quad (0,043) \quad (0,019) \quad (0,006) \\ n &= 506, R^2 = 0,581. \end{aligned}$$

Todas as estimativas de inclinação têm os sinais esperados. Cada coeficiente é estatisticamente diferente de zero a níveis de significância muito pequenos, incluindo o coeficiente de $\log(\text{oxn})$. No entanto, não queremos testar $\beta_1 = 0$. A hipótese nula de interesse é $H_0: \beta_1 = -1$, com a estatística t correspondente $(-0,954 + 1)/0,117 = 0,393$. Quando a estatística t é pequena como essa, há pouca necessidade de olhar a tabela t de um valor crítico: a elasticidade estimada não é estatisticamente diferente de -1 , mesmo a níveis de significância bastante altos. Controlando fatores que incluímos, há pouca evidência de que a elasticidade seja diferente de -1 .

Cálculos dos *p*-Valores dos Testes *t*

Até agora, falamos sobre como testar hipóteses ao usar uma abordagem clássica: após formular a hipótese alternativa, escolhemos um nível de significância, que então determina um valor crítico. Uma vez que o valor crítico tenha sido identificado, o valor da estatística t é comparado com o valor crítico, e a hipótese nula é rejeitada ou não, ao nível de significância dado.

Mesmo após decidir sobre a alternativa apropriada, há um componente de arbitrariedade na abordagem clássica, resultante da necessidade de escolher um nível de significância com antecedência. Diferentes pesquisadores preferem níveis de significância diferentes, dependendo da aplicação particular. Não há nível de significância “correto”.

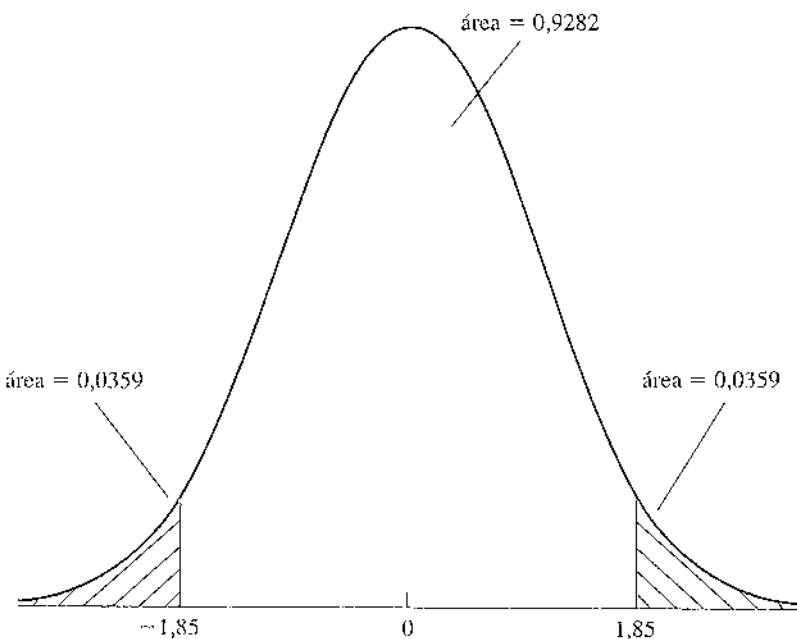
Comprometer-se com um nível de significância antecipadamente pode esconder informações úteis sobre o resultado de um teste de hipóteses. Por exemplo, suponha que desejamos testar a hipótese nula de que um parâmetro seja zero contra uma hipótese alternativa bilateral, e que com 40 graus de liberdade, obtivemos uma estatística t igual a 1,85. A hipótese nula não é rejeitada ao nível de 5%, visto que a estatística t é menor que o valor crítico bicaudal de $c = 2,021$. Um pesquisador cujo propósito é não rejeitar a hipótese nula poderia simplesmente descrever esse resultado juntamente com a estimativa: a hipótese nula não é rejeitada ao nível de 5%. Evidentemente, se a estatística t , o coeficiente e seu erro-

padrão fossem informados, então poderíamos também determinar que a hipótese nula seria rejeitada ao nível de 10%, já que o valor crítico de 10% é $c = 1,684$.

Em vez de fazer o teste a diferentes níveis de significância, é mais informativo responder à seguinte questão: dado o valor observado da estatística t , qual é o menor nível de significância ao qual a hipótese nula seria rejeitada? Esse nível é conhecido como o *p-valor* do teste (veja o Apêndice C, disponível no site da Thomson). No exemplo anterior, sabemos que o *p-valor* é maior que 0,05, visto que a hipótese nula não é rejeitada ao nível de 5%, e sabemos que o *p-valor* é menor que 0,10, já que a hipótese nula é rejeitada ao nível de 10%. Obtemos o *p-valor* real ao calcular a probabilidade de que uma variável aleatória t , com $40 gl$, seja maior que 1,85 em valor absoluto. Isto é, o *p-valor* é o nível de significância do teste quando usamos o valor da estatística de teste (1,85 no exemplo anterior) como o valor crítico do teste. Esse *p-valor* é mostrado na Figura 4.6.

Figura 4.6

Obtendo o *p-valor* contra uma alternativa bilateral, quando $t = 1,85$ e $gl = 40$.



Como um *p-valor* é uma probabilidade, seu valor está sempre entre zero e um. A fim de calcular os *p-valores*, precisamos de tabelas impressas extremamente detalhadas da distribuição t — o que não é muito prático — ou um programa de computador que calcule as áreas sob a função densidade de probabilidade da distribuição t . A maioria dos programas de regressão modernos tem essa capacidade. Alguns deles calculam os *p-valores* rotineiramente a cada regressão de MQO, mas somente para certas hipóteses. Se um programa de regressão informa um *p-valor* juntamente com o resultado padrão de MQO, esse valor é, quase certamente, o *p-valor* de testar a hipótese nula $H_0: \beta_j = 0$ contra a hipótese alternativa bilateral. O *p-valor*, nesse caso, é

$$P(|T| > |t|), \quad (4.15)$$

em que, por clareza, T representa uma variável aleatória com distribuição t , com $n - k - 1$ graus de liberdade, e t é o valor numérico da estatística de teste.

O p -valor resume, com precisão, a força e a fraqueza da evidência empírica contra a hipótese nula. Talvez a interpretação mais útil seja a seguinte: o p -valor é a probabilidade de observar uma estatística t tão extrema quanto aceitariamos se a hipótese nula fosse verdadeira. Isso significa que p -valores pequenos são evidências contra a hipótese nula; p -valores grandes fornecem pouca evidência contra H_0 . Por exemplo, se p -valor = 0,50 (informado sempre como um decimal, e não uma percentagem), observariamos um valor da estatística t tão extremo quanto o faríamos em 50% de todas as amostras aleatórias quando a hipótese nula fosse verdadeira; essa é uma evidência bastante fraca contra H_0 .

No exemplo com $gl = 40$ e $t = 1,85$, o p -valor é calculado como

$$p\text{-valor} = P(|T| > 1,85) = 2P(T > 1,85) = 2(0,0359) = 0,0718,$$

em que $P(T > 1,85)$ é a área à direita de 1,85 da distribuição t com 40 gl . (Esse valor foi calculado usando o programa econométrico Stata; ele não está disponível na Tabela G.2.) Isso significa que, se a hipótese nula for verdadeira, observariamos um valor absoluto da estatística t tão grande quanto 1,85 em cerca de 7,2% das vezes. Isso fornece alguma evidência contra a hipótese nula, mas não a rejeitariammos ao nível de significância de 5%.

O exemplo anterior ilustra que, uma vez que o p -valor foi calculado, um teste clássico pode ser realizado a qualquer nível desejado. Se α é o nível de significância do teste (na forma decimal), então H_0 é rejeitada se p -valor < α ; de outro modo, H_0 não é rejeitada ao nível de $100 \cdot \alpha\%$.

Calcular os p -valores de alternativas unilaterais também é muito simples. Suponha, por exemplo, que vamos testar $H_0: \beta_j = 0$ contra $H_1: \beta_j > 0$. Se $\hat{\beta}_j < 0$, então calcular um p -valor não é importante: sabemos que o p -valor é maior que 0,50, o que nunca nos fará rejeitar H_0 em favor de H_1 . Se $\hat{\beta}_j > 0$, então $t > 0$ e o p -valor é exatamente a probabilidade de uma variável aleatória, com os gl apropriados, exceder o valor t . Alguns programas de regressão somente calculam os p -valores para alternativas bilaterais. No entanto, é simples obter o p -valor unilateral: apenas divide o p -valor bilateral por 2.

Se a hipótese alternativa for $H_1: \beta_j < 0$, faz sentido calcular um p -valor se $\hat{\beta}_j < 0$ (e portanto $t < 0$): p -valor = $P(T < t) = P(T > |t|)$, porque a distribuição t é simétrica em torno de zero. Uma vez mais, isso pode ser obtido como a metade do p -valor de um teste bicaudal.

Suponha que você tenha estimou um modelo de regressão e obteve $\hat{\beta}_1 = 0,56$ e p -valor = 0,086 para testar $H_0: \beta_1 = 0$ contra $H_1: \beta_1 \neq 0$. Qual é o p -valor para testar $H_0: \beta_1 = 0$ contra $H_1: \beta_1 > 0$?

Como as magnitudes das estatísticas t que levam à significância estatística se tornarão rapidamente familiares, especialmente para tamanhos de amostras grandes, não é sempre crucial descrever os p -valores das estatísticas t . No entanto, não é incorreto informá-las. Além disso, quando discutirmos o teste F na Seção 4.5, veremos que é importante calcular os p -valores, porque os valores críticos dos testes F não são facilmente memorizados.

Lembrete sobre a Linguagem do Teste de Hipóteses Clássico

Quando H_0 não é rejeitada, preferimos usar a linguagem “não é possível rejeitar H_0 ao nível de $x\%$ ” em vez de “ H_0 é aceita ao nível de $x\%$ ”. Podemos usar o Exemplo 4.5 para ilustrar o porquê de a primeira afirmação ser preferida. Naquele exemplo, a elasticidade estimada do preço em relação a oxn é $-0,954$, e a estatística t para testar $H_0: \beta_{oxn} = -1$ é $t = 0,393$; portanto, não podemos rejeitar H_0 . Porém, há muitos outros valores de β_{oxn} (mais do que podemos contar) que não podem ser rejeitados. Por exemplo, a estatística t para $H_0: \beta_{oxn} = -0,9$ é $(-0,954 + 0,9)/0,117 = -0,462$, e portanto essa hipótese nula também não é rejeitada. Claramente, $\beta_{oxn} = -1$ e $\beta_{oxn} = -0,9$ não podem ser ambas verdadeiras, de modo que não faz sentido dizer que “aceitamos” uma dessas hipóteses. Tudo o que podemos dizer é que os dados não nos permitem rejeitar uma dessas hipóteses ao nível de significância de 5%.

Significância Econômica ou Prática versus Significância Estatística

Após termos enfatizado a *significância estatística* ao longo desta seção, agora é um bom momento para lembrar que devemos prestar atenção na magnitude das estimativas dos *coeficientes*, além do tamanho das estatísticas t . A significância estatística de uma variável x_j é determinada completamente pelo tamanho de $t_{\hat{\beta}_j}$, enquanto a **significância econômica** ou a **significância prática** da variável está relacionada ao tamanho (e sinal) de $\hat{\beta}_j$.

Lembre-se de que a estatística t para testar $H_0: \beta_j = 0$ é definida ao dividirmos a estimativa por seu erro-padrão: $t_{\hat{\beta}_j} = \hat{\beta}_j / \text{ep}(\hat{\beta}_j)$. Assim, $t_{\hat{\beta}_j}$ pode indicar significância estatística porque $\hat{\beta}_j$ é “grande” ou porque $\text{ep}(\hat{\beta}_j)$ é “pequeno”. É importante, na prática, distinguir entre essas duas razões das estatísticas t estatisticamente significantes. Colocar muita ênfase sobre a significância estatística pode levar à conclusão falsa de que uma variável é “importante” para explicar y embora seu efeito estimado seja moderado.

EXEMPLO 4.6

[Taxas de Participação nos Planos de Pensão]

No Exemplo 3.3, usamos os dados dos planos de pensão para estimar um modelo que descreve as taxas de participação em termos de taxas de complementação das empresas e das idades dos planos. Vamos incluir, agora, uma medida de tamanho das empresas, o número total de empregados das empresas, (*totemp*). A equação estimada é

$$\begin{aligned} \text{taxap} &= 80,29 + 5,44 \text{ taxicont} + 0,269 \text{ idade} + 0,00013 \text{ totemp} \\ (0,78) &\quad (0,52) \quad (0,045) \quad (0,00004) \\ n &= 1,534, R^2 = 0,100. \end{aligned}$$

A menor estatística t , em valor absoluto, é a da variável *totemp*: $t = -0,00013/0,00004 = -3,25$, e ela é estatisticamente significante a níveis de significância muito pequenos. (O p -valor bicaudal dessa estatística t é cerca de 0,001.) Assim, todas as variáveis são estatisticamente significantes a níveis de significância bem pequenos.

Qual o tamanho, em um sentido prático, do coeficiente de *totemp*? Mantendo fixos *taxcomp* e *idade*, se uma firma cresce em 10.000 empregados, a taxa de participação cai em $10.000(0,00013) = 1,3$ pontos percentuais. Isso é um crescimento enorme no número de empregados, com um efeito somente modesto na taxa de participação. Assim, embora o tamanho da firma afete, de fato, a taxa de participação, o efeito não é, na prática, muito grande.

O exemplo anterior mostra que é particularmente importante interpretar a magnitude do coeficiente, além de olhar as estatísticas t , ao trabalhar com amostras grandes. Com tamanhos de amostras grandes, os parâmetros podem ser estimados com muita precisão: os erros-padrão são, em geral, muito pequenos em relação às estimativas dos coeficientes, o que frequentemente resulta em significância estatística.

Alguns pesquisadores insistem em usar níveis de significância pequenos quando o tamanho da amostra cresce, em parte como uma maneira de compensar o fato de que os erros-padrão estão ficando menores. Por exemplo, se nos sentimos confortáveis com um nível de 5% quando n corresponde a algumas centenas, de observações, deveríamos usar o nível de 1% quando n corresponde a alguns milhares. Usar um nível de significância menor significa que as significâncias econômica e estatística são mais prováveis de coincidir, mas não há garantias: no exemplo anterior, mesmo se usarmos um nível de significância tão pequeno quanto 0,1% (um décimo de um por cento), ainda concluiríamos que $totemp$ é estatisticamente significante.

A maior parte dos pesquisadores também está disposta a considerar níveis de significância maiores em aplicações com tamanhos de amostra pequenos, refletindo o fato de que é difícil achar significância com tamanhos de amostra menores (os valores críticos são maiores em magnitude, e os estimadores são menos precisos). Infelizmente, se esse é ou não o caso pode depender dos planos subjacentes do pesquisador.

EXEMPLO 4.7

(Efeitos das Subvenções a Treinamento de Trabalho sobre as Taxas de Rejeição de Produtos das Empresas)

A taxa de rejeição de produtos de uma firma manufatureira é o número de itens defeituosos que devem ser descartados de cada 100 itens produzidos. Assim, uma diminuição nessa taxa de rejeição reflete maior produtividade.

Podemos usar a taxa de rejeição para mensurar o efeito do treinamento dos trabalhadores sobre a produtividade. Para uma amostra de firmas manufatureiras de Michigan em 1987, estimou-se a seguinte equação:

$$\begin{aligned} \log(rejei) &= 13,72 - 0,028 \text{ hrsemp} - 1,21 \log(vendas) + 1,48 \log(empreg) \\ (4,91) &\quad (0,019) \quad (0,41) \quad (0,43) \\ n = 30, R^2 &= 0,431. \end{aligned}$$

(Essa regressão usa um subconjunto de dados de JTRAIN.RAW.) A variável $hrsemp$ corresponde às horas anuais de treinamento por trabalhador, $vendas$ corresponde às vendas anuais da firma (em dólares), e $empreg$ é o número de empregados da firma. A taxa média de rejeição na amostra é cerca de 3,5, e $hrsemp$ médio é cerca de 7,3.

A principal variável de interesse é $hrsemp$. Uma hora a mais de treinamento por trabalhador diminui $\log(rejei)$ em 0,028, o que significa que a taxa de rejeição é cerca de 2,8% menor. Assim, se $hrsemp$ aumenta em 5 — cada empregado é treinado 5 horas a mais por ano —, estima-se que a taxa de rejeição caia em $5(2,8\%) = 14\%$. Isso parece ser um efeito razoavelmente grande mas, saber se o treinamento adicional vale a pena para a firma, depende do custo de treinamento e dos benefícios de uma taxa de rejeição menor. Não temos os números necessários para fazer uma análise custo/benefício, mas o efeito estimado não parece trivial.

E o que dizer sobre a *significância estatística* da variável de treinamento? A estatística t de $hrsemp$ é $-0,028/0,019 = -1,47$, e agora você provavelmente reconhece como não sendo suficientemente grande em magnitude para concluir que $hrsemp$ é estatisticamente significante ao nível de 5%. De fato, com $30 - 4 = 26$ graus de liberdade para a alternativa unilateral, $H_1: \beta_{hrsemp} < 0$, o valor crítico de 5% é cerca de $-1,71$.

EXEMPLO 4.7 (continuação)

Assim, usando um teste de nível estrito a 5%, devemos concluir que $hrsemp$ não é estatisticamente significante, mesmo usando uma alternativa unilateral.

Como o tamanho da amostra é bastante pequeno, poderíamos ser mais liberais com o nível de significância. O valor crítico de 10% é $-1,32$, e portanto $hrsemp$ é significante contra a alternativa unilateral ao nível de 10%. O p -valor é facilmente calculado como $P(T_{26} < -1,47) = 0,077$. Esse pode ser um p -valor suficientemente pequeno para concluir que o efeito estimado do treinamento não se deve apenas ao erro de amostragem, mas alguns economistas teriam opiniões diferentes a respeito do assunto.

Lembre-se de que erros-padrão grandes podem também ser um resultado da multicolinearidade (alta correlação entre algumas das variáveis independentes), mesmo que o tamanho da amostra pareça razoavelmente grande. Como discutimos na Seção 3.4, não há muito que possamos fazer sobre esse problema além de coletar mais dados ou mudar o escopo da análise excluindo certas variáveis independentes do modelo. Como no caso de um tamanho de amostra pequeno, pode ser difícil estimar precisamente os efeitos parciais quando algumas das variáveis explicativas são altamente correlacionadas. (A Seção 4.5 contém um exemplo.)

Finalizamos esta seção com algumas instruções para discutir as significâncias econômica e estatística de uma variável em um modelo de regressão múltipla:

1. Cheque a significância estatística. Se a variável é estatisticamente significante, discuta a magnitude do coeficiente para ter uma idéia de sua importância prática ou econômica. Esse último passo pode requerer algum cuidado, dependendo de como as variáveis independentes e dependentes aparecem na equação. (Em particular, quais são as unidades de medida? As variáveis aparecem na forma logarítmica?)
2. Se uma variável não é estatisticamente significante aos níveis usuais (10%, 5% ou 1%), você poderia ainda perguntar se a variável tem o efeito esperado sobre y e se tal efeito é, na prática, grande. Se ele é grande, você deve calcular um p -valor para a estatística t . Para tamanhos de amostras pequenos, você pode, às vezes, construir um argumento para p -valores tão grandes quanto 0,20 (mas não há regras rigorosas). Com p -valores grandes, isto é, estatísticas t pequenas, estamos pisando em gelo fino, porque as estimativas grandes, na prática, podem ser devidas ao erro de amostragem: uma amostra aleatória diferente poderia resultar em uma estimativa muito diferente.
3. É comum encontrar variáveis com estatísticas t pequenas que têm o sinal “errado”. Para propósitos práticos, elas podem ser ignoradas: concluímos que as variáveis são estatisticamente não significantes. Uma variável importante que tem sinal não esperado e um efeito prático grande é um problema muito mais preocupante e difícil de resolver. Em geral, deve-se pensar mais sobre o modelo e a natureza dos dados, a fim de solucionar tais problemas. Frequentemente, uma estimativa contra-intuitiva e significante resulta da omissão de uma variável fundamental ou de um dos problemas importantes que discutiremos nos Capítulos 9 e 15.

4.3 INTERVALOS DE CONFIANÇA

Sob as hipóteses do modelo linear clássico, podemos facilmente construir um **intervalo de confiança** (IC) para o parâmetro populacional β_j . Os intervalos de confiança são também chamados *estimativas de intervalo*, porque eles dão uma extensão dos valores prováveis do parâmetro populacional, e não somente uma estimativa pontual.

Usando o fato de que $(\hat{\beta}_j - \bar{\beta}_j)/ep(\hat{\beta}_j)$ tem uma distribuição t com $n - k - 1$ graus de liberdade [veja (4.3)], uma simples manipulação algébrica leva a um IC do β_j desconhecido. Um *intervalo de confiança de 95%*, é dado por

$$\hat{\beta}_j \pm c \cdot ep(\hat{\beta}_j), \quad (4.16)$$

em que a constante c é o 97,5º percentil de uma distribuição $t_{n - k - 1}$. Mais precisamente, os limites inferiores e superiores do intervalo de confiança são dados por

$$\underline{\beta}_j = \hat{\beta}_j - c \cdot ep(\hat{\beta}_j)$$

e

$$\bar{\beta}_j = \hat{\beta}_j + c \cdot ep(\hat{\beta}_j),$$

respectivamente.

Neste ponto, é útil rever o significado de um intervalo de significância. Se as amostras aleatórias fossem obtidas repetidas vezes, com $\underline{\beta}_j$ e $\bar{\beta}_j$ calculados a cada vez, então o valor populacional (desconhecido) β_j estaria dentro do intervalo $(\underline{\beta}_j, \bar{\beta}_j)$ em 95% das amostras. Infelizmente, para a única amostra que usamos para construir o IC, não sabemos se β_j está, realmente, contido no intervalo. Esperamos que tenhamos obtido uma amostra que seja uma das 95% de todas as amostras em que a estimativa de intervalo contém β_j , mas não temos essa garantia.

Construir um intervalo de confiança é muito simples quando se usa a tecnologia computacional atual. São necessárias três quantidades: $\hat{\beta}_j$, $ep(\hat{\beta}_j)$ e c . A estimativa do coeficiente e seu erro-padrão são informados por qualquer programa de regressão. Para obter o valor de c , devemos conhecer os graus de liberdade, $n - k - 1$, e o nível de confiança – 95% neste caso. Portanto, o valor de c é obtido da distribuição $t_{n - k - 1}$.

Como um exemplo, para $gl = n - k - 1 = 25$, um intervalo de confiança de 95% para qualquer β_j é dado por $[\hat{\beta}_j - 2,06 \cdot ep(\hat{\beta}_j), \hat{\beta}_j + 2,06 \cdot ep(\hat{\beta}_j)]$.

Quando $n - k - 1 > 120$, a distribuição $t_{n - k - 1}$ está suficientemente próxima da normal para usar o 97,5º percentil de uma distribuição normal padrão para construir um IC de 95%: $\hat{\beta}_j \pm 1,96 \cdot ep(\hat{\beta}_j)$. De fato, quando $n - k - 1 > 50$, o valor de c está próximo demais de 2, de modo que podemos usar uma *regra de bolso* simples para intervalos de confiança de 95%: $\hat{\beta}_j$ mais ou menos duas vezes seu desvio-padrão. Para graus de liberdade pequenos, os percentis exatos devem ser obtidos das tabelas t .

É fácil construir intervalos de confiança para qualquer outro nível de confiança. Por exemplo, um IC de 90% é obtido ao escolher c como o 95º percentil da distribuição $t_{n - k - 1}$. Quando $gl = n - k - 1 = 25$, $c = 1,71$, e portanto o IC de 90% é $\hat{\beta}_j \pm 1,71 \cdot ep(\hat{\beta}_j)$, que é necessariamente mais estreito que o IC de 95%. Para um IC de 99%, c é o 99,5º percentil da distribuição t_{25} . Com $gl = 25$, o IC de 99% é aproximadamente $\hat{\beta}_j \pm 2,79 \cdot ep(\hat{\beta}_j)$, que é inevitavelmente mais largo que o IC de 95%.

Muitos programas de regressão modernos pouparam-nos de fazer quaisquer cálculos ao informar um IC de 95% juntamente com cada coeficiente e seu erro-padrão. Visto que um intervalo de confiança é construído, é fácil realizar um teste de hipóteses bicaudal. Se a hipótese nula for $H_0: \beta_j = a_j$, então H_0 é rejeitada contra $H_1: \beta_j \neq a_j$ ao nível de significância de (por exemplo) 5% se, e somente se, a_j não está no intervalo de confiança de 95%.

EXEMPLO 4.8**(Modelo de Preço Hedônico de Casas)**

Um modelo que explica o preço de um bem em termos das características desse bem é chamado *modelo de preço hedônico*. A equação seguinte é um modelo de preço hedônico de preços de casas; as características são área (*arquad*), número de quartos (*qtdorm*) e número de banheiros (*banhos*). Em geral, preço aparece na forma logarítmica, assim como algumas das variáveis explicativas. Usando $n = 19$ observações sobre casas que foram vendidas em Waltham, Massachusetts, em 1990, a equação estimada (com os erros-padrão entre parênteses abaixo das estimativas dos coeficientes) é

$$\log(\hat{preço}) = 7,46 + 0,634 \log(arquad) - 0,066 qtdorm + 0,158 banhos$$

(1,15)	(0,184)	(0,059)	(0,075)
--------	---------	---------	---------

$$n = 19, R^2 = 0,806.$$

Como tanto *preço* como *arquad* aparecem na forma logarítmica, a elasticidade preço em relação à área é 0,634, de modo que, mantendo o número de quartos e banheiros fixos, um aumento de 1% na área aumenta o preço previsto da casa em cerca de 0,634%. Podemos construir um intervalo de confiança de 95% para a elasticidade populacional usando o fato de que o modelo estimado tem $n - k - 1 = 19 - 3 - 1 = 15$ graus de liberdade. Da Tabela G.2, achamos o 97,5º percentil de uma distribuição t_{25} : $c = 2,131$. Assim, o intervalo de confiança para $\beta_{\log(arquad)}$ é $0,634 \pm 2,131(0,184)$, ou $(0,242, 1,026)$. Como zero está excluído desse intervalo de confiança, rejeitamos $H_0: \beta_{\log(arquad)} = 0$ contra a alternativa bilateral ao nível de 5%.

O coeficiente de *qtdorm* é negativo, o que parece contra-intuitivo. Entretanto, é importante lembrar a natureza *ceteris paribus* do coeficiente: ele mede o efeito de mais um quarto, mantendo fixos o tamanho da casa e o número de banheiros. Se duas casas têm o mesmo tamanho, mas uma tem mais quartos, então a casa com mais quartos tem quartos menores; mais quartos menores não é, necessariamente, uma coisa boa. Em qualquer caso, podemos ver que o intervalo de confiança de 95% para β_{qtdorm} é um pouco largo, e ele contém o valor zero: $-0,66 \pm 2,131(0,059)$ ou $(-0,192, 0,060)$. Assim, quartos não tem um efeito *ceteris paribus* estatisticamente significante sobre o preço das casas.

Dados o tamanho e o número de quartos, prevê-se que um banheiro a mais aumenta o preço da casa em cerca de 15,8%. (Lembre-se de que devemos multiplicar o coeficiente de *banhos* por 100 para obter o efeito em percentagem.) O intervalo de confiança de 95% para β_{banhos} é $(-0,002, 0,318)$. Nesse caso, zero está por muito pouco dentro do intervalo de confiança; assim, tecnicamente falando, β_{banhos} não é estatisticamente significante ao nível de 5% contra uma hipótese alternativa bilateral. Como ele está muito próximo de ser significante, provavelmente concluiríamos que o número de banheiros tem um efeito sobre $\log(\hat{preço})$.

Você deve lembrar que um intervalo de confiança é tão bom quanto as hipóteses subjacentes feitas para construí-lo. Se omitirmos fatores importantes que são correlacionados com as variáveis explicativas, então as estimativas dos coeficientes não são confiáveis: MQO é viesado. Se a heteroscedasticidade está presente — por exemplo, no exemplo anterior, se a variância de $\log(\hat{preço})$ depende de qualquer uma das variáveis explicativas —, então o erro-padrão não é válido como uma estimativa de ep ($\hat{\beta}_j$) (como discutido na Seção 3.4), e o intervalo de confiança calculado ao se usar esses erros-padrão não será, verdadeiramente, um IC de 95%. Também usamos a hipótese de normalidade dos erros para obter esses ICs mas, como veremos no Capítulo 5, isso não é tão importante para aplicações que envolvem centenas de observações.

4.4 TESTES DE HIPÓTESES SOBRE UMA COMBINAÇÃO LINEAR DOS PARÂMETROS

As duas seções anteriores mostraram como usar o teste de hipóteses clássico ou os intervalos de confiança para testar hipóteses sobre um único β_j de cada vez. Nas aplicações, devemos freqüentemente testar hipóteses que envolvem mais de um dos parâmetros da população. Nesta seção, vamos mostrar como testar uma única hipótese envolvendo mais de um dos β_j . A Seção 4.5 mostrará como testar hipóteses múltiplas.

Para ilustrar a abordagem geral, consideraremos um modelo simples para comparar os retornos da educação de cursos superiores profissionalizantes de dois anos (*junior colleges*) e de cursos superiores de quatro anos (*four-year colleges*); por simplicidade, vamos nos referir ao último como “universidades”. [Kane e Rouse (1995) fazem uma análise detalhada dos retornos dos *junior colleges* e dos *four-year colleges*.] A população inclui as pessoas com o ensino médio completo que trabalham, e o modelo é

$$\log(\text{salário}) = \beta_0 + \beta_1 cp; \beta_2 univ + \beta_3 exper + u, \quad (4.17)$$

em que cp é o número de anos freqüentados em um curso superior profissionalizante de dois anos e $univ$ é o número de anos freqüentados em um curso superior de quatro anos. Note que qualquer combinação de curso profissionalizante e curso de quatro anos é permitida, incluindo $cp = 0$ e $univ = 0$.

A hipótese de interesse é se um ano no curso profissionalizante é equivalente a um ano na universidade: isso é expresso como

$$H_0: \beta_1 = \beta_2. \quad (4.18)$$

Sob H_0 , um ano a mais no curso profissionalizante e um ano a mais na universidade levam ao mesmo aumento percentual *ceteris paribus* em *salário*. Na maioria dos casos, a alternativa de interesse é unilateral: um ano no curso profissionalizante é menos valioso do que um ano na universidade. Isso é expresso como

$$H_1: \beta_1 < \beta_2. \quad (4.19)$$

As hipóteses (4.18) e (4.19) dizem respeito a *dois* parâmetros, β_1 e β_2 , uma situação com a qual não tínhamos nos deparado ainda. Não podemos simplesmente usar as estatísticas t individuais de $\hat{\beta}_1$ e $\hat{\beta}_2$ para testar H_0 . Entretanto, conceitualmente, não há dificuldade em construir uma estatística t para testar (4.18). A fim de fazer isso, vamos reescrever a hipótese nula e a alternativa como $H_0: \beta_1 - \beta_2 = 0$ e $H_1: \beta_1 - \beta_2 < 0$, respectivamente. A estatística t é baseada em se a diferença estimada $\hat{\beta}_1 - \hat{\beta}_2$ é suficientemente menor que zero para assegurar a rejeição de (4.18) em favor de (4.19). Para considerar o erro de nossos estimadores, padronizamos essa diferença ao dividi-la pelo erro-padrão:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\text{ep}(\hat{\beta}_1 - \hat{\beta}_2)}. \quad (4.20)$$

Uma vez que temos a estatística t de (4.20), o teste segue o procedimento anterior. Escolhemos um nível de significância para o teste ϵ , com base nos gl , obtemos um valor crítico. Como a alternativa é da forma (4.19), a regra de rejeição é da forma $t < -c$, em que c é um valor positivo escolhido de uma distribuição t apropriada. Ou então calculamos a estatística t e, em seguida, o p -valor (veja a Seção 4.2).

A única coisa que faz com que o teste da igualdade de dois parâmetros diferentes seja mais difícil do que testar um único β_j é a obtenção do erro-padrão do denominador de (4.20). Obter o numerador é trivial, uma vez que tenhamos computado a regressão de MQO. Ao usar os dados do arquivo TWOYEAR.RAW, provenientes de Kane e Rouse (1995), estimamos a equação (4.17):

$$\begin{aligned} \log(\hat{\text{Salário}}) &= 1,472 + 0,0667 \text{ cp} + 0,0769 \text{ univ} + 0,0049 \text{ exper} \\ &\quad (0,021) \quad (0,0068) \quad (0,0023) \quad (0,0002) \\ n &= 6,763, R^2 = 0,222. \end{aligned} \quad [4.21]$$

De (4.21), fica claro que cp e $univ$ têm ambos os efeitos – econômico e estatístico – significantes sobre o salário. Isso é, certamente, de interesse, mas estamos mais interessados em testar se a diferença estimada dos coeficientes é estatisticamente significante. A diferença é estimada como $\hat{\beta}_1 - \hat{\beta}_2 = -0,0102$, de modo que o retorno de um ano em um curso profissionalizante é cerca de um ponto percentual menor que um ano na universidade. Economicamente, isso não é uma diferença trivial. A diferença de $-0,0102$ é o numerador da estatística t em (4.20).

Infelizmente, os resultados da regressão em (4.21) *não* contêm informações suficientes para obter o erro-padrão de $\hat{\beta}_1 - \hat{\beta}_2$. Pode ser tentador afirmar que $\text{ep}(\hat{\beta}_1 - \hat{\beta}_2) = \text{ep}(\hat{\beta}_1) - \text{ep}(\hat{\beta}_2)$, mas isso não é verdade. De fato, se invertêssemos os papéis de $\hat{\beta}_1$ e $\hat{\beta}_2$, terminariamos com um erro-padrão negativo da diferença ao usar a diferença dos erros-padrão. Estes devem *sempre* ser positivos porque eles são estimativas dos desvios-padrão. Embora o erro-padrão da diferença $\hat{\beta}_1 - \hat{\beta}_2$ dependa, certamente, de $\text{ep}(\hat{\beta}_1)$ e $\text{ep}(\hat{\beta}_2)$, ele depende de uma maneira um pouco complicada. Para encontrar $\text{ep}(\hat{\beta}_1 - \hat{\beta}_2)$, primeiro obtemos a variância da diferença. Ao usar os resultados das variâncias do Apêndice B (disponível no site da Thomson), temos

$$\text{Var}(\hat{\beta}_1 - \hat{\beta}_2) = \text{Var}(\hat{\beta}_1) + \text{Var}(\hat{\beta}_2) - 2 \text{Cov}(\hat{\beta}_1, \hat{\beta}_2). \quad [4.22]$$

Observe, cuidadosamente, como as duas variâncias são *somadas* e, então, a covariância é subtraída duas vezes. O desvio-padrão de $\hat{\beta}_1 - \hat{\beta}_2$ é exatamente a raiz quadrada de (4.22), e, como $[\text{ep}(\hat{\beta}_1)]^2$ é um estimador não-viesado de $\text{Var}(\hat{\beta}_1)$, e similarmente para $[\text{ep}(\hat{\beta}_2)]^2$, temos

$$\text{ep}(\hat{\beta}_1 - \hat{\beta}_2) = \{[\text{ep}(\hat{\beta}_1)]^2 + [\text{ep}(\hat{\beta}_2)]^2 - 2s_{12}\}^{1/2}, \quad [4.23]$$

em que s_{12} é uma estimativa de $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$. Não mostramos uma fórmula para $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$. Alguns programas de regressão têm características que nos permitem obter s_{12} , caso em que se pode calcular o erro-padrão em (4.23) e, em seguida, a estatística t em (4.20). O Apêndice E mostra com usar a álgebra matricial para obter s_{12} .

Vamos sugerir uma outra rota que é muito mais simples de calcular, menos provável de levar a erro e prontamente aplicável a uma variedade de problemas. Em vez de tentar calcular $\text{ep}(\hat{\beta}_1 - \hat{\beta}_2)$ a partir de (4.23), é muito mais fácil estimar um modelo diferente que produz, diretamente, o erro-padrão

de interesse. Defina um novo parâmetro como a diferença entre β_1 e β_2 : $\theta_1 = \beta_1 - \beta_2$. Então, queremos testar

$$H_0: \theta_1 = 0 \text{ contra } H_1: \theta_1 < 0. \quad (4.24)$$

A estatística t em (4.20), em termos de $\hat{\theta}_1$, é exatamente $t = \hat{\theta}_1/\text{ep}(\hat{\theta}_1)$. O desafio é encontrar $\text{ep}(\hat{\theta}_1)$.

Podemos obter isso ao reescrever o modelo, de modo que θ_1 aparece diretamente como o coeficiente de uma das variáveis independentes. Como $\theta_1 = \beta_1 - \beta_2$, podemos também escrever $\beta_1 = \theta_1 + \beta_2$. Inserindo em (4.17) e rearranjando, resulta a equação

$$\begin{aligned} \log(\text{salário}) &= \beta_0 + (\theta_1 + \beta_2)\text{cp} + \beta_2\text{univ} + \beta_3\text{exper} + u \\ &= \beta_0 + \theta_1\text{cp} + \beta_2(\text{cp} + \text{univ}) + \beta_3\text{exper} + u. \end{aligned} \quad (4.25)$$

A ideia fundamental é que o parâmetro θ_1 , cuja hipótese estamos interessados em testar, multiplica agora a variável cp . O intercepto ainda é β_0 , e exper também aparece multiplicado por β_3 . Mais importante, há uma nova variável multiplicando β_2 , a saber, $\text{pc} + \text{univ}$. Assim, se quisermos estimar diretamente θ_1 e obter o erro-padrão $\hat{\theta}_1$, então devemos construir a nova variável $\text{pc} + \text{univ}$ e incluí-la no modelo de regressão no lugar de univ . Nesse exemplo, a nova variável tem uma interpretação natural: ela é o *total* de anos de curso superior; assim, defina $\text{totgrad} = \text{pc} + \text{univ}$ e escreva (4.25) como

$$\log(\text{salário}) = \beta_0 + \theta_1\text{cp} + \beta_2\text{totgrad} + \beta_3\text{exper} + u. \quad (4.26)$$

O parâmetro β_1 desapareceu do modelo, enquanto θ_1 aparece explicitamente. Esse modelo é, de fato, uma maneira diferente de escrever o modelo original. A única razão pela qual definimos esse modelo é que, quando o estimamos, o coeficiente de cp ($\hat{\theta}_1$), e, mais importante, $\text{ep}(\hat{\theta}_1)$ é informado juntamente com a estimativa. A estatística t que queremos é a que está relacionada a cp (e *não* à totgrad), e é informada por qualquer programa de regressão.

Quando fazemos isso com as 6.763 observações utilizadas anteriormente, o resultado é

$$\begin{aligned} \log(\hat{\text{salário}}) &= 1,472 - 0,0102 \text{ cp} + 0,0769 \text{ totgrad} + 0,0049 \text{ exper} \\ &\quad (0,021) (0,0069) \quad (0,0023) \quad (0,0002) \\ &n = 6,763, R^2 = 0,222. \end{aligned} \quad (4.27)$$

Nessa equação, o único número que não poderíamos obter de (4.21) é o erro-padrão da estimativa $-0,0102$, que é igual a $0,0069$. A estatística t para testar (4.18) é $-0,0102/0,0069 = -1,48$. Contra a alternativa unilateral (4.19), o p -valor é cerca de $0,070$; assim, há alguma, mas não forte, evidência contra (4.18).

O intercepto e a estimativa de inclinação de exper , juntamente com os erros-padrão, são os mesmos de (4.21). Esse fato *deve* ser exato, e ele fornece uma maneira de checar se a equação transformada foi apropriadamente estimada. O coeficiente da nova variável, totgrad , é a mesma do coeficiente de univ em (4.21), e o erro-padrão também é o mesmo. Sabemos que isso deve acontecer ao comparar (4.17) e (4.25).

É bastante simples calcular um intervalo de confiança de 95% para $\theta_1 = \beta_1 - \beta_2$. Usando a aproximação normal padronizada, o IC é obtido da maneira usual: $\hat{\theta}_1 \pm 1,96 \text{ ep}(\hat{\theta}_1)$, que, nesse caso, leva a $-0,0102 \pm 0,0135$.

A estratégia de reescrever o modelo, de modo que ele contenha o parâmetro de interesse, funciona em todos os casos e é fácil de implementar. (Veja os problemas 4.12 e 4.14 para outros exemplos.)

4.5 TESTES DE RESTRIÇÕES LINEARES MÚLTIPLAS: O TESTE F

A estatística t associada com qualquer coeficiente de MQO pode ser usada para testar se o parâmetro desconhecido correspondente na população é igual a qualquer constante dada (freqüentemente, mas não sempre, zero). Acabamos de mostrar como testar hipóteses sobre uma única combinação linear dos β_j ao rearranjar a equação e computar uma regressão usando variáveis transformadas. No entanto, até agora, somente cobrimos hipóteses que envolvem uma *única* restrição. Freqüentemente, desejamos testar hipóteses *múltiplas* sobre os parâmetros subjacentes $\beta_0, \beta_1, \dots, \beta_k$. Vamos começar com o caso de testar se um conjunto de variáveis independentes não tem efeito parcial sobre uma variável dependente.

Teste de Restrições de Exclusão

Já sabemos como testar se uma variável particular não tem efeito sobre a variável dependente, usando a estatística t . Agora, queremos testar se um *grupo* de variáveis não tem efeito sobre a variável dependente. Mais precisamente, a hipótese nula é que um conjunto de variáveis não tem efeito sobre y , já que outro conjunto de variáveis foi controlado.

Como uma ilustração do porquê de testar a significância de um grupo é útil, vamos considerar o seguinte modelo que explica os salários dos jogadores da principal liga de beisebol dos Estados Unidos:

$$\begin{aligned} \log(\text{salário}) = & \beta_0 + \beta_1 \text{anos} + \beta_2 \text{jogosano} + \beta_3 \text{rebmed} + \\ & \beta_4 \text{hrunano} + \beta_5 \text{rebrunano} + u, \end{aligned} \quad (4.28)$$

em que *salário* é o salário total do jogador em 1993, *anos* corresponde aos anos do jogador na liga, *jogosano* é a média de partidas jogadas por ano, *rebmed* é a média de rebatidas na carreira do jogador, *hrunano* corresponde a rebatidas que redundaram em pontos (volta completa por todas as bases) por ano, e *rebrunano* corresponde a rebatidas que redundaram em corrida até a próxima base por ano. Suponha que queiramos testar a hipótese nula de que, uma vez que anos na liga e jogos por ano foram controlados, as estatísticas que medem o desempenho – *rebmed*, *hrunano* e *rebrunano* – não têm efeito sobre o salário. Essencialmente, a hipótese nula expressa que a produtividade, medida pelas estatísticas do beisebol, não tem efeito sobre o salário.

Em termos dos parâmetros do modelo, a hipótese nula é formulada como

$$H_0: \beta_3 = 0, \beta_4 = 0, \beta_5 = 0. \quad (4.29)$$

A hipótese nula (4.29) constitui três **restrições de exclusão**: se (4.29) é verdadeira, então *rebmed*, *hrunano* e *rebrunano* não têm efeito sobre $\log(\text{salário})$ após *anos* e *jogosano* terem sido controlados e, portanto, deveriam ser excluídos do modelo. Esse é um exemplo de conjunto de **restrições múltiplas**

porque estamos colocando mais de uma restrição sobre os parâmetros de (4.28); posteriormente, veremos mais exemplos gerais de restrições múltiplas. Um teste de restrições múltiplas é chamado **teste de hipóteses múltiplas** ou o **teste de hipóteses conjuntas**.

Qual seria a alternativa a (4.29)? Se o que temos em mente é que “estatísticas de desempenho importam, mesmo após controlar as variáveis anos na liga e jogos por ano”, então a hipótese alternativa é simplesmente

$$H_1: H_0 \text{ não é verdadeira.} \quad (4.30)$$

A alternativa (4.30) se mantém quando pelo menos um dos β_3 , β_4 ou β_5 for diferente de zero. (Qualquer um deles ou todos poderiam ser diferentes de zero.) O teste que estudamos aqui é construído para detectar qualquer violação de H_0 . Ele também é válido quando a hipótese alternativa é algo como $H_1: \beta_3 > 0$, ou $\beta_4 > 0$, ou $\beta_5 > 0$, mas ele não será o melhor teste possível sob essas alternativas. Não temos o espaço ou a formação estatística necessários para cobrir testes mais poderosos sob hipóteses alternativas unilaterais múltiplas.

Como devemos proceder para testar (4.29) contra (4.30)? É tentador testar (4.29) usando as estatísticas t das variáveis *rebmed*, *hrunano* e *rebrunano* para determinar se cada variável é *individualmente* significante. Essa opção não é apropriada. Uma estatística t particular testa uma hipótese que não coloca restrições sobre os outros parâmetros. Além disso, teríamos três resultados para resolver o problema – um para cada estatística t . Qual deles constituiria a rejeição de (4.29) ao nível dc, por exemplo, 5%? Dever-se-ia exigir que todas as três estatísticas t são significantes ao nível de 5% ou somente uma das três? Essas são questões difíceis, e felizmente não temos de respondê-las. Além do mais, usar estatísticas t separadas para testar uma hipótese múltipla como (4.29) pode ser muito enganoso. Precisamos de uma maneira para testar as restrições de exclusão *conjuntamente*.

Para ilustrar essas questões, estimamos a equação (4.28) usando os dados do arquivo MLB1.RAW. Obtemos

$$\begin{aligned} \log(\text{salary}) = & 11,10 + 0,0689 \text{ anos} + 0,0126 \text{ jogosano} \\ & (0,29) \quad (0,0121) \quad (0,0026) \\ & + 0,00098 \text{ rebmed} + 0,0144 \text{ hrunano} + 0,0108 \text{ rebrunano} \\ & (0,00110) \quad (0,0161) \quad (0,0072) \\ & n = 353, \text{ SQR} = 183,186, R^2 = 0,6278, \end{aligned} \quad (4.31)$$

em que SQR é a soma dos resíduos quadrados. (vamos usá-lo mais tarde.) A fim de facilitar futuras comparações, deixamos vários números após a vírgula em SQR e no R -quadrado. A equação (4.31) revela que, enquanto *anos* e *jogosano* são estatisticamente significantes, nenhuma das variáveis *rebmed*, *hrunano* e *rebrunano* tem uma estatística t estatisticamente significante contra uma alternativa bilateral ao nível de significância de 5%. (A estatística t de *rebrunano* está muito próxima de ser significante; seu p -valor bilateral é 0,134.) Assim, baseados nas três estatísticas t , parece que não podemos rejeitar H_0 .

Essa conclusão revela-se errada. A fim de ver isso, devemos derivar um teste de restrição múltipla cuja distribuição seja conhecida e tabelada. A soma dos resíduos quadrados aparece, agora, para dar uma base muito conveniente para testar hipóteses múltiplas. Também mostraremos como o R -quadrado pode ser usado no caso especial de testar restrições de exclusão.

Conhecer a soma dos resíduos quadrados em (4.31) não nos diz nada sobre a decisão quanto à hipótese nula em (4.29). O que nos dirá algo é: saber de quanto aumenta SQR quando retiramos as variáveis *rebmed*, *hrunano* e *rebrunano* do modelo. Lembre-se de que, como as estimativas de MQO são escolhidas para minimizar a soma dos resíduos quadrados, o SQR *sempre* aumenta quando variáveis são retiradas do modelo; esse é um fato algébrico. A questão é saber se esse aumento é suficientemente grande, *relativamente* ao SQR do modelo com todas as variáveis, para garantir a rejeição da hipótese nula.

O modelo sem as três variáveis em questão é simplesmente

$$\log(salário) = \beta_0 + \beta_1 anos + \beta_2 jogosano + u. \quad (4.32)$$

No contexto do teste de hipóteses, a equação (4.32) é o **modelo restrito** para testar (4.29); o modelo (4.28) é chamado **modelo irrestrito**. O modelo restrito sempre tem menos parâmetros que o modelo irrestrito.

Quando estimarmos o modelo restrito usando os dados do arquivo MLB1.RAW, obtemos

$$\begin{aligned} \log(salário) &= 11,22 + 0,0713 \text{ anos} + 0,0202 \text{ jogosano} \\ &\quad (0,11) \quad (0,0125) \quad (0,0013) \\ n &= 353, \text{ SQR} = 198,311, R^2 = 0,5971. \end{aligned} \quad (4.33)$$

Como imaginamos, o SQR de (4.33) é maior que o SQR de (4.31), e o *R*-quadrado do modelo restrito é menor que o *R*-quadrado do modelo irrestrito. O que precisamos decidir é se, ao passarmos do modelo irrestrito para o modelo restrito, o aumento em SQR (183,186 para 198,311) é suficientemente grande para garantir a rejeição de (4.29). Como em todo teste, a resposta depende do nível de significância do teste. No entanto, não podemos realizar o teste a um determinado nível de significância até que tenhamos uma estatística cuja distribuição seja conhecida, e possa ser tabelada, sob H_0 . Assim, precisamos de uma maneira de combinar as informações dos dois SQRs para obter uma estatística de teste com uma distribuição conhecida sob H_0 .

Podemos derivar o teste para o caso geral, visto que isso não é tão difícil. Escreva o modelo *irrestrito* com k variáveis independentes como

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u; \quad (4.34)$$

o número de parâmetros no modelo irrestrito é $k + 1$. (Lembre-se de adicionar um por causa do intercepto.) Suponha que temos q restrições de exclusão para testar: isto é, a hipótese nula afirma que q variáveis em (4.34) têm coeficientes zero. Por simplicidade notacional, assuma que sejam as q últimas variáveis da lista de variáveis independentes: x_{k-q+1}, \dots, x_k . (A ordem das variáveis, evidentemente, é arbitrária e não importa.) A hipótese nula é formulada como

$$H_0: \beta_{k-q+1} = 0, \dots, \beta_k = 0, \quad (4.35)$$

que coloca q restrições de exclusão sobre o modelo (4.34). A hipótese alternativa a (4.35) é simplesmente que H_0 é falsa; isso significa que pelo menos um dos parâmetros listados em (4.35) é diferente de zero. Quando impomos as restrições sob H_0 , ficamos com o modelo restrito:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_{k-q} x_{k-q} + u. \quad (4.36)$$

Nesta seção, vamos assumir que ambos os modelos irrestrito e restrito contêm um intercepto, já que esse é o caso mais amplamente encontrado na prática.

Agora, vamos à estatística de teste propriamente dita. Anteriormente, sugerimos que olhar para o aumento relativo em SQR quando nos movemos do modelo irrestrito para o restrito deveria ser informativo para testar a hipótese (4.35). A **estatística F** (ou razão F) é definida como

$$F \equiv \frac{(SQR_r - SQR_{ir})/q}{SQR_{ir}/(n - k - 1)}, \quad (4.37)$$

em que SQR_r é a soma dos resíduos quadrados do modelo restrito, e SQR_{ir} é a soma dos resíduos quadrados do modelo irrestrito.

Você deveria observar imediatamente que, como SQR_r não pode ser maior que SQR_{ir} , a estatística F é sempre não-negativa (e quase sempre estritamente positiva). Assim, se você calcular uma estatística F negativa, algo está errado; em geral, a ordem dos SQRs no numerador de F é equivocadamente invertida. Também, o SQR no denominador de F é o SQR do modelo irrestrito. A maneira mais fácil de lembrar onde os SQRs aparecem é pensar em F medindo o aumento relativo em SQR quando nos movemos do modelo irrestrito para o restrito.

Exemplo 4.10

Considere a possibilidade de relacionar o desempenho individual em um teste padronizado, *nota*, a uma variedade de outras variáveis. Fatores relativos à escola incluem o tamanho médio da classe, os gastos por estudante, o salário médio dos professores e o total de matrículas escolares. Outras variáveis específicas aos estudantes são a renda familiar, a educação da mãe, a educação do pai e o número de irmãos. O modelo é

$$\begin{aligned} \text{nota} = & \beta_0 + \beta_1 tclasse + \beta_2 gasto + \beta_3 totalsalp + \\ & \beta_4 matricl + \beta_5 rendfam + \beta_6 educm + \\ & \beta_7 educp + \beta_8 irmãos + u. \end{aligned}$$

Formule a hipótese nula de que as variáveis específicas aos estudantes não têm efeito sobre o desempenho no teste padronizado, uma vez que os fatores relativos à escola sejam controlados. Quais são os valores de k e q nesse exemplo? Escreva a versão restrita do modelo.

A diferença nos SQRs no numerador de F é dividida por q , o qual é o número de restrições impostas ao nos movermos do modelo irrestrito para o restrito (q variáveis independentes foram retiradas). Portanto, podemos escrever

$$q = \text{graus de liberdade do numerador} = gl_r - gl_{ir}, \quad (4.38)$$

o que também mostra que q é a diferença nos graus de liberdade entre os modelos restrito e irrestrito. (Lembre-se de que gl = número de observações-número de parâmetros estimados.) Visto que o modelo restrito tem menos parâmetros — e cada modelo é estimado usando as mesmas n observações —, gl_r é sempre maior que gl_{ir} .

O SQR no denominador de F é dividido pelos graus de liberdade do modelo irrestrito:

$$n - k - 1 = \text{graus de liberdade do denominador} = gl_{ir}. \quad (4.39)$$

De fato, o denominador de F é exatamente o estimador não-viesado de $\sigma^2 = \text{Var}(u)$ do modelo irrestrito.

Em uma aplicação particular, calcular a estatística F é mais fácil do que ler penosamente a notação um pouco incômoda usada para descrever o caso geral. Em primeiro lugar, obtemos os graus de liberdade do modelo irrestrito, gl_{ir} . Então, contamos quantas variáveis estão excluídas no modelo restrito; esse é o valor de q . Os SQRs são informados em toda regressão de MQO e, portanto, é simples compor a estatística F .

Na regressão do salário da principal liga de beisebol, $n = 353$, e o modelo completo (4.28) contém seis parâmetros. Assim, $n - k - 1 = gl_{ir} = 353 - 6 = 347$. O modelo restrito (4.32) contém menos três variáveis independentes que (4.28), e portanto, $q = 3$. Assim, temos todos os ingredientes para calcular a estatística F ; vamos adiar o cálculo até que saibamos o que fazer com ele.

A fim de usar a estatística F , devemos conhecer sua distribuição amostral sob a hipótese nula para escolher os valores críticos e as regras de rejeição. Pode ser mostrado que, sob H_0 (e assumindo que as hipóteses do MCL se mantêm), F é distribuído como uma variável aleatória F com $(q, n - k - 1)$ graus de liberdade. Escrevemos isso como

$$F \sim F_{q, n - k - 1}.$$

A distribuição de $F_{q, n - k - 1}$ está tabelada e disponível em tabelas estatísticas (veja a Tabela G.3) e, ainda mais importante, em programas estatísticos.

Não vamos derivar a distribuição F porque a matemática é muito complicada. Basicamente, pode ser mostrado que a equação (4.37) é, de fato, a razão de duas variáveis aleatórias qui-quadradas independentes, divididas por seus respectivos graus de liberdade. A variável aleatória qui-quadrada do numerador tem q graus de liberdade, e a qui-quadrada do denominador tem $n - k - 1$ graus de liberdade. Essa é a definição de uma variável aleatória com distribuição F (veja o Apêndice B, disponível no site da Thomson).

Da definição de F , é bastante claro que rejeitaremos H_0 em favor de H_1 quando F for suficientemente “grande”. A grandeza depende de nosso nível de significância escolhido. Suponha que decidimos por um teste ao nível de 5%. Seja c o 95º percentil da distribuição $F_{q, n - k - 1}$. O valor crítico depende de q (os gl do numerador) e $n - k - 1$ (os gl do denominador). É importante guardar corretamente os graus de liberdade do numerador e do denominador.

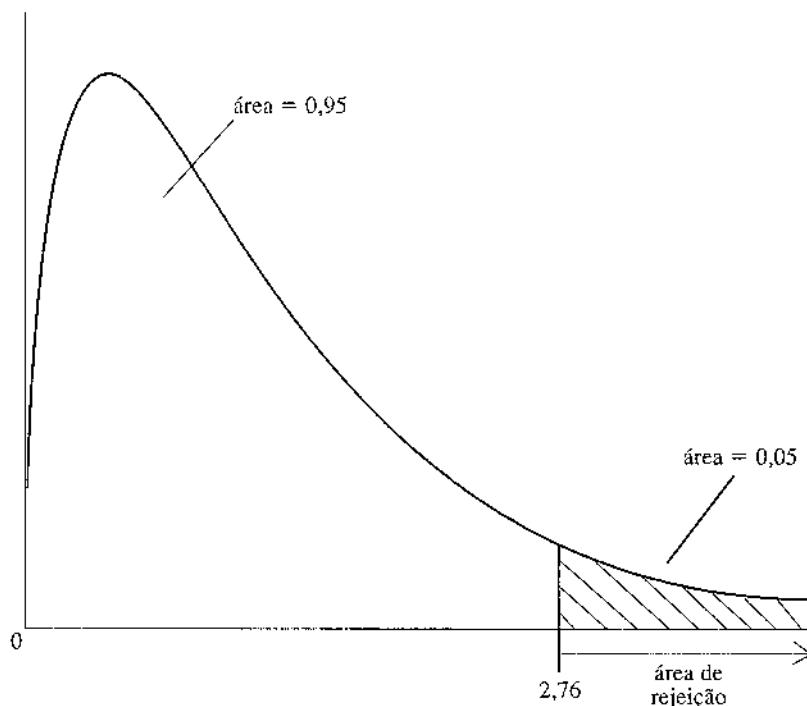
Os valores críticos de 10%, 5% e 1% da distribuição F são dados na Tabela 1.3. A regra de rejeição é simples. Uma vez obtido c , rejeitaremos H_0 em favor de H_1 , ao nível de significância escolhido se

$$F > c. \quad (4.40)$$

Com um nível de significância de 5%, $q = 3$ e $n - k - 1 = 60$, o valor crítico é $c = 2,76$. Rejeitariamós H_0 ao nível de significância de 5% se o valor calculado da estatística F excedesse 2,76. O valor crítico a 5% e a região de rejeição são apresentados na Figura 4.7. Para os mesmos graus de liberdade, o valor crítico a 1% é 4,13.

Figura 4.7

O valor crítico de 5% e a região de rejeição em uma distribuição $F_{3,60}$.



Na maioria das aplicações, os graus de liberdade do numerador (q) serão notadamente menores que os graus de liberdade do denominador ($n - k - 1$). As aplicações em que $n - k - 1$ menor têm menos probabilidade de serem bem-sucedidas porque os parâmetros do modelo irrestrito provavelmente não serão estimados com precisão. Quando os gl do denominador alcançam cerca de 120, a distribuição F não é mais sensível a eles. (Isso é totalmente semelhante à distribuição t aproximada pela distribuição normal padronizada quando os gl tornam-se grandes.) Assim, há uma entrada na tabela para o denominador $gl = \infty$, e isso é o que usamos com amostras grandes (visto que $n - k - 1$ é, então, grande). Uma formulação semelhante é válida para os gl do numerador grandes, mas isso raramente ocorre nas aplicações.

Se H_0 é rejeitada, dizemos que x_{k-q+1}, \dots, x_k são **estatisticamente significantes conjuntamente** (ou apenas *conjuntamente significantes*) ao nível de significância apropriado. Esse teste sozinho não nos permite dizer quais das variáveis têm um efeito parcial sobre y ; todas elas podem afetar y ou talvez somente uma afeta. Se a hipótese nula não for rejeitada, as variáveis são **conjuntamente não significantes**, o que, em geral, justifica retirá-las do modelo.

No exemplo da principal liga de beisebol com três graus de liberdade do numerador e 347 graus de liberdade do denominador, o valor crítico a 5% é 2,60, e o valor crítico a 1% é 3,78. Rejeitamos H_0 ao nível de 1% se F está acima de 3,78; rejeitamos H_0 ao nível de 5% se F está acima de 2,60.

Estamos agora em posição para testar a hipótese com a qual iniciamos esta seção: após controlar *anos* e *jogosano*, as variáveis *rebmed*, *hrunano* e *rebrunano* não têm efeito sobre os salários dos jogadores. Na prática, é mais fácil, em primeiro lugar, calcular $(SQR_r - SQR_{ir})/SQR_{ir}$ e então multiplicar o resultado por $(n - k - 1)/q$; a razão pela qual a fórmula é expressa como em (4.37) é que ela torna mais fácil guardar corretamente os graus de liberdade do numerador e do denominador. Usando os SQRs em (4.31) e (4.33), temos

$$F = \frac{(198,311 - 183,186)}{183,186} \cdot \frac{347}{3} \approx 9,55.$$

Esse número está bem acima do valor crítico de 1% da distribuição F com 3 e 347 graus de liberdade, e, portanto, rejeitamos completamente a hipótese de que *rebmed*, *hrunano* e *rebrunano* não têm efeito sobre o salário.

O resultado do teste conjunto pode parecer surpreendente à luz das estatísticas t não significantes das três variáveis. O que está acontece é que as variáveis *hrunano* e *rebrunano* são altamente correlacionadas, e essa multicolinearidade torna difícil descobrir o efeito parcial de cada variável; isso é refletido nas estatísticas t individuais. A estatística F testa se essas variáveis (incluindo *rebmed*) são conjuntamente significantes, e a multicolinearidade entre *hrunano* e *rebrunano* é muito menos relevante para testar essa hipótese. No Problema 4.16, pediremos que você estime novamente o modelo retirando *rebrunano*, caso em que *hrunano* torna-se muito significante. O mesmo é verdadeiro para *rebrunano* quando *hrunano* é retirado do modelo.

A estatística F é freqüentemente útil para testar a exclusão de um grupo de variáveis quando as variáveis do grupo são altamente correlacionadas. Por exemplo, suponha que queiramos testar se o desempenho da empresa afeta os salários dos seus diretores executivos. Há muitas maneiras de medir o desempenho das empresas, e não é claro dizer, antecipadamente, qual medida é a mais importante. Como as medidas de desempenho das empresas são, provavelmente, altamente correlacionadas, esperar encontrar medidas individualmente significantes pode ser pedir demais, devido à multicolinearidade. No entanto, um teste F pode ser usado para determinar se, como um grupo, as variáveis de desempenho das empresas afetam o salário.

Relação entre as Estatísticas F e t

Vimos nesta seção como a estatística F pode ser usada para testar se um grupo de variáveis deve ser incluído em um modelo. O que aconteceria se aplicássemos a estatística F ao caso de testar a significância de uma única variável independente? Esse caso certamente não é excluído pelo desenvolvimento anterior. Por exemplo, podemos descrever a hipótese nula como $H_0: \beta_k = 0$ e $q = 1$ (para testar a única restrição de exclusão, de que x_k pode ser excluído do modelo). Da Seção 4.2, sabemos que a estatística t de β_k pode ser usada para testar essa hipótese. A questão, então, é: temos duas maneiras separadas de testar hipóteses sobre um único coeficiente? A resposta é não. É possível mostrar que a estatística F para testar a exclusão de uma única variável é igual ao quadrado da estatística t correspondente. Como t_{n-k-1}^2 tem uma distribuição $F_{1, n-k-1}$, as duas abordagens levam exatamente ao mesmo resultado, desde que a hipótese alternativa seja bilateral. A estatística t é mais flexível para testar uma única hipótese porque ela pode ser usada para testar alternativas unilaterais. Visto que as estatísticas t também são mais fáceis de serem obtidas do que as estatísticas F , não há razão para usar uma estatística F para testar hipóteses sobre um único parâmetro.

Na regressão dos jogadores da principal liga de beisebol, vimos que duas (ou mais) variáveis que têm, cada uma, estatísticas t não significantes podem ser conjuntamente muito significantes. Também é possível que, em um grupo de muitas variáveis explicativas, uma variável tenha uma estatística t significante, mas o grupo de variáveis é conjuntamente não significante aos níveis de significância usuais. O que devemos fazer com esse tipo de resultado? Em termos concretos, suponha que, em um modelo com muitas variáveis explicativas, não possamos rejeitar a hipótese nula de que $\beta_1, \beta_2, \beta_3, \beta_4$ e β_5 são todos iguais a zero ao nível de 5%, ainda que a estatística t de $\hat{\beta}_1$ seja significante ao nível de 5%. Logicamente, não podemos ter $\beta_1 \neq 0$ e também ter $\beta_1, \beta_2, \beta_3, \beta_4$ e β_5 todos iguais a zero! Contudo, quando se trata de fazer um teste, é possível que agrupemos um punhado de variáveis não significantes juntamente com uma variável significante e concluamos que o conjunto inteiro de variáveis é conjuntamente não significante. (Tais possíveis conflitos entre um teste t e um teste F conjunto dão outro exemplo da razão de não devermos “aceitar” hipóteses nulas; podemos somente não rejeitá-las.) Espera-se que a estatística F revele se qualquer combinação de um conjunto de coeficientes é diferente de zero, mas ele nunca é o melhor teste para determinar se um único coeficiente é diferente de zero. O teste t é o mais apropriado para testar uma única hipótese. (Tecnicamente, uma estatística F para restrições conjuntas que incluem $\beta_1 = 0$ tem menos poder de detectar $\beta_1 \neq 0$ do que a estatística t usual. Para uma discussão do poder de um teste, veja a Seção C.6 do Apêndice C, disponível no site da Thomson.)

Infelizmente, o fato de podermos às vezes ocultar uma variável estatisticamente significante entre algumas variáveis não significantes pode levar a equívocos se os resultados da regressão não forem cuidadosamente descritos. Por exemplo, suponha que, num estudo dos determinantes das taxas de aprovação de empréstimos de uma cidade, x_1 é a fração de famílias negras na cidade. Suponha que as variáveis x_2, x_3, x_4 e x_5 sejam as frações de famílias chefiadas por diferentes grupos de idade. Ao explicar as taxas de empréstimos, incluiríamos medidas de renda, riqueza, avaliação de crédito, e assim por diante. Suponha que a idade do chefe de família não tenha efeito sobre as taxas de aprovação de empréstimos, uma vez que as outras variáveis sejam controladas. Mesmo se a raça tiver um efeito marginalmente significante, é possível que as variáveis raça e idade sejam conjuntamente não significantes. Alguém que queira concluir que raça não é um fator importante poderia simplesmente escrever algo como “As variáveis raça e idade foram acrescentadas à equação, mas elas foram conjuntamente não significantes ao nível de 5%”. Felizmente, a revisão atenta impede esses tipos de conclusões enganosas, mas você deve estar consciente de que elas podem ocorrer.

Freqüentemente, quando uma variável é estatisticamente muito significante e ela é testada conjuntamente com outro conjunto de variáveis, o conjunto será conjuntamente significante. Em tais casos, não há mais inconsistência lógica em rejeitar ambas as hipóteses nulas.

A Forma R -quadrado da Estatística F

Para testar restrições de exclusão é freqüentemente mais conveniente ter uma forma da estatística F que possa ser calculada usando os R -quadrados dos modelos restrito e irrestrito.

Uma razão para isso é que o R -quadrado está sempre entre zero e um, enquanto os SQRs podem ser muito grandes, dependendo da unidade de y , o que faz dos cálculos baseados nos SQRs algo entediante. Usando o fato de que $SQR_r = SQT(1 - R_r^2)$ e $SQR_{ir} = SQT(1 - R_{ir}^2)$, podemos substituir esses termos em (4.37) para obter

$$F = \frac{(R_{ir}^2 - R_r^2)/q}{(1 - R_{ir}^2)/(n - k - 1)} \quad (4.41)$$

(observe que os termos SQT são cancelados). Isso se chama a **forma R-quadrado da estatística F**. [Neste ponto, você deve ser advertido de que embora a equação (4.41) seja muito conveniente para testar restrições de exclusão, ela não pode ser aplicada para testar todas as restrições lineares. Como veremos ao discutir como testar restrições lineares gerais, a forma soma dos resíduos quadrados da estatística F é, às vezes, necessária.]

Como o R -quadrado é um resultado informado em quase todas as regressões (embora o SQR não seja), é fácil usar os R -quadrados dos modelos irrestrito e restrito para testar a exclusão de algumas variáveis. Atenção particular deve ser colocada à ordem dos R -quadrados do numerador: o R -quadrado *irrestrito* vem primeiro [compare com os SQRs em (4.37)]. Como $R_u^2 > R_r^2$, isso mostra novamente que F sempre será positivo.

Ao usar a forma R -quadrado para testar a exclusão de um conjunto de variáveis, é importante *não* elevar ao quadrado o R -quadrado antes de colocá-lo na fórmula (4.41), pois isso já foi feito. Todas as regressões informam o R^2 , e esses números são colocados diretamente em (4.41). No exemplo do salário dos jogadores de beisebol, podemos usar (4.41) para obter a estatística F :

$$F = \frac{(0,6278 - 0,5971)}{(1 - 0,6278)} \cdot \frac{347}{3} \approx 9,54,$$

que está muito próxima da que obtivemos anteriormente. (A diferença se deve a erros de arredondamento.)

EXEMPLO 4.9

(Educação dos Pais em uma Equação do Peso de Nascimentos)

Como outro exemplo de cálculo de uma estatística F , considere o seguinte modelo para explicar o peso de recém-nascidos em termos de vários fatores:

$$\text{pesonas} = \beta_0 + \beta_1 \text{cigs} + \beta_2 \text{ordnas} + \beta_3 \text{rendfam} + \beta_4 \text{educm} + \beta_5 \text{educp} + u, \quad (4.42)$$

em que pesonas é o peso de nascimento, em libras, cigs é o número médio de cigarros que a mãe fumou por dia durante a gravidez, ordnas é a ordem de nascimento dessa criança, rendfam é a renda familiar anual, educm corresponde aos anos de escolaridade formal da mãe e educp corresponde aos anos de escolaridade formal do pai. Vamos testar a hipótese nula de que, após controlar cigs , ordnas e rendfam , a educação dos pais não tem efeito sobre o peso de nascimento. Isso é expresso como $H_0: \beta_4 = 0, \beta_5 = 0$, e portanto há $q = 2$ restrições de exclusão para serem testadas. Há $k + 1 = 6$ parâmetros no modelo irrestrito (4.42), de modo que os g /do modelo irrestrito são $n - 6$, em que n é o tamanho da amostra.

Vamos testar essa hipótese usando os dados em BWGHT.RAW. Esse conjunto de dados contém informações de 1.388 nascimentos, mas devemos ser cuidadosos ao contar as observações usadas no teste da hipótese nula. Ocorre que em pelo menos uma das variáveis educm e educp estão faltando informações de 197 nascimentos na amostra; essas observações não podem ser incluídas ao estimar o modelo irrestrito. Assim, temos realmente $n = 1.191$ observações e, portanto, há $1.191 - 6 = 1.185$ g/ no modelo irrestrito. Devemos estar seguros de usar essas mesmas 1.191 observações quando estimarmos o modelo restrito (e não o total das 1.388 observações que estão disponíveis). Em geral, ao estimar o modelo restrito para calcular um teste F , devemos usar as mesmas observações para estimar o modelo irrestrito; de outro modo, o teste não é válido. Quando não faltarem dados, isso deixa de ser um problema.

EXEMPLO 4.9 (continuação)

Os g_f do numerador são iguais a 2, e os g_d do denominador, a 1.185; da Tabela G.3, o valor crítico a 5% é $c = 3,0$. Por brevidade, em vez de informar os resultados completos, vamos apresentar somente os R -quadradinhos. O R -quadrado do modelo completo é $R^2_{\text{tot}} = 0,0387$. Quando educm e educp são retirados da regressão, o R -quadrado cai para $R^2_r = 0,0364$. Assim, a estatística F é $F = [(0,0387 - 0,0364)/(1 - 0,0387)](1,185/2) = 1,42$; como esse valor está bem abaixo do valor crítico de 5%, não é possível rejeitar H_0 . Em outras palavras, educm e educp são conjuntamente não significantes na equação do peso de nascimento.

Cálculo dos p -Valores para Testes F

Para apresentar os resultados dos testes F , os p -valores são especialmente úteis. Como a distribuição F depende dos g_f do numerador e do denominador, é difícil obter uma impressão de quanto é forte ou fraca a evidência contra a hipótese nula simplesmente olhando para o valor da estatística F e um ou dois valores críticos.

No contexto do teste F , o p -valor é definido como

$$p\text{-valor} = P(\mathcal{F} > F), \quad (4.43)$$

em que, para enfatizar, \mathcal{F} representa uma variável aleatória F com $(q, n - k - 1)$ graus de liberdade, e F é o valor real da estatística do teste. O p -valor ainda tem a mesma interpretação que ele tinha para a estatística t : ele é a probabilidade de observarmos um valor de F pelo menos tão grande quanto aquele que encontramos, *dado* que a hipótese nula é verdadeira. Um p -valor pequeno é evidência contra H_0 . Por exemplo, o p -valor = 0,016 significa que a probabilidade de observarmos um valor de F tão grande quanto aquele para o qual a hipótese nula é verdadeira é somente 1,6%; em geral, rejeitamos H_0 em tais casos. Se o p -valor = 0,314, então a probabilidade de observarmos um valor da estatística F tão grande quanto aquele sob a hipótese nula é 31,4%. A maioria acharia esse valor uma evidência bastante fraca contra H_0 .

Os dados do arquivo ATTEND.RAW foram usados para estimar as duas equações

$$\begin{aligned} \text{taxafreq} &= 47,13 + 13,37 \text{ nmgradp} \\ &\quad (2,87) \quad (1,09) \\ n &= 680, R^2 = 0,183, \end{aligned}$$

e

$$\begin{aligned} \text{taxafreq} &= 75,70 + 17,26 \text{ nmgradp} - 1,72 \text{ tac} \\ &\quad (3,88) \quad (1,08) \quad (?) \\ n &= 680, R^2 = 0,291, \end{aligned}$$

em que, como sempre, os erros-padrão estão entre parênteses; o erro-padrão de tac está faltando na segunda equação. Qual é a estatística t do coeficiente de tac? (Sugestão: Primeiro calcule a estatística F da significância de tac.)

Assim como com o teste t , uma vez calculado o p -valor, o teste F pode ser realizado para qualquer nível de significância. Por exemplo, se o p -valor = 0,024, rejeitamos H_0 ao nível de significância de 5%, mas não ao nível de 1%.

O p -valor do teste F no Exemplo 4.9 é 0,238, e portanto a hipótese nula de β_{educm} e β_{educp} serem ambos zero não é rejeitada mesmo ao nível de significância de 20%.

Muitos programas econométricos têm recursos embutidos para testar restrições múltiplas de exclusão. Neles, o cálculo computacional tem várias vantagens sobre o cálculo manual das estatísticas: provavelmente cometemos menos erros, os p -valores são calculados automaticamente e o problema de falta de dados, como no Exemplo 4.9, é tratado sem qualquer trabalho adicional de nossa parte.

A Estatística F para a Significância Geral de uma Regressão

Um conjunto especial de restrições de exclusão é rotineiramente testado por muitos programas de regressão. Essas restrições têm a mesma interpretação, independentemente do modelo. No modelo com k variáveis independentes, podemos escrever a hipótese nula como

$$H_0: x_1, x_2, \dots, x_k \text{ não ajudam a explicar } y.$$

Essa hipótese nula é, de certa maneira, muito pessimista. Ela afirma que *nenhuma* das variáveis explicativas tem um efeito sobre y . Expressa em termos dos parâmetros, a hipótese nula é que todos os parâmetros de inclinação são zero:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0, \quad (4.44)$$

e a hipótese alternativa é que pelo menos um dos β_j seja diferente de zero. Outra maneira útil de formular a hipótese nula é que $H_0: E(y|x_1, x_2, \dots, x_k) = E(y)$, de modo que conhecer os valores de x_1, x_2, \dots, x_k não afeta o valor esperado de y .

Há k restrições em (4.44), e quando as impomos, obtemos o modelo restrito

$$y = \beta_0 + u; \quad (4.45)$$

todas as variáveis independentes foram retiradas da equação. Agora, o R -quadrado da estimativa de (4.45) é zero; nada da variação em y está sendo explicado porque não há variáveis explicativas. Portanto, a estatística F para testar (4.44) pode ser escrita como

$$\frac{R^2/k}{(1 - R^2)/(n - k - 1)}, \quad (4.46)$$

em que R^2 é exatamente o R -quadrado usual da regressão de y sobre x_1, x_2, \dots, x_k .

A maioria dos programas de regressão informa a estatística F em (4.46) automaticamente, o que torna tentador usar essa estatística para testar restrições de exclusão gerais. Você deve evitar essa tentação. A estatística F em (4.41) é usada para restrições de exclusão gerais; ela depende dos R -quadrados dos modelos restrito e irrestrito. A forma especial (4.46) é válida somente para testar a exclusão conjunta de *todas* as variáveis independentes. Às vezes, isso é chamado de teste de **significância geral da regressão**.

Se não for possível rejeitar (4.44), não há evidência de que qualquer uma das variáveis independentes ajude a explicar y . Isso usualmente significa que devemos procurar outras variáveis para explicar y . Para o Exemplo 4.9, a Estatística F para testar (4.44) é cerca de 9,55 com $k = 5$ e $n - k - 1 = 1.185$ gl. O p -valor é zero para quatro casas após o ponto decimal, de modo que (4.44) é fortemente rejeitada. Assim, concluímos que as variáveis na equação *pesonas* explicam, *de fato*, alguma variação em *pesonas*. A quantidade explicada não é grande: somente 3,87%. No entanto, o R-quadrado aparentemente pequeno resulta em uma estatística F altamente significante. Essa é a razão de devermos calcular a estatística F para testar a significância conjunta e não apenas olhar o tamanho do R-quadrado.

Ocasionalmente, a estatística F para a hipótese de que todas as variáveis independentes são conjuntamente não significantes pode ser o foco de um estudo. O Problema 4.10 pedirá a você para usar dados de retorno de ações para testar se os retornos das ações ao longo de um horizonte de quatro anos são previsíveis, com base em informações conhecidas somente no início do período. Sob a hipótese de mercados eficientes, os retornos não deveriam ser previsíveis; a hipótese nula é precisamente (4.44).

Teste de Restrições Lineares Gerais

Testar restrições de exclusão é, de longe, a mais importante aplicação da estatística F . Às vezes, entretanto, as restrições implicadas por uma teoria são mais complicadas do que apenas excluir algumas variáveis independentes. É ainda simples usar a estatística F para um teste dessa natureza.

Como um exemplo, considere a seguinte equação:

$$\begin{aligned} \log(\text{preço}) &= \beta_0 + \beta_1 \log(\text{aval}) + \beta_2 \log(\text{tamterr}) \\ &\quad + \beta_3 \log(\text{arquad}) + \beta_4 \text{qtdorm} + u, \end{aligned} \tag{4.47}$$

em que *preço* é o preço das casas, *aval* é o valor avaliado das casas (antes de elas serem vendidas), *tamterr* é o tamanho dos terrenos, em pés quadrados, *arquad* é a área da casa em pés quadrados e *qtdorm* é o número de quartos. Agora, suponha que gostaríamos de testar se o preço de avaliação das casas é uma avaliação racional. Nesse caso, uma variação de 1% em *aval* deve estar associada a uma variação de 1% em *preço*; isto é, $\beta_1 = 1$. Além disso, *tamterr*, *arquad* e *qtdorm* não devem ajudar a explicar $\log(\text{preço})$, uma vez que o valor de avaliação tenha sido controlado. Juntas, essas hipóteses podem ser expressas como

$$H_0: \beta_1 = 1, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0. \tag{4.48}$$

Aqui há quatro restrições a serem testadas; três são restrições de exclusão, mas $\beta_1 = 1$ não é. Como podemos testar essa hipótese usando a estatística F ?

Como no caso da restrição de exclusão, estimamos o modelo irrestrito, (4.47) nesse caso, e, em seguida, impomos as restrições em (4.48) para obter o modelo restrito. O segundo passo pode ser um pouquinho complicado. Porém, tudo o que fazemos é inserir as restrições. Se escrevermos (4.47) como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + u, \tag{4.49}$$

então o modelo restrito é $y = \beta_0 + x_1 + u$. Agora, a fim de impor a restrição de que o coeficiente de x_1 é a unidade, devemos estimar o seguinte modelo:

$$y - x_1 = \beta_0 + u. \quad (4.50)$$

Esse é apenas um modelo com um intercepto (β_0), mas com uma variável dependente diferente daquela em (4.49). O procedimento para calcular a estatística F é o mesmo: estime (4.50), obtenha o SQR (SQR_r) e utilize-o com o SQR de (4.49) na estatística F (4.37). Estamos testando $q = 4$ restrições, e há $n - 5$ gl no modelo irrestrito. A estatística F é simplesmente $[(SQR_r - SQR_{ir})/SQR_{ir}][(n - 5)/4]$.

Antes de ilustrar esse teste usando um conjunto de dados devemos enfatizar um ponto: não podemos usar a forma R -quadrado da estatística F nesse exemplo porque a variável dependente em (4.50) é diferente daquela em (4.49). Isso significa que a soma dos quadrados total das duas regressões será diferente, e (4.41) não é mais equivalente a (4.37). Como regra geral, a forma SQR da estatística F deve ser usada se uma variável dependente diferente for necessária para computar a regressão restrita.

O modelo irrestrito estimado usando os dados em HPRICE1.RAW é

$$\begin{aligned} \log(\hat{p}\text{reço}) &= 0,264 + 1,043 \log(aval) + 0,0074 \log(tamterr) \\ (0,570) &\quad (0,151) \quad (0,0386) \\ &- 0,1032 \log(arquad) + 0,0338 qtdorm \\ &\quad (0,1384) \quad (0,0221) \\ n &= 88, SQR = 1,822, R^2 = 0,773. \end{aligned}$$

Se usarmos separadamente as estatísticas t para testar cada hipótese em (4.48), não será possível rejeitar cada uma delas. Contudo, a racionalidade da avaliação é uma hipótese conjunta, de modo que devemos testar as restrições conjuntamente. O SQR do modelo restrito é $SQR_r = 1,880$ e, portanto, a estatística F é $[(1,880 - 1,822)/1,822](83/4) = 0,661$. O valor crítico de 5% em uma distribuição F com (4,83) gl é cerca de 2,50, e portanto não podemos rejeitar H_0 . Não há, essencialmente, evidência contra a hipótese de que os valores de avaliação sejam racionais.

4.6 DESCRIÇÃO DOS RESULTADOS DA REGRESSÃO

Finalizamos este capítulo dando algumas instruções de como descrever os resultados da regressão múltipla para projetos empíricos relativamente complicados. Isso deve ensiná-lo a ler trabalhos publicados nas ciências sociais aplicadas, ao mesmo tempo em que prepara você para escrever seus próprios artigos empíricos. Expandiremos este tópico no restante do livro ao descrever os resultados de vários exemplos, mas muitos dos pontos fundamentais podem ser apresentados agora.

Naturalmente, os coeficientes estimados de MQO devem ser sempre informados. Das variáveis fundamentais de uma análise, você deve *interpretar* os coeficientes estimados (o que, freqüentemente, requer conhecer as unidades de medida das variáveis). Por exemplo, ele é uma elasticidade ou tem alguma outra interpretação que necessita de explicação? A importância econômica ou prática das estimativas das variáveis-chave devem ser discutidas.

Os erros-padrão devem sempre ser incluídos juntamente com os coeficientes estimados. Alguns autores preferem informar as estatísticas t em vez dos erros-padrão (e, freqüentemente, apenas o valor absoluto das estatísticas t). Embora não haja, realmente, nada de errado com isso, há alguma preferê-

cia por informar os erros-padrão. Primeiro, isso nos força a pensar, cuidadosamente, sobre a hipótese nula que está sendo testada; a hipótese nula nem sempre corresponde a dizer que o parâmetro populacional é zero. Segundo, ter os erros-padrão torna mais fácil calcular os intervalos de confiança.

O R -quadrado da regressão deve sempre ser incluído. Vimos que, além de dar uma medida do grau de ajuste, ele faz com que os cálculos das estatísticas F para as restrições de exclusão fiquem simples. Informar a soma dos resíduos quadrados e o erro-padrão da regressão às vezes é uma boa idéia, mas não é crucial. O número de observações usado na estimativa de qualquer equação deve aparecer próximo da equação estimada.

Se somente alguns poucos modelos são estimados, os resultados podem ser resumidos na forma de equações, como fizemos até aqui. Entretanto, em muitos trabalhos, várias equações são estimadas para muitos conjuntos diferentes de variáveis independentes. Podemos estimar a mesma equação para diferentes grupos de pessoas, ou mesmo ter equações que explicam diferentes variáveis dependentes. Em tais casos, é melhor resumir os resultados em uma ou mais tabelas. A variável dependente deve ser indicada claramente na tabela, e as variáveis independentes, listadas na primeira coluna. Os erros-padrão (ou as estatísticas t) podem ser colocados em parênteses abaixo das estimativas.

EXEMPLO 4.10

(A Relação Salário-Benefícios de Professores)

Façamos $totrem$ representar a remuneração média anual total de um professor, incluindo o salário e todos os benefícios adicionais (pensão, seguro-saúde etc.). Ampliando a equação dos salários, a remuneração total deve ser uma função da produtividade e talvez de outras características. Como é padrão, vamos usar a forma logarítmica:

$$\log(totrem) = f(\text{características da produtividade, outros fatores}),$$

em que $f(\cdot)$ é alguma função (não-especificada por enquanto). Escreva

$$totrem = \text{salário} + \text{benefícios} = \text{salário} \left(1 + \frac{\text{benefícios}}{\text{salário}} \right).$$

Essa equação mostra que a remuneração total é o produto de dois termos: salário e $1 + b/s$, em que b/s é a abreviação para "razão benefícios-salário". Tirando o log dessa equação resulta em $\log(totrem) = \log(\text{salário}) + \log(1 + b/s)$. Agora, para um b/s "pequeno", $\log(1 + b/s) \approx b/s$; vamos usar essa aproximação. Isso leva ao modelo econométrico

$$\log(\text{salário}) = \beta_0 + \beta_1(b/s) + \text{outros fatores}.$$

Testar a relação salário-benefícios é, então, o mesmo que testar $H_0: \beta_1 = -1$ contra $H_1: \beta_1 \neq -1$.

Vamos usar os dados do arquivo MEAP93.RAW para testar essa hipótese. Esses dados são ponderados por escola, e não observamos muitos outros fatores que poderiam afetar a remuneração total. Incluiremos os controles para o tamanho da escola (*matricl*), número de funcionários por mil estudantes e medidas como taxas de evasão escolar (*taxevas*) e de formatura (*taxform*). O b/s médio na amostra é cerca de 0,205, e o maior valor é 0,450.

As equações estimadas são apresentadas na Tabela 4.1, na qual os erros-padrão aparecem entre parênteses, abaixo das estimativas dos coeficientes. A variável-chave é b/s , a razão benefícios-salário.

EXEMPLO 4.10 (continuação)

Na primeira coluna da Tabela 4.1, vemos que, sem controlar quaisquer outros fatores, o coeficiente de MQO de b/s é $-0,825$. A estatística t para testar a hipótese nula $H_0: \beta_1 = -1$ é $t = (-0,825 + 1)/0,200 = 0,875$, e portanto a regressão simples não permite rejeitar H_0 . Após adicionar controles para o tamanho da escola e o tamanho do corpo docente (o qual captura, mais ou menos, o número de estudantes por professor), a estimativa do coeficiente de b/s passa a ser $-0,605$. Agora, o teste de $\beta_1 = -1$ resulta em uma estatística t igual a cerca de $2,39$; assim, H_0 é rejeitada ao nível de 5% contra uma alternativa bilateral. As variáveis $\log(matrix)$ e $\log(staff)$ são estatisticamente muito significantes.

De que modo o acréscimo de $taxevas$ e $taxform$ afeta a estimativa da relação salário-benefícios? Essas variáveis são conjuntamente significantes ao nível de 5% ? E ao nível de 10% ?

Tabela 4.1

Teste da Relação Salário-Benefícios

Variável Dependente: $\log(salário)$			
Variáveis Independentes	(1)	(2)	(3)
b/s	-0,825 (0,200)	-0,605 (0,165)	-0,589 (0,165)
$\log(matrix)$	—	0,874 (0,0073)	0,0881 (0,0073)
$\log(staff)$	—	-0,222 (0,050)	-0,218 (0,050)
$taxevas$	—	—	-0,00028 (0,00161)
$taxfor$	—	—	0,00097 (0,00066)
<i>intercepto</i>	10,523 (0,042)	10,884 (0,252)	10,738 (0,258)
Observações	408	408	408
R -quadrado	0,040	0,353	0,361

Neste capítulo, cobrimos o tópico muito importante da inferência estatística, o qual nos permite obter conclusões sobre o modelo populacional a partir de uma amostra aleatória. Vamos resumir os pontos principais:

1. Sob as hipóteses do modelo linear clássico RLM.1 a RLM.6, os estimadores de MQO são normalmente distribuídos.
2. Sob as hipóteses do MLC, as estatísticas t têm distribuições t sob a hipótese nula.
3. Usamos as estatísticas t para testar hipóteses sobre um único parâmetro contra alternativas unilaterais ou bilaterais, usando testes monocaudais ou bicaudais, respectivamente. A hipótese nula mais comum é $H_0: \beta_j = 0$, mas, às vezes, queremos testar outros valores de β_j sob H_0 .
4. No teste de hipótese clássico, primeiro escolhemos um nível de significância que, juntamente com os gl e a hipótese alternativa, determina o valor crítico contra o qual compararmos a estatística t . É mais informativo calcular o p -valor de um teste t – o nível de significância menor ao qual a hipótese nula é rejeitada –, de modo que a hipótese pode ser testada a qualquer nível de significância.
5. Sob as hipóteses do MLC, os intervalos de confiança podem ser construídos para cada β_j . Esses ICs podem ser usados para testar qualquer hipótese nula relativa a β_j contra uma alternativa bilateral.
6. Testes de hipóteses simples relativos a mais de um β_j podem sempre ser testados, reescrevendo o modelo de tal forma que ele contenha o parâmetro de interesse. Em seguida, uma estatística t padrão pode ser usada.
7. A estatística F é usada para testar restrições múltiplas de exclusão, e há duas formas equivalentes do teste. Uma está baseada nos SQRs dos modelos restrito e irrestrito. Uma forma mais conveniente está baseada nos R -quadrados dos dois modelos.
8. Ao calcular uma estatística F , os gl do numerador correspondem ao número de restrições que estão sendo testadas, enquanto os gl do denominador são os graus de liberdade do modelo irrestrito.
9. A hipótese alternativa do teste F é bilateral. Na abordagem clássica especificamos um nível de significância que, juntamente com o gl do numerador e o gl do denominador, determina o valor crítico. A hipótese nula é rejeitada quando a estatística F excede o valor crítico c . Alternativamente, podemos calcular o p -valor para resumir a evidência contra H_0 .
10. Restrições lineares múltiplas gerais podem ser testadas usando a forma soma dos resíduos quadrados da estatística F .
11. A estatística F da significância geral de uma regressão testa a hipótese nula de que *todos* os parâmetros de inclinação são zero, com o intercepto irrestrito. Sob H_0 , as variáveis explicativas não têm efeito sobre o valor esperado de y .

4.1 Quais dos seguintes itens podem fazer com que as estatísticas de MQO não sejam válidas (isto é, que elas não tenham distribuições t sob H_0)?

- (i) Heteroscedasticidade.
- (ii) Um coeficiente de correlação de 0,95 entre duas variáveis independentes que estão no modelo.
- (iii) Omitir uma variável explicativa importante.

4.2 Considere uma equação para explicar os salários dos diretores executivos em termos das vendas anuais das empresas (*vendas*), dos retornos das ações sobre o patrimônio (*rma*, na forma percentual) e dos retornos das ações sobre o valor das ações das empresas (*raf*, na forma percentual):

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

- (i) Em termos dos parâmetros do modelo, formule a hipótese nula em que, após controlar *vendas* e *rma*, *raf* não tem efeito sobre o salário dos diretores executivos. Formule a hipótese alternativa de que um melhor desempenho de mercado das ações aumenta o salário dos diretores executivos.
- (ii) Usando os dados em CEOSAL1.RAW, obteve-se a seguinte equação por MQO:

$$\begin{aligned}\hat{\log}(\text{salário}) &= 4,32 + 0,280 \log(\text{vendas}) + 0,0174 rma + 0,00024 raf \\ &\quad (0,32) \quad (0,035) \quad (0,0041) \quad (0,00054) \\ &n = 209, R^2 = 0,283.\end{aligned}$$

Se *raf* aumenta em 50 pontos, qual é a variação percentual prevista em *salário*? Na prática, *raf* tem um efeito grande sobre *salário*?

- (iii) Teste a hipótese nula de que *raf* não tem efeito sobre *salário* contra a alternativa de que *raf* tem um efeito positivo. Faça o teste ao nível de significância de 10%.
- (iv) Você incluiria *raf* no modelo final que explica a remuneração dos diretores executivos em termos do desempenho das empresas? Explique.

4.3 A variável *pdintens* corresponde a gastos com pesquisa e desenvolvimento (P&D) como uma percentagem das vendas. As vendas são mensuradas em milhões de dólares. A variável *lucrmarg* corresponde a lucros como uma percentagem das vendas.

Usando os dados do arquivo RDCHEM.RAW de 32 empresas da indústria química, estimou-se a seguinte equação:

$$\begin{aligned}pdintens &= 0,472 + 0,321 \log(\text{vendas}) + 0,050 lucrmarg \\ &\quad (1,369) \quad (0,216) \quad (0,046) \\ &n = 32, R^2 = 0,099.\end{aligned}$$

- (i) Interprete o coeficiente de $\log(\text{vendas})$. Em particular, se *vendas* aumenta em 10%, qual é a variação percentual estimada em *pdintens*? Esse efeito é economicamente grande?
- (ii) Teste a hipótese de que a intensidade de P&D não varia com *vendas* contra a alternativa de que P&D aumenta com as vendas. Teste aos níveis de 5% e 10%.
- (iii) *lucrmarg* tem um efeito estatisticamente significante sobre *pdintens*?

4.4 As taxas de aluguel são influenciadas pela população de estudantes em uma cidade onde há universidades? Seja *alug* o aluguel médio mensal pago pela unidade alugada em uma cidade nos Estados Unidos, onde há universidades. Seja *pop* o total da população da cidade, *rendmed*, a renda média da cidade e *pctestu*, a população de estudantes como um percentual da população total. Um modelo para testar uma relação é

$$\log(\text{alug}) = \beta_0 + \beta_1 \log(\text{pop}) + \beta_2 \log(\text{rendmed}) + \beta_3 \text{pctestu} + u.$$

- (i) Formule a hipótese nula de que o tamanho da população estudantil relativo à população das cidades não tem efeito *ceteris paribus* sobre os aluguéis mensais. Formule a alternativa de que há um efeito.
- (ii) Quais sinais você espera para β_1 e β_2 ?
- (iii) A equação estimada, usando 1.990 dados de 64 cidades com universidades do arquivo RENTAL.RAW, é

$$\hat{\log}(alug) = 0,043 + 0,066 \log(pop) + 0,507 \log(rendmed) + 0,0056 pctestu$$

(0,844)	(0,039)	(0,081)	(0,0017)
$n = 64, R^2 = 0,458.$			

O que está errado com a seguinte afirmação: “Um aumento de 10% na população está associado a um aumento de cerca de 6,6% no aluguel”?

- (iv) Teste a hipótese formulada na parte (i) ao nível de 1%.

4.5 Considere a equação estimada do Exemplo 4.3 (que poderia também ser usada para estudar os efeitos de faltar às aulas sobre a nota média em curso superior):

$$nmgrad = 1,39 + 0,412 nmemb + 0,15 tac - 0,083 faltas$$

(0,33)	(0,94)	(0,011)	(0,026)
$n = 141, R^2 = 0,234.$			

- (i) Usando a aproximação normal padronizada, encontre o intervalo de confiança de 95% para β_{nmemb} .
- (ii) Você pode rejeitar a hipótese $H_0: \beta_{nmemb} = 0,4$ contra a hipótese alternativa bilateral ao nível de 5%?
- (iii) Você pode rejeitar a hipótese $H_0: \beta_{nmemb} = 1$ contra a hipótese alternativa bilateral ao nível de 5%?

4.6 Na Seção 4.5 usamos, como exemplo, o teste da razionalidade da avaliação dos preços de casas. Lá, usamos um modelo log-log em *preço* e *aval* [veja equação (4.47)]. Aqui, vamos usar uma formulação nível-nível.

- (i) No modelo de regressão simples

$$preço = \beta_0 + \beta_1 aval + u,$$

a avaliação é racional se $\beta_1 = 1$ e $\beta_0 = 0$. A equação estimada é

$$\hat{preço} = -14,47 + 0,976 aval$$

(16,27)	(0,049)
$n = 88, SQR = 165,644.51, R^2 = 0,820.$	

Primeiro, teste a hipótese $H_0: \beta_0 = 0$ contra a hipótese alternativa bilateral. Em seguida, teste $H_0: \beta_1 = 1$ contra a hipótese alternativa bilateral. O que você conclui?

- (ii) Para testar a hipótese conjunta $\beta_0 = 0$ e $\beta_1 = 1$, precisamos do SQR do modelo restrito. Isso é igual a calcular $\sum_{i=1}^n (\hat{preço}_i - \hat{aval}_i)^2$, em que $n = 88$, visto que os resíduos do modelo

restrito são exatamente $\hat{preço}_i = \hat{aval}_i$. (Nenhuma estimação é necessária para o modelo restrito porque ambos os parâmetros estão especificados sob H_0 .) Isso tem como resultado SQR = 209.448,99. Faça o teste F para a hipótese conjunta.

- (iii) Agora, teste $H_0: \beta_2 = 0, \beta_3 = 0$ e $\beta_4 = 0$ no modelo

$$\hat{preço} = \beta_0 + \beta_1 \hat{aval} + \beta_2 \hat{tamterr} + \beta_3 \hat{arquad} + \beta_4 \hat{qtdorm} + u.$$

O R -quadrado da estimativa desse modelo usando as mesmas 88 residências é 0,829.

- (iv) Se a variância de $\hat{preço}$ varia com $\hat{aval}, \hat{tamterr}, \hat{arquad}$ ou \hat{qtdorm} , o que você pode dizer sobre o teste F da parte (iii)?

4.7 No Exemplo 4.7, usamos dados de empresas manufatureiras de Michigan para estimar a relação entre a taxa de rejeição e outras características da firma. Agora, vamos olhar esse exemplo mais de perto e usar uma amostra maior de empresas.

- (i) O modelo populacional estimado no Exemplo 4.7 pode ser escrito como

$$\log(\hat{rejei}) = \beta_0 + \beta_1 \log(hrsemp) + \beta_2 \log(vendas) + \beta_3 \log(empreg) + u.$$

Usando as 43 observações disponíveis para 1987, a equação estimada é

$$\begin{aligned} \log(\hat{rejei}) &= 11,74 - 0,042 \log(hrsemp) - 0,951 \log(vendas) + 0,992 \log(empreg) \\ (4,57) &\quad (0,019) \quad (0,370) \quad (0,360) \\ n &= 43, R^2 = 0,310. \end{aligned}$$

Compare essa equação com aquela estimada com somente 30 firmas na amostra.

- (ii) Mostre que o modelo populacional pode também ser escrito como

$$\log(\hat{rejei}) = \beta_0 + \beta_1 \log(hrsemp) + \beta_2 \log(vendas/empreg) + \theta_3 \log(empreg) + u.$$

em que $\theta_3 = \beta_2 + \beta_3$. [Sugestão: Lembre-se de que $\log(x_2/x_3) = \log(x_2) - \log(x_3)$.]

Interprete a hipótese $H_0: \theta_3 = 0$.

- (iii) Quando a equação da parte (ii) é estimada, obtemos

$$\begin{aligned} \log(\hat{rejei}) &= 11,74 - 0,042 \log(hrsemp) - 0,951 \log(vendas/empreg) + 0,041 \log(empreg) \\ (4,57) &\quad (0,019) \quad (0,370) \quad (0,205) \\ n &= 43, R^2 = 0,310. \end{aligned}$$

Controlando o treinamento dos trabalhadores e a razão vendas-empregados, as empresas maiores têm taxas de rejeição maiores estatisticamente significantes?

- (iv) Teste a hipótese de que um aumento de 1% em $vendas/empreg$ está associado a uma queda de 1% na taxa de rejeição.

4.8 Considere o modelo de regressão múltipla com três variáveis independentes, sob as hipóteses do modelo linear clássico RLM.1 a RLM.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

Você deseja testar a hipótese nula $H_0: \beta_1 - 3\beta_2 = 1$.

- (i) Sejam $\hat{\beta}_1$ e $\hat{\beta}_2$ os estimadores de MQO de β_1 e β_2 . Encontre $\text{Var}(\hat{\beta}_1 - 3\hat{\beta}_2)$ em termos das variâncias de $\hat{\beta}_1$ e $\hat{\beta}_2$ e a covariância entre eles. Qual é o erro-padrão de $\hat{\beta}_1 - 3\hat{\beta}_2$?
- (ii) Escreva a estatística t para testar $H_0: \beta_1 - 3\beta_2 = 1$.
- (iii) Defina $\theta_1 = \beta_1 - 3\beta_2$ e $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$. Escreva uma equação de regressão que envolva β_0 , θ_1 , β_2 e β_3 , que permita que você obtenha diretamente $\hat{\theta}_1$ e seu erro-padrão.

4.9 No Problema 3.3, estimamos a equação

$$\begin{aligned} \hat{do\text{\'o}mir} &= 3.638,25 - 0,148 \text{ trabtot} - 11,13 \text{ educ} + 2,20 \text{ idade} \\ &\quad (112,28) \quad (0,017) \quad (5,88) \quad (1,45) \\ &n = 706, R^2 = 0,113, \end{aligned}$$

para a qual informamos, agora, os erros-padrão juntamente com as estimativas.

- (i) *educ* ou *idade* são individualmente significantes ao nível de 5% contra uma hipótese alternativa bilateral? Mostre como você chegou à resposta.
- (ii) Ao retirar *educ* e *idade* da equação, temos

$$\begin{aligned} \hat{do\text{\'o}mir} &= 3.586,25 - 0,151 \text{ trabtot} \\ &\quad (38,91) \quad (0,017) \\ &n = 706, R^2 = 0,103, \end{aligned}$$

É possível afirmar que *educ* e *idade* são conjuntamente significantes na equação original ao nível de 5%? Justifique sua resposta.

- (iii) Incluir *educ* e *idade* no modelo afeta muito a relação estimada entre dormir e trabalhar?
- (iv) Suponha que a equação de dormir contenha heteroscedasticidade. O que isso significa para os testes calculados nas partes (i) e (ii)?

4.10 A análise de regressão pode ser usada para testar se o mercado usa eficientemente as informações ao avaliar ações. Seja *retorno* o retorno total de manter ações de uma firma ao longo de um período de quatro anos, do final de 1990 até o final de 1994. A *hipótese de mercados eficientes* diz que esses retornos não devem estar sistematicamente relacionados à informação conhecida em 1990. Se as características conhecidas no início do período ajudarem a prever os retornos das ações, poderíamos usar essas informações para escolher as ações.

Para 1990, seja *rdc* a relação dívida-capital de uma empresa, seja *gpa* os ganhos por ação, seja *rendliq* a renda líquida e seja *salário* a remuneração total dos diretores executivos da empresa.

- (i) Usando os dados do arquivo RETURN.RAW, estimou-se a seguinte equação:

$$\begin{aligned} \hat{retorno} &= -14,37 + 0,321 \text{ rdc} + 0,043 \text{ gpa} - 0,0051 \text{ rendliq} + 0,0035 \text{ salário} \\ &\quad (6,89) \quad (0,201) \quad (0,078) \quad (0,0047) \quad (0,0022) \\ &n = 142, R^2 = 0,0395. \end{aligned}$$

Teste se as variáveis explicativas são conjuntamente significantes ao nível de 5%? Alguma variável explicativa é individualmente significante?

- (ii) Agora, nova estimativa do modelo usando a forma log para *rendliq* e *salário* forneceu a seguinte equação:

$$\hat{retorno} = -36,30 + 0,327 rdc + 0,069 gpa - 4,74 \log(rendliq) + 7,24 \log(salário)$$

(39,37)	(0,203)	(0,080)	(3,39)	(6,31)
---------	---------	---------	--------	--------

$$n = 142, R^2 = 0,0330.$$

Alguma de suas conclusões da parte (i) mudou?

- (iii) Por que não usamos também os logs de *rdc* e *gpa* na parte (ii)?
(iv) Em geral, a evidência da previsibilidade dos retornos é forte ou fraca?

4.11 A tabela seguinte foi criada ao usar os dados do arquivo CEOSAL2.RAW:

Variável Dependente: <i>log(salário)</i>			
Variáveis Independentes	(1)	(2)	(3)
<i>log(vendas)</i>	0,224 (0,027)	0,158 (0,040)	0,188 (0,040)
<i>log(valmerc)</i>	—	0,112 (0,050)	0,100 (0,049)
<i>lucrmarg</i>	—	-0,0023 (0,0022)	-0,0022 (0,0021)
<i>permceo</i>	—	—	0,0171 (0,0055)
<i>percomp</i>	—	—	-0,0092 (0,0033)
<i>intercepto</i>	4,94 (0,20)	4,62 (0,25)	4,57 (0,25)
Observações	177	177	177
<i>R</i> -quadrado	0,281	0,304	0,353

A variável *valmerc* é o valor de mercado da firma, *lucrmarg* é o lucro como percentagem das vendas, *permceo* corresponde aos anos trabalhando como diretor executivo na atual companhia e *percomp* é o total de anos na companhia.

- (i) Comente os efeitos de *lucrmarg* sobre o salário dos diretores executivos.
(ii) O valor de mercado tem um efeito significante? Explique.
(iii) Interprete os coeficientes de *permceo* e *percomp*. As variáveis são estatisticamente significantes?
(iv) O que você entende do fato de que a permanência muito longa na companhia, mantendo fixos os outros fatores, está associada a salários mais baixos?

Análise de Regressão Múltipla: MQO Assimptótico

os capítulos 3 e 4 estudamos o que chamamos propriedades de *amostra finita*, de *amostra pequena* ou *exatas* dos estimadores de MQO do modelo populacional

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u . \quad (5.1)$$

Por exemplo, a inexistência de viés de MQO (derivada no Capítulo 3), sob as quatro primeiras hipóteses de Gauss-Markov, é uma propriedade de amostra finita porque ela é válida para *qualquer* amostra de tamanho n (sujeita à restrição amena de que n deve ser pelo menos tão grande quanto o número total de parâmetros no modelo de regressão, $k + 1$). Semelhantemente, o fato de que MQO é o melhor estimador não-viesado linear sob o conjunto completo das hipóteses de Gauss-Markov (RLM.1 a RLM.6) é uma propriedade de amostra finita.

No Capítulo 4 acrescentamos a hipótese do modelo linear clássico RLM.6, a qual afirma que o termo erro u é normalmente distribuído e independente das variáveis explicativas. Isso nos permitiu derivar as distribuições amostrais *exatas* dos estimadores de MQO (condicionados às variáveis explicativas da amostra). Em particular, o Teorema 4.1 mostrou que os estimadores de MQO têm distribuições amostrais normais, o que levou diretamente às distribuições t e F das estatísticas t e F . Se o erro não é normalmente distribuído, a distribuição de uma estatística t não é exatamente t , e uma estatística F não tem uma distribuição F exata para qualquer tamanho de amostra.

Além das propriedades de amostra finita é importante conhecer as **propriedades assimptóticas** ou **propriedades de amostras grandes** dos estimadores e das estatísticas de testes. Essas propriedades não são definidas para um tamanho particular de amostra; pelo contrário, elas são definidas quando o tamanho da amostra cresce sem limites. Felizmente, sob as hipóteses que fizemos, o método MQO tem propriedades de amostra grande satisfatórias. Uma constatação importante na prática é que mesmo sem a hipótese de normalidade (hipótese RLM.6), as estatísticas t e F têm distribuições *aproximadamente* t e F , pelo menos em amostras grandes. Vamos discutir esse assunto com mais detalhes na Seção 5.2, após compreendermos a consistência do método MQO na Seção 5.1.

5.1 CONSISTÊNCIA

A inexistência de viés dos estimadores, embora importante, não pode ser conseguida sempre. Por exemplo, como discutimos no Capítulo 3, o erro-padrão da regressão, $\hat{\sigma}$, não é um estimador não-viesado de σ , o desvio-padrão do erro u em um modelo de regressão múltipla.

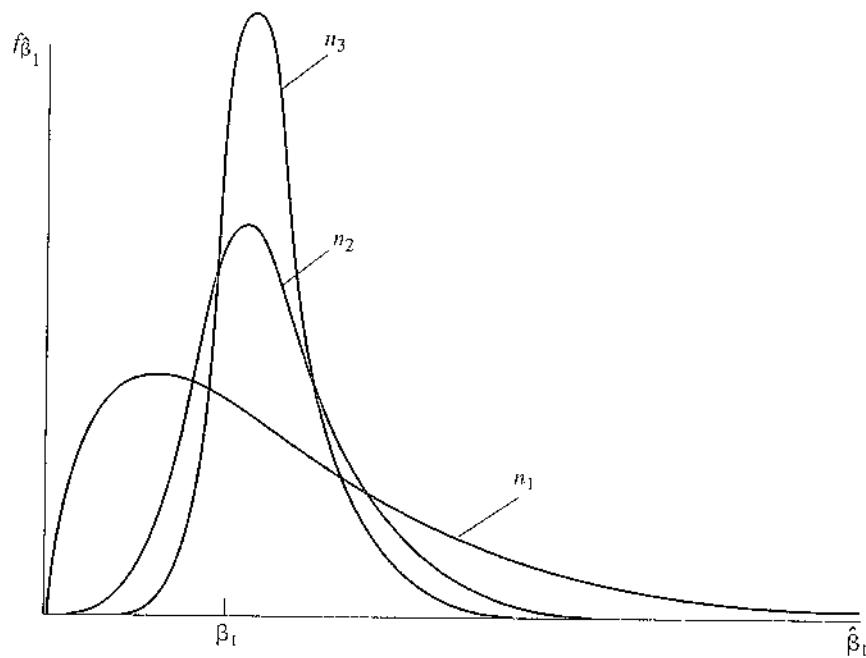
Embora os estimadores de MQO sejam não-viesados sob RLM.1 a RLM.4, descobriremos, no Capítulo 11, que há regressões de séries de tempo em que os estimadores de MQO não são não-viesados. Além disso, na Parte 3 do livro, encontraremos vários outros estimadores que são viesados.

Embora nem todos os estimadores úteis sejam não-viesados, virtualmente todos os economistas concordam que a **consistência** é um requisito mínimo de um estimador. O famoso econometrista Clive W. J. Granger observou certa vez: “Se você não puder obter a consistência apropriadamente quando n tende ao infinito, você não deveria se envolver com isso”. A implicação é que, quando seu estimador de um parâmetro populacional particular não for consistente, você estará desperdiçando seu tempo.

Há maneiras um pouco diferentes de descrever a consistência. As definições e os resultados formais estão apresentados no Apêndice C, disponível em www.thomsonlearning.com.br, aqui vamos dar ênfase a um entendimento intuitivo. Mais concretamente, seja $\hat{\beta}_j$ o estimador de MQO de β_j para algum j . Para cada n , $\hat{\beta}_j$ tem uma distribuição de probabilidades (representando seus valores possíveis em diferentes amostras aleatórias de tamanho n). Como $\hat{\beta}_j$ é não-viesado sob as hipóteses RLM.1 a RLM.4, essa distribuição tem valor médio β_j . Se esse estimador for consistente, a distribuição de $\hat{\beta}_j$ se torna mais e mais estreitamente distribuída ao redor de β_j quando o tamanho da amostra cresce. Quando n tende ao infinito, a distribuição de $\hat{\beta}_j$ encontra-se no ponto único β_j . De fato, isso significa que podemos fazer com que nosso estimador, arbitrariamente, aproxime-se de β_j se pudermos coletar tantos dados quanto desejarmos. Essa convergência está ilustrada na Figura 5.1.

Figura 5.1

Distribuições amostrais de $\hat{\beta}_1$ para amostras de tamanhos $n_1 < n_2 < n_3$.



Naturalmente, para qualquer aplicação, temos um tamanho de amostra fixo, que é a razão pela qual uma propriedade assintótica tal como a consistência pode ser difícil de entender. **Consistência** envolve um experimento mental sobre o que aconteceria se o tamanho da amostra se tornasse grande (enquanto, ao mesmo tempo, obtemos muitas amostras aleatórias para cada tamanho de amostra). Se a obtenção de mais e mais dados não nos levar, em geral, para perto do valor do parâmetro de interesse, estamos usando um procedimento de estimação insatisfatório.

Convenientemente, o mesmo conjunto de hipóteses implica tanto a inexistência de viés como a consistência de MQO. Vamos resumir o assunto com um teorema.

TEOREMA 5.1 (CONSISTÊNCIA DE MQO)

Sob as hipóteses RLM.1 a RLM.4, o estimador de MQO $\hat{\beta}_j$ é um estimador consistente de β_j , para todo $j = 0, 1, \dots, k$.

Uma prova geral desse resultado é mais facilmente desenvolvida usando os métodos da álgebra matricial descritos nos Apêndices D e E, disponíveis na página do livro, no site www.thomsonlearning.com.br. No entanto, podemos provar o Teorema 5.1 sem dificuldades no caso do modelo de regressão simples. Vamos nos concentrar no estimador de inclinação, $\hat{\beta}_1$.

A prova começa do mesmo modo que a prova da inexistência de viés: escrevemos a fórmula de $\hat{\beta}_1$, e em seguida a inserimos em $y_i = \beta_0 + \beta_1 x_{i1} + u_i$:

$$\begin{aligned}\hat{\beta}_1 &= \left(\sum_{i=1}^n (x_{i1} - \bar{x}_1) y_i \right) / \left(\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 \right) \\ &= \beta_1 + \left(n^{-1} \sum_{i=1}^n (x_{i1} - \bar{x}_1) u_i \right) / \left(n^{-1} \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 \right).\end{aligned}\tag{5.2}$$

Podemos aplicar a lei dos grandes números ao numerador e ao denominador, os quais convergem em probabilidade para as quantidades populacionais, $\text{Cov}(x_1, u)$ e $\text{Var}(x_1)$, respectivamente. Como $\text{Var}(x_1) \neq 0$ — o que é assumido em RLM.4 — podemos usar as propriedades dos *limites de probabilidade* (veja o Apêndice C, disponível no site da Thomson) para obter

$$\begin{aligned}\text{plim } \hat{\beta}_1 &= \beta_1 + \text{Cov}(x_1, u) / \text{Var}(x_1) \\ &= \beta_1, \text{ porque } \text{Cov}(x_1, u) = 0.\end{aligned}\tag{5.3}$$

Usamos o fato, discutido nos capítulos 2 e 3, de que $E(u|1) = 0$ implica que x_1 e u são não-correlacionados (têm covariância zero).

Como questão técnica, para garantir que os limites de probabilidade existam, devemos assumir que $\text{Var}(x_1) < \infty$ e $\text{Var}(u) < \infty$ (o que significa que suas distribuições de probabilidade não são muito espalhadas), mas não nos preocuparemos com casos em que essas hipóteses não se mantenham.

Os argumentos anteriores, e a equação (5.3) em particular, mostram que MQO é consistente no caso da regressão simples se assumirmos somente correlação zero. Isso também é verdadeiro no caso geral. Vamos agora formular isso como uma hipótese.

H I P Ó T E S E R L M . 3 ' (MÉDIA ZERO E CORRELAÇÃO ZERO)

$E(u) = 0$ e $\text{Cov}(x_j, u) = 0$, para $j = 1, 2, \dots, k$.

No Capítulo 3 discutimos por que a hipótese RLM.3 implica RLM.3', mas não vice-versa. O fato de que MQO é consistente sob a hipótese mais fraca RLM.3' se revelará útil no Capítulo 15 e em outras situações. É interessante observar que, embora MQO seja não-viesado sob RLM.3, esse não é o caso sob a hipótese RLM.3'. (Essa é a razão primeira de termos assumido RLM.3.)

A Derivação da Inconsistência no Método MQO

Do mesmo modo que a não-observância de $E(u|x_1, \dots, x_k) = 0$ causa viés dos estimadores de MQO, a correlação entre u e *qualquer* das variáveis x_1, x_2, \dots, x_k faz com que, em geral, *todos* os estimadores de MQO sejam inconsistentes. Essa simples, mas importante, observação é freqüentemente resumida como: *se o erro é correlacionado com qualquer uma das variáveis independentes, MQO é viesado e inconsistente*. Isso representa muita falta de sorte, porque significa que qualquer viés persiste quando o tamanho da amostra cresce.

No caso da regressão simples, podemos obter a inconsistência da primeira parte da equação (5.3), que se mantém sejam u e x_1 não-correlacionados ou não. A **inconsistência** em $\hat{\beta}_1$ (às vezes, imprecisamente chamada de **viés assumptótico**) é

$$\text{plim } \hat{\beta}_1 - \beta_1 = \text{Cov}(x_1, u)/\text{Var}(x_1). \quad (5.4)$$

Como $\text{Var}(x_1) > 0$, a inconsistência em $\hat{\beta}_1$ é positiva se x_1 e u são positivamente correlacionados, e a inconsistência é negativa se x_1 e u são negativamente correlacionados. Se a covariância entre x_1 e u é pequena relativamente à variância em x_1 , a inconsistência pode ser desprezível; infelizmente, não podemos nem mesmo estimar quão grande é a covariância porque u não é observado.

Podemos usar (5.4) para derivar o análogo assumptótico do viés de variável omitida (veja Tabela 3.2 no Capítulo 3). Suponha que o modelo verdadeiro,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \nu,$$

satisfaça as quatro primeiras hipóteses de Gauss-Markov. Então, ν tem média zero e é não-correlacionado com x_1 e x_2 . Se $\hat{\beta}_0, \hat{\beta}_1$ e $\hat{\beta}_2$ foram os estimadores de MQO da regressão de y sobre x_1 e x_2 , o Teorema 5.1 implica que esses estimadores são consistentes. Se omitirmos x_2 da regressão e fizermos a regressão simples de y sobre x_1 , $u = \beta_2 x_2 + \nu$. Seja $\tilde{\beta}_1$ o estimador de inclinação da regressão simples. Então

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \beta_2 \delta_1 \quad (5.5)$$

em que

$$\delta_1 = \text{Cov}(x_1, x_2)/\text{Var}(x_1). \quad (5.6)$$

Assim, para propósitos práticos, podemos ver a inconsistência como idêntica ao viés. A diferença é que a inconsistência é expressa em termos da variância populacional de x_1 e da covariância populacional

entre x_1 e x_2 , enquanto o viés é baseado em suas contrapartes amostrais (porque estabelecemos condicionamento aos valores de x_1 e x_2 na amostra).

Se x_1 e x_2 forem não-correlacionados (na população), então $\delta_1 = 0$ e $\tilde{\beta}_1$ é um estimador consistente de β_1 (embora não necessariamente não-viesado). Se x_2 tiver um efeito parcial positivo sobre y , de modo que $\beta_2 > 0$, e x_1 e x_2 forem positivamente correlacionados, de modo que $\delta_1 > 0$, a inconsistência em $\tilde{\beta}_1$ é positiva, e assim por diante. Podemos obter a direção da inconsistência ou o viés assimptótico a partir da Tabela 3.2. Se a covariância entre x_1 e x_2 for pequena relativamente à variância de x_1 , a inconsistência pode ser pequena.

EXEMPLO 5.1

{Preços de Casas e Distância de um Incinerador}

Seja y o preço de uma casa (preço), x_1 a distância da casa a um novo incinerador de lixo (distância), e x_2 a "qualidade" da casa (qualidade). A variável *qualidade* é imprecisa, de modo que ela pode incluir coisas como o tamanho da casa e do terreno, número de quartos e de banheiros, e, intangíveis, coisas como a atratividade da vizinhança. Se o incinerador deprecia os preços das casas, então β_1 deve ser positivo: tudo mais sendo igual, uma casa que está mais distante do incinerador é mais valiosa. Por definição, β_2 é positivo, visto que casas de qualidade maior são vendidas por preços maiores, mantendo outros fatores iguais. Se o incinerador estivesse mais longe, em média, das casas melhores, a distância e a qualidade seriam positivamente correlacionadas, e portanto $\delta_1 > 0$. Uma regressão simples de *preço* sobre *distância* [ou $\log(\text{preço})$ sobre $\log(\text{distância})$] tenderá a superestimar o efeito do incinerador: $\beta_1 + \beta_2\delta_1 > \beta_1$.

Suponha que o modelo

$$\text{nota} = \beta_0 + \beta_1 \text{faltas} + \beta_2 \text{nmgradp} + u$$

satisfaz as quatro primeiras hipóteses de Gauss-Markov, e onde *nota* é a nota de um exame final, *faltas* é o número de faltas e *nmgrad* é uma nota média acumulada até o penúltimo semestre. Se $\tilde{\beta}_1$ for o estimador de regressão simples de *nota* sobre *faltas*, qual será a direção do viés assimptótico em $\tilde{\beta}_1$?

Um ponto importante sobre a inconsistência dos estimadores de MQO é que, por definição, o problema não desaparece ao adicionarmos mais observações à amostra. O problema fica pior com mais dados: o estimador de MQO fica mais e mais próximo de $\beta_1 + \beta_2\delta_1$ quando aumenta o tamanho da amostra.

Derivar o sinal e a magnitude da inconsistência no caso geral de k regressores é mais difícil, do mesmo modo que derivar o viés é mais difícil. Precisamos lembrar que, se tivermos o modelo da equação (5.1), em que, por exemplo, x_1 é correlacionado com u , mas as outras variáveis independentes são não-correlacionadas com u , *todos* os estimadores de MQO serão geralmente inconsistentes. Por exemplo, no caso em que $k = 2$,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u,$$

suponha que x_2 e u sejam não-correlacionados, mas x_1 e u sejam correlacionados. Então, os estimadores de MQO $\hat{\beta}_1$ e $\hat{\beta}_2$ serão, em geral, inconsistentes. (O intercepto também será inconsistente.) A inconsistência em $\hat{\beta}_2$ surge quando x_1 e x_2 são correlacionados, como é normalmente o caso. Se x_1 e x_2 forem não-

correlacionados, então qualquer correlação entre x_1 e u não resulta na inconsistência de $\hat{\beta}_2$: $\text{plim } \hat{\beta}_2 = \beta_2$. Além disso, a inconsistência em $\hat{\beta}_1$ é a mesma que em (5.4). A mesma formulação se mantém no caso geral: se x_1 for correlacionado com u , mas x_1 e u não forem correlacionados com as outras variáveis independentes, então somente $\hat{\beta}_1$ é inconsistente, e a inconsistência é dada por (5.4). O caso geral é muito semelhante ao caso de variável omitida da Seção 3A.4 do Apêndice 3A, disponível no site da Thomson.

5.2 NORMALIDADE ASSIMPTÓTICA E INFERÊNCIA EM AMOSTRAS GRANDES

A consistência de um estimador é uma importante propriedade, mas ela sozinha não nos permite trabalhar com inferência estatística. Saber simplesmente que o estimador está se aproximando do valor populacional quando o tamanho da amostra cresce não nos permite testar hipóteses sobre os parâmetros. Para tanto, precisamos da distribuição amostral dos estimadores de MQO. Sob as hipóteses do modelo linear clássico, RLM.1 a RLM.6, o Teorema 4.1 mostra que as distribuições amostrais são normais. Esse resultado é a base para derivar as distribuições t e F que usamos com muita freqüência na econometria aplicada.

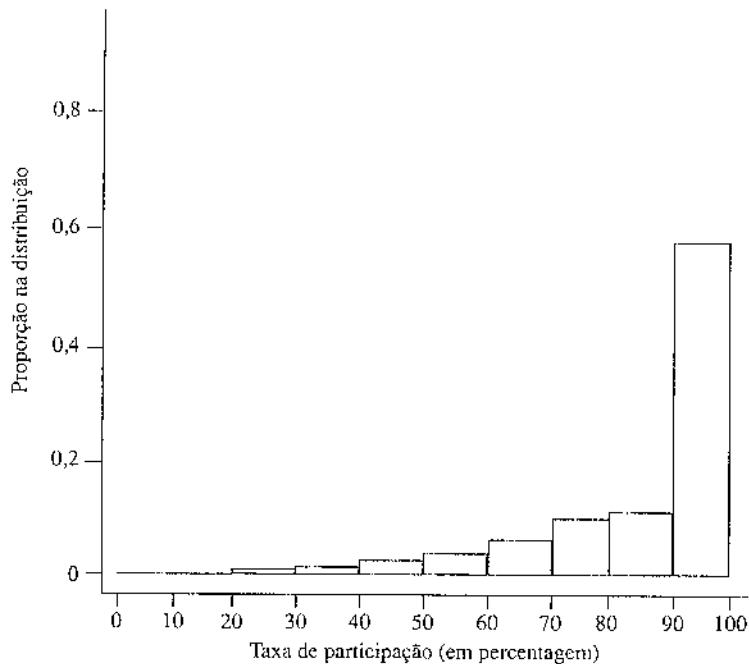
A normalidade exata dos estimadores de MQO depende crucialmente da normalidade da distribuição do erro, u , na população. Se os erros u_1, u_2, \dots, u_n forem extrações aleatórias de alguma distribuição, diferente da normal, o $\hat{\beta}$, não será normalmente distribuído, o que significa que as estatísticas t não terão distribuições t e as estatísticas F não terão distribuições F . Esse é um problema potencialmente sério porque nossa inferência depende de sermos capazes de obter p -valores das distribuições t e F .

Lembre-se de que a hipótese RLM.6 é equivalente a dizer que a distribuição de y , dados x_1, x_2, \dots, x_k , é normal. Como y é observado e u não é, em uma aplicação particular, é muito mais fácil pensar se é provável que a distribuição de y seja normal. De fato, já vimos alguns exemplos em que y definitivamente não poderia ter uma distribuição condicional normal. Uma variável aleatória normalmente distribuída é distribuída simetricamente ao redor de sua média, pode assumir qualquer valor positivo ou negativo (mas com probabilidade zero), e mais de 95% da área sob a distribuição está dentro de dois desvios-padrão.

No Exemplo 3.4 estimamos um modelo que explica o número de prisões de homens jovens durante um determinado ano (*npre86*). Na população, a maioria dos homens não estava presa durante o ano, e a maioria tinha sido presa uma vez no máximo. (Na amostra de 2.725 homens nos dados do arquivo CRIMEI.RAW, menos de 8% foram presos mais que uma vez durante 1986.) Como *npre86* assume somente dois valores para 92% da amostra, ela não pode estar próxima de ser normalmente distribuída na população.

No Exemplo 4.6 estimamos um modelo que explica as percentagens de participação (*taxap*) nos planos de pensão nos Estados Unidos. A distribuição de freqüência (também chamada *histograma*) na Figura 5.2 mostra que a distribuição de *taxap* é fortemente inclinada para a direita, em vez de ser normalmente distribuída. De fato, mais de 40% das observações de *taxap* são iguais a 100, indicando participação de 100%. Isso viola a hipótese de normalidade, mesmo que condicional às variáveis explicativas.

Sabemos que a normalidade não tem nenhum papel na inexistência de viés de MQO, nem afeta a conclusão de que MQO é o melhor estimador linear não-viesado sob as hipóteses de Gauss-Markov. No entanto, a inferência exata baseada nas estatísticas t e F necessita de RML.6. Isso significa que, em nossa análise de *taxap* do Exemplo 4.6, devemos abandonar as estatísticas t para determinar quais variáveis são estatisticamente significantes? Felizmente, a resposta a essa questão é *não*. Ainda que os y_i não sejam provenientes de uma distribuição normal, podemos usar o teorema do limite central do Apêndice C, disponível no site da Thomson, para concluir que os estimadores de MQO satisfazem a **normalidade assimptótica**, o que significa que eles são, de maneira aproximada, normalmente distribuídos em amostras de tamanhos suficientemente grandes.

Figura 5.2Histograma de t_{taxap} , usando dados do arquivo 401K.RAW.**TEOREMA 5.2 (NORMALIDADE ASSIMPTÓTICA DE MQO)**

Sob as hipóteses de Gauss-Markov RLM.1 a RLM.5,

- (i) $\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{d} \text{Normal}(0, \sigma^2/a_j^2)$, em que $\sigma^2/a_j^2 > 0$ é a **variância assimptótica** de $\sqrt{n}(\hat{\beta}_j - \beta_j)$; para os coeficientes de inclinação, $a_j^2 = \text{plim} \left(n^{-1} \sum_{i=1}^n \hat{r}_{ij}^2 \right)$, em que os \hat{r}_{ij} são os resíduos da regressão de x_j sobre outras variáveis independentes. Dizemos que $\hat{\beta}_j$ é assimptótica e normalmente distribuído (veja Apêndice C, disponível no site de Thomson);
- (ii) $\hat{\sigma}^2$ é um estimador consistente de $\sigma^2 = \text{Var}(u)$;
- (iii) Para cada j ,

$$(\hat{\beta}_j - \beta_j)/\text{ep}(\hat{\beta}_j) \xrightarrow{d} \text{Normal}(0, 1), \quad (5.7)$$

em que $\text{ep}(\hat{\beta}_j)$ é o erro-padrão usual de MQO.

A prova da normalidade assimptótica é um pouco complicada e está delineada para o caso da regressão linear simples no apêndice deste capítulo. A parte (ii) provém da lei dos grandes números e a parte (iii) decorre das partes (i) e (ii) e das propriedades assimptóticas discutidas no Apêndice C, disponível no site de Thomson.

O teorema 5.2 é útil porque a hipótese de normalidade RLM.6 foi excluída; a única restrição sobre a distribuição do erro é que ele tenha variância finita, algo que sempre assumiremos. Também assumimos média condicional zero e homoscedasticidade de u .

Observe como a distribuição normal padronizada aparece em (5.7), em oposição à distribuição t_{n-k-1} . Isso ocorre porque a distribuição é somente aproximada. Em contraste, no Teorema 4.2, a distribuição da relação em (5.7) era *exatamente* t_{n-k-1} para qualquer tamanho de amostra. De uma perspectiva prática, essa diferença é irrelevante. De fato, é válido escrever

$$(\hat{\beta}_j - \beta_j)/\text{cp}(\hat{\beta}_j) \stackrel{d}{\rightarrow} t_{n-k-1}, \quad (5.8)$$

visto que t_{n-k-1} aproxima-se da distribuição normal padronizada quando os graus de liberdade tornam-se grandes.

A equação (5.8) nos diz que o teste t e a construção dos intervalos de confiança são realizados *exatamente* como sob as hipóteses do modelo linear clássico. Isso significa que nossa análise das variáveis dependentes, como *taxap* e *npre86*, não tem absolutamente de mudar se as hipóteses de Gauss-Markov se mantêm; em ambos os casos, temos pelo menos 1.500 observações, o que certamente é suficiente para justificar a aproximação pelo teorema do limite central.

Se o tamanho da amostra não é muito grande, então a distribuição t pode ser uma aproximação insatisfatória da distribuição da estatística t quando u não é normalmente distribuído. Infelizmente, antes de saber se a aproximação é suficientemente boa, não há prescrições gerais de quão grande deve ser o tamanho da amostra. Alguns econometristas pensam que $n = 30$ é satisfatório, mas esse valor pode não ser suficiente para todas as possíveis distribuições de u . Dependendo da distribuição de u , podem ser necessárias mais observações antes de o teorema do limite central começar a fazer efeito. Além disso, a qualidade da aproximação não depende apenas de n , mas dos gl , $n - k - 1$: com mais variáveis independentes no modelo, um tamanho de amostra maior é usualmente necessário para usar a aproximação t . Os métodos para inferência com graus de liberdade e erros não normais estão fora do escopo deste livro. Usaremos as estatísticas t como sempre usamos, sem nos preocuparmos com a hipótese de normalidade.

É muito importante ver que o Teorema 5.2, *de fato*, exige a hipótese de homoscedasticidade (juntamente com a hipótese de média condicional zero). Se $\text{Var}(y|x)$ não é constante, as estatísticas t usuais e os intervalos de confiança não são válidos, não importa quão grande seja o tamanho da amostra; na presença da heteroscedasticidade, o teorema do limite central em nada nos ajuda. Por essa razão, dedicaremos todo o Capítulo 8 à discussão do que pode ser feito na presença de heteroscedasticidade.

Uma conclusão do Teorema 5.2 é que $\hat{\sigma}^2$ é um estimador consistente de σ^2 ; já sabemos do Teorema 3.3 que $\hat{\sigma}^2$ é não-viesado para σ^2 sob as hipóteses de Gauss-Markov. A consistência implica que $\hat{\sigma}$ é um estimador consistente de σ , o que é importante para estabelecer o resultado da normalidade assimptótica na equação (5.7).

Lembre-se de que $\hat{\sigma}$ aparece no erro-padrão de cada $\hat{\beta}_j$. De fato, a variância estimada de $\hat{\beta}_j$ é

$$\text{Vár}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{\text{SQT}_j(1 - R_j^2)}, \quad (5.9)$$

em que SQT_j é a soma dos quadrados total de x_j na amostra, e R_j^2 é o R -quadrado da regressão de x_j sobre todas as outras variáveis independentes. Na Seção 3.4 estudamos cada componente de (5.9), os quais vamos expor agora no contexto da análise assintótica. Quando o tamanho da amostra aumenta, $\hat{\sigma}^2$ converge em probabilidade para a constante σ^2 . Além disso, R_j^2 se aproxima de um número estritamente entre zero e um (de modo que $1 - R_j^2$ converge para algum número entre zero e um). A variância amos-

tral de x_j é SQT/n , e portanto SQT/n converge para $\text{Var}(x_j)$ quando o tamanho da amostra aumenta. Isso quer dizer que SQT_j cresce aproximadamente à mesma taxa que o tamanho da amostra: $SQT_j \approx n\sigma_j^2$, em que σ_j^2 é a variância populacional de x_j . Quando combinamos esses fatos, vemos que $\text{Var}(\hat{\beta}_j)$ se contrai para zero à taxa de $1/n$; essa é a razão de tamanhos maiores de amostra serem melhores.

Em um modelo de regressão com um tamanho de amostra grande, qual é o intervalo de confiança de 95% aproximado para $\hat{\beta}_j$ sob RLM.1 a RLM.5? Ele é chamado de **intervalo de confiança assimptótico**.

Quando u não é normalmente distribuído, a raiz quadrada de (5.9) é, às vezes, chamada de **erro-padrão assimptótico**, e as estatísticas t são chamadas de **estatísticas t assimptóticas**. Como essas quantidades são as mesmas que estudamos no Capítulo 4, vamos chamá-las apenas de erros-padrão e estatísticas t , com o entendimento de que elas, algumas vezes, têm somente justificativa de amostra grande.

Ao usar o argumento anterior sobre a variância estimada, podemos escrever

$$\text{ep}(\hat{\beta}_j) \approx c_j / \sqrt{n}, \quad (5.10)$$

em que c_j é uma constante positiva que *não* depende do tamanho da amostra. A equação (5.10) é somente uma aproximação, mas ela é uma regra de bolso útil: pode-se esperar que os erros-padrão diminuam a uma taxa que é o inverso da *raiz quadrada* do tamanho da amostra.

EXEMPLO 5.2

(Erros-Padrão em uma Equação do Peso de Nascimentos)

Usamos os dados do arquivo BWGHT.RAW para estimar uma relação em que o log do peso de nascimento é a variável dependente, e os cigarros fumados por dia (*cigs*) e o log da renda familiar são as variáveis independentes. O número total de observações é 1.388. Ao usar a primeira metade das observações (694), o erro-padrão de $\hat{\beta}_{cigs}$ é cerca de 0,0013. Ao usar todas as observações, o erro-padrão é cerca de 0,00086. A razão entre o último erro-padrão e o primeiro é $0,00086/0,0013 \approx 0,662$. Isso está bastante próximo de $\sqrt{694/1.388} \approx 0,707$, a razão obtida pela aproximação em (5.10). Em outras palavras, a equação (5.10) implica que o erro-padrão, ao usar o tamanho de amostra maior, deve ser 70,7% do erro-padrão obtido ao usar a amostra menor. Essa percentagem está muito perto dos 66,2% que realmente calculamos a partir da relação entre os erros-padrão.

A normalidade assimptótica dos estimadores de MQO também implica que as estatísticas F têm distribuições F aproximadas em tamanhos de amostras grandes. Assim, para testar as restrições de exclusão ou outras hipóteses múltiplas, nada muda em relação ao que tínhamos feito antes.

Outros Testes de Amostras Grandes: A Estatística Multiplicador de Lagrange

Visto que entramos no domínio da análise assimptótica, outras estatísticas de testes podem ser usadas para testar hipóteses. Para a maioria dos propósitos, há pouca razão para ir além das estatísticas t e F .

usuais: como acabamos de ver, essas estatísticas têm justificativa de amostra grande sem a hipótese de normalidade. No entanto, algumas vezes é útil ter outras maneiras de testar restrições de exclusão múltiplas; por isso, vamos agora estudar a **estatística multiplicador de Lagrange (LM)**, que vem alcançando alguma popularidade na econometria moderna.

O nome “estatística multiplicador de Lagrange” provém da otimização com restrição, um tópico além do escopo deste livro. [Veja Davidson e MacKinnon (1993).] Também é usado o nome **estatística de escore** – o qual também é proveniente da otimização utilizada em cálculo. Felizmente, no arcabouço da regressão linear, é simples explicar a estatística *LM* sem se aprofundar na matemática mais complicada.

A forma da estatística *LM* que vamos derivar aqui apóia-se nas hipóteses de Gauss-Markov, as mesmas hipóteses que justificam a estatística *F* em amostras grandes. Não precisamos da hipótese de normalidade.

Para derivar a estatística *LM*, considere o modelo de regressão múltipla habitual com k variáveis independentes:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u. \quad (5.11)$$

Gostaríamos de testar se, por exemplo, todas as últimas q dessas variáveis têm parâmetros populacionais zero: a hipótese nula é

$$H_0: \beta_{k-q+1} = 0, \dots, \beta_k = 0, \quad (5.12)$$

a qual coloca q restrições de exclusão sobre o modelo (5.11). Assim como no teste *F*, a hipótese alternativa a (5.12) é que pelo menos um dos parâmetros é diferente de zero.

A estatística *LM* exige somente a estimação do modelo *restrito*. Assim, assuma que computamos a regressão

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \dots + \tilde{\beta}_{k-q} x_{k-q} + \tilde{u}, \quad (5.13)$$

em que “~” indica que as estimativas são do modelo restrito. Em particular, \tilde{u} representa os resíduos do modelo restrito. (Como sempre, isso é apenas uma maneira de escrever para indicar que obtivemos o resíduo restrito de cada observação da amostra.)

Se as variáveis omitidas x_{k-q+1} até x_k tiverem, realmente, coeficientes populacionais zero, então \tilde{u} deve ser, pelo menos aproximadamente, não-correlacionado com cada uma dessas variáveis na amostra. Isso sugere computar uma regressão desses resíduos sobre aquelas variáveis independentes excluídas sob H_0 , que é semelhante ao que o teste *LM* faz. Entretanto, para obter uma estatística de teste que possa ser usada, devemos incluir *todas* as variáveis independentes na regressão. (A razão pela qual devemos incluir todos os regressores é que, em geral, os regressores omitidos no modelo restrito são correlacionados com os regressores que aparecem no modelo restrito.) Assim, computamos a regressão de

$$\tilde{u} \text{ sobre } x_1, x_2, \dots, x_k. \quad (5.14)$$

Essa é um exemplo de **regressão auxiliar**, uma regressão usada para calcular uma estatística de teste, mas cujos coeficientes não são de interesse direto.

Como podemos usar o resultado da regressão de (5.14) para testar (5.12)? Se (5.12) for verdadeira, o R -quadrado de (5.14) deve estar “próximo” de zero, sujeito ao erro amostral, porque \tilde{u} será aproximadamente não-correlacionado com todas as variáveis independentes. A questão – como sempre ocorre com os testes de hipóteses – é como determinar quando a estatística é suficientemente grande para rejeitar a hipótese nula a um nível de significância escolhido. Isso resulta que, sob a hipótese nula, o tamanho da amostra multiplicado pelo R -quadrado da regressão auxiliar (5.14) é distribuído assintoticamente como uma variável aleatória qui-quadrada com q graus de liberdade. Isso leva a um procedimento simples para testar a significância conjunta de um grupo de q variáveis independentes.

A ESTATÍSTICA MULTIPLICADOR DE LAGRANGE PARA q RESTRIÇÕES DE EXCLUSÃO:

- (i) Regreda y sobre o conjunto *restrito* de variáveis independentes e salve os resíduos, \tilde{u} .
- (ii) Regreda \tilde{u} sobre *todas* as variáveis independentes e obtenha o R -quadrado, por exemplo $R_{\tilde{u}}^2$ (para distingui-lo dos R -quadrados obtidos com y como variável dependente).
- (iii) Calcule $LM = nR_{\tilde{u}}^2$ [o tamanho amostral vezes o R -quadrado obtido no passo (ii)].
- (iv) Compare o LM com o valor crítico apropriado, c , de uma distribuição χ_q^2 ; se $LM > c$, a hipótese nula é rejeitada. Melhor ainda, obtenha o p -valor como a probabilidade de que uma variável aleatória χ_q^2 exceda o valor da estatística de teste. Se o p -valor for menor que o nível de significância desejado, então H_0 é rejeitada. Se não for, não podemos rejeitar H_0 . A regra de rejeição é essencialmente a mesma do teste F .

Por causa de sua forma, a estatística LM é às vezes referida como a **estatística R -n-quadrado**. Diferentemente da estatística F , os graus de liberdade do modelo irrestrito não têm qualquer papel na realização do teste LM . Tudo o que importa é o número de restrições que estão sendo testadas (q), o tamanho do R -quadrado auxiliar ($R_{\tilde{u}}^2$) e o tamanho da amostra (n). Os gl do modelo irrestrito não têm qualquer papel por causa da natureza assintótica da estatística LM . No entanto, devemos estar certos de multiplicar $R_{\tilde{u}}^2$ pelo tamanho da amostra para obter LM ; um valor aparentemente baixo do R -quadrado pode ainda levar à significância conjunta se n for grande.

Antes de dar um exemplo, uma palavra de precaução se faz necessária. Se, no passo (i), inadvertidamente regredirmos y sobre todas as variáveis independentes e utilizarmos os resíduos obtidos dessa regressão irrestrita no passo (ii), não vamos ter uma estatística interessante: o R -quadrado resultante será exatamente zero! Isso ocorre porque MQO escolhe as estimativas de modo que os resíduos sejam não-correlacionados nas amostras com todas as variáveis independentes incluídas [veja as equações (3.13)]. Assim, podemos somente testar (5.12) ao regredir os resíduos restritos sobre todas as variáveis independentes. (Regredir os resíduos restritos sobre o conjunto restrito de variáveis independentes também produzirá $R^2 = 0$.)

EXEMPLO 5.3

(Modelo Econômico do Crime)

Vamos ilustrar o teste LM ao usar uma versão ligeiramente mais extensa do modelo do crime do Exemplo 3.4:

$$npre86 = \beta_0 + \beta_1 pcond + \beta_2 sentmed + \beta_3 temptot + \beta_4 ptemp86 + \beta_5 sempr86 + u,$$

em que $npre86$ é o número de vezes que um homem foi preso, $pcond$ é a proporção de prisões anteriores que levaram à condenação, $sentmed$ é a sentença média cumprida de condenações passadas, $temptot$ é o

EXEMPLO 5.3 (continuação)

tempo total que o homem passou na prisão em 1986 desde que atingiu a idade de 18 anos, $p\text{temp}86$ corresponde aos meses passados na prisão em 1986 e $\text{empr}86$ é o número de trimestres, em 1986, durante os quais o homem esteve legalmente empregado. Vamos usar a estatística LM para testar a hipótese nula de que sentmed e temptot não têm efeito sobre $n\text{pre}86$, uma vez que os outros fatores foram controlados.

No passo (i), estimamos o modelo restrito ao regredir $n\text{pre}86$ sobre $p\text{cond}$, $p\text{temp}86$ e $\text{empr}86$ – as variáveis sentmed e temptot são excluídas dessa regressão – e obtemos os resíduos \hat{u} dessa regressão (2.725 resíduos). Então, computamos a regressão de

$$\hat{u} \text{ sobre } p\text{cond}, p\text{temp}86, \text{empr}86, \text{sentmed} \text{ e } \text{temptot}; \quad (5.15)$$

como sempre, a ordem na qual listamos as variáveis independentes é irrelevante. Essa segunda regressão gera $R^2_{\hat{u}\hat{u}}$ que é cerca de 0,0015. Esse valor pode parecer pequeno, mas devemos multiplicá-lo por n para obter a estatística $LM = 2.725(0,0015) \approx 4,09$. O valor crítico de 10% em uma distribuição qui-quadrada com dois graus de liberdade é cerca de 4,61 (arredondado para duas casas decimais; veja a Tabela G.4). Assim, não é possível rejeitar a hipótese nula de que $\beta_{\text{sentmed}} = 0$ e $\beta_{\text{temptot}} = 0$ ao nível de 10%. O p -valor é $P(\chi^2_2 > 4,09) \approx 0,129$, de modo que rejeitamos H_0 ao nível de 15%.

Como comparação, o teste F para a significância conjunta de sentmed e temptot resulta em um p -valor de cerca de 0,131, bastante próximo daquele obtido ao usar a estatística LM . Isso não é surpreendente, visto que assimptoticamente as duas estatísticas têm a mesma probabilidade de erro Tipo I. (Isto é, elas rejeitam a hipótese nula com a mesma freqüência quando a hipótese nula é verdadeira.)

Como o exemplo anterior sugere, com uma amostra grande raramente vemos discrepâncias importantes entre os resultados dos testes LM e F . Usaremos a estatística F para a maior parte dos problemas porque ela é rotineiramente calculada pela maioria dos pacotes de regressão. No entanto, você deve estar consciente sobre a estatística LM para reconhecê-la, pois ela é usada em trabalhos aplicados.

Um último comentário sobre a estatística LM . Assim como com a estatística F , devemos estar seguros de usar as mesmas observações nos passos (i) e (ii). Se faltarem dados para algumas das variáveis independentes excluídas sob a hipótese nula, os resíduos do passo (i) devem ser obtidos de uma regressão sobre o conjunto de dados reduzido.

5.3 EFICIÊNCIA ASSIMPTÓTICA DE MQO

Sabemos que, sob as hipóteses de Gauss-Markov, os estimadores de MQO são os melhores estimadores não-viesados lineares. MQO também é, sob as hipóteses de Gauss-Markov, **assimptoticamente eficiente** dentro uma classe de estimadores. Um tratamento geral requer álgebra matricial e análise assintótica avançada. Em primeiro lugar, vamos descrever o resultado para o caso da regressão simples.

No modelo

$$y = \beta_0 + \beta_1 x + u, \quad (5.16)$$

u tem média condicional zero sob RLM.3: $E(u|x) = 0$. Isso dá lugar a uma variedade de estimadores consistentes de β_0 e β_1 ; como habitual, vamos nos concentrar no parâmetro de inclinação, β_1 . Seja $g(x)$ qual-

quer função de x ; por exemplo, $g(x) = x^2$ ou $g(x) = 1/(1 + |x|)$. Então u é não-correlacionado com $g(x)$ (veja a Propriedade CE.5 no Apêndice B, disponível no site da Thomson). Seja $z_i = g(x_i)$ para todas as observações i . Então, o estimador

$$\tilde{\beta}_1 = \left(\sum_{i=1}^n (z_i - \bar{z})y_i \right) / \left(\sum_{i=1}^n (z_i - \bar{z})x_i \right) \quad (5.17)$$

é consistente para β_1 , desde que $g(x)$ e x sejam correlacionados. [Lembre-se: é possível que $g(x)$ e x sejam não-correlacionados porque a correlação mensura a dependência *linear*.] Para ver isso, podemos colocar $y_i = \beta_0 + \beta_1 x_i + u_i$ em (5.17) e escrever $\tilde{\beta}_1$ como

$$\tilde{\beta}_1 = \beta_1 + \left(n^{-1} \sum_{i=1}^n (z_i - \bar{z})u_i \right) / \left(n^{-1} \sum_{i=1}^n (z_i - \bar{z})x_i \right). \quad (5.18)$$

Agora, podemos aplicar a lei dos grandes números ao numerador e denominador, os quais convergem em probabilidade para $\text{Cov}(z,u)$ e $\text{Cov}(z,x)$, respectivamente. Na condição de que $\text{Cov}(z,x) \neq 0$ – de modo que z e x sejam correlacionados –, temos

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \text{Cov}(z,u)/\text{Cov}(z,x) = \beta_1,$$

porque $\text{Cov}(z,u) = 0$ sob RLM.3.

É mais difícil mostrar que $\tilde{\beta}_1$ é assimptoticamente normal. No entanto, usando argumentos semelhantes àqueles do apêndice deste capítulo, pode ser mostrado que $\sqrt{n}(\tilde{\beta}_1 - \beta_1)$ é assimptoticamente normal com média zero e variância assimptótica $\sigma^2 \text{Var}(z)/[\text{Cov}(z,x)]^2$. A variância assimptótica do estimador de MQO é obtida quando $z = x$, caso em que $\text{Cov}(z,x) = \text{Cov}(x,x) = \text{Var}(x)$. Portanto, a variância assimptótica de $\sqrt{n}(\hat{\beta}_1 - \beta_1)$, em que $\hat{\beta}_1$ é o estimador de MQO, é $\sigma^2 \text{Var}(x)/[\text{Var}(x)]^2 = \sigma^2/\text{Var}(x)$. Agora, a desigualdade de Cauchy-Schwartz (veja Apêndice B.4, disponível no site da Thomson) implica que $[\text{Cov}(z,x)]^2 \leq \text{Var}(z)\text{Var}(x)$, o que implica que a variância assimptótica de $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ não é maior do que a de $\sqrt{n}(\tilde{\beta}_1 - \beta_1)$. Assim, para o caso da regressão simples, mostramos que, sob as hipóteses de Gauss-Markov, o estimador de MQO tem uma variância assimptótica menor do que qualquer outro estimador da forma (5.17). [O estimador em (5.17) exemplifica um *estimador de variáveis instrumentais*, que estudaremos extensivamente no Capítulo 15.] Se a hipótese de homoscedasticidade não for válida, então há estimadores da forma (5.17) que têm uma variância assimptótica menor do que a de MQO. Veremos isso no Capítulo 8.

O caso geral é semelhante, mas matematicamente muito mais difícil. No caso de k regressores, a classe de estimadores consistentes é obtida ao generalizar as condições de primeira ordem de MQO:

$$\sum_{i=1}^n g_j(x_i)(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \dots - \tilde{\beta}_k x_{ik}) = 0, j = 0, 1, \dots, k, \quad (5.19)$$

em que $g_j(x_i)$ representa qualquer função de todas as variáveis explicativas para a observação i . Como pode ser visto ao comparar (5.19) com as condições de primeira ordem de MQO em (3.13), obtemos os estimadores de MQO quando $g_0(x_i) = 1$ e $g_j(x_i) = x_{ij}$, para $j = 1, 2, \dots, k$. A classe dos estimadores em (5.19) é infinita, pois podemos usar qualquer função de x_{ij} que quisermos.

T E O R E M A 5 . 3 (EFICIÊNCIA ASSIMPTÓTICA DE MQO)

Sob as hipóteses de Gauss-Markov, sejam $\tilde{\beta}_j$ os estimadores que solucionam as equações da forma (5.19) e sejam $\hat{\beta}_j$ os estimadores de MQO. Então, para $j = 0, 1, 2, \dots, k$, os estimadores de MQO têm as menores variâncias assimptóticas: $\text{Avar } \sqrt{n}(\hat{\beta}_j - \beta_j) \leq \text{Avar } \sqrt{n}(\tilde{\beta}_j - \beta_j)$.

Provar a consistência dos estimadores em (5.19), sem mostrar que eles são assintoticamente normais, é matematicamente difícil. [Veja Wooldridge (2002, Capítulo 5).]

As afirmações subjacentes ao material deste capítulo são razoavelmente técnicas, mas suas implicações práticas são diretas. Mostramos que as quatro primeiras hipóteses de Gauss-Markov implicam que MQO é consistente. Além disso, todos os métodos de testar e construir intervalos de confiança que aprendemos no Capítulo 4 são aproximadamente válidos, sem assumir que os erros são extraídos de uma distribuição normal (equivalentemente, a distribuição de y , dadas as variáveis explicativas, não é normal). Isso significa que podemos aplicar MQO e usar os métodos anteriores para um conjunto de aplicações em que a variável dependente não é de fato aproximadamente normalmente distribuída. Também mostramos que, em vez da estatística F , a estatística LM pode ser usada para testar restrições de exclusão.

Antes de deixarmos este capítulo, devemos observar que coisas como o Exemplo 5.3 podem muito bem apresentar problemas que, *de fato*, exigem atenção especial. Para uma variável como $npre86$, que é zero ou um para a maioria dos homens na população, um modelo linear pode não ser capaz de adequadamente capturar a relação funcional entre $npre86$ e as variáveis explicativas. Além do mais, mesmo se um modelo linear descreve o valor esperado das prisões, a heteroscedasticidade poderia ser um problema. Problemas como esses não são mitigados quando o tamanho da amostra aumenta, e portanto retornaremos a eles em capítulos posteriores.

5.1 No modelo de regressão simples sob RLM.1 a RLM.4, dissemos que o estimador de inclinação, $\hat{\beta}_1$, é consistente para β_1 . Usando $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$, mostre que $\text{plim } \hat{\beta}_0 = \beta_0$. [Você precisa usar a consistência de $\hat{\beta}_1$ e a lei dos grandes números, juntamente com o fato de que $\beta_0 = E(y) - \beta_1 E(x_1)$.]

5.2 Suponha que o modelo

$$pctação = \beta_0 + \beta_1 funds + \beta_2 riscotol + u$$

satisfaca as quatro primeiras hipóteses de Gauss-Markov, onde $pctação$ é a percentagem da pensão de um trabalhador investida no mercado de ações, $funds$ é o número de fundos mútuos que o trabalhador pode escolher e $riscotol$ é alguma medida de tolerância de risco ($riscotol$ maior significa que a pessoa tem uma tolerância maior ao risco). Se $funds$ e $riscotol$ são positivamente correlacionados, qual é a inconsistência em $\hat{\beta}_1$, o coeficiente de inclinação da regressão de $pctação$ sobre $funds$?

5.3 O conjunto de dados do arquivo SMOKE.RAW contém informações sobre o comportamento tabagista e outras variáveis para uma amostra aleatória de adultos solteiros dos Estados Unidos. A variável *cigs* é o número (médio) de cigarros fumados por dia. Você acha que *cigs* tem uma distribuição normal na população adulta dos Estados Unidos? Explique.

5.4 No modelo de regressão simples (5.16), sob as quatro primeiras hipóteses de Gauss-Markov, mostramos que os estimadores da forma (5.17) são consistentes para a inclinação, β_1 . Dado tal estimador, defina um estimador de β_0 como $\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1 \bar{x}$. Mostre que $\text{plim } \tilde{\beta}_0 = \beta_0$.

Vamos delinear uma prova da normalidade assimptótica de MQO [Teorema 5.2(i)] no caso da regressão simples. Escreva o modelo de regressão simples como na equação (5.16). Em seguida, por meio da álgebra usual da regressão simples, podemos escrever

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = (1/s_x^2)[n^{-1/2}\sum_{i=1}^n(x_i - \bar{x})u_i],$$

em que usamos s_x^2 para representar a variância amostral de $\{x_i; i = 1, 2, \dots, n\}$. Pela lei dos grandes números (veja o Apêndice C, disponível no site de Thomson), $s_x^2 \xrightarrow{P} \sigma_x^2 = \text{Var}(x)$. A hipótese RLM.4 exclui a perfeita colinearidade, o que significa que $\text{Var}(x) > 0$ (x_i varia na amostra, e portanto x não é constante na população). Em seguida, $n^{-1/2}\sum_{i=1}^n(x_i - \bar{x})u_i = n^{-1/2}\sum_{i=1}^n(x_i - \mu)u_i + (\mu - \bar{x})[n^{-1/2}\sum_{i=1}^nu_i]$, em que $\mu = E(x)$ é a média populacional de x . Agora $\{u_i\}$ é a seqüência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média zero e variância σ^2 , e portanto $n^{-1/2}\sum_{i=1}^nu_i$ converge para a distribuição Normal($0, \sigma^2$) quando $n \rightarrow \infty$; isso é exatamente o teorema do limite central do Apêndice C, disponível no site da Thomson. Pela lei dos grandes números, $\text{plim}(\mu - \bar{x}) = 0$. Um resultado padrão da teoria assimptótica é que se $\text{plim}(w_n) = 0$ e z_n tem uma distribuição normal assimptótica, então $\text{plim}(w_n z_n) = 0$. [Veja Wooldridge (2002, Capítulo 3) para mais discussão.] Isso implica que $(\mu - \bar{x})[n^{-1/2}\sum_{i=1}^nu_i]$ tem plim zero. Em seguida, $\{(x_i - \mu)u_i; i = 1, 2, \dots\}$ é uma seqüência indefinida de variáveis aleatórias i.i.d. com média zero – porque u e x são não-correlacionados sob RLM.3 – e variância $\sigma^2\sigma_x^2$, pela hipótese de homoscedasticidade RLM.5. Portanto, $n^{-1/2}\sum_{i=1}^n(x_i - \mu)u_i$ tem uma distribuição Normal($0, \sigma^2\sigma_x^2$) assimptótica. Acabamos de mostrar que a diferença entre $n^{-1/2}\sum_{i=1}^n(x_i - \bar{x})u_i$ e $n^{-1/2}\sum_{i=1}^n(x_i - \mu)u_i$ tem plim zero. Um resultado da teoria assimptótica é que se z_n tem uma distribuição normal e $\text{plim}(v_n - z_n) = 0$, então v_n tem a mesma distribuição normal assimptótica que z_n . Em decorrência $n^{-1/2}\sum_{i=1}^n(x_i - \bar{x})u_i$ também tem uma distribuição Normal($0, \sigma^2\sigma_x^2$) assimptótica. Colocando todas essas peças juntas, temos

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = (1/\sigma_x^2)[n^{-1/2}\sum_{i=1}^n(x_i - \bar{x})u_i]$$

$$+ [(1/s_x^2) - (1/\sigma_x^2)][n^{-1/2} \sum_{i=1}^n (x_i - \bar{x}) u_i],$$

e como $\text{plim}(1/s_x^2) = 1/\sigma_x^2$, o segundo termo tem plim zero. Portanto, a distribuição assimptótica de $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ é $\text{Normal}(0, (\sigma^2 \sigma_x^2)/(\sigma_x^2)^2) = \text{Normal}(0, \sigma^2/\sigma_x^2)$. Isso completa a prova para o caso da regressão simples, quando $\sigma_1^2 = \sigma_x^2$ neste caso. Veja Wooldridge (2002, Capítulo 4) para o caso geral.

Análise de Regressão Múltipla: Problemas Adicionais

ste capítulo articula vários problemas da análise de regressão múltipla que não foram tratados convenientemente nos capítulos anteriores. Estes tópicos não são tão fundamentais quanto os discutidos nos capítulos 3 e 4, mas são importantes para a aplicação da regressão múltipla em uma ampla gama de problemas empíricos.

6.1 EFEITOS DA DIMENSÃO DOS DADOS NAS ESTATÍSTICAS MQO

No Capítulo 2, sobre regressão bivariada, discutimos de forma sucinta os efeitos da mudança nas unidades de medida sobre os interceptos e as estimativas de inclinação do MQO. Também mostramos que a mudança nas unidades de medida não afeta o R -quadrado. Agora retornaremos ao problema da dimensão dos dados e examinaremos o efeito do redimensionamento das variáveis dependente ou independente sobre os erros-padrão, estatísticas t , estatísticas F e intervalos de confiança.

Veremos que tudo o que esperamos acontecer, acontecerá. Quando as variáveis são redimensionadas, os coeficientes, erros-padrão, intervalos de confiança, estatísticas t e F mudam de tal maneira que preservam todos os efeitos mensurados e os resultados dos testes. Embora isso não seja uma grande surpresa – aliás, ficaríamos muito preocupados se não fosse assim – é útil ver o que ocorre explicitamente. Muitas vezes, o redimensionamento de dados é usado com finalidade cosmética, tal como reduzir o número de zeros depois da vírgula, em um coeficiente estimado. Escolhendo-se criteriosamente as unidades de medida, podemos melhorar a aparência de uma equação estimada sem alterar nada que seja essencial.

Poderíamos tratar deste problema de maneira generalizada, mas ele é mais bem ilustrado com exemplos. Da mesma forma, será de pouca valia neste ponto introduzirmos uma notação abstrata.

Começamos com uma equação relacionando o peso dos recém-nascidos com o hábito de fumar e a renda familiar:

$$pes\acute{o}nas = \hat{\beta}_0 + \hat{\beta}_1 cigs + \hat{\beta}_2 rendfam \quad (6.1)$$

onde $pes\acute{o}nas$ é o peso dos recém-nascidos, em onças, $cigs$ é o número médio de cigarros que a mãe fumou por dia durante a gravidez, e $rendfam$ é a renda anual familiar, em milhares de dólares. As estimativas desta equação, obtidas utilizando dados contidos no arquivo BWGHT.RAW, são dadas na primeira coluna da Tabela 6.1. Os erros-padrão estão relacionados entre parênteses. A estimativa de $cigs$ mostra que se uma mulher fumar cinco ou mais cigarros por dia, o peso previsto dos recém-

nascidos deve estar em torno de $0,4634(5) = 2,317$ onças a menos. A estatística t de $cigs$ é $-5,06$, de modo que a variável é estatisticamente bastante significante.

Tabela 6.1

Efeitos da Dimensão dos Dados

Variável Dependente	(1) <i>pesonas</i>	(2) <i>pesonaslb</i>	(3) <i>pesonas</i>
Variáveis Independentes			
<i>cigs</i>	-0,4634 (0,0916)	-0,0289 (0,0057)	—
<i>maços</i>	—	—	-9,268 (1,832)
<i>rendfam</i>	0,0927 (0,0292)	0,0058 (0,0018)	0,0927 (0,0292)
<i>intercepto</i>	116,974 (1,049)	7,3109 (0,0656)	116,974 (1,049)
Observações	1.388	1.388	1.388
R-quadrado	0,0298	0,0298	0,0298
SQR	557.485,51	2.177,6778	557.485,51
EPR	20,063	1,2539	20,063

Agora, suponha que decidimos medir o peso dos recém-nascidos em libras, em vez de onças. Façamos $\text{pesonaslb} = \text{pesonas}/16$ ser o peso dos recém-nascidos em libras. O que acontece com nossas estatísticas MQO se usarmos essa variável dependente em nossa equação? É fácil verificar o efeito no coeficiente da estimativa pela simples manipulação da equação (6.1). Divida a equação inteira por 16:

$$\text{pesonas}/16 = \hat{\beta}_0/16 + (\hat{\beta}_1/16)cigs + (\hat{\beta}_2/16)\text{rendfam}.$$

Como o termo da esquerda é o peso dos recém-nascidos em libras, segue-se que cada novo coeficiente corresponderá ao coeficiente antigo dividido por 16. Para verificar isso, a regressão de pesonaslb sobre $cigs$ e rendfam está registrada na coluna (2) da Tabela 6.1. Até quatro dígitos, o intercepto e as inclinações da coluna (2) são exatamente os da coluna (1) divididos por 16. Por exemplo, o coeficiente de $cigs$ é agora $-0,0289$; isso significa que, se $cigs$ fosse cinco vezes mais alto, o peso de nascimento sceria $0,0289(5) = 0,1445$ libras mais baixo. Em termos de onças, temos $0,1445(16) = 2,312$, que é um pouco diferente dos 2,317 que obtivemos anteriormente devido ao erro de arredondamento. A questão importante é que, uma vez que os efeitos tenham sido transformados nas mesmas unidades, obtemos exatamente a mesma resposta, independentemente de como a variável dependente seja medida.

E quanto à significância estatística? Como esperado, a alteração da variável dependente de onças para libras não tem efeito sobre o quanto são estatisticamente importantes as variáveis independentes. Os erros-padrão na coluna (2) são 16 vezes menores que os da coluna (1). Alguns cálculos rápidos mostram que as estatísticas t na coluna (2) são, realmente, idênticas às da coluna (1). Os pontos extremos dos intervalos de confiança na coluna (2) são exatamente os pontos extremos na coluna (1) divididos por 16. Isso ocorre porque os ICs mudam pelos mesmos fatores dos erros-padrão. (Lembre-se de que o IC de 95% neste caso é $\hat{\beta}_j \pm 1,96 \text{ ep}(\hat{\beta}_j)$.)

Em termos de grau de ajuste, os R -quadrados das duas regressões são idênticos, como esperado. Observe que a soma dos resíduos quadrados, SQR, e o erro-padrão da regressão, EPR, diferem nas equações. Essas diferenças são facilmente explicadas. Seja \hat{u}_i o resíduo da observação i na equação original (6.1). Então, quando pesonaslb é a variável dependente, o resíduo é simplesmente $\hat{u}_i/16$. Assim, o resíduo quadrado na segunda equação é $(\hat{u}_i/16)^2 = \hat{u}_i^2/256$. Essa é a razão pela qual a soma dos resíduos quadrados na coluna (2) é igual à SQR na coluna (1) dividida por 256.

Como $\text{EPR} = \hat{\sigma} = \sqrt{\text{SQR}/(n-k-1)} = \sqrt{\text{SQR}/1.385}$, SQR na coluna (2) é 16 vezes menor do que na coluna (1). Outra maneira de ver isso é que o erro na equação com pesonaslb como a variável dependente tem um desvio-padrão 16 vezes menor do que o desvio-padrão do erro original. Isso não significa que tenhamos reduzido o erro ao alterarmos a maneira pela qual o peso dos recém-nascidos é medido: o EPR menor simplesmente reflete uma diferença nas unidades de medida.

Continuando, retornemos à unidade de medida original da variável dependente: pesonas é medido em onças. Vamos alterar a unidade de medida de uma das variáveis independentes, $cigs$. Defina maços como sendo a quantidade de maços de cigarros fumados por dia. Assim, $\text{maços} = \text{cigs}/20$. Agora, o que acontece com os coeficientes e outras estatísticas MQO? Dessa forma, podemos escrever

$$\text{pesonas} = \hat{\beta}_0 + (20\hat{\beta}_1)(\text{cigs}/20) + \hat{\beta}_2 \text{rendfam} = \hat{\beta}_0 + (20\hat{\beta}_1)\text{maços} + \hat{\beta}_2 \text{rendfam}.$$

Portanto, o intercepto e o coeficiente de inclinação de rendfam não se alteraram, mas o coeficiente de maços é 20 vezes o de cigs . Isso é intuitivamente atraente. Os resultados da regressão de pesonas sobre maços e rendfam estão na coluna (3) da Tabela 6.1. A propósito, lembre-se de que não teria sentido incluir tanto cigs como maços na mesma equação; isso induziria à multicolinearidade perfeita e não teria nenhum significado interessante.

Na equação original sobre o peso dos recém-nascidos (6.1), suponha que rendfam seja medida em dólares em lugar de milhares de dólares. Desse modo, defina a variável $\text{rendfam dol} = 1.000 \cdot \text{rendfam}$. Como mudam as estatísticas MQO quando rendfam dol substitui rendfam ? Para o propósito de apresentar os resultados da regressão, você acha melhor medir a renda em dólares ou em milhares de dólares?

Além do coeficiente de maços , existe outra estatística na coluna (3) que difere da mostrada na coluna (1): o erro-padrão de maços é 20 vezes maior que o de cigs na coluna (1). Isso significa que a estatística t para verificar a significância do hábito de fumar é a mesma, quer ele seja medido em cigarros ou em maços. Isso é natural.

O exemplo anterior explica claramente a maioria das possibilidades que surgem quando a variável dependente e as variáveis independentes são redimensionadas. O redimensionamento muitas

vezes é feito com os valores monetários em economia, especialmente quando os montantes são muito grandes.

No Capítulo 2, argumentamos que, se a variável dependente aparecer na forma logarítmica, a alteração na unidade de medida não afetará o coeficiente de inclinação. Isso também acontece aqui: a alteração na unidade de medida da variável dependente, quando aparece na forma logarítmica, não afeta qualquer das estimativas de inclinação. Isso resulta do simples fato de que $\log(c_1 y_i) = \log(c_1) + \log(y_i)$ para qualquer constante $c_1 > 0$. O novo intercepto será $\log(c_1) + \hat{\beta}_0$. De forma semelhante, a alteração da unidade de medida de qualquer x_j , onde $\log(x_j)$ aparece na regressão, afeta somente o intercepto. Isso corresponde ao que conhecemos sobre alterações em porcentagens e, em particular, em elasticidades: elas não sofrem alterações quando mudam as unidades de medida de y ou de x_j . Por exemplo, se tivéssemos especificado a variável dependente em (6.1) como $\log(pesonas)$, estimássemos a equação, e depois a tivéssemos reestimado com $\log(pesonaslb)$ como a variável dependente, os coeficientes de $cigs$ e $rendfam$ seriam os mesmos em ambas as regressões; somente o intercepto seria diferente.

Os Coeficientes Beta

Algumas vezes, em aplicações econometrísticas, uma variável-chave é medida em uma dimensão de difícil interpretação. Economistas especializados na área de trabalho freqüentemente incluem a pontuação de testes de conhecimentos em equações salariais, e a dimensão em que tais testes são registrados muitas vezes é arbitrária e de difícil interpretação (pelo menos para os economistas!). Em quase todos os casos estamos interessados em saber como a pontuação de um indivíduo em particular se compara com a população. Assim, em lugar de perguntarmos a respeito do efeito sobre o salário por hora se, digamos, a pontuação do teste for dez pontos mais alta, faz mais sentido perguntar o que acontece quando a pontuação do teste for um desvio-padrão mais alto.

Nada impede que vejamos o que acontece com a variável dependente quando uma variável independente em um modelo estimado aumenta certo número de desvios-padrão, supondo que tenhamos obtido o desvio-padrão da amostra (o que é fácil na maioria dos programas de regressão). Geralmente, essa é uma boa idéia. Assim, por exemplo, quando observamos o efeito de uma pontuação de teste padronizada, como o SAT (nota de ingresso em curso superior nos Estados Unidos), sobre a nota média em curso superior, podemos encontrar o desvio-padrão de SAT e verificar o que acontece quando essa pontuação aumenta em um ou dois desvios-padrão.

Algumas vezes é útil obter resultados de regressão quando todas as variáveis envolvidas, a dependente e todas as independentes, tenham sido padronizadas. Uma variável é padronizada em uma amostra pela subtração de sua média e dividindo o resultado por seu desvio-padrão (veja Apêndice C disponível no site do livro, no site www.thomsonlearning.com.br). Isso significa que computamos a transformação z de cada variável na amostra. Depois, fazemos a regressão usando os valores de z .

Por que a padronização é útil? É mais fácil começarmos com a equação MQO original, com as variáveis em suas formas originais:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik} + \hat{u}_i. \quad (6.2)$$

Incluímos o subscrito de observação i para enfatizar que nossa padronização é aplicada a todos os valores da amostra. Agora, se ao calcularmos a média de (6.2), usarmos o fato de que \hat{u}_i tem uma média de amostra zero, e subtrairmos o resultado de (6.2), temos

$$y_i - \bar{y} = \hat{\beta}_1(x_{i1} - \bar{x}_1) + \hat{\beta}_2(x_{i2} - \bar{x}_2) + \dots + \hat{\beta}_k(x_{ik} - \bar{x}_k) + \hat{u}_i$$

Em seguida, definamos $\hat{\sigma}_y$ como o desvio-padrão da amostra da variável dependente, $\hat{\sigma}_1$ como o dp da amostra da x_1 , $\hat{\sigma}_2$ como o dp da amostra de x_2 , e assim sucessivamente. Agora, um pouco de álgebra produz a equação

$$(y_i - \bar{y})/\hat{\sigma}_y = (\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1[(x_{i1} - \bar{x}_1)/\hat{\sigma}_1] + \dots \\ + (\hat{\sigma}_k/\hat{\sigma}_y)\hat{\beta}_k[(x_{ik} - \bar{x}_k)/\hat{\sigma}_k] + (\hat{u}_i/\hat{\sigma}_y). \quad (6.3)$$

Cada variável em (6.3) foi padronizada pela substituição de suas médias por seus valores de z , e isso resultou em novos coeficientes de inclinação. Por exemplo, o coeficiente de inclinação de $(x_{i1} - \bar{x}_1)/\hat{\sigma}_1$ é $(\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1$. Isso é simplesmente o coeficiente original, $\hat{\beta}_1$, multiplicado pela razão do desvio-padrão de x_1 sobre o desvio-padrão de y . O intercepto simplesmente desapareceu.

É útil reescrever (6.3), eliminando o subscrito i , como

$$z_y = \hat{\beta}_1 z_1 + \hat{\beta}_2 z_2 + \dots + \hat{\beta}_k z_k + erro, \quad (6.4)$$

onde z_y é o valor de z de y , z_1 é o valor de z de x_1 , e assim por diante. Os novos coeficientes são

$$\hat{\beta}_j = (\hat{\sigma}_j/\hat{\sigma}_y)\hat{\beta}_j \text{ para } j = 1, \dots, k. \quad (6.5)$$

Esses $\hat{\beta}_j$ são tradicionalmente chamados de **coeficientes padronizados ou coeficientes beta**. (Esta última denominação é mais comum, mas um pouco inadequada, já que temos usado o beta chapéu para representar as estimativas MQO *usuais*.)

Os coeficientes beta recebem seus interessantes significados a partir da equação (6.4): Se x_j aumentar em um desvio-padrão, \hat{y} , então, será alterado em $\hat{\beta}_j$ desvios-padrão. Assim, estamos medindo os efeitos não em termos das unidades originais de y ou de x_j , mas em unidades de desvios-padrão. Como isso torna a dimensão dos regressores irrelevante, essa equação coloca as variáveis explicativas em pé de igualdade. Em uma equação MQO padrão, não é possível simplesmente verificar o tamanho dos diferentes coeficientes e concluir que a variável explicativa com o maior coeficiente é “a mais importante”. Acabamos de ver que a magnitude dos coeficientes pode ser mudada à vontade pela alteração das unidades de medida das variáveis x_j . Mas, quando cada x_j é padronizado, a comparação das magnitudes dos coeficientes beta resultantes é mais convincente.

Para obter os coeficientes beta, podemos sempre padronizar y , x_1, \dots, x_k e em seguida computar a regressão MQO do valor de z de y sobre os valores de z de x_1, \dots, x_k – no qual não é necessário incluir um intercepto, já que ele será zero. Isso pode ser tedioso com muitas variáveis independentes. Alguns programas econométricos produzem coeficientes beta com um simples comando. O exemplo seguinte ilustra o uso de coeficientes beta.

EXEMPLO 6.1**(Efeitos da Poluição sobre os Preços de Imóveis)**

Utilizamos os dados do Exemplo 4.5 (do arquivo HPRICE2.RAW) para ilustrar o uso de coeficientes beta. Lembre-se de que a principal variável independente é *oxn*, uma medida do óxido nitroso no ar em cada comunidade. Uma maneira de entender o tamanho do efeito da poluição – sem entrar na questão científica do efeito do óxido de nitrogênio sobre a qualidade do ar é computar os coeficientes beta. (O Exemplo 4.5 contém um método alternativo: obtivemos uma elasticidade-preço em relação a *oxn* usando *preço* e *oxn* em forma logarítmica.)

A equação populacional é o modelo nível-nível

$$\text{preço} = \beta_0 + \beta_1 \text{oxn} + \beta_2 \text{crime} + \beta_3 \text{comods} + \beta_4 \text{dist} + \beta_5 \text{razestud} + u,$$

onde todas as variáveis exceto *crime* foram definidas no Exemplo 4.5; *crime* é o número de crimes registrados *per capita*. Os coeficientes beta aparecem na seguinte equação (portanto cada variável foi convertida ao seu valor de *z*):

$$z\text{preço} = -0,340 z\text{oxn} - 0,143 z\text{crime} + 0,514 z\text{comods} - 0,235 z\text{dist} - 0,270 z\text{razestud}.$$

Esta equação mostra que o aumento de um desvio-padrão em *oxn* reduz o preço em 0,34 desvio-padrão; o aumento de um desvio-padrão em *crime* reduz o preço em 0,14 desvio-padrão. Assim, o mesmo movimento relativo da poluição na população tem um efeito maior sobre os preços dos imóveis do que o da criminalidade. O tamanho do imóvel, medido pelo número de cômodos (*comods*), tem o maior efeito padronizado. Se quisermos saber os efeitos de cada variável independente sobre o valor da mediana dos preços dos imóveis, teremos que usar as variáveis não padronizadas.

O uso de variáveis padronizadas ou não padronizadas não afetará a significância estatística: as estatísticas *t* serão as mesmas, em ambos os casos.

6.2 UM POCO MAIS SOBRE A FORMA FUNCIONAL

Em vários dos exemplos anteriores, encontramos o artifício mais comum em econometria para permitir relações não lineares entre a variável explicada e as variáveis explicativas: o uso de logaritmos das variáveis dependentes ou independentes. Também vimos modelos contendo os quadrados de algumas variáveis explicativas, mas ainda precisamos discorrer sobre um tratamento sistemático desses tópicos. Nesta seção, trataremos de algumas variações e extensões sobre formas funcionais que surgem freqüentemente em trabalhos aplicados.

Um pouco mais sobre o Uso de Formas Funcionais Logarítmicas

Começamos revendo como interpretar os parâmetros no modelo

$$\log(\text{preço}) = \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 \text{comods} + u, \quad (6.6)$$

onde essas variáveis são as mesmas do Exemplo 4.5. Lembre-se de que em todo o texto $\log(x)$ é o log natural de x . O coeficiente β_1 é a elasticidade do preço em relação a oxn (poluição). O coeficiente β_2 é a mudança em $\log(\text{preço})$, quando $\Delta comods = 1$; como vimos muitas vezes, quando multiplicada por 100, essa é a percentagem aproximada de mudança em preço. Lembre-se de que $100 \cdot \beta_2$ é algumas vezes chamado de semi-elasticidade do preço em relação a $comods$.

Quando estimamos utilizando os dados do arquivo HPRICE2.RAW, obtemos

$$\begin{aligned} \log(\hat{\text{preço}}) &= 9,23 - 0,718 \log(oxn) + 0,306 comods \\ (0,19) &\quad (0,66) \qquad \quad (0,019) \end{aligned} \quad (6.7)$$

$n = 506, R^2 = 0,514.$

Assim, quando oxn aumenta em 1%, preço cai em 0,718%, mantendo-se apenas $comods$ fixo. Quando $comods$ aumenta em um, preço aumenta em aproximadamente $100(0,306) = 30,6\%$.

A estimativa de que um cômodo a mais aumenta o preço em cerca de 30,6% acaba por ser de certa forma impreciso para esta aplicação. O erro de aproximação ocorre porque, como a mudança em $\log(y)$ se torna cada vez maior, a aproximação $\% \Delta y \approx 100 \cdot \Delta \log(y)$ se mostra cada vez mais imprecisa. Felizmente, existe um cálculo simples para computar a percentagem exata de mudança.

Para descrever o procedimento, consideremos o modelo estimado de forma geral

$$\hat{\log}(y) = \hat{\beta}_0 + \hat{\beta}_1 \log(x_1) + \hat{\beta}_2 x_2.$$

(A inclusão de variáveis independentes adicionais não altera o procedimento.) Agora, fixando x_1 , temos $\Delta \hat{\log}(y) = \hat{\beta}_2 \Delta x_2$. O uso de simples propriedades algébricas das funções exponenciais e logarítmicas produz a percentagem exata de mudança no y estimado como

$$\% \hat{\Delta}y = 100 \cdot [\exp(\hat{\beta}_2 \Delta x_2) - 1], \quad (6.8)$$

onde a multiplicação por 100 transforma a mudança proporcional em uma mudança percentual. Quando $\Delta x_2 = 1$,

$$\% \hat{\Delta}y = 100 \cdot [\exp(\hat{\beta}_2) - 1]. \quad (6.9)$$

Aplicada ao exemplo dos preços dos imóveis com $x_2 = comods$ e $\hat{\beta}_2 = 0,306$, $\% \Delta \text{preço} = 100[\exp(0,306) - 1] = 35,8\%$, que é visivelmente maior do que a percentagem aproximada de mudança, 30,6%, obtida diretamente de (6.7). {A propósito, esse não é um estimador não-viesado, pois $\exp(\cdot)$ é uma função não-linear; ele é, porém, um estimador consistente de $100[\exp(\beta_2) - 1]$. Isso é assim porque o limite de probabilidade é calculado por meio de funções contínuas, enquanto o operador valor esperado não é calculado dessa forma. Veja Apêndice C, no site da Thomson.}

O ajuste na equação (6.8) não é tão crucial para pequenas mudanças percentuais. Por exemplo, quando incluímos na equação (6.7) a relação aluno-professor, seu coeficiente estimado é $-0,052$, o que significa que, se *razestud* aumentar em um, *preço* diminui em aproximadamente 5,2%. A mudança proporcional exata é $\exp(-0,052) - 1 \approx -0,051$, ou $-5,1\%$. De outro lado, se aumentarmos *razestud* em

cinco, então a mudança percentual aproximada em *preço* será -26% , enquanto a mudança exata obtida da equação (6.8) é $100[\exp(-0,26) - 1] \approx -22,9\%$.

Vimos que o uso de logs naturais leva a coeficientes com interpretações interessantes e podemos ignorar o fato de as unidades de medida das variáveis aparecerem em forma logarítmica, pois os coeficientes de inclinação são invariantes em relação a redimensionamentos. Existem várias outras razões pelas quais os logs são tão usados em trabalhos aplicados. Em primeiro lugar, quando $y > 0$, os modelos que usam $\log(y)$ como a variável dependente geralmente satisfazem as hipóteses do MLC mais apropriadamente do que os modelos que usam o nível de y . Variáveis estritamente positivas freqüentemente possuem distribuições condicionais que são heteroscedásticas ou concentradas; o uso do log pode aliviar, se não eliminar, ambos os problemas.

Além disso, o uso de logs normalmente estreita a amplitude dos valores das variáveis, em alguns casos em quantidade considerável. Isso torna as estimativas menos sensíveis a observações díspares (ou extremas) na variável dependente ou nas variáveis independentes. Abordaremos a questão das observações extremas no Capítulo 9.

Existem algumas regras práticas padronizadas para o uso de logs, embora nenhuma definitiva. Quando a variável é um valor monetário positivo, ele freqüentemente é transformado em log. Temos visto isso para variáveis como salários, vendas de empresas e valores de mercado das empresas. Variáveis como população, número total de empregados e matrículas escolares freqüentemente aparecem em forma logarítmica; elas têm a característica comum de serem grandes valores inteiros.

Variáveis que são medidas em anos — como educação, experiência, tempo de permanência, idade etc. — normalmente aparecem em sua forma original. Uma variável que seja uma proporção ou uma percentagem — como a taxa de desemprego, a taxa de participação em planos de aposentadoria, a taxa de estudantes aprovados em um exame padronizado e a taxa de detenção sobre crimes registrados — pode aparecer tanto em sua forma original como logarítmica, embora haja uma tendência em usá-la em forma de nível. Isso se deve ao fato de que quaisquer coeficientes de regressão envolvendo a variável *original* — seja ela a variável dependente ou independente — terão uma interpretação de mudança de *pontos percentuais*. (Veja Apêndice A, no site da Thomson, para uma revisão sobre a distinção entre mudança percentual e mudança de pontos percentuais.) Se usarmos, digamos, $\log(\text{desemp})$ em uma regressão, onde *desemp* é a percentagem de indivíduos desempregados, precisaremos ter muito cuidado para distinguir entre uma mudança de pontos percentuais e uma mudança percentual. Lembre-se, quando *desemp* aumenta de oito para nove, isso é um acréscimo de um ponto percentual, equivalente a um incremento de 12,5% sobre o nível de desemprego inicial. O uso do log significa que queremos saber a mudança percentual da taxa de desemprego: $\log(9) - \log(8) \approx 0,118$ ou 11,8%, que é a aproximação logarítmica do aumento efetivo de 12,5%.

Suponha que o número anual de prisões por direção de veículo sob embriaguez¹ seja determinado por

$$\log(prisões) = \beta_0 + \beta_1 \log(pop) + \beta_2 idade16_25 + \text{outros fatores},$$

onde *idade16_25* é a proporção da população entre 16 e 25 anos de idade. Mostre que β_2 tem a seguinte interpretação (*ceteris paribus*): ela é a mudança percentual em *prisões* quando a percentagem da população com idade entre 16 e 25 anos aumenta em um ponto percentual.

¹ NE: Nos Estados Unidos.

Uma limitação do log é que ele não pode ser usado, caso uma variável assuma valor zero ou negativo. Em casos nos quais a variável y não seja negativa, mas pode assumir o valor 0, $\log(1 + y)$ é algumas vezes usado. As interpretações de mudança percentual são, em geral, estritamente preservadas, exceto para mudanças começando em $y = 0$ (em que a percentagem de mudança não é sequer definida). Geralmente, usar $\log(1 + y)$ e depois interpretar as estimativas como se a variável fosse $\log(y)$ é aceitável quando os dados em y não são dominados por zeros. Um exemplo pode ser o de y representar horas de treinamento por funcionário da população industrial, quando uma grande fração das empresas oferece treinamento a, pelo menos, um de seus funcionários.

Uma desvantagem de usar uma variável dependente na forma logarítmica está na dificuldade de se prever a variável original. O modelo original nos permite prever $\log(y)$, e não y . No entanto, é razoavelmente fácil transformar uma previsão de $\log(y)$ em uma previsão de y (veja Seção 6.4). Uma questão relacionada é que não é válido comparar R -quadrados de modelos nos quais y é a variável dependente em um caso e $\log(y)$ é a variável dependente no outro. Essas medidas explicam variações em diferentes variáveis. Discutimos como computar medidas comparáveis de graus de ajuste na Seção 6.4.

Modelos com Funções Quadráticas

As funções quadráticas também são usadas com bastante freqüência em economia aplicada para capturar efeitos marginais crescentes ou decrescentes. Seria interessante rever as propriedades das funções quadráticas no Apêndice A, no site da Thomson.

No caso mais simples, y depende de um único fator observado x , mas de uma forma quadrática:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u.$$

Por exemplo, considere $y = \text{salário}$ e $x = \text{exper}$. Como discutimos no Capítulo 3, esse modelo não se enquadra na análise de regressão simples, mas é facilmente trabalhado em regressão múltipla.

É importante lembrar que β_1 não mede a mudança em y em relação a x ; não faz sentido manter x^2 fixo quando se altera x . Se escrevermos a equação estimada como

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2, \quad (6.10)$$

teremos a aproximação

$$\Delta\hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x)\Delta x, \text{ e assim } \Delta\hat{y}/\Delta x \approx \hat{\beta}_1 + 2\hat{\beta}_2 x. \quad (6.11)$$

Isso nos mostra que a inclinação da relação entre x e y depende do valor de x ; a inclinação estimada é $\hat{\beta}_1 + 2\hat{\beta}_2 x$. Se inserirmos $x = 0$, veremos que $\hat{\beta}_1$ pode ser interpretado como a inclinação aproximada na alteração de $x = 0$ para $x = 1$. Para outras mudanças no valor de x , o segundo termo, $2\hat{\beta}_2 x$, deve ser levado em conta.

Se estivermos interessados em somente computar a mudança prevista em y dado um valor inicial de x e uma mudança de x , poderíamos usar (6.10) diretamente: não há nenhuma razão para usar cálculos de aproximação. Porém, normalmente estamos mais interessados em resumir rapidamente o efeito de x em y , e a interpretação de $\hat{\beta}_1$ e de $\hat{\beta}_2$ na equação (6.11) fornece esse resumo. Em geral, podemos

inserir o valor médio de x na amostra, ou outros valores de interesse, como a mediana ou os valores dos quartis inferior ou superior de x .

Em muitas aplicações, $\hat{\beta}_1$ é positivo, e $\hat{\beta}_2$ é negativo. Por exemplo, utilizando os dados de salários contidos no arquivo WAGE1.RAW, obtemos

$$\begin{aligned} \text{salárioh} &= 3,73 + 0,298 \text{ exper} - 0,0061 \text{ exper}^2 \\ &\quad (0,35) (0,041) \quad (0,0009) \quad [6.12] \\ n &= 526, R^2 = 0,093. \end{aligned}$$

A equação estimada sugere que exper tem um efeito de redução sobre salárioh . O primeiro ano de experiência vale aproximadamente 30 centavos de dólar por hora (0,298 dólares). O segundo ano de experiência vale menos [cerca de $0,298 - 2(0,0061)(1) \approx 0,286$, ou 28,6 centavos de dólar, de acordo com a aproximação em (6.11) com $x = 1$]. Aumentando de 10 para 11 os anos de experiência, a previsão de aumento do salário-hora é de cerca de $0,298 - 2(0,0061)(10) \approx 0,176$ ou 17,6 centavos de dólar. E assim por diante.

Quando o coeficiente de x é positivo e o coeficiente de x^2 é negativo, a função quadrática tem um formato parabólico. Sempre existe um valor positivo de x , no qual o efeito de x sobre y é zero; antes desse ponto, x tem um efeito positivo sobre y ; após esse ponto, x tem um efeito negativo sobre y . Na prática, pode ser importante saber onde fica esse ponto crítico.

Na equação estimada (6.10) com $\hat{\beta}_1 > 0$ e $\hat{\beta}_2 < 0$, esse ponto crítico (ou o máximo da função) é sempre alcançado na relação entre o coeficiente de x e duas vezes o valor absoluto do coeficiente de x^2 :

$$x^* = |\hat{\beta}_1 / (2\hat{\beta}_2)|. \quad [6.13]$$

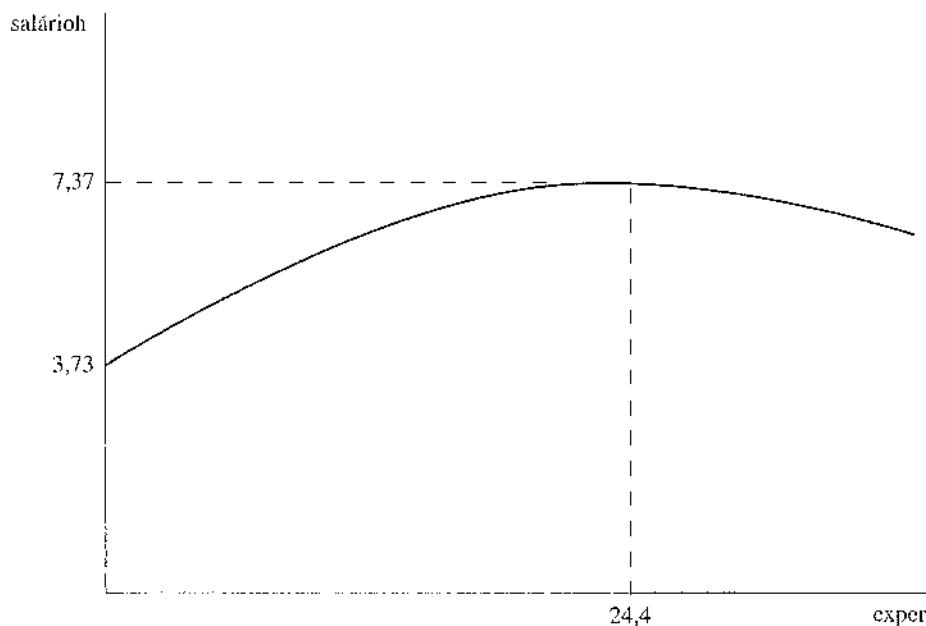
No exemplo dos salários, $x^* = \text{exper}^*$ é $0,298/[2(0,0061)] \approx 24,4$. (Observe como simplesmente eliminamos o sinal de menos em $-0,0061$ ao fazermos esse cálculo.) Esta relação quadrática está ilustrada na Figura 6.1.

Na equação dos salários (6.12), o retorno da experiência passa a ser zero por volta dos 24,4 anos. O que devemos concluir disso? Existem, pelo menos, três explicações possíveis. Primeiro, pode ser que poucas pessoas na amostra tenham mais de 24 anos de experiência, e assim a parte da curva à direita de 24 pode ser ignorada. A consequência de se usar uma função quadrática para capturar efeitos decrescentes é que a partir de certo ponto ela acabará fazendo um movimento inverso. Se esse ponto estiver além de uma pequena percentagem das pessoas na amostra, isso não será motivo para grande preocupação. Mas no conjunto de dados do arquivo WAGE1.RAW, cerca de 28% das pessoas na amostra têm mais de 24 anos de experiência; essa é uma percentagem alta demais para se ignorar.

É possível que o retorno de exper realmente se torne negativo em algum ponto, mas é difícil acreditar que isso aconteça aos 24 anos de experiência. Uma possibilidade mais provável é que o efeito estimado de exper sobre salárioh seja viesado, por não termos controlado outros fatores ou porque a relação funcional entre salárioh e exper na equação (6.12) não está totalmente correta. O Problema 6.9 pede que você explore essa possibilidade controlando a educação, além de usar $\log(\text{salárioh})$ como a variável dependente.

Figura 6.1

Relação quadrática entre salárioh e exper.



Quando um modelo tem uma variável dependente na forma logarítmica e uma variável explicativa como uma função quadrática, é necessário certo cuidado para fazer uma boa interpretação. O exemplo seguinte também mostra que a função quadrática pode ter um formato em U, em vez de uma forma parabólica. A forma em U surge na equação (6.10) quando $\hat{\beta}_1$ é negativo e $\hat{\beta}_2$ é positivo; isso captura um efeito crescente de x sobre y .

EXEMPLO 6.2**(Efeitos da Poluição sobre os Preços dos Imóveis)**

Modificamos o modelo dos preços dos imóveis do Exemplo 4.5 para incluir um termo quadrático em comods:

$$\begin{aligned} \log(\text{preço}) = & \beta_0 + \beta_1 \log(\text{oxn}) + \beta_2 \log(\text{dist}) + \beta_3 \text{comods} \\ & + \beta_4 \text{comods}^2 + \beta_5 \text{razestud} + u. \end{aligned} \quad (6.14)$$

O modelo estimado utilizando os dados contidos no arquivo HPRICE2.RAW é

$$\begin{aligned} \log(\hat{\text{preço}}) = & 13,39 - 0,902 \log(\text{oxn}) - 0,087 \log(\text{dist}) \\ & (0,57) \quad (0,115) \quad (0,043) \end{aligned}$$

EXEMPLO 6.2 (continuação)

$$= 0,545 \text{ comods} + 0,062 \text{ comods}^2 - 0,48 \text{ razestud}$$

(0,165) (0,013) (0,006)

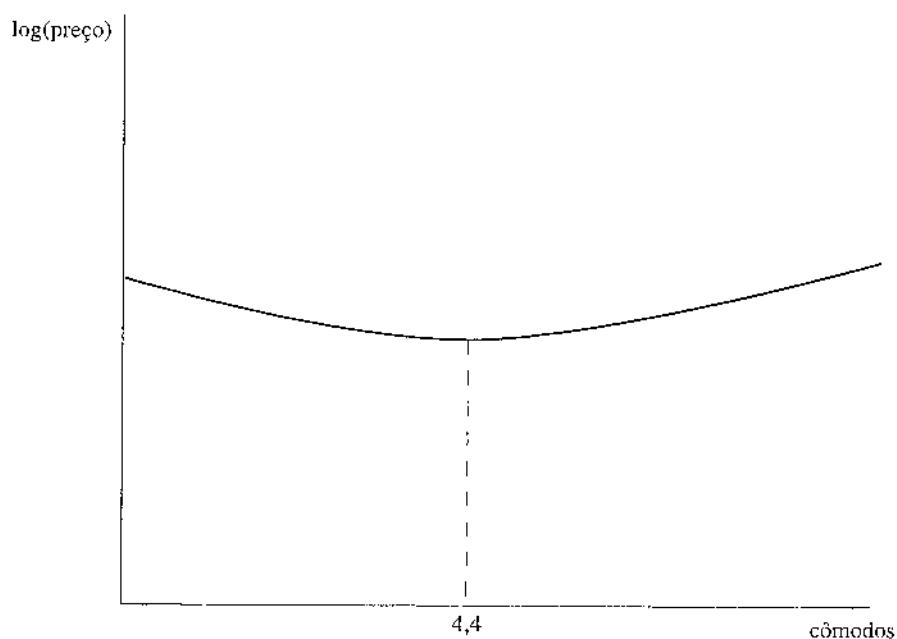
$n = 506, R^2 = 0,603.$

O termo quadrático comods^2 tem uma estatística t em torno de 4,77 e, portanto, é estatisticamente bastante significante. Mas o que é possível afirmar sobre a interpretação do efeito de comods sobre $\log(\text{preço})$? Inicialmente, o efeito parece ser estranho. Como o coeficiente de comods é negativo e o coeficiente de comods^2 é positivo, a equação literalmente sugere que, com valores baixos de comods , um cômodo adicional tem um efeito negativo sobre $\log(\text{preço})$. Em algum ponto, o efeito se torna positivo, e a forma quadrática significa que a semi-elasticidade de preço em relação a comods cresce na mesma proporção do crescimento de comods . Esta situação é mostrada na Figura 6.2.

Obtemos o valor do ponto crítico de comods usando a equação (6.13) (embora $\hat{\beta}_1$ seja negativo e $\hat{\beta}_2$ seja positivo). O valor absoluto do coeficiente de comods , 0,545, dividido pelo dobro do coeficiente de comods^2 , 0,062, resulta em $\text{comods}' = 0,545/[2(0,062)] \approx 4,4$; este ponto está marcado na Figura 6.2.

Figura 6.2

$\log(\text{preço})$ como uma função quadrática de comods .



EXEMPLO 6.2 {continuação}

Será que podemos acreditar que se iniciarmos com três cômodos e aumentarmos para quatro isso efetivamente reduzirá o valor esperado do imóvel? Provavelmente não. Acontece que somente cinco das 506 comunidades na amostra possuem imóveis com média de 4,4 cômodos ou menos, cerca de 1% da amostragem. Isso é tão pequeno que a função quadrática à esquerda de 4,4 pode, para fins práticos, ser ignorada. À direita de 4,4, vemos que a adição de outro cômodo tem um efeito crescente na mudança percentual no preço:

$$\Delta \log(\hat{preço}) \approx \{[-0,545 + 2(0,062)]comods\} \Delta comods$$

e assim

$$\begin{aligned}\% \Delta \hat{preço} &\approx 100 \{[-0,545 + 2(0,062)]comods\} \Delta comods \\ &= (-54,5 + 12,4comods) \Delta comods.\end{aligned}$$

Portanto, um aumento em *comods*, digamos de cinco para seis, aumenta o preço em aproximadamente $-54,5 + 12,4(5) = 7,5\%$; o aumento de seis para sete aumenta o preço em aproximadamente $-54,5 + 12,4(6) = 19,9\%$. Esse é um efeito crescente bastante forte.

Existem muitas outras possibilidades de usar funções quadráticas juntamente com logaritmos. Por exemplo, uma extensão de (6.14) que permita uma elasticidade não-constante entre *preço* e *oxn* é

$$\begin{aligned}\log(\hat{preço}) &= \beta_0 + \beta_1 \log(oxn) + \beta_2 [\log(oxn)]^2 \\ &+ \beta_3 crime + \beta_4 comods + \beta_5 comods^2 + \beta_6 razestud + u.\end{aligned}\quad (6.15)$$

Se $\beta_2 = 0$, β_1 será a elasticidade do *preço* em relação a *oxn*. Caso contrário, essa elasticidade dependerá do nível de *oxn*. Para verificar isso, podemos combinar os argumentos dos efeitos parciais nos modelos quadrático e logarítmico para mostrar que

$$\% \Delta \hat{preço} \approx [\beta_1 + 2\beta_2 \log(oxn)] \% \Delta oxn; \quad (6.16)$$

portanto, a elasticidade do *preço* em relação a *oxn* é $\beta_1 + 2\beta_2 \log(oxn)$, de forma que ela depende de $\log(oxn)$.

Finalmente, outros termos polinomiais podem ser incluídos nos modelos de regressão. Certamente a função quadrática é vista com mais freqüência, mas um termo cúbico ou até de quarta potência aparece de vez em quando. Uma forma funcional freqüentemente razoável de uma função de custo total é

$$custo = \beta_0 + \beta_1 quantidade + \beta_2 quantidade^2 + \beta_3 quantidade^3 + u.$$

Não é complicado estimar este modelo. A interpretação dos parâmetros é mais trabalhosa (embora objetiva com o uso de cálculo infinitesimal); não estudaremos este modelo com mais detalhes.

Modelos com Termos de Interação

Algumas vezes, é natural que o efeito parcial, a elasticidade, ou a semi-elasticidade da variável dependente, em relação a uma variável explicativa, dependa da magnitude de *outra* variável explicativa. Por exemplo, no modelo

$$\text{preço} = \beta_0 + \beta_1 \text{arquad} + \beta_2 \text{qtdorm} + \beta_3 \text{arquad} \cdot \text{qtdorm} + \beta_4 \text{banhos} + u,$$

o efeito parcial de *qtdorm* sobre *preço* (mantendo fixas todas as outras variáveis) é

$$\frac{\Delta \text{preço}}{\Delta \text{qtdorm}} = \beta_2 + \beta_3 \text{arquad}. \quad (6.17)$$

Se $\beta_3 > 0$, (6.17) sugere que um quarto a mais produz um aumento maior no preço dos imóveis maiores. Em outras palavras, existe um **efeito de interação** entre a área do imóvel e o número de quartos. Ao resumirmos o efeito de *qtdorm* sobre *preço*, devemos avaliar (6.17) quanto aos valores de interesse de *arquad*, como o valor médio, ou os quartis inferior ou superior na amostra. Se β_3 é zero ou não, é algo que podemos verificar facilmente.

Pode ser complicado interpretar os parâmetros das variáveis originais quando incluímos um termo de interação. Por exemplo, na equação anterior sobre preços de imóveis, a equação (6.17) mostra que β_2 é o efeito de *qtdorms* sobre *preço* para um preço com zero de área construída! Esse efeito, obviamente, não é muito interessante. Em vez disso, devemos ser cuidadosos ao inserirmos valores de interesse da área do imóvel, como o valor médio ou a mediana da amostra, na versão estimada da equação (6.17).

Freqüentemente, é vantajoso reparametrizar um modelo para que os coeficientes das variáveis originais tenham significados interessantes. Considere um modelo com duas variáveis explicativas e uma interação:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u.$$

Como acabamos de mencionar, β_2 é o efeito parcial de x_2 quando $x_1 = 0$. Muitas vezes, isso não é de interesse. Em vez disso, podemos reparametrizar o modelo como

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3(x_1 - \mu_1)(x_2 - \mu_2) + u,$$

onde μ_1 é a média populacional de x_1 , e μ_2 é a média populacional de x_2 . Podemos facilmente ver que agora o coeficiente de x_2 , δ_2 , é o efeito parcial de x_2 sobre y no valor médio de x_1 . (Multiplicando a interação na segunda equação e comparando os coeficientes, podemos mostrar com facilidade que $\delta_2 = \beta_2 + \beta_3 \mu_1$. O parâmetro δ_1 tem uma interpretação semelhante.) Portanto, se subtraímos as médias das variáveis — na prática, elas seriam, tipicamente, as médias da amostra — antes de criarmos o termo de interação, os coeficientes das variáveis originais terão uma interpretação útil. E mais, obteremos imediatamente os erros-padrão dos efeitos parciais ao nível dos valores médios. Nada nos impede de substituir μ_1 ou μ_2 por outros valores das variáveis explicativas que possam ser de interesse. O exemplo seguinte ilustra como podemos usar os termos de interação.

EXEMPLO 6.3**(Efeitos da Freqüência Escolar no Desempenho de Exames Finais)**

Um modelo para explicar o resultado padronizado de um exame final (*respad*) em termos da taxa de freqüência escolar, da nota média anterior* ao curso superior até o penúltimo semestre, e da nota do teste de avaliação de conhecimentos para ingresso em curso superior é

$$\begin{aligned} \text{respad} = & \beta_0 + \beta_1 \text{taxafreq} + \beta_2 \text{nmgradp} + \beta_3 \text{tac} + \beta_4 \text{nmgradp}^2 \\ & + \beta_5 \text{tac}^2 + \beta_6 \text{nmgradp} \cdot \text{taxafreq} + u. \end{aligned} \quad (6.18)$$

(Utilizamos o resultado padronizado do exame pelos motivos discutidos na Seção 6.1: é mais fácil interpretar o desempenho de um aluno em relação ao restante da classe.) Além dos termos quadráticos em *nmgradp* e *tac*, este modelo inclui uma interação entre *nmgradp* e *taxafreq*. A idéia é que a freqüência às aulas pode ter um efeito diferente nos alunos que tiveram desempenhos diferentes no passado, como medido por *nmgradp*. Estamos interessados nos efeitos da freqüência sobre as notas do exame final: $\Delta\text{respad}/\Delta\text{taxafreq} = \beta_1 + \beta_6 \text{nmgradp}$.

Usando as 680 observações do arquivo ATTEND.RAW para estudantes da área de economia, a equação estimada é

$$\begin{aligned} \hat{\text{respad}} = & 2,05 - 0,0067 \text{taxafreq} - 1,63 \text{nmgradp} - 0,128 \text{tac} \\ & (1,36) \quad (0,0102) \quad (0,48) \quad (0,098) \\ & + 0,296 \text{nmgradp}^2 + 0,0045 \text{tac}^2 + 0,0056 \text{nmgradp} \cdot \text{taxafreq} \\ & (0,101) \quad (0,0022) \quad (0,0043) \\ n = & 680, R^2 = 0,229, \bar{R}^2 = 0,222. \end{aligned} \quad (6.19)$$

Devemos interpretar essa equação com extremo cuidado. Se simplesmente olharmos o coeficiente de *taxafreq*, concluiremos de forma errônea que a freqüência tem um efeito negativo na nota do exame final. Porém, esse coeficiente supostamente mede o efeito quando *nmgradp* = 0, o que não é interessante (nessa amostra, a menor nota média do ensino médio é cerca de 0,86). Também devemos ter cuidado para não examinarmos separadamente as estimativas de β_1 e β_6 e concluirmos que, como cada estatística *t* é não significante, não podemos rejeitar $H_0: \beta_1 = 0, \beta_6 = 0$. Aliás, o *p*-valor do teste *F* dessa hipótese conjunta é 0,014, de modo que com certeza rejeitamos H_0 ao nível de 5%. Este é um bom exemplo de como o exame em separado de estatísticas *t*, quando estamos testando uma hipótese conjunta, pode nos levar a equívocos.

Como devemos estimar o efeito parcial de *taxafreq* sobre *respad*? Devemos inserir valores de interesse de *nmgradp* para obter o efeito parcial. O valor médio de *nmgradp* na amostra é 2,59, de modo que nesse valor médio o efeito de *taxafreq* sobre *respad* é $-0,0067 + 0,0056(2,59) \approx 0,0078$. Qual o significado disso? Como *taxafreq* é medida como um percentual, isso significa que um aumento de dez pontos percentuais em *taxafreq* aumenta *respad* em 0,078 desvios-padrão da nota média do exame final.

Como podemos dizer se a estimativa 0,0078 é estatisticamente diferente de zero? Temos que computar novamente a regressão, substituindo $\text{nmgradp} \cdot \text{taxafreq}$ por $(\text{nmgradp} - 2,59) \cdot \text{taxafreq}$. Isso produz, como o novo coeficiente de *taxafreq*, o efeito estimado quando *nmgradp* = 2,59, juntamente com seu erro-padrão; nada mais é alterado na regressão. (Descrevemos esse mecanismo na Seção 4.4.) A execução dessa nova regressão fornece o erro-padrão de $\hat{\beta}_1 + \hat{\beta}_6(2,59) = 0,0078$ como 0,0026, o que produz

EXEMPLO 6.3 [continuação]

$t = 0,0078/0,026 = 3$. Portanto, na $nmgradp$ média, concluímos que a taxa de freqüência às aulas tem um efeito positivo estatisticamente significante nas notas do exame final.

É ainda mais complicado encontrar o efeito de $nmgradp$ sobre $respad$, devido ao termo quadrático $nmgradp^2$. Para encontrar o efeito no valor médio de $nmgradp$ e na taxa média de freqüência, 0,82, substituímos $nmgradp^2$ por $(nmgradp - 2,59)^2$ e $nmgradp \cdot taxafreq$ por $nmgradp \cdot (taxafreq - 0,82)$. O coeficiente de $nmgradp$ se tornará o efeito parcial nos valores médios e obteremos seu erro-padrão. (Veja o Problema 6.14.)

Se adicionarmos o termo $\beta_7 \cdot taxafreq$ à equação (6.18), qual será o efeito parcial de $taxafreq$ sobre $respad$?

6.3 UM POUCO MAIS SOBRE O GRAU DE AJUSTE E A SELEÇÃO DE REGRESSORES

Até agora, não dedicamos muita atenção ao tamanho do R^2 na avaliação de nossos modelos de regressão, porque estudantes iniciantes tendem a colocar muito peso no R -quadrado. Como em breve veremos, a seleção de um conjunto de variáveis explicativas com base no tamanho do R -quadrado pode levar a modelos absurdos. No Capítulo 10 descobriremos que R -quadrados obtidos de regressões de séries temporais podem ser artificialmente altos e podem resultar em conclusões enganosas.

Nada nas hipóteses do modelo linear clássico exige que o R^2 esteja acima de qualquer valor em particular; o R^2 é simplesmente uma estimativa do quanto da variação em y é explicado por x_1, x_2, \dots, x_k na população. Vimos várias regressões que tinham R -quadrados bastante pequenos. Embora isso signifique que não tenhamos avaliado vários fatores que afetam y , isso não quer dizer que os fatores em u sejam correlacionados com as variáveis independentes. A hipótese de média condicional zero RLM.3 é a que determina se obteremos estimadores não-viesados dos efeitos *ceteris paribus* das variáveis independentes, e o tamanho do R -quadrado não tem influência direta nisso.

Um R -quadrado pequeno sugere que a variância do erro é grande em relação à variância de y , o que significa que podemos ter muito trabalho para estimar β_j com precisão. Porém, lembre-se: vimos na Seção 3.4 que uma variância grande do erro pode ser compensada por uma amostra de tamanho grande: se tivermos dados suficientes, podemos ter condições de estimar com precisão os efeitos parciais, mesmo que não tenhamos controlado muitos dos fatores não-observados. Se podemos ou não obter estimativas suficientemente precisas, depende da aplicação que estamos pesquisando. Por exemplo, suponha que alguns alunos ingressantes de uma grande universidade recebam, aleatoriamente, subsídios para a compra de computadores. Se o montante do subsídio for determinado de forma realmente aleatória, podemos estimar o efeito *ceteris paribus* do montante do subsídio sobre a nota de aproveitamento dos alunos, com o uso de uma análise de regressão simples. (Devido à atribuição aleatória, todos os outros fatores que afetam a nota de aproveitamento seriam não-correlacionados com o montante do subsídio.) Parece provável que o montante de subsídio explique pouco da variação na nota de aproveitamento, de modo que o R -quadrado de tal regressão provavelmente será muito pequeno. Mas se tivermos uma amostra de grande tamanho, ainda poderemos ter condições de obter uma estimativa razoavelmente precisa do efeito do subsídio.

Lembre-se, porém, de que a *mudança* relativa no R -quadrado, quando variáveis são adicionadas à equação, é muito útil: a estatística F em (4.41) para testar a significância conjunta depende de forma crucial da diferença nos R -quadrados entre o modelo sem restrições e o modelo restrito.

O R -Quadrado Ajustado

A maioria dos programas econométricos registra, juntamente com o R -quadrado, uma estatística chamada **R -quadrado ajustado**. Como o R -quadrado ajustado é descrito em muitos trabalhos aplicados, e como ele tem algumas características úteis, trataremos dele nesta subseção.

Para verificar como o R -quadrado usual pode ser ajustado, é útil escrevê-lo como

$$R^2 = 1 - (\text{SQR}/n)/(\text{SQT}/n), \quad (6.20)$$

onde SQR é a soma dos resíduos quadrados e SQT é a soma dos quadrados total; comparada com a equação (3.28), tudo o que fizemos foi dividir tanto SQR como SQT por n . Essa expressão revela o que R^2 está realmente estimando. Defina σ_y^2 como a variância populacional de y e faça com que σ_u^2 represente a variância populacional do termo erro, u . (Até agora temos usado σ^2 para representar σ_u^2 , mas é vantajoso ser mais específico neste caso.) O **R -quadrado da população** é definido como $1 - \sigma_u^2/\sigma_y^2$; essa é a proporção da variação em y na população explicada pelas variáveis independentes. Isso é o que, supostamente, R^2 deve estar estimando.

O R^2 estima σ_u^2 por SQR/ n , que sabemos ser viesado. Então, por que não substituir SQR/ n por SQR/ $(n - k - 1)$? Além disso, podemos usar SQT/ $(n - 1)$ em lugar de SQT/ n , já que o primeiro é o estimador não-viesado de σ_y^2 . Usando esses estimadores, chegamos ao R -quadrado ajustado:

$$\begin{aligned} \bar{R}^2 &= 1 - [\text{SQR}/(n - k - 1)]/[\text{SQT}/(n - 1)] \\ &= 1 - \hat{\sigma}^2/[\text{SQT}/(n - 1)], \end{aligned} \quad (6.21)$$

já que $\hat{\sigma}^2 = \text{SQR}/(n - k - 1)$. Devido à notação usada para representar o R -quadrado ajustado, ele é, algumas vezes, chamado de *R-barra-quadrado*.

O R -quadrado ajustado algumas vezes é chamado de *R-quadrado corrigido*, mas esse não é um bom nome, pois sugere que \bar{R}^2 é de alguma forma melhor que R^2 como um estimador do R -quadrado da população. Infelizmente, \bar{R}^2 não é reconhecido, de forma geral, como um melhor estimador. É tentador imaginar que \bar{R}^2 corrige o viés de R^2 na estimativa do R -quadrado da população, mas ele não faz isso: a razão de dois estimadores não-viesados não é um estimador não-viesado.

O ponto mais atraente do \bar{R}^2 é que ele impõe uma penalidade à inclusão de variáveis independentes adicionais em um modelo. Sabemos que R^2 nunca pode diminuir quando uma nova variável independente é incluída em uma equação de regressão: isso ocorre porque SQR nunca aumenta (e normalmente diminui) quando novas variáveis independentes são adicionadas. Mas a fórmula do \bar{R}^2 mostra que ele depende explicitamente de k , o número de variáveis independentes. Se uma variável independente for adicionada a uma regressão, SQR diminui, mas o mesmo acontece com os gl na regressão, $n - k - 1$. Portanto, SQR/ $(n - k - 1)$ pode aumentar ou diminuir quando uma nova variável independente é adicionada a uma regressão.

Um fato algébrico interessante é o seguinte: se adicionarmos uma nova variável independente a uma equação de regressão, \bar{R}^2 aumenta se, e somente se, a estatística t da nova variável for maior que um em valor absoluto. (Uma extensão disto é que \bar{R}^2 aumenta quando um grupo de variáveis é adicionado a uma regressão se, e somente se, a estatística F da significância conjunta das novas variáveis for maior

que a unidade.) Assim, vemos imediatamente que usar o \bar{R}^2 para decidir se determinada variável independente (ou conjunto de variáveis) pertence a um modelo nos fornece uma resposta diferente daquelas fornecidas pelos testes usuais t ou F (já que uma estatística t ou F igual à unidade não é estatisticamente significante aos níveis tradicionais de significância).

Algumas vezes é útil termos uma fórmula do \bar{R}^2 em termos de R^2 . A álgebra simples mostra que

$$\bar{R}^2 = 1 - (1 - R^2)(n - 1)/(n - k - 1). \quad (6.22)$$

Por exemplo, se $R^2 = 0,30$, $n = 51$, e $k = 10$, então $\bar{R}^2 = 1 - 0,70(50)/40 = 0,125$. Assim, para n pequeno e k grande, \bar{R}^2 pode estar substancialmente abaixo de R^2 . De fato, se o R -quadrado normal for pequeno, e $n - k - 1$ for pequeno, \bar{R}^2 pode, na realidade, ser negativo! Por exemplo, podemos considerar $R^2 = 0,10$, $n = 51$, e $k = 10$ para verificar que $\bar{R}^2 = -0,125$. Um \bar{R}^2 negativo indica uma adaptação muito pobre do modelo relativamente ao número de graus de liberdade.

O R -quadrado ajustado algumas vezes é descrito junto com o R -quadrado habitual em regressões, e algumas vezes o R^2 é descrito em lugar do \bar{R}^2 . É importante lembrar que é o R^2 , e não o \bar{R}^2 , que aparece na estatística F em (4.41). A mesma fórmula com \bar{R}_r^2 e \bar{R}_{ir}^2 não é válida.

O Uso do R -quadrado Ajustado para a Escolha entre Modelos Não-Aninhados

Na Seção 4.5 aprendemos como calcular uma estatística F para testar a significância conjunta de um grupo de variáveis; isso nos possibilita decidir, em um nível particular de significância, se pelo menos uma variável no grupo afeta a variável dependente. Esse teste não nos permite decidir *qual* das variáveis tem um efeito. Em alguns casos, queremos escolher um modelo sem variáveis independentes redundantes, e o R -quadrado ajustado pode nos ajudar nessa tarefa.

No exemplo dos salários dos jogadores da principal liga de beisebol na Seção 4.5, vimos que nem *hrunano* nem *rebrunano* eram individualmente significantes. Essas duas variáveis são altamente correlacionadas, de modo que podemos querer optar entre os modelos

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{anos} + \beta_2 \text{jogosano} + \beta_3 \text{rebmed} + \beta_4 \text{hrunano} + u$$

e

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{anos} + \beta_2 \text{jogosano} + \beta_3 \text{rebmed} + \beta_4 \text{rebrunano} + u.$$

Esses dois exemplos são **modelos não-aninhados**, pois nenhuma equação é um caso especial da outra. A estatística F , que estudamos no Capítulo 4, nos permite testar somente modelos *aninhados*: um modelo (o modelo restrito) é um caso especial do outro modelo (o modelo sem restrições). Veja as equações (4.32) e (4.28) para exemplos dos modelos restritos e sem restrições. Uma possibilidade é criar um modelo combinado que contenha *todas* as variáveis explicativas dos modelos originais e depois testar cada modelo contra o modelo geral usando o teste F . O problema deste processo é que ambos os modelos poderão ser rejeitados, ou nenhum modelo poderá ser rejeitado (como acontece com o exemplo dos salários dos jogadores da principal liga de beisebol na Seção 4.5). Assim, esse processo nem sempre fornece uma maneira de fazermos a distinção entre modelos com regressores não-aninhados.

Na regressão dos salários dos jogadores de beisebol, o \bar{R}^2 da regressão contendo *hrunano* é 0,6211 e o \bar{R}^2 da regressão contendo *rebrunano* é 0,6226. Portanto, com base no *R*-quadrado ajustado, existe uma preferência muito pequena para o modelo com *rebrunano*. Mas a diferença, na prática, é muito pequena, e podemos obter uma resposta diferente controlando algumas das variáveis no Problema 4.16. (Como ambos os modelos não-aninhados contêm cinco parâmetros, o *R*-quadrado habitual pode ser usado para fornecer a mesma conclusão.)

A comparação dos \bar{R}^2 para optarmos entre os diferentes conjuntos não-aninhados de variáveis independentes pode ser de grande valia quando essas variáveis representam formas funcionais diferentes. Considere dois modelos relacionando a intensidade de P&D às vendas de uma empresa:

$$pdintens = \beta_0 + \beta_1 \log(vendas) + u. \quad (6.23)$$

$$pdintens = \beta_0 + \beta_1 vendas + \beta_2 vendas^2 + u. \quad (6.24)$$

O primeiro modelo capture um rendimento decrescente pela inclusão de *vendas* na forma logarítmica; o segundo modelo faz isso com o uso de um termo quadrático. Assim, o segundo modelo contém um parâmetro a mais que o primeiro.

Quando as equações são estimadas usando as 32 observações das empresas de produtos químicos do arquivo RDCHEM.RAW, o R^2 é 0,061, e o R^2 da equação (6.24) é 0,148. Portanto, parece que a função quadrática faz um ajuste muito melhor. Mas uma comparação dos *R*-quadrados habituais com o primeiro modelo é injusta, porque ele contém um parâmetro a menos que a equação (6.24). Isto é, (6.23) é um modelo mais parcimonioso que (6.24).

Tudo o mais igual, modelos mais simples são melhores. Como o *R*-quadrado habitual não penaliza modelos mais complicados, é melhor usar o \bar{R}^2 . O \bar{R}^2 de (6.23) é 0,030, enquanto o \bar{R}^2 de (6.24) é 0,090. Portanto, mesmo depois dos ajustes das diferenças nos graus de liberdade, o modelo quadrático é o melhor. O modelo quadrático também é o preferido quando margens de lucro são incluídas em cada regressão.

Existe uma limitação importante no uso do \bar{R}^2 para escolher entre modelos não-aninhados: não podemos usá-lo para a escolha entre diferentes formas funcionais da variável dependente. Isso é uma pena, pois muitas vezes queremos saber se *y* ou $\log(y)$ (ou talvez alguma outra transformação) deve ser usada como a variável dependente com base no grau de ajuste. Mas nem o R^2 nem o \bar{R}^2 podem ser usados para esse fim. A razão é simples: esses *R*-quadrados medem a proporção explicada do total da variação de qualquer variável dependente que estejamos usando na regressão, e diferentes funções da variável dependente terão diferentes montantes de variação a ser explicadas. Por exemplo, as variações totais em *y* e $\log(y)$ não são as mesmas. A comparação dos *R*-quadrados ajustados dessas regressões com essas diferentes formas das variáveis dependentes não nos dá nenhuma informação sobre qual modelo se adapta melhor; eles estimam duas variáveis dependentes separadas.

Explique por que escolher um modelo maximizando \bar{R}^2 ou minimizando $\hat{\sigma}$ (o erro-padrão da regressão) é a mesma coisa.

EXEMPLO 6.4**(Remuneração de Diretores Executivos e Desempenho de Empresas)**

Considere dois modelos estimados relacionando a remuneração de diretores executivos ao desempenho de empresas:

$$\begin{aligned} \hat{\text{salário}} &= 830,63 + 0,0163 \text{ vendas} + 19,63 \text{ rma} \\ &\quad (223,90) \quad (0,0089) \quad (11,08) \\ n &= 209, R^2 = 0,029, \bar{R}^2 = 0,20 \end{aligned} \tag{6.25}$$

e

$$\begin{aligned} \hat{\text{Isalário}} &= 4,36 + 0,275 \text{ lvendas} + 0,0179 \text{ rma} \\ &\quad (0,29) \quad (0,033) \quad (0,0040) \\ n &= 209, R^2 = 0,282, \bar{R}^2 = 0,275, \end{aligned} \tag{6.26}$$

onde *rma* é o retorno das ações, discutido no Capítulo 2. Para simplificar, *Isalário* e *lvendas* representam os logs naturais de *salário* e *vendas*. Já sabemos como interpretar essas diferentes equações estimadas. Mas podemos dizer se um modelo ajusta os dados melhor que o outro?

O *R*-quadrado da equação (6.25) mostra que *vendas* e *rma* explicam somente cerca de 2,9% da variação do salário dos diretores executivos na amostra. Tanto *vendas* como *rma* têm significância estatística marginal.

A equação (6.26) mostra que *log(vendas)* e *rma* explicam cerca de 28,2% da variação do *log(salário)*. Em termos de grau de ajuste, esse *R*-quadrado bem mais alto parece sugerir que o modelo (6.26) é bem melhor, mas esse não é necessariamente o caso. A soma dos quadrados total de *salário* na amostra é 391.732.982, enquanto a soma dos quadrados total de *log(salário)* é somente 66,72. Assim, há muito menos variação em *log(salário)* que precisa ser explicada.

Neste ponto, podemos usar outros recursos além do *R*² e do *R̄*² para optar entre esses modelos. Por exemplo, *log(vendas)* e *rma* são muito mais significantes, estatisticamente, em (6.26) do que são *vendas* e *rma* em (6.25), e os coeficientes em (6.26) provavelmente são de maior interesse. Para termos certeza, porém, precisaremos fazer uma comparação válida dos graus de ajuste.

Na Seção 6.4 fornecemos um indicador que efetivamente nos permite comparar modelos nos quais só aparece tanto na forma em nível como na forma logarítmica.

O Controle de muitos Fatores na Análise de Regressão

Em muitos dos exemplos que tratamos, e certamente em nossa discussão sobre o viés de variáveis omitidas no Capítulo 3, temos nos preocupado com a omissão de fatores importantes em modelos que possam estar correlacionados com as variáveis independentes. Também é possível controlarmos grande quantidade de variáveis em uma análise de regressão.

Se enfatizarmos exageradamente o grau de ajuste, estaremos nos propondo a controlar fatores em um modelo de regressão que não deveriam ser controlados. Para evitar este equívoco, precisamos nos lembrar da interpretação *ceteris paribus* de modelos de regressão múltipla.

Para ilustrar esse problema, suponha que estejamos fazendo um estudo para avaliar o impacto dos impostos estaduais sobre a cerveja em acidentes fatais de trânsito. A idéia é que um imposto mais elevado sobre a cerveja reduzirá o consumo de bebidas alcoólicas e, da mesma forma, o hábito de dirigir sob embriaguez, resultando em menos acidentes fatais de trânsito. Para medirmos o efeito *ceteris paribus* dos impostos sobre esses acidentes, podemos modelar *fatalidades* como uma função de diversos fatores, inclusive o *imposto* sobre a cerveja:

$$\text{fatalidades} = \beta_0 + \beta_1 \text{imposto} + \beta_2 \text{milhas} + \beta_3 \text{percmasc} + \beta_4 \text{perc16_21} + \dots,$$

onde *milhas* é o total de milhas dirigidas, *percmasc* é a percentagem masculina da população do Estado, e *perc16_21* é a percentagem da população entre 16 e 21 anos de idade, e assim por diante. Observe que não incluímos uma variável medindo o consumo *per capita* de cerveja. Estaremos cometendo um erro de variáveis omitidas? A resposta é não. Se controlarmos o consumo de cerveja nessa equação, de que forma o imposto sobre cerveja afetará as fatalidades no trânsito? Na equação

$$\text{fatalidades} = \beta_0 + \beta_1 \text{imposto} + \beta_2 \text{conscerv} + \dots,$$

β_1 mede a diferença nas fatalidades devido ao aumento de um ponto percentual no imposto, mantendo *conscerv* fixo. É difícil de entender por que isso seria de interesse. Não deveríamos controlar as diferenças de *conscerv* entre os Estados, a menos que quiséssemos verificar algum tipo de efeito indireto do imposto sobre a cerveja. Outros fatores, como a distribuição por sexo e idade, deveriam ser controlados.

A questão de decidir se devemos ou não controlar certos fatores nem sempre é bem definida. Por exemplo, Betts (1995) estuda o efeito da qualidade do ensino médio sobre a renda subsequente. Ele salienta que, se qualidade melhor de ensino resulta em mais educação, então controlar a educação na regressão juntamente com avaliação da qualidade subestimarão o retorno da qualidade. Betts faz a análise com e sem anos de escolaridade na equação para obter uma gama de efeitos estimados da qualidade de ensino.

Para verificar claramente como a ênfase em *R*-quadrados altos pode criar problemas, considere o exemplo do preço dos imóveis da Seção 4.5 que ilustra a verificação de múltiplas hipóteses. Naquele caso, queríamos verificar a razoabilidade da avaliação dos preços dos imóveis. Fizemos a regressão de *log(preço)* sobre *log(aval)*, *log(tamterr)*, *log(arquad)*, e *qtdorm* e verificamos se as três últimas variáveis tinham coeficientes populacionais zero enquanto *log(aval)* tinha um coeficiente unitário. Entretanto, se quisermos estimar um modelo de preço hedônico, como no Exemplo 4.8, onde os valores marginais de várias características dos imóveis são obtidos? Devemos incluir *log(aval)* na equação? O *R*-quadrado ajustado da regressão com *log(aval)* é 0,762, enquanto o *R*-quadrado ajustado sem ele é 0,630. Com base somente no grau de ajuste, devemos incluir *log(aval)*. Contudo, isso será incorreto se nossa meta for determinar os efeitos do tamanho da propriedade, área construída e número de quartos nos valores dos imóveis. A inclusão de *log(aval)* na equação equivale a manter um indicador de valor fixo e indagar quanto a adição de um quarto alterará outro indicador de valor. Essa medida não faz sentido na avaliação das características dos imóveis.

Se lembrarmos que modelos diferentes servem a propósitos diferentes, e nos concentrarmos na interpretação *ceteris paribus* da regressão, não incluiremos os fatores errados em um modelo de regressão.

A Adição de Regressores para Reduzir a Variância do Erro

Acabamos de ver alguns exemplos nos quais certas variáveis independentes não devem ser incluídas em um modelo de regressão, mesmo que elas sejam correlacionadas com a variável dependente. Do Capítulo 3, sabemos que a adição de uma nova variável independente em uma regressão pode exa-

cerbar o problema da multicolinearidade. De outro lado, como estamos retirando algo do termo erro, a adição de uma variável geralmente reduz a variância do erro. De forma geral, não podemos saber que efeito será dominante.

Porém, há um caso que é óbvio: devemos sempre incluir variáveis independentes que afetem y e que sejam *não-correlacionadas* com todas as variáveis independentes de interesse. A razão para esta inclusão é simples: a adição dessa variável não induz multicolinearidade na população (e, portanto, a multicolinearidade na amostra deve ser desprezível), mas reduzirá a variância do erro. Em amostras de tamanho grande, os erros-padrão de todos os estimadores MQO serão reduzidos.

Como exemplo, considere estimar a demanda individual por cerveja como uma função do preço médio da cerveja no município. Pode ser razoável assumir que as características individuais sejam não-correlacionadas com os preços em nível de municípios, e assim uma regressão simples do consumo de cerveja sobre o preço nos municípios seria suficiente para estimar o efeito do preço sobre a demanda individual. Entretanto, é possível obter uma estimativa mais precisa da elasticidade-preço da demanda por cerveja com a inclusão de características individuais, como a idade e o grau de escolaridade. Se esses fatores afetarem a demanda e forem não-correlacionados com o preço, o erro-padrão do coeficiente do preço será menor, pelo menos em amostras grandes.

Como segundo exemplo, considere o subsídio para equipamentos de computação dado no início da Seção 6.3. Se, além da variável do subsídio, controlarmos outros fatores que possam explicar a nota média em curso superior, poderemos provavelmente conseguir uma estimativa mais precisa do efeito do subsídio. Variáveis indicadoras da nota média no ensino médio, a classificação da instituição, as pontuações sat e tac e os antecedentes familiares são bons candidatos. Como os montantes do subsídio são determinados aleatoriamente, todas as variáveis de controle adicionais serão não-correlacionadas com o montante de subsídio; nessa amostra, a multicolinearidade entre o montante do subsídio e as outras variáveis independentes deve ser mínima. Porém, a adição de controles extras pode reduzir显著emente a variância do erro, conduzindo a uma estimativa mais precisa do efeito do subsídio. Lembre-se, neste caso, de que o problema não é a inexistência de viés: obteremos um estimador não-viesado e consistente, quer incluamos ou não as variáveis de desempenho no ensino médio e de antecedentes familiares. O problema está na obtenção de um estimador com uma menor variância amostral.

Infelizmente, casos em que temos informações sobre as variáveis explicativas adicionais que sejam não-correlacionadas com as variáveis explicativas de interesse são raros no campo das ciências sociais. Porém, vale a pena lembrar que, quando essas variáveis estão disponíveis, elas poderão ser incluídas em um modelo para reduzir a variância do erro sem induzir multicolinearidade.

6.4 PREVISÃO E ANÁLISE DE RESÍDUOS

No Capítulo 3 definimos os valores previstos ou estimados do MQO e os resíduos do MQO. As previsões certamente são úteis, mas estão sujeitas à variação amostral, já que elas são obtidas com o uso dos estimadores MQO. Assim, nesta seção, mostramos como obter intervalos de confiança de previsões da linha de regressão MQO.

Sabemos, dos capítulos 3 e 4, que os resíduos são usados para obter a soma dos resíduos quadrados e o R -quadrado, de modo que eles são importantes para o grau de ajuste e os testes de hipóteses.

Algumas vezes, os economistas estudam os resíduos de uma observação específica para obter informações sobre os indivíduos (ou empresas, imóveis etc.) na amostra.

Intervalos de Confiança de Previsões

Suponha que tenhamos estimado a equação

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (6.27)$$

Quando inserimos valores específicos das variáveis independentes nessa equação, obtemos uma previsão de y , que é uma estimativa do *valor esperado* de y , dados os valores específicos das variáveis explicativas. Para enfatizar, sejam c_1, c_2, \dots, c_k valores particulares de cada uma das k variáveis independentes; elas poderão ou não corresponder a um ponto efetivo dos dados em nossa amostra. O parâmetro que gostaríamos de estimar é

$$\begin{aligned} \theta_0 &= \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \dots + \beta_k c_k \\ &= E(y|x_1 = c_1, x_2 = c_2, \dots, x_k = c_k). \end{aligned} \quad (6.28)$$

O estimador de θ_0 é

$$\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \dots + \hat{\beta}_k c_k. \quad (6.29)$$

Na prática, isso é fácil de ser computado. Entretanto, se o que quisermos for um indicador da incerteza nesse valor previsto? É natural construir um intervalo de confiança de θ_0 que seja centrado em $\hat{\theta}_0$.

Para obter um intervalo de confiança de θ_0 precisamos de um erro-padrão de $\hat{\theta}_0$. Então, com um grande gl , poderemos construir um intervalo de confiança de 95% utilizando a regra prática $\hat{\theta}_0 \pm 2 \cdot ep(\hat{\theta}_0)$. (Como sempre, podemos usar os percentis exatos em uma distribuição t .)

Como obtemos o erro-padrão de $\hat{\theta}_0$? Este é o mesmo problema que encontramos na Seção 4.4: precisamos obter um erro-padrão de uma combinação linear dos estimadores MQO. Aqui, o problema é ainda mais complicado, pois todos os estimadores MQO geralmente aparecem em $\hat{\theta}_0$ (a menos que algum c_j seja zero). No entanto, o mesmo truque que usamos na Seção 4.4 funcionará aqui. Escreva $\beta_0 = \theta_0 - \beta_1 c_1 - \dots - \beta_k c_k$ e agregue isso à equação

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

para obter

$$y = \theta_0 + \beta_1(x_1 - c_1) + \beta_2(x_2 - c_2) + \dots + \beta_k(x_k - c_k) + u. \quad (6.30)$$

Em outras palavras, subtraímos o valor c_j de cada observação de x_j , e depois computamos a regressão de

$$y_i \text{ sobre } (x_{i1} - c_1), \dots, (x_{ik} - c_k), i = 1, 2, \dots, n. \quad (6.31)$$

O valor previsto em (6.29) e, mais importante, seu erro-padrão, são obtidos do *intercepto* (ou constante) na regressão (6.31).

Como exemplo, obtemos um intervalo de confiança de uma previsão a partir de uma regressão de nota média em curso superior, das quais usamos informações do ensino médio.

EXEMPLO 6.5**(Intervalo de Confiança de *nmgrad* Previsto)**

Utilizando os dados contidos no arquivo GPA2.RAW, obtemos a seguinte equação para prever *nmgrad*:

$$\begin{aligned}
 \hat{nmgrad} &= 1,493 + 0,00149 \text{sat} - 0,01386 \text{emperc} \\
 &\quad (0,075) \quad (0,00007) \quad (0,00056) \\
 &\quad - 0,06088 \text{tamclas} + 0,00546 \text{tamclas}^2 \\
 &\quad (0,1650) \quad (0,00227) \\
 n &= 4.137, R^2 = 0,278, \bar{R}^2 = 0,277, \hat{\sigma} = 0,560,
 \end{aligned} \tag{6.32}$$

onde apresentamos as estimativas com várias casas decimais para reduzir o erro de arredondamento. Qual a previsão de *nmgrad* quando *sat* = 1.200, *emperc* = 30 e *tamclas* = 5 (o que significaria 500)? Isso é fácil de ser obtido, incorporando esses valores na equação (6.32): $\hat{nmgrad} = 2,70$ (arredondado para duas casas decimais). Infelizmente, não podemos usar diretamente a equação (6.32) para obter um intervalo de confiança da *nmgrad* esperada com os valores dados das variáveis independentes. Uma maneira simples de obter um intervalo de confiança é definir um novo conjunto de variáveis independentes: *sat0* = *sat* - 1.200, *emperc0* = *emperc* - 30, *tamclas0* = *tamclas* - 5 e *tamclasquadrado* = *tamclas*² - 25. Quando fazemos a regressão de *nmgrad* sobre essas novas variáveis independentes, obtemos

$$\begin{aligned}
 \hat{nmgrad} &= 2,700 + 0,00149 \text{sat0} - 0,01386 \text{emperc0} \\
 &\quad (0,020) \quad (0,00007) \quad (0,00056) \\
 &\quad - 0,06088 \text{tamclas0} + 0,00546 \text{tamclasquadrado} \\
 &\quad (0,1650) \quad (0,00227) \\
 n &= 4.137, R^2 = 0,278, \bar{R}^2 = 0,277, \hat{\sigma} = 0,560.
 \end{aligned}$$

A única diferença entre esta regressão e aquela em (6.32) é o intercepto, que é a previsão que queremos, juntamente com seu erro-padrão, 0,020. Não é por acidente que os coeficientes de inclinação, seus erros-padrão, *R*-quadrado etc. são os mesmos de antes; esse fato fornece uma maneira de verificarmos se foram feitas as transformações adequadas. Podemos construir com facilidade um intervalo de confiança de 95% da nota média esperada: $2,70 \pm 1,96(0,020)$ ou em torno de 2,66 a 2,74. Este intervalo de confiança é suficientemente estreito devido ao tamanho bastante grande da amostra.

Como a variância do estimador do intercepto é a menor quando cada variável explicativa tem média amostral zero (veja a Questão 2.5 para o caso da regressão simples), segue da regressão em (6.31) que a variância da previsão nos valores médios de *x_j* (isto é, *c_j* = *bar{x_j* para todo *j*) é a menor. Este resultado não é tão surpreendente, já que o ponto de maior confiabilidade em nossa linha de regressão está próximo ao centro dos dados. Na medida em que os valores de *c_j* se afastam de *bar{x_j*, Var(*ŷ*) se torna cada vez maior.

O método anterior nos possibilita colocar um intervalo de confiança em torno da estimativa MQO de $E(y|x_1, \dots, x_k)$ para quaisquer valores das variáveis explicativas. Em outras palavras, obtemos um

intervalo de confiança do valor médio de y da subpopulação com determinado conjunto de covariadas. Entretanto, um intervalo de confiança da média pessoal na subpopulação não é a mesma coisa que um intervalo de confiança de uma unidade particular (indivíduo, família, empresa etc.) da população. Na formação de um intervalo de confiança de um resultado desconhecido de y , devemos avaliar outra fonte muito importante de variação: a variância no erro não observado, que registra nosso desconhecimento dos fatores não observados que afetam y .

Seja y^0 o valor para o qual gostaríamos de construir um intervalo de confiança, que algumas vezes chamamos de **intervalo de previsão**. Por exemplo, y^0 poderia representar uma pessoa ou uma empresa que não esteja em nossa amostra original. Façamos x_1^0, \dots, x_k^0 serem os novos valores das variáveis independentes, que assumimos observar, e u^0 ser o erro não observado. Portanto, temos

$$y^0 = \beta_0 + \beta_1 x_1^0 + \beta_2 x_2^0 + \dots + \beta_k x_k^0 + u^0. \quad (6.33)$$

Como antes, nossa melhor previsão de y^0 é o valor esperado de y^0 , dadas as variáveis explicativas que estimamos da linha de regressão MQO: $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_1^0 + \hat{\beta}_2 x_2^0 + \dots + \hat{\beta}_k x_k^0$. O **erro de previsão** com o uso de \hat{y}^0 para prever y^0 é

$$\hat{e}^0 = y^0 - \hat{y}^0 = (\beta_0 + \beta_1 x_1^0 + \dots + \beta_k x_k^0) + u^0 - \hat{y}^0. \quad (6.34)$$

Agora, $E(\hat{y}^0) = E(\hat{\beta}_0) + E(\hat{\beta}_1)x_1^0 + E(\hat{\beta}_2)x_2^0 + \dots + E(\hat{\beta}_k)x_k^0 = \beta_0 + \beta_1 x_1^0 + \dots + \beta_k x_k^0$, porque os $\hat{\beta}$ são não-viesados. (Como antes, essas expectativas são todas condicionais aos valores amostrais das variáveis independentes.) Como u^0 tem média zero, $E(\hat{e}^0) = 0$. Mostramos que o erro de previsão esperado é zero.

Ao encontrar a variância de \hat{e}^0 , observe que u^0 é não-correlacionado com cada $\hat{\beta}_j$, porque u^0 é não-correlacionado com os erros na amostra usada para a obtenção de $\hat{\beta}_j$. Pelas propriedades básicas da covariância (veja Apêndice B, no site da Thomson), u^0 e \hat{y}^0 são não-correlacionados. Portanto, a **variância do erro de previsão** (condicional a todos os valores das variáveis independentes incluídas na amostra) é a soma das variâncias:

$$\text{Var}(\hat{e}^0) = \text{Var}(\hat{y}^0) + \text{Var}(u^0) = \text{Var}(\hat{y}^0) + \sigma^2, \quad (6.35)$$

onde $\sigma^2 = \text{Var}(u^0)$ é a variância do erro. Existem duas fontes de variância em \hat{e}^0 . A primeira é o erro de amostragem em \hat{y}^0 , que surge por termos estimado β_j . Como cada $\hat{\beta}_j$ tem uma variância proporcional a $1/n$, na qual n é o tamanho da amostra, $\text{Var}(\hat{y}^0)$ é proporcional a $1/n$. Isso significa que, para amostras grandes, $\text{Var}(\hat{y}^0)$ pode ser muito pequena. Em contraposição, σ^2 é a variância do erro na população; ela não muda com o tamanho da amostra. Em muitos exemplos, σ^2 será o termo dominante em (6.35).

Sob as hipóteses do modelo linear clássico, $\hat{\beta}_j$ e u^0 são normalmente distribuídos, e assim \hat{e}^0 também é normalmente distribuído (condicional a todos os valores amostrais das variáveis explicativas). Anteriormente, descrevemos como obter um estimador não-viesado de $\text{Var}(\hat{y}^0)$, e obtivemos nosso estimador não-viesado de σ^2 no Capítulo 3. Com o uso desses estimadores, podemos definir o erro-padrão de \hat{e}^0 como

$$ep(\hat{e}^0) = \{\{ep(\hat{y}^0)\}^2 + \hat{\sigma}^2\}^{1/2}. \quad (6.36)$$

Utilizando o mesmo raciocínio para as estatísticas t de $\hat{\beta}_j$, $\hat{e}^0 / ep(\hat{e}^0)$ tem uma distribuição t com $n - (k + 1)$ graus de liberdade. Portanto,

$$P[-t_{0.025} \leq \hat{e}^0 / ep(\hat{e}^0) \leq t_{0.025}] = 0.95,$$

onde $t_{0.025}$ é o 97,5º percentil na distribuição t_{n-k-1} . Para $n = k + 1$ grande, lembre-se de que $t_{0.025} \approx 1,96$. Inserindo $\hat{e}^0 = \hat{y}^0 - \hat{y}^0$ e fazendo a reordenação, obtemos um intervalo de previsão de 95% para \hat{y}^0 :

$$\hat{y}^0 \pm t_{0.025} \cdot ep(\hat{e}^0); \quad (6.37)$$

como sempre, exceto para gl pequeno, uma boa regra prática é $\hat{y}^0 \pm 2ep(\hat{e}^0)$. Isso é mais amplo que o próprio intervalo de confiança de \hat{y}^0 , devido a $\hat{\sigma}^2$ em (6.36); normalmente ela é muito mais ampla para refletir os fatores em u^0 que não tenhamos controlado.

EXEMPLO 6.6

(Intervalo de Confiança de Notas Médias Futuras)

Suponha que desejamos um IC de 95% de $nmgrad$ futuro de um aluno do ensino médio com $sat = 1.200$, $emperc = 30$ e $tamclasse = 5$. No Exemplo 6.5 obtivemos um intervalo de confiança de 95% da média da nota média entre todos os alunos com as características particulares $sat = 1.200$, $emperc = 30$ e $tamclasse = 5$. Agora, queremos um intervalo de confiança de 95% de qualquer aluno que especificamente tenha essas características. O intervalo de previsão de 95% deve registrar a variação na característica individual, não-observada, que afeta o desempenho universitário. Temos tudo que é preciso para obter um IC de $nmgrad$. Sabemos que $ep(\hat{y}^0) = 0,020$ e $\hat{\sigma} = 0,560$ e, portanto, de (6.36), $ep(\hat{e}^0) = \{(0,020)^2 + (0,560)^2\}^{1/2} \approx 0,560$. Observe o quanto $ep(\hat{y}^0)$ é pequeno em relação a $\hat{\sigma}$: virtualmente, toda a variação em \hat{e}^0 vem da variação em u^0 . O IC de 95% é $2,70 \pm 1,96(0,560)$ ou está entre 1,60 e 3,80. Este é um intervalo de confiança enorme, e mostra que, com base nos fatores que incluímos na regressão, não podemos definir com clareza a futura nota de graduação de determinado indivíduo. (Em certo sentido, isso é bom, por significar que a classificação no curso médio e o desempenho no teste de aptidão acadêmica não predeterminam o desempenho de alguém na faculdade.) Evidentemente, as características não observadas variam amplamente de um indivíduo para o outro com as mesmas notas no teste de aptidão acadêmica e na classificação no curso médio observadas.

Análise de Resíduos

Algumas vezes, é útil examinar as observações individuais para verificar se o valor efetivo da variável dependente está acima ou abaixo do valor previsto; isto é, examinar os resíduos das observações individuais. Este processo é chamado **análise de resíduos**. Os economistas são conhecidos por examinarem os resíduos de uma regressão para auxiliá-los, por exemplo, na compra de um imóvel.

O exemplo seguinte sobre preços de imóveis ilustra a análise de resíduos. Os preços dos imóveis estão relacionados a várias características observadas do imóvel. Podemos relacionar todas as características que julgarmos importantes, como tamanho, número de quartos, número de banheiros, e assim por diante. Podemos usar uma amostra de imóveis para estimar o relacionamento entre o preço e as características, e terminamos obtendo um valor previsto e um valor real de cada imóvel. Então, podemos construir os resíduos, $\hat{u}_i = y_i - \hat{y}_i$. O imóvel com o maior resíduo negativo é, pelo menos com base nos fatores que estamos controlando, o mais barato em relação às suas características *observadas*. É claro que um preço de venda substancialmente inferior ao seu preço previsto poderia indicar alguma característica indesejável do imóvel que deixamos de avaliar, e que portanto está contido no erro não-observado. Além da obtenção da previsão e do resíduo, também faz sentido computar o intervalo de confiança de qual seria o preço de venda do imóvel no futuro, utilizando o método descrito na equação (6.37).

Utilizando os dados contidos no arquivo HPRICE1.RAW, computamos a regressão de *preço* sobre *tamterr*, *arquad* e *qtdorm*. Na amostra de 88 imóveis, o resíduo mais negativo é -120,206, do 81º imóvel. Portanto, o preço pedido por esse imóvel está US\$120.206,00 abaixo de seu preço previsto.

Existem muitos outros usos da análise de resíduos. Uma maneira de classificar as faculdades de direito é fazer a regressão da mediana dos salários iniciais sobre uma variedade de características dos alunos (como a mediana das notas de ingresso nos cursos para novos alunos, a mediana das notas médias de graduação para novos alunos etc.) e obter um valor previsto e um resíduo de cada faculdade de direito. A faculdade de direito com o maior resíduo terá o maior valor agregado previsto. (Naturalmente, ainda existirá muita incerteza sobre como o salário inicial de um indivíduo se compararia com a mediana geral de uma faculdade de direito.) Esses resíduos poderão ser usados juntamente com mensalidades cobradas pelas faculdades de direito para determinarmos o valor mais vantajoso; isso exigirá um desconto apropriado dos ganhos futuros.

A análise de resíduos também tem participação em decisões judiciais. Um artigo publicado no jornal *The New York Times* intitulado “Juiz Diz que Pobreza de Alunos, e Não a Segregação, Prejudica Aproveitamento Escolar” (28.06.95) descreve um importante processo legal. A questão era se o fraco desempenho nos exames padronizados do Distrito Escolar de Hartford, em relação ao desempenho nos distritos vizinhos, era devido à baixa qualidade de ensino nas escolas altamente segregadas. O juiz concluiu que “a disparidade nas notas de aproveitamento escolar não indica que Hartford esteja fazendo um trabalho inadequado ou insuficiente na educação de seus alunos ou que suas escolas sejam deficientes, pois as notas de aproveitamento previstas com base em relevantes fatores socioeconômicos estão próximas dos níveis esperados”. Esta conclusão foi, quase com certeza, baseada em uma análise de regressão das notas de aproveitamento médias ou de suas medianas sobre as características socioeconômicas de vários distritos escolares de Connecticut. A conclusão do juiz sugere que, considerando o nível de pobreza dos alunos das escolas de Hartford, as notas de aproveitamento efetivas dos alunos eram semelhantes às previstas em uma análise de regressão: o resíduo de Hartford não era suficientemente negativo para se concluir que as escolas em si mesmas eram responsáveis pelas baixas notas de aproveitamento escolar.

Como você poderia usar a análise de resíduos para determinar quais atores cinematográficos são remunerados em níveis acima da bilheteria dos filmes em que atuam?

Previsão de y quando a Variável Dependente é $\log(y)$

Como a transformação do log natural é usada com tanta freqüência na variável dependente em economia empírica, dedicamos esta subseção ao problema de prognosticar y quando a variável dependente é $\log(y)$. Como um subproduto, obtaremos um indicador de grau de ajuste do modelo log que possa ser comparado com o R -quadrado do modelo em nível.

Para obter uma previsão útil definirmos $\hat{\log}y = \log(\hat{y})$; isso realça o fato de que é o log de y que será previsto no modelo

$$\hat{\log}y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u. \quad (6.38)$$

Nesta equação os x_j poderão ser transformações de outras variáveis; por exemplo, poderíamos ter $x_1 = \log(vendas)$, $x_2 = \log(valmerc)$, $x_3 = permceo$ no exemplo do salário dos diretores executivos.

Dados os estimadores MQO, sabemos como prever $\hat{\log}y$ para qualquer valor das variáveis independentes:

$$\hat{\log}y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k. \quad (6.39)$$

Agora, como o exponencial desfaz o log, nossa primeira suposição para prever y é simplesmente exponenciar o valor previsto de $\log(y)$: $\hat{y} = \exp(\hat{\log}y)$. Isso não funciona; aliás, isso sistematicamente *subestimar*á o valor esperado de y . De fato, se o modelo (6.38) obedecer às hipóteses do modelo linear clássico RLM.1 até RLM.6, pode ser demonstrado que

$$E(y|x) = \exp(\sigma^2/2) \cdot \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)$$

onde x representa as variáveis independentes e σ^2 é a variância de u . [Se $u \sim \text{Normal}(0, \sigma^2)$, o valor esperado de $\exp(u)$ será $\exp(\sigma^2/2)$]. Esta equação mostra que um ajuste simples é necessário para prevermos y :

$$\hat{y} = \exp(\hat{\sigma}^2/2) \exp(\hat{\log}y), \quad (6.40)$$

onde $\hat{\sigma}^2$ é simplesmente o estimador não-viesado de σ^2 . Como $\hat{\sigma}$, o erro-padrão da regressão, é sempre conhecido, a obtenção de valores previstos de y será fácil. Como $\hat{\sigma}^2 > 0$, $\exp(\hat{\sigma}^2/2) > 1$. Para um $\hat{\sigma}^2$ grande, esse fator de ajuste poderá ser substancialmente maior que a unidade.

A previsão em (6.40) não é não-viesada, mas consistente. Não existem previsões não-viesadas de y , e em muitos casos (6.40) funciona bem. Porém, ela depende da normalidade do termo erro, u . No Capítulo 5, mostramos que o MQO possui propriedades desejáveis, mesmo quando u não for normalmente distribuído. Portanto, é vantajoso ter uma previsão que não dependa da normalidade. Se simplesmente assumirmos que u é independente das variáveis explicativas, teremos

$$E(y|x) = \alpha_0 \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k), \quad (6.41)$$

onde α_0 é o valor esperado de $\exp(u)$, que deve ser maior que a unidade.

Dada uma estimativa de $\hat{\alpha}_0$, podemos prever y como

$$\hat{y} = \hat{\alpha}_0 \exp(\hat{l}\log y), \quad (6.42)$$

que mais uma vez simplesmente requer que façamos a exponenciação do valor previsto do modelo log e que multipliquemos o resultado por $\hat{\alpha}_0$.

O resultado é que um estimador consistente de $\hat{\alpha}_0$ é facilmente obtido.

PREVISÃO DE y QUANDO A VARIÁVEL DEPENDENTE É LOG(y):

- (i) Obtenha os valores estimados $\hat{l}\log y_i$ da regressão de $\log y$ sobre x_1, \dots, x_k .
- (ii) Para cada observação i , crie $\hat{m}_i = \exp(\hat{l}\log y_i)$.
- (iii) Agora, faça a regressão de y sobre a variável única \hat{m} sem um intercepto; isto é, faça uma regressão simples passando pela origem. O coeficiente de \hat{m} , o único coeficiente que existe, é a estimativa de α_0 .

Quando $\hat{\alpha}_0$ for obtido, ele poderá ser usado juntamente com previsões de $\log y$ para prever y . Os passos são os seguintes:

- (i) Para valores dados de x_1, x_2, \dots, x_k , obtenha $\hat{l}\log y$ de (6.39).
- (ii) Obtenha a previsão de \hat{y} de (6.42).

EXEMPLO 6.7

(Previsão de Salários de Diretores Executivos)

O modelo de interesse é

$$\log(salário) = \beta_0 + \beta_1 \log(vendas) + \beta_2 \log(valmerc) + \beta_3 permceo + u,$$

de forma que β_1 e β_2 são elasticidades e $100 \cdot \beta_3$ é uma semi-elasticidade. A equação estimada utilizando os dados contidos no arquivo CEOSEL2.RAW é

$$\begin{aligned} Isalário &= 4,504 + 0,163 lvendas + 0,109 lvalmerc + 0,0117 permceo \\ (0,257) &\quad (0,039) \quad (0,050) \quad (0,0053) \end{aligned} \quad (6.43)$$

$$n = 177, R^2 = 0,318,$$

onde, para maior clareza, $Isalário$ representa o log de $salário$ e, de forma semelhante, $lvendas$ e $lvalmerc$ representam o log de $vendas$ e $valmerc$. A seguir, obtemos $\hat{m}_i = \exp(Isalário_i)$ de cada observação na amostra. A regressão de $salário$ sobre \hat{m} (sem uma constante) produz $\hat{\alpha}_0 \approx 1,117$.

Podemos usar esse valor de $\hat{\alpha}_0$ juntamente com (6.43) para prever $salário$ de qualquer valor de $vendas$, $valmerc$ e $permceo$. Encontremos a previsão de $vendas = 5.000$ (que significa 5 bilhões de dólares, já que $vendas$ está em milhões de dólares), $valmerc = 10.000$ (10 bilhões de dólares), e $permceo = 10$. Da equação (6.43), a previsão de $Isalário$ é $4,504 + 0,163 \cdot \log(5.000) + 0,109 \cdot \log(10.000) + 0,0117(10) \approx 7,013$. O salário previsto é, portanto, $1,117 \cdot \exp(7,013) \approx 1.240,967$, ou 1.240.967 dólares. Se esquecermos de multiplicar por $\hat{\alpha}_0 = 1,117$, obteremos uma previsão de 1.110.983 dólares.

Podemos usar o método anterior de obter previsões para determinar o quanto o modelo com $\log(y)$ como variável dependente explica bem a variável y . Já temos indicadores para modelos quando y é a variável dependente: o R -quadrado e o R -quadrado ajustado. O objetivo é encontrar um bom indicador de grau de ajuste no modelo $\log(y)$ que possa ser comparado com um R -quadrado de um modelo no qual y é a variável dependente.

Existem várias maneiras de encontrarmos esse indicador, mas apresentamos um método que é fácil de ser implementado. Após computar a regressão de y sobre \hat{m} passando pela origem no passo (iii), obtemos os valores estimados dessa regressão, $\hat{y}_i = \hat{\alpha}_0\hat{m}_i$. Em seguida, encontramos a correlação amostral entre \hat{y}_i e o verdadeiro y_i na amostra. O quadrado dessa correlação amostral pode ser comparado com o R -quadrado que obtemos com o uso de y como a variável dependente em um modelo de regressão linear. Lembre-se de que o R -quadrado na equação estimada

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$$

é exatamente a correlação quadrada entre y_i e \hat{y}_i (veja a Seção 3.2).

EXEMPLO 6.8

(Previsão de Salários de Diretores Executivos)

Após o passo (iii) no procedimento precedente, obtemos os valores estimados $\text{salário} = \hat{\alpha}_0\hat{m}_i$. A correlação simples entre salário_i e salário na amostra é 0,493; o quadrado desse valor está por volta de 0,243. Esse é o nosso indicador de quanto da variação de salário é explicada pelo modelo log; não se trata do R -quadrado de (6.43), que é 0,318.

Suponha que estimemos um modelo com todas as variáveis em nível:

$$\text{salário} = \beta_0 + \beta_1 \text{vendas} + \beta_2 \text{valmerc} + \beta_3 \text{permceo} + u.$$

O R -quadrado obtido estimando esse modelo, usando as mesmas 177 observações, é 0,201. Assim, o modelo log explica mais da variação em salário , e assim é o nosso preferido com base no grau de ajuste. O modelo log também é o escolhido porque parece ser mais realista e seus parâmetros são mais fáceis de ser interpretados.

Neste capítulo, tratamos de alguns tópicos importantes sobre a análise de regressão múltipla.

A Seção 6.1 mostrou que uma mudança nas unidades de medida de uma variável independente altera o coeficiente do MQO da maneira esperada: se x_i for multiplicado por c , seu coeficiente será dividido por c . Se a variável dependente é multiplicada por c , todos os coeficientes de MQO são multiplicados por c . Nem a estatística t nem a F são alteradas pela mudança das unidades de medida de quaisquer variáveis.

Discutimos os coeficientes beta, que medem os efeitos das variáveis independentes sobre a variável dependente em unidades de desvios-padrão. Os coeficientes beta são obtidos de uma regressão

MQO padrão depois de as variáveis dependente e independentes terem sido transformadas em valores padronizados.

Como vimos em vários exemplos, a forma funcional logarítmica produz coeficientes com interpretações de efeitos percentuais. Comentamos sobre suas vantagens adicionais na Seção 6.2. Também vimos como computar o efeito percentual exato quando um coeficiente em um modelo log-nível é grande. Modelos com termos quadráticos consideram efeitos marginais decrescentes ou crescentes. Modelos com interações possibilitam que o efeito marginal de uma variável explicativa dependa do nível de outra variável explicativa.

Introduzimos o R -quadrado ajustado, \bar{R}^2 , como uma alternativa ao R -quadrado habitual para medir o grau de ajuste. Enquanto o R^2 nunca pode cair quando outra variável é adicionada na regressão, o \bar{R}^2 penaliza o número de regressores e pode cair quando uma variável independente é adicionada. Isso faz do \bar{R}^2 o preferido para a opção entre modelos não-aninhados com diferentes quantidades de variáveis explicativas. Nem o R^2 nem o \bar{R}^2 podem ser usados para comparar modelos com variáveis dependentes diferentes. No entanto, é bastante fácil obter indicadores de graus de ajuste para optarmos entre y e $\log(y)$ como a variável dependente, como mostrado na Seção 6.4.

Na Seção 6.3 discutimos o problema de certa forma sutil de dependermos demasiadamente do R^2 ou do \bar{R}^2 para chegarmos a um modelo final: é possível controlarmos grandes quantidades de fatores em um modelo de regressão. Por essa razão, é importante pensar à frente sobre a especificação de modelos, particularmente sobre a natureza *ceteris paribus* da equação de regressão múltipla. Variáveis explicativas que afetem y e sejam não-correlacionadas com todas as outras variáveis explicativas podem ser usadas para reduzir a variância do erro sem induzir multicolinearidade.

Na Seção 6.4 demonstramos como obter um intervalo de confiança de uma previsão feita de uma linha de regressão MQO. Também mostramos como um intervalo de confiança pode ser construído para um valor futuro e desconhecido de y .

Ocasionalmente, queremos prever y quando $\log(y)$ é usado como a variável dependente em um modelo de regressão. A Seção 6.4 explica esse método simples. Finalmente, algumas vezes estamos interessados em conhecer o sinal e a magnitude dos resíduos de observações específicas. A análise de resíduos pode ser usada para determinarmos se elementos específicos da amostra possuem valores previstos que estejam bem acima, ou bem abaixo, dos verdadeiros resultados.

6.1 A seguinte equação foi estimada utilizando os dados contidos no arquivo CEOSAL1.RAW:

$$\log(\hat{\text{salário}}) = 4,322 + 0,276 \log(\text{vendas}) + 0,0215 rma - 0,00008 rma^2$$

(0,324) (0,33) (0,129) (0,00026)

$$n = 209, R^2 = 0,282.$$

Esta equação permite que rma tenha um efeito decrescente sobre $\log(\text{salário})$. Essa generalidade é necessária? Justifique.

6.2 Sejam $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ as estimativas MQO da regressão de y_i sobre x_{i1}, \dots, x_{ik} , $i = 1, 2, \dots, n$. Para constantes diferentes de zero c_1, \dots, c_k , argumente que o intercepto e as inclinações MQO da regressão de c_0y_i sobre $c_1x_{i1}, \dots, c_kx_{ik}$, $i = 1, 2, \dots, n$ são dados por $\tilde{\beta}_0 = c_0\hat{\beta}_0, \tilde{\beta}_1 = (c_0/c_1)\hat{\beta}_1, \dots, \tilde{\beta}_k = (c_0/c_k)\hat{\beta}_k$

$\hat{\beta}_k$. (Sugestão: Use o fato de que $\hat{\beta}_j$ soluciona as condições de primeira ordem em (3.13), e que $\tilde{\beta}_j$ deve solucionar as condições de primeira ordem envolvendo as variáveis dependentes e independentes redimensionadas.)

6.3 Utilizando os dados contidos no arquivo RDCHEM.RAW, a seguinte equação foi obtida por MQO:

$$\begin{aligned} pdintens &= 2,613 + 0,00030 vendas - 0,0000000070 vendas^2 \\ &\quad (0,429) \quad (0,00014) \quad (0,000000037) \\ n &= 32, R^2 = 0,1484. \end{aligned}$$

- (i) Em que ponto o efeito marginal de *vendas* sobre *pdintens* se torna negativo?
- (ii) Você manteria o termo quadrático no modelo? Explique.
- (iii) Defina *vendasbil* como vendas expressas em bilhões de dólares: $vendasbil = vendas/1.000$. Reescreva a equação com *vendasbil* e $vendasbil^2$ como as variáveis independentes. Certifique-se de descrever os erros-padrão e o *R*-quadrado. [Sugestão: Observe que $vendasbil^2 = vendas^2/(1.000)^2$.]
- (iv) Com o propósito de descrever os resultados, qual equação você prefere?

6.4 O seguinte modelo permite que o retorno da educação dependa da educação total dos pais, chamada *edupais*:

$$\log(salário) = \beta_0 + \beta_1 educ + \beta_2 educ \cdot edupais + \beta_3 exper + \beta_4 perm + u.$$

- (i) Mostre que, em forma decimal, o retorno de mais um ano de educação nesse modelo é

$$\Delta \log(salário)/\Delta educ = \beta_1 + \beta_2 edupais.$$

Que sinal você espera para β_2 ? Por quê?

- (ii) Utilizando os dados contidos no arquivo WAGE2.RAW, a equação estimada é

$$\begin{aligned} \log(\hat{salário}) &= 5,65 + 0,047 educ + 0,00078 educ \cdot edupais + \\ &\quad (0,13) \quad (0,010) \quad (0,00021) \\ &\quad 0,019 exper + 0,010 perm \\ &\quad (0,004) \quad (0,003) \\ n &= 722, R^2 = 0,169. \end{aligned}$$

(Somente 722 observações contêm todas as informações sobre a educação dos pais.) Interprete o coeficiente do termo de interação. Pode ser interessante escolher dois valores específicos para *edupais*, por exemplo, *edupais* = 32 se ambos tiverem educação superior, ou *edupais* = 24 se ambos tiverem educação de nível médio — e comparar o retorno estimado de *educ*.

(iii) Quando *edupais* é adicionada como uma variável separada na equação, obtemos:

$$\begin{aligned} \log(\hat{\text{salário}}) = & 4,94 + 0,097 \text{ educ} + 0,033 \text{ edupais} + 0,0016 \text{ educ} \cdot \text{edupais} \\ & (0,38) \quad (0,027) \quad (0,017) \quad (0,0012) \\ & + 0,020 \text{ exper} + 0,010 \text{ perm} \\ & (0,004) \quad (0,003) \\ & n = 722, R^2 = 0,174. \end{aligned}$$

O retorno da educação agora depende positivamente da educação dos pais? Teste a hipótese nula de que o retorno da educação não depende da educação dos pais.

6.5 No Exemplo 4.2, no qual a percentagem de alunos aprovados em um exame de matemática do 10º ano (*mate10*) é a variável dependente, faz sentido incluir *cien11* — a percentagem de alunos do 11º ano aprovados em um exame de ciências — como uma variável explicativa adicional?

6.6 Quando *taxafreq*² e *tac* · *taxafreq* são adicionadas à equação (6.19), o *R*-quadrado passa a ser 0,232. Esses termos adicionais são conjuntamente significantes no nível de 10%? Você os incluiria no modelo?

6.7 As três seguintes equações foram estimadas utilizando-se as 1,534 observações contidas no arquivo 401K.RAW.

$$\begin{aligned} \hat{\text{taxap}} = & 80,29 + 5,44 \text{ taxcomp} + 0,269 \text{ idade} - 0,00013 \text{ totemp} \\ & (0,78) \quad (0,52) \quad (0,045) \quad (0,00004) \\ & R^2 = 0,100, \bar{R}^2 = 0,098. \end{aligned}$$

$$\begin{aligned} \hat{\text{taxap}} = & 97,32 + 5,02 \text{ taxcomp} + 0,314 \text{ idade} - 2,66 \log(\text{totemp}) \\ & (1,95) \quad (0,51) \quad (0,044) \quad (0,28) \\ & R^2 = 0,144, \bar{R}^2 = 0,142. \end{aligned}$$

$$\begin{aligned} \hat{\text{taxap}} = & 80,62 + 5,34 \text{ taxcomp} + 0,290 \text{ idade} - 0,00043 \text{ totemp} \\ & (0,78) \quad (0,52) \quad (0,045) \quad (0,00009) \\ & + 0,000000039 \text{ totemp}^2 \\ & \quad (0,000000010) \\ & R^2 = 0,108, \bar{R}^2 = 0,106. \end{aligned}$$

Qual desses três modelos você prefere? Por quê?

Análise de Regressão Múltipla com Informações Qualitativas: Variáveis Binárias (ou Dummy)

os capítulos anteriores, as variáveis dependentes e independentes em nossos modelos de regressão múltipla tinham significado *quantitativo*. Alguns exemplos incluíam taxas de salário por hora, anos de escolaridade, nota média em curso superior, quantidade da poluição do ar, níveis de vendas de empresas e número de detenções. Em cada caso, a magnitude da variável carrega informações valiosas. No trabalho empírico também devemos incorporar fatores *qualitativos* nos modelos de regressão. O sexo ou a raça de um indivíduo, o ramo de atividade de uma empresa (fabricante, varejista etc.) e a região onde uma cidade está localizada (sul, norte, oeste etc.) são todos considerados fatores qualitativos.

A maior parte deste capítulo é dedicada a variáveis *independentes* qualitativas. Após discutirmos as maneiras apropriadas de descrever informações qualitativas na Seção 7.1, mostraremos como variáveis explicativas qualitativas podem ser facilmente incorporadas em modelos de regressão múltipla nas Seções 7.2, 7.3 e 7.4. Essas seções tratam de quase todos os modos conhecidos nos quais as variáveis independentes qualitativas são usadas na análise de regressão de corte transversal.

Na Seção 7.5 examinaremos uma variável dependente binária, que é um tipo especial de variável dependente qualitativa. O modelo de regressão múltipla tem uma interpretação bastante interessante neste caso e é chamado de modelo de probabilidade linear. Embora muito criticado por alguns econometristas, a simplicidade do modelo de probabilidade linear faz dele uma ferramenta útil em muitos contextos empíricos. Também descreveremos suas falhas na Seção 7.5, embora elas sejam freqüentemente secundárias no trabalho empírico.

7.1 A DESCRIÇÃO DAS INFORMAÇÕES QUALITATIVAS

Fatores qualitativos freqüentemente aparecem na forma de informação binária: uma pessoa é do sexo feminino ou masculino; alguém possui ou não um computador pessoal; uma firma oferece ou não certo tipo de plano de pensão a seus empregados; um estado adota ou não a pena capital. Em todos esses exemplos, a informação relevante pode ser capturada pela definição de uma **variável binária** ou uma variável zero-um. Em econometria, as variáveis binárias são em geral chamadas **variáveis dummy**, embora esse nome não seja muito descritivo.

Ao definirmos uma variável *dummy*, precisamos decidir a qual evento será atribuído o valor um e a qual será atribuído o valor zero. Por exemplo, em um estudo sobre a determinação do salário individual, podemos definir *feminino* como a variável binária que assumirá o valor um quando a pessoa for mulher e zero, quando homem. Neste caso, o nome indica o evento cujo valor é um. A mesma informação é transmitida se definirmos que masculino será um, se a pessoa for homem, e zero, se mulher.

Qualquer uma dessas formas é melhor que usarmos *gênero*, porque esse nome não deixa claro quando a variável *dummy* é um: gênero = 1 corresponde a homem ou a mulher? A maneira pela qual denominamos nossas variáveis não tem importância para obtermos os resultados da regressão, mas sempre ajuda a escolhermos nomes que deixem claras as equações e as explicações.

Exemplos de variáveis binárias

Suponha que, em um estudo que compara resultados de eleições entre candidatos democratas e republicanos, você queira indicar o partido de cada candidato. Um nome como *partido* será uma boa escolha para uma variável binária neste caso? Qual seria um nome melhor?

Suponha que no exemplo do salário tenhamos escolhido o nome *feminino* para indicar o sexo. Além disso, definimos uma variável binária *casado* como igual a um se a pessoa for casada, e zero, caso contrário. A Tabela 7.1 fornece uma listagem parcial de um possível conjunto de dados sobre salários. Vemos que a Pessoa 1 é do sexo feminino e não é casada, a Pessoa 2 é do sexo feminino e é casada, a Pessoa 3 é do sexo masculino e não é casada, e assim por diante.

Por que usamos os valores zero e um para descrever informações qualitativas? Em certo sentido, esses valores são arbitrários: quaisquer dois valores diferentes serviriam. O benefício real de capturar informação qualitativa usando variáveis zero-um é que elas levam a modelos de regressão nos quais os parâmetros têm interpretações bastante naturais, como veremos agora.

Tabela 7.1

Uma Listagem Parcial dos Dados do Arquivo WAGE1.RAW

pessoa	salárioh	educ	exper	feminino	casado
1	3,10	11	2	1	0
2	3,24	12	22	1	1
3	3,00	11	2	0	0
4	6,00	8	44	0	1
5	5,30	12	7	0	1
.
.
.
525	11,56	16	5	0	1
526	3,50	14	5	1	0

7.2 UMA ÚNICA VARIÁVEL DUMMY INDEPENDENTE

Como incorporamos informações binárias em modelos de regressão? No caso mais simples, com somente uma variável *dummy* explicativa, simplesmente adicionamos a variável à equação como uma variável independente. Por exemplo, considere o seguinte modelo simples de determinação de salários por hora:

$$\text{salário}_h = \beta_0 + \delta_0 \text{feminino} + \beta_1 \text{educ} + u. \quad (7.1)$$

Usamos δ_0 como o parâmetro da variável *feminino* de maneira a ressaltar a interpretação dos parâmetros que multiplicam variáveis *dummy*; mais adiante, usaremos a notação que for mais conveniente.

No modelo (7.1), somente dois fatores observados afetam os salários: gênero e educação. Como *feminino* = 1 quando a pessoa é mulher e *feminino* = 0 quando a pessoa é homem, o parâmetro δ_0 tem a seguinte interpretação: δ_0 é a diferença no salário por hora entre mulheres e homens, *dado* o mesmo grau de educação (e o mesmo termo erro u). Assim, o coeficiente δ_0 determina se existe discriminação contra as mulheres: se $\delta_0 < 0$, então, para o mesmo nível dos outros fatores, as mulheres ganham menos que os homens, em média.

Em termos de expectativas, se assumirmos a hipótese de média condicional zero $E(u|\text{feminino}, \text{educ}) = 0$, então

$$\delta_0 = E(\text{salário}_h | \text{feminino} = 1, \text{educ}) - E(\text{salário}_h | \text{feminino} = 0, \text{educ}).$$

Como *feminino* = 1 corresponde a mulheres e *feminino* = 0 corresponde a homens, podemos escrever essa expressão de forma mais simples:

$$\delta_0 = E(\text{salário}_h | \text{feminino}, \text{educ}) - E(\text{salário}_h | \text{masculino}, \text{educ}). \quad (7.2)$$

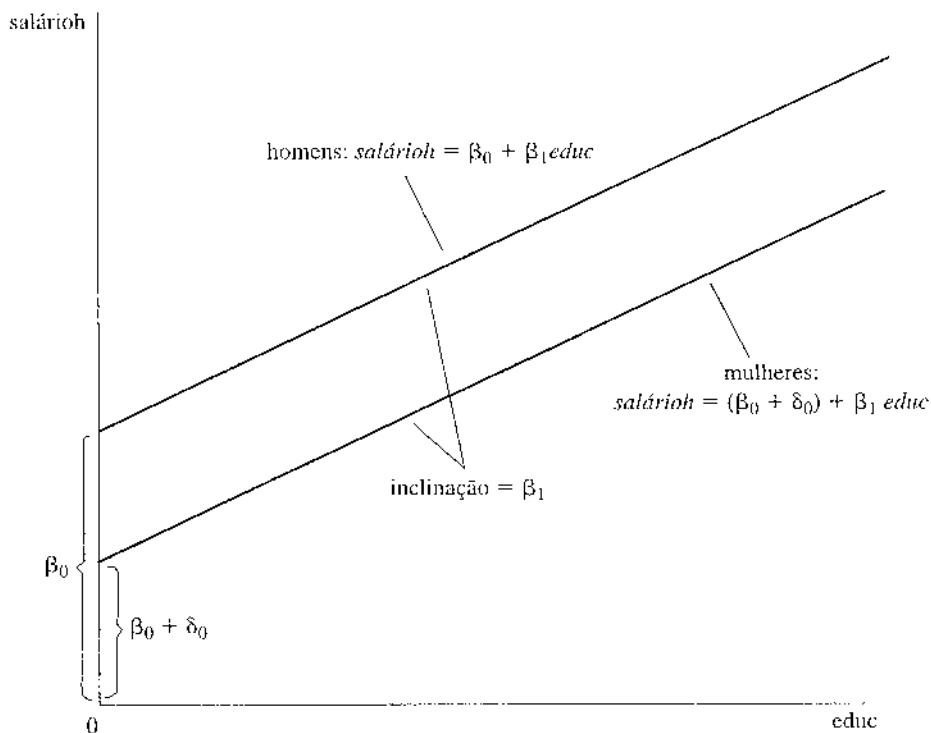
O importante aqui é que o nível de educação é o mesmo em ambas as expectativas; a diferença, δ_0 , deve-se somente ao gênero.

A situação pode ser descrita graficamente como um **deslocamento de intercepto** entre as linhas que representam homens e mulheres. Na Figura 7.1, o caso $\delta_0 < 0$ é mostrado, de modo que os homens ganham um montante fixo por hora a mais que as mulheres. A diferença não depende do nível de educação, e isso explica a razão de os perfis salário-educação das mulheres e dos homens serem paralelos.

Neste ponto, você pode estar se perguntando por que não incluímos, também, em (7.1) uma variável *dummy*, digamos, *masculino*, que seria um para homens e zero para mulheres. A razão é que isso seria redundante. Na equação (7.1), o intercepto para homens é β_0 , enquanto o intercepto para mulheres é $\beta_0 + \delta_0$. Como existem apenas dois grupos, precisamos de apenas dois interceptos diferentes. Isso significa que, além de β_0 , precisamos usar somente *uma* variável *dummy*; decidimos incluir a variável *dummy* para mulheres. O uso de duas variáveis *dummy* introduziria colinearidade perfeita, porque *feminino* + *masculino* = 1, o que significa que *masculino* é uma função linear perfeita de *feminino*. A inclusão de variáveis *dummy* para ambos os sexos é o exemplo mais simples da chamada **armadilha da variável dummy**, que surge quando um grande número de variáveis *dummy* descreve determinado número de grupos. Discutiremos esse assunto mais adiante.

Figura 7.1

Gráfico de $\text{salárioh} = \beta_0 + \delta_0 \text{feminino} + \beta_1 \text{educ}$ para $\delta_0 < 0$.



Na equação (7.1) acima, escolhemos homens para ser o **grupo base** ou o **grupo de referência**, isto é, o grupo contra o qual as comparações são feitas. Esta é a razão pela qual β_0 é o intercepto para os homens, e δ_0 é a *diferença* dos interceptos entre mulheres e homens. Poderíamos ter escolhido as mulheres como o grupo base, escrevendo o modelo como

$$\text{salárioh} = \alpha_0 + \gamma_0 \text{masculino} + \beta_1 \text{educ} + u,$$

onde o intercepto para mulheres é α_0 e o intercepto para homens é $\alpha_0 + \gamma_0$; isso implica que $\alpha_0 = \beta_0 + \delta_0$ e $\alpha_0 + \gamma_0 = \beta_0$. Em qualquer aplicação, não importa como escolhemos o grupo base, mas é importante estar atento para qual é o grupo base.

Alguns pesquisadores preferem eliminar o intercepto global do modelo e incluir variáveis *dummy* para cada grupo. A equação então seria $\text{salárioh} = \beta_0 \text{masculino} + \alpha_0 \text{feminino} + \beta_1 \text{educ} + u$, onde o intercepto para os homens é β_0 e o intercepto para as mulheres é α_0 . Não existe armadilha da variável *dummy* neste caso, porque não temos um intercepto global. Porém, essa formulação tem pouco a oferecer, já que é mais difícil verificar diferenças nos interceptos, e não existe uma maneira consensual de computar o *R*-quadrado em regressões sem intercepto. Portanto, sempre incluiremos um intercepto global para o grupo base.

Nada mais muda muito quando mais variáveis explicativas estão envolvidas. Considerando os homens como o grupo base, um modelo que controla a experiência e a permanência, além da educação é

$$\text{salárioh} = \beta_0 + \delta_0 \text{feminino} + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{perm} + u. \quad (7.3)$$

Se *educ*, *exper* e *perm* forem todas características relevantes da produtividade, a hipótese nula de não-existência de diferença entre homens e mulheres será $H_0: \delta_0 = 0$. A hipótese alternativa de que existe discriminação contra as mulheres será $H_1: \delta_0 < 0$.

Como podemos efetivamente testar a discriminação salarial? A resposta é simples: simplesmente estimamos o modelo por MQO, *exatamente* como antes, e usamos a estatística *t* habitual. Nada muda na inecônica do MQO ou na teoria estatística quando algumas das variáveis independentes são definidas como variáveis *dummy*. A única diferença em relação ao que vínhamos fazendo até agora é a interpretação do coeficiente da variável *dummy*.

EXEMPLO 7.1

(Equação dos Salários por Hora)

Utilizando os dados contidos no arquivo WAGE1.RAW, estimamos o modelo (7.3). Por enquanto, usamos *salárioh*, em vez de $\log(\text{salárioh})$, como a variável dependente:

$$\begin{aligned} \text{salárioh} = & -1,57 - 1,81 \text{feminino} + 0,572 \text{educ} \\ & (0,72) \quad (0,26) \quad (0,049) \\ & + 0,025 \text{exper} + 0,141 \text{perm} \\ & (0,012) \quad (0,021) \\ & n = 526, R^2 = 0,364. \end{aligned} \quad (7.4)$$

O intercepto negativo — o intercepto para os homens, neste caso — não é muito significativo, já que ninguém na amostra tem anos de *educ*, *exper* e *perm* próximos de zero. O coeficiente de *feminino* é interessante, porque ele registra a diferença média no salário por hora entre uma mulher e um homem, dados os mesmos níveis de *educ*, *exper* e *perm*. Se compararmos uma mulher e um homem com os mesmos níveis de educação, experiência e permanência, a mulher ganha, em média, 1,81 dólares por hora a menos que o homem. (Não se esqueça de que estamos tratando de salários de 1976.)

É importante lembrarmos que, como fizemos uma regressão múltipla e controlamos *educ*, *exper* e *perm*, o diferencial de 1,81 dólares no salário não pode ser explicado por diferentes níveis médios de educação, experiência ou permanência entre homens e mulheres. Podemos concluir que o diferencial de 1,81 dólares é devido ao gênero ou a fatores associados ao gênero que não tenhamos controlado na regressão.

É esclarecedor comparar o coeficiente de *feminino* na equação (7.4) com a estimativa que obtemos quando todas as outras variáveis explicativas são eliminadas da equação:

EXEMPLO 7.1 (continuação)

$$\begin{aligned} \text{salárioh} &= 7,10 - 2,51 \text{feminino} \\ &\quad (0,21) \quad (0,30) \\ n &= 526, R^2 = 0,116 \end{aligned} \tag{7.5}$$

Os coeficientes em (7.5) têm uma interpretação simples. O intercepto é o salário-hora médio dos homens na amostra (*feminino* = 0), de modo que os homens ganham, em média, 7,10 dólares por hora. O coeficiente de *feminino* é a diferença no salário médio entre homens e mulheres. Assim, o salário médio das mulheres, na amostra, é $7,10 - 2,51 = 4,59$, ou 4,59 dólares por hora. (A propósito, existem 274 homens e 252 mulheres na amostra.)

A equação (7.5) apresenta um modo simples de realizar um teste de comparação de médias entre os dois grupos, que neste caso são homens e mulheres. A diferença estimada $-2,51$ tem uma estatística *t* de $-8,37$, que é, estatisticamente, bastante significante (e, claro, 2,51 dólares também é grande economicamente). Geralmente, a regressão simples sobre uma constante e uma variável *dummy* é uma maneira bastante objetiva de comparar as médias de dois grupos. Para que o teste *t* habitual seja válido, temos que assumir a manutenção da hipótese de homoscedasticidade, o que significa que a variância populacional dos salários dos homens é a mesma dos salários das mulheres.

O diferencial salarial estimado entre homens e mulheres é maior em (7.5) que na equação (7.4), porque (7.5) não controla as diferenças em educação, experiência e permanência, e esses fatores são mais baixos, em média, para as mulheres do que para os homens nessa amostra. A equação (7.4) fornece uma estimativa mais confiável da discrepância salarial *ceteris paribus* entre os sexos; ela ainda indica um diferencial bastante grande.

Em muitos casos, variáveis *dummy* independentes refletem escolhas de indivíduos ou de outras unidades econômicas (em oposição a algo predeterminado, como gênero). Em tais situações, a questão da causalidade é novamente crucial. No exemplo seguinte, gostaríamos de saber se o fato de possuir um computador pessoal *causa* uma nota média mais elevada em curso superior.

EXEMPLO 7.2**{Efeitos de se Possuir Computadores na Avaliação em Cursos Superiores}**

Para determinar os efeitos de se possuir um computador na nota média em curso superior, estimamos o modelo

$$nmgrad = \beta_0 + \delta_0 PC + \beta_1 nmem + \beta_2 tac + u,$$

onde a variável *dummy* *PC* é igual a um se um aluno possui um computador pessoal e zero caso contrário. Existem várias razões pelas quais a propriedade de um computador pessoal pode ter um efeito sobre *nmgrad*. O trabalho de um aluno pode ser de melhor qualidade se feito em um computador e ele pode ganhar tempo por não ter que ficar esperando sua vez em um laboratório de informática. É claro que o aluno pode estar mais inclinado a brincar com jogos ou a navegar na Internet se ele, ou ela, possuir seu próprio PC,

EXEMPLO 7.2 {continuação}

de modo que não é óbvio que δ_0 será positivo. As variáveis *nmem* (nota média do ensino médio) e *tac* (nota do teste de avaliação para ingresso em curso superior) são usadas como controles: pode ser que alunos mais fortes, medidos pelas nota média do ensino médio e nota do teste de avaliação, provavelmente possuam computadores. Controlamos esses fatores porque gostaríamos de saber o efeito médio sobre *nmgrad* se escolhermos um aluno aleatoriamente e dermos a ele um computador pessoal.

Utilizando os dados contidos no arquivo GPA1.RAW, obtemos

$$\begin{aligned} \hat{nmgrad} &= 1,26 + 0,157 PC + 0,447 nmem + 0,0087 tac \\ (0,33) &\quad (0,057) \quad (0,094) \quad (0,0105) \end{aligned} \tag{7.6}$$

$n = 141, R^2 = 0,219.$

Esta equação sugere para um aluno que possua um computador pessoal uma nota média prevista em torno de 0,16 pontos mais alta quando comparada com a de um aluno que não possui um PC (lembre-se, tanto *nmgrad* como *nmem* estão em uma escala de quatro pontos). O efeito também é, estatisticamente, bastante significante, com $t_{PC} = 0,157/0,057 \approx 2,75$.

O que acontece se eliminarmos *nmem* e *tac* da equação? Claramente, a eliminação da última variável deve ter um efeito muito pequeno, já que seu coeficiente e a respectiva estatística *t* são muito pequenos. Mas *nmem* é bastante significante, e assim sua eliminação pode afetar a estimativa de β_{PC} . A regressão de *nmgrad* sobre *PC* produz a estimativa do coeficiente de *PC* igual a, aproximadamente, 0,170, com um erro-padrão de 0,063; neste caso, $\hat{\beta}_{PC}$ e sua estatística *t* não mudam muito.

Nos exercícios no final do capítulo será solicitado o controle de outros fatores na equação para verificar se o efeito de se possuir um computador desaparece, ou se pelo menos se torna significativamente menor.

Cada um dos exemplos anteriores pode ser entendido como relevante para a **análise de políticas públicas**. No primeiro exemplo estávamos interessados na discriminação sexual na força de trabalho. No segundo exemplo estávamos preocupados com o efeito de se possuir um computador sobre o desempenho no curso superior. Um caso especial de análise de políticas públicas é a **avaliação de programas**, na qual gostaríamos de saber o efeito de programas econômicos ou sociais sobre os indivíduos, empresas, vizinhança, cidades etc.

No caso mais simples existem dois grupos de objetos de estudo. O **grupo de controle** não participa do programa. O **grupo experimental** ou **grupo de tratamento** faz parte do programa. Esses nomes provêm da literatura das ciências experimentais, não devem ser interpretados literalmente. Exceto em casos raros, a escolha dos grupos de controle e de tratamento não é feita aleatoriamente. Porém, em alguns casos, a análise de regressão múltipla pode ser usada para controlar um número suficiente de outros fatores para estimar o efeito causal do programa.

EXEMPLO 7.3**{Efeitos da Concessão de Subsídios sobre as Horas de Treinamento}**

Utilizando os dados de 1988 das indústrias de Michigan contidas no arquivo JTRAIN.RAW, obteremos a seguinte equação estimada:

EXEMPLO 7.3 (continuação)

$$\begin{aligned}
 hrs\hat{emp} &= 46,67 + 26,25 subs - 0,98 \log(vendas) \\
 (43,41) &\quad (5,59) \quad (3,54) \\
 &- 6,07 \log(empreg) \quad (7.7) \\
 &\quad (3,88) \\
 n &= 105, R^2 = 0,237.
 \end{aligned}$$

A variável dependente é horas de treinamento por empregado, ao nível da empresa. A variável *subs* é uma variável *dummy* igual a um se a firma recebeu um subsídio para treinamento em 1988 e zero, caso contrário. As variáveis *vendas* e *empreg* representam as vendas anuais e o número de empregados, respectivamente. Não podemos usar *hrsmp* na forma logarítmica porque ela tem valor zero para 20 das 105 empresas usadas na regressão.

A variável *subs* é estatisticamente bastante significante, com $t_{subs} = 4,70$. Controlando vendas e emprego, as empresas que receberam subsídios treinaram cada um de seus empregados, em média, 26,25 horas mais que as outras. Como o número médio de horas de treinamento por empregado na amostra está em torno de 17, com um valor máximo de 164, *subs* tem um grande efeito sobre o treinamento, como o esperado.

O coeficiente de $\log(vendas)$ é pequeno e não significante. O coeficiente de $\log(empreg)$ significa que, se a empresa for 10% maior, ela treinará seus empregados cerca de 0,61 horas menos. Sua estatística *t* é $-1,56$, que é somente marginalmente significante, em termos estatísticos.

Assim como em relação a qualquer outra variável independente, devemos perguntar se o efeito mensurado de uma variável qualitativa é causal. Na equação (7.7), a diferença de treinamento entre as empresas que recebem subsídios e as que não recebem se deve aos subsídios, ou o recebimento de um subsídio é simplesmente um indicador de alguma outra coisa? É possível que as empresas que recebem subsídios tenham, em média, treinado seus empregados de modo mais regular, mesmo sem os subsídios. Nada nesta análise nos informa se estimamos um efeito causal; precisamos saber como foram determinadas as empresas que receberiam subsídios. Somente podemos esperar que tenhamos controlado tantos fatores quanto possível de modo que possam estar relacionados à questão de a empresa ter recebido subsídios e com os seus níveis de treinamento.

Retornaremos à análise de políticas públicas com variáveis *dummy* na Seção 7.6, como também em outros capítulos.

A Interpretação dos Coeficientes de Variáveis Dummy Explicativas quando a Variável Dependente É Expressa como $\log(y)$

Uma especificação comum em trabalhos aplicados tem a variável dependente aparecendo na forma logarítmica, com uma ou mais variáveis *dummy* aparecendo como variáveis independentes. Como interpretamos os coeficientes das variáveis *dummy* neste caso? Não surpreendentemente, os coeficientes têm uma interpretação *percentual*.

EXEMPLO 7.4**(Regressão dos Preços de Imóveis)**

Utilizando os dados contidos no arquivo HPRICE1.RAW, obtemos a equação

$$\begin{aligned}
 \log(\text{preço}) &= 5,56 + 0,168 \log(\text{tamterr}) + 0,707 \log(\text{arquad}) \\
 (0,65) &\quad (0,038) && (0,093) \\
 &+ 0,027 \log(\text{qtdorm}) + 0,054 \text{ colonial} \\
 (0,029) &\quad (0,045) \\
 n = 88, R^2 &= 0,649.
 \end{aligned} \tag{7.8}$$

Todas as variáveis são auto-explicativas, exceto *colonial*, que é uma variável binária igual a um se o imóvel tiver estilo colonial. O que significa o coeficiente de *colonial*? Para níveis dados de *tamterr*, *arquad* e *qtdorm*, a diferença em log (preço) entre um imóvel de estilo colonial e outro de outro estilo é 0,054. Isso significa prever que um imóvel de estilo colonial seja vendido por cerca de 5,4% a mais, mantendo-se todos os outros fatores fixos.

Este exemplo mostra que, quando $\log(y)$ é a variável dependente em um modelo, o coeficiente de uma variável *dummy*, quando multiplicado por 100, é interpretado como a diferença percentual em y , mantendo fixos todos os outros fatores. Quando o coeficiente de uma variável *dummy* sugere uma grande mudança proporcional em y , a diferença percentual exata pode ser obtida exatamente como no cálculo da semi-elasticidade na Seção 6.2.

EXEMPLO 7.5**(Equação Log do Salário-Hora)**

Reestimemos a equação salarial do exemplo 7.1, usando $\log(\text{salárioh})$ como a variável dependente e adicionando termos quadráticos em *exper* e *perm*:

$$\begin{aligned}
 \log(\text{salárioh}) &= 0,417 - 0,297 \text{feminino} + 0,080 \text{educ} + 0,029 \text{exper} \\
 (0,099) &\quad (0,036) && (0,007) && (0,005) \\
 &- 0,00058 \text{exper}^2 + 0,032 \text{perm} - 0,00059 \text{perm}^2 \\
 (0,00010) &\quad (0,007) && (0,00023) \\
 n = 526, R^2 &= 0,441.
 \end{aligned} \tag{7.9}$$

Usando a mesma aproximação do Exemplo 7.4, o coeficiente de *feminino* implica que, para os mesmos níveis de *educ*, *exper* e *perm*, as mulheres ganham cerca de $100(0,297) = 29,7\%$ a menos que os homens. Podemos fazer melhor que isso ao computarmos a diferença percentual exata nos salários previstos. O que

EXEMPLO 7.5 (continuação)

queremos é a diferença proporcional nos salários entre mulheres e homens, mantendo fixos todos os outros fatores: $(\text{salário}_{hM} - \text{salário}_{hH})/\text{salário}_{hH}$. O que temos a partir de (7.9) é

$$\hat{\log}(\text{salário}_{hM}) - \hat{\log}(\text{salário}_{hH}) = 0,297.$$

Fazendo a exponenciação e a subtração temos

$$(\text{salário}_{hM} - \text{salário}_{hH})/\text{salário}_{hH} = \exp(-0,297) - 1 \approx -0,257.$$

Esta estimativa mais exata implica que o salário de uma mulher é, em média, 25,7% menor que o salário de um homem nas mesmas condições.

Se tivéssemos feito a mesma correção no Exemplo 7.4, teríamos obtido $\exp(0,054) - 1 \approx 0,0555$, ou cerca de 5,6%. A correção tem um efeito menor no Exemplo 7.4 do que no exemplo salarial, porque a magnitude do coeficiente da variável *dummy* é muito menor em (7.8) do que em (7.9).

De forma geral, se $\hat{\beta}_1$ for o coeficiente de uma variável *dummy*, digamos, x_1 , quando $\log(y)$ é a variável dependente, a diferença percentual exata em y previsto quando $x_1 = 1$ versus quando $x_1 = 0$ é

$$100 \cdot [\exp(\hat{\beta}_1) - 1]. \quad (7.10)$$

O coeficiente $\hat{\beta}_1$ estimado pode ser positivo ou negativo, e é importante preservar seu sinal ao computar (7.10).

7.3 O USO DE VARIÁVEIS DUMMY PARA CATEGORIAS MÚLTIPLAS

Podemos usar diversas variáveis *dummy* independentes na mesma equação. Por exemplo, poderíamos adicionar a variável *dummy casado* na equação (7.9). O coeficiente de *casado* fornece o diferencial proporcional (aproximado) nos salários entre aqueles que são, ou não, casados, mantendo fixos gênero, *educ*, *exper* e *perm*. Quando estimamos esse modelo, o coeficiente de *casado* (com o erro-padrão entre parênteses) é 0,053 (0,041), e o coeficiente de *feminino* passa a ser -0,290 (0,036). Assim, o “prêmio” por ser casado é estimado em torno de 5,3%, mas não é estatisticamente diferente de zero ($t = 1,29$). Uma limitação importante deste modelo é que o prêmio por ser casado é assumido como o mesmo para homens e mulheres; isso é relaxado no exemplo a seguir.

EXEMPLO 7.6**(Equação do Log do Salário-Hora)**

Estimemos um modelo que considere diferenças salariais entre quatro grupos: homens casados, mulheres casadas, homens solteiros e mulheres solteiras. Para fazermos isso temos que selecionar um grupo base; escolhemos homens solteiros. Então, devemos definir as variáveis *dummy* para cada um dos demais grupos. Vamos chamá-los *hcasados*, *mcasadas* e *msoalteiras*. Colocando essas três variáveis na equação (7.9) (e, claro, eliminando *feminino*, já que agora ela é redundante) produz

EXEMPLO 7.6 (continuação)

$$\begin{aligned}
 \log(salário_{ih}) = & 0,321 + 0,213 hcasados - 0,198 mcasadas \\
 & (0,100) (0,055) \quad (0,058) \\
 & - 0,110 msolteiras + 0,079 educ + 0,027 exper - 0,00054 exper^2 \\
 & (0,056) \quad (0,007) \quad (0,005) \quad (0,00011) \quad \boxed{7.11} \\
 & + 0,029 perm - 0,00053 perm^2 \\
 & (0,007) \quad (0,00023) \\
 n = 526, R^2 = & 0,461.
 \end{aligned}$$

Todos os coeficientes, exceto o de *msolteiras*, têm estatísticas *t* bem acima de dois, em valores absolutos. A estatística *t* de *msolteiras* está em torno de $-1,96$, que é significante apenas ao nível de 5% contra uma alternativa bilateral.

Para interpretar os coeficientes das variáveis *dummy*, devemos nos lembrar de que o grupo base é o de homens solteiros. Assim, as estimativas das três variáveis *dummy* medem a diferença proporcional nos salários *relativamente* aos homens solteiros. Por exemplo, estima-se que os homens casados ganhem cerca de 21,3% mais que os homens solteiros, mantendo fixas educação, experiência e permanência. [A estimativa mais precisa a partir de (7.10) está em torno de 23,7%.] Uma mulher casada, no entanto, deve ganhar 19,8% menos que um homem solteiro com os mesmos níveis das outras variáveis.

Como o grupo base é representado pelo intercepto na equação (7.11), incluímos variáveis *dummy* para apenas três dos quatro grupos. Se tivéssemos incluído uma variável *dummy* para homens solteiros em (7.11) cairíamos na armadilha da variável *dummy*, por termos introduzido colinearidade perfeita. Alguns programas de regressão corrigem automaticamente esse engano, enquanto outros apenas informam a existência de colinearidade perfeita. É melhor especificar cuidadosamente as variáveis *dummy*, pois isso nos forçará a interpretar apropriadamente o modelo final.

Embora os homens solteiros sejam o grupo base em (7.11), podemos usar essa equação para obter a diferença estimada entre dois grupos quaisquer. Como o intercepto global é comum a todos os grupos, podemos ignorá-lo quando procuramos diferenças. Assim, a diferença proporcional estimada entre as mulheres solteiras e as casadas é $-0,110 - (-0,198) = 0,088$, o que significa que as mulheres solteiras ganham cerca de 8,8% mais que as mulheres casadas. Infelizmente, não podemos usar a equação (7.11) para verificar se a diferença estimada entre as mulheres solteiras e as casadas é estatisticamente significante. O conhecimento dos erros-padrão de *mcasadas* e *msolteiras* não é suficiente para realizar o teste (veja Seção 4.4). O mais fácil a fazer é selecionar um desses grupos para ser o grupo base e reestimar a equação. Nada de substancial mudará, mas obteremos a estimativa necessária e seu erro-padrão de forma direta. Quando usarmos as mulheres casadas como o grupo base, obtemos

$$\begin{aligned}
 \log(salário_{ih}) = & 0,123 + 0,411 hcasados + 0,198 hsolteiros + 0,088 msolteiras + \dots, \\
 & (0,106) \quad (0,056) \quad (0,058) \quad (0,052)
 \end{aligned}$$

onde, é claro, nenhum dos coeficientes ou erros-padrão não descritos sofreram alterações. A estimativa do coeficiente de *msolteiras* é, como esperado, 0,088. Agora temos um erro-padrão para acompanhar essa esti-

EXEMPLO 7.6 {continuação}

inativa. A estatística t para a hipótese nula de que não existe diferença na população entre mulheres casadas e solteiras é $t_{msolteiras} = 0,88/0,052 \approx 1,69$. Essa é uma evidência marginal contra a hipótese nula. Também vemos que a diferença estimada entre os homens casados e as mulheres casadas é, estatisticamente, muito significante ($t_{casados} = 7,34$).

O exemplo anterior ilustra um princípio geral para a inclusão de variáveis *dummy* que indicam grupos diferentes: se o modelo de regressão deve ter diferentes interceptos para, digamos, g grupos ou categorias, precisamos incluir $g - 1$ variáveis *dummy* no modelo, juntamente com um intercepto. O intercepto do grupo base é o intercepto global no modelo, e o coeficiente da variável *dummy* de um determinado grupo representa a diferença estimada nos interceptos entre aquele grupo e o grupo base. A inclusão de g variáveis *dummy* juntamente com um intercepto resultará na armadilha da variável *dummy*. Uma alternativa é incluir g variáveis *dummy* e excluir um intercepto global. Isso não é recomendável, pois o teste de diferenças relativas a um grupo base se tornará difícil, e alguns programas de regressão alteram a maneira como o *R*-quadrado é computado quando a regressão não contém um intercepto.

Nos dados sobre salários dos jogadores de beisebol encontrados no arquivo MLB1.RAW, os jogadores ocupam uma de seis posições: *pibase*, *segbase*, *terbase*, *interbase*, *jardext* ou *receptor*. Para possibilitar diferenças salariais entre as posições, com determinados defensores (*jardext*) como o grupo base, quais variáveis *dummy* você incluiria como variáveis independentes?

Incorporação de Informações Ordinais com o Uso de Variáveis Dummy

Suponha que gostaríamos de estimar o efeito do risco de crédito das cidades sobre as taxas de juros dos títulos públicos municipais (*TTM*). Várias instituições financeiras, como a Moody's Investment Service e a Standard and Poor's, classificam a qualidade da dívida de governos locais, na qual a classificação depende de fatores como a probabilidade de inadimplência (governos locais preferem taxas menores de juros para reduzir seus custos de empréstimos). À guisa de simplicidade, suponha que a classificação varie de zero a quatro, na qual zero é o pior risco de crédito e quatro, o melhor. Este é um exemplo de uma variável ordinal. Vamos chamar essa variável de *CR* por definição. A questão com a qual devemos tratar é: Como incorporamos a variável *CR* em um modelo para explicar a variável *TTM*?

Uma possibilidade é apenas incluir *CR* como incluiríamos qualquer outra variável explicativa:

$$TTM = \beta_0 + \beta_1 CR + \text{outros fatores},$$

onde deliberadamente não mostramos quais são os outros fatores. Neste caso, β_1 é a mudança em pontos percentuais em *TTM* quando *CR* aumenta uma unidade, mantendo fixos todos os outros fatores. Infelizmente, é bastante difícil interpretar um aumento de uma unidade em *CR*. Sabemos o significado quantitativo de mais um ano de educação, ou de um dólar a mais gasto por aluno, mas fatores como risco de crédito, em geral, têm apenas significado ordinal. Sabemos que um *CR* de quatro é melhor que

um CR de três, mas será que a diferença entre quatro e três é a mesma que a diferença entre um e zero? Se não, não fará sentido assumir que um aumento de uma unidade em CR terá um efeito constante sobre TTM .

Uma abordagem melhor que podemos implementar, pois CR assume relativamente poucos valores, é definir variáveis *dummy* para cada valor de CR . Assim, definimos $CR_1 = 1$ se $CR = 1$, e, caso contrário, $CR_1 = 0$; $CR_2 = 1$ se $CR = 2$ e, caso contrário, $CR_2 = 0$, e assim por diante. Na realidade, levamos em conta o risco de crédito e o transformamos em cinco categorias. Desta forma podemos estimar o modelo

$$TTM = \beta_0 + \delta_1 CR_1 + \delta_2 CR_2 + \delta_3 CR_3 + \delta_4 CR_4 + \text{outros fatores.} \quad (7.12)$$

Seguindo nossa regra sobre inclusão de variáveis *dummy* em um modelo, incluímos quatro variáveis *dummy*, já que temos cinco categorias. A categoria aqui omitida é risco de crédito zero, e portanto ela é o grupo base. (Esta é a razão pela qual não precisamos definir uma variável *dummy* para esta categoria.) Os coeficientes são de fácil interpretação: δ_1 é a diferença em TTM (outros fatores fixos) entre uma cidade com um risco de crédito um e uma cidade com um risco de crédito zero; δ_2 é a diferença em TTM entre uma cidade com um risco de crédito dois e uma cidade com um risco de crédito zero; e assim por diante. O movimento entre os índices de risco de crédito tem efeitos diferentes, de modo que o uso de (7.12) é muito mais flexível do que simplesmente considerarmos CR uma variável única. Uma vez definidas as variáveis *dummy*, estimar (7.12) é simples.

No modelo (7.12), como você testaria a hipótese nula de que o risco de crédito não tem efeito sobre TTM ?

A equação (7.12) contém o modelo com um efeito parcial constante, sendo este um caso especial. Uma maneira de escrever as três restrições que implicam um efeito parcial constante é $\delta_2 = 2\delta_1$, $\delta_3 = 3\delta_1$, e $\delta_4 = 4\delta_1$. Quando incorporamos essas restrições à equação (7.12) e a reorganizamos, obtemos $TTM = \beta_0 + \delta_1 (CR_1 + 2CR_2 + 3CR_3 + 4CR_4) + \text{outros fatores}$. Agora, o termo que multiplica δ_1 é simplesmente a variável original do risco de crédito, CR . Ao obtermos a estatística F para testar as restrições do efeito parcial constante, obtemos o R -quadrado irrestrito de (7.12) e o R -quadrado restrito a partir da regressão de TTM sobre CR e os outros fatores que tenhamos controlado. A estatística F é obtida como na equação (4.41) com $q = 3$.

EXEMPLO 7.7

(Efeitos da Atratividade Física sobre os Salários)

Hamermesh e Biddle (1994) usaram indicadores de boa aparência física em uma equação de salários. Cada pessoa da amostra foi classificada por um entrevistador quanto à aparência, utilizando cinco categorias (feia, comum, média, bonita e muito bonita). Como pouca gente se classifica nos dois extremos, os autores colocaram as pessoas em um dos três grupos para a análise de regressão, média, abaixo da média e acima da média, na qual o grupo base era a média. Utilizando os dados da Quality of Employment Survey (Pesquisa de Qualidade do Emprego) de 1977, após terem sido controladas as características de produtividade habituais, Hamermesh e Biddle estimaram uma equação para homens:

EXEMPLO 7.7 (continuação)

$$\log(\hat{\text{salário}}) = \hat{\beta}_0 + 0,164 \text{abaixomed} + 0,016 \text{acimamed} + \text{outros fatores}$$

$$(0,046) \qquad \qquad \qquad (0,033)$$

$$n = 700, \bar{R}^2 = 0,403$$

e uma equação para mulheres:

$$\log(\hat{\text{salário}}) = \hat{\beta}_0 + 0,124 \text{abaixomed} + 0,035 \text{acimamed} + \text{outros fatores}$$

$$(0,066) \qquad \qquad \qquad (0,049)$$

$$n = 409, \bar{R}^2 = 0,330.$$

Os outros fatores controlados na regressão incluem educação, experiência, permanência, estado civil e raça; veja Tabela 3 no artigo de Hamermesh e Biddle para uma lista mais completa. Para economizar espaço, os coeficientes das outras variáveis e do intercepto não são descritos no artigo.

Entre os homens, o artigo prevê que aqueles com aparência abaixo da média ganham cerca de 16,4% menos que os com aparência média com os mesmos fatores (inclusive educação, experiência, permanência, estado civil e raça). O efeito é estatisticamente diferente de zero, com $t = -3,57$. De modo semelhante, homens com aparência acima da média ganham 1,6% mais, embora o efeito não seja estatisticamente significante ($t < 0,5$).

Uma mulher com aparência abaixo da média ganha cerca de 12,4% menos que outra com aparência média, com $t = -1,88$. Como aconteceu com os homens, a estimativa do coeficiente de *acimamed* não é estatisticamente diferente de zero.

Em alguns casos, a variável ordinal assume um número muito grande de valores, de maneira que não é possível incluir uma variável *dummy* para cada valor. Por exemplo, o arquivo LAW SCH85.RAW contém dados sobre a mediana dos salários iniciais dos formados em faculdades de direito. Uma das principais variáveis explicativas é a classificação da faculdade de direito. Como cada faculdade tem uma posição diferente, evidentemente não podemos incluir uma variável *dummy* para cada posição. Se não quisermos colocar a classificação diretamente na equação, podemos dividi-la em categorias. O exemplo seguinte mostra como isso é feito.

EXEMPLO 7.8**(Efeitos da Classificação das Faculdades de Direito sobre Salários Iniciais)**

Defina as variáveis *dummy*, *r11_25*, *r26_40*, *r41_60* e *r61_100*, assumindo valor unitário quando a variável *rank* cair na faixa apropriada. Definimos o grupo base como sendo os das faculdades classificadas abaixo de 100. A equação estimada é

EXEMPLO 7.8 (continuação)

$$\begin{aligned}
 \log(salário) = & 9,17 + 0,700 top10 + 0,594 r11_25 + 0,375 r26_40 \\
 & (0,41) (0,053) (0,039) (0,034) \\
 & + 0,263 r41_60 + 0,132 r61_100 + 0,0057 lsat \\
 & (0,028) (0,021) (0,0031) \\
 & + 0,014 nmdgrad + 0,036 \log(volbib) + 0,0008 \log(custo) \\
 & (0,074) (0,026) (0,0251) \\
 n = 136, R^2 = & 0,911, \tilde{R}^2 = 0,905.
 \end{aligned} \tag{7.13}$$

Vemos imediatamente que todas as variáveis *dummy* que definem as diferentes classificações são estatisticamente bastante significantes. A estimativa do coeficiente de *r61_100* significa que, mantendo fixos *lsat*, *nmdgrad*, *volbib* e *custo*, o salário mediano de ex-alunos formados em uma escola classificada entre as posições 61 e 100 é cerca de 13,2% mais alto do que aqueles de uma escola classificada abaixo de 100. A diferença entre as dez primeiras e as abaixo de 100 é bastante elevada. Fazendo os cálculos exatos dados na equação (7.10) temos $\exp(0,700) - 1 \approx 1,014$, e assim o salário mediano previsto é mais de 100% mais alto nas dez primeiras escolas do que nas abaixo de 100.

Para indicar se a divisão da classificação em diferentes grupos é um aperfeiçoamento, podemos comparar o *R*-quadrado ajustado de (7.13) com o *R*-quadrado ajustado com *rank* incluído como uma variável única: o primeiro é 0,905 e o segundo é 0,836, de modo que a flexibilidade adicional de (7.13) está garantida.

Curiosamente, quando a classificação é colocada nas categorias dadas (admitidamente arbitrárias), todas as outras variáveis tornam-se não significantes. Aliás, um teste para verificar a significância conjunta de *lsat*, *nmdgrad*, $\log(volbib)$ e $\log(custo)$ produz um *p*-valor de 0,055, que está no limite da significância. Quando *rank* é incluída em sua forma original, o *p*-valor da significância conjunta é zero até quatro casas decimais.

Um comentário final sobre este exemplo: na derivação das propriedades dos mínimos quadrados ordinários, assumimos que tínhamos uma amostra aleatória. A aplicação atual infringe essa hipótese devido à maneira como *rank* foi definida: a classificação de uma faculdade necessariamente depende da classificação das outras escolas da amostra, e portanto os dados não podem representar extrações independentes da população de faculdades de direito. Isso não causa qualquer problema mais grave, desde que o termo erro seja não-correlacionado com as variáveis explicativas.

7.4 INTERAÇÕES ENVOLVENDO VARIÁVEIS DUMMY

Interações entre Variáveis Dummy

Assim como as variáveis com significados quantitativos podem interagir em modelos de regressão, as variáveis *dummy* também podem. Vimos uma ilustração disso no Exemplo 7.6, no qual definimos quatro categorias com base em estado civil e gênero. Aliás, podemos reformular aquele modelo

adicionando um **termo de interação** entre *feminino* e *casado*, onde essas variáveis apareçam separadamente. Isso possibilita que o prêmio por ser casado dependa do gênero, como era o caso em (7.11). Com o propósito de comparação, o modelo estimado com o termo de interação *feminino-casado* é

$$\begin{aligned} \log(\hat{\text{salário}}) &= 0,321 - 0,110 \text{feminino} + 0,213 \text{casado} \\ &\quad (0,100) \quad (0,056) \quad (0,055) \\ &\quad - 0,301 \text{feminino} \cdot \text{casado} + \dots, \\ &\quad (0,072) \end{aligned} \tag{7.14}$$

onde o restante da regressão será necessariamente idêntico a (7.11). A equação (7.14) mostra explicitamente que existe uma interação estatisticamente significante entre gênero e estado civil. Este modelo também permite obter o diferencial estimado de salários entre todos os quatro grupos, mas aqui devemos ter o cuidado de inserir a correta combinação de valores zero e um.

A definição *feminino* = 0 e *casado* = 0 corresponde ao grupo de homens solteiros, que é o grupo base, já que isso elimina *feminino*, *casado*, e *feminino-casado*. Podemos encontrar o intercepto de homens casados definindo *feminino* = 0 e *casado* = 1 em (7.14); isso produz um intercepto de $0,321 + 0,213 = 0,534$, e assim por diante.

A equação (7.14) é apenas uma maneira diferente de encontrar diferenciais de salários entre todas as combinações de gênero e estado civil. Ela nos possibilita facilmente testar a hipótese nula de que o diferencial de gênero não depende do estado civil (e, de forma equivalente, que o diferencial de estado civil não depende do gênero). A equação (7.11) é mais conveniente para testarmos os diferenciais salariais entre qualquer grupo e o grupo base de homens solteiros.

EXEMPLO 7.9

(Efeitos da Utilização de Computadores nos Salários)

Krueger (1993) estima os efeitos da utilização de computadores sobre os salários. Ele define uma variável *dummy*, que chamaremos *comptrab*, igual a um se a pessoa usa um computador no trabalho. Outra variável *dummy*, *compcasa*, é igual a um se a pessoa usa um computador em casa. Utilizando 13.379 pessoas da Current Population Survey (Censo Populacional Corrente) de 1989, Krueger (1993, Tabela 4) obtém

$$\begin{aligned} \log(\hat{\text{salário}}) &= \hat{\beta}_0 + 0,177 \text{comptrab} + 0,070 \text{compcasa} \\ &\quad (0,009) \quad (0,019) \\ &\quad + 0,017 \text{comptrab} \cdot \text{compcasa} + \text{outros fatores.} \\ &\quad (0,023) \end{aligned} \tag{7.15}$$

(Os outros fatores são os padrões para regressões de salários, inclusive educação, experiência, gênero e estado civil; veja o ensaio de Krueger para a lista exata.) Krueger não descreve o intercepto porque ele não é importante; tudo que precisamos saber é que o grupo base consiste de pessoas que não usam computador em casa ou no trabalho. Vale a pena observar que o retorno estimado do uso de computador no trabalho

EXEMPLO 7.9 (continuação)

(mas não em casa) está em torno de 17,7%. (A estimativa mais precisa é de 19,4%.) De forma semelhante, as pessoas que usam o computador em casa, não no trabalho, têm um prêmio salarial em torno de 7% sobre as que não usam computador, em casa ou no trabalho. O diferencial entre os que usam o computador em ambos os locais e os que simplesmente não usam computador é de cerca de 26,4% (obtida com a adição dos três coeficientes e multiplicando o resultado por 100), ou a estimativa mais precisa de 30,2% obtida da equação (7.10).

O termo de interação em (7.15) não é estatisticamente significante, e tampouco muito grande economicamente. Entretanto, sua inclusão prejudica muito a equação.

Consideração de Inclinações Diferentes

Vimos vários exemplos de como permitir diferentes interceptos para qualquer número de grupos em um modelo de regressão múltipla. Também existem casos de interação de variáveis *dummy* com variáveis explicativas que não são *dummy* para permitir uma **diferença nas inclinações**. Continuando com o exemplo salarial, suponha que queremos verificar se o retorno da educação é o mesmo para homens e mulheres, considerando um diferencial de salários constante entre homens e mulheres (um diferencial do qual já encontramos comprovação). Para simplificar, incluímos somente educação e gênero no modelo. Que tipo de modelo leva em conta retornos diferentes em educação? Considere o modelo

$$\log(\text{salárioh}) = (\beta_0 + \delta_0 \text{feminino}) + (\beta_1 + \delta_1 \text{feminino})\text{educ} + u. \quad (7.16)$$

Se fizermos $\text{feminino} = 0$ em (7.16), veremos que o intercepto de homens é β_0 , enquanto a inclinação na educação dos homens é β_1 . Para as mulheres, usamos $\text{feminino} = 1$; assim, o intercepto para as mulheres será $\beta_0 + \delta_0$ e a inclinação será $\beta_1 + \delta_1$. Portanto, δ_0 mede a diferença nos interceptos entre mulheres e homens, enquanto δ_1 mede a diferença no retorno da educação entre mulheres e homens. Dois dos quatro casos dos sinais de δ_0 e δ_1 são apresentados na Figura 7.2.

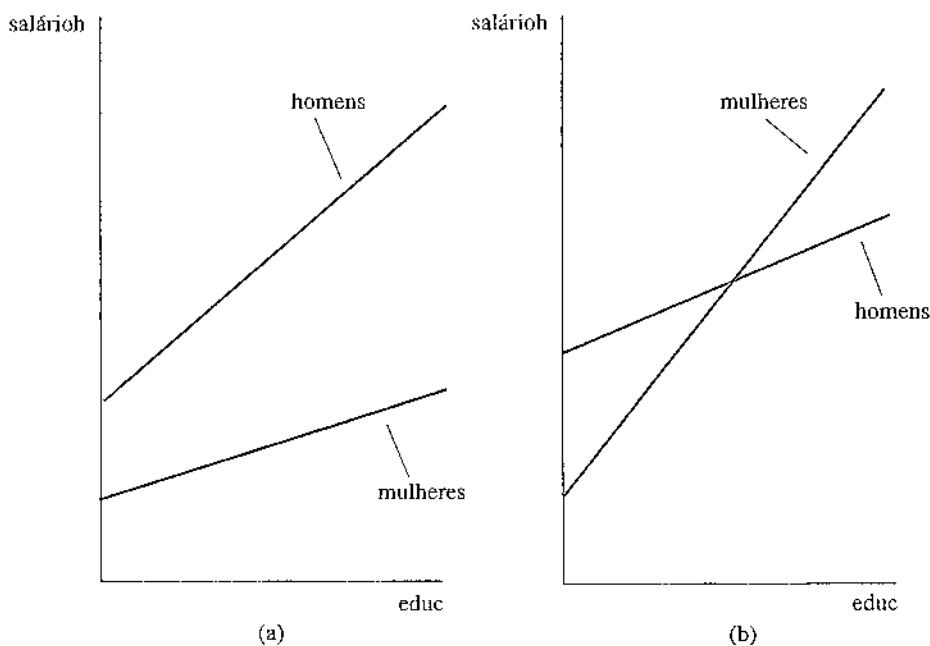
O gráfico (a) mostra o caso em que o intercepto das mulheres está abaixo do intercepto dos homens, enquanto a inclinação da linha é menor para as mulheres do que para os homens. Isso significa que as mulheres ganham menos que os homens em todos os níveis de educação, e a diferença aumenta conforme educ se torna maior. No gráfico (b), o intercepto das mulheres está abaixo do intercepto dos homens, mas a inclinação da educação é maior para as mulheres. Isso significa que as mulheres ganham menos que os homens em baixos níveis de educação, mas a diferença diminui conforme a educação aumenta. Em algum ponto, uma mulher ganhará mais que um homem, dado o mesmo nível de educação (e esse ponto é facilmente encontrado, dada a equação estimada).

Como podemos estimar o modelo (7.16)? Para aplicar o MQO, devemos escrever o modelo com uma interação entre feminino e educ :

$$\log(\text{salárioh}) = \beta_0 + \delta_0 \text{feminino} + \beta_1 \text{educ} + \delta_1 \text{feminino} \cdot \text{educ} + u. \quad (7.17)$$

Figura 7.2

Gráficos da equação (7.16). (a) $\delta_0 < 0, \delta_1 < 0$; (b) $\delta_0 < 0, \delta_1 > 0$.



Os parâmetros agora podem ser estimados a partir da regressão de $\log(\text{salárioh})$ sobre *feminino*, *educ* e *feminino-educ*. A obtenção do termo de interação é fácil com o uso de qualquer programa de regressão. Não se assuste com a natureza estranha de *feminino-educ*, que será zero para qualquer homem na amostra e igual ao nível de educação de qualquer mulher na amostra.

Uma hipótese importante é que o retorno da educação é o mesmo para mulheres e homens. Em termos do modelo (7.17), isso é declarado como $H_0: \delta_1 = 0$, o que significa que a inclinação de $\log(\text{salárioh})$ em relação a *educ* é a mesma para homens e mulheres. Observe que esta hipótese não faz nenhuma restrição sobre a diferença nos interceptos, δ_0 . Um diferencial de salários entre homens e mulheres é admitido nessa hipótese nula, mas ele deve ser o mesmo em todos os níveis de educação. Essa circunstância é descrita pela Figura 7.1.

Também estamos interessados na hipótese de que os salários médios são idênticos para homens e mulheres que tenham os mesmos níveis de educação. Isso significa que δ_0 e δ_1 devem *ambos* ser zero sob a hipótese nula. Na equação (7.17) precisamos usar um teste *F* para testar $H_0: \delta_0 = 0, \delta_1 = 0$. No modelo com apenas uma diferença de interceptos, rejeitamos essa hipótese, pois $H_0: \delta_0 = 0$ é completamente rejeitada contra $H_1: \delta_0 < 0$.

EXEMPLO 7.10**(Equação do Log dos Salários-Hora)**

Adicionemos termos quadráticos de experiência e permanência à equação (7.17):

$$\begin{aligned}
 \hat{\log(\text{salário}_h)} = & 0,389 - 0,227 \text{feminino} + 0,082 \text{educ} \\
 & (0,119) \quad (0,168) \quad (0,008) \\
 & - 0,0056 \text{feminino} \cdot \text{educ} + 0,029 \text{exper} - 0,00058 \text{exper}^2 \\
 & (0,0131) \quad (0,005) \quad (0,00011) \quad (7.18) \\
 & + 0,032 \text{perm} - 0,00059 \text{perm}^2 \\
 & (0,007) \quad (0,00024) \\
 n = & 526, R^2 = 0,441.
 \end{aligned}$$

O retorno estimado da educação dos homens nesta equação é 0,082, ou 8,2%. Para mulheres, o retorno é $0,082 - 0,0056 \approx 0,0764$, ou cerca de 7,6%. A diferença, $-0,56\%$, ou pouco mais de meio ponto percentual a menos para as mulheres, não é economicamente grande nem estatisticamente significante: a estatística t é $-0,0056/0,0131 \approx -0,43$. Assim, concluímos que não há comprovação contra a hipótese de que o retorno da educação seja o mesmo para homens e mulheres.

O coeficiente de *feminino*, embora permaneça economicamente grande, não é mais significante aos níveis convencionais ($t = -1,35$). Seu coeficiente e a estatística t na equação sem a interação eram $-0,297$ e $-8,25$, respectivamente [veja a equação (7.9)]. Devemos agora concluir que não existe evidência estatisticamente significante de salários mais baixos para mulheres nos mesmos níveis de *educ*, *exper* e *perm*? Isso seria um erro grave. Como adicionamos a interação *feminino*·*educ* à equação, o coeficiente de *feminino* é agora estimado com muito menos precisão do que na equação (7.9): o erro-padrão aumentou em quase cinco vezes ($0,168/0,036 \approx 4,67$). A razão disto é que *feminino* e *feminino*·*educ* são altamente correlacionados na amostra. Neste exemplo, existe uma maneira proveitosa de pensar sobre a multicolinearidade: na equação (7.17) e na equação mais geral estimada em (7.18), δ_0 mede o diferencial de salários entre mulheres e homens quando *educ* = 0. Como não existe ninguém na amostra com anos de educação sequer próximo de zero, não é surpreendente que temos muito trabalho para estimar o diferencial em *educ* = 0 (tampouco o diferencial em zero anos de educação é muito informativo). Mais interessante seria estimar o diferencial por gênero no, digamos, nível médio de educação da amostra (cerca de 12,5). Para fazer isso temos que substituir *feminino*·*educ* por *feminino*·(*educ* - 12,5) e computar novamente a regressão; isso muda apenas o coeficiente de *feminino* e seu erro-padrão. (Veja o Exercício 7.15.)

Se computarmos a estatística F de $H_0: \delta_0 = 0, \delta_1 = 0$, obteremos $F = 34,33$, que é um enorme valor para uma variável aleatória F com numerador $gl = 2$ e denominador $gl = 518$: o p -valor é zero até quatro casas decimais. No final, preferimos o modelo (7.9), que considera um diferencial de salários constante entre mulheres e homens.

Como você ampliaria o modelo estimado em (7.18) para possibilitar que o retorno de *perm* difere por gênero?

Como um exemplo mais complicado envolvendo interações, examinamos agora os efeitos da raça e da composição racial das cidades sobre os salários dos jogadores de beisebol da liga principal desse esporte nos Estados Unidos.

EXEMPLO 7.11

(Efeitos da Raça sobre os Salários dos Jogadores de Beisebol)

A equação seguinte é estimada para os 330 jogadores de beisebol da liga principal das cidades onde estão disponíveis estatísticas sobre a composição racial. As variáveis *negro* e *hispan* são indicadores binários dos jogadores individuais. (O grupo base é formado pelos jogadores brancos.) A variável *porcnegro* é a percentagem de negros na cidade da equipe, enquanto *porchisp* é a percentagem de hispânicos. As outras variáveis indicam aspectos da produtividade e da longevidade dos jogadores. Neste caso, estamos interessados no efeito da raça após termos controlado esses fatores.

Além de termos incluído *negro* e *hispan* na equação, adicionamos as interações *negro*·*porcnegro* e *hispan*·*porchisp*. A equação estimada é

$$\begin{aligned}
 \hat{\log(\text{salário})} = & 10,34 + 0,0673 \text{ anos} + 0,0089 \text{ jogosano} \\
 & (2,18) \quad (0,0129) \quad (0,0034) \\
 & + 0,00095 \text{ rebmed} + 0,0146 \text{ hrunano} + 0,0045 \text{ rbisyr} \\
 & (0,00151) \quad (0,0164) \quad (0,0076) \\
 & + 0,0072 \text{ runsano} + 0,0011 \text{ perccap} + 0,0075 \text{ porcest} \quad (7.19) \\
 & (0,0046) \quad (0,0021) \quad (0,0029) \\
 & - 0,198 \text{ negro} - 0,190 \text{ hispan} + 0,0125 \text{ negro} \cdot \text{porcnegro} \\
 & (0,125) \quad (0,153) \quad (0,0050) \\
 & + 0,0201 \text{ hispan} \cdot \text{porchisp}, \quad n = 330, R^2 = 0,638. \\
 & \quad (0,0098)
 \end{aligned}$$

Primeiro, devemos verificar se as quatro variáveis raciais, *negro*, *hispan*, *negro*·*porcnegro* e *hispan*·*porchisp* são conjuntamente significantes. Usando os mesmos 330 jogadores, o *R*-quadrado quando as quatro variáveis raciais são eliminadas é 0,626. Como existem quatro restrições e $gl = 330 - 13$ no modelo sem restrições, a estatística *F* está em torno de 2,63, o que produz um *p*-valor de 0,034. Assim, essas variáveis são conjuntamente significantes ao nível de 5% (embora não o sejam ao nível de 1%).

Como interpretaremos os coeficientes das variáveis raciais? Na discussão seguinte, todos os fatores de produtividade são mantidos fixos. Primeiro, considere o que acontece com jogadores negros, mantendo fixo *porchisp*. O coeficiente $-0,198$ de *negro* literalmente significa que se um jogador negro estiver em uma cidade onde não haja negros (*porcnegro* = 0), então o jogador negro ganhará cerca de 19,8% menos do que um jogador branco nas mesmas condições. Na medida em que *porcnegro* aumenta — o que significa que a população branca diminui, já que *porchisp* é mantida fixo — os salários dos negros aumentam comparado aos dos brancos. Em uma cidade com 10% de negros, $\log(\text{salário})$ dos negros comparado com o dos brancos é $-0,198 + 0,0125(10) = -0,073$, ou seja, os salários dos negros serão cerca de 7,3% menores

EXEMPLO 7.11 (continuação)

que os dos brancos em tal cidade. Quando $porcnegro = 20$, os negros ganham cerca de 5,2% mais que os brancos. A maior percentagem de negros em uma cidade está em torno de 74% (Detroit).

De forma semelhante, os hispânicos ganham menos que os brancos em cidades com um baixo percentual de hispânicos. Mas podemos facilmente encontrar o valor de $porchisp$ que torna o diferencial entre brancos e hispânicos igual a zero: ele deve ser $-0,190 + 0,0201 porchisp = 0$, o que produz $porchisp \approx 9,45$. Em cidades nas quais a percentagem de hispânicos for menor que 9,45%, é possível prever que os hispânicos ganharão menos que os brancos (para uma determinada população de negros), e o oposto é verdadeiro se o número de hispânicos estiver acima de 9,45%. Doze das vinte e duas cidades representadas na amostra possuem população hispânica menor que 6% da população total. A maior percentagem de hispânicos está em torno de 31%.

Como interpretar esses resultados? Não podemos simplesmente alegar que existe discriminação contra negros e hispânicos, porque as estimativas indicam que os brancos ganham menos que os negros e os hispânicos em cidades densamente povoadas por minorias. A importância da composição racial de uma cidade sobre os salários pode ser devida às preferências dos jogadores: talvez os melhores jogadores negros vivam em cidades com mais negros e os melhores jogadores hispânicos tendam a viver em cidades com maior concentração de hispânicos. As estimativas em (7.19) nos possibilitam determinar a presença de alguma relação, mas não temos condições de fazer a distinção entre essas duas hipóteses.

Verificação de Diferenças nas Funções de Regressão entre Grupos

Os exemplos anteriores ilustram que a interação de variáveis *dummy* com outras variáveis independentes pode ser uma ferramenta poderosa. Algumas vezes, queremos testar a hipótese nula de que duas populações, ou grupos, seguem a mesma função de regressão, contra a hipótese alternativa de que uma ou mais das inclinações diferem entre os grupos. Também veremos exemplos disso no Capítulo 13, quando discutiremos o agrupamento de diferentes cortes transversais ao longo do tempo.

Suponha que queiramos testar se o mesmo modelo de regressão descreve a nota média no curso superior de atletas universitários masculinos e femininos. A equação é

$$nmgradac = \beta_0 + \beta_1 sat + \beta_2 emperc + \beta_3 tothrs + u,$$

onde *sat* é a nota obtida no exame de ingresso em curso superior, *emperc* é o percentil da classificação no ensino médio, e *tothrs* é o total de horas do curso superior. Sabemos que para considerar uma diferença nos interceptos podemos incluir uma variável *dummy* para masculino ou feminino. Se quisermos que qualquer uma das inclinações dependa do gênero, simplesmente fazemos a interação da variável apropriada com, digamos, *feminino*, e a incluímos na equação.

Se estivermos interessados em verificar se existe *qualquer* diferença entre homens e mulheres, então devemos admitir um modelo no qual o intercepto e todas as inclinações possam ser diferentes entre os grupos:

$$\begin{aligned} nmgradac = & \beta_0 + \delta_0 feminino + \beta_1 sat + \delta_1 feminino \cdot sat + \beta_2 emperc \\ & + \delta_2 feminino \cdot emperc + \beta_3 tothrs + \delta_3 feminino \cdot tothrs + u. \end{aligned} \tag{7.20}$$

O parâmetro δ_0 é a diferença nos interceptos entre mulheres e homens, δ_1 é a diferença de inclinações em relação a *sat* entre mulheres e homens, e assim por diante. A hipótese nula de que *nmgradac* segue o mesmo modelo para homens e mulheres é escrita como

$$H_0: \delta_0 = 0, \delta_1 = 0, \delta_2 = 0, \delta_3 = 0. \quad (7.21)$$

Se um dos δ_j for diferente de zero, então os modelos são diferentes para homens e mulheres.

Utilizando os dados do arquivo GPA3.RAW, o modelo completo é estimado como

$$\begin{aligned} \hat{nmgradac} = & 1,48 - 0,353 \text{feminino} + 0,0011 \text{sat} + 0,00075 \text{feminino} \cdot \text{sat} \\ & (0,21) \quad (0,411) \quad (0,0002) \quad (0,00039) \\ & - 0,0085 \text{emperc} - 0,00055 \text{feminino} \cdot \text{emperc} + 0,0023 \text{tohrs} \\ & (0,0014) \quad (0,00316) \quad (0,0009) \quad (7.22) \\ & - 0,00012 \text{feminino} \cdot \text{tohrs} \\ & (0,00163) \end{aligned}$$

$n = 366, R^2 = 0,406, \bar{R}^2 = 0,394.$

A variável *dummy feminino* e todos os termos de interação não são muito significantes; somente a interação *feminino*·*sat* tem uma estatística *t* próxima de dois. Entretanto, sabemos que não devemos confiar nas estatísticas *t* individuais para testar uma hipótese conjunta como (7.21). Para computar a estatística *F* devemos estimar o modelo restrito, que resulta da eliminação de *feminino* e de todas as interações; isso produz um R^2 (o R^2 restrito) em torno de 0,352, de modo que a estatística *F* está em torno de 8,14; o *p*-valor é zero até cinco casas decimais, o que nos leva a rejeitar completamente (7.21). Assim, os modelos que especificam *nmgradac* de atletas masculinos e femininos são diferentes, embora cada termo em (7.22), que permitem que homens e mulheres sejam diferentes, sejam individualmente não significantes ao nível de 5%.

O grande erro-padrão da variável *feminino* e os termos de interação tornam difícil dizer com precisão como diferem homens e mulheres. Precisamos ter muito cuidado na interpretação da equação (7.22), pois, na obtenção das diferenças entre homens e mulheres, os termos de interação devem ser levados em conta. Se olharmos somente a variável *feminino*, concluiremos erroneamente que *nmgradac* é cerca de 0,353 menor para mulheres do que para homens, mantendo fixos os outros fatores. Esta é a diferença estimada somente quando *sat*, *emperc* e *tohrs* são definidas como zero, o que não é um cenário interessante. Com *sat* = 1.100, *emperc* = 10 e *tohrs* = 50, a diferença prevista entre uma mulher e um homem é $-0,353 + 0,00075(1.100) - 0,00055(10) - 0,00012(50) \approx 0,461$. Ou seja, é possível prever que a atleta feminina tem *nmgradac* quase meio ponto mais alta que um atleta masculino nas mesmas condições.

Em um modelo com três variáveis, *sat*, *emperc* e *tohrs*, é muito simples adicionar todas as interações para testar diferenças entre grupos. Em alguns casos, muito mais variáveis explicativas estão envolvidas, e portanto é conveniente termos uma maneira diferente de computar a estatística. A soma dos resíduos quadrados da estatística *F* pode ser computada facilmente mesmo quando muitas variáveis independentes estão envolvidas.

No modelo geral com k variáveis explicativas e um intercepto, suponha que temos dois grupos, que chamaremos de $g = 1$ e $g = 2$. Gostaríamos de verificar se o intercepto e todas as inclinações são os mesmos nos dois grupos. Escreva o modelo como

$$y = \beta_{g,0} + \beta_{g,1}x_1 + \beta_{g,2}x_2 + \dots + \beta_{g,k}x_k + u, \quad (7.23)$$

para $g = 1$ e $g = 2$. A hipótese de que cada beta em (7.23) é o mesmo nos dois grupos envolve $k + 1$ restrições (no exemplo de *nmgradac*, $k + 1 = 4$). O modelo sem restrições, que pode ser entendido como tendo uma variável *dummy* de grupo e k termos de interação, além do intercepto e das próprias variáveis, tem $n - 2(k + 1)$ graus de liberdade. [No exemplo da *nmgradac*, $n - 2(k + 1) = 366 - 2(4) = 358$.] Até aqui, não há nenhuma novidade. A percepção básica é que a soma dos resíduos quadrados do modelo sem restrições pode ser obtida de duas regressões *separadas*, uma para cada grupo. Seja SQR_1 a soma dos resíduos quadrados obtida ao estimar (7.23) para o primeiro grupo; isso envolve n_1 observações. Seja SQR_2 a soma dos resíduos quadrados obtida ao estimar o modelo usando o segundo grupo (n_2 observações). No exemplo anterior, se o grupo 1 for de mulheres, $n_1 = 90$ e $n_2 = 276$. Agora, a soma dos resíduos quadrados do modelo sem restrições é simplesmente $SQR_{\text{pr}} = SQR_1 + SQR_2$. A soma dos resíduos quadrados restrita é somente a SQR do agrupamento dos grupos e da estimativa de uma única equação, digamos SQR_p . Uma vez calculados esses termos, computarmos a estatística F da forma habitual:

$$F = \frac{[SQR_p - (SQR_1 + SQR_2)]}{SQR_1 + SQR_2} \cdot \frac{[n - 2(k + 1)]}{k + 1} \quad (7.24)$$

onde n é o número *total* de observações. Esta estatística F específica é usualmente chamada em econometria de **estatística de Chow**. Como o teste de Chow é apenas um teste F , ele só é válido sob homoscedasticidade. Em particular, sob a hipótese nula, as variâncias dos erros dos dois grupos devem ser iguais. Como sempre, a normalidade não é necessária para a análise assimptótica.

Para aplicarmos a estatística de Chow no exemplo de *nmgradac*, precisamos da SQR da regressão que reuniu os grupos: ela é $SQR_{\text{pr}} = 85,515$. A SQR das 90 mulheres na amostra é $SQR_1 = 19,603$ e a SQR dos homens é $SQR_2 = 58,752$. Portanto, $SQR_{\text{pr}} = 19,603 + 58,752 = 78,355$. A estatística F é $[(85,515 - 78,355)/78,355](358/4) \approx 8,18$; naturalmente, sujeito ao erro de arredondamento, isso é o que obtemos usando a forma R -quadrado do teste nos modelos com e sem os termos de interação. (Uma advertência: não existe uma forma R -quadrado simples do teste se regressões separadas forem estimadas para cada grupo; a forma R -quadrado do teste poderá ser usada somente se tiverem sido incluídas interações para criar o modelo sem restrições.)

Uma limitação importante do teste de Chow, independentemente do método usado para implementá-lo, é a hipótese nula não permitir nenhuma diferença entre os grupos. Em muitos casos, é mais interessante considerar uma diferença nos interceptos entre os grupos e depois verificar as diferenças das inclinações; viemos uma ilustração disso na equação salarial no Exemplo 7.10. Há duas maneiras de fazermos com que os interceptos difiram sob a hipótese nula. Uma delas é incluir a *dummy* do grupo e todos os termos de interação, como na equação (7.22), mas apenas testar a significância conjunta dos termos de interação. A segunda é calcular uma estatística F como na equação (7.24), mas onde a soma dos quadrados restrita, chamada “ SQR_p ” na equação (7.24), é obtida pela regressão que permite somente um deslocamento do intercepto. Em outras palavras, computarmos uma regressão agrupada e apenas incluímos as variáveis *dummy* que distinguem os dois grupos. No exemplo da nota média do

curso superior, fazemos a regressão de *nmgradac* sobre *feminino*, *sat*, *emperc* e *tothrs*, usando os dados dos alunos-atletas femininos e masculinos. No exemplo de *nmgradac*, usamos o primeiro método, e assim a hipótese nula é $H_0: \delta_1 = 0, \delta_2 = 0, \delta_3 = 0$ na equação (7.20). (δ_0 não é restrita sob a hipótese nula.) A estatística F para essas três restrições está em torno de 1,53, o que produz um p -valor igual a 0,205. Portanto, não rejeitamos a hipótese nula.

A impossibilidade de rejeitar a hipótese de que os parâmetros que multiplicam os termos de interação são todos zero sugere que o melhor modelo permite somente uma diferença de interceptos:

$$\begin{aligned}
 nm\hat{g}radac &= 1,39 + 0,310 \text{feminino} + 0,0012 \text{sat} - 0,0084 \text{emperc} \\
 (0,18) &\quad (0,059) \quad (0,0002) \quad (0,0012) \\
 &\quad + 0,0025 \text{tothrs} \\
 &\quad (0,0007) \\
 n &= 366, R^2 = 0,398, \bar{R}^2 = 0,392.
 \end{aligned} \tag{7.25}$$

Os coeficientes das inclinações em (7.25) estão próximos daqueles do grupo base (homens) em (7.22); a eliminação das interações altera pouca coisa. Porém, *feminino* em (7.25) é altamente significativa: sua estatística t está acima de cinco, sugerindo que, em determinados níveis de *sat*, *emperc* e *tothrs*, uma atleta mulher tem uma *nmgradac* prevista que é 0,31 pontos mais alta que a de um atleta homem. Essa é, de fato, uma diferença importante.

7.5 UMA VARIÁVEL DEPENDENTE BINÁRIA: O MODELO DE PROBABILIDADE LINEAR

Até agora, aprendemos bastante sobre as propriedades e a aplicabilidade do modelo de regressão linear múltipla. Nas últimas seções, estudamos como podemos incorporar informações qualitativas, por exemplo, variáveis explicativas em um modelo de regressão múltipla, por meio do uso de variáveis independentes binárias. Em todos os modelos vistos até agora, a variável dependente y teve um significado *quantitativo* (por exemplo, y é um montante em dólares, uma pontuação em um teste, uma percentagem, ou seus logs). O que acontece se quisermos usar regressão múltipla para *explicar* um evento qualitativo?

No caso mais simples o evento que gostaríamos de explicar, e que aparece com muita freqüência na prática, é um resultado binário. Em outras palavras, nossa variável dependente, y , assume somente um dos dois valores: zero ou um. Por exemplo, y pode ser definido para indicar se um adulto conchuiu o ensino médio; y pode indicar se um aluno do curso superior usou drogas ilegais durante determinado ano escolar; ou y pode indicar se uma empresa foi absorvida por outra durante determinado ano. Em cada um desses exemplos, podemos definir que $y = 1$ represente um dos resultados e $y = 0$, o outro.

Isso significaria escrever um modelo de regressão múltipla, tal como

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u, \tag{7.26}$$

quando y for uma variável binária? Como y pode assumir somente dois valores, β_j não pode ser interpretado como a mudança em y devido ao aumento de uma unidade em x_j , mantendo fixos todos os

outros fatores: y somente muda de zero para um ou de um para zero. No entanto, os coeficientes β_j ainda têm interpretações úteis. Se assumirmos que a hipótese de média condicional zero RLM.3 é válida, isto é, $E(u|x_1, \dots, x_k) = 0$, então teremos, como sempre,

$$E(y|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k,$$

onde x é uma forma abreviada que representa todas as variáveis explicativas.

O ponto principal é que, quando y é uma variável binária assumindo os valores zero e um, é sempre verdade que $P(y = 1|x) = E(y|x)$: a probabilidade de “sucesso” — isto é, a probabilidade de que $y = 1$ — é a mesma do valor esperado de y . Assim, temos a importante equação

$$P(y = 1|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (7.27)$$

que mostra a probabilidade de sucesso, digamos, $p(x) = P(y = 1|x)$, uma função linear de x_j . A equação (7.27) é um exemplo de modelo de resposta binária, e $P(y = 1|x)$ também é chamado de **probabilidade de resposta**. (Trataremos de outros modelos de resposta binária no Capítulo 17.) Como a soma das probabilidades deve ser um, $P(y = 0|x) = 1 - P(y = 1|x)$ também é uma função linear de x_j .

O modelo de regressão linear múltipla com uma variável dependente binária é chamado de **modelo de probabilidade linear (MPL)** porque a probabilidade de resposta é linear nos parâmetros β_j . No MPL, β_j mede a mudança na probabilidade de sucesso quando x_j muda, mantendo fixos os outros fatores:

$$\Delta P(y = 1|x) = \beta_j \Delta x_j. \quad (7.28)$$

Com isso em mente, o modelo de regressão múltipla pode nos permitir estimar o efeito de diversas variáveis explicativas sobre eventos qualitativos. A mecânica do MQO é a mesma de antes.

Se escrevermos a equação estimada como

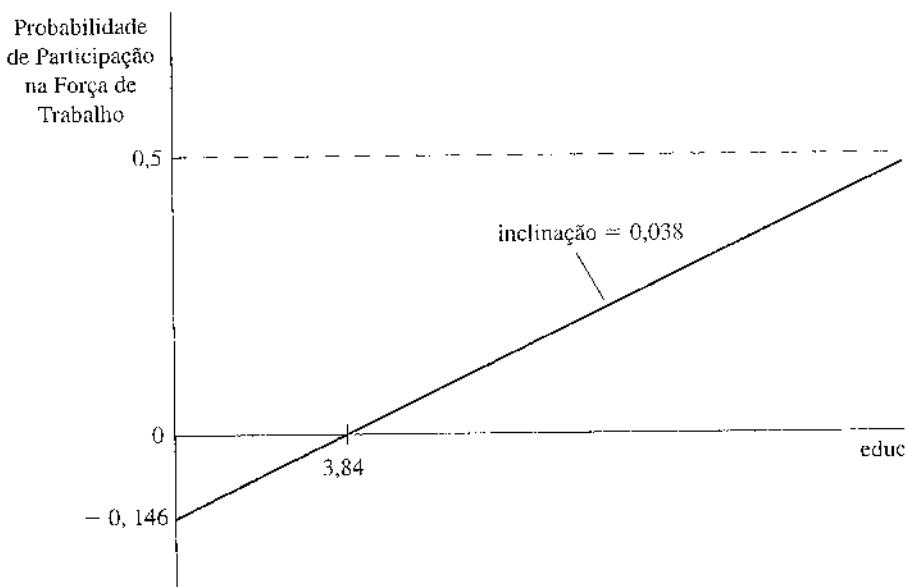
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k,$$

temos que nos lembrar que \hat{y} é a probabilidade de sucesso prevista. Portanto, $\hat{\beta}_0$ é a probabilidade de sucesso prevista quando cada x_j é definido como zero, o que pode, ou não, ser interessante. O coeficiente de inclinação $\hat{\beta}_1$ mede a mudança prevista na probabilidade de sucesso quando x_1 aumenta em uma unidade.

Para interpretarmos corretamente um modelo de probabilidade linear, precisamos saber o que constitui um “sucesso”. Assim, é uma boa idéia dar à variável dependente um nome que descreva o evento $y = 1$. Como exemplo, definimos *naft* (“na força de trabalho”) como uma variável binária indicando a participação na força de trabalho de uma mulher casada, durante 1975: *naft* = 1 se a mulher informar ter trabalhado com remuneração fora de casa em algum período do ano, e zero, caso contrário. Assumimos que a participação na força de trabalho depende de outras fontes de renda, inclusive a renda do marido (*nespend*, expressa em milhares de dólares), anos de estudo (*educ*), experiência anterior no mercado de trabalho (*exper*), *idade*, número de filhos menores de seis anos (*crianmed6*) e número de filhos entre 6 e 18 anos (*crianma6*). Utilizando os dados de Mroz (1987), estimamos o seguinte modelo de probabilidade linear, no qual 428 das 753 mulheres da amostra informam terem estado na força de trabalho em algum período do ano de 1975:

Figura 7.3

Relação estimada entre a probabilidade de estar na força de trabalho e anos de educação, com outras variáveis explicativas fixas.



$$\begin{aligned}
 \hat{naft} &= 0,586 - 0,0034 \text{ nesprend} + 0,038 \text{ educ} + 0,039 \text{ exper} \\
 &\quad (0,154) \quad (0,0014) \quad (0,007) \quad (0,006) \\
 &\quad - 0,00060 \text{ exper}^2 - 0,016 \text{ idade} - 0,262 \text{ crianmed6} + 0,0130 \text{ crianma6} \\
 &\quad (0,00018) \quad (0,002) \quad (0,034) \quad (0,0132) \\
 &n = 753, R^2 = 0,264.
 \end{aligned} \tag{7.29}$$

Usando as estatísticas *t* habituais, todas as variáveis em (7.29) exceto *crianma6* são estatisticamente significantes, e todas as variáveis significantes têm os efeitos que esperaríamos baseados na teoria econômica (ou no bom senso).

Para interpretar as estimativas, devemos nos lembrar que uma alteração na variável independente muda a probabilidade de que *naft* = 1. Por exemplo, o coeficiente de *educ* significa que, tudo o mais em (7.29) mantido fixo, mais um ano de educação, aumenta a probabilidade de participação na força de trabalho em 0,038. Se interpretarmos essa equação literalmente, mais dez anos de educação aumentariam a probabilidade de estar na força de trabalho em $0,038(10) = 0,38$, o que é um aumento bastante grande em uma probabilidade. A relação entre a probabilidade de participação na força de trabalho e *educ* está traçada na Figura 7.3. As outras variáveis independentes são fixadas nos valores *nesprend* = 50, *exper* = 5, *idade* = 30, *crianmed6* = 1 e *crianma6* = 0, para fins ilustrativos. A probabilidade prevista é negativa até que o nível de educação iguale 3,84 anos. Isso não deve causar muita preocupação

porque, na amostra, nenhuma mulher tem menos de cinco anos de estudo. O nível de educação mais alto informado é de 17 anos, e isso leva a uma probabilidade prevista de 0,5. Se definíssemos as outras variáveis independentes com diferentes valores, a gama de probabilidades previstas se alteraria. Contudo, o efeito marginal de mais um ano de educação na probabilidade de participação na força de trabalho será sempre 0,038.

O coeficiente de *nesprend* sugere que, se $\Delta_{nesprend} = 10$ (o que significa um aumento de 10.000 dólares), a probabilidade de que uma mulher esteja na força de trabalho diminui em 0,034.

Esse não é um efeito especialmente grande, considerando que um aumento na renda de 10.000 dólares é bastante significativo em termos de dólares de 1975. A experiência foi incluída como um termo quadrático para possibilitar que o efeito da experiência anterior tenha um efeito decrescente na probabilidade de participação na força de trabalho. Mantendo fixos outros fatores, a mudança estimada na probabilidade será aproximadamente $0,039 - 2(0,0006)exper = 0,039 - 0,0012exper$. O ponto no qual a experiência anterior não tem efeito sobre a probabilidade de participação na força de trabalho é $0,039/0,0012 = 32,5$, que é um alto nível de experiência: somente 13 das 753 mulheres na amostra têm mais de 32 anos de experiência.

Ao contrário do número de filhos mais velhos, o número de filhos mais novos tem um enorme impacto na participação na força de trabalho. Ter mais um filho com menos de seis anos de idade reduz a probabilidade de participação na força de trabalho em $-0,262$, nos níveis dados das outras variáveis. Na amostra, pouco menos de 20% das mulheres têm pelo menos um filho nessa faixa de idade.

Este exemplo ilustra o quanto é fácil estimar e interpretar os modelos de probabilidade linear, mas também destaca algumas de suas deficiências. Primeiro, é fácil verificar que, se agregarmos certas combinações de valores das variáveis independentes em (7.29), podemos obter previsões menores que zero ou maiores que um. Como estamos falando de probabilidades previstas, e probabilidades devendo estar entre zero e um, isso pode ser um pouco complicado. Por exemplo, qual seria o sentido de prever que uma mulher está na força de trabalho com uma probabilidade de $-0,10$? Aliás, das 753 mulheres na amostra, 16 dos valores estimados usando (7.29) são menores que zero, e 17 dos valores estimados são maiores que um.

Um problema relacionado é que a probabilidade não pode ser linearmente relacionada com as variáveis independentes em todos os seus possíveis valores. Por exemplo, a equação (7.29) prevê que o efeito de passar de zero filho para um filho menor de seis anos reduz a probabilidade de trabalhar em 0,262. Essa também é a redução se a mulher passar de um filho para dois. Pareceria mais realista que o primeiro filho reduzisse a probabilidade em grande escala, enquanto os filhos subsequentes tivessem um efeito marginal menor. De fato, quando levada ao extremo, a equação (7.29) sugere que passar de zero para quatro filhos reduz a probabilidade de trabalhar em $\Delta_{n\hat{a}ft} = 0,262(\Delta_{crianmed6}) = 0,262(4) = 1,048$, o que é impossível.

Mesmo com esses problemas, o modelo de probabilidade linear é útil e freqüentemente aplicado em economia. Normalmente, ele funciona bem com os valores das variáveis independentes que estejam próximos das médias na amostra. No exemplo da participação na força de trabalho, não existe nenhuma mulher na amostra que tenha quatro filhos menores de seis anos; aliás, somente três mulheres têm três filhos pequenos. Mais de 96% das mulheres não têm filhos ou têm apenas um, e assim provavelmente deveríamos restringir nossa atenção neste caso, quando interpretarmos a equação estimada.

Probabilidades previstas fora do intervalo da unidade são um pouco problemáticas quando queremos fazer previsões, mas muito raramente esse é o ponto central de uma análise. Normalmente, queremos saber o efeito *ceteris paribus* de certas variáveis sobre a probabilidade.

Devido à natureza binária de y , o modelo de probabilidade linear infringe uma das hipóteses de Gauss-Markov. Quando y é uma variável binária, sua variância condicional em x , é

$$\text{Var}(y|x) = p(x)[1 - p(x)], \quad (7.30)$$

onde $p(x)$ é uma forma abreviada da probabilidade de sucesso: $p(x) = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k$.

Isso significa que, com exceção do caso em que a probabilidade não depende de qualquer das variáveis independentes, deve haver heteroscedasticidade no modelo de probabilidade linear. Sabemos, do Capítulo 3, que isso não causa viés nos estimadores MQO de β_j . Entretanto, sabemos também, dos Capítulos 4 e 5, que a homoscedasticidade é crucial para justificar as estatísticas t e F habituais, mesmo em amostras grandes. Como os erros-padrão em (7.29) não são, de forma geral, válidos, devemos usá-los com cuidado. Mostraremos como corrigir os erros-padrão quanto à heteroscedasticidade no Capítulo 8. Em muitas aplicações, as estatísticas MQO habituais não ficam muito distorcidas, e ainda é aceitável no trabalho aplicado apresentar uma análise MQO padrão de um modelo de probabilidade linear.

EXEMPLO 7.12

(Um Modelo de Probabilidade Linear de Prisões)

Façamos $pris86$ ser uma variável binária igual à unidade se um homem foi preso durante o ano de 1986, e zero caso contrário. A população é um grupo de homens na Califórnia nascidos em 1960 ou 1961 que tenham sido presos pelo menos uma vez antes de 1986. Um modelo de probabilidade linear para descrever $pris86$ é

$$pris86 = \beta_0 + \beta_1 pcond + \beta_2 sentmed + \beta_3 temptot + \beta_4 ptemp86 + \beta_5 empr86 + u,$$

onde $pcond$ é a proporção de condenações anteriores, $sentmed$ é a duração média da sentença por condenações anteriores (em meses), $temptot$ é o número de meses passado na prisão desde os 18 anos de idade antes de 1986, $ptemp86$ é o número de meses passados na prisão durante 1986, e $empr86$ é o número de trimestres (0 a 4) que o homem esteve legalmente empregado em 1986.

Os dados que usamos estão no arquivo CRIME1.RAW, o mesmo conjunto de dados usado para o Exemplo 3.5. Neste caso, usamos uma variável dependente binária, porque somente 7,2% dos homens da amostra foram presos mais de uma vez. Cerca de 27,7% foram presos pelo menos uma vez durante 1986. A equação estimada é

$$\begin{aligned} pris86 &= 0,441 - 0,162 pcond + 0,0061 sentmed - 0,0023 temptot \\ &\quad (0,017) \quad (0,021) \quad (0,0065) \quad (0,0050) \\ &\quad - 0,022 ptemp86 - 0,043 empr86 \\ &\quad (0,005) \quad (0,005) \\ &n = 2.725, R^2 = 0,0474. \end{aligned} \quad (7.31)$$

O intercepto 0,441 é a probabilidade esperada de prisão de alguém que não tenha sido condenado (e portanto $pcond$ e $sentmed$ são ambas zero), não tenha passado tempo na prisão desde a idade de 18 anos, não tenha passado tempo na prisão em 1986, e esteve desempregado durante o ano inteiro. As variáveis $sentmed$ e $temptot$ não são significantes tanto individual como conjuntamente (o teste F produz p -valor = 0,347), e $sentmed$ terá um sinal antiintuitivo se sentenças mais longas supostamente desencorajarem a criminalidade. Grogger (1991), utilizando um conjunto ampliado desses dados e métodos econométricos

EXEMPLO 7.12 (continuação)

diferentes, verificou que *temptot* tem um efeito positivo estatisticamente significante sobre as prisões e concluiu que *temptot* é um indicador do capital humano construído em torno de atividades criminais.

O aumento da probabilidade de condenação não reduz a probabilidade de prisões, mas precisamos ser cuidadosos quando interpretarmos a magnitude do coeficiente. A variável *pcond* é uma proporção entre zero e um; assim, a alteração de *pcond* de zero para um significará mudar de não haver possibilidade de ser condenado para ser condenado com certeza. Mesmo essa grande mudança reduz a probabilidade de ser preso em somente 0,162; aumentando *pcond* em 0,5 decresce a probabilidade de ser preso em 0,081.

O efeito das prisões é dado pelo coeficiente de *ptemp86*. Se um homem estiver na cadeia ele não pode ser preso. Como *ptemp86* é medido em meses, seis meses a mais na prisão reduzem a probabilidade de ser preso em $0,022(6) = 0,132$. A equação (7.31) fornece outro exemplo no qual o modelo de probabilidade linear não pode ser verdadeiro em toda a amplitude das variáveis independentes. Se um homem esteve preso durante todos os doze meses do ano de 1986, ele não pôde ser preso nesse ano. Definindo todas as outras variáveis iguais a zero, a probabilidade prevista de prisão quando *temp86* = 12 será $0,441 - 0,022(12) = 0,177$, que é diferente de zero. Mesmo assim, se partirmos da probabilidade incondicional de prisão, 0,277, doze meses na prisão reduzem a probabilidade a praticamente zero: $0,277 - 0,022(12) = 0,13$.

Finalmente, o emprego reduz a probabilidade de ser preso de maneira significante. Com todos os outros fatores mantidos fixos, um homem empregado em todos os quatro trimestres tem 0,172 menos probabilidade de ser preso do que outro desempregado.

Também podemos incluir variáveis *dummy* independentes em modelos com variável *dummy* dependente. O coeficiente mede a diferença prevista na probabilidade quando a variável *dummy* vai de zero a um. Por exemplo, se adicionarmos duas *dummies* de raça, *negro* e *hispan*, à equação sobre prisões, obteremos

$$\begin{aligned}
 \text{pris86} &= 0,380 - 0,152 \text{ } pcond + 0,0046 \text{ } sentmed - 0,0026 \text{ } temptot \\
 &\quad (0,019) \quad (0,021) \quad (0,0064) \quad (0,0049) \\
 &- 0,024 \text{ } ptemp86 - 0,038 \text{ } empr86 + 0,170 \text{ } negro + 0,096 \text{ } hispan \\
 &\quad (0,005) \quad (0,005) \quad (0,024) \quad (0,021) \quad (7.32) \\
 &n = 2.725, R^2 = 0,0682.
 \end{aligned}$$

O coeficiente de *negro* significa que, todos os outros fatores iguais, um homem negro tem uma probabilidade 0,17 maior de ser preso que um homem branco (o grupo base). Outra maneira de expressar isso é dizer que a probabilidade de prisão é 17 pontos percentuais mais elevada para os negros do que para os brancos. A diferença também é estatisticamente significante. De forma semelhante, homens hispânicos têm uma probabilidade 0,096 maior de ser presos que um homem branco.

Qual é a probabilidade de prisão prevista para um homem negro que nunca tenha sido condenado — portanto *pcond*, *sentmed*, *temptot* e *ptemp86* são todas zero — e que esteve empregado em todos os quatro trimestres de 1986? Isso parece razoável?

7.6 UM POUCO MAIS SOBRE ANÁLISE E AVALIAÇÃO DE POLÍTICAS E PROGRAMAS GOVERNAMENTAIS

Vimos alguns exemplos de modelos contendo variáveis *dummy* que podem ser úteis para a avaliação das políticas de governo. O Exemplo 7.3 deu um exemplo de avaliação de programa governamental, na qual algumas empresas receberam subsídios para treinamento de pessoal e outras não.

Como mencionamos anteriormente, precisamos ser cuidadosos ao avaliarmos programas de governo, pois na maioria dos exemplos em ciências sociais, os grupos de controle e de tratamento não são determinados aleatoriamente. Considere novamente o estudo de Holzer et al. (1993), de acordo com o qual estamos agora interessados no efeito dos subsídios para treinamento de pessoal sobre a produtividade dos trabalhadores (em contraposição à quantidade de treinamento). A equação de interesse é

$$\log(rejei) = \beta_0 + \beta_1 subs + \beta_2 \log(vendas) + \beta_3 \log(empreg) + u,$$

onde *rejei* é a taxa de rejeição dos produtos da empresa, e as últimas duas variáveis são incluídas como controles. A variável binária *subs* indica se a empresa recebeu subsídios para treinamento de pessoal em 1988.

Antes de examinarmos as estimativas, devemos nos preocupar em saber se os fatores não observados que afetam a produtividade dos trabalhadores — tais como os níveis médios de educação, aptidão, experiência e permanência no emprego — podem estar correlacionados com o fato de a empresa ter recebido subsídios. Holzer et al. destacam que os subsídios foram concedidos na ordem de chegada das solicitações. Isso não é a mesma coisa que distribuir os subsídios de forma aleatória. Pode ser que empresas com trabalhadores menos produtivos tenham visto uma oportunidade de melhorar a produtividade e, portanto, tenham sido mais diligentes na solicitação de subsídios.

Utilizando os dados contidos no arquivo JTRAIN.RAW de 1988 — quando as empresas foram efetivamente qualificadas a receber subsídios — obtemos

$$\begin{aligned} \log(\hat{rejei}) &= 4,99 - 0,052 subs - 0,455 \log(vendas) \\ &\quad (4,66) \quad (0,431) \quad (0,373) \\ &\quad + 0,639 \log(empreg) \\ &\quad (0,365) \\ n &= 50, R^2 = 0,072. \end{aligned} \tag{7.33}$$

(Dezesete das cinqüenta empresas receberam subsídios para treinamento, e o índice médio de rejeição entre todas as empresas é de 3,47.) A estimativa de $-0,052$ de *subs* significa que, para *vendas* e *empreg* dados, as empresas que receberam subsídios têm índice de rejeição cerca de 5,2% menor que o daquelas que não receberam. Esta é a tendência do efeito esperado se os subsídios ao treinamento forem eficientes, mas a estatística *t* é muito pequena. Assim, desta análise de corte transversal, devemos concluir que os subsídios não tiveram efeito sobre a produtividade das empresas. Retornaremos a este exemplo no Capítulo 9 e mostraremos como a adição de informações de um ano anterior conduz a conclusões bastante diferentes.

Mesmo nos casos em que a análise das políticas de governo não envolve a atribuição de unidades para um grupo de controle e para um grupo de tratamento, devemos ser cuidadosos ao incluirmos fatores que possam estar sistematicamente relacionados com as variáveis independentes binárias de

interesse. Um bom exemplo disso é fazer um teste sobre discriminação racial. Raça não é algo determinado por um indivíduo ou por administradores governamentais. De fato, raça parece ser um exemplo perfeito de uma variável explicativa exógena, já que é determinada quando do nascimento da pessoa. Porém, por razões históricas, raça não é necessariamente exógena: existem diferenças culturais sistemáticas entre as raças, e essas diferenças podem ser importantes em um teste sobre a discriminação corrente.

Como um exemplo, considere a possibilidade de fazermos um teste para verificar a discriminação na aprovação de empréstimos. Se pudermos colher informações sobre, digamos, solicitações de empréstimos hipotecários individuais, poderemos definir a variável dummy dependente *aprovado* como igual a um se uma solicitação foi aprovada, e zero, caso contrário. Uma diferença sistemática nas taxas de aprovação entre as raças será uma indicação de discriminação. Porém, como a aprovação depende de muitos outros fatores, inclusive renda, riqueza, classificação do risco de crédito e capacidade de pagamento do empréstimo, devemos controlar esses fatores se neles houver diferenças sistemáticas entre as raças. Um modelo de probabilidade linear para testar a discriminação pode parecer com o seguinte:

$$\text{aprovado} = \beta_0 + \beta_1 \text{não-branco} + \beta_2 \text{renda} + \beta_3 \text{riqueza} + \beta_4 \text{risicodecrédito} + \text{outros fatores},$$

A discriminação contra as minorias é indicada por uma rejeição de $H_0: \beta_1 = 0$ em favor de $H_1: \beta_1 < 0$, porque β_1 é o montante pelo qual a probabilidade de uma pessoa não-branca obter uma aprovação difere da probabilidade de uma pessoa branca obtê-la, dados os mesmos níveis das outras variáveis na equação. Se *renda*, *riqueza* etc. forem sistematicamente diferentes entre as raças, então será importante controlar esses fatores em uma análise de regressão múltipla.

Outro problema que freqüentemente surge em avaliações de políticas e programas de governo é que os indivíduos (ou empresas ou cidades) escolhem participar ou não de certos procedimentos ou programas. Por exemplo, as pessoas escolhem fazer uso de drogas ilegais ou de bebidas alcoólicas. Se quisermos examinar os efeitos de tais comportamentos sobre a situação de desemprego, a renda, ou a atividade criminal, deveremos nos preocupar com o fato de que o uso de drogas pode estar correlacionado a outros fatores que possam afetar os resultados do emprego e da criminalidade. Crianças qualificadas para participar de programas de desenvolvimento da saúde infantil são incluídas nesses programas com base na decisão de seus pais. Como a formação familiar é levada em conta nas decisões desses programas e afeta os resultados dos alunos, devemos controlar esses fatores ao verificarmos seus efeitos [veja, por exemplo, Currie e Thomas (1995)]. Indivíduos selecionados por empregadores ou por agências governamentais para participarem de programas de treinamento no emprego podem participar ou não, e essa decisão provavelmente não será aleatória [veja, por exemplo, Lynch (1991)]. Cidades e estados decidem implementar certas leis de controle de armas, e muito provavelmente essa decisão estará sistematicamente relacionada a outros fatores que afetam os crimes violentos [veja, por exemplo, Kleck e Patterson (1993)].

O parágrafo anterior dá exemplos do que, de forma geral, é conhecido como problemas de *auto-seleção* em economia. Literalmente, o termo advém do fato de que os indivíduos se auto-selecionam para certos procedimentos ou programas: a participação não é determinada de forma aleatória. O termo é geralmente usado quando um indicador binário de participação puder estar sistematicamente relacionado com fatores não-observados. Assim, se escrevermos o modelo simples

$$y = \beta_0 + \beta_1 \text{partic} + u, \quad (7.34)$$

onde y é uma variável de resultado e $partic$ é uma variável binária igual à unidade se o indivíduo, empresa ou cidade participa de um procedimento, programa, ou tem certo tipo de lei, então estamos preocupados com que o valor médio de u dependa da participação: $E(u|partic = 1) \neq E(u|partic = 0)$. Como sabemos, isso levará o estimador β_1 da regressão simples a ser viesado, e portanto não encontraremos o efeito verdadeiro da participação. Assim o problema da auto-seleção é outra maneira de uma variável explicativa ($partic$, neste caso) poder ser endógena.

Por enquanto sabemos que a análise de regressão múltipla pode, até certo ponto, aliviar o problema de auto-seleção. Fatores no termo erro na equação (7.34) que sejam correlacionados com $partic$ podem ser incluídos em uma equação de regressão múltipla, assumindo, é claro, que possamos coletar os dados desses fatores. Infelizmente, em muitos casos, preocupa-nos que fatores não-observados estejam relacionados com a participação, caso em que a regressão múltipla produzirá estimadores viesados.

Na análise padrão de regressão múltipla que usa dados de corte transversal, devemos ficar atentos com o aparecimento de efeitos espúrios de programas nas variáveis de resultado devido ao problema de auto-seleção. Um bom exemplo disto está contido em Currie e Cole (1993). Esses autores examinam os efeitos da participação, em um programa específico de auxílio para famílias com dependentes menores de idade, sobre o peso dos recém-nascidos. Mesmo após controlar uma variedade de famílias e características culturais, os autores obtêm estimativas MQO que sugerem que a participação nesse programa específico *diminui* o peso dos recém-nascidos. Como os autores ressaltam, é difícil acreditar que a participação no programa, por si própria, seja a *causa* do peso menor dos recém-nascidos. [Veja Currie (1995) para exemplos adicionais.] Utilizando um método econométrico diferente, que discutiremos no Capítulo 15, Currie e Cole encontraram evidência tanto para nenhum efeito como para um efeito positivo da participação no programa de auxílio para famílias com dependentes menores de idade, sobre o peso dos recém-nascidos.

Quando o problema de auto-seleção faz com que a análise de regressão múltipla seja viesada devido à falta de suficientes variáveis de controle, métodos mais avançados, tratados nos Capítulos 13, 14 e 15 poderão ser usados.

Neste capítulo aprendemos como usar informações qualitativas na análise de regressão. No caso mais simples, uma variável *dummy* é definida para fazer a distinção entre dois grupos, e o coeficiente estimado da variável *dummy* estima a diferença *ceteris paribus* entre os dois grupos. A admissão de mais de dois grupos é realizada pela definição de um conjunto de variáveis *dummy*: se houver g grupos, então $g - 1$ variáveis *dummy* são incluídas no modelo. Todas as estimativas das variáveis *dummy* são interpretadas em relação ao grupo base ou referencial (o grupo para o qual nenhuma variável *dummy* é incluída no modelo).

As variáveis *dummy* também são úteis para incorporar informações ordinais, como classificações de crédito e de aparência pessoal, em modelos de regressão. Simplesmente definimos um conjunto de variáveis *dummy* representando os diferentes resultados da variável ordinal, admitindo uma das categorias como grupo base.

Para possibilitar diferenças de inclinações entre os diferentes grupos, as variáveis *dummy* podem interagir com variáveis quantitativas. No caso extremo, podemos permitir que cada grupo tenha sua própria inclinação em todas as variáveis, como também seu próprio intercepto. O teste de Chow pode ser usado para detectar se existem quaisquer diferenças entre os grupos. Em muitos casos, é mais interessante verificar se, após termos permitido uma diferença de interceptos, as inclinações de dois gru-

pos diferentes são as mesmas. Um teste F padrão pode ser usado para esse propósito em um modelo irrestrito que inclua interações entre a *dummy* do grupo e todas as variáveis.

O modelo de probabilidade linear, que é simplesmente estimado pelo MQO, possibilita explicar uma resposta binária usando análise de regressão. As estimativas MQO agora são interpretadas como alterações na probabilidade de "sucesso" ($y = 1$), dado um aumento de uma unidade na variável explicativa correspondente. O MPL tem algumas inconveniências: pode produzir probabilidades previstas menores que zero ou maiores que um, implica um efeito marginal constante de cada variável explicativa que aparece em sua forma original, e contém heteroscedasticidade. Os primeiros dois problemas muitas vezes não são graves quando estamos obtendo estimativas dos efeitos parciais das variáveis explicativas na faixa intermediária dos dados. A heteroscedasticidade invalida os erros-padrão usuais do MQO e as estatísticas de testes mas, como veremos no próximo capítulo, isso é facilmente corrigido em amostras suficientemente grandes.

Terminamos este capítulo com uma discussão de como variáveis binárias são usadas para a avaliação de políticas e programas de governo. Como em toda análise de regressão, devemos nos lembrar que a participação em programas, ou algum outro regressor binário com implicações relacionadas a políticas governamentais, pode ser correlacionada com fatores não observados que afetem a variável dependente, resultando no viés usual de variáveis omitidas.

7.1 Utilizando os dados contidos no arquivo SLEEP75.RAW (veja também o Problema 3.3), obtemos a equação estimada

$$\begin{aligned} \hat{dormir} &= 3.840,83 - 0,163 \text{ trabtot} - 11,71 \text{ educ} - 8,70 \text{ idade} \\ &\quad (235,11) \quad (0,018) \quad (5,86) \quad (11,21) \\ &\quad + 0,128 \text{ idade}^2 + 87,75 \text{ masculino} \\ &\quad (0,134) \quad (34,33) \\ n &= 706, R^2 = 0,123, \bar{R}^2 = 0,117. \end{aligned}$$

A variável *dormir* é o total de minutos gastos por semana dormindo durante a noite, *trabtot* é o total de minutos semanais gastos trabalhando, *educ* e *idade* são medidas em anos e *masculino* é uma variável *dummy* de gênero.

- (i) Supondo todos os outros fatores iguais, existe evidência de que os homens durmam mais que as mulheres? O quanto essa evidência é forte?
- (ii) Existe uma relação de substituição estatisticamente significante entre trabalhar e dormir? Qual é a relação de substituição estimada?
- (iii) Que outras regressões você precisa executar para testar a hipótese nula de que, mantendo fixos os outros fatores, a idade não tem efeito sobre dormir?

7.2 As seguintes equações foram estimadas utilizando os dados contidos no arquivo BWGHT.RAW:

$$\begin{aligned} \log(pesonas) = & 4,66 - 0,0044 cigs + 0,0093 \log(rendfam) + 0,016 ordnas \\ (0,22) & (0,0009) \quad (0,0059) \quad (0,006) \\ & + 0,027 masculino + 0,055 branco \\ (0,010) & \quad (0,013) \\ n = & 1.388, R^2 = 0,0472 \end{aligned}$$

e

$$\begin{aligned} \log(pesonas) = & 4,65 - 0,0052 cigs + 0,0110 \log(rendfam) + 0,017 ordnas \\ (0,38) & (0,0010) \quad (0,0085) \quad (0,006) \\ & + 0,034 masculino + 0,045 branco - 0,0030 educm + 0,0032 educp \\ (0,011) & (0,015) \quad (0,0030) \quad (0,0026) \\ n = & 1.191, R^2 = 0,0493. \end{aligned}$$

As variáveis são definidas como no Exemplo 4.9, mas adicionamos uma variável *dummy* para o caso de a criança ser do sexo masculino e uma variável *dummy* indicando se a criança é classificada como branca.

- (i) Na primeira equação, interprete o coeficiente da variável *cigs*. Particularmente, qual é o efeito no peso dos recém-nascidos se a mãe fumar dez ou mais cigarros por dia?
- (ii) Quanto se espera que um recém-nascido branco pesará mais que uma criança não-branca, mantendo fixos todos os outros fatores na primeira equação? A diferença é estatisticamente significante?
- (iii) Comente sobre o efeito estimado e a significância estatística de *educm*.
- (iv) Com a informação dada, por que você não terá condições de computar a estatística *F* da significância conjunta de *educm* e *educp*? O que você teria que fazer para computar a estatística *F*?

7.3 Utilizando os dados contidos no arquivo GPA2.RAW, a seguinte equação foi estimada:

$$\begin{aligned} sat = & 1.028,10 + 19,30 tamclas - 2,19 tamclas^2 - 45,09 feminino \\ (6,29) & (3,83) \quad (0,53) \quad (4,29) \\ & - 169,81 negro + 62,31 feminino \cdot negro \\ (12,71) & \quad (18,15) \\ n = & 4.137, R^2 = 0,0858. \end{aligned}$$

A variável *sat* é a nota combinada de matemática e habilidade verbal do estudante para ingresso em curso superior, *tamclas* é o tamanho da classe do aluno no ensino médio, em centenas, *feminino* é uma variável *dummy* para gênero e *negro* é uma variável *dummy* da raça igual a um para negros e zero, caso contrário.

- (i) Existe evidência forte que *tamclas*² deveria ser incluída no modelo? Desta equação, qual é o tamanho ótimo da classe no ensino médio?
- (ii) Mantendo fixo *tamclas*, qual é a diferença estimada na nota *sat* entre mulheres não-negras e homens não-negros? O quanto é estatisticamente significante essa diferença estimada?
- (iii) Qual é a diferença estimada na *sat* entre homens não-negros e homens negros? Teste a hipótese nula de que não há diferença entre suas notas, contra a hipótese alternativa de que existe uma diferença.
- (iv) Qual é a diferença estimada na nota *sat* entre mulheres negras e mulheres não-negras? O que você necessaria fazer para verificar se a diferença é estatisticamente significante?

7.4 Uma equação que explica os salários dos diretores executivos é

$$\begin{aligned} \log(\text{salário}) = & 4,59 + 0,257 \log(\text{vendas}) + 0,011 \text{rma} + 0,158 \text{financeira} \\ & (0,30) \quad (0,032) \qquad \qquad \qquad (0,004) \quad (0,089) \\ & + 0,181 \text{prodcons} - 0,283 \text{serv} \\ & (0,085) \qquad \qquad \qquad (0,099) \\ & n = 209, R^2 = 0,357. \end{aligned}$$

Os dados usados estão no arquivo CEOSAL1.RAW, no qual *financeira*, *prodcons* e *serv* são variáveis binárias indicando as empresas financeiras, de produtos de consumo e de serviços públicos. O ramo de atividade omitido foi o de transportes.

- (i) Calcule a diferença percentual aproximada no salário estimado entre os setores de serviços públicos e de transportes, mantendo fixos *vendas* e *rma*. A diferença é estatisticamente significante ao nível de 1%?
- (ii) Use a equação (7.10) para obter a diferença percentual exata no salário estimado entre os setores de serviços públicos e de transportes e compare-a com a resposta obtida na parte (i).
- (iii) Qual é a diferença percentual aproximada no salário estimado entre as indústrias de produtos de consumo e financeiros? Escreva uma equação que possibilite verificar se a diferença é estatisticamente significante.

7.5 No Exemplo 7.2 defina *semPC* como uma variável *dummy* igual a um se o aluno não possuir um PC, e zero caso contrário.

- (i) Se *semPC* for usada no lugar de *PC* na equação (7.6), o que acontece com o intercepto na equação estimada? Qual será o coeficiente de *semPC*? (Sugestão: Escreva $PC = 1 - semPC$ e agregue isso na equação $\hat{nmgrad} = \hat{\beta}_0 + \hat{\delta}_0 PC + \hat{\beta}_1 nmem + \hat{\beta}_2 tac$.)
- (ii) O que acontecerá com o *R*-quadrado se *semPC* for usado em lugar de *PC*?
- (iii) As variáveis *PC* e *semPC* deveriam ser incluídas como variáveis independentes no modelo? Explique.

7.6 Para testar a eficiência de um programa de treinamento de pessoal sobre os subsequentes salários dos trabalhadores, especificamos o modelo

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

onde $trein$ é uma variável binária igual à unidade se um trabalhador participou do programa. Pense no termo erro u como contendo a aptidão não observada do trabalhador. Se trabalhadores menos aptos tiverem maior oportunidade de serem selecionados para o programa, e você usar uma análise MQO, o que você pode dizer sobre o provável viés no estimador MQO de β_1 ? (Sugestão: Consulte o Capítulo 3.)

7.7 No exemplo na equação (7.29), suponha que definamos $foraft$ como sendo um se a mulher estiver fora da força de trabalho, e zero caso contrário.

- (i) Se fizermos a regressão de $foraft$ sobre todas as variáveis independentes na equação (7.29), o que acontecerá com as estimativas do intercepto e da inclinação? (Sugestão: $naft = 1 - foraft$. Agregue essa expressão na equação populacional $naft = \beta_0 + \beta_1 nesprend + \beta_2 educ + \dots$ e reorganize.)
- (ii) O que acontecerá com os erros-padrão das estimativas do intercepto e da inclinação?
- (iii) O que acontecerá com o R -quadrado?

7.8 Suponha que você colete dados de uma pesquisa sobre salários, educação, experiência e gênero. Além disso, você solicita informações sobre o uso de maconha. A pergunta original é: “Em quantas ocasiões distintas, no mês passado, você fumou maconha?”.

- (i) Escreva uma equação que permita a você estimar os efeitos do uso de maconha sobre os salários com todos os outros fatores controlados. Você deve ter condições de fazer declarações do tipo, “Estima-se que fumar maconha cinco vezes ou mais por mês altera os salários em $x\%$.”
- (ii) Escreva um modelo que permita verificar se o uso de drogas tem efeitos diferentes sobre os salários dos homens e das mulheres. Como você verificaria que não existem diferenças nos efeitos do uso de drogas nos homens e nas mulheres?
- (iii) Suponha que você considere ser melhor avaliar o uso de maconha colocando as pessoas em uma de quatro categorias: não-usuário, usuário leve (um a cinco vezes por mês), usuário moderado (seis a dez vezes por mês) e usuário inveterado (mais de dez vezes por mês). Agora escreva um modelo que permita estimar os efeitos da maconha sobre os salários.
- (iv) Usando o modelo da parte (iii), explique em detalhes como testar a hipótese nula de que o uso de maconha não tem efeito sobre o salário. Seja bastante específico e inclua uma relação cuidadosa de graus de liberdade.
- (v) Quais são alguns dos problemas potenciais de procurar inferência causal utilizando os dados da pesquisa que você coletou?

Heteroscedasticidade

hipótese de homoscedasticidade, apresentada no Capítulo 3 para a regressão múltipla, significa que a variância do erro não observável, u , condicional nas variáveis explicativas, é constante. A homoscedasticidade não se mantém sempre que a variância dos fatores não-observáveis muda ao longo de diferentes segmentos da população, nos quais os segmentos são determinados pelos diferentes valores das variáveis explicativas. Por exemplo, em uma equação de poupança, a heteroscedasticidade está presente se a variância dos fatores não-observados que afetam a poupança aumenta com a renda.

Nos Capítulos 4 e 5, vimos que a homoscedasticidade é necessária para justificar os habituais testes t e F , bem como os intervalos de confiança da estimativa MQO do modelo de regressão linear, mesmo com amostras de tamanhos grandes. Neste capítulo discutiremos as soluções disponíveis quando ocorre heteroscedasticidade, e também mostraremos como verificar sua presença. Iniciamos fazendo uma breve revisão das consequências da heteroscedasticidade para a estimativa de mínimos quadrados ordinários.

8.1 CONSEQUÊNCIAS DA HETEROSCEDASTICIDADE PARA O MÉTODO MQO

Considere novamente o modelo de regressão linear múltipla:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u. \quad (8.1)$$

No Capítulo 3, provamos a inexistência de viés dos estimadores de MQO $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ sob as quatro primeiras hipóteses de Gauss-Markov, RLM.1 a RLM.4. No Capítulo 5 mostramos que as mesmas quatro hipóteses implicam consistência dos estimadores de MQO. A hipótese de homoscedasticidade RLM.5, estabelecida em termos da variância do erro como $\text{Var}(u|x_1, x_2, \dots, x_k) = \sigma^2$, não teve participação para mostrar se os estimadores MQO eram não-viesados ou consistentes. É importante lembrar que a heteroscedasticidade não provoca viés ou inconsistência nos estimadores MQO de β_j , enquanto algo como a omissão de uma variável importante teria esse efeito.

A interpretação de nossas medidas dos graus-de-ajuste, R^2 e \bar{R}^2 , também não é afetada pela presença de heteroscedasticidade. Por quê? Lembre-se de que na Seção 6.3 o R -quadrado usual e o R -quadrado ajustado são modos diferentes de estimar o R -quadrado da população, que é simplesmente $1 - \sigma_u^2/\sigma_y^2$, onde σ_u^2 é a variância do erro da população e σ_y^2 é a variância populacional de y . A questão crucial é

que, como ambas as variâncias no R -quadrado da população são variâncias incondicionais, o R -quadrado da população não é afetado pela presença de heteroscedasticidade em $\text{Var}(u|x_1, \dots, x_k)$. Além disso, SQR/n estima consistentemente σ_u^2 , e SQT/n estima consistentemente σ_y^2 , seja $\text{Var}(u|x_1, \dots, x_k)$ constante ou não. O mesmo é verdadeiro quando usamos os ajustes dos graus de liberdade. Portanto, R^2 e \bar{R}^2 são ambos estimadores consistentes do R -quadrado da população mantendo-se ou não a hipótese de homoscedasticidade.

Se a heteroscedasticidade não provoca viés ou inconsistência nos estimadores MQO, por que a introduzimos como uma das hipóteses de Gauss-Markov? Lembre-se do Capítulo 3 que os estimadores de *variâncias*, $\text{Var}(\hat{\beta}_j)$, são viesados sem a hipótese de homoscedasticidade. Como os erros-padrão dos estimadores MQO são baseados diretamente nessas variâncias, eles não mais são válidos para construirmos intervalos de confiança e estatísticas t . As estatísticas t habituais dos estimadores MQO não têm distribuições t na presença de heteroscedasticidade, e o problema não será resolvido com o uso de amostras de tamanho grande. Veremos isso claramente no caso de regressão simples na próxima seção, na qual derivaremos a variância do estimador MQO da inclinação sob heteroscedasticidade e proporemos um estimador válido na presença de heteroscedasticidade. De maneira semelhante, as estatísticas F não têm distribuição F , e a estatística LM não tem uma distribuição qui-quadrada assumptórica. Em resumo, as estatísticas que usamos para testar hipóteses sob as hipóteses de Gauss-Markov não são válidas na presença de heteroscedasticidade.

Também sabemos que o teorema de Gauss-Markov, que diz que os estimadores MQO são os melhores estimadores lineares não-viesados, vale-se de forma crucial da hipótese de homoscedasticidade. Se $\text{Var}(u|x)$ não for constante, os estimadores MQO não mais serão BLUE. Além disso, os estimadores MQO não mais serão assumptivamente eficientes na classe dos estimadores descritos no Teorema 5.3. Como veremos na Seção 8.4, é possível encontrar estimadores que são mais eficientes que os MQO na presença de heteroscedasticidade (embora seja necessário o conhecimento da forma da heteroscedasticidade). Com amostras de tamanhos relativamente grandes, pode não ser tão importante obter um estimador eficiente. Na próxima seção mostraremos como os testes estatísticos usuais dos estimadores MQO podem ser modificados de forma a serem válidos, pelo menos assumptivamente.

8.2 INFERÊNCIA ROBUSTA EM RELAÇÃO À HETEROSCEDASTICIDADE APÓS A ESTIMAÇÃO MQO

Como os testes de hipóteses são um componente importante de qualquer análise econométrica e a inferência habitual dos estimadores MQO geralmente é imperfeita na presença de heteroscedasticidade, temos que decidir se abandonaremos de vez o método MQO. Felizmente, ele ainda é útil. Nas últimas duas décadas, os econometristas aprenderam como ajustar erros-padrão, estatísticas t , F e LM de forma a torná-las válidas na presença de **heteroscedasticidade de forma desconhecida**. Isto é muito conveniente, pois significa que podemos descrever novas estatísticas que funcionam independentemente do tipo de heteroscedasticidade presente na população. Os métodos desta seção são conhecidos como procedimentos *robustos em relação à heteroscedasticidade*, porque eles são válidos — pelo menos em amostras grandes — tenham ou não os erros variância constante, e não precisamos saber qual é o caso.

Começamos esboçando como as variâncias, $\text{Var}(\hat{\beta}_j)$, podem ser estimadas na presença de heteroscedasticidade. Uma derivação cuidadosa da teoria está bem além do escopo desta obra, mas a aplicação de métodos robustos em relação à heteroscedasticidade é bastante fácil, pois muitos programas estatísticos e econométricos computam essas estatísticas como uma opção.

Primeiro, considere o modelo com uma única variável independente, na qual incluímos um subscrito i por ênfase:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Assumimos que as primeiras quatro hipóteses de Gauss-Markov se sustentam. Se os erros contiverem heteroscedasticidade, então

$$\text{Var}(u_i | x_i) = \sigma_i^2,$$

onde colocamos um subscrito i em σ^2 para indicar que a variância do erro depende do valor particular de x_i .

Escreva o estimador MQO como

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Sob as hipóteses RLM.1 a RLM.4 (isto é, sem a hipótese de homoscedasticidade), e condicionado aos valores de x_i na amostra, podemos usar os mesmos argumentos do Capítulo 2 para mostrar que

$$\text{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{\text{SQT}_x^2}, \quad \{8.2\}$$

onde $\text{SQT}_x = \sum_{i=1}^n (x_i - \bar{x})^2$ é a soma dos quadrados total de x_i . Quando $\sigma_i^2 = \sigma^2$ para todo i , essa fórmula se reduz à forma habitual, σ^2/SQT_x . A equação (8.2) mostra explicitamente que, no caso da regressão simples, a fórmula de variância derivada sob homoscedasticidade não mais é válida quando a heteroscedasticidade está presente.

Como o erro-padrão de $\hat{\beta}_1$ é baseado diretamente na estimativa de $\text{Var}(\hat{\beta}_1)$, precisamos de um modo de estimar a equação (8.2) quando a heteroscedasticidade está presente. White (1980) mostrou como isso pode ser feito. Façamos \hat{u}_i representar os resíduos MQO da regressão inicial de y sobre x . Então, um estimador válido de $\text{Var}(\hat{\beta}_1)$, para a heteroscedasticidade de qualquer forma (inclusive homoscedasticidade), é

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2 \hat{u}_i^2}{\text{SQT}_x^2}, \quad \{8.3\}$$

que é facilmente calculado após a regressão MQO.

Em que sentido a expressão (8.3) é um estimador válido de $\text{Var}(\hat{\beta}_1)$? Isso é bastante sutil. Resumidamente, pode ser mostrado que quando a equação (8.3) é multiplicada pelo tamanho da amostra, n , ela converge em probabilidade para $E[(x_i - \mu_x)^2 u_i^2]/(\sigma_x^2)^2$, que é o limite de probabilidade de n vezes (8.2). Em última análise, isso é o necessário para justificar o uso de erros-padrão para construir

intervalos de confiança e estatísticas t . A lei dos grandes números e o teorema do limite central desempenham papéis importantes no estabelecimento dessas convergências. Você pode consultar o artigo original de White para detalhes, mas ele é bastante técnico. Veja também Wooldridge (2002, Capítulo 4).

Uma fórmula semelhante funciona no modelo geral de regressão múltipla

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u,$$

Pode ser mostrado que um estimador válido de $\text{Var}(\hat{\beta}_j)$, sob as hipóteses RLM.1 a RLM.4, é

$$\text{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_i^2}{\text{SQR}_j^2}, \quad (8.4)$$

onde \hat{r}_{ij} representa o i -ésimo resíduo da regressão de x_j sobre todas as outras variáveis independentes, e SQR_j é a soma dos resíduos quadrados dessa regressão (veja Seção 3.2 para a representação parcial das estimativas MQO). A raiz quadrada de (8.4) é chamada de **erro-padrão robusto em relação à heteroscedasticidade** de $\hat{\beta}_j$. Em econometria, esses erros-padrão robustos são em geral atribuídos a White (1980). Trabalhos anteriores em estatística, notavelmente os de Eicker (1967) e Huber (1967), indicaram a possibilidade de obter tais erros-padrão robustos. Em trabalhos aplicados, eles são algumas vezes chamados de *erros-padrão de White, de Huber, ou de Eicker* (ou alguma combinação hifenizada desses nomes). Faremos referência a eles apenas como erros-padrão robustos em relação à heteroscedasticidade, ou mesmo apenas como erros-padrão robustos, quando o contexto for claro.

Algumas vezes, como uma correção de graus de liberdade, a equação (8.4) é multiplicada por $n/(n - k - 1)$ antes de extrairmos a raiz quadrada. O raciocínio para esse ajuste é que, se os resíduos quadrados MQO \hat{u}_i^2 fossem os mesmos para todas as observações i — a forma mais forte possível de homoscedasticidade em uma amostra — obteríamos os erros-padrão habituais MQO. Outras modificações de (8.4) são estudadas em MacKinnon e White (1985). Como todas as formas têm apenas justificativas assintóticas e são assintoticamente equivalentes, nenhuma delas é uniformemente preferida às outras. Em geral, utilizamos qualquer forma que seja computada pelo programa econométrico em uso.

Uma vez que os erros-padrão robustos em relação à heteroscedasticidade tenham sido obtidos, é fácil construir uma **estatística t robusta em relação à heteroscedasticidade**. Lembre-se de que a forma geral da estatística t é

$$t = \frac{\text{estimativa} - \text{valor hipotético}}{\text{erro-padrão}}. \quad (8.5)$$

Como ainda estamos usando as estimativas MQO e escolhemos o valor hipotético antecipadamente, a única diferença entre a estatística t usual de MQO e a estatística t robusta em relação à heteroscedasticidade é como o erro-padrão é calculado.

EXEMPLO 8.1**(Equação do Log dos Salários com Erros-Padrão Robustos em Relação à Heteroscedasticidade)**

Estimamos o modelo no Exemplo 7.6, mas escrevemos os erros-padrão robustos em relação à heteroscedasticidade juntamente com os erros-padrão do MQO. Algumas das estimativas estão registradas com mais dígitos para podermos comparar os erros-padrão habituais com os erros-padrão robustos em relação à heteroscedasticidade:

$$\begin{aligned}
 \log(\hat{\text{salário}}) = & 0,321 + 0,213 \text{ hcasados} - 0,198 \text{ mcasadas} - 0,110 \text{ insolteiras} \\
 & (0,100) \quad (0,055) \quad (0,058) \quad (0,056) \\
 & [0,109] \quad [0,057] \quad [0,058] \quad [0,057] \\
 & + 0,0789 \text{ educ} + 0,0268 \text{ exper} - 0,00054 \text{ exper}^2 \\
 & (0,0067) \quad (0,0055) \quad (0,00011) \\
 & [0,0074] \quad [0,0051] \quad [0,00011] \tag{8.6} \\
 & + 0,0291 \text{ perm} - 0,00053 \text{ perm}^2 \\
 & (0,0068) \quad (0,00023) \\
 & [0,0069] \quad [0,00024] \\
 n = 526, R^2 = 0,461.
 \end{aligned}$$

Os erros-padrão usuais de MQO estão entre parênteses, (), abaixo das estimativas MQO correspondentes, e os erros-padrão robustos em relação à heteroscedasticidade estão entre colchetes, []. Os números entre colchetes são as únicas novidades, já que a equação ainda é estimada por MQO.

Diversos fatores são aparentes a partir da equação (8.6). Primeiro, nesta aplicação particular, qualquer variável que era estatisticamente significante com o uso da estatística t habitual, continua estatisticamente significante com o uso da estatística t robusta em relação à heteroscedasticidade. Isto é devido ao fato de os dois conjuntos de erros-padrão não serem muito diferentes. (Os p -valores associados serão levemente diferentes devido ao fato de as estatísticas t robustas não serem idênticas às estatísticas t habituais, não robustas.) A maior mudança relativa nos erros-padrão está no coeficiente de educ : o erro-padrão usual é 0,0067, e o erro-padrão robusto é 0,0074. Não obstante, o erro-padrão robusto implica uma estatística t robusta acima de 10.

A equação (8.6) também mostra que os erros-padrão robustos podem ser maiores, ou menores, do que os erros-padrão usuais. Por exemplo, o erro-padrão robusto da variável exper é 0,0051, enquanto o erro-padrão usual é 0,0055. Não sabemos qual será maior antecipadamente. Como um tópico empírico, os erros-padrão robustos são freqüentemente maiores do que os erros-padrão usuais.

Antes de sairmos deste exemplo, devemos enfatizar que não sabemos, neste ponto, sequer se a heteroscedasticidade está presente no modelo populacional básico da equação (8.6). Tudo que fizemos foi descrever, juntamente com os erros-padrão usuais, aqueles que são válidos (assimptoticamente) haja ou não presença de heteroscedasticidade. Podemos ver que nenhuma conclusão importante é destruída pelo uso dos erros-padrão robustos neste exemplo. Isso acontece com freqüência em trabalhos aplicados, mas, em outros casos, as diferenças entre os erros-padrão usuais e os robustos são muito maiores. Para um exemplo no qual as diferenças são substanciais, veja o Problema 8.7.

Neste ponto, você deve estar se perguntando: se os erros-padrão robustos em relação à heteroscedasticidade são válidos com maior freqüência que os erros-padrão usuais MQO, por que nos preocuparmos com os erros-padrão usuais, afinal? Esta é uma pergunta sensata. Uma das razões para eles

ainda serem usados em trabalhos de corte transversal é que, se a hipótese de homoscedasticidade se mantiver e os erros forem normalmente distribuídos, as estatísticas t usuais têm distribuições t exatas, independentemente do tamanho da amostra (veja Capítulo 4). Os erros-padrão robustos e as estatísticas t robustas são justificadas somente quando o tamanho da amostra se torna grande. Com amostras de tamanho pequeno, as estatísticas t robustas podem ter distribuições que não sejam muito próximas da distribuição t , e isso pode ofuscar nossa inferência.

Em amostras de tamanho grande podemos tomar a decisão de sempre levar em conta somente os erros-padrão robustos em relação à heteroscedasticidade em aplicações de corte transversal, e esta prática vem sendo seguida cada vez mais em trabalhos aplicados. Também é comum usar ambos os erros-padrão, como na equação (8.6), de maneira que um leitor possa determinar se alguma conclusão é sensível ao erro-padrão em uso.

Também é possível obter estatísticas F e LM que sejam robustas em relação à heteroscedasticidade de forma desconhecida, e mesmo arbitrária. A **estatística F robusta em relação à heteroscedasticidade** (ou uma transformação simples dela) também é chamada de *estatística de Wald robusta em relação à heteroscedasticidade*. Uma abordagem geral da estatística de Wald requer álgebra matricial e está esboçada no Apêndice E, disponível na página do livro, no site www.thomsonlearning.com.br; veja Wooldridge (2002, Capítulo 4) para um tratamento mais detalhado. No entanto, a utilização de estatísticas robustas em relação à heteroscedasticidade para restrições de exclusões múltiplas é simples, pois muitos programas econométricos computam essas estatísticas rotineiramente.

EXEMPLO 8.2

(Estatística F Robusta em Relação à Heteroscedasticidade)

Utilizando os dados s contidos no arquivo GPA3.RAW, estimamos a seguinte equação:

$$\begin{aligned}
 nmgradac &= 1,47 + 0,00114 \text{ sat} - 0,00857 \text{ emperc} + 0,00250 \text{ tothrs} \\
 &\quad (0,23) (0,00018) (0,00124) (0,00073) \\
 &\quad [0,22] [0,00019] [0,00140] [0,00073] \\
 &\quad + 0,303 \text{ feminino} - 0,128 \text{ negro} - 0,059 \text{ branco} \\
 &\quad (0,059) (0,147) (0,141) \\
 &\quad [0,059] [0,118] [0,110] \\
 n &= 366, R^2 = 0,4006, \bar{R}^2 = 0,3905.
 \end{aligned} \tag{8.7}$$

Novamente, as diferenças entre os erros-padrão usuais e os erros-padrão robustos em relação à heteroscedasticidade não são muito grandes e o uso da estatística t robusta não altera a significância estatística de qualquer variável independente. Tampouco os testes de significância conjunta são muito afetados. Suponha que queiramos testar a hipótese nula de que, depois de termos todos os outros fatores controlados, não haja diferenças em $nmgradac$ por raça. Isso é escrito como $H_0: \beta_{negro} = 0, \beta_{branco} = 0$. A estatística F habitual é facilmente obtida, uma vez que tenhamos o R -quadrado do modelo restrito; o cálculo resulta em 0,3983. A estatística F então é $[(0,4006 - 0,3983)/(1 - 0,4006)](359/2) \approx 0,69$. Se houver presença de heteroscedasticidade, essa versão do teste não é válida. A versão robusta em relação à heteroscedasticidade não tem uma forma simples, mas pode ser computada utilizando-se certos programas estatísticos. O valor da estatística F robusta em relação à heteroscedasticidade é de 0,75, diferenciando-se apenas levemente do valor da versão não robusta. O p -valor do teste robusto é 0,474, que não está próximo dos níveis padrão de significância. Não é possível rejeitar a hipótese nula usando qualquer um dos testes.

Computando Testes *LM* Robustos em Relação à Heteroscedasticidade

Nem todos os programas econômétricos calculam estatísticas *F* que sejam robustas em relação à heteroscedasticidade. Portanto, algumas vezes é conveniente ter um meio de obter um teste de restrições de exclusões múltiplas que seja robusto em relação à heteroscedasticidade e não exija um tipo especial de programa econômétrico. Uma **estatística *LM* robusta em relação à heteroscedasticidade** é facilmente obtida usando qualquer programa econômétrico.

Avalie a seguinte declaração: Os erros-padrão robustos em relação à heteroscedasticidade são sempre maiores que os erros-padrão usuais.

Para ilustrar o cálculo da estatística *LM* robusta, considere o modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u,$$

e suponha que queiramos testar $H_0: \beta_4 = 0, \beta_5 = 0$. Para obter a estatística *LM* usual, primeiro estimamos o modelo restrito (isto é, o modelo sem x_4 e x_5) para obter os resíduos, \tilde{u} . Então fazemos a regressão de \tilde{u} sobre todas as variáveis independentes e a estatística $LM = n \cdot R_{\tilde{u}}^2$, onde $R_{\tilde{u}}^2$ é o *R*-quadrado usual desta regressão.

Obter uma versão que seja robusta em relação à heteroscedasticidade exige mais trabalho. Um método de calcular a estatística requer somente regressões MQO. Precisamos dos resíduos, digamos \tilde{r}_1 , da regressão de x_4 sobre x_1, x_2, x_3 . Também necessitamos dos resíduos, digamos \tilde{r}_2 , da regressão de x_5 sobre x_1, x_2, x_3 . Assim, fazemos a regressão de cada uma das variáveis independentes excluídas, conforme a hipótese nula, sobre todas as variáveis independentes incluídas. A cada vez, guardamos os resíduos. O último passo parece estranho, mas ele é, no final das contas, apenas um artifício de cálculo. Compute a regressão de

$$I \text{ sobre } \tilde{r}_1 \tilde{u}, \tilde{r}_2 \tilde{u}, \quad (8.8)$$

sem um intercepto. Sim, na verdade definimos uma variável dependente com valor um para todas as observações. Fazemos a regressão dessa constante sobre os produtos $\tilde{r}_1 \tilde{u}$ e $\tilde{r}_2 \tilde{u}$. A estatística *LM* robusta acaba sendo $n - SQR_1$, onde SQR_1 é exatamente a soma dos resíduos quadrados usual da regressão (8.8).

A razão pela qual isso funciona é algo técnico. Basicamente, isso significa fazer para o teste *LM* o que os erros-padrão robustos fazem para o teste *t*. [Veja Wooldridge (1991b) ou Davidson e MacKinnon (1993) para uma discussão mais detalhada.]

Agora resumimos o cálculo da estatística *LM* robusta em relação à heteroscedasticidade no caso geral.

UMA ESTATÍSTICA *LM* ROBUSTA EM RELAÇÃO À HETEROSCEDASTICIDADE:

1. Obtenha os resíduos \tilde{u} do modelo restrito.

2. Faça a regressão de cada uma das variáveis independentes excluídas, conforme a hipótese nula, sobre todas as variáveis independentes incluídas; se houver q variáveis excluídas, isso levará a q conjuntos de resíduos ($\tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_q$).
3. Encontre os produtos de cada \tilde{r}_j por \tilde{u} (para todas as observações).
4. Faça a regressão de 1 sobre $\tilde{r}_1\tilde{u}, \tilde{r}_2\tilde{u}, \dots, \tilde{r}_q\tilde{u}$, sem um intercepto. A estatística LM robusta é $n - SQR_1$, onde SQR_1 é exatamente a soma dos resíduos quadrados desta última regressão. Sob H_0 , a estatística LM é distribuída aproximadamente como χ^2_q .

Uma vez que a estatística LM robusta tenha sido obtida, a regra de rejeição e a computação dos p -valores são as mesmas da estatística LM usual da Seção 5.2.

EXEMPLO 8.3

(Estatística LM Robusta em Relação à Heteroscedasticidade)

Usamos os dados do arquivo CRIME1.RAW para verificar se a média de tempo das penas cumpridas de condenações passadas afeta o número de prisões no ano atual (1986). O modelo estimado é

$$\begin{aligned}
 npre86 = & 0,567 - 0,136 pcond + 0,0178 sentmed - 0,00052 sentmed^2 \\
 & (0,036) (0,040) (0,0097) (0,00030) \\
 & [0,040] [0,034] [0,0101] [0,00021] \\
 & - 0,0394 ptemp86 - 0,0505 empr86 - 0,00148 rend86 \\
 & (0,0087) (0,0144) (0,00034) \\
 & [0,0062] [0,0142] [0,00023] \quad (8.9) \\
 & + 0,325 negro + 0,193 hispan \\
 & (0,045) (0,040) \\
 & [0,058] [0,040] \\
 n = 2,725, R^2 = 0,0728.
 \end{aligned}$$

Neste exemplo existem diferenças mais substanciais entre alguns dos erros-padrão usuais e os erros-padrão robustos. Por exemplo, a estatística t habitual de $sentmed^2$ é aproximadamente $-1,73$, enquanto a estatística robusta está em torno de $-2,48$. Assim, $sentmed^2$ é mais significante usando o erro-padrão robusto.

O efeito de $sentmed$ sobre $npre86$ é um pouco difícil de reconciliar. Como a relação é quadrática, podemos descobrir onde $sentmed$ tem um efeito positivo em $npre86$ e onde o efeito se torna negativo. O ponto de reversão é $0,0178[2(0,00052)] \approx 17,12$; observe que esta é uma medida de meses. Literalmente, isso significa que $npre86$ é positivamente relacionado com $sentmed$ quando $sentmed$ é menor que 17 meses; então $sentmed$ tem o efeito desencorajador esperado após 17 meses.

Para verificar se a média de tempo das penas cumpridas tem um efeito estatisticamente significante sobre $npre86$, devemos testar as hipóteses $H_0: \beta_{sentmed} = 0, \beta_{sentmed^2} = 0$. Utilizando a estatística LM usual (veja Seção 5.2), obtemos $LM = 3,54$, em uma distribuição qui-quadrada com dois gl , isso produz um p -valor = 0,170. Assim, não podemos rejeitar H_0 mesmo no nível de 15%. A estatística LM robusta em relação à heteroscedasticidade é $LM = 4,00$ (arredondada para duas casas decimais), com um p -valor = 0,135. Isso ainda não é uma evidência muito forte contra H_0 ; $sentmed$ não parece ter um efeito forte sobre $npre86$. [A propósito, quando $sentmed$ aparece sozinha em (8.9), isto é, sem o termo quadrático, a estatística t usual é 0,658, e a estatística t robusta é 0,592.]

8.3 O TESTE DA EXISTÊNCIA DE HETEROSCEDASTICIDADE

Os erros-padrão robustos em relação à heteroscedasticidade oferecem um método simples para calcular estatísticas t que sejam assimptoticamente distribuídas como t , haja ou não a presença de heteroscedasticidade. Também já vimos que as estatísticas F e LM robustas em relação à heteroscedasticidade estão disponíveis. A implementação desses testes não exige o conhecimento prévio da presença ou não de heteroscedasticidade. Não obstante, ainda existem algumas boas razões para fazermos alguns testes simples que possam detectar sua presença. Primeiro, como mencionamos na seção anterior, as estatísticas t habituais têm distribuições t exatas sob as hipóteses do modelo linear clássico. Por essa razão, muitos econometristas ainda preferem utilizar os erros-padrão MQO usuais e testar as estatísticas informadas, a menos que haja evidência de heteroscedasticidade. Segundo, se houver a presença de heteroscedasticidade, o estimador MQO não mais será o melhor estimador linear não-viesado. Como veremos na Seção 8.4, é possível obter um estimador melhor que o MQO quando a forma da heteroscedasticidade é conhecida.

Muitos testes de heteroscedasticidade têm sido sugeridos ao longo dos anos. Alguns deles, embora tenham a capacidade de detectar a heteroscedasticidade, não testam diretamente a hipótese de que a variância dos erros não depende das variáveis independentes. Nós nos restringiremos a testes mais modernos, que detectam os tipos de heteroscedasticidade que invalidam as estatísticas MQO habituais. Isso também traz o benefício de colocar todos os testes na mesma estrutura.

Como sempre, iniciamos com o modelo linear

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u, \quad (8.10)$$

onde as hipóteses RLM.1 a RLM.4 são mantidas nesta seção. Particularmente, assumimos que $E(u|x_1, x_2, \dots, x_k) = 0$, de forma que o MQO seja não-viesado e consistente.

Consideraremos como hipótese nula que a hipótese RLM.5 é verdadeira:

$$H_0: \text{Var}(u|x_1, x_2, \dots, x_k) = \sigma^2. \quad (8.11)$$

Ou seja, assumimos que a hipótese ideal de homoscedasticidade se mantém, e precisamos que os dados nos informem se isso é adequado ou não. Se não pudermos rejeitar (8.11) em um nível de significância suficientemente pequeno, normalmente concluímos que a heteroscedasticidade não será um problema. Porém, lembre-se de que nunca aceitamos H_0 ; simplesmente não podemos rejeitá-la.

Como estamos assumindo que u tem uma esperança condicional zero, $\text{Var}(u|x) = E(u^2|x)$, e assim a hipótese nula de homoscedasticidade é equivalente a

$$H_0: E(u^2|x_1, x_2, \dots, x_k) = E(u^2) = \sigma^2.$$

Isso mostra que, para testar a violação da hipótese de homoscedasticidade, queremos verificar se u^2 está relacionado (em valor esperado) a uma ou mais das variáveis explicativas. Se H_0 for falsa, o valor esperado de u^2 , dadas as variáveis independentes, pode ser virtualmente qualquer função de x_j . Um método simples é assumir uma função linear:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + v, \quad (8.12)$$

onde v é um termo erro com média zero, dados x_j . Preste bastante atenção na variável dependente nesta equação: ela é o quadrado do erro na equação de regressão original, (8.10). A hipótese nula de homoscedasticidade é

$$H_0: \delta_1 = \delta_2 = \dots = \delta_k = 0. \quad (8.13)$$

Sob a hipótese nula, freqüentemente é razoável assumir que o erro em (8.12), v , é independente de x_1, x_2, \dots, x_k . Então, sabemos da Seção 5.2 que a estatística F ou a estatística LM , sobre a significância geral das variáveis independentes na explicação de u^2 , podem ser usadas para testar (8.13). Ambas as estatísticas terão justificação assimptótica, mesmo que u^2 não possa ser normalmente distribuído. (Por exemplo, se u for normalmente distribuído, então u^2/σ^2 é distribuído como χ^2_1 .) Se pudéssemos observar u^2 na amostra, então poderíamos calcular com facilidade essa estatística, computando a regressão MQO de u^2 sobre x_1, x_2, \dots, x_k , usando todas as n observações.

Como enfatizamos anteriormente, nunca conhecemos os erros efetivos no modelo populacional, mas temos estimativas deles: o resíduo MQO, \hat{u}_i é uma estimativa do erro u_i para a observação i . Assim, podemos estimar a equação

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + \text{erro} \quad (8.14)$$

e calcular as estatísticas F ou LM da significância conjunta de x_1, \dots, x_k . Constata-se que o uso dos resíduos MQO em lugar dos erros não afeta a distribuição de amostra grande das estatísticas F ou LM , embora mostrar isso seja bastante complicado.

Ambas as estatísticas F e LM dependem do R -quadrado da regressão (8.14); chamemos isso de $R_{\hat{u}^2}^2$ para distingui-lo do R -quadrado na estimação da equação (8.10). Então, a estatística F será

$$F = \frac{R_{\hat{u}^2}^2/k}{(1 - R_{\hat{u}^2}^2)/(n - k - 1)}, \quad (8.15)$$

onde k é o número de regressores em (8.14); este é o mesmo número de variáveis independentes em (8.10). Calcular (8.15) raramente será necessário, já que a maioria dos programas econométricos calcula automaticamente a estatística F da significância geral de uma regressão. A estatística F tem (aproximadamente) uma distribuição $F_{k, n-k-1}$ sob a hipótese nula de homoscedasticidade.

A estatística LM para a heteroscedasticidade é simplesmente o tamanho da amostra multiplicado pelo R -quadrado de (8.14):

$$LM = n \cdot R_{\hat{u}^2}^2. \quad (8.16)$$

Sob a hipótese nula, a estatística LM é distribuída assintoticamente como χ^2_k . Isso também é obtido com muita facilidade após computarmos a regressão (8.14).

A versão LM do teste é geralmente chamada **teste de Breusch-Pagan da heteroscedasticidade (teste BP)**. Breusch e Pagan (1980) sugeriram uma forma diferente do teste o qual assume que os erros são normalmente distribuídos. Koenker (1983) sugeriu a forma da estatística LM em (8.16), que é em geral preferida devido a sua maior aplicabilidade.

Resumimos os passos para verificar a existência de heteroscedasticidade usando o teste BP:

O TESTE DE BREUSCH-PAGAN DA HETEROSCEDASTICIDADE

1. Estime o modelo (8.10) por MQO, como usual. Obtenha os resíduos quadrados MQO, \hat{u}^2 (um para cada observação).
2. Compute a regressão (8.14). Guarde o R -quadrado desta regressão, $R_{\hat{u}^2}^2$.
3. Construa a estatística F ou a estatística LM e calcule o p -valor (usando a distribuição $F_{k,n-k-1}$ no caso anterior e a distribuição χ_k^2 neste último caso). Se o p -valor for suficientemente pequeno, isto é, abaixo do nível de significância selecionado, então rejeitamos a hipótese nula de homoscedasticidade.

Se o teste BP resultar em um p -valor suficientemente pequeno, alguma medida corretiva deve ser tomada. Uma possibilidade é usar os erros-padrão robustos em relação à heteroscedasticidade e as estatísticas de testes discutidas na seção anterior. Outra possibilidade será discutida na Seção 8.4.

EXEMPLO 8.4

(Heteroscedasticidade nas Equações de Preços de Imóveis)

Utilizaremos os dados contidos no arquivo HPRICE1.RAW para verificar a existência de heteroscedasticidade em uma equação simples de preços de imóveis. A equação estimada usando os níveis de todas as variáveis é

$$\begin{aligned} \hat{\text{preço}} &= -21,77 + 0,00207 \text{ tamterr} + 0,123 \text{ arquad} + 13,85 \text{ qtdorm} \\ &\quad (29,48) \quad (0,00064) \quad (0,013) \quad (9,01) \\ &\quad n = 88, R^2 = 0,672. \end{aligned} \tag{8.17}$$

Esta equação não nos diz nada sobre se o erro no modelo populacional é heteroscedástico. Necessitamos fazer a regressão dos resíduos quadrados MQO sobre as variáveis independentes. O R -quadrado da regressão de \hat{u}^2 sobre tamterr , arquad e qtdorm é $R_{\hat{u}^2}^2 = 0,1601$. Com $n = 88$ e $k = 3$, isso produzirá uma estatística F para a significância das variáveis independentes de $F = [0,1601/(1 - 0,1601)](84/3) \approx 5,34$. O p -valor associado é 0,002, o que é forte evidência contra a hipótese nula. A estatística LM é $88(0,1601) \approx 14,09$; isso dá um p -valor $\approx 0,0028$ (usando a distribuição χ_3^2), produzindo essencialmente a mesma conclusão da estatística F . Isso significa que os erros-padrão usuais informados em (8.17) não são confiáveis.

No Capítulo 6 mencionamos que um dos benefícios de usar a forma funcional logarítmica da variável dependente é que a heteroscedasticidade é muitas vezes reduzida. Neste exemplo, coloquemos preço , tamterr e arquad em forma logarítmica, de forma que as elasticidades do preço em relação a tamterr e arquad sejam constantes. A equação estimada é

$$\begin{aligned} \log(\hat{\text{preço}}) &= -1,30 + 0,168 \log(\text{tamterr}) + 0,700 \log(\text{arquad}) + 0,037 \text{ qtdorm} \\ &\quad (0,65) \quad (0,038) \quad (0,093) \quad (0,028) \\ &\quad n = 88, R^2 = 0,643. \end{aligned} \tag{8.18}$$

Fazendo a regressão dos resíduos quadrados MQO desta regressão sobre $\log(\text{tamterr})$, $\log(\text{arquad})$ e qtdorm gera $R_{\hat{u}^2}^2 = 0,0480$. Assim, $F = 1,41$ (p -valor = 0,245) e $LM = 4,22$ (p -valor = 0,239). Portanto, não podemos rejeitar a hipótese nula de homoscedasticidade no modelo com a forma funcional logarítmica. A ocorrência de menos heteroscedasticidade com a variável dependente em forma logarítmica tem sido observada em muitas aplicações empíricas.

Considere a equação de salários (7.11), na qual você acredita que a variância condicional de $\log(\text{salário})$ não depende de $educ$, $exper$ ou $perm$. Porém, você está preocupado porque a variância de $\log(\text{salário})$ pode diferir ao longo dos quatro grupos demográficos de homens casados, mulheres casadas, homens solteiros e mulheres solteiras. Que regressão você faria para verificar a existência de heteroscedasticidade? Quais são os graus de liberdade no teste F ?

Se suspeitarmos que a heteroscedasticidade depende somente de certas variáveis independentes, podemos, com facilidade, modificar o teste de Breusch-Pagan: simplesmente fazemos a regressão de \hat{u}^2 sobre quaisquer variáveis independentes que escolhamos e aplicamos o teste F ou LM apropriado. Lembre-se de que os graus de liberdade apropriados dependem do número de variáveis independentes na regressão com \hat{u}^2 como variável dependente; o número de variáveis independentes que aparece na equação (8.10) é irrelevante.

Se os resíduos quadrados forem regredidos somente sobre uma única variável independente, o teste de heteroscedasticidade será a estatística t habitual da variável. Uma estatística t significativa sugere que a heteroscedasticidade é um problema.

O Teste de White para a Heteroscedasticidade

No Capítulo 5 mostramos que tanto os erros-padrão como as estatísticas de testes estimados habitualmente por MQO são assimptoticamente válidos, desde que todas as hipóteses de Gauss-Markov se mantenham. Acontece que a hipótese de homoscedasticidade, $\text{Var}(u_1|x_1, \dots, x_k) = \sigma^2$, pode ser substituída pela hipótese mais fraca de que o erro quadrado, u^2 , é *não-correlacionado* com todas as variáveis independentes (x_j), com os quadrados das variáveis independentes, (x_j^2), e com todos os produtos cruzados ($x_j x_h$ para $j \neq h$). Essa observação motivou White (1980) a propor um teste para a heteroscedasticidade que adiciona quadrados e produtos cruzados de todas as variáveis independentes à equação (8.14). O teste é explicitamente destinado a testar formas de heteroscedasticidade que invalidem os erros-padrão e as estatísticas de testes habituais, estimados por MQO.

Quando o modelo contém $k = 3$ variáveis independentes, o teste de White baseia-se na estimativa de

$$\begin{aligned}\hat{u}^2 = & \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 \\ & + \delta_7 x_1 x_2 + \delta_8 x_1 x_3 + \delta_9 x_2 x_3 + \text{erro}.\end{aligned}\quad (8.19)$$

Comparado com o teste de Breusch-Pagan, esta equação tem seis regressores a mais. O teste de White para a heteroscedasticidade é a estatística LM para testarmos que todos os δ_j na equação (8.19) sejam zero, exceto o intercepto. Assim, nove restrições estão sendo testadas neste caso. Também podemos usar um teste F desta hipótese; ambos os testes têm justificativa assimptótica.

Com somente três variáveis independentes no modelo original, a equação (8.19) tem nove variáveis independentes. Com seis variáveis independentes no modelo original, a regressão de White envolveria geralmente 27 regressores (a menos que alguns sejam redundantes). Essa abundância de regressores é uma fraqueza na forma pura do teste de White: ele usa muitos graus de liberdade para modelos com um número moderado de variáveis independentes.

É possível obter um teste que seja mais facilmente implementado que o teste de White e menos prejudicial quanto aos graus de liberdade. Para criar esse teste, observe que a diferença entre os testes de White e de Breusch-Pagan é que o primeiro inclui os quadrados e os produtos cruzados das variáveis independentes. Pode-se obter o mesmo resultado utilizando menos funções das variáveis independentes. Uma sugestão é usar os valores estimados MQO para verificar a existência de heteroscedasticidade. Lembre-se de que os valores estimados são definidos para cada observação i por

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}.$$

Os valores estimados são apenas funções lineares das variáveis independentes. Se eles forem elevados ao quadrado, obteremos uma função particular de todos os quadrados e produtos cruzados das variáveis independentes. Isso sugere testar a heteroscedasticidade estimando a equação

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \text{erro}, \quad (8.20)$$

onde \hat{y} representa os valores estimados. É importante não confundir \hat{y} com y nesta equação. Usamos os valores estimados porque eles são funções das variáveis independentes (e dos parâmetros estimados); o uso de y em (8.20) não produz um teste válido para a heteroscedasticidade.

Podemos usar as estatísticas F ou LM para a hipótese nula $H_0: \delta_1 = 0, \delta_2 = 0$ na equação (8.20). Isso resultará em duas restrições ao testar a hipótese nula da homoscedasticidade, a despeito do número de variáveis independentes no modelo original. Preservar os graus de liberdade desta maneira é quase sempre uma boa idéia, e também faz com que o teste seja implementado mais facilmente.

Como \hat{y} é uma estimativa do valor esperado de y , dadas as variáveis x_j , usar (8.20) para testar a heteroscedasticidade é útil nos casos em que há suspeita de que a variância muda com o nível do valor esperado, $E(y|x)$. O teste a partir de (8.20) pode ser visto como um caso especial do teste de White, já que pode ser mostrado que a equação (8.20) impõe restrições sobre os parâmetros da equação (8.19).

UM CASO ESPECIAL DO TESTE DE WHITE PARA A HETEROSCEDASTICIDADE:

1. Estime o modelo (8.10) por MQO, da maneira habitual. Obtenha os resíduos \hat{u} e os valores estimados \hat{y} do MQO. Calcule os resíduos quadrados \hat{u}^2 e os quadrados dos valores estimados \hat{y}^2 do MQO.
2. Compute a regressão da equação (8.20). Guarde o R -quadrado desta regressão, $R_{\hat{u}^2}^2$.
3. Construa as estatísticas F ou LM e calcule o p -valor (usando a distribuição $F_{2,n-3}$ no primeiro caso e a distribuição χ^2_2 no último).

EXEMPLO 8.5

(Forma Especial do Teste de White na Equação do Log dos Preços de Imóveis)

Aplicamos o caso especial do teste de White na equação (8.18) na qual usamos a estatística LM . O importante a lembrar é que a distribuição qui-quadrado sempre tem dois gr. A regressão de \hat{u}^2 sobre $lpreço$, $(lpreço)^2$, onde $lpreço$ representa os valores estimados em (8.18), produz $R_{\hat{u}^2}^2 = 0,0392$; assim, $LM = 88(0,0392) \approx 3,45$, e o p -valor = 0,178. Isso é evidência de heteroscedasticidade mais forte do que a fornecida pelo teste de Breusch-Pagan, mas ainda não podemos rejeitar a homoscedasticidade, mesmo ao nível de 15%.

Antes de deixarmos esta seção, devemos comentar uma importante limitação. Temos interpretado uma rejeição usando um dos testes de heteroscedasticidade como evidência de heteroscedasticidade. Isso está correto desde que mantenhamos as hipóteses RLM.1 a RLM.4. Mas se RLM.3 for violada — especialmente se a forma funcional de $E(y|x)$ estiver mal-especificada — então um teste de heteroscedasticidade pode rejeitar H_0 , mesmo se $\text{Var}(y|x)$ for constante. Por exemplo, se omitirmos um ou mais termos quadráticos em um modelo de regressão ou usarmos o modelo em nível quando deveríamos usar em log, um teste de heteroscedasticidade pode ser significante. Isso tem levado alguns economistas a verem os testes de heteroscedasticidade como testes generalizados de má especificação. Porém, existem testes melhores e mais diretos para testar a má especificação de formas funcionais, e estudaremos alguns deles na Seção 9.1. É melhor usar, primeiro, testes específicos de formas funcionais, já que a má especificação da forma funcional é mais importante que a heteroscedasticidade. Em seguida, uma vez que estejamos satisfeitos com a forma funcional, podemos fazer o teste para verificar a existência de heteroscedasticidade.

8.4 ESTIMAÇÃO DE MÍNIMOS QUADRADOS PONDERADOS

Se for detectada heteroscedasticidade com o uso de um dos testes da Seção 8.3, sahemos da Seção 8.2 que uma reação possível é usar estatísticas robustas em relação à heteroscedasticidade após a estimação MQO. Antes do desenvolvimento das estatísticas robustas em relação à heteroscedasticidade, a resposta à descoberta de heteroscedasticidade era modelar e estimar sua forma específica. Como veremos, isso leva a um estimador mais eficiente que o MQO, e produz estatísticas t e F que têm distribuições t e F . Embora isso pareça atraente, requer mais trabalho de nossa parte, pois temos de ser muito específicos sobre a natureza de qualquer heteroscedasticidade.

A Heteroscedasticidade É Percebida como uma Constante Multiplicativa

Considere que x representa todas as variáveis explicativas na equação (8.10) e assuma que

$$\text{Var}(u|x) = \sigma^2 h(x), \quad (8.21)$$

onde $h(x)$ é alguma função das variáveis explicativas que determina a heteroscedasticidade. Como variâncias devem ser positivas, $h(x) > 0$ para todos os valores possíveis das variáveis independentes. Supomos nesta subseção que a função $h(x)$ é conhecida. O parâmetro populacional σ^2 é desconhecido, mas teremos condições de estimá-lo a partir de uma amostra de dados.

Para uma extração aleatória da população, podemos escrever $\sigma_i^2 = \text{Var}(u_i|x_i) = \sigma^2 h(x_i) = \sigma^2 h_i$, onde novamente usamos a notação x_i para representar todas as variáveis independentes para as observações i , enquanto h_i muda a cada observação porque as variáveis independentes mudam ao longo das observações. Por exemplo, considere a função simples de poupança

$$poup_i = \beta_0 + \beta_1 renda_i + u_i \quad (8.22)$$

$$\text{Var}(u_i|renda_i) = \sigma^2 renda_i \quad (8.23)$$

Aqui, $h(\mathbf{x}) = h(\text{renda}) = \text{renda}$; a variância do erro é proporcional ao nível da renda. Isso significa que, conforme a renda aumenta, a variabilidade da poupança cresce. (Se $\beta_1 > 0$, o valor esperado da poupança também aumenta com a renda.) Como renda é sempre positiva, a variância na equação (8.23) garantidamente será sempre positiva. O desvio-padrão de u_i , condicional em renda_i , é $\sigma\sqrt{\text{renda}}_i$.

Como podemos usar a informação da equação (8.21) para estimar β_j ? Essencialmente, levamos em conta a equação original,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i, \quad (8.24)$$

que contém erros heteroscedásticos, e a transformamos em uma equação que não contenha tais erros (e satisfaça as outras hipóteses de Gauss-Markov). Como h_i é apenas uma função de x_i , $\mu_i/\sqrt{h_i}$ tem zero como valor esperado condicional em x_i . Além disso, como $\text{Var}(u_i|x_i) = E(u_i^2|x_i) = \sigma^2 h_i$, a variância de $u_i/\sqrt{h_i}$ (condicional em x_i) é σ^2 :

$$E((u_i/\sqrt{h_i})^2) = E(u_i^2)/h_i = (\sigma^2 h_i)/h_i = \sigma^2,$$

onde suprimimos a condicionalidade em x_i , por simplicidade. Podemos dividir a equação (8.24) por $\sqrt{h_i}$ para obter

$$\begin{aligned} y_i/\sqrt{h_i} &= \beta_0/\sqrt{h_i} + \beta_1(x_{i1}/\sqrt{h_i}) + \beta_2(x_{i2}/\sqrt{h_i}) + \dots \\ &\quad + \beta_k(x_{ik}/\sqrt{h_i}) + (u_i/\sqrt{h_i}) \end{aligned} \quad (8.25)$$

ou

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \dots + \beta_k x_{ik}^* + u_i^*, \quad (8.26)$$

onde $x_{i0}^* = 1/\sqrt{h_i}$ e as outras variáveis sobrescritas com * representam as variáveis originais correspondentes divididas por $\sqrt{h_i}$.

A equação (8.26) parece um pouco peculiar, mas o importante a ser lembrado é que a derivamos para podermos obter os estimadores de β_j que tenham propriedades de eficiência melhores que MQO. O intercepto β_0 na equação original (8.24) agora está multiplicando a variável $x_{i0}^* = 1/\sqrt{h_i}$. Cada parâmetro de inclinação em β_j multiplica uma nova variável que raramente tem interpretação útil. Isso não deve causar problemas se lembrarmos que, na interpretação dos parâmetros e do modelo, sempre queremos retornar à equação original (8.24).

No exemplo precedente da poupança, a equação transformada se assemelha a

$$\text{poup}_i/\sqrt{\text{renda}}_i = \beta_0(1/\sqrt{\text{renda}}_i) + \beta_1\sqrt{\text{renda}}_i + u_i^*,$$

onde usamos o fato de que $\text{renda}_i/\sqrt{\text{renda}}_i = \sqrt{\text{renda}}_i$. Contudo, β_1 é a propensão marginal a poupar, uma interpretação que obtemos da equação (8.22).

A equação (8.26) é linear em seus parâmetros (portanto satisfaz RLM.1), e a hipótese de amostragem aleatória não se alterou. Além disso, u_i^* tem uma média zero e uma variância constante (σ^2), condicional em x_i^* . Isso significa que se a equação original satisfizer as quatro primeiras hipóteses de Gauss-Markov, então a equação transformada (8.26) satisfará todas as cinco hipóteses de Gauss-Markov. Além disso, se u_i tiver uma distribuição normal, então u_i^* terá uma distribuição normal com variância σ^2 . Portanto, a equação transformada satisfará as hipóteses do modelo linear clássico (RLM.1 a RLM.6), se o modelo original também o fizer, com exceção da hipótese de homoscedasticidade.

Como sabemos que o MQO tem propriedades atraentes (BLUE, por exemplo) sob as hipóteses de Gauss-Markov, a discussão no parágrafo anterior sugere estimarmos os parâmetros da equação (8.26) por mínimos quadrados ordinários. Esses estimadores, $\beta_0^*, \beta_1^*, \dots, \beta_k^*$, serão diferentes dos estimadores MQO na equação original. Os β_j^* são exemplos de **estimadores de mínimos quadrados generalizados (MQG)**. Neste caso, os estimadores MQG são usados para explicar a heteroscedasticidade nos erros. Encontraremos outros estimadores MQG no Capítulo 12.

Como a equação (8.26) satisfaz todas as hipóteses ideais, erros-padrão, estatísticas t e estatísticas F podem ser obtidas de regressões que usem as variáveis transformadas. A soma dos quadrados dos resíduos em (8.26) dividida pelos graus de liberdade é um estimador não-viesado de σ^2 . Além disso, os estimadores MQG, por serem os melhores estimadores lineares não-viesados de β_j , são necessariamente mais eficientes que os estimadores MQO $\hat{\beta}_j$ obtidos da equação não transformada. Em essência, após termos transformado as variáveis, simplesmente usamos a análise padrão MQO. Porém, devemos nos lembrar de interpretar as estimativas à luz da equação original.

O R -quadrado obtido da estimação em (8.26), embora seja útil para computar estatísticas F , não é especialmente informativo como uma medida do grau de ajuste: ele nos informa quanto da variação em y^* é explicado por x_j^* , e isso raramente é significativo.

Os estimadores MQG para a correção da heteroscedasticidade são chamados de **estimadores de mínimos quadrados ponderados (MQP)**. Esse nome advém do fato de que β_j^* minimiza a soma *ponderada* dos quadrados dos resíduos, onde cada resíduo quadrado é ponderado por $1/h_i$. A idéia é colocar menos peso nas observações com uma variância de erro mais alta; o método MQO dá a cada observação o mesmo peso, pois isso é melhor quando a variância do erro é idêntica para todas as partições da população. Matematicamente, os estimadores MQP são os valores de b_j que tornam a expressão

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \dots - b_k x_{ik})^2 / h_i \quad (8.27)$$

tão pequena quanto possível. Levar a raiz quadrada de $1/h_i$ para dentro do resíduo quadrado mostra que a soma ponderada dos quadrados dos resíduos é idêntica à soma dos quadrados dos resíduos nas variáveis transformadas:

$$\sum_{i=1}^n (y_i^* - b_0 x_{i0}^* - b_1 x_{i1}^* - b_2 x_{i2}^* - \dots - b_k x_{ik}^*)^2.$$

Como MQO minimiza a soma dos resíduos quadrados (a despeito das definições das variáveis dependente e independentes), concluímos que os estimadores MQP que minimizam (8.27) são simplesmente os estimadores MQO de (8.26). Observe cuidadosamente que os resíduos quadrados em (8.27) são ponderados por $1/h_i$, enquanto as variáveis transformadas em (8.26) são ponderadas por $1/\sqrt{h_i}$.

Um estimador de mínimos quadrados ponderados pode ser definido para qualquer conjunto de pesos positivos. O MQO é o caso especial que atribui pesos iguais a todas as observações. O procedi-

mento eficiente, MQG, pondera cada resíduo quadrado pelo inverso da variância condicional de u_i , dado x_i .

A obtenção das variáveis transformadas na equação (8.25) com o objetivo de calcular manualmente os mínimos quadrados ponderados pode ser entediante, e não se pode desprezar a possibilidade de cometer erros. Felizmente, a maioria dos programas econometrícios modernos tem um recurso para computar mínimos quadrados ponderados. Em geral, juntamente com as variáveis dependentes e independentes no modelo original, apenas especificamos a função de ponderação, $1/h_i$, que aparece em (8.27). Isto é, especificamos pesos proporcionais ao inverso da variância, e não ao desvio-padrão. Além de haver menor possibilidade de cometermos erros, isso nos força a interpretar as estimativas de mínimos quadrados ponderados no modelo original. Aliás, podemos escrever a equação estimada da maneira habitual. As estimativas e os erros-padrão serão diferentes do MQO, mas a maneira como interpretamos essas estimativas, erros-padrão e estatísticas de testes é a mesma.

EXEMPLO 8.6

(Equação de Poupança Familiar)

A Tabela 8.1 contém estimativas de funções de poupança construídas a partir do arquivo de dados SAVING.RAW (sobre 100 famílias, desde 1970). Estimamos o modelo de regressão simples (8.22) por MQO e por mínimos quadrados ponderados, assumindo no último caso que a variância é dada por (8.23). Em seguida, adicionamos as variáveis tamanho da família, idade do chefe da família, anos de escolaridade do chefe da família e uma variável *dummy* indicando se o chefe da família é negro.

No modelo de regressão simples, a estimativa MQO da propensão marginal a poupar (PMP) é 0,147, com uma estatística t de 2,53. (Os erros-padrão MQO na Tabela 8.1 são os erros-padrão não-robustos. Se realmente considerássemos que a heteroscedasticidade seria um problema, provavelmente computaríamos também os erros-padrão robustos em relação à heteroscedasticidade, o que não faremos neste caso.) A estimativa MQP da PMP é um pouco mais alta: 0,172, com $t = 3,02$. Os erros-padrão das estimativas MQO e MQP são muito semelhantes para este coeficiente. As estimativas MQO e MQP do intercepto são bastante diferentes, mas isso não deve causar preocupação já que as estatísticas t são ambas muito pequenas. Encontrar mudanças razoavelmente grandes em coeficientes que não são significantes não é incomum quando se comparam as estimativas MQO e MQP. Os R -quadrados nas colunas (1) e (2) não são comparáveis.

A adição de variáveis demográficas reduz a PMP, quer usemos o MQO ou o MQP; os erros-padrão também aumentam em quantidade razoável (devido à multicolinearidade que é induzida pela inclusão dessas variáveis adicionais). É fácil verificar, usando as estimativas do MQO ou do MQP, que nenhuma das variáveis adicionais é, individualmente, significativa. Serão elas significativas conjuntamente? O teste F baseado na estimativa MQO usa os R -quadrados das colunas (1) e (3). Com 94 g_f no modelo irrestrito e com quatro restrições, a estatística F é $F = [(0,0828 - 0,0621)/(1 - 0,0828)](94/4) \approx 0,53$ e o p -valor = 0,715. O teste F , usando as estimativas MQP, usa os R -quadrados das colunas (2) e (4): $F \approx 0,50$ e o p -valor = 0,739. Assim, usando o MQO ou o MQP, as variáveis demográficas são conjuntamente não significantes. Isso sugere que o modelo de regressão simples relacionando poupança à renda é suficiente.

Qual deveríamos escolher como nossa melhor estimativa da propensão marginal a poupar? Neste caso, não importa muito se usarmos a estimativa MQO de 0,147 ou a estimativa MQP de 0,172. Lembre-se de que ambas são apenas estimativas de uma amostra relativamente pequena, e o intervalo de confiança de 95% do MQO contém a estimativa MQP, e vice-versa.

Tabela 8.1

Variável Dependente: poup

Variáveis Independentes	(1) MQO	(2) MQP	(3) MQO	(4) MQP
<i>renda</i>	0,147 (0,058)	0,172 (0,057)	0,109 (0,071)	0,101 (0,077)
<i>tamanho</i>	—	—	67,66 (222,96)	-6,87 (168,43)
<i>educ</i>	—	—	151,82 (117,25)	139,48 (100,54)
<i>idade</i>	—	—	0,286 (50,031)	21,75 (41,31)
<i>negro</i>	—	—	518,39 (1.308,06)	137,28 (844,59)
<i>intercepto</i>	124,84 (655,39)	-124,95 (480,86)	-1.605,42 (2.830,71)	-1.854,81 (2.351,80)
<i>Observações</i>	100	100	100	100
<i>R-quadrado</i>	0,0621	0,0853	0,0828	0,1042

Usando os resíduos MQO obtidos da regressão MQO descrita na coluna (1) da Tabela 8.1, a regressão de \hat{u}^2 sobre *renda* produz uma estatística *t* do coeficiente de *renda* de 0,96. Existe alguma necessidade de se usar mínimos quadrados ponderados no Exemplo 8.6?

Na prática, raramente sabemos na forma simples como a variância depende de uma determinada variável independente. Por exemplo, na equação de poupança que inclui todas as variáveis demográficas, como sabemos que a variância de *poup* não muda com a idade ou os níveis de educação? Na maioria das aplicações ficamos inseguros sobre $\text{Var}(y|x_1, x_2, \dots, x_k)$.

Existe um caso no qual os pesos necessários para o MQP surgem naturalmente de um modelo econométrico subjacente. Isso acontece quando, em vez de usarmos dados em nível individual, somente temos médias de dados de algum grupo ou região geográfica. Por exemplo, suponha que estejamos interessados em determinar a relação entre o montante que um trabalhador contribui para seu plano de pensão como uma função da generosidade do plano. Suponhamos que *i* seja uma empresa em particular e que *e* represente um empregado dessa empresa. Um modelo simples é

$$\text{contrib}_{i,e} = \beta_0 + \beta_1 \text{ganhos}_{i,e} + \beta_2 \text{idade}_{i,e} + \beta_3 \text{taxcont}_i + u_{i,e}, \quad [8.28]$$

onde $contrib_{i,e}$ é a contribuição anual por empregado e que trabalha na empresa i , $ganhos_{i,e}$ é o ganho anual dessa pessoa, e $idade_{i,e}$ é a idade dessa pessoa. A variável $taxcont_i$ é o montante que a empresa deposita na conta do empregado para cada dólar pago em contribuição pelo empregado.

Se (8.28) satisfizer as hipóteses de Gauss-Markov, então poderemos estimá-la, dada uma amostra de indivíduos entre vários empregadores. Suponha, porém, que somente temos valores *médios* de contribuições, ganhos e idade, por empregador. Em outras palavras, dados em nível individual não estão disponíveis. Assim, façamos com que $contrib_i$ represente a contribuição média do pessoal da empresa i e, semelhantemente, $ganhos_i$ e $idade_i$, também representem médias. Seja m_i o número de empregados da empresa i ; assumimos que esse é um número conhecido. Então, se computarmos a média da equação (8.28) para todos os empregados da empresa i , obteremos a equação dela

$$\overline{contrib}_i = \beta_0 + \beta_1 \overline{ganhos}_i + \beta_2 \overline{idade}_i + \beta_3 taxcont_i + \bar{u}_i, \quad (8.29)$$

onde $\bar{u}_i = m_i^{-1} \sum_{e=1}^{m_i} u_{i,e}$ é o erro médio entre todos os empregados da empresa i . Se tivermos n empresas na nossa amostra, então (8.29) será apenas um modelo de regressão linear múltipla que pode ser estimado por MQO. Os estimadores serão não-viesados se o modelo original (8.28) satisfizer as hipóteses de Gauss-Markov e os erros individuais $u_{i,e}$ serão independentes do tamanho da empresa, m_i [porque então o valor esperado de \bar{u}_i , dadas as variáveis explicativas em (8.29), será zero].

Se a equação no nível individual satisfizer a hipótese de homoscedasticidade, então a equação da empresa (8.29) deverá ter heteroscedasticidade. Assim, se $Var(u_{i,e}) = \sigma^2$ para todo i e e , então $Var(\bar{u}_i) = \sigma^2/m_i$. Em outras palavras, para empresas maiores, a variância do termo erro \bar{u}_i diminui com o tamanho da empresa. Neste caso, $h_i = 1/m_i$, e, portanto, o procedimento mais eficiente será o dos mínimos quadrados ponderados, com pesos correspondentes ao número de empregados das empresas ($1/h_i = m_i$). Isso garante que empresas grandes recebam maior peso, o que nos oferece um método eficiente de estimação dos parâmetros no modelo em nível individual quando somente temos médias para as empresas.

Uma ponderação semelhante surge quando estamos usando dados *per capita* no nível de cidade, município, estado ou país. Se a equação no nível individual satisfizer as hipóteses de Gauss-Markov, o erro na equação *per capita* terá uma variância proporcional a um sobre o tamanho da população. Portanto, mínimos quadrados ponderados com pesos iguais à população serão apropriados. Por exemplo, suponha que temos dados em nível de cidades sobre o consumo *per capita* de cerveja (em onças*), a percentagem de pessoas com mais de 21 anos na população, níveis médios de educação dos adultos, níveis médios de renda e o preço da cerveja nas cidades. Então, o modelo no nível de cidades

$$cervpc = \beta_0 + \beta_1 perc21 + \beta_2 educmed + \beta_3 rendapc + \beta_4 preco + u$$

pode ser estimado por mínimos quadrados ponderados, com pesos iguais à população das cidades.

A vantagem de fazer a ponderação pelos tamanhos da empresa, população da cidade, e assim por diante, depende de a equação individual subjacente ser homoscedástica. Se existir heteroscedasticidade no nível individual, então a ponderação adequada dependerá da forma da heteroscedasticidade. Esta é uma razão que explica porque um número cada vez maior de pesquisadores simplesmente computa erros-padrão e estatísticas de teste robustas na estimação de modelos que usam dados *per capita*. Uma

* NT: 1 onça = 29,574 mililitros.

alternativa é ponderar pela população, mas registrar as estatísticas robustas em relação à heteroscedasticidade na estimação MQP. Isso assegura que, embora a estimação seja eficiente se o modelo no nível individual satisfizer as hipóteses de Gauss-Markov, qualquer heteroscedasticidade no nível individual seja representada pela inferência robusta.

A Necessidade de Estimar a Função de Heteroscedasticidade: O MQG Factível

Na subseção anterior, vimos alguns exemplos nos quais a heteroscedasticidade é conhecida como uma forma multiplicativa. Na maioria dos casos, a forma exata de heteroscedasticidade não é óbvia. Em outras palavras, é difícil encontrar a função $h(x_i)$ da seção anterior. Contudo, em muitos casos podemos modelar a função h e utilizar os dados para estimar os parâmetros desconhecidos nesse modelo. Isso resulta em uma estimativa de cada h_i , indicada por \hat{h}_i . O uso de \hat{h}_i em lugar de h_i na transformação MQG produz um estimador chamado **estimador MQG factível (MQGF)**. O MQG factível algumas vezes é chamado de *MQG estimado* ou MQGE.

Existem várias maneiras de modelar a heteroscedasticidade, mas estudaremos um método particular razoavelmente flexível. Assuma que

$$\text{Var}(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k), \quad (8.30)$$

onde x_1, x_2, \dots, x_k são variáveis independentes que aparecem no modelo de regressão [veja equação (8.1)], e δ_j são parâmetros desconhecidos. Outras funções de x_j podem aparecer, mas nos concentraremos primariamente em (8.30). Na notação da subseção anterior, $h(x) = \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k)$.

Você pode estar curioso para saber por que usamos a função exponencial em (8.30). Afinal de contas, quando fizemos o teste para verificar a existência de heteroscedasticidade usando o teste de Breusch-Pagan, assumimos que a heteroscedasticidade era uma função linear de x_j . Alternativas lineares como (8.12) são boas quando verificamos a existência de heteroscedasticidade, mas elas podem ser problemáticas quando fazemos a correção da heteroscedasticidade usando mínimos quadrados ponderados. Já encontramos a razão para esse problema antes: modelos lineares não asseguram que os valores previstos sejam positivos, e nossas variâncias estimadas devem ser positivas para podermos usar o método MQP.

Se os parâmetros δ_j fossem conhecidos, apenas aplicaríamos o MQP, como na subseção anterior. Isso não é muito realista. É melhor usar os dados para estimar esses parâmetros, e então utilizar essas estimativas para construir os pesos. Como podemos estimar os δ_j ? Basicamente, transformaremos essa equação em uma forma linear que, com pequenas modificações, poderá ser estimada por MQO.

Sob a hipótese (8.30), podemos escrever

$$u^2 = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k) v,$$

onde v tem uma média igual à unidade, condicional em $x = (x_1, x_2, \dots, x_k)$. Se assumirmos que v é realmente independente de x , podemos escrever

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k + e, \quad (8.31)$$

onde e tem média zero e é independente de x ; o intercepto nessa equação é diferente de δ_0 , mas isso não é importante. A variável dependente é o log do erro quadrado.

Como (8.31) satisfaz as hipóteses de Gauss-Markov, podemos obter estimadores não-viesados de δ_j utilizando MQO.

Como sempre, devemos substituir o u não-observado pelos resíduos MQO. Portanto, computamos a regressão de

$$\log(\hat{u}^2) \text{ sobre } x_1, x_2, \dots, x_k. \quad (8.32)$$

Na realidade, o que necessitamos dessa regressão são os valores estimados; vamos chamá-los de \hat{g}_i . Então, as estimativas de h_i serão simplesmente

$$\hat{h}_i = \exp(\hat{g}_i). \quad (8.33)$$

Agora usamos o método MQP com pesos $1/\hat{h}_i$ em lugar de $1/h_i$ na equação (8.27). Façamos um resumo dos passos.

UM PROCEDIMENTO MQG FACTÍVEL PARA CORRIGIR A HETEROSCEDASTICIDADE

1. Execute a regressão de y sobre x_1, x_2, \dots, x_k e obtenha os resíduos \hat{u} .
2. Crie $\log(\hat{u}^2)$ primeiramente elevando ao quadrado os resíduos MQO e depois calculando seu log natural.
3. Execute a regressão na equação (8.32) e obtenha os valores estimados, \hat{g} .
4. Calcule o exponencial dos valores estimados a partir de (8.32): $\hat{h} = \exp(\hat{g})$.
5. Estime a equação

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

pelo método MQP, usando pesos $1/\hat{h}$.

Se pudéssemos usar h_i em vez de \hat{h}_i no procedimento MQP, saberíamos que nossos estimadores seriam não-viesados; de fato, eles seriam os melhores estimadores lineares não-viesados, supondo que tenhamos modelado apropriadamente a heteroscedasticidade. Estimar h_i usando os mesmos dados significa que o estimador MQGV deixa de ser não-viesado (portanto, ele tampouco pode ser BLUE). No entanto, o estimador MQGV é consistente e assintoticamente mais eficiente que o MQO. Isso é difícil de demonstrar devido à estimação dos parâmetros da variância. Entretanto, se ignorarmos isso — o que, no fim das contas, faremos — a prova é semelhante à demonstração de que o MQO é eficiente na classe de estimadores no Teorema 5.3. De qualquer forma, para amostras de tamanho grande, o MQGV é uma atraente alternativa ao MQO quando existe evidência de heteroscedasticidade que infla os erros-padrão das estimativas MQO.

Devemos lembrar que os estimadores MQGV são estimadores dos parâmetros na equação

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u.$$

Assim como as estimativas MQO medem o impacto marginal de cada x_j sobre y , as estimativas MQGV também têm essa característica. Usamos as estimativas MQGV em lugar das estimativas MQO porque

elas são mais eficientes e possuem estatísticas de testes associadas às distribuições t e F usuais, pelo menos em amostras grandes. Se tivermos alguma dúvida sobre a variância especificada na equação (8.30), poderemos usar os erros-padrão e as estatísticas de testes robustos em relação à heteroscedasticidade na equação transformada.

Outra alternativa útil para estimar h_i é substituir as variáveis independentes na regressão (8.32) pelos valores estimados por MQO e seus quadrados. Em outras palavras, obter os \hat{g}_i , como os valores estimados da regressão de

$$\log(\hat{u}^2) \text{ sobre } \hat{y}, \hat{y}^2 \quad (8.34)$$

e depois obter os \hat{h}_i , exatamente como na equação (8.33). Isso altera apenas o passo (3) do procedimento anterior.

Se usarmos a regressão (8.32) para estimar a função da variância, você vai querer saber se podemos simplesmente fazer o teste para verificar a existência de heteroscedasticidade usando essa mesma regressão (um teste F ou LM pode ser usado). Aliás, Park (1966) fez essa sugestão. Infelizmente, quando comparado com os testes discutidos na Seção 8.3, o teste de Park tem alguns problemas. Primeiro, a hipótese nula deve ser algo mais forte que a homoscedasticidade: efetivamente, u e x devem ser independentes. Isso não é exigido pelos testes de Breusch-Pagan ou White. Segundo, o uso dos resíduos MQO \hat{u} em lugar de u em (8.32) pode fazer com que a estatística F se desvie da distribuição F , mesmo em amostras de tamanho grande. Isso não é um problema nos outros testes que tratamos. Por essas razões, o teste de Park não é recomendado quando estivermos fazendo testes para verificar a existência de heteroscedasticidade. A razão pela qual a regressão (8.32) funciona bem para mínimos quadrados ponderados é que somente precisamos de estimadores consistentes de δ_j , e a regressão (8.32) certamente os produz.

EXEMPLO 8.7

(Demanda de Cigarros)

Utilizamos os dados contidos no arquivo SMOKE.RAW para estimar uma função de demanda de consumo diário de cigarros. Como a maioria das pessoas não fuma, a variável dependente $cigs$ é zero para a maioria das observações. Um modelo linear não é o ideal, pois ele pode produzir valores previstos negativos. Mesmo assim, podemos aprender alguma coisa sobre os determinantes do hábito de fumar utilizando um modelo linear.

A equação estimada por mínimos quadrados ponderados, com os erros-padrão MQO usuais entre parênteses, é

$$\begin{aligned} \hat{cigs} = & -3,64 + 0,880 \log(renda) - 0,751 \log(precig) \\ & (24,08) \quad (0,728) \quad (5,773) \\ & -0,501 educ + 0,771 idade - 0,0090 idade^2 - 2,83 restaurn \\ & (0,167) \quad (0,160) \quad (0,0017) \quad (1,11) \\ & n = 807, R^2 = 0,0526, \end{aligned} \quad (8.35)$$

EXEMPLO 8.7 (continuação)

onde $cigs$ é o número de cigarros fumados por dia, $renda$ é a renda anual, $precig$ é o preço por maço de cigarros (em centavos de dólar), $educ$ representa anos de escolaridade formal, $idade$ é medida em anos e $restaurn$ é um indicador binário igual a um se a pessoa residir em um estado com restrições de fumar em restaurantes. Como também vamos trabalhar com o método de mínimos quadrados ponderados, não registraremos os erros-padrão robustos em relação à heteroscedasticidade do MQO. (A propósito, 13 dos 807 valores estimados são menores que zero; isso é menos de 2% da amostra e não é motivo importante para preocupação.)

Nem a renda nem o preço dos cigarros são estatisticamente significantes em (8.35), e seus efeitos não são grandes na prática. Por exemplo, se a renda aumenta em 10%, $cigs$ aumenta previsivelmente em $(0,880/100)(10) = 0,088$, ou menos de um décimo de cigarro por dia. A magnitude do efeito do preço é semelhante.

Cada ano de educação formal reduz a média de cigarros fumados por dia à metade, e o efeito é estatisticamente significante. O hábito de fumar também está relacionado com a idade, de um modo quadrático, aumenta com a idade até a $idade = 0,771/[2(0,009)] \approx 42,83$, e depois diminui com a idade. Os dois termos relacionados à idade são estatisticamente significantes. A presença de uma restrição quanto a fumar em restaurantes reduz o hábito de fumar em quase três cigarros por dia, em média.

Os erros subjacentes na equação (8.35) contêm heteroscedasticidade? A regressão de Breusch-Pagan dos resíduos quadrados MQO sobre as variáveis independentes em (8.35) [veja equação (8.14)] produz $R^2_{\text{BP}} = 0,040$. Esse R -quadrado pequeno pode parecer indicar a não-existência de heteroscedasticidade, mas devemos nos lembrar de calcular a estatística F ou a LM. Se o tamanho da amostra for grande, um R^2_{BP} aparentemente pequeno pode resultar em uma rejeição muito forte da homoscedasticidade. A estatística LM é $807(0,040) = 32,28$, e esse é o resultado de uma variável aleatória χ^2_6 . O p -valor é menor que 0,000015, o que é uma evidência muito forte de heteroscedasticidade.

Portanto, estimamos a equação usando o procedimento MQG factível anterior. A equação estimada é

$$\begin{aligned} \hat{cigs} &= 5,64 + 1,30 \log(renda) - 2,94 \log(precig) \\ &\quad (17,80) \quad (0,44) \qquad \qquad \qquad (4,46) \\ &- 0,463 educ + 0,482 idade - 0,0056 idade^2 - 3,46 restaurn \\ &\quad (0,120) \quad (0,097) \quad (0,0009) \quad (0,80) \\ &\qquad \qquad \qquad n = 807, R^2 = 0,1134. \end{aligned} \tag{8.36}$$

O efeito da renda agora é estatisticamente significante e maior em magnitude. O efeito do preço é, também, notavelmente maior, mas continua sendo estatisticamente não significante. [Uma razão para isso é que $precig$ varia somente entre estados na amostra, e assim existe muito menos variação em $\log(precig)$ do que em $\log(renda)$, $educ$ e $idade$.]

As estimativas sobre as outras variáveis, naturalmente, mudaram um pouco, mas a história continua a mesma. Fumar está relacionado negativamente com escolaridade, tem uma relação quadrática com a idade, e é negativamente afetado pelas restrições de fumar em restaurantes.

Devemos ter algum cuidado ao calcularmos estatísticas F para testar hipóteses múltiplas após a estimativa por MQP. (Isso é válido se a soma dos resíduos quadrados ou a forma R -quadrada da estatística F for usada.) É importante que os mesmos pesos sejam usados para estimar os modelos com e

sem restrições. Devemos primeiro estimar o modelo sem restrições por MQO. Uma vez que tenhamos obtido os pesos, poderemos usá-los para também estimar o modelo restrito. A estatística F pode ser calculada da maneira habitual. Felizmente, muitos programas econométricos possuem um comando simples para testar restrições conjuntas após a estimação MQP, de modo que não precisamos, nós mesmos, calcular manualmente a regressão restrita.

Suponha que o modelo da heteroscedasticidade na equação (8.30) não esteja correto, mas usamos o procedimento MQG factível baseado nessa variância. O MQP ainda é coerente, mas os usuais erros-padrão, estatísticas t etc. não serão válidos, mesmo assimptoticamente. O que podemos fazer? [Sugestão: Veja a equação (8.26), na qual u_i^* contém heteroscedasticidade se $\text{Var}(u|x) \neq \sigma^2 h(x)$.]

O Exemplo 8.7 sugere uma questão que às vezes surge em aplicações de mínimos quadrados ponderados: as estimativas MQO e MQP podem ser substancialmente diferentes. Isso não é um grande problema na equação de demanda de cigarros, pois todos os coeficientes mantêm os mesmos sinais, e as maiores mudanças são nas variáveis que eram estatisticamente não significantes quando a equação foi estimada por MQO. As estimativas MQO e MQP sempre serão diferentes devido ao erro amostral. O problema é quando suas diferenças são suficientes para alterar conclusões importantes.

Se os métodos MQO e MQP produzirem estimativas estatisticamente significantes que sejam diferentes nos sinais — por exemplo, a elasticidade-preço estimada por MQO é significante e positiva, enquanto a estimada por MQP é significante e negativa — ou se as diferenças em magnitude das estimativas forem de fato grandes, devemos ficar desconfiados. Em geral, isso indica que uma das outras hipóteses de Gauss-Markov é falsa, particularmente a hipótese de média condicional dos erros nula (RLM.3). A correlação entre u e qualquer variável independente causa viés e inconsistência no MQO e no MQP, e os vieses normalmente serão diferentes. O teste de Hausman [Hausman (1978)] pode ser usado para comparar formalmente as estimativas MQO e MQP para verificar se elas diferem mais do que é sugerido pelo erro amostral. Esse teste está além do escopo deste livro. Em muitos casos, um exame informal das estimativas é suficiente para detectar o problema.

8.5 O MODELO DE PROBABILIDADE LINEAR REVISITADO

Como vimos na Seção 7.5, quando a variável dependente y é binária, o modelo deve conter heteroscedasticidade, a menos que todos os parâmetros de inclinação sejam nulos. Estamos agora em posição de lidar com esse problema.

A maneira mais simples de tratar a heteroscedasticidade no modelo de probabilidade linear é continuar a usar a estimação MQO, mas também calcular os erros-padrão robustos nas estatísticas de testes. Isso ignora o fato de que efetivamente conhecemos a forma da heteroscedasticidade do MPL. Contudo, as estimativas MQO do MPL são simples e geralmente produzem resultados satisfatórios.

EXEMPLO 8.8**{Participação de Mulheres Casadas na Força de Trabalho}**

No exemplo da participação na força de trabalho na Seção 7.5 [veja equação (7.29)], registramos os erros-padrão MQO habituais. Agora, também computamos os erros-padrão robustos em relação à heteroscedasticidade. Eles estão registrados entre parênteses abaixo dos erros-padrão usuais:

$$\begin{aligned}
 \hat{naft} &= 0,586 - 0,0034 \text{ nesprend} + 0,038 \text{ educ} + 0,039 \text{ exper} \\
 (0,154) &\quad (0,0014) \quad (0,007) \quad (0,006) \\
 [0,151] &\quad [0,0015] \quad [0,007] \quad [0,006] \\
 &- 0,00060 \text{ exper}^2 - 0,016 \text{ idade} - 0,262 \text{ crianme6} + 0,0130 \text{ crianma6} \quad (8.37) \\
 (0,00018) &\quad (0,002) \quad (0,034) \quad (0,0132) \\
 [0,00019] &\quad [0,002] \quad [0,032] \quad [0,0135] \\
 n &= 753, R^2 = 0,264.
 \end{aligned}$$

Vários dos erros-padrão robustos e MQO são os mesmos para o grau de precisão registrado; em todos os casos, as diferenças são de fato muito pequenas. Portanto, embora a heteroscedasticidade seja um problema na teoria, ela não é na prática, pelo menos neste exemplo. Muitas vezes constata-se que os erros-padrão e estatísticas de testes usuais MQO são semelhantes aos seus correspondentes robustos em relação à heteroscedasticidade. E mais ainda, o esforço exigido para computar ambos é mínimo.

Geralmente, os estimadores MQO são inefficientes no MPL. Lembre-se de que a variância condicional de y no MPL é

$$\text{Var}(y|x) = p(x)[1 - p(x)], \quad (8.38)$$

onde

$$p(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (8.39)$$

é a probabilidade de resposta (probabilidade de sucesso, $y = 1$). Parece natural usar mínimos quadrados ponderados, mas existem alguns obstáculos. A probabilidade $p(x)$ claramente depende de parâmetros desconhecidos da população, β_j . No entanto, realmente temos estimadores não-viesados desses parâmetros, ou seja, os estimadores MQO. Quando os estimadores MQO são integrados na equação (8.39), obtemos os valores estimados do MQO. Assim, para cada observação i , $\text{Var}(y_i|x_i)$ é estimada por

$$\hat{h}_i = \hat{y}_i(1 - \hat{y}_i), \quad (8.40)$$

onde \hat{y}_i é o valor estimado por MQO da observação i . Agora, aplicamos o MQG factível, exatamente como na Seção 8.4.

Infelizmente, ter condição de estimar \hat{h}_i de cada i não significa que poderemos prosseguir diretamente com a estimação MQP. O problema é o que discutimos brevemente na Seção 7.5: os valores estimados \hat{y}_i não precisam cair no intervalo unitário. Se $\hat{y}_i < 0$ ou $\hat{y}_i > 1$, a equação (8.40) mostra que \hat{h}_i será negativo. Como o método MQP prossegue multiplicando a observação i por $1/\sqrt{\hat{h}_i}$, o método falhará se \hat{h}_i for negativo (ou zero) em qualquer observação. Em outras palavras, todos os pesos do método MQP devem ser positivos.

Em alguns casos, $0 < \hat{y}_i < 1$ para todos os i , quando o MQP pode ser usado para estimar o MPL. Nos casos com muitas observações e pequenas probabilidades de sucesso ou fracasso, é muito comum encontrarmos alguns valores estimados fora do intervalo unitário. Se isso acontecer, como no exemplo da participação na força de trabalho na equação (8.37), é mais fácil abandonar o método MQP e registrar as estatísticas robustas em relação à heteroscedasticidade. Uma alternativa é ajustar os valores estimados menores que zero ou maiores que a unidade, e depois aplicar o método MQP. Uma sugestão é definir $\hat{y}_i = 0,01$ se $\hat{y}_i < 0$ e $\hat{y}_i = 0,99$ se $\hat{y}_i > 1$. Infelizmente, isso exigirá uma escolha arbitrária por parte do pesquisador — por exemplo, por que não usar 0,001 e 0,999 como os valores estimados? Se muitos valores estimados estiverem fora do intervalo unitário, o ajuste pode afetar os resultados; nesta situação, provavelmente será melhor usar somente o método MQO.

ESTIMANDO O MODELO DE PROBABILIDADE LINEAR POR MÍNIMOS QUADRADOS PONDERADOS

1. Estime o modelo por MQO e obtenha os valores estimados \hat{y} .
2. Determine se todos os valores estimados estão dentro do intervalo unitário. Se assim for, prossiga para o passo (3). Caso contrário, alguns ajustes serão necessários para trazer todos os valores estimados para dentro do intervalo unitário.
3. Construa as variâncias estimadas na equação (8.40).
4. Estime a equação

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$

por MQP, usando pesos $1/\hat{h}$.

EXEMPLO 8.9

(Determinantes da Propriedade de Computadores Pessoais)

Utilizamos os dados contidos no arquivo GPA1.RAW para estimar a probabilidade de possuir um computador. PC é um indicador binário igual à unidade se o aluno possuir um computador, e zero caso contrário. A variável $nmem$ é a nota média do ensino médio, tac é a nota do teste de avaliação de conhecimentos para ingresso em curso superior e $paifac$ é um indicador binário igual a um se pelo menos um dos genitores tem curso superior. (Indicadores separados para a mãe e para o pai não produzem resultados individualmente significantes, já que são bastante correlacionados.)

A equação estimada por MQO é

$$\begin{aligned} \hat{PC} = & -0,0004 + 0,065 nmem + 0,0006 ACT + 0,221 paifac \\ (0,4905) & (0,137) \quad (0,0155) \quad (0,093) \\ [0,4888] & [0,139] \quad [0,0158] \quad [0,087] \end{aligned} \tag{8.41}$$

$n = 141, R^2 = 0,0415.$

EXEMPLO 8.9 (continuação)

Como no exemplo 8.8, não existem diferenças substanciais entre os erros-padrão usuais e os robustos. Não obstante, também estimamos o modelo por MQP. Como todos os valores estimados se encontram dentro do intervalo unitário, não são necessários ajustes:

$$\hat{PC} = 0,026 + 0,033 nmem + 0,0043 ACT + 0,215 paisfac \quad (8.42)$$

(0,477)	(0,130)	(0,0155)	(0,086)
$n = 141, R^2 = 0,0464.$			

Não existem diferenças importantes nas estimativas MQO e MQP. A única variável explicativa significante é *paisfac*, e em ambos os casos estimamos que a probabilidade de propriedade de um computador pessoal é cerca de 0,22 maior se pelo menos um dos genitores tem curso superior.

Iniciamos revendo as propriedades dos mínimos quadrados ordinários na presença de heteroscedasticidade. Esta não causa viés ou inconsistência nos estimadores MQO, mas os erros-padrão e as estatísticas de testes usuais não serão mais válidos. Mostramos como computar erros-padrão e estatísticas *t* robustos em relação à heteroscedasticidade, algo que é feito rotineiramente por muitos programas econométricos. A maioria dos programas também computa uma estatística *F* robusta em relação à heteroscedasticidade.

Discutimos duas maneiras comuns de verificar a existência de heteroscedasticidade: o teste de Breusch-Pagan e um caso especial do teste de White. Essas duas estatísticas envolvem a regressão dos resíduos *quadrados* do MQO sobre as variáveis independentes (BP) ou sobre os valores estimados e o quadrado destes (White). Um teste *F* simples é assintoticamente válido; existem também versões do multiplicador de Lagrange dos testes.

O MQO não é mais o melhor estimador linear não-viesado na presença de heteroscedasticidade. Quando a forma da heteroscedasticidade é conhecida, a estimação por mínimos quadrados generalizados (MQG) pode ser usada. Isso conduz aos mínimos quadrados ponderados como um meio de obter estimadores BLUE. As estatísticas de testes da estimação MQP são perfeitamente válidas quando o termo erro é normalmente distribuído ou assintoticamente válido sob não normalidade. Isso supõe, é claro, que temos o modelo de heteroscedasticidade apropriado.

Mais comumente, devemos estimar um modelo quanto à heteroscedasticidade antes de aplicarmos o MQP. O estimador MQG *fáctível* resultante não mais será não-viesado, mas será consistente e assintoticamente eficiente. As estatísticas habituais da regressão MQP são assintoticamente válidas. Discutimos um método para assegurar que as variâncias estimadas sejam estritamente positivas para todas as observações, o que é necessário para aplicar o método MQP.

Como discutimos no Capítulo 7, o modelo de probabilidade linear de uma variável dependente binária terá, necessariamente, um termo erro heteroscedástico. Uma maneira simples de lidar com esse problema é computar estatísticas robustas em relação à heteroscedasticidade. Alternativamente, se todos os valores estimados (isto é, as probabilidades estimadas) estiverem estritamente entre zero e um, os mínimos quadrados ponderados poderão ser usados para a obtenção de estimadores assintoticamente eficientes.

8.1 Quais das seguintes alternativas são consequências da heteroscedasticidade?

- (i) Os estimadores MQO, $\hat{\beta}_j$, são inconsistentes.
- (ii) A estatística F usual não mais tem uma distribuição F .
- (iii) Os estimadores MQO não mais são BLUE.

8.2 Considere um modelo linear para explicar o consumo mensal de cerveja:

$$\begin{aligned} \text{cerveja} &= \beta_0 + \beta_1 \text{renda} + \beta_2 \text{preço} + \beta_3 \text{educ} + \beta_4 \text{feminino} + u \\ \text{E}(u|\text{renda}, \text{preço}, \text{educ}, \text{feminino}) &= 0 \\ \text{Var}(u|\text{renda}, \text{preço}, \text{educ}, \text{feminino}) &= \sigma^2 \text{renda}^2. \end{aligned}$$

Escreva a equação transformada que tenha um termo erro homoscedástico.

8.3 Verdadeiro ou Falso: O método MQP é preferido ao MQO quando uma variável importante for omitida do modelo.**8.4** Utilizando os dados contidos no arquivo GPA3.RAW, a seguinte equação foi estimada para os alunos de uma universidade:

$$\begin{aligned} \hat{nsgrad} &= -2,12 + 0,900 npgrad + 0,193 nmgradac + 0,0014 tohrs \\ &\quad (0,55) (0,175) (0,064) (0,0012) \\ &\quad [0,55] [0,166] [0,074] [0,0012] \\ &+ 0,0018 sat - 0,0039 emperc + 0,351 feminino - 0,157 estac \\ &\quad (0,0002) (0,0018) (0,085) (0,098) \\ &\quad [0,0002] [0,0019] [0,079] [0,080] \\ n &= 269, R^2 = 0,465. \end{aligned}$$

Aqui, $nsgrad$ é a nota obtida pelo aluno no exame final do curso, no semestre corrente, $npgrad$ é uma média ponderada das notas nas diversas disciplinas cursadas no semestre, $nmgradac$ é a nota do exame de final de semestre, no semestre anterior, $tohrs$ é o total de créditos em horas, acumuladas até o semestre anterior, sat é a nota do aluno no exame de ingresso na Universidade, $emperc$ é o percentil do aluno no curso médio na escola em que o aluno se formou antes de ingressar na Universidade, $feminino$ é uma *dummy* de gênero e $estac$ é uma *dummy* igual a um se o esporte praticado pelo aluno for praticado durante o outono. Os erros-padrão usuais e os robustos em relação à heteroscedasticidade estão registrados entre parênteses e colchetes, respectivamente.

- (i) As variáveis $npgrad$, $nmgradac$ e $tohrs$ têm os efeitos estimados esperados? Quais dessas variáveis são estatisticamente significantes ao nível de 5%? Importa quais erros-padrão são usados?
- (ii) Por que a hipótese $H_0: \beta_{npgrad} = 1$ faz sentido? Teste esta hipótese contra a alternativa bicaudal ao nível de 5%, usando ambos os erros-padrão. Descreva suas conclusões.
- (iii) Verifique se existe um efeito sazonal sobre a variável $nsgrad$, usando ambos os erros-padrão. O nível de significância no qual a hipótese nula pode ser rejeitada depende do erro-padrão usado?

8.5 A variável *fuma* é uma variável binária igual a um se a pessoa fuma, e zero caso contrário. Utilizando os dados contidos no arquivo SMOKE.RAW, estimamos um modelo de probabilidade linear de *fuma*:

$$\begin{aligned}
 \hat{fuma} = & 0,656 - 0,069 \log(precig) + 0,012 \log(renda) - 0,029 educ \\
 & (0,855) \quad (0,204) \quad \quad \quad (0,026) \quad \quad \quad (0,006) \\
 & [0,856] \quad [0,207] \quad \quad \quad [0,026] \quad \quad \quad [0,006] \\
 & + 0,020 idade - 0,00026 idade^2 - 0,101 restaurn - 0,026 branco \\
 & (0,006) \quad (0,00006) \quad (0,039) \quad \quad \quad (0,052) \\
 & [0,005] \quad [0,00006] \quad [0,038] \quad \quad \quad [0,050] \\
 & n = 807, R^2 = 0,062.
 \end{aligned}$$

A variável *branco* é igual a um se a pessoa for branca, e zero caso contrário; as outras variáveis independentes foram definidas no Exemplo 8.7. Tanto os erros-padrão usuais como os robustos em relação à heteroscedasticidade estão informados.

- (i) Existe alguma diferença importante entre os dois conjuntos de erros-padrão?
- (ii) Mantendo os outros fatores fixos, se a educação aumentar em quatro anos, o que acontece com a probabilidade de fumar estimada?
- (iii) Em que ponto mais um ano de idade reduz a probabilidade de fumar?
- (iv) Interprete o coeficiente da variável binária *restaurn* (uma variável dummy igual a um se a pessoa viver em um estado em que há restrições de fumar em restaurantes).
- (v) A pessoa número 206 no conjunto de dados tem as seguintes características: *precig* = 67,44, *renda* = 6.500, *educ* = 16, *idade* = 77, *restaurn* = 0, *branco* = 0 e *fuma* = 0. Calcule a probabilidade de fumar dessa pessoa e comente o resultado.

Problemas Adicionais de Especificação e de Dados

o Capítulo 8 estudamos uma violação das hipóteses de Gauss-Markov. A heteroscedasticidade nos erros pode ser vista como uma má especificação de modelo, porém é um problema de menor importância. A presença de heteroscedasticidade não causa viés ou inconsistência nos estimadores MQO. Além disso, é razoavelmente fácil ajustar intervalos de confiança e estatísticas t e F para obter inferência válida após a estimação MQO, ou mesmo para obter estimadores mais eficientes com o uso de mínimos quadrados ponderados.

Neste capítulo, retornamos a um problema muito mais sério da correlação entre o erro, u , e uma ou mais das variáveis explicativas. Lembre-se do Capítulo 3 em que, se u for, por qualquer razão, correlacionado com a variável explicativa x_j , então dizemos que x_j é uma **variável explicativa endógena**. Também fazemos uma discussão mais detalhada sobre três razões pelas quais uma variável explicativa pode ser endógena; em alguns casos discutimos possíveis correções.

Já vimos nos Capítulos 3 e 5 que a omissão de uma variável importante pode causar correlação entre o erro e algumas das variáveis explicativas, o que geralmente conduz a viés e inconsistência em *todos* os estimadores MQO. No caso especial em que a variável omitida é uma função de uma variável explicativa no modelo, este sofrerá de **má especificação da forma funcional**.

Iniciamos a primeira seção discutindo as consequências da má especificação da forma funcional e como testar sua existência. Na Seção 9.2, mostramos como o uso de variáveis *proxy* pode resolver, ou pelo menos aliviar, o viés de variáveis omitidas. Na Seção 9.3, derivamos e explicamos o viés no método MQO que pode aparecer sob certas formas de **erro de medida**. Problemas adicionais de dados são discutidos na Seção 9.4.

Todos os procedimentos descritos neste capítulo são baseados na estimação MQO. Como vemos, certos problemas que causam correlação entre o erro e algumas variáveis explicativas não podem ser resolvidos usando MQO em estudos de corte transversal. Postergamos uma abordagem sobre métodos de estimação alternativos para a Parte 3.

9.1 MÁ ESPECIFICAÇÃO DA FORMA FUNCIONAL

Um modelo de regressão múltipla sofre de má especificação da forma funcional quando não explica de maneira apropriada a relação entre as variáveis explicativas e a dependente observadas. Por exemplo, se o salário por hora é determinado por $\log(\text{salárioh}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + u$, mas omitimos o termo de experiência elevado ao quadrado, exper^2 , então estamos cometendo uma má especificação da forma funcional. Já sabemos, como vimos no Capítulo 3, que isso geralmente conduz a estimadores viesados de β_0 , β_1 e β_2 . (Não estimamos β_3 porque o termo exper^2 foi excluído do modelo).

Desse modo, especificando incorretamente como exper afeta $\log(\text{salárioh})$ geralmente resulta em um estimador viesado do retorno da educação, β_1 . A magnitude desse viés depende do tamanho de β_2 e da correlação entre educ , exper e exper^2 .

A situação é pior ao estimar o retorno da experiência: mesmo que pudéssemos conseguir um estimador não-viesado de β_2 , não seríamos capazes de estimar o retorno da experiência, pois ela é equivalente a $\beta_2 + 2\beta_3\text{exper}$ (em forma decimal). Usar apenas o estimador viesado de β_2 pode ser enganoso, especialmente nos valores extremos de exper .

Como outro exemplo, suponha que a equação $\log(\text{salárioh})$ seja

$$\begin{aligned}\log(\text{salárioh}) = & \beta_0 + \beta_1\text{educ} + \beta_2\text{exper} + \beta_3\text{exper}^2 \\ & + \beta_4\text{feminino} + \beta_5\text{feminino} \cdot \text{educ} + u,\end{aligned}\tag{9.1}$$

onde feminino é uma variável binária. Se omitirmos o termo de interação, $\text{feminino} \cdot \text{educ}$, estaremos especificando mal a forma funcional. De maneira geral, não obteremos estimadores não-viesados de nenhum dos outros parâmetros, e como o retorno da educação depende do gênero, não fica claro que tipo de retorno estariamos estimando quando omitimos o termo de interação.

A omissão de funções de variáveis independentes não é a única maneira de um modelo sofrer o problema da má especificação da forma funcional. Por exemplo, se (9.1) for o modelo verdadeiro para satisfazer as primeiras quatro hipóteses de Gauss-Markov, mas utilizarmos salárioh , em lugar de $\log(\text{salárioh})$, como variável dependente, não obteremos estimadores não-viesados ou consistentes dos efeitos parciais. Os testes a seguir têm certa capacidade de detectar esse tipo de problema da forma funcional, mas existem testes melhores que serão mencionados nas subseções de testes contra alternativas não-aninhadas.

A má especificação da forma funcional de um modelo pode, certamente, trazer sérias consequências. No entanto, em um aspecto importante, o problema é secundário: por definição, temos dados de todas as variáveis necessárias para obter uma relação funcional que se ajuste bem aos dados. Isso pode ser contrastado com o problema tratado na próxima seção, na qual uma variável importante será omitida e sobre a qual não poderemos coletar dados.

Já temos uma ferramenta poderosa para detectar uma forma funcional mal-especificada: o teste F para restrições de exclusões conjuntas. Muitas vezes faz sentido adicionar termos quadráticos de quaisquer variáveis significantes a um modelo e executar um teste conjunto de significância. Se os termos quadráticos adicionados forem significantes, eles podem ser adicionados ao modelo (ao custo de complicar sua interpretação). Porém, termos quadráticos significantes podem ser sintomáticos de outros problemas de formas funcionais, como, por exemplo, usar uma variável em nível quando o logaritmo é mais apropriado, ou vice-versa. Pode ser difícil localizar a razão exata pela qual uma forma funcional está mal-especificada. Felizmente, em muitos casos, o uso de logaritmos de certas variáveis e a adição de termos quadráticos são suficientes para detectar muitas relações não-lineares importantes em economia.

EXEMPLO 9.1**[Modelo Econômico do Crime]**

A Tabela 9.1 contém estimativas MQO do modelo econômico do crime (veja Exemplo 8.3). Primeiro estimamos o modelo sem nenhum termo quadrático; os resultados estão na coluna (1).

Tabela 9.1

Variável Dependente: $npre86$

Variáveis Independentes	(1)	(2)
$pcond$	-0,133 (0,040)	0,533 (0,154)
$pcond^2$	—	-0,730 (0,156)
$sentmed$	-0,011 (0,012)	-0,017 (0,012)
$temptot$	0,012 (0,009)	0,012 (0,009)
$ptemp86$	-0,041 (0,009)	0,287 (0,004)
$ptemp86^2$	—	-0,0296 (0,0039)
$empr86$	-0,051 (0,014)	-0,014 (0,017)
$rend86$	-0,0015 (0,0003)	-0,0034 (0,0008)
$rend86^2$	—	0,000007 (0,000003)
$negro$	0,327 (0,045)	0,292 (0,045)
$hispan$	0,194 (0,040)	0,164 (0,039)
$intercepto$	0,596 (0,036)	0,505 (0,037)
Observações	2725	2725
R-quadrado	0,0723	0,1035

Na coluna (2) os quadrados de $pcond$, $ptemp86$ e $rend86$ foram adicionados; decidimos incluir os quadrados dessas variáveis porque cada termo em nível é significante na coluna (1). A variável $empr86$ é uma variável discreta, ao assumirmos somente cinco valores, de modo que não incluímos seu quadrado na coluna (2).

EXEMPLO 9.1 (continuação)

Cada um dos termos quadráticos é significante e simultaneamente eles são de maneira conjunta muito significantes ($F = 31,37$, com $g_f = 3$ e 2713 ; o p -valor é basicamente zero). Portanto, parece que o modelo inicial deixou de fora algumas não-linearidades potencialmente importantes.

A presença dos termos quadráticos faz com que a interpretação do modelo seja um pouco difícil. Por exemplo, $pcond$ não tem mais um efeito estritamente de dissusão: a relação entre $npre86$ e $pcond$ é positiva até $pcond = 0,365$ e, a partir daí, a relação é negativa. Podemos concluir que existe pouco ou nenhum efeito de dissusão em valores mais baixos de $pcond$; o efeito somente aparece com taxas de condenações anteriores mais altas. Teríamos que utilizar formas funcionais mais sofisticadas do que a quadrática para verificar essa conclusão. Pode ser que $pcond$ não seja inteiramente exógena. Por exemplo, pessoas que não tenham sido condenadas no passado (de modo que $pcond = 0$) são, talvez, criminosos acidentais, e, portanto, é menos provável que tenham sido presos em 1986. Isto poderia estar causando um viés nas estimativas.

Similarmente, a relação entre $npre86$ e $ptemp86$ é positiva até $ptemp86 = 4,85$ (quase cinco meses na prisão), e a partir daí é negativa. A maioria das pessoas na amostra não passou nenhum tempo na prisão em 1986, de modo que, novamente, devemos ser cuidadosos ao interpretar os resultados.

A renda legal tem um efeito negativo sobre $npre86$ até $rend86 = 242,85$; como a renda é medida em centenas de dólares, isso representa uma renda anual de 24.285 dólares. Somente 46 das pessoas na amostra têm rendimentos acima deste nível. Portanto, podemos concluir que $npre86$ e $rend86$ são negativamente relacionadas, com um efeito decrescente.

Por que não incluímos os quadrados de *negro* e *hispan* na coluna (2) da Tabela 9.1?

O exemplo 9.1 é um problema delicado de forma funcional, devido à natureza da variável dependente. Outros modelos que, teoricamente, são mais apropriados para lidar com variáveis dependentes empregam um pequeno número de valores inteiros. Discutiremos resumidamente esses modelos no Capítulo 17.

O Teste RESET como um Teste Geral da Má Especificação da Forma Funcional

Alguns testes têm sido propostos para detectar a má especificação da forma funcional. O teste de erro de especificação da regressão (RESET) de Ramsey (1969) tem se mostrado útil a esse respeito.

A idéia por trás do teste RESET é bastante simples. Se o modelo original

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u \quad (9.2)$$

satisfizer RLM.3, nenhuma função não-linear das variáveis independentes deve ser significante quando adicionada à equação (9.2). No Exemplo 9.1 adicionamos termos quadráticos às variáveis explicativas significantes. Embora isso muitas vezes detecte problemas de forma funcional, tem a desvantagem de gastar muitos graus de liberdade se houver muitas variáveis explicativas no modelo original (tanto quanto a forma direta do teste de White da heteroscedasticidade consome graus de liberdade). Além

disso, certos tipos de não-linearidades negligenciadas não serão detectadas pela adição de termos quadráticos. O teste RESET adiciona polinômios aos valores estimados MQO na equação (9.2) para detectar tipos gerais de má especificação de formas funcionais.

Para implementar o teste RESET, temos que decidir quantas funções dos valores estimados devem ser incluídas na regressão expandida. Não existe uma resposta certa para esta questão, mas os termos quadráticos e cúbicos têm demonstrado utilidade na maior parte das aplicações.

Sejam \hat{y} os valores estimados MQO ao estimar (9.2). Considere a equação expandida

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + erro. \quad (9.3)$$

Esta equação parece um tanto estranha, pois funções dos valores estimados na estimação inicial agora aparecem como variáveis explicativas. Na realidade, não estaremos interessados nos parâmetros estimados em (9.3); apenas usamos esta equação para testar se (9.2) tem não-linearidades importantes ausentes. O que devemos lembrar é que \hat{y}^2 e \hat{y}^3 são apenas funções não-lineares de x_j .

A hipótese nula é que (9.2) está corretamente especificada. Portanto, a estatística do teste RESET é a estatística F para testar $H_0: \delta_1 = 0, \delta_2 = 0$ no modelo expandido (9.3). Uma estatística F significante sugere algum tipo de problema na forma funcional. A distribuição da estatística F é, aproximadamente, $F_{2, n-k-3}$ em amostras grandes sob a hipótese nula (e sob as hipóteses de Gauss-Markov). Os gl na equação expandida (9.3) são $n - k - 1 - 2 = n - k - 3$. Uma versão LM também está disponível (e a distribuição qui-quadrado terá dois gl). Além disso, o teste pode ser transformado em robusto em relação à heteroscedasticidade utilizando os métodos discutidos na Seção 8.2.

EXEMPLO 9.2

(Equação dos Preços de Imóveis)

Estimamos dois modelos para os preços de imóveis. O primeiro tem todas as variáveis em forma de nível:

$$preço = \beta_0 + \beta_1 tamterr + \beta_2 arquad + \beta_3 qtdorm + u. \quad (9.4)$$

O segundo utiliza os logaritmos de todas as variáveis, exceto $qtdorm$:

$$lpreço = \beta_0 + \beta_1 ltamterr + \beta_2 larquad + \beta_3 qtdorm + u. \quad (9.5)$$

Usando $n = 88$ imóveis do arquivo HPRICE.RAW, constata-se que a estatística do teste RESET da equação (9.4) é 4,67; este é o valor de uma variável aleatória $F_{2,82}$ ($n = 88, k = 3$), e o p -valor associado é 0,012. Isto é uma evidência de má especificação da forma funcional em (9.4).

A estatística do teste RESET em (9.5) é 2,56, com p -valor = 0,084. Portanto, não rejeitamos (9.5) no nível de significância de 5% (embora o faríamos ao nível de 10%). Com base no teste RESET, o modelo log-log em (9.5) é preferido.

No exemplo anterior, tentamos dois modelos para explicar os preços de imóveis. Um foi rejeitado pelo teste RESET, enquanto o outro não o foi (pelo menos ao nível de 5%). Muitas vezes, as coisas

não são tão simples. Uma desvantagem do teste RESET é que ele não fornece uma orientação prática de como proceder se o modelo for rejeitado. A rejeição de (9.4), pelo uso do teste RESET não sugere diretamente que (9.5) seja o passo seguinte. A equação (9.5) foi estimada porque modelos de elasticidade constante são fáceis de serem interpretados e podem apresentar boas propriedades estatísticas. Neste exemplo, o modelo também passa no teste da forma funcional.

Algumas pessoas argumentaram que o teste RESET é demasiadamente generalizado da má-especificação de modelos, incluindo variáveis omitidas não-observadas e heteroscedasticidade. Infelizmente, tal uso do teste é bastante equívoco. Pode ser demonstrado que o teste RESET não tem poder para detectar variáveis omitidas sempre que houver expectativa de que elas sejam lineares nas variáveis independentes incluídas no modelo [veja Wooldridge (1995) para um enunciado preciso]. Além disso, se a forma funcional for apropriadamente especificada, o teste RESET não tem poder para detectar heteroscedasticidade. O ponto principal é que o teste RESET é um teste da forma funcional, e nada mais que isso.

Testes Contra Alternativas Não-Aninhadas

Obter testes para outros tipos de má especificação da forma funcional — por exemplo, tentar decidir se uma variável independente deveria aparecer em nível ou em forma logarítmica — nos leva para fora do âmbito dos testes de hipótese clássicos. É possível testar o modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \quad (9.6)$$

contra o modelo

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + u, \quad (9.7)$$

e vice-versa. Porém, esses são **modelos não-aninhados** (veja Capítulo 6), e portanto não podemos simplesmente usar um teste *F* padrão. Dois métodos diferentes podem ser sugeridos. O primeiro é construir um modelo abrangente que contenha cada modelo como um caso especial e, em seguida, testar as restrições que conduziram a cada um dos modelos. No exemplo, o modelo abrangente é

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 \log(x_1) + \gamma_4 \log(x_2) + u. \quad (9.8)$$

Podemos primeiro testar $H_0: \gamma_3 = 0, \gamma_4 = 0$, como um teste de (9.6). Podemos também testar $H_0: \gamma_1 = 0, \gamma_2 = 0$ como um teste de (9.7). Esta abordagem foi sugerida por Mizon e Richard (1986).

Outro método foi sugerido por Davidson e MacKinnon (1981). Eles salientam que, se (9.6) for verdadeira, então os valores estimados do modelo (9.7) deveriam ser não significantes em (9.6). Assim, para testar (9.6), primeiro estimamos o modelo (9.7) por MQO para obtermos os valores estimados. Chamemos esses valores de \hat{y} . Então, o **teste de Davidson-MacKinnon** baseia-se na estatística *t* sobre \hat{y} na equação

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{y} + erro.$$

Uma estatística *t* significante (contra uma alternativa bicaudal) é uma rejeição de (9.6).

Similarmente, se \hat{y} representar os valores estimados da estimação de (9.6), o teste de (9.7) é a estatística t sobre \hat{y} no modelo

$$y = \beta_0 + \beta_1 \log(x_1) + \beta_2 \log(x_2) + \theta_1 \hat{y} + \text{erro};$$

uma estatística t significante é evidência contra (9.7). Os mesmos dois testes podem ser usados para testar quaisquer dois modelos não-aninhados com a mesma variável dependente.

Existem alguns problemas com testes não-aninhados. Primeiro, não necessariamente, um dos modelos será claramente o escolhido. Ambos os modelos, ou nenhum deles, podem ser rejeitados. Se nenhum deles for rejeitado, podemos usar o R -quadrado ajustado para selecionar um deles. Se ambos os modelos forem rejeitados, teremos mais trabalho. Porém, é importante sabermos das consequências práticas ao utilizarmos cada um deles: se os efeitos de importantes variáveis independentes sobre y não forem muito diferentes, então não importa qual dos modelos será usado.

Um segundo problema é que a rejeição de (9.6) pela utilização, digamos, do teste de Davidson-MacKinnon, não significa que (9.7) seja o modelo correto. O modelo (9.6) pode ser rejeitado por uma diversidade de más especificações da forma funcional.

Um problema ainda mais difícil é obter testes não-aninhados quando os modelos concorrentes têm variáveis dependentes diferentes. O caso principal é y versus $\log(y)$. Vimos no Capítulo 6 que apenas a obtenção de medidas de graus de ajustes que possam ser comparadas necessita de algum cuidado. Alguns testes foram propostos para resolver este problema, mas estão além do escopo deste texto. [Veja Wooldridge (1994a) para um teste que tem uma interpretação simples e é fácil de ser implementado.]

9.2 UTILIZANDO VARIÁVEIS PROXY PARA VARIÁVEIS EXPLICATIVAS NÃO-OBSERVADAS

Um problema mais difícil surge quando um modelo exclui uma variável importante, normalmente devendo à não-disponibilidade de dados. Considere uma equação de salários que explicitamente reconheça que a aptidão (*aptid*) afeta $\log(\text{salário})$:

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{aptid} + u. \quad (9.9)$$

Este modelo mostra explicitamente que queremos manter fixa a aptidão quando medimos os retornos de *educ* e de *exper*. Se, digamos, *educ* for correlacionada com *aptid*, colocar *aptid* no termo erro fará com que o estimador MQO de β_1 (e β_2) sejam viesados, um tema que tem aparecido repetidamente.

Nosso interesse primordial na equação (9.9) está nos parâmetros de inclinação β_1 e β_2 . Realmente, não nos interessa se obteremos um estimador não-viesado ou consistente do intercepto β_0 ; como veremos em breve, normalmente isso não é possível. Tampouco podemos esperar estimar β_3 , pois *aptid* não é observada; na verdade, de qualquer forma, não saberíamos como interpretar β_3 , pois a aptidão é, na melhor das hipóteses, um conceito vago.

Como podemos resolver, ou pelo menos aliviar, o problema do viés de variáveis omitidas em uma equação como (9.9)? Uma possibilidade é obter uma variável *proxy* da variável omitida. *Grosso modo*, uma variável *proxy* é algo que está relacionado com a variável não-observada que gostaríamos de controlar em nossa análise. Na equação do salário, uma possibilidade é usar o quociente de inteligência, ou QI, como uma *proxy* da aptidão. Isso não exige que QI seja a mesma coisa que aptidão; o que precisamos é que QI seja correlacionado com aptidão, o que esclareceremos na discussão a seguir.

Todas essas idéias podem ser ilustradas em um modelo com três variáveis independentes, duas das quais são observadas:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3^* + u. \quad (9.10)$$

Assumimos que os dados estão disponíveis para y , x_1 e x_2 — no exemplo do salário, eles são *log(salário)*, *educ* e *exper*, respectivamente. A variável explicativa x_3^* é não-observada, mas temos uma variável proxy de x_3^* . Chamemos essa variável proxy de x_3 .

O que necessitamos de x_3 ? No mínimo, ela deve ter alguma relação com x_3^* . Isso é capturado pela equação de regressão simples

$$x_3^* = \delta_0 + \delta_3 x_3 + v_3, \quad (9.11)$$

onde v_3 é um erro pelo fato de que x_3^* e x_3 não são exatamente relacionadas. O parâmetro δ_3 mede a relação entre x_3^* e x_3 ; em geral, pensamos em x_3^* e x_3 como positivamente relacionadas, de forma que $\delta_3 > 0$. Se $\delta_3 = 0$, então x_3 não é uma proxy adequada de x_3^* . O intercepto δ_0 em (9.11), que pode ser positivo ou negativo, simplesmente permite que x_3^* e x_3 sejam medidas em diferentes escalas. (Por exemplo, certamente não é necessário que a aptidão não-observada tenha o mesmo valor médio do QI na população dos Estados Unidos.)

Como podemos usar x_3 para obtermos estimadores não-viesados (ou pelo menos consistentes) de β_1 e β_2 ? A proposta é simular que x_3 e x_3^* sejam as mesmas, de forma que possamos computar a regressão de

$$y \text{ sobre } x_1, x_2, x_3. \quad (9.12)$$

Chamamos a isso de **solução plugada do problema de variáveis omitidas** porque a variável x_3 está “plugada” em x_3^* antes de executarmos o MQO. Se x_3 for verdadeiramente relacionada com x_3^* , isso parece sensato. Porém, como x_3 e x_3^* não são as mesmas, devemos determinar quando esse procedimento produzirá, de fato, estimadores consistentes de β_1 e β_2 .

As hipóteses necessárias para que a solução plugada forneça estimadores consistentes de β_1 e β_2 podem ser decompostas em hipóteses sobre u e v_3 :

- 1) O erro u é não-correlacionado com x_1 , x_2 e x_3^* , que justamente é a hipótese padrão no modelo (9.10). Além disso, u é não-correlacionado com x_3 . Esta última hipótese significa exatamente que x_3 é irrelevante no modelo populacional, já que as variáveis x_1 , x_2 e x_3^* foram incluídas. Isso é basicamente verdadeiro por definição, visto que x_3 é uma variável proxy de x_3^* : é x_3^* que diretamente afeta y , não x_3 . Assim, a hipótese de que u é não-correlacionada com x_1 , x_2 , x_3^* e x_3 não é muito controversa. (Outra maneira de explicar essa hipótese é que o valor esperado de u , dadas todas essas variáveis, é zero.)
- 2) O erro v_3 é não-correlacionado com x_1 , x_2 e x_3 . Supor que v_3 é não-correlacionado com x_1 e x_2 exige que x_3 seja uma “boa” proxy de x_3^* . Pode-se ver isso de maneira mais fácil, escrevendo-se a expressão análoga dessas hipóteses em termos de expectativas condicionais:

$$E(x_3^*|x_1, x_2, x_3) = E(x_3^*|x_3) = \delta_0 + \delta_3 x_3. \quad (9.13)$$

A primeira igualdade, que é a mais importante, diz que, como x_3 é controlada, o valor esperado de x_3^* não depende de x_1 ou de x_2 . Alternativamente, x_3^* tem correlação zero com x_1 e com x_2 , dado que x_3 é arbitrária.

Na equação de salários (9.9), onde QI é a proxy da aptidão, a condição (9.13) torna-se

$$E(aptid|educ,exper,QI) = E(aptid|QI) = \delta_0 + \delta_3 QI.$$

Assim, a média de aptidão somente muda com QI , não com $educ$ e $exper$. Isso é razoável? Talvez não seja exatamente verdade, mas está perto de ser. Certamente, vale a pena incluir QI na equação de salários para vermos o que acontece com o retorno estimado da variável educação.

Podemos facilmente verificar a razão de as hipóteses anteriores serem suficientes para que a solução plugada funcione. Se integrarmos a equação (9.11) na equação (9.10) e aplicarmos álgebra simples teremos

$$y = (\beta_0 + \beta_3 \delta_0) + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \delta_3 x_3 + u + \beta_3 v_3.$$

Chamemos o erro composto nesta equação de $e = u + \beta_3 v_3$; ele depende do erro no modelo de interesse (9.10) e do erro na equação da variável proxy, v_3 . Como tanto u quanto v_3 têm média zero e ambos são não-correlacionados com x_1 , x_2 e x_3 , o erro e também tem média zero e é não-correlacionado com x_1 , x_2 e x_3 . Escreva esta equação como

$$y = \alpha_0 + \beta_1 x_1 + \beta_2 x_2 + \alpha_3 x_3 + e,$$

onde $\alpha_0 = (\beta_0 + \beta_3 \delta_0)$ é o novo intercepto e $\alpha_3 = \beta_3 \delta_3$ é o parâmetro de inclinação da variável proxy x_3 . Como mencionamos anteriormente, quando computarmos a regressão em (9.12), não obteremos estimadores não-viesados de β_0 e β_3 ; em vez disso, obteremos estimadores não-viesados (ou pelo menos consistentes) de α_0 , β_1 , β_2 e α_3 . O importante é que obtenhamos boas estimativas dos parâmetros β_1 e β_2 .

Em muitos casos, a estimativa de α_3 é efetivamente mais interessante do que uma estimativa de β_3 . Por exemplo, na equação de salários, α_3 mede o retorno do salário se um ou mais pontos forem atribuídos à pontuação do QI.

EXEMPLO 9.3

(QI como Proxy de Aptidão)

O arquivo WAGE2.RAW, de Blackburn e Neumark (1992), contém informações sobre renda mensal, educação, diversas variáveis demográficas e pontuação de QI para 935 homens, em 1980. Como um método para explicar o viés da variável omitida aptidão, adicionamos QI a uma equação de log salário padrão. Os resultados estão mostrados na Tabela 9.2.

Nosso principal interesse está no que acontece com o retorno estimado da educação. A coluna (1) contém as estimativas sem a utilização de QI como uma variável proxy. O retorno estimado da educação é de 6,5%. Se imaginarmos que a aptidão omitida é positivamente correlacionada com $educ$, admitimos que esta estimativa é alta demais. (Mais precisamente, as estimativas médias de todas as amostras aleatórias seriam altas demais.) Quando QI é adicionada à equação, o retorno da educação cai para 5,4%, o que corresponde a nossa opinião anterior sobre o viés de omitir a variável aptidão.

EXEMPLO 9.3 (continuação)**Tabela 9.2**Variável Dependente: $\log(\text{salário})$

Variáveis Independentes	(1)	(2)	(3)
<i>educ</i>	0,065 (0,006)	0,054 (0,007)	0,018 (0,041)
<i>exper</i>	0,014 (0,003)	0,014 (0,003)	0,014 (0,003)
<i>perm</i>	0,012 (0,002)	0,011 (0,002)	0,011 (0,002)
<i>casado</i>	0,199 (0,039)	0,200 (0,039)	0,201 (0,039)
<i>sul</i>	-0,091 (0,026)	-0,080 (0,026)	-0,080 (0,026)
<i>urban</i>	0,184 (0,027)	0,182 (0,027)	0,184 (0,027)
<i>negro</i>	-0,188 (0,038)	-0,143 (0,039)	-0,147 (0,040)
<i>QI</i>	—	-0,0036 (0,0010)	-0,0009 (0,0052)
<i>educ · QI</i>	—	—	0,00034 (0,00038)
<i>intercepto</i>	5,395 (0,113)	5,176 (0,128)	5,648 (0,546)
Observações	935	935	935
R-quadrado	0,253	0,263	0,263

O efeito do QI nos resultados socioeconômicos foi recentemente documentado no controverso livro *The Bell Curve* (A curva em forma de sino), de Herrnstein e Murray (1994). A coluna (2) mostra que a variável QI tem um efeito estatisticamente significante e positivo sobre a renda, após vários outros fatores terem sido controlados. Todos os outros fatores permanecendo inalterados, um aumento de dez pontos no QI aumenta a renda em 3,6%. O desvio-padrão do QI na população dos Estados Unidos é 15, de modo que um aumento de desvio-padrão no QI está associado a uma elevação na renda de 5,4%. Essa elevação é idêntica ao aumento previsto em salário motivado por um ano a mais de educação. Fica claro, a partir da coluna (2),

EXEMPLO 9.3 (continuação)

que a educação ainda tem um papel importante no aumento da renda, embora o efeito não seja tão grande quanto originalmente se estimava.

Outras observações interessantes surgem do exame das colunas (1) e (2). Adicionar a variável QI à equação somente aumenta o R -quadrado de 0,253 para 0,263. A maior parte da variação em $\log(salário)$ não é explicada pelos fatores da coluna (2). Além disso, a adição de QI à equação não elimina a diferença da renda estimada entre negros e brancos: estima-se que um negro, com o mesmo QI , educação, experiência etc. de um branco ganhe cerca de 14,3% a menos, e essa diferença é estatisticamente bastante significante.

A coluna (3) na Tabela 9.2 inclui o termo de interação $educ \cdot QI$. Ele possibilita que $educ$ e $aptid$ interajam na determinação de $\log(salário)$. Podemos pensar que o retorno da educação seja mais alto para pessoas com mais aptidão mas este acaba não sendo o caso: o termo de interação não é significante e sua adição torna $educ$ e QI individualmente não significantes, além de complicar o modelo. Portanto, as estimativas da coluna (2) são preferidas.

Não existe razão para usarmos somente uma variável *proxy* da aptidão neste exemplo. O conjunto de dados do arquivo WAGE2.RAW também contém registros da pontuação de cada pessoa no teste KWW — *Knowledge of the World of Work* (Conhecimento do Mundo do Trabalho)*. Essa pontuação produz uma medida diferente da aptidão, que possa ser usada isoladamente ou em conjunto com o QI para estimar o retorno da educação (veja o Exercício 9.7).

O que é possível concluir do pequeno e estatisticamente não significante coeficiente da variável $educ$ na coluna (3) da Tabela 9.2? (Sugestão: Quando $educ \cdot QI$ está na equação, qual é a interpretação do coeficiente de $educ$?)

É fácil observar como o uso de uma variável *proxy* ainda pode conduzir a viés, se ela não satisfizer as hipóteses precedentes. Suponha que, em lugar de (9.11), a variável não-observada, x_3^* , seja relacionada com todas as variáveis observadas por

$$x_3^* = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + v_3, \quad (9.14)$$

onde v_3 tem média zero e é não-correlacionada com x_1 , x_2 e x_3 . A equação (9.11) assume que δ_1 e δ_2 são ambos zero. Plugando a equação (9.14) a (9.10), obtemos

$$y = (\beta_0 + \beta_3 \delta_0) + (\beta_1 + \beta_3 \delta_1)x_1 + (\beta_2 + \beta_3 \delta_2)x_2 + \beta_3 \delta_3 x_3 + u + \beta_3 v_3, \quad (9.15)$$

da qual segue que $\text{plim}(\hat{\beta}_1) = \beta_1 + \beta_3 \delta_1$ e $\text{plim}(\hat{\beta}_2) = \beta_2 + \beta_3 \delta_2$. [Isso acontece porque o erro em (9.15), $u + \beta_3 v_3$, tem média zero e é não-correlacionado com x_1 , x_2 e x_3 .] No exemplo anterior, no qual

* NRT: Pesquisa patrocinada e dirigida pelo *Bureau of Labor Statistics* (Agência de Estatísticas do Trabalho) do U.S. *Department of Labor* (Departamento de Trabalho dos Estados Unidos).

$x_1 = \text{educ}$ e $x_3^* = \text{aptid}$, $\beta_3 > 0$, de modo que existe um viés positivo (inconsistência), se aptid tem uma correlação parcial com educ ($\delta_1 > 0$). Desse modo, ainda poderíamos continuar obtendo um viés para cima no retorno da educação, utilizando QI como proxy de aptid , se QI não for uma boa proxy. Porém, podemos ter alguma esperança de que esse viés será menor do que se ignorarmos totalmente o problema da aptidão omitida.

Variáveis proxy também podem aparecer na forma de informação binária. No exemplo 7.9 [veja equação (7.15)], discutimos as estimativas de Krueger (1993) do retorno do uso de computador no trabalho. Krueger também incluiu uma variável binária indicando se o trabalhador utiliza um computador em casa (como também um termo de interação entre a utilização de computador no trabalho e em casa). Sua principal razão para incluir a utilização de computador em casa na equação foi a de substituir “aptidão técnica” não-observada que pudesse afetar diretamente o salário e estar relacionada com a utilização de computador no trabalho.

O Uso de Variáveis Dependentes Defasadas como Variáveis Proxy

Em algumas aplicações, como no exemplo anterior dos salários, temos pelo menos uma vaga idéia de qual fator não-observado gostaríamos de controlar. Isto facilita a escolha das variáveis proxy. Em outras aplicações, suspeitamos que uma ou mais variáveis independentes seja correlacionada com uma variável omitida, mas não temos a menor idéia de como obter uma proxy para a variável omitida. Nesses casos, podemos incluir, como um controle, o valor da variável dependente de um período anterior. Isso é especialmente útil para a análise de políticas públicas.

O uso de uma variável dependente defasada em uma equação de corte transversal aumenta os requisitos de dados, mas também fornece uma maneira simples de explicar fatores históricos que causam diferenças correntes na variável dependente que são difíceis de explicar de outras maneiras. Por exemplo, algumas cidades apresentaram altas taxas de criminalidade no passado. Muitos dos mesmos fatores não-observados contribuem para as taxas de criminalidade tanto atuais como passadas. Da mesma forma, algumas universidades são, de modo tradicional, academicamente melhores que outras. Efeitos inertiais também são capturados trabalhando-se com defasagens de y .

Considere uma equação simples para explicar as taxas de criminalidade de uma cidade:

$$\text{crime} = \beta_0 + \beta_1 \text{desemp} + \beta_2 \text{gasto} + \beta_3 \text{crime}_{-1} + u, \quad (9.16)$$

onde crime é uma medida do crime *per capita*, desemp é a taxa de desemprego da cidade, gasto é o dispêndio *per capita* para a imposição da lei e crime_{-1} indica a taxa de criminalidade medida em algum ano anterior (que poderá ser o ano passado ou vários anos atrás). Estamos interessados nos efeitos de desemp sobre crime , como também nos efeitos dos dispêndios com a imposição da lei sobre a variável crime.

Qual o propósito de se incluir crime_{-1} na equação? Certamente, esperamos que $\beta_3 > 0$, já que crime possui inércia. Entretanto, a principal razão para colocá-la na equação é o fato de que cidades com taxas históricas elevadas de criminalidade devem gastar mais com a prevenção do crime. Assim, fatores não-observados para nós (econometristas) que afetem crime são propensos a estarem correlacionados com gasto (e desemp). Se usarmos uma análise de corte transversal pura é improvável que obtenhamos um estimador não-viesado do efeito causal dos dispêndios com a imposição da lei sobre o crime. Porém, ao incluirmos crime_{-1} na equação, podemos, no mínimo, fazer a seguinte experiência: se duas cidades têm as mesmas taxas, anterior de criminalidade e atual de desemprego, então β_2 mede o efeito do gasto de mais um dólar com a imposição da lei sobre o crime.

EXEMPLO 9.4**(Taxes de Criminalidade em Cidades)**

Estimamos uma versão de elasticidade constante do modelo do crime na equação (9.16) ($desemp$, por ser uma percentagem, é deixada em forma de nível). Os dados do arquivo CRIME2.RAW são de 46 cidades para o ano de 1987. A taxa de criminalidade também está disponível para 1982, e a utilizamos como variável independente adicional na tentativa de controlar as variáveis não-observáveis das cidades que afetem o crime e possam estar correlacionadas com os dispêndios atuais com a imposição da lei. A Tabela 9.3 contém os resultados.

Tabela 9.3Variável Dependente: $\log(txcrim_{87})$

Variáveis Independentes	(1)	(2)
$desemp_{87}$	-0,029 (0,032)	0,009 (0,020)
$\log(disppclei_{87})$	0,203 (0,173)	-0,140 (0,109)
$\log(txcrim_{82})$	—	1,194 (0,132)
<i>intercepto</i>	3,34 (1,25)	0,076 (0,821)
Observações	46	46
R-quadrado	0,057	0,680

Sem a taxa de criminalidade defasada na equação, os efeitos da taxa de desemprego e dos dispêndios com a imposição da lei não são intuitivos; nenhum dos coeficientes é estatisticamente significante, embora a estatística t da variável $\log(disppclei_{87})$ seja 1,17. É possível que o aumento dos dispêndios com a imposição da lei melhore a burocracia dos registros, e assim mais crimes serão informados. Entretanto, também é provável que cidades com taxas elevadas de criminalidade recentes gastem mais com a imposição da lei.

A adição do log da taxa de criminalidade de cinco anos atrás produz um grande efeito no coeficiente dos dispêndios. A elasticidade da taxa de criminalidade em relação aos dispêndios passa a ser de -0,14 com $t = -1,28$. Isso não é muito significante, mas sugere que um modelo mais sofisticado com mais cidades na amostra poderia produzir resultados melhores.

Não surpreende que as taxas atuais de criminalidade sejam fortemente relacionadas às taxas passadas. A estimativa indica que se a taxa de criminalidade em 1982 fosse 1% mais alta, então a taxa de criminalidade prevista de 1987 seria cerca de 1,19% mais alta. Não podemos rejeitar a hipótese de que a elasticidade da criminalidade corrente em relação à criminalidade passada seja unitária [$t = (1,194 - 1)/0,132 \approx 1,47$]. A adição da taxa de criminalidade passada aumenta o poder explicativo da regressão de maneira marcante, mas isso não surpreende. A razão principal para incluir a taxa de criminalidade defasada é a obtenção de uma melhor estimativa do efeito *ceteris paribus* de $\log(disppclei_{87})$ sobre $\log(txcrim_{87})$.

A prática de usar uma variável y defasada como um método geral para controlar variáveis não-observadas está longe de ser perfeita. Porém, ela pode auxiliar na obtenção de uma melhor estimativa dos efeitos das variáveis de políticas de governo em diversos resultados.

A adição de um valor defasado de y não é a única maneira de utilizarmos dois anos de dados para controlar fatores omitidos. Quando discutirmos sobre métodos de dados de painel nos Capítulos 13 e 14, abordaremos outras maneiras sobre o uso de dados repetidos nas mesmas unidades de corte transversal em diferentes pontos no tempo.

9.3 PROPRIEDADES DO MÉTODO MQO QUANDO HÁ ERROS DE MEDIDA

Algumas vezes, em aplicações econômicas, não podemos coletar dados da variável que verdadeiramente afetam o comportamento econômico. Um bom exemplo é o ganho marginal do imposto de renda com que se defronta uma família que esteja tentando determinar quanto contribuir para instituições de caridade em determinado ano. O ganho marginal pode ser difícil de ser obtido ou resumido como um número único para todos os níveis de renda. Em vez disso, podemos calcular o ganho médio baseado na renda total e no pagamento do imposto.

Quando utilizamos uma medida imprecisa de uma variável econômica em um modelo de regressão, nosso modelo conterá um erro de medida. Nesta seção derivamos as consequências do erro de medida para a estimação dos mínimos quadrados ordinários. O método MQO será coerente sob certas hipóteses, mas existem outras sob as quais ele será incoerente. Em alguns desses casos, podemos inferir o tamanho do viés assimptótico.

Como veremos, o problema do erro de medida tem uma estrutura estatística similar ao problema variável omitida-variável *proxy* discutido na seção anterior, mas eles são conceitualmente diferentes. No caso da variável *proxy*, estamos procurando uma variável que de certo modo é associada à variável não-observada. No caso do erro de medida, a variável que não observamos tem um significado quantitativo bem definido (como o ganho marginal do imposto ou a renda anual), mas as medidas sobre elas registradas por nós, podem conter erros. Por exemplo, a renda anual registrada é uma medida da renda anual efetiva, enquanto a pontuação de QI é uma *proxy* da aptidão.

Outra diferença importante entre os problemas da variável *proxy* e do erro de medida, é que, no último caso, muitas vezes a variável independente mal medida é a de maior interesse. No caso da variável *proxy*, o efeito parcial da variável omitida raramente é de interesse central: normalmente estamos preocupados com os efeitos das outras variáveis independentes.

Antes de considerarmos os detalhes, devemos nos lembrar que o erro de medida é um problema somente quando as variáveis cujos dados o econometrista pode coletar diferem das variáveis que influenciam as decisões de indivíduos, famílias, firmas etc.

Erro de Medida na Variável Dependente

Começaremos com o caso no qual somente a variável dependente é medida com erro. Vamos chamar de y^* a variável (na população, como sempre) que queremos explicar. Por exemplo, y^* poderia ser a poupança familiar anual. O modelo de regressão tem a forma usual

$$y^* = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u, \quad (9.17)$$

e assumimos que satisfaz as hipóteses de Gauss-Markov. Seja y a medida observável de y^* . No caso da poupança, y é a poupança anual registrada. Infelizmente, as famílias não declaram com perfeição

susas poupanças anuais; é fácil deixar categorias de fora ou superestimar o montante contribuído para determinado fundo. Geralmente, podemos esperar que y e y^* sejam diferentes, pelo menos em alguns subconjuntos de famílias na população.

O erro de medida (na população) é definido como a diferença entre o valor observado e o valor real:

$$e_0 = y - y^*. \quad (9.18)$$

Para uma extração aleatória i na população, podemos escrever $e_{i0} = y_i - y_i^*$, mas o importante é como o erro de medida na população está relacionado a outros fatores. Para obter um modelo que pode ser estimado, escrevemos $y^* = y - e_0$, inserimos essa expressão na equação (9.17) e reorganizamos esta última:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u + e_0. \quad (9.19)$$

O termo erro na equação (9.19) é $u + e_0$. Como y, x_1, x_2, \dots, x_k são observados, podemos estimar este modelo por MQO. Na verdade, simplesmente ignoramos o fato de que y é uma medida imperfeita de y^* e prosseguimos da maneira habitual.

Quando o método MQO com y em lugar de y^* produz estimadores consistentes de β_j ? Como o modelo original (9.17) satisfaz as hipóteses de Gauss-Markov, u tem média zero e é não-correlacionado com cada x_j . É natural assumir que o erro de medida tem média zero; se não for assim, simplesmente obteremos um estimador viesado do intercepto, β_0 , o que raramente é motivo de preocupação. Muito mais importante é nossa suposição sobre a relação entre o erro de medida, e_0 , e as variáveis explicativas, x_j . A suposição habitual é que o erro de medida em y é estatisticamente independente de cada variável explicativa. Se isso for verdade, então os estimadores MQO de (9.19) são não-viesados e consistentes. Além disso, os procedimentos de inferência do método MQO (estatísticas t , F e LM) são válidos.

Se e_0 e u forem não-correlacionados, como normalmente se supõe, então $\text{Var}(u + e_0) = \sigma_u^2 + \sigma_0^2 > \sigma_u^2$. Isso significa que o erro de medida na variável dependente resulta em uma variância de erro maior do que quando não ocorre nenhum erro; isso produz, evidentemente, variâncias maiores dos estimadores MQO. Esses resultados devem ser esperados, e não há nada que possamos fazer (exceto coletar dados melhores). O ponto principal é que se o erro de medida for não-correlacionado com as variáveis independentes, a estimação MQO possuirá boas propriedades.

EXEMPLO 9.5

(Função de Poupança com Erro de Medida)

Considere uma função de poupança

$$poup^* = \beta_0 + \beta_1 rend + \beta_2 tam + \beta_3 educ + \beta_4 idade + u,$$

mas na qual a poupança real ($poup^*$) possa desviar-se da poupança registrada ($poup$). A questão é saber se o tamanho do erro de medida em $poup$ está ou não sistematicamente relacionado com as outras variáveis. Pode ser razoável assumir que o erro de medida não esteja correlacionado com $rend$, tam , $educ$ e $idade$. De outro lado, podemos pensar que famílias com rendas mais elevadas, ou mais educação, declaram suas poupanças com mais precisão. Não podemos ter certeza se o erro de medida está correlacionado com $rend$ ou $educ$, a menos que possamos coletar dados de $poup^*$; então o erro de medida poderá ser calculado para cada observação como $ei_0 = poup - poup^*$.

Quando a variável dependente está na forma logarítmica, de forma que $\log(y^*)$ é a variável dependente, é natural que a equação do erro de medida seja da forma

$$\log(y) = \log(y^*) + e_0. \quad (9.20)$$

Isso é proveniente de um **erro de medida multiplicativo** de y : $y = y^*a_0$, onde $a_0 > 0$ e $e_0 = \log(a_0)$.

EXEMPLO 9.6

(Erro de Medida nas Taxas de Rejeição de Produtos Industriais)

Na Seção 7.6 discutimos um exemplo no qual queríamos determinar se a concessão de subsídios para treinamento de pessoal reduzia a taxa de rejeição de produtos das indústrias. Podemos certamente imaginar que a taxa de rejeição registrada pela empresa seja medida com erro. (De fato, a maioria das empresas da amostra sequer computa taxas de rejeição de seus produtos.) Em uma estrutura de regressão simples isso é capturado pela equação

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

onde $rejei^*$ é a rejeição verdadeira e $subs$ é a variável *dummy* indicando se uma empresa recebeu subsídios. A equação do erro de medida é

$$\log(rejei) = \log(rejei^*) + e_0.$$

O erro de medida e_0 é independente de a empresa ter, ou não, recebido subsídios? De um ponto de vista crítico, é possível pensar que uma empresa que tenha recebido subsídios está mais propensa a esconder sua taxa de rejeição para fazer com que os subsídios pareçam efetivos. Se isso acontecer, então, na equação estimada

$$\log(rejei) = \beta_0 + \beta_1 subs + u + e_0,$$

o erro $u + e_0$ é negativamente correlacionado com $subs$. Isso produziria um viés para baixo em β_1 , o que tenderia a fazer com que o programa de treinamento parecesse mais efetivo do que na realidade foi. (Lembre-se: um β_1 mais negativo significa que o programa foi mais efetivo, pois uma melhor produtividade do trabalhador está associada a uma taxa de rejeição mais baixa.)

O ponto principal desta subseção é que o erro de medida na variável dependente pode causar vieses no método MQO se ele for sistematicamente relacionado com uma ou mais das variáveis explicativas. Se o erro de medida for apenas um erro de informação aleatório que seja independente das variáveis explicativas, como muitas vezes é assumido, o método MQO é perfeitamente apropriado.

Erro de Medida em uma Variável Explicativa

Tradicionalmente, o erro de medida em uma variável explicativa tem sido considerado um problema muito mais importante do que o erro de medida em uma variável dependente. Nesta subseção veremos a razão de isso ser assim.

Comecemos com o modelo de regressão simples

$$y = \beta_0 + \beta_1 x_1^* + u, \quad (9.21)$$

assumindo que ele satisfaz pelo menos as primeiras quatro hipóteses de Gauss-Markov. Isso significa que a estimação de (9.21) por MQO produziria estimadores de β_0 e β_1 não-viesados e consistentes. O problema é que x_1^* não é observado. Em vez disso, temos uma medida de x_1^* , que pode ser chamada de x_1 . Por exemplo, x_1^* poderia ser a verdadeira renda e x_1 poderia ser a renda registrada.

O erro de medida na população é simplesmente

$$e_1 = x_1 - x_1^*, \quad (9.22)$$

e pode ser positivo, negativo ou zero. Assumimos que o erro de medida médio na população é zero: $E(e_1) = 0$. Isso é natural e, de qualquer forma, não afeta a importante conclusão a seguir. Uma suposição sustentada no que segue é que u é não-correlacionado com x_1^* e x_1 . Em termos de expectativa condicional, podemos escrevê-la como $E(y|x_1^*, x_1) = E(y|x_1^*)$, que apenas diz que x_1 não afeta y após ter-se controlado x_1^* . Usamos a mesma suposição no caso da variável *proxy* e isso não é controverso; ela se mantém quase que por definição.

Queremos saber as propriedades de MQO se simplesmente substituirmos x_1^* por x_1 e computarmos a regressão de y sobre x_1 . Elas dependerão crucialmente das suposições que fizermos sobre o erro de medida. Duas hipóteses têm sido enfatizadas na literatura econométrica, e ambas representam extremos opostos. A primeira hipótese é que e_1 é não-correlacionado com a medida observada, x_1 :

$$\text{Cov}(x_1, e_1) = 0. \quad (9.23)$$

Da relação em (9.22), se a hipótese (9.23) for verdadeira, então e_1 deve ser correlacionado com a variável não-observada x_1^* . Para determinar as propriedades de MQO neste caso, escrevemos $x_1^* = x_1 - e_1$, e inserimos esta expressão na equação (9.21):

$$y = \beta_0 + \beta_1 x_1 + (u - \beta_1 e_1). \quad (9.24)$$

Como assumimos que tanto u quanto e_1 têm média zero e são não-correlacionados com x_1 , $u - \beta_1 e_1$ tem média zero e é não-correlacionado com x_1 . Em consequência, a estimação de MQO com x_1 em lugar de x_1^* produz um estimador consistente de β_1 (e também de β_0). Como u é não-correlacionado com e_1 , a variância do erro em (9.24) é $\text{Var}(u - \beta_1 e_1) = \sigma_u^2 + \beta_1^2 \sigma_e^2$. Assim, exceto quando $\beta_1 = 0$, o erro de medida aumenta a variância do erro. Porém, isso não afeta nenhuma das propriedades de MQO (exceto pelo fato de que as variâncias da $\hat{\beta}_j$ serão maiores do que se observarmos x_1^* diretamente).

A hipótese de que e_1 é não-correlacionada com x_1 é análoga à hipótese da variável *proxy* que fizemos na Seção 9.2. Como esta hipótese implica que o método MQO tenha todas as suas propriedades perfeitas, não é isso o que os economistas têm em mente quando se referem ao erro de medida em uma variável explicativa. A suposição de **erro clássico nas variáveis (ECV)** é que o erro de medida é não-correlacionado com a variável explicativa *não-observada*:

$$\text{Cov}(x_1^*, e_1) = 0. \quad (9.25)$$

Esta hipótese provém de ter-se escrito a medida observada como a soma da variável explicativa verdadeira com o erro de medida,

$$x_1 = x_1^* + e_1,$$

e em seguida assumindo que os dois componentes de x_1 são não-correlacionados. (Isso não tem nada a ver com as hipóteses sobre u ; sempre supomos que u é não-correlacionado com x_1^* e x_1 , e, portanto, com e_1 .)

Se a hipótese (9.25) for válida, então x_1 e e_1 devem ser correlacionadas:

$$\text{Cov}(x_1, e_1) = E(x_1 e_1) = E(x_1^* e_1) + E(e_1^2) = 0 + \sigma_{e_1}^2 = \sigma_{e_1}^2. \quad (9.26)$$

Assim, a covariância entre x_1 e e_1 é igual à variância do erro de medida sob a hipótese ECV.

Com referência à equação (9.24), podemos ver que a correlação entre x_1 e e_1 causará problemas. Como u e x_1 são não-correlacionados, a covariância entre x_1 e o erro composto $u - \beta_1 e_1$ é

$$\text{Cov}(x_1, u - \beta_1 e_1) = -\beta_1 \text{Cov}(x_1, e_1) = -\beta_1 \sigma_{e_1}^2.$$

Assim, no caso ECV, a regressão de MQO de y sobre x_1 produz um estimador viesado e inconsistente.

Utilizando os resultados assintóticos do Capítulo 5, podemos determinar o montante de inconsistência no método MQO. O limite de probabilidade de $\hat{\beta}_1$ é β_1 mais a razão da covariância entre x_1 e $u - \beta_1 e_1$ e a variância de x_1 :

$$\begin{aligned} \text{plim}(\hat{\beta}_1) &= \beta_1 + \frac{\text{Cov}(x_1, u - \beta_1 e_1)}{\text{Var}(x_1)} \\ &= \beta_1 - \frac{\beta_1 \sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} = \beta_1 \left(1 - \frac{\sigma_{e_1}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right) \\ &= \beta_1 \left(-\frac{\sigma_{x_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2}\right), \end{aligned} \quad (9.27)$$

onde usamos o fato de que $\text{Var}(x_1) = \text{Var}(x_1^*) + \text{Var}(e_1)$.

A equação (9.27) é bastante interessante. O termo que multiplica β_1 , que é a razão $\text{Var}(x_1^*)/\text{Var}(e_1)$, é sempre menor que um [uma implicação da hipótese ECV (9.25)]. Assim, $\text{plim}(\hat{\beta}_1)$ está sempre mais perto de zero que β_1 . Isso é chamado de **viés de atenuação** no MQO devido ao erro clássico nas variáveis: em médias (ou grandes) amostras, o efeito estimado de MQO será *atenuado*. Em particular, se β_1 for positivo, $\hat{\beta}_1$ tenderá a subestimar β_1 . Esta é uma conclusão importante, porém ela depende da hipótese ECV.

Se a variância de x_1^* for grande, em relação à variância no erro de medida, então a inconsistência no MQO será pequena. Isso é devido ao fato de $\text{Var}(x_1^*)/\text{Var}(x_1)$ ficar próximo da unidade, quando $\sigma_{x_1^*}^2/\sigma_{e_1}^2$ for grande. Portanto, dependendo do volume de variação em x_1^* , com relação a e_1 , o erro de medida não causará, necessariamente, grandes vieses.

As coisas se complicam quando adicionamos mais variáveis explicativas. Como ilustração, considere o modelo

$$y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + \beta_3 x_3 + u, \quad (9.28)$$

onde a primeira das três variáveis explicativas é medida com erro. Naturalmente, supomos que u é não-correlacionado com x_1^*, x_2, x_3 e x_1 . Novamente, a hipótese crucial refere-se ao erro de medida e_1 . Em quase todos os casos, assume-se que e_1 é não-correlacionado com x_2 e x_3 — as variáveis explicativas não medidas com erro. O grande problema é se e_1 é não-correlacionado com x_1 . Se for, então a regressão MQO de y sobre x_1, x_2 e x_3 produzirá estimadores consistentes. Pode-se ver isso facilmente escrevendo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u - \beta_1 e_1, \quad (9.29)$$

onde u e e_1 são ambos não-correlacionados com todas as variáveis explicativas.

Sob a hipótese ECV em (9.25), o MQO será viesado e inconsistente, pois e_1 é correlacionado com x_1 na equação (9.29). Lembre-se de que isso significa que, em geral, *todos* os estimadores MQO serão viesados, e não somente $\hat{\beta}_1$. E quanto ao viés de atenuação derivado na equação (9.27)? Ainda existe um viés de atenuação ao se estimar β_1 : pode ser demonstrado que

$$\text{plim}(\hat{\beta}_1) = \beta_1 \left(\frac{\sigma_{r_1^*}^2}{\sigma_{x_1^*}^2 + \sigma_{e_1}^2} \right), \quad (9.30)$$

onde r_1^* é o erro populacional na equação $x_1^* = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3 + r_1^*$. A fórmula (9.30) também funciona no caso da variável geral k , quando x_1 for a única variável medida erroneamente.

As coisas são menos nítidas ao se estimar β_j nas variáveis não medidas com erro. No caso especial em que x_1^* é não-correlacionado com x_2 e x_3 , $\hat{\beta}_2$ e $\hat{\beta}_3$ são consistentes. Entretanto, na prática isso é raro. Geralmente, o erro de medida em uma única variável provoca inconsistência em todos os estimadores. Infelizmente, os tamanhos, e até mesmo as direções dos vieses, não são facilmente derivados.

EXEMPLO 9.7**(Equação da Nota Média com Erro de Medida)**

Considere o problema de estimar o efeito da renda familiar na nota média da graduação, após ter-se controlado $nmem$ (nota média do ensino médio) e sat (teste de aptidão acadêmica).

Pode ser que, embora a renda familiar seja importante para o desempenho escolar antes da faculdade, ela não tenha efeito direto no desempenho na faculdade. Para testarmos isso, podemos postular o modelo

$$nmgrad = \beta_0 + \beta_1 rendfam^* + \beta_2 nmem + \beta_3 sat + u,$$

onde $rendfam^*$ é a renda anual familiar efetiva. (Elas podem aparecer na forma logarítmica, mas, para fins de ilustração, deixaremos na forma em nível.) Dados precisos sobre $nmgrad$, $nmem$ e sat são relativamente fáceis de ser obtidos. Porém, a renda familiar, especialmente as informadas pelos estudantes, podem facilmente ser mal medidas. Se $rendfam = rendfam^* + e_1$, e a hipótese ECV for válida, então o uso da renda familiar informada em lugar da renda familiar efetiva viesará o estimador MQO de β_1 em direção a zero. Uma consequência disso é que um teste de $H_0: \beta_1 = 0$ terá menos possibilidade de detectar $\beta_1 > 0$.

Evidentemente, o erro de medida pode estar presente em mais de uma variável explicativa, ou em algumas das variáveis explicativas e na variável dependente. Como discutido anteriormente, qualquer erro de medida na variável dependente é usualmente assumido como não-correlacionado com todas as variáveis explicativas, seja ele observado ou não. Derivar o viés nos estimadores MQO sob extensões das hipóteses ECV é complicado e não leva a resultados claros.

Em alguns casos, fica evidente que a hipótese ECV em (9.25) não pode ser verdadeira. Considere uma variante do Exemplo 9.7:

$$nmgrad = \beta_0 + \beta_1 fumou^* + \beta_2 nmem + \beta_3 sat + u,$$

onde $fumou^*$ é o número efetivo de vezes que um estudante fumou maconha nos últimos 30 dias. A variável $fumou$ é a resposta à questão: em quantas ocasiões distintas um estudante fumou maconha nos últimos 30 dias? Suponha que postulemos o modelo padrão de erro de medida

$$fumou = fumou^* + e_1.$$

Mesmo que admitamos que os estudantes informem a verdade, é pouco provável que a hipótese ECV se mantenha. Pessoas que nunca fumam maconha — de forma que $fumou^* = 0$ — provavelmente responderão $fumou = 0$, de modo que o erro de medida será, provavelmente, zero para os estudantes que nunca fumaram maconha. Quando $fumou^* > 0$ é muito mais provável que o estudante tenha errado na contagem de quantas vezes ele, ou ela, fumou maconha nos últimos 30 dias. Isso significa que o erro de medida e_1 e o número efetivo de vezes em que ele fumou, $fumou^*$, são correlacionados, o que infringe a hipótese ECV em (9.25). Infelizmente, derivar as implicações do erro de medida que não satisfazem (9.23) ou (9.25) é difícil e está além do escopo deste livro.

Seja educ^* o grau efetivo de escolaridade, medido em anos (que pode ser um número não-inteiro) e seja educ o número de anos mais elevado de educação formal. Você acha que educ e educ^* são relacionados pelo modelo clássico de erro nas variáveis?

Antes de sairmos desta seção, enfatizamos que, *a priori*, a hipótese ECV (9.25) não é melhor ou pior que a hipótese (9.23), o que implica que MQO é consistente. A verdade está, provavelmente, em algum ponto entre as duas, e se e_1 for correlacionado com x_1^* e x_1 , MQO é inconsistente. Isto levanta uma questão importante: temos que conviver com estimadores inconsistentes sob o modelo clássico de erro nas variáveis, ou com outros tipos de erros de medida que são correlacionados com x_1 ? Felizmente, a resposta é não. O Capítulo 15 mostra como, sob certas hipóteses, os parâmetros podem ser consistentemente estimados na presença de erros gerais de medida. Adiamos a discussão para mais tarde, porque ela exige que abandonemos o âmbito da estimativa MQO.

9.4 AUSÊNCIA DE DADOS, AMOSTRAS NÃO-ALEATÓRIAS E OBSERVAÇÕES EXTREMAS

O problema do erro de medida discutido na seção anterior pode ser visto como um problema de dados: não podemos obter dados sobre as variáveis de interesse. Além disso, sob o modelo clássico de erro nas variáveis, o termo erro composto é correlacionado com a variável dependente mal medida, violando as hipóteses de Gauss-Markov.

Outro problema de dados que discutimos várias vezes em capítulos anteriores é a multicolinearidade entre as variáveis explicativas. Lembremo-nos de que a correlação entre variáveis explicativas não infringe nenhuma hipótese. Quando duas variáveis independentes são altamente correlacionadas, pode ser difícil estimar o efeito parcial de cada uma delas. Entretanto, isto é adequadamente refletido nas estatísticas de MQO usuais.

Nesta seção apresentamos uma introdução aos problemas de dados que podem violar a hipótese de amostragem aleatória, RLM.2. Podemos isolar casos nos quais a amostragem não-aleatória não tem efeito prático sobre o método MQO. Em outros casos, a amostragem não-aleatória faz com que os estimadores MQO sejam viesados e inconsistentes. Um tratamento mais completo que comprova várias das afirmações feitas aqui é apresentado no Capítulo 17.

Ausência de Dados

O problema de **ausência de dados** pode surgir de várias formas. Muitas vezes coletamos uma amostra aleatória de pessoas, escolas, cidades etc. e mais tarde descobrimos que estão faltando informações de algumas variáveis importantes para diversas unidades na amostra. Por exemplo, no conjunto de dados do arquivo BWGHT.RAW, 197 das 1.388 observações não têm nenhuma informação sobre a educação do pai, da mãe ou de ambos. No conjunto de dados sobre salários medianos iniciais dos recém-formados em faculdades de direito, no arquivo LAWSCH85.RAW, seis das 156 faculdades não têm informações sobre as medianas da pontuação *lsat* dos alunos, relativas a classes novas; também faltam outras variáveis de algumas faculdades de direito.

Quando estão faltando dados de uma observação na variável dependente ou em uma das variáveis independentes, a observação não pode ser usada em uma análise de regressão múltipla padrão. Aliás, desde que os dados ausentes tenham sido adequadamente indicados, todos os modernos programas de

regressão rastreiam os dados e simplesmente ignoram as observações ao calcularem uma regressão. Vimos isso claramente na situação do peso de nascimento no Exemplo 4.9, quando 197 observações foram eliminadas devido à não-existência de informações sobre o nível de educação dos pais.

Além de reduzir o tamanho da amostra disponível para uma regressão, há consequências *estatísticas* provocadas pela ausência de dados? Depende do motivo da ausência dos dados. Se estes estiverem faltando aleatoriamente, então o tamanho da amostra aleatória disponível da população será simplesmente reduzido. Embora isso torne os estimadores menos precisos, não produz nenhum viés: a hipótese de amostragem aleatória, RLM.2, ainda é válida. Existem maneiras de usar as informações das observações nas quais somente algumas variáveis estão faltando, mas, na prática, não se faz isso com frequência. A melhoria nos estimadores normalmente é pequena, embora o método seja um pouco complicado. Na maioria dos casos, simplesmente ignoramos as observações que representam falta de informação.

Amostras Não-Aleatórias

A ausência de dados é mais problemática quando resulta de uma **amostra não-aleatória** da população. Por exemplo, no conjunto de dados sobre pesos de nascimento, o que acontecerá se a probabilidade de que estejam faltando informações sobre o nível de educação for mais alta para as pessoas cujo nível de educação seja mais baixo que a média? Ou, na Seção 9.2, tenhamos usado um conjunto de dados que tenha incluído pontuações de QI. Esse conjunto de dados foi construído com a omissão de várias pessoas da amostra para as quais não havia informações sobre a pontuação do QI. Se a obtenção de escores de QI é mais fácil para as pessoas com QI mais elevado, a amostra não será representativa da população. A hipótese de amostragem aleatória RLM.2 está sendo violada e devemos nos preocupar com suas consequências durante a estimação MQO.

Felizmente, certos tipos de amostragens não-aleatórias *não* causam viés ou inconsistência no MQO. Sob as hipóteses de Gauss-Markov (mas sem a RLM.2), a amostra pode ser escolhida com base nas variáveis *independentes* sem causar nenhum problema estatístico. Isso é chamado de *seleção da amostra com base nas variáveis independentes*, e é um exemplo de *seleção amostral exógena*. Para ilustrar, suponha que estejamos estimando uma função de poupança, na qual a poupança anual depende da renda, idade, tamanho da família e talvez de alguns outros fatores. Um modelo simples é

$$\text{poupança} = \beta_0 + \beta_1 \text{renda} + \beta_2 \text{idade} + \beta_3 \text{tamanho} + u. \quad (9.31)$$

Suponha que nosso conjunto de dados foi montado com base em uma pesquisa feita com pessoas com mais de 35 anos de idade, com isso deixando-nos com uma amostra não-aleatória de todos os adultos. Embora isso não seja o ideal, ainda podemos obter estimadores não-viesados e consistentes dos parâmetros no modelo populacional (9.31), utilizando a amostra não-aleatória. Não demonstraremos isso formalmente aqui, mas a razão pela qual o MQO na amostra não-aleatória é não-viesado é o fato de a função de regressão $E(\text{poupança} | \text{renda}, \text{idade}, \text{tamanho})$ ser a mesma para qualquer subconjunto da população descrito por *renda*, *idade* ou *tamanho*. Desde que haja variação suficiente nas variáveis independentes na subpopulação, a seleção com base nas variáveis independentes não será um problema sério, exceto pelo fato de resultar em estimadores ineficientes.

No exemplo mencionado da pontuação do QI, as coisas não são tão nítidas, porque nenhuma regra fixa baseada no QI foi utilizada para incluir alguém na amostra. Ao contrário, a *probabilidade* de estar na amostra aumenta com o QI. Se os outros fatores que determinam a seleção na amostra forem independentes do termo erro na equação de salários, teremos outro caso de seleção amostral exógena,

e o MQO usando a amostra selecionada terá todas as propriedades desejáveis sob as outras hipóteses de Gauss-Markov.

O assunto é bem diferente quando a seleção é baseada na variável dependente, y , que é chamada de *seleção de amostra com base na variável dependente* e é um exemplo de **seleção amostral endógena**. Se a amostra tiver como base o fato de a variável dependente estar acima ou abaixo de determinado valor, sempre ocorrerá viés no MQO ao estimarmos o modelo populacional. Por exemplo, suponha que queremos estimar a relação entre a riqueza individual e vários outros fatores na população adulta:

$$\text{riqueza} = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{idade} + u. \quad (9.32)$$

Suponha que somente pessoas com riqueza abaixo de 75.000 dólares sejam incluídas na amostra. Essa é uma amostra não-aleatória da população de interesse, e se baseia no valor da variável dependente. A utilização de uma amostra de pessoas com riqueza abaixo de 75.000 dólares resultará em estimadores viesados e inconsistentes dos parâmetros em (9.32). Resumidamente, a razão é que a regressão populacional $E(\text{riqueza}|\text{educ}, \text{exper}, \text{idade})$ não é a mesma que o valor esperado condicional da *riqueza* ser menor que 75.000 dólares.

Outros esquemas de amostragem levam a amostras não-aleatórias da população, em geral intencionalmente. Um método comum de coleta de dados é a **amostragem estratificada**, na qual a população é dividida em grupos ou estratos não sobrepostos. Então, alguns grupos aparecem com mais freqüência do que a determinada por sua representação populacional e outros aparecem com menor freqüência. Por exemplo, algumas pesquisas propositalmente superdimensionam grupos minoritários ou grupos de baixa renda. Quando forem necessários métodos especiais, isso dependerá outra vez de ser a estratificação exógena (baseada em variáveis explicativas exógenas) ou endógena (baseada na variável dependente). Suponha que uma pesquisa sobre o contingente militar superdimensionou mulheres porque o interesse inicial era estudar os fatores que determinam o pagamento às mulheres no serviço militar. (Superdimensionar um grupo que seja relativamente pequeno na população é comum na coleta de amostras estratificadas.) Desde que os homens também tenham sido representados na amostra, podemos usar o MQO na amostra estratificada para estimar qualquer diferencial de gênero, juntamente com os retornos da educação e da experiência de todo o contingente militar. (Podemos querer pressupor que retorno da educação e da experiência não sejam específicos quanto ao gênero.) A razão pela qual o MQO é não-viesado e consistente prende-se ao fato de a estratificação ser feita com relação a uma variável explicativa, ou seja, o gênero.

Se, em vez disso, a pesquisa superdimensionou o contingente militar de salários mais baixos, então o MQO utilizando a amostra estratificada não estima consistentemente os parâmetros da equação de salários militares, porque a estratificação é endógena. Em tais casos, são necessários métodos econométricos especiais [veja Wooldridge (2002, Capítulo 17)].

A amostragem estratificada é uma forma bastante óbvia de amostragem não-aleatória. Outros problemas de seleção de amostras são mais sutis. Por exemplo, em vários dos exemplos anteriores estimamos os efeitos de diversas variáveis, particularmente educação e experiência, sobre os salários por hora. O conjunto de dados do arquivo WAGE1.RAW, que utilizamos amplamente, é, essencialmente, uma amostra aleatória de indivíduos *trabalhadores*. Economistas especializados em trabalho com freqüência estão interessados em estimar os efeitos da, digamos, educação sobre os salários ofertados. A ideia é a seguinte: toda pessoa em idade de trabalhar se confronta com uma oferta de salários por hora e ele, ou ela, pode trabalhar, ou não, por aquele salário. Para alguém que trabalhe, o salário ofertado é o salário ganho. Para pessoas que não trabalhem não podemos, normalmente, observar o salário oferecido. Agora, como a equação da oferta salarial

$$\log(salárioh^0) = \beta_0 + \beta_1 educ + \beta_2 exper + u \quad (9.33)$$

representa a população de trabalhadores de todas as idades, não podemos estimá-la utilizando uma amostra aleatória desta população; em vez disso, temos informações sobre a oferta salarial somente para pessoas que trabalham (embora possamos obter dados sobre *educ* e *exper* de pessoas que não trabalham). Se utilizarmos uma amostra aleatória de pessoas que trabalham para estimar (9.33), obteremos estimadores não-viesados? Este caso não é muito claro. Como a amostra é selecionada com base na decisão de alguém de trabalhar (em oposição ao tamanho do salário ofertado), este não é como o caso anterior. Porém, como a decisão de trabalhar pode estar relacionada com fatores não-observados que afetam a oferta salarial, a seleção pode ser endógena, e isso pode resultar em um viés de seleção de amostra nos estimadores MQO. Trataremos os métodos que podem ser usados para testar e corrigir vieses de seleção de amostras no Capítulo 17.

Suponha que estejamos interessados nos efeitos dos gastos com campanha eleitoral feitos por candidatos, sobre o apoio dos eleitores. Alguns candidatos decidem não apresentar sua candidatura à reeleição. Se somente pudermos coletar os resultados da votação e dos gastos dos candidatos que efetivamente participaram da eleição, existe a possibilidade de ocorrer uma seleção endógena da amostra?

Observações Extremas ou Atípicas

Em algumas aplicações, especialmente, mas não apenas nelas, com conjuntos de dados pequenos, as estimativas MQO são influenciadas por uma ou diversas observações. Tais observações são chamadas de **observações extremas ou atípicas**. De maneira imprecisa, uma observação é extrema se sua eliminação de uma análise de regressão produzir mudança de “grande” monta nas estimativas MQO.

O MQO é suscetível a observações extremas porque ele minimiza a soma dos quadrados dos resíduos: grandes resíduos (positivos ou negativos) recebem muita carga no problema de minimização de mínimos quadrados. Se as estimativas mudarem em quantidade significativa quando modificarmos ligeiramente nossa amostra, devemos nos preocupar.

Quando estatísticos e econometristas estudam teoricamente o problema dessas observações extremas, algumas vezes os dados são vistos como provenientes de uma amostra aleatória de determinada população — embora com uma distribuição pouco comum que pode resultar em valores extremos — e às vezes se presume que as observações extremas provêm de uma população diferente. De uma perspectiva prática, tais observações podem ocorrer por duas razões. O caso mais fácil de tratar é quando um engano foi cometido na entrada dos dados. A adição de zeros extras a um número ou a má colocação do ponto decimal podem mascarar as estimativas MQO, especialmente em amostras de pequeno tamanho. É sempre uma boa idéia computar as estatísticas sumárias, especialmente os mínimos e os máximos, de maneira a detectar enganos da entrada dos dados. Infelizmente, entradas incorretas nem sempre são óbvias.

Os valores extremos também podem surgir quando se faz a amostragem de uma pequena população, se um ou vários membros da população forem muito diferentes em alguns aspectos relevantes do resto da população. Pode ser difícil tomar a decisão de manter ou desprezar tais observações em uma análise de regressão, e as propriedades estatísticas dos estimadores resultantes são complicadas. Observações extremas podem fornecer informações importantes aumentando-se a variação das variá-

veis explicativas (o que reduz os erros-padrão). Porém, os resultados de MQO deverão ser descritos com e sem as observações extremas, nos casos em que um ou vários pontos dos dados alteram substancialmente os resultados.

EXEMPLO 9.8

(Intensidade de Pesquisa e Desenvolvimento e Tamanho das Empresas)

Suponha que os gastos com pesquisa e desenvolvimento como uma percentagem das vendas (*pdiintens*) sejam relacionados a *vendas* (em milhões) e a lucros como uma percentagem das vendas (*lucrmarg*):

$$pdiintens = \beta_0 + \beta_1 vendas + \beta_2 lucrmarg + u. \quad (9.34)$$

A equação MQO utilizando os dados de 32 empresas da indústria química, contidos no arquivo RDCHEM.RAW é

$$\begin{aligned} pdiintens &= 2,625 + 0,000053 vendas + 0,446 lucrmarg \\ &\quad (0,586) \quad (0,000044) \quad (0,0462) \\ n &= 32, R^2 = 0,0761, \bar{R}^2 = 0,0124. \end{aligned}$$

Nem *vendas* nem *lucrmarg* são estatisticamente significantes, mesmo ao nível de 10% nesta regressão.

Das 32 empresas, 31 têm vendas anuais abaixo de 20 bilhões de dólares. Uma empresa tem vendas anuais de quase 40 bilhões de dólares. A Figura 9.1 mostra a distância desta empresa do resto da amostra. Em termos de vendas, esta empresa é quase duas vezes maior do que qualquer outra, de modo que deve ser uma boa idéia estimar o modelo sem ela. Quando fazemos isso, obtemos

$$\begin{aligned} pdiintens &= 2,297 + 0,000186 vendas + 0,0478 lucrmarg \\ &\quad (0,592) \quad (0,000084) \quad (0,0445) \\ n &= 31, R^2 = 0,1728, \bar{R}^2 = 0,1137. \end{aligned}$$

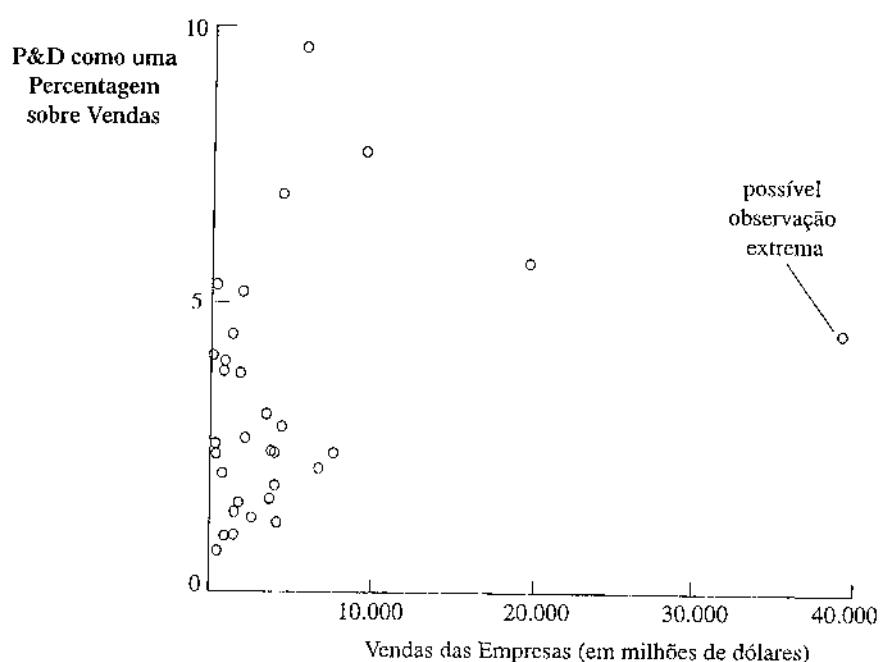
Se a maior empresa for eliminada da regressão, o coeficiente de *vendas* mais do que triplica, e agora ela tem uma estatística *t* acima de dois. Usando a amostra de empresas menores, concluíramos que existe um efeito positivo estatisticamente significante entre a intensidade de pesquisa e desenvolvimento e o tamanho das empresas. A margem de lucro ainda não é significante, e seu coeficiente não mudou muito.

Algumas vezes, as observações extremas são definidas pelo tamanho do resíduo em uma regressão MQO, na qual todas as observações são usadas. Essa não é uma boa idéia. No exemplo anterior, utilizando todas as empresas na regressão, uma empresa com vendas pouco abaixo de 4,6 bilhões de dólares tinha o maior resíduo de todas (cerca de 6,37). O resíduo para a maior empresa foi de -1,62, que é menos que um desvio-padrão estimado de zero ($\hat{\sigma} = 1,82$). A eliminação da observação com o maior resíduo não altera muito os resultados.

Algumas formas funcionais são menos sensíveis a observações extremas. Na Seção 6.2, mencionamos que, para a maioria das variáveis econômicas, a transformação logarítmica estreita significantemente a amplitude dos dados e também produz formas funcionais — tais como os modelos de elasticidade constante — que podem explicar uma gama mais ampla de dados.

Figura 9.1

Gráfico da intensidade de pesquisa e desenvolvimento e as vendas das empresas.

**EXEMPLO 9.9****(Intensidade de P&D)**

Podemos testar se a intensidade de P&D aumenta com o tamanho das empresas, começando com o modelo

$$pq = \text{vendas}^{\beta_1} \exp(\beta_0 + \beta_2 \text{lucrmarg} + u). \quad (9.35)$$

Assim, mantendo fixos os outros fatores, a intensidade de P&D aumenta com *vendas* se, e somente se, $\beta_1 > 1$. Considerando o log de (9.35) teremos

$$\log(pq) = \beta_0 + \beta_1 \log(\text{vendas}) + \beta_2 \text{lucrmarg} + u. \quad (9.36)$$

Quando usamos todas as 32 empresas, a equação de regressão é

$$\begin{aligned} \hat{\log}(pq) &= -4,378 + 1,084 \log(\text{vendas}) + 0,0217 \text{lucrmarg}, \\ &\quad (0,468) \quad (0,062) \quad (0,0128) \\ n &= 32, R^2 = 0,9180, \bar{R}^2 = 0,9123, \end{aligned}$$

EXEMPLO 9.9 (continuação)

enquanto retirando-se a maior empresa resulta na equação

$$\begin{aligned}\hat{\log}(pq) &= -4,404 + 1,088 \log(vendas) + 0,0218 lucrmarg, \\ &\quad (0,511) \quad (0,067) \quad (0,0130) \\ n &= 31, R^2 = 0,9037, \bar{R}^2 = 0,8968.\end{aligned}$$

Esses resultados são praticamente os mesmos. Em nenhum dos casos rejeitamos a hipótese nula $H_0: \beta_1 = 1$ contra $H_1: \beta_1 > 1$ (Por quê?).

Em alguns casos, certas observações são suspeitas, desde o princípio, de serem fundamentalmente diferentes do restante da amostra. Isso acontece com freqüência quando utilizamos dados em níveis muito agregados, como os níveis de cidades, municípios ou estados. O que segue é um exemplo disso.

EXEMPLO 9.10**(Taxas Estaduais de Mortalidade Infantil)**

Informações sobre mortalidade infantil, renda *per capita* e saúde podem ser obtidas, em nível de estados, no *Statistical Abstracts of the United States* (Sumários Estatísticos dos Estados Unidos). Faremos aqui uma análise simples apenas para ilustrar o efeito das observações extremas. Os dados são para o ano de 1990, e temos todos os 50 estados dos Estados Unidos, mais o Distrito de Colúmbia (D.C.). A variável *mortinf* é o número de mortes no primeiro ano de vida por 1.000 nascimentos, *rendpc* é a renda *per capita*, *medic* representa médicos por 100.000 habitantes, e *popul* é a população (em milhares). Os dados estão contidos no arquivo INFMRT.RAW. Incluímos todas as variáveis independentes na forma logarítmica:

$$\begin{aligned}\hat{mortinf} &= 33,86 - 4,68 \log(rendpc) + 4,15 \log(medic) \\ &\quad (20,43) \quad (2,60) \quad (1,51) \\ &\quad - 0,088 \log(popul) \\ &\quad (0,287) \\ n &= 51, R^2 = 0,139, \bar{R}^2 = 0,084.\end{aligned}\tag{9.37}$$

A renda *per capita* tem uma relação estimada inversa em relação à mortalidade infantil, um resultado esperado. Porém, a variável médicos *per capita* está associada com taxas maiores de mortalidade infantil, o que é contrário à intuição. As taxas de mortalidade infantil parecem não estar relacionadas ao tamanho da população.

O Distrito de Colúmbia é incomum por ter bolsões de extrema pobreza e de grande riqueza em uma área tão pequena. Aliás, a taxa de mortalidade infantil do D.C. em 1990 foi de 20,7, comparado com 12,4 do estado seguinte com maior taxa. Ele também tem 615 médicos por 100.000 habitantes, comparados com 337 do segundo estado. A grande quantidade de médicos complementada pela elevada taxa de mortalidade infantil no D.C. poderia certamente influenciar os resultados. Se retirarmos o D.C. da regressão, teremos

EXEMPLO 9.10 (continuação)

$$\begin{aligned}
 \hat{mortinf} = & 23,95 - 0,57 \log(rendpc) + 2,74 \log(medic) \\
 & (12,42) \quad (1,64) \quad (1,19) \\
 & + 0,629 \log(popul) \\
 & (0,191) \\
 n = & 50, R^2 = 0,273, \bar{R}^2 = 0,226.
 \end{aligned} \tag{9.38}$$

Agora verificamos que o número maior de médicos *per capita* reduz a mortalidade infantil, e a estimativa é estatisticamente diferente de zero ao nível de 5%. O efeito da renda *per capita* caiu drasticamente e não mais é estatisticamente significante. Na equação (9.38), as taxas de mortalidade infantil são mais elevadas nos estados mais populosos, e a relação é estatisticamente bastante significante. Além disso, muito mais variação em *mortinf* é explicada quando o D.C. é retirado da regressão. Claramente, o D.C. tem uma influência substancial nas estimativas iniciais, e provavelmente o deixaríamos de fora de qualquer análise futura.

Em vez de termos que, pessoalmente, determinar a influência de certas observações, algumas vezes é útil ter estatísticas que detectem tais observações atípicas. Essas estatísticas existem, mas estão além do escopo deste livro. [Veja, por exemplo, Belsley, Kuh e Welsh (1980).]

Antes de encerrarmos esta seção, mencionamos outro método de trabalhar com observações atípicas. Em vez de tentar encontrar observações extremas nos dados antes de aplicar os mínimos quadrados, podemos usar um método de estimação que seja menos sensível a elas do que o MQO. Isso evita a necessidade de se procurar, explicitamente, por atipicidades antes ou durante a estimação. Um desses métodos, que vem se tornando cada vez mais popular entre os econometristas, é chamado de **mínimos desvios absolutos (MDA)**. O estimador MDA minimiza a soma dos desvios absolutos dos resíduos, em lugar da soma dos resíduos quadrados. Sabe-se que o MDA foi construído para estimar os efeitos das variáveis explicativas sobre a *mediana* condicional, em vez da média condicional, da variável dependente. Como a mediana não é afetada por grandes alterações em observações extremas, as estimativas de parâmetros obtidas pelo MDA são resistentes a observações extremas. (Veja a Seção A.1 para uma breve discussão sobre a mediana da amostra.) Na escolha das estimativas, o MQO atribui muito mais importância a grandes resíduos, pois cada resíduo é quadrado.

Embora o MDA auxilie na proteção contra atipicidades, ele tem algumas inconveniências. Primeiro, não existem fórmulas para os estimadores; eles só podem ser encontrados com o uso de métodos iterativos em computadores. Um problema relacionado é o fato de a obtenção de erros-padrão das estimativas ser algo mais complicado do que obter os erros-padrão das estimativas MQO. Hoje em dia, com computadores tão poderosos, preocupações deste tipo não são muito importantes, a não ser que o MDA seja aplicado em conjuntos de dados muito grandes, com muitas variáveis explicativas. Uma segunda inconveniência, pelo menos em amostras menores, é que todas as inferências estatísticas que envolvam os estimadores MDA são justificadas apenas assintoticamente. Com o MQO, sabemos que, sob as hipóteses clássicas do modelo linear, estatísticas *t* têm distribuições *t* exatas, e estatísticas *F* têm distribuições *F* exatas. Embora versões assintóticas desses testes estejam disponíveis para o MDA, elas se justificam apenas para amostras grandes.

Outra inconveniência sutil, mas importante, do MDA é o fato de que nem sempre ele estima consistentemente os parâmetros que aparecem na função de média condicional, $E(y|x_1, \dots, x_k)$. Como mencionado anteriormente, o MDA foi construído para estimar os efeitos sobre a mediana condicional.

Geralmente, a média e a mediana são as mesmas somente quando a distribuição de y , dadas as covariadas x_1, \dots, x_k , é simétrica em relação a $\beta_0 + \beta_1x_1 + \dots + \beta_kx_k$. (Equivalentemente, o termo erro da população, u , é simétrico em relação a zero.) Lembre-se de que o MQO produz estimadores não-viesados e consistentes dos parâmetros na média condicional, seja ou não simétrica a distribuição do erro; a simetria não aparece entre as hipóteses de Gauss-Markov. Quando MDA e MQO são aplicados em casos com distribuições assimétricas, o efeito parcial estimado de, digamos, x_1 , obtido com o MDA pode ser muito diferente do obtido com o MQO. Porém, tal diferença pode apenas refletir a diferença entre a mediana e a média e pode não ter nada a ver com atipicidade nas observações.

Se assumirmos que o erro populacional u no modelo (9.2) é *independente* de (x_1, \dots, x_k) , as estimativas de inclinação MQO e MDA devem diferir apenas por erro de amostragem, seja ou não simétrica a distribuição de u . As estimativas dos interceptos geralmente serão diferentes, refletindo o fato de que, se a média de u for zero, sua mediana é diferente de zero sob assimetria. Infelizmente, a independência entre o erro e as variáveis explicativas é, muitas vezes, inacreditavelmente forte quando o MDA é aplicado. Em particular, a independência impede a heteroscedasticidade, um problema que muitas vezes surge em aplicações com distribuições assimétricas.

Mínimos desvios absolutos são um caso especial do que muitas vezes é chamado de *regressão robusta*. Infelizmente, a maneira como “robusta” está sendo usada aqui pode criar confusão. Na literatura estatística, um estimador de regressão robusta é relativamente insensível a observações extremas.

Efetivamente, observações com grandes resíduos recebem menos peso do que nos mínimos quadrados. [Berk (1990) contém um tratamento introdutório de estimadores que são robustos em relação a observações extremas.] Com base em nossa discussão anterior, no linguajar característico econômico, o MDA não é um estimador robusto da média condicional, pois ele exige hipóteses extras para estimar consistentemente os parâmetros da média condicional. Na equação (9.2), a distribuição de u dado (x_1, \dots, x_k) tem que ser simétrica em relação a zero, ou u deve ser independente de (x_1, \dots, x_k) . Nada disso é exigido pelo MQO.

Investigamos mais detalhadamente alguns problemas importantes de especificação e de dados que muitas vezes surgem na análise de corte transversal empírica. Formas funcionais mal especificadas fazem com que a equação estimada seja difícil de ser interpretada. No entanto, a forma funcional incorreta pode ser detectada pela adição de termos quadráticos, pelo cálculo do teste RESET, ou fazendo-se o teste contra um modelo alternativo não-aninhado, utilizando o teste de Davidson-MacKinnon. Não é necessária nenhuma coleta adicional de dados.

Resolver o problema das variáveis omitidas é mais difícil. Na Seção 9.2 discutimos uma possível solução com base no uso de uma variável *proxy* substituindo a variável omitida. Sob hipóteses razoáveis, a inclusão da variável *proxy* em uma regressão MQO elimina ou pelo menos reduz o viés. A dificuldade em aplicar este método é que variáveis *proxy* podem ser difíceis de encontrar. Uma possibilidade geral é usar dados de uma variável dependente de um ano anterior.

Os economistas que trabalham em áreas aplicadas estão freqüentemente preocupados com os erros de medida. Sob as hipóteses do erro clássico nas variáveis (ECV), o erro de medida na variável dependente não tem efeito nas propriedades estatísticas do MQO. De outro lado, sob as hipóteses ECV para uma variável independente, o estimador MQO do coeficiente na variável mal medida é viesado em direção a zero. O viés nos coeficientes das outras variáveis pode ser para qualquer lado e é difícil de ser determinado.

Amostras não-aleatórias de uma população subjacente podem levar a vieses no MQO. Quando a seleção da amostra está correlacionada com o termo erro u , o MQO é geralmente viésado e inconsistente. De outro lado, a seleção amostral exógena — que se baseia nas variáveis explicativas ou, se não for assim, é independente de u — não causa problemas para o MQO. Observações extremas em conjuntos de dados podem produzir grandes impactos nas estimativas MQO, especialmente em amostras pequenas. É importante, pelo menos informalmente, identificar as atípicidades e reestimar os modelos com as observações extremas suspeitas excluídas.

9.1 No Exercício 4.11 o R -quadrado da estimativa do modelo

$$\begin{aligned} \log(\text{salário}) = & \beta_0 + \beta_1 \log(\text{vendas}) + \beta_2 \log(\text{valmerc}) + \beta_3 \text{lucrmarg} \\ & + \beta_4 \text{perceo} + \beta_5 \text{percomp} + u, \end{aligned}$$

usando os dados contidos no arquivo CEOSAL2.RAW, era $R^2 = 0,353$ ($n = 177$). Quando perceo^2 e percomp^2 são adicionados, $R^2 = 0,375$. Existe evidência de má especificação da forma funcional neste modelo?

9.2 Modifiquemos o Exercício 8.9 utilizando os resultados das eleições de 1990 dos candidatos que foram eleitos em 1988. O candidato A foi eleito em 1988 e buscava a reeleição em 1990; $votoA90$ é a percentagem de votos do Candidato A na eleição de 1990. A percentagem de votos do Candidato A na eleição de 1988 é usada como uma variável proxy da qualidade do candidato. Todas as demais variáveis são das eleições de 1990. As seguintes equações foram estimadas, utilizando os dados contidos no arquivo VOTE2.RAW:

$$\begin{aligned} \hat{votoA90} = & 75,71 + 0,312 \text{ forpartA} + 4,93 \text{ democA} \\ & (9,25) (0,046) (1,01) \\ & - 0,929 \log(gastoA) - 1,950 \log(gastoB) \\ & (0,684) (0,281) \\ & n = 186, R^2 = 0,495, \bar{R}^2 = 0,483, \end{aligned}$$

e

$$\begin{aligned} \hat{votoA90} = & 70,81 + 0,282 \text{ forpartA} + 4,52 \text{ democA} \\ & (10,01) (0,052) (1,06) \\ & - 0,839 \log(gastoA) - 1,846 \log(gastoB) + 0,067 \text{ votoA88} \\ & (0,687) (0,292) (0,053) \\ & n = 186, R^2 = 0,499, \bar{R}^2 = 0,485. \end{aligned}$$

- (i) Interprete o coeficiente de $votoA88$ e comente sobre sua significância estatística.
- (ii) A adição de $votoA88$ tem muito efeito sobre os outros coeficientes?

9.3 Seja $mate10$ a percentagem de aprovação em um teste padrão de matemática de estudantes de uma escola secundária de Michigan (veja também o Exemplo 4.2). Estamos interessados em estimar o efeito do gasto por estudante no desempenho em matemática. Um modelo simples é

$$mate10 = \beta_0 + \beta_1 \log(gasto) + \beta_2 \log(matricl) + \beta_3 pobreza + u,$$

onde $pobreza$ é a percentagem de estudantes vivendo em condições de pobreza.

- (i) A variável $prgalm$ é a percentagem de estudantes qualificados para o programa de merenda escolar financiado pelo governo federal. Por que ela é uma variável proxy razoável de *pobreza*?
- (ii) A tabela seguinte contém estimativas MQO, com e sem $prgalm$ como uma variável explicativa.

Variável Dependente: $mate10$

Variáveis Independentes	(1)	(2)
$\log(gasto)$	11,13 (3,30)	7,75 (3,04)
$\log(matricl)$	0,022 (0,615)	-1,26 (0,58)
$prgalm$	—	-0,324 (0,036)
<i>intercepto</i>	-69,24 (26,72)	-23,14 (24,99)
Observações	428	428
<i>R</i> -quadrado	0,0297	0,1893

Explique por que o efeito dos gastos sobre $mate10$ é menor na coluna (2) do que na coluna (1). O efeito na coluna (2) ainda é estatisticamente maior que zero?

- (iii) Parece que as taxas de aprovação são menores em escolas maiores, com os outros fatores sendo iguais?
- (iv) Interprete o coeficiente de $prgalm$ na coluna (2).
- (v) O que você deduz do substancial aumento de R^2 da coluna (1) para a coluna (2)?

9.4 A equação seguinte explica o número de horas por semana que uma criança passa assistindo televisão, em termos da idade da criança, educação da mãe, educação do pai e número de irmãos:

$$tvhoras^* = \beta_0 + \beta_1 idade + \beta_2 idade^2 + \beta_3 educm + \beta_4 educp + \beta_5 irmos + u.$$

Estamos preocupados com a possibilidade de que $tvhoras^*$ tenha sido medida com erro em nossa pesquisa. Seja $tvhoras$ o número de horas por semana que se gasta assistindo televisão.

- (i) O que as hipóteses do erro clássico nas variáveis (ECV) requerem nesta aplicação?

(ii) Você acha que as hipóteses ECV têm possibilidades de se manter? Explique.

9.5 No Exemplo 4.4 estimamos um modelo relacionando número de crimes no *campus* às matrículas de estudantes em um grupo de faculdades. A amostra que usamos não era uma amostra aleatória de faculdades nos Estados Unidos, pois muitas escolas em 1992 não registraram crimes no *campus*. Você acha que a falha das faculdades em informar os crimes pode ser vista como uma seleção amostral exógena? Explique.



Análise de Regressão com Dados de Séries Temporais

gora que temos uma sólida compreensão de como usar o modelo de regressão múltipla em aplicações de corte transversal, podemos passar para a análise econométrica de dados de séries temporais. Como vamos nos apoiar muito no método dos mínimos quadrados ordinários, a maior parte do trabalho relativo à mecânica e à inferência já foi feito. Entretanto, como vimos no Capítulo 1, os dados de séries temporais têm certas características não presentes em dados de corte transversal, que podem exigir uma atenção especial no momento da aplicação do MQO.

O Capítulo 10 aborda o básico da análise de regressão e dá uma atenção especial a problemas específicos de dados de séries temporais. Apresentamos um conjunto de hipóteses de Gauss-Markov e do modelo linear clássico para aplicações de séries temporais. São discutidos também os problemas da forma funcional, variáveis *dummy*, tendência e sazonalidade.

Como certos modelos de séries temporais necessariamente violam as hipóteses de Gauss-Markov, o Capítulo 11 descreve a natureza dessas violações e apresenta as propriedades de amostras grandes do método de mínimos quadrados ordinários. Como não podemos mais assumir amostragens aleatórias, devemos considerar condições que restrinjam a correlação temporal em uma série de tempo, a fim de garantir que a análise assimptótica usual seja válida.

O Capítulo 12 enfoca um novo e importante problema: a correlação serial nos erros nas regressões de séries temporais. Discutimos as consequências, maneiras de testar e métodos para trabalhar com a presença de correlação serial. O Capítulo 12 também contém uma explicação de como a heteroscedasticidade pode surgir em modelos de séries temporais.

O Básico da Análise de Regressão com Dados de Séries Temporais

este capítulo começamos a estudar as propriedades de MQO para estimar modelos de regressão linear usando dados de séries temporais. Na Seção 10.1 discutiremos algumas diferenças conceituais entre séries temporais e dados de corte transversal. A Seção 10.2 apresentará alguns exemplos de regressões de séries temporais que geralmente são estimadas nas ciências sociais empíricas. Depois voltaremos nossa atenção para as propriedades da amostra finita dos estimadores de MQO e definiremos as hipóteses de Gauss-Markov e as hipóteses do modelo linear clássico para a regressão de séries temporais. Embora essas hipóteses tenham aspectos comuns em relação às do caso de corte transversal, elas também têm algumas diferenças importantes que precisarão ser destacadas.

Além disso, voltaremos a algumas questões que tratamos na regressão com dados de corte transversal, como, por exemplo, de que maneira usar e interpretar a forma funcional logarítmica e as variáveis *dummy*. Os tópicos importantes sobre como incorporar tendência e explicar a sazonalidade em regressão múltipla serão discutidos na Seção 10.5.

10.1 A NATUREZA DOS DADOS DAS SÉRIES TEMPORAIS

Uma característica óbvia dos dados de séries temporais, que os distingue dos dados de corte transversal, é que um conjunto de dados de séries temporais tem uma ordenação temporal. Por exemplo, no Capítulo 1 falamos resumidamente de um conjunto de dados de séries temporais relativos ao emprego, salário mínimo e outras variáveis econômicas em Porto Rico. Nesse conjunto de dados, devemos saber que os dados de 1970 precedem imediatamente os dados de 1971. Para analisar dados de séries temporais nas ciências sociais devemos reconhecer que o passado pode afetar o futuro, mas o contrário não acontece (ao contrário do universo do filme *Jornada nas Estrelas*). Para enfatizar a ordenação apropriada dos dados de séries temporais, a Tabela 10.1 apresenta uma lista parcial de dados sobre taxas de inflação e desemprego nos Estados Unidos contidas em várias edições do *Economic Report of the President* (Relatório Econômico da Presidência), incluindo o de 2001 (Tabelas B-43 e B-63).

Outra diferença entre dados de corte transversal e dados de séries temporais é mais sutil. Nos Capítulos 3 e 4 estudamos as propriedades estatísticas dos estimadores de MQO baseados na noção de que as amostras foram colhidas aleatoriamente de determinada população. É muito fácil entender porque os dados de corte transversal devem ser vistos como resultados aleatórios: uma amostra diferente colhida da população em geral vai produzir valores diferentes das variáveis dependentes e independentes (como educação, experiência, salário etc.). Portanto, as estimativas de MQO calculadas a partir de diferentes amostras aleatórias geralmente serão diferentes, e é por isso que consideramos os estimadores de MQO como variáveis aleatórias.

Tabela 10.1

Lista Parcial de Dados sobre Taxas de Inflação e Desemprego nos Estados Unidos, de 1948 a 1999

Ano	Inflação	Desemprego
1948	8,1	3,8
1949	-1,2	5,9
1950	1,3	5,3
1951	7,9	3,3
:	:	:
1994	2,6	6,1
1995	2,8	5,6
1996	3,0	5,4
1997	2,3	4,9
1998	1,6	4,5
1999	2,2	4,2

Como devemos entender a aleatoriedade em dados de séries temporais? Certamente, as séries temporais econômicas satisfazem os requisitos intuitivos de serem resultados de variáveis aleatórias. Por exemplo, hoje não sabemos qual será a média do Índice Industrial Dow Jones no fechamento da próxima sessão. Não sabemos qual será a taxa de crescimento anual da produção do Canadá no ano que vem. Como os resultados dessas variáveis não são conhecidos previamente, elas devem ser claramente vistas como variáveis aleatórias.

Formalmente, uma sequência de variáveis aleatórias indexadas pelo tempo é chamada de **processo estocástico** ou **processo de série temporal**. ("Estocástico" é sinônimo de aleatório.) Quando coletamos um conjunto de dados de séries temporais, obtemos um resultado possível, ou *realização*, do processo estocástico. Só podemos ver uma única realização porque não podemos retornar no tempo e iniciar o processo novamente. (Isto é análogo a uma análise de corte transversal em que podemos coletar apenas uma amostra aleatória.) Entretanto, se determinadas condições históricas tivessem sido diferentes, em geral conseguíramos uma realização diferente para o processo estocástico, e por isso consideramos os dados de séries temporais como o resultado de variáveis aleatórias. O conjunto de todas as realizações possíveis de um processo de séries temporais faz o papel da população na análise de corte transversal. O tamanho da amostra de um conjunto de dados de séries temporais é o número de períodos nos quais observamos as variáveis de interesse.

10.2 EXEMPLOS DE MODELOS DE REGRESSÃO DE SÉRIES TEMPORAIS

Nesta seção discutiremos dois exemplos de modelos de séries de tempo que têm sido úteis na análise empírica e que são facilmente estimados por mínimos quadrados ordinários. No Capítulo 11 estudaremos mais modelos.

Modelos Estáticos

Suponha que temos dados de séries temporais disponíveis para duas variáveis, digamos y e z , onde y , e z , são datadas contemporaneamente. Um **modelo estático** que relaciona y a z é:

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad t = 1, 2, \dots, n. \quad (10.1)$$

O nome “modelo estático” deriva do fato de estarmos modelando uma relação contemporânea entre y e z . Em geral, um modelo estático é postulado quando se acredita que uma mudança em z no período t terá um efeito imediato em y : $\Delta y_t = \beta_1 \Delta z_t$, quando $\Delta u_t = 0$. Os modelos de regressão estáticos também são usados quando temos interesse em conhecer a relação de trocas entre y e z .

Um exemplo de modelo estático é a *curva de Phillips estática*, representada por:

$$inf_t = \beta_0 + \beta_1 desemp_t + u_t, \quad (10.2)$$

onde inf é a taxa de inflação anual e $desemp$ é a taxa de desemprego no mesmo período. Esta forma da curva de Phillips pressupõe uma *taxa natural de desemprego* e expectativas inflacionárias constantes, e ela pode ser usada para estudar a relação contemporânea entre inflação e desemprego. [Veja, por exemplo, Mankiw (1994, Seção 11.2)].

Naturalmente, podemos ter diversas variáveis explicativas em um modelo estático de regressão. Sejam $txhom_i$ o número de homicídios por 10.000 habitantes em uma determinada cidade durante o ano t , $txcond$ a taxa de condenações em homicídios, $desemp$ a taxa de desemprego local e $jovmasc$ a proporção da população masculina na faixa de 18 a 25 anos de idade. Portanto, um modelo estático de regressão múltipla que explica as taxas de homicídios é

$$txhom_i = \beta_0 + \beta_1 txcond_i + \beta_2 desemp_i + \beta_3 jovmasc_i + u_i. \quad (10.3)$$

Usando um modelo como este, poderemos estimar, por exemplo, o efeito *ceteris paribus* de um aumento na taxa de condenações sobre a atividade criminal.

Modelos de Defasagens Distribuídas Finitas

Em um **modelo de defasagens distribuídas finitas (DDF)**, permitimos que uma ou mais variáveis afetem y com defasagens. Por exemplo, para observações anuais considere o seguinte modelo:

$$tgf_t = \alpha_0 + \delta_0 ip_t + \delta_1 ip_{t-1} + \delta_2 ip_{t-2} + u_t, \quad (10.4)$$

onde tgf é a taxa geral de fertilidade (número de crianças nascidas por 1.000 mulheres em idade fértil) e ip é o valor real em dólares da isenção de impostos pessoais. A idéia é ver se, em termos agregados, a decisão de ter filhos está relacionada ao valor da isenção de impostos de ter uma criança. A equação (10.4) reconhece que, tanto por motivos biológicos como comportamentais, a decisão de ter filhos não resulta imediatamente de mudanças na isenção de impostos.

A equação (10.4) é um exemplo do modelo

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t, \quad (10.5)$$

que é um modelo DDF de *ordem dois*. Para interpretar os coeficientes em (10.5), suponha que z seja uma constante igual a c , em todos os períodos antes de t . Em t , z aumenta em uma unidade para $c + 1$ e, em seguida, retorna ao seu nível anterior em $t + 1$. (Isto é, o aumento em z é temporário.) Mais precisamente,

$$\dots, z_{t-2} = c, z_{t-1} = c, z_t = c + 1, z_{t+1} = c, z_{t+2} = c, \dots$$

Para enfatizar o efeito *ceteris paribus* de z sobre y , fazemos com que o termo erro em cada período seja zero. Então,

$$\begin{aligned} y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c, \\ y_t &= \alpha_0 + \delta_0(c + 1) + \delta_1 c + \delta_2 c, \\ y_{t+1} &= \alpha_0 + \delta_0 c + \delta_1(c + 1) + \delta_2 c, \\ y_{t+2} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2(c + 1), \\ y_{t+3} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c, \end{aligned}$$

e assim por diante. Das duas primeiras equações, $y_t - y_{t-1} = \delta_0$, mostra como δ_0 é a mudança imediata em y devido ao aumento de uma unidade em z no tempo t . δ_0 usualmente é chamada de **propensão de impacto** ou **multiplicador de impacto**.

Da mesma forma, $\delta_1 = y_{t+1} - y_{t-1}$ é a mudança em y um período após a mudança temporária, e $\delta_2 = y_{t+2} - y_{t-1}$ é a mudança em y dois períodos após a mudança. Em $t + 3$, y retornou ao seu nível inicial: $y_{t+3} = y_{t-1}$. Isso ocorre porque assumimos que apenas duas defasagens de z aparecem em (10.5). Quando traçamos um gráfico de δ_j como uma função de j , obtemos a **distribuição de defasagens**, que resume o efeito dinâmico que um aumento temporário em z tem em y . Uma possível distribuição de defasagens para o modelo DDF de ordem dois é apresentada na Figura 10.1. (Naturalmente, nunca conheceremos os parâmetros δ_j ; em vez disso, estimaremos o valor de δ_j e, em seguida, traçaremos a curva de distribuição de defasagens estimada.)

A distribuição de defasagens na Figura 10.1 implica que o efeito maior está na primeira defasagem. A distribuição de defasagens tem uma interpretação útil. Se padronizarmos o valor inicial de y em $y_{t-1} = 0$, a distribuição de defasagens define todos os valores subsequentes de y por causa de um aumento temporário, de uma unidade, em z .

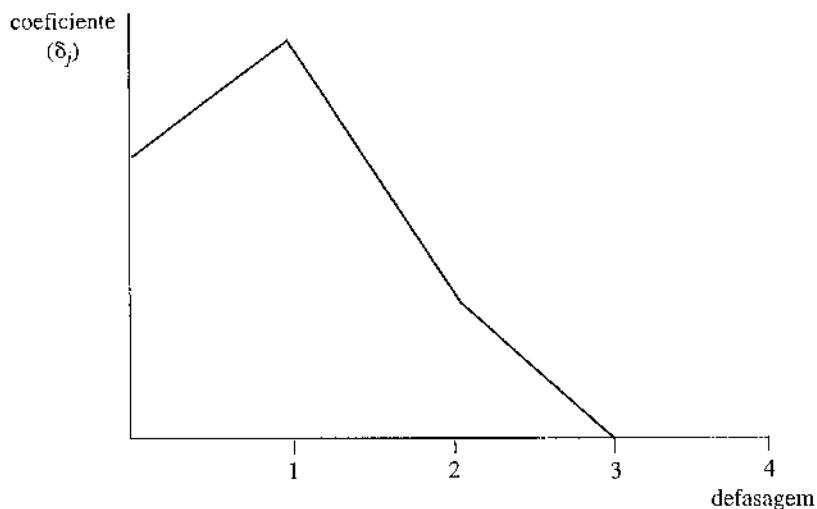
Estamos também interessados na mudança em y em virtude de um aumento *permanente* em z . Antes do tempo t , z é igual à constante c . Em t , z aumenta permanentemente para $c + 1$: $z_s = c$, $s < t$ e $z_s = c + 1$, $s \geq t$. Novamente, fazendo os erros serem iguais a zero, temos

$$\begin{aligned} y_{t-1} &= \alpha_0 + \delta_0 c + \delta_1 c + \delta_2 c, \\ y_t &= \alpha_0 + \delta_0(c + 1) + \delta_1 c + \delta_2 c, \\ y_{t+1} &= \alpha_0 + \delta_0(c + 1) + \delta_1(c + 1) + \delta_2 c, \\ y_{t+2} &= \alpha_0 + \delta_0(c + 1) + \delta_1(c + 1) + \delta_2(c + 1), \end{aligned}$$

e assim por diante. Com o aumento permanente em z , depois de um período, y aumentou $\delta_0 + \delta_1$, e depois de dois períodos, y aumentou $\delta_0 + \delta_1 + \delta_2$. Não há outras mudanças em y depois de dois períodos. Isso demonstra que a soma dos coeficientes das z atual e defasada, $\delta_0 + \delta_1 + \delta_2$, é a mudança de *longo prazo* em y quando há um aumento permanente em z e é chamada de **propensão de longo prazo** (PLP) ou **multiplicador de longo prazo**. A PLP é geralmente um indicador de grande interesse em modelos de defasagens distribuídas.

Figura 10.1

Uma distribuição de defasagens com duas defasagens diferentes de zero. O efeito máximo está na primeira defasagem.



Como exemplo, na equação (10.4), δ_0 mede a mudança imediata na fertilidade em virtude de um aumento de um dólar em ip . Como dissemos anteriormente, há razões para acreditar que δ_0 é pequeno, ou mesmo igual a zero. Porém, δ_1 , ou δ_2 , ou ambos, podem ser positivos. Se ip aumentar *permanentemente* em um dólar, então, depois de dois anos, tgf terá mudado em $\delta_0 + \delta_1 + \delta_2$. Este modelo presume que não há mais mudanças depois de dois anos. Saber se esse realmente é o caso, ou não, é uma questão empírica.

Um modelo de defasagens distribuídas finitas de ordem q é apresentado como:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \dots + \delta_q z_{t-q} + u_t. \quad (10.6)$$

A equação (10.6) contém o modelo estático como um caso especial fazendo $\delta_1, \delta_2, \dots, \delta_q = 0$. Às vezes, a principal finalidade de estimar um modelo de defasagens distribuídas finitas é testar se z tem um efeito defasado sobre y . A propensão de impacto é sempre o coeficiente de z contemporâneo, δ_0 . Ocionalmente, omitimos z_t de (10.6), e nesse caso a propensão de impacto é zero. A distribuição de defasagens é novamente δ_j grafado como uma função de j . A propensão de longo prazo é a soma de todos os coeficientes das variáveis z_{t-j} :

$$\text{PLP} = \delta_0 + \delta_1 + \dots + \delta_q. \quad (10.7)$$

Devido à correlação freqüentemente substancial em z em diferentes defasagens, ou seja, em virtude da multicolinearidade em (10.6), pode ser difícil conseguir estimativas individuais precisas de δ_j . É inte-

ressante o fato de que, mesmo quando os δ_j não podem ser estimados com precisão, geralmente podemos ter boas estimativas da PLP. Mais adiante veremos um exemplo.

Podemos ter mais de uma variável explicativa com defasagens ou podemos acrescentar variáveis contemporâneas a um modelo de DDF. Por exemplo, o nível de educação média de mulheres na idade fértil poderia ser adicionado à equação (10.4), o que nos permitiria levar em conta as mudanças nos níveis de educação das mulheres.

Em uma equação de dados anuais, supondo que

$$jur_t = 1,6 + 0,48 \inf_t - 0,15 \inf_{t-1} + 0,32 \inf_{t-2} + u_t$$

onde jur é uma taxa de juros e \inf é a taxa de inflação, quais são as propensões de impacto e de longo prazo?

Convenção sobre o Índice Temporal

Quando os modelos têm variáveis explicativas defasadas (e, como veremos no próximo capítulo, modelos com y defasado), pode surgir alguma confusão relativa ao tratamento das observações iniciais. Por exemplo, se em (10.5), assumirmos que a equação se mantém, iniciando em $t = 1$, então as variáveis explicativas para o primeiro período de tempo são z_1, z_0 e z_{-1} . Nossa convenção será que esses são os valores iniciais de nossa amostra, de forma que sempre poderemos iniciar o índice temporal em $t = 1$. Na prática, isso não é muito importante, pois os programas de regressões automaticamente acompanham as observações disponíveis para estimar modelos com defasagens. Porém, para este e para os próximos dois capítulos, precisamos de alguma convenção relativa ao primeiro período de tempo que está representado pela equação de regressão.

10.3 PROPRIEDADES DE AMOSTRA FINITA DO MQO SOB AS HIPÓTESES CLÁSSICAS

Nesta seção apresentamos uma relação completa das propriedades de amostra finita, ou de pequena amostra, do MQO sob as hipóteses padrão. Prestaremos atenção especial em como as hipóteses devem ser alteradas a partir de nossas análises de corte transversal para cobrir regressões de séries temporais.

Inexistência de Viés do MQO

A primeira hipótese simplesmente estabelece que o processo da série temporal segue um modelo linear em seus parâmetros.

H I P Ó T E S E S T . 1 (LINEAR NOS PARÂMETROS)

O processo estocástico $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t); t = 1, 2, \dots, n\}$ segue o modelo linear

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t, \quad (10.8)$$

onde $\{u_t; t = 1, 2, \dots, n\}$ é a seqüência de erros ou disturbâncias. Aqui, n é o número de observações (períodos de tempo).

Tabela 10.2

Exemplo de X para as Variáveis Explicativas na Equação (10.3)

<i>t</i>	<i>txcond</i>	<i>desemp</i>	<i>jovmasc</i>
1	0,46	0,074	0,12
2	0,42	0,071	0,12
3	0,42	0,063	0,11
4	0,47	0,062	0,09
5	0,48	0,060	0,10
6	0,50	0,059	0,11
7	0,55	0,058	0,12
8	0,56	0,059	0,13

Na notação x_{ij} , t representa o período de tempo e j é, como sempre, um indicador de uma das k variáveis explicativas. A terminologia usada em regressões de corte transversal se aplica neste caso: y_t é a variável dependente, variável explicada ou regressando; as x_{ij} são as variáveis independentes, variáveis explicativas ou regressores.

Devemos considerar a hipótese ST.1 como essencialmente igual à hipótese RLM.1 (a primeira hipótese da análise de corte transversal), porém agora estamos especificando um modelo linear para dados de séries temporais. Os exemplos da Seção 10.2 podem ser descritos na forma de (10.8) definindo x_{ij} apropriadamente. Por exemplo, a equação (10.5) é obtida definindo-se $x_{t1} = z_t$, $x_{t2} = z_{t-1}$ e $x_{t3} = z_{t-2}$.

Para estabelecer e discutir muitas das hipóteses restantes, assumimos que $x_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ representa o conjunto de todas as variáveis independentes da equação em t . Além disso, X representa o conjunto de todas as variáveis independentes de todos os períodos de tempo. É útil entender que X é um arranjo de n linhas e k colunas. Isso mostra como os dados de séries temporais são armazenados em programas de econometria: a t -ésima linha de X é x_t , formada por todas as variáveis independentes do período de tempo t . Portanto, a primeira linha de X corresponde a $t = 1$, a segunda a $t = 2$ e a última a $t = n$. Um exemplo é dado na Tabela 10.2 com $n = 8$ e as variáveis explicativas na equação (10.3).

A próxima hipótese é a análoga (das séries de tempo) da hipótese RLM.3, e ela também elimina a hipótese de amostragem aleatória na hipótese RLM.2.

H I P Ó T E S E S T . 2 (MÉDIA CONDICIONAL ZERO)

Para cada t , o valor esperado do erro u_t , dadas as variáveis explicativas de todos os períodos de tempo, é zero. Matematicamente,

$$E(u_t | X) = 0, t = 1, 2, \dots, n. \quad (10.9)$$

Essa é uma hipótese crucial, e precisamos ter uma compreensão intuitiva de seu significado. Como no caso do corte transversal, é bem mais fácil ver a hipótese em termos de não-correlação: a hipótese ST.2 implica que o erro no tempo t , u_t , é não-correlacionado com cada variável explicativa em *todos* os períodos de tempo. O fato de isso ser colocado em termos de expectativa condicional significa que devemos também especificar corretamente a relação funcional entre y_t e as variáveis explicativas. Se u_t for independente de X e $E(u_t) = 0$, a hipótese ST.2 se mantém automaticamente.

Dada a análise de corte transversal do Capítulo 3, não é surpresa que exijamos que u_t seja não-correlacionado com as variáveis explicativas também datadas no tempo t : em termos de média condicional,

$$E(u_t | x_{t1}, \dots, x_{tk}) = E(u_t | x_t) = 0. \quad (10.10)$$

Quando (10.10) se mantém, dizemos que os x_{ij} são **contemporaneamente exógenos**. A equação (10.10) implica que u_t e as variáveis explicativas são contemporaneamente não-correlacionadas: $\text{Corr}(x_{ij}, u_t) = 0$, para todo j .

A hipótese ST.2 exige mais que exogeneidade contemporânea: u_t deve ser não-correlacionado com x_{sj} , mesmo quando $s \neq t$. Esse é o sentido de como as variáveis explicativas devem ser exógenas, e quando ST.2 se mantém, dizemos que as variáveis explicativas são **estritamente exógenas**. No Capítulo 11 demonstraremos que (10.10) é suficiente para provar a consistência do estimador de MQO. Porém, para mostrar que o MQO é não-viesado, precisamos da hipótese de exogeneidade estrita.

No caso do corte transversal, não declaramos explicitamente como o termo erro da, digamos, pessoa i , u_i , está relacionado com as variáveis explicativas de *outras* pessoas na amostra. O motivo de isso ser desnecessário é que, com amostragem aleatória (hipótese RLM.2), u_i é *automaticamente* independente das variáveis explicativas para outras observações além de i . Em um contexto de séries de tempo, a amostragem aleatória quase nunca é apropriada, de modo que devemos assumir explicitamente que o valor esperado de u_t não é relacionado com as variáveis explicativas em qualquer período de tempo.

É importante ver que a hipótese ST.2 não restringe a correlação nas variáveis independentes ou em u_t ao longo do tempo. A hipótese ST.2 apenas diz que o valor médio de u_t é não-relacionado com as variáveis independentes em qualquer período de tempo.

Qualquer elemento que leve os fatores não-observáveis no tempo t a serem correlacionados com qualquer das variáveis explicativas, em qualquer período de tempo, fará com que a hipótese ST.2 seja violada. Dois importantes candidatos para essa violação são variáveis omitidas e erros de medida em alguns dos regressores. Porém, a hipótese de exogeneidade estrita também pode ser violada, em virtude de outras razões menos óbvias. No modelo estático de regressão simples

$$y_t = \beta_0 + \beta_1 z_t + u_t$$

a hipótese ST.2 exige não apenas que u_t e z_t sejam não-correlacionados, mas que u_t também seja não-correlacionado com valores passados e futuros de z . Isso tem duas implicações. A primeira é que z pode não ter nenhum efeito defasado sobre y . Se z tiver um efeito defasado sobre y , devemos, então, estimar um modelo de defasagens distribuídas. Um aspecto mais sutil é que a exogeneidade estrita exclui a possibilidade de que mudanças no termo erro, hoje, possam provocar alterações futuras em z . Isso efetivamente impede uma realimentação de y sobre futuros valores de z . Por exemplo, considere um modelo estático simples para explicar uma taxa de homicídios em uma cidade em termos de número de policiais *per capita*:

$$txhom_i = \beta_0 + \beta_1 polpc_i + u_i$$

Pode ser razoável assumirmos que u_i seja não-correlacionado com $polpc_i$, e mesmo com valores passados de $polpc_i$; para fins de argumentação, consideremos que esse seja o caso. Porém, suponhamos que a cidade ajuste o tamanho de sua força policial baseada em valores passados da taxa de homicídios. Isso significa que, digamos, $polpc_{i+1}$ poderia estar correlacionado com u_i (já que um u_i maior resulta em $txhom_i$ mais alto). Se este for o caso, a hipótese ST.2 geralmente é violada.

Existem considerações semelhantes nos modelos de defasagens distribuídas. Geralmente, não nos preocupamos que u_t possa ser correlacionado com z passados porque estamos controlando z passados no modelo. Porém, a realimentação de u para z futuros é sempre um problema.

As variáveis explicativas que são estritamente exógenas não podem reagir ao que aconteceu com y no passado. Um fator como o volume de chuva em uma função de produção agrícola satisfaz este requisito: a chuva em qualquer ano futuro não será influenciada pelo ocorrido neste ano ou em anos passados. Entretanto, algo como a quantidade da força de trabalho pode não ser estritamente exógeno, já que é escolhido pelo fazendeiro, e este pode ajustar a força de trabalho com base na produção do ano anterior. Variáveis de políticas governamentais, como o crescimento na oferta de moeda, gastos com assistência social e limites de velocidade nas rodovias são, muitas vezes, influenciadas pelo que tenha acontecido com a variável no passado. Nas ciências sociais, muitas variáveis explicativas podem muito bem violar a hipótese de exogeneidade estrita.

Embora a hipótese ST.2 possa não ser realista, iniciamos com ela o propósito de concluir que os estimadores MQO são não-viesados. A maioria dos tratamentos de modelos estáticos e de defasagens distribuídas finitas admite ST.2 elaborando a hipótese mais forte de que as variáveis explicativas são não-aleatórias, ou fixas, em amostras repetidas. A hipótese de não-aleatoriedade é obviamente falsa para observações de séries temporais; a hipótese ST.2 tem a vantagem de ser mais realista a respeito da natureza aleatória de x_{ij} , ao mesmo tempo em que isola a hipótese necessária sobre como u_t e as variáveis explicativas são relacionadas de maneira que os MQO sejam não-viesados.

A última hipótese necessária para a não-existência de viés do MQO é a hipótese padrão de inexistência de colinearidade perfeita.

H I P Ó T E S E S T . 3 (INEXISTÊNCIA DE COLINEARIDADE PERFEITA)

Na amostra (e portanto no processo subjacente da série temporal), nenhuma variável independente é constante ou é uma combinação linear perfeita das outras.

Discutimos esta hipótese detalhadamente no contexto de dados de corte transversal no Capítulo 3. As questões são essencialmente as mesmas quando trabalhamos com séries de tempo. Lembre-se, a hipótese ST.3 permite que as variáveis explicativas sejam correlacionadas, mas não permite a correlação *perfeita* na amostra.

T E O R E M A 10 . 1 (INEXISTÊNCIA DE VIÉS DO MQO)

Sob as hipóteses ST.1, ST.2 e ST.3 os estimadores de MQO são não-viesados condicionados em X e, portanto, também incondicionalmente: $E(\hat{\beta}_j) = \beta_j, j = 0, 1, \dots, k$.

A prova deste teorema é essencialmente a mesma do Teorema 3.1 no Capítulo 3, de modo que a omitimos. Ao compararmos o Teorema 10.1 ao Teorema 3.1 tivemos condições de eliminar a hipótese de amostragem aleatória ao assumirmos que, para cada t , u_t tem média zero, dadas as variáveis explicativas em todos os períodos de tempo. Se esta hipótese não se sustentar, os estimadores MQO não poderão ser não-viesados.

A análise de viés de variáveis omitidas, que tratamos na Seção 3.3, é essencialmente a mesma no caso de séries temporais. Particularmente, a Tabela 3.2 e a discussão em torno dela podem ser usadas para determinar as direções do viés devido às variáveis omitidas.

No modelo DDF $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + u_t$, o que precisamos assumir sobre a seqüência $\{z_0, z_1, \dots, z_n\}$ para que a hipótese ST.1 seja válida?

As Variâncias dos Estimadores MQO e o Teorema de Gauss-Markov

Necessitamos adicionar duas hipóteses para completar as hipóteses de Gauss-Markov para regressões de séries temporais. A primeira é familiar, da análise de corte transversal.

H I P Ó T E S E S T . 4 (HOMOSCEDASTICIDADE)

Condisional em X , a variância de u_t é a mesma para todo t : $\text{Var}(u_t | X) = \text{Var}(u_t) = \sigma^2, t = 1, 2, \dots, n$.

Esta hipótese significa que $\text{Var}(u_t | X)$ não pode depender de X — é suficiente que u_t e X sejam independentes — e que $\text{Var}(u_t)$ seja constante ao longo do tempo. Quando ST.4 não se mantém, dizemos que os erros são *heteroscedásticos*, exatamente como no caso da análise de corte transversal. Por exemplo, considere uma equação para determinar as taxas de títulos do Tesouro norte-americano de três meses ($i3_t$), com base na taxa de inflação (inf_t) e no déficit federal como uma percentagem do produto interno bruto (def_t):

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t. \quad (10.11)$$

Entre outras coisas, a hipótese ST.4 exige dos fatores não-observáveis que estejam afetando a taxa de juros uma variância constante ao longo do tempo. Como sabemos que as alterações da política econômica afetam a variação das taxas de juros, essa hipótese pode muito bem ser falsa. Além disso, pode ser que a variação nas taxas de juros dependa do nível de inflação ou do tamanho relativo do déficit. Isso também violaria a hipótese de homoscedasticidade.

Quando $\text{Var}(u_t | X)$ depende de X , ela freqüentemente depende das variáveis explicativas no momento t , x_t . No Capítulo 12 veremos que os testes para verificar a existência de heteroscedasticidade do Capítulo 8 também podem ser usados nas regressões de séries temporais, pelo menos sob certas hipóteses.

A última hipótese de Gauss-Markov de análise de séries temporais é nova.

H I P Ó T E S E S T . 5 (INEXISTÊNCIA DE CORRELAÇÃO SERIAL)

Condisional em X , os erros em dois períodos de tempo diferentes são não-correlacionados: $\text{Corr}(u_t, u_s | X) = 0$, para todo $t \neq s$.

A maneira mais fácil de entender esta hipótese é ignorar a condicionalidade em X . Então, a hipótese ST.5 passa a ser simplesmente

$$\text{Corr}(u_t, u_s) = 0, \text{ para } t \neq s. \quad (10.12)$$

(Esta é a forma como a hipótese de inexistência de correlação serial é definida quando X é tratada como não-aleatória.) Quando consideramos se a hipótese ST.5 tem possibilidades de se sustentar, nós nos concentramos na equação (10.12) devido a sua interpretação simples.

Quando (10.12) for falsa, dizemos que os erros em (10.8) sofrem de **correlação serial**, ou **autocorrelação**, porque são correlacionados ao longo do tempo. Considere o caso de erros dos períodos de tempo adjacentes. Suponha que, quando $u_{t-1} > 0$, na média, o erro no próximo período de tempo, u_t , também seja positivo. Então, $\text{Corr}(u_t, u_{t-1}) > 0$, e os erros sofrem de correlação serial. Na equação (10.11), isso significa que, se as taxas de juros forem inesperadamente altas para esse período, elas estarão propensas a ficar acima da média (dos níveis dados de inflação e déficit) do próximo período. Isso acaba sendo uma razoável caracterização dos termos erros em muitas aplicações de séries temporais, que veremos no Capítulo 12. Por enquanto, assumiremos a hipótese ST.5.

É importante observar que a hipótese ST.5 não presume nada sobre correlação temporal nas variáveis *independentes*. Por exemplo, na equação (10.11), inf , é, quase com certeza, correlacionada ao longo do tempo. Porém, isso não tem nada a ver com o fato de a hipótese ST.5 se manter ou não.

Uma questão natural que surge é: nos Capítulos 3 e 4, por que não assumimos que os erros para diferentes observações de corte transversal eram não-correlacionados? A resposta vem da hipótese da amostragem aleatória: sob amostragem aleatória, u_i e u_h são independentes para quaisquer duas observações i e h . Também pode ser mostrado que, sob amostragem aleatória, os erros para observações diferentes são independentes, condicionais às variáveis explicativas na amostra. Assim, para nosso propósito, consideraremos que a correlação serial somente será um problema potencial nas regressões com dados de séries temporais. (Nos Capítulos 13 e 14, o problema da correlação serial surgirá novamente, em conexão com a análise de dados de painel.)

As hipóteses ST.1 a ST.5 são as hipóteses de Gauss-Markov apropriadas para aplicações de séries temporais, mas elas também podem ser usadas para outras finalidades. Algumas vezes as hipóteses ST.1 a ST.5 são satisfeitas em aplicações de corte transversal, mesmo quando a amostragem aleatória

não é uma hipótese razoável, como quando as unidades de corte transversal são grandes em relação à população. Suponha que tenhamos um conjunto de dados de corte transversal ao nível de cidades. Pode ser que exista correlação entre cidades do mesmo Estado, para algumas das variáveis explicativas, como impostos sobre a propriedade ou despesas de assistência social *per capita*. A correlação das variáveis explicativas ao longo das observações não causa problema para verificarmos as hipóteses de Gauss-Markov, desde que os termos erros sejam não-correlacionados entre as cidades. Porém, neste capítulo estamos, antes de mais nada, interessados em aplicar as hipóteses de Gauss-Markov nos problemas de regressão de séries temporais.

T E O R E M A 10 . 2 (VARIÂNCIAS AMOSTRAIS DO MQO)

Sob as hipóteses de séries temporais ST.1 a ST.5 de Gauss-Markov, a variância de $\hat{\beta}_j$, condicional em X , é

$$\text{Var}(\hat{\beta}_j|X) = \sigma^2 / [\text{SQT}_j(1 - R_j^2)], j = 1, \dots, k, \quad (10.13)$$

onde SQT_j é a soma dos quadrados total de x_{ij} e R_j^2 é o *R*-quadrado da regressão de x_j sobre as outras variáveis independentes.

A equação (10.13) é exatamente a mesma variância que derivamos no Capítulo 3 sob as hipóteses de Gauss-Markov de corte transversal. Como sua prova é muito semelhante à do Teorema 3.2, ela será omitida. A discussão do Capítulo 3 sobre os fatores causadores de grandes variâncias, inclusive a multicolinearidade entre as variáveis explicativas, aplica-se diretamente ao caso de séries temporais.

O estimador usual da variância do erro também é não-viesado sob as hipóteses ST.1 a ST.5, e o Teorema de Gauss-Markov se mantém.

T E O R E M A 10 . 3 (ESTIMAÇÃO NÃO-VIESADA DE σ^2)

Sob as hipóteses ST.1 a ST.5, o estimador $\hat{\sigma}^2 = \text{SQR}/g$ é um estimador não-viesado de σ^2 , onde $g = n - k - 1$.

T E O R E M A 10 . 4 (TEOREMA DE GAUSS-MARKOV)

Sob as hipóteses ST.1 a ST.5, os estimadores MQO são os melhores estimadores lineares não-viesados condicionais em X .

O ponto principal é que o MQO tem as mesmas propriedades de amostra finita desejáveis sob as hipóteses ST.1 a ST.5 que ele tem sob as hipóteses RLM.1 a RLM.5.

No modelo DDF $y_t = \alpha_0 + \delta_0 z_r + \delta_1 z_t + u_t$, explique a natureza de qualquer multicolinearidade nas variáveis explicativas.

Inferência sob as Hipóteses do Modelo Linear Clássico

Para podermos usar os usuais erros-padrão, estatísticas t e estatísticas F do método MQO, precisamos adicionar uma última hipótese que é análoga à hipótese de normalidade usada para a análise de corte transversal.

H I P Ó T E S E S T . 6 (NORMALIDADE)

Os erros u_t são independentes de X e são idêntica e independentemente distribuídos como $\text{Normal}(0, \sigma^2)$.

A hipótese ST.6 implica ST.2, ST.4 e ST.5, mas é mais forte devido às hipóteses de independência e normalidade.

T E O R E M A 10 . 5 (DISTRIBUIÇÕES AMOSTRAIS NORMAIS)

Sob as hipóteses ST.1 a ST.6, as hipóteses MLC para as séries temporais, os estimadores MQO são normalmente distribuídos, condicionais em X . Além disso, sob a hipótese nula, cada estatística t tem uma distribuição t , e cada estatística F tem uma distribuição F . A construção habitual de intervalos de confiança também é válida.

As implicações do teorema 10.5 são da maior importância. Quando as hipóteses ST.1 a ST.6 se sustentam, tudo que aprendemos sobre estimadores e inferência das regressões de corte transversal aplica-se diretamente às regressões de séries temporais. Assim, estatísticas t podem ser usadas para testar a significância estatística de variáveis explicativas individuais, e estatísticas F podem ser usadas para testar a significância conjunta.

Da mesma forma que no caso da análise de corte transversal, os procedimentos normais de inferência serão, apenas, tão bons quanto os das hipóteses subjacentes. As hipóteses do modelo linear clássico para dados de séries temporais são muito mais restritivas que para dados de corte transversal — particularmente, as hipóteses de exogeneidade estrita e de inexistência de correlação serial podem ser irreais. Todavia, a estrutura MLC é um bom ponto de partida para muitas aplicações.

EXEMPLO 10.1

(A Curva de Phillips Estática)

Para determinar se existe uma relação, em média, entre desemprego e inflação, podemos testar $H_0: \beta_1 = 0$ contra $H_1: \beta_1 < 0$ na equação (10.2). Se as hipóteses do modelo linear clássico se mantiverem, podemos

EXEMPLO 10.1 (continuação)

usar a estatística t usual do MQO. Utilizando os dados anuais dos Estados Unidos contidos no arquivo PHILIPS.RAW, para os anos de 1948 a 1996, obtemos

$$\begin{aligned} \hat{inf}_t &= 1,42 + 0,468 \text{ desemp}_t \\ &\quad (1,72) \quad (0,289) \\ n &= 49, R^2 = 0,053, \bar{R}^2 = 0,033. \end{aligned} \tag{10.14}$$

Esta equação não sugere uma relação de substituição entre desemp e inf , pois $\hat{\beta}_1 > 0$. A estatística t de $\hat{\beta}_1$ está em torno de 1,62, o que dá um p -valor contra a alternativa bilateral de cerca de 0,11. Portanto, na verdade, existe um relacionamento positivo entre inflação e desemprego.

Existem alguns problemas com esta análise que não podemos avaliar em detalhes agora. No Capítulo 12 veremos que as hipóteses MLC não permanecem. Além disso, a curva de Phillips estática talvez não seja o melhor modelo para determinar se existe uma relação de substituição de curto prazo entre inflação e desemprego. Os macroeconomistas geralmente preferem a curva de Phillips de expectativas aumentadas, da qual será dado um exemplo simples no Capítulo 11.

Como um segundo exemplo, estimamos a equação (10.11) utilizando os dados anuais da economia dos Estados Unidos.

EXEMPLO 10.2**(Efeitos da Inflação e dos Déficits sobre as Taxas de Juros)**

Os dados do arquivo INTDEF.RAW provêm do Relatório Econômico da Presidência (REP) dos Estados Unidos para 1997 e abrangem os anos de 1948 a 1996. A variável $i3$ é a taxa de juros de títulos do Tesouro norte-americano de três meses, inf é a taxa anual de inflação baseada no índice de preços ao consumidor (IPC) e def é o déficit orçamentário federal como uma percentagem do PIB. A equação estimada é

$$\begin{aligned} \hat{i3}_t &= 1,25 + 0,613 \text{ inf}_t + 0,700 \text{ def}_t \\ &\quad (0,44) \quad (0,076) \quad (0,118) \\ n &= 49, R^2 = 0,697, \bar{R}^2 = 0,683. \end{aligned} \tag{10.15}$$

Essas estimativas mostram que aumentos na inflação e o tamanho relativo do déficit trabalham juntos para aumentar as taxas de juros de curto prazo, o que é esperado a partir dos fundamentos da teoria econômica. Por exemplo, um aumento de um ponto percentual *ceteris paribus* na taxa de inflação aumenta $i3$ em 0,613 pontos. Tanto inf como def são estatisticamente bastante significantes, supondo, é claro, que as hipóteses MLC se mantenham.

10.4 FORMA FUNCIONAL, VARIÁVEIS DUMMY E NÚMEROS-ÍNDICES

Todas as formas funcionais que aprendemos nos capítulos anteriores podem ser usadas nas regressões de séries de tempo. A mais importante delas é o logaritmo natural: regressões de séries temporais com efeitos percentuais constantes aparecem com freqüência no trabalho aplicado.

EXEMPLO 10.3

(Emprego e Salário Mínimo em Porto Rico)

Dados anuais sobre a taxa de emprego, salário mínimo e outras variáveis são usados por Castillo-Freeman e Freeman (1992) para estudar os efeitos do salário mínimo norte-americano sobre o emprego em Porto Rico. Uma versão simplificada do modelo deles é

$$\log(\text{prepop}_t) = \beta_0 + \beta_1 \log(\text{mincov})_t + \beta_2 \log(\text{uspnb})_t + u_t, \quad (10.16)$$

onde prepop_t é a taxa de emprego em Porto Rico durante o ano t (razão entre as pessoas com trabalho e a população total), uspnb_t é o produto nacional bruto real dos Estados Unidos (em bilhões de dólares) e mincov mede a importância do salário mínimo em relação aos salários médios. Especificamente, $\text{mincov} = (\text{minmed}/\text{salmed}) \cdot \text{cobmed}$, onde minmed é o salário mínimo médio, salmed é o salário médio total e cobmed é a taxa média de cobertura (a proporção dos trabalhadores efetivamente cobertos pela lei do salário mínimo).

Utilizando dados dos anos de 1950 a 1987 obtemos

$$\begin{aligned} \log(\text{prepop}_t) &= -1,05 - 0,154 \log(\text{mincov}_t) - 0,012 \log(\text{uspnb}_t) \\ &\quad (0,77) \quad (0,065) \quad (0,089) \\ n &= 38, R^2 = 0,661, \bar{R}^2 = 0,641. \end{aligned} \quad (10.17)$$

A elasticidade estimada de prepop em relação a mincov é $-0,154$, e é estatisticamente significante com $t = -2,37$. Portanto, um salário mínimo maior reduz a taxa de emprego, algo que é previsto pela teoria econômica clássica. A variável do PNB não é estatisticamente significante, mas isso mudará quando levarmos em conta uma tendência temporal na próxima seção.

Podemos também usar formas funcionais logarítmicas em modelos de defasagens distribuídas. Por exemplo, para dados trimestrais, suponha que a demanda por moeda (M_t) e o produto interno bruto (PIB_t) sejam relacionados por

$$\begin{aligned} \log(M_t) &= \alpha_0 + \delta_0 \log(PIB_t) + \delta_1 \log(PIB_{t-1}) + \delta_2 \log(PIB_{t-2}) \\ &\quad + \delta_3 \log(PIB_{t-3}) + \delta_4 \log(PIB_{t-4}) + u_t. \end{aligned}$$

A propensão de impacto na equação, δ_0 , também é chamada de **elasticidade de curto prazo**: ela mede a mudança percentual imediata na demanda por moeda, dado um aumento de 1% no PIB . A propensão de longo prazo, $\delta_0 + \delta_1 + \dots + \delta_4$, é, algumas vezes, chamada de **elasticidade de longo prazo**: ela mede o aumento percentual na demanda por moeda após quatro trimestres, dado um aumento permanente de 1% no PIB .

Variáveis independentes binárias ou *dummy* também são bastante úteis em aplicações de séries temporais. Como a unidade de observação é o tempo, uma variável *dummy* mostra, em cada período de tempo, se ocorreu determinado evento. Por exemplo, para dados anuais, podemos indicar em cada

ano se um Democrata ou um Republicano é o presidente dos Estados Unidos definindo a variável $democ$, que será um se o presidente for um Democrata e zero, caso contrário. Ou, examinando os efeitos da pena capital sobre as taxas de homicídios no estado norte-americano do Texas, podemos definir uma variável *dummy* para cada ano, igual a um se o Texas tiver pena capital durante determinado ano e zero, caso contrário.

Muitas vezes, variáveis *dummy* são usadas para isoler certos períodos que possam ser sistematicamente diferentes de outros períodos cobertos pelo conjunto de dados.

EXEMPLO 10.4

(Efeitos da Isenção de Impostos nas Taxas de Fertilidade)

A taxa geral de fertilidade (tgf) é o número de crianças nascidas para cada 1.000 mulheres em idade fértil. Para os anos de 1913 a 1984, a equação

$$tgf_t = \beta_0 + \beta_1 ip_t + \beta_2 ww2_t + \beta_3 pill_t + u_t,$$

explica tgf em termos do valor médio real em dólares da taxa de isenção de impostos (ip) e duas variáveis binárias. A variável $ww2$ recebe o valor um durante os anos de 1941 a 1945, quando os Estados Unidos estavam envolvidos na Segunda Guerra Mundial. A variável $pill$ terá valor um de 1963 em diante, quando a pílula anticoncepcional foi disponibilizada para o controle da natalidade.

Utilizando os dados contidos no arquivo FERTIL3.RAW, que foram retirados do artigo escrito por Whittington, Alm e Peters (1990), temos

$$\begin{aligned} \hat{tgf}_t &= 98,68 + 0,083 ip_t - 24,24 ww2_t - 31,59 pill_t, \\ (3,21) &\quad (0,030) \quad (7,46) \quad (4,08) \\ n &= 72, R^2 = 0,473, \bar{R}^2 = 0,450. \end{aligned} \tag{10.18}$$

Cada variável é estatisticamente significante ao nível de 1%, contra uma alternativa bilateral. Vemos que a taxa de fertilidade era baixa durante a Segunda Guerra Mundial: dado ip , houve cerca de 24 nascimentos a menos para cada 1.000 mulheres em idade fértil, o que é uma grande redução. (De 1913 a 1984, a tgf oscilou entre 65 e 127.) De maneira semelhante, a taxa de fertilidade tem sido substancialmente mais baixa desde o lançamento da pílula anticoncepcional.

A variável de interesse econômico é ip . A ip média nesse período de tempo é US\$ 100,40, variando de zero a US\$ 243,83. O coeficiente de ip sugere que um aumento de 12 dólares em ip aumenta tgf em cerca de um nascimento para cada 1.000 mulheres em idade fértil. Este efeito não é nada comum.

Na Seção 10.2, observamos que a taxa de fertilidade pode reagir a mudanças em ip com uma defasagem. Estimando um modelo de defasagens distribuídas com duas defasagens, obtemos

$$\begin{aligned} \hat{tgf}_t &= 95,87 + 0,073 ip_t - 0,0058 ip_{t-1} + 0,034 ip_{t-2} \\ (3,28) &\quad (0,126) \quad (0,1557) \quad (0,126) \\ &\quad - 22,13 ww2_t - 31,30 pill_t, \\ (10,73) &\quad (3,98) \\ n &= 70, R^2 = 0,499, \bar{R}^2 = 0,459. \end{aligned} \tag{10.19}$$

EXEMPLO 10.4 (continuação)

Nesta regressão, temos somente 70 observações porque perdemos duas quando fizemos a defasagem de ip duas vezes. Os coeficientes das variáveis ip são estimados de forma bastante imprecisa, e cada um deles é individualmente não significante. Constatase que existe substancial correlação entre ip_t , ip_{t-1} e ip_{t-2} , e essa multicolinearidade torna difícil estimar o efeito em cada defasagem. Porém, ip_t , ip_{t-1} e ip_{t-2} são conjuntamente significantes: a estatística F tem um p -valor = 0,012. Assim, ip tem um efeito em tgf [como já vimos na equação (10.18)], mas não temos estimativas suficientemente boas para determinar se ela é contemporânea ou tem uma ou duas defasagens (ou um pouco de cada uma). Efetivamente, ip_{t-1} e ip_{t-2} , são conjuntamente não significantes nesta equação (p -valor = 0,95), de forma que, neste ponto, seria justificável usarmos o modelo estático. Entretanto, para fins ilustrativos, vamos obter um intervalo de confiança da propensão de longo prazo neste modelo.

A PLP estimada em (10.19) é $0,073 - 0,0058 + 0,034 \approx 0,101$. Porém, não temos informação suficiente em (10.19) para obter o erro-padrão desta estimativa. Para obter o erro-padrão da PLP estimada, usamos o truque sugerido na Seção 4.4. Suponha que $\theta_0 = \delta_0 + \delta_1 + \delta_2$ é a PLP e escreva δ_0 em termos de θ_0 , δ_1 e δ_2 como $\delta_0 = \theta_0 - \delta_1 - \delta_2$. A seguir, substitua δ_0 no modelo

$$tgf_t = \alpha_0 + \delta_0 ip_t + \delta_1 ip_{t-1} + \delta_2 ip_{t-2} + \dots$$

para obter

$$\begin{aligned} tgf_t &= \alpha_0 + (\theta_0 - \delta_1 - \delta_2) ip_t + \delta_1 ip_{t-1} + \delta_2 ip_{t-2} + \dots \\ &= \alpha_0 + \theta_0 ip_t + \delta_1 (ip_{t-1} - ip_t) + \delta_2 (ip_{t-2} - ip_t) + \dots \end{aligned}$$

Desta última equação, podemos obter $\hat{\theta}_0$ e seu erro-padrão fazendo a regressão de tgf_t sobre ip_t , $(ip_{t-1} - ip_t)$, $(ip_{t-2} - ip_t)$, $ww2_t$ e $pill_t$. O coeficiente e o erro-padrão associado de ip_t são os elementos de que precisamos. Computando esta regressão obtemos $\hat{\theta}_0 = 0,101$ como o coeficiente de ip_t (como já sabíamos) e $ep(\hat{\theta}_0) = 0,030$ [que não podíamos calcular de (10.19)]. Portanto, a estatística t de $\hat{\theta}_0$ está em torno de 3,37, de forma que $\hat{\theta}_0$ é estatisticamente diferente de zero a níveis reduzidos de significância. Embora nenhum dos $\hat{\delta}_j$ seja individualmente significante, a PLP é muito significante. O intervalo de confiança de 95% da PLP está entre 0,041 e 0,160.

Whittington, Alm e Peters (1990) permitem defasagens adicionais, mas restringem os coeficientes para ajudar a aliviar o problema de multicolinearidade que dificulta a estimação dos δ_j individuais. (Veja o Problema 10.6 para um exemplo de como fazer isso.) Para estimar a PLP, que pareceria ser de principal interesse neste caso, tais restrições são desnecessárias. Whittington, Alm e Peters também mantêm controle das variáveis adicionais, como o salário médio das mulheres e a taxa de desemprego.

As variáveis explicativas binárias são as principais componentes do que é chamado de **estudo de evento**. Em um estudo de evento, o objetivo é ver se determinado evento influencia algum resultado. Economistas que estudam a organização industrial têm analisado os efeitos de certos eventos sobre os preços das ações das empresas. Por exemplo, Rose (1985) estudou os efeitos da nova regulamentação do transporte rodoviário sobre os preços das ações das transportadoras.

Uma versão simplificada de uma equação usada para tais estudos de evento é

$$R_t^F = \beta_0 + \beta_1 R_t''' + \beta_2 d_t + u_t$$

onde R_t^f é o retorno das ações da empresa f durante o período t (normalmente uma semana ou um mês), R_t^m é o retorno de mercado (normalmente computado como um índice amplo do mercado de ações) e d_t é uma variável *dummy* indicando quando ocorreu o evento. Por exemplo, se a empresa for de transporte aéreo, d_t pode mostrar se ela passou pela experiência de algum acidente, ou quase acidente, ter sido divulgado na semana t . A inclusão de R_t^m na equação controla a possibilidade de que movimentos grandes no mercado de ações possam coincidir com os acidentes da empresa. Algumas vezes são usadas variáveis *dummy* múltiplas. Por exemplo, se o evento for a imposição de uma nova regulamentação que possa afetar determinada empresa, podemos incluir uma variável *dummy* que assuma o valor um algumas semanas antes de a regulamentação ter sido publicada e uma segunda variável simulada para algumas semanas após a publicação da regulamentação. A primeira variável *dummy* pode detectar a presença de informações privilegiadas.

Antes de exemplificarmos um estudo de evento, precisamos discutir a noção de **número-índice** e a diferença entre variáveis econômicas nominais e reais. Um número-índice tipicamente agrega um vasto grau de informações em uma única quantidade. Números-índices são usados regularmente em análise de séries temporais, especialmente em aplicações macroeconômicas. Um exemplo de número-índice é o índice de produção industrial (IPRI), calculado mensalmente pelo Banco Central dos Estados Unidos. O IPRI é uma medida da produção pesquisada junto a uma ampla gama de indústrias e, como tal, sua magnitude em um ano particular não tem significado quantitativo. Para interpretar a magnitude do IPRI, precisamos conhecer o **período base** e o **valor base**. No *Relatório Econômico da Presidência (REP)* dos Estados Unidos de 1997, o ano-base é 1987, e o valor base é 100. (Fazer IPRI igual a 100 no período base é apenas uma convenção; teria o mesmo sentido fazer IPRI = 1 em 1987, e alguns índices são definidos tendo um como valor base.) Como o IPRI foi 107,7 em 1992, podemos dizer que a produção industrial foi, em 1992, 7,7% mais alta que em 1987. Podemos usar o IPRI em dois anos quaisquer para calcular a diferença percentual na produção industrial durante esses dois anos. Por exemplo, como IPRI = 61,4 em 1970 e IPRI = 85,7 em 1979, a produção industrial cresceu cerca de 39,6% durante os anos setenta.

É fácil alterar o período base de qualquer número-índice, e algumas vezes temos de fazer isso para definir uma base comum para números-índices descritos com anos-base diferentes. Por exemplo, se quisermos alterar o ano-base do IPRI de 1987 para 1982, simplesmente dividimos o IPRI de cada ano pelo valor de 1982 e multiplicamos por 100, para fazer com que o valor do período base seja 100. De modo geral, a fórmula é

$$indnovo_t = 100(indantigo_t / indantigo_{basenova}), \quad (10.20)$$

onde $indantigo_{basenova}$ é o valor original do índice no novo ano-base. Por exemplo, fazendo 1987 o ano-base, o IPRI em 1992 é 107,7; se mudarmos o ano-base para 1982, o IPRI em 1992 passa a ser $100(107,7/81,9) = 131,5$ (pois o IPRI em 1982 era 81,9).

Outro exemplo importante de um número índice é um *índice de preços*, como o índice de preços ao consumidor (IPC). Já utilizamos o IPC para calcular taxas inflacionárias anuais no Exemplo 10.1. Como no caso do índice de produção industrial, o IPC somente tem significado quando o comparamos com outros anos (ou meses, se estivermos usando dados mensais). No *REP* de 1997, IPC = 38,8 em 1970 e IPC = 130,7 em 1990. Assim, o nível geral de preços cresceu quase 237% nesse período de vinte anos. (Em 1997, o IPC é definido de tal maneira que sua média em 1982, 1983 e 1984 é igual a 100; assim, o período base é apresentado como 1982-1984).

Além de serem usados para calcular taxas de inflação, os índices de preços são necessários para transformar uma série temporal medida em *valores nominais* (ou *valores correntes*), em *valores reais*

(ou *valores constantes*). A maior parte do comportamento econômico é considerada como sendo influenciada por variáveis reais, e não nominais. Por exemplo, a economia do trabalho clássica assume que a oferta de mão-de-obra está baseada no salário por hora real, e não no salário nominal. Obter o salário real a partir do salário nominal é fácil se tivermos um índice de preços como o IPC. Devemos ter algum cuidado e primeiro dividir o IPC por 100, de forma que o valor no ano-base seja um. Então, se s for o salário médio por hora em valores nominais e $p = \text{IPC}/100$, o *salário real* será simplesmente s/p . Esse salário é medido em dólares do período base do IPC. Por exemplo, na Tabela B-45 do REP de 1997, os salários médios por hora são informados em termos nominais e em valores de 1982 (o que significa que o IPC usado para calcular o salário real tinha 1982 como ano-base). Essa tabela informa que o salário nominal por hora em 1960 era de US\$ 2,09 mas, medido em valores de 1982, o salário era de US\$ 6,79. O salário por hora real atingiu um pico em 1973, de US\$ 8,55 em valores de 1982, e caiu para US\$ 7,40 em 1995. Assim, houve um declínio bastante significativo nos salários reais nos últimos 20 anos. (Se compararmos os salários nominais de 1973 e 1995, teremos um quadro enganoso: US\$ 3,94 em 1973 e US\$ 11,44 em 1995. Como o salário real efetivamente caiu, o aumento no salário nominal é devido inteiramente à inflação.)

Medidas padrão de resultados econômicos são feitas em termos reais. A mais importante delas é o *produto interno bruto*, ou *PIB*. Quando é divulgado pela imprensa, um crescimento no PIB é sempre um crescimento real. Na Tabela B-9 do REP de 1997, o PIB está registrado em bilhões de dólares de 1992. Utilizamos uma medida de resultados semelhante, produto nacional bruto real, no Exemplo 10.3.

Fatos interessantes acontecem quando variáveis de valores reais são usadas em combinação com logaritmos naturais. Suponha, por exemplo, que a média de horas trabalhadas por semana esteja relacionada com o salário real como

$$\log(\text{horas}) = \beta_0 + \beta_1 \log(s/p) + u.$$

Usando o fato de que $\log(s/p) = \log(s) - \log(p)$, podemos escrever isso como

$$\log(\text{horas}) = \beta_0 + \beta_1 \log(s) + \beta_2 \log(p) + u, \quad (10.21)$$

mas com a restrição de que $\beta_2 = -\beta_1$. Portanto, a hipótese de que somente o salário real influencia a oferta de mão-de-obra impõe uma restrição nos parâmetros do modelo (10.21). Se $\beta_2 \neq -\beta_1$, o nível de preços não tem efeito sobre a oferta de mão-de-obra, o que pode acontecer se os trabalhadores não entenderem totalmente a distinção entre os salários real e nominal.

Existem muitos aspectos práticos quanto ao cálculo de números-índices mas discutí-los aqui nos desviaria muito de nosso objetivo. Argumentações detalhadas sobre índices de preços podem ser encontradas na maioria dos textos de nível intermediário sobre macroeconomia, como Mankiw (1994, Capítulo 2). Para nós, o importante é que sejamos capazes de usar números-índices em análise de regressão. Como mencionado anteriormente, já que as magnitudes dos números-índices não são especialmente informativas, elas freqüentemente aparecem em forma logarítmica, de modo que os coeficientes de regressão têm interpretações de mudanças percentuais.

Exemplificaremos agora um estudo de evento que também utiliza números-índices.

EXEMPLO 10.5**{Ações Anti-dumping e Importações de Produtos Químicos}**

Krupp e Pollard (1996) analisaram os efeitos das ações anti-*dumping* impetradas pela indústria química dos Estados Unidos contra a importação de vários produtos químicos. Enfatizaremos aqui apenas um produto químico, o cloreto de bário, um agente de limpeza utilizado em vários processos químicos e na produção de gasolina. Os dados estão contidos no arquivo BARIUM.RAW. No início dos anos 1980, os produtores norte-americanos de cloreto de bário acreditavam que a China estava oferecendo seu produto aos Estados Unidos a um preço deslealmente baixo (uma atitude conhecida como *dumping*), e a indústria de cloreto de bário moveu uma ação na Comissão de Comércio Internacional (CCI) em outubro de 1983. A CCI decidiu a favor da indústria norte-americana de cloreto de bário em outubro de 1984. Existem várias questões de interesse neste caso, mas abordaremos apenas algumas delas. Primeira, as importações são anormalmente elevadas no período que imediatamente precede uma ação inicial? Segunda, as importações mudam de forma considerável após a impetração de uma ação anti-*dumping*? Finalmente, qual é a redução nas importações após uma decisão favorável à indústria dos Estados Unidos?

Para responder essas questões, seguimos Krupp e Pollard definindo três variáveis *dummy*: *befile6* é igual a um durante os seis meses que antecedem a impetração da ação, *affile6* indica os seis meses após a impetração da ação e *afdec6* representa os seis meses após a decisão positiva. A variável dependente é o volume das importações de cloreto de bário da China, *chnimp*, que usamos na forma logarítmica. Incluímos como variáveis explicativas, todas em forma logarítmica, um índice da produção química, *chemp1* (para controlar a demanda total de cloreto de bário), o volume da produção de gasolina, *gas* (outra variável de demanda), e um índice da taxa cambial, *rtwex*, que mede a força do dólar frente a várias outras moedas. O índice de produção química foi definido como 100 em junho de 1977. A análise aqui difere um pouco da de Krupp e Pollard pelo fato de usarmos logaritmos naturais de todas as variáveis (exceto das variáveis *dummy*, é claro), e incluirmos todas as três variáveis na mesma regressão.

Utilizando dados mensais de fevereiro de 1978 a dezembro de 1988 obtemos o seguinte resultado:

$$\begin{aligned}
 \log(\hat{chnimp}) = & -17,80 + 3,12 \log(chemp1) + 0,196 \log(gas) \\
 & (21,05) \quad (0,48) \quad (0,907) \\
 & + 0,983 \log(rtwex) + 0,060 \text{befile6} - 0,032 \text{affile6} - 0,565 \text{afdec6} \\
 & (0,400) \quad (0,261) \quad (0,264) \quad (0,286) \\
 & n = 131, R^2 = 0,305, \bar{R}^2 = 0,271.
 \end{aligned} \tag{10.22}$$

A equação mostra que *befile6* é estatisticamente não significante, de modo que não há evidência de que as importações chinesas fossem singularmente altas durante os seis meses que antecederam a impetração da ação. Além disso, embora a estimativa de *affile6* seja negativa, o coeficiente é pequeno (indicando uma queda de aproximadamente 3,2% nas importações chinesas), e é estatisticamente bastante não significante. O coeficiente de *afdec6* mostra uma queda substancial nas importações chinesas de cloreto de bário após a decisão a favor da indústria dos Estados Unidos, o que não surpreende. Como o efeito é tão grande, calculamos a percentagem exata de mudança: $100[\exp(-0,565) - 1] \approx -43,2\%$. O coeficiente é estatisticamente significante ao nível de 5% contra uma alternativa bilateral.

Os sinais dos coeficientes nas variáveis de controle são os que esperávamos: um aumento na produção geral de produtos químicos aumenta a demanda por agentes de limpeza. A produção de gasolina não afeta de maneira significante as importações chinesas. O coeficiente de $\log(rtwex)$ mostra que um aumento no valor do dólar em relação a outras moedas aumenta a demanda por importações da China, conforme previsto pela teoria econômica. (Aliás, a elasticidade não é, estatisticamente, diferente de um. Por quê?)

Interações entre variáveis qualitativas e quantitativas também são usadas na análise de séries temporais. Segue um exemplo com importância prática.

EXEMPLO 10.6

(Resultados de Eleições e Desempenho Econômico)

Fair (1996) resume seu trabalho explicando os resultados das eleições presidenciais em termos de desempenho econômico. Ele explica a proporção dos votos dos dois partidos favoráveis ao candidato democrata utilizando os dados dos anos de 1916 a 1992 (cada quatro anos), em um total de 20 observações. Estimamos uma versão simplificada do modelo de Fair (usando nomes de variáveis que são mais descritivas do que os dele):

$$demvoto = \beta_0 + \beta_1 partiCB + \beta_2 cand + \beta_3 partiCB \cdot pepos + \beta_4 partiCB \cdot inf + u,$$

onde *demvoto* é a proporção da votação dos dois partidos favoráveis ao candidato democrata. A variável explicativa *partiCB* é semelhante a uma variável *dummy*, mas assume o valor um se um Democrata estiver na Casa Branca e –1 se for um Republicano. Fair usa essa variável para impor a restrição de que o efeito de um Republicano estar na Casa Branca tem a mesma magnitude, mas com sinal oposto ao de um Democrata estar na Casa Branca. Esta é uma restrição natural, já que a participação dos partidos deve totalizar um, por definição. Ela também poupa dois graus de liberdade, o que é importante, com tão poucas observações. De forma semelhante, a variável *cand* é definida como um se um candidato Democrata estiver concorrendo, –1 se um candidato Republicano estiver concorrendo, e zero, se nenhuma das duas. A variável *pepos* é o número de trimestres, durante os 15 primeiros trimestres da administração da época, em que o crescimento trimestral da produção real *per capita* foi superior a 2,9% (taxa anual) e *inf* é a taxa média anual de inflação nos primeiros 15 trimestres da administração. Veja Fair (1996) para definições precisas.

Os economistas estão mais interessados nos termos de interação *partiCB-pepos* e *partiCB-inf*. Como *partiCB* é igual a um quando um Democrata está na Casa Branca, β_3 mede o efeito de boas notícias econômicas sobre o partido no poder; esperamos $\beta_3 > 0$. De forma análoga, β_4 mede o efeito que a inflação tem sobre o partido no poder. Como a inflação durante uma administração é considerada má notícia, esperamos $\beta_4 < 0$.

A equação estimada utilizando os dados de FAIR.RAW é

$$\begin{aligned} demvoto &= 0,481 - 0,0435 partiCB + 0,0544 cand \\ &\quad (0,012) \quad (0,0405) \quad (0,0234) \\ &+ 0,0108 partiCB \cdot pepos - 0,0077 partiCB \cdot inf \\ &\quad (0,0041) \quad (0,0033) \\ &n = 20, R^2 = 0,663, \bar{R}^2 = 0,573. \end{aligned} \tag{10.23}$$

Todos os coeficientes, exceto o de *partiCB*, são estatisticamente significantes ao nível de 5%. A candidatura vale cerca de 5,4 pontos percentuais na participação nos votos. (Lembre-se, *demvoto* é medida como uma proporção.) Além disso, a variável sobre as notícias econômicas tem um efeito positivo: um trimestre a mais com boas notícias vale cerca de 1,1 pontos percentuais. A inflação, como esperado, tem um efeito negativo: se a média anual de inflação for, digamos, dois pontos percentuais mais alta, o partido no poder perde cerca de 1,5 pontos percentuais dos votos.

EXEMPLO 10.6 (continuação)

Poderíamos ter usado essa equação para prever os resultados das eleições presidenciais de 1996 entre Bill Clinton, o Democrata, e Bob Dole, o Republicano. (O candidato independente, Ross Perot, é excluído porque a equação de Fair é para uma eleição de somente dois partidos.) Como Clinton concorreu como candidato do partido que estava no poder, $partiCB = 1$ e $cand = 1$. Para prever os resultados da eleição, precisamos das variáveis $pepos$ e inf . Durante os primeiros 15 trimestres da administração Clinton, o PIB real *per capita* ultrapassou 2,9% por três vezes, portanto $pepos = 3$. Além disso, usando o deflator de preços do PIB registrado na Tabela B-4 do REP de 1997, a taxa média anual de inflação (calculada usando a fórmula de Fair) do quarto trimestre de 1991 ao terceiro trimestre de 1996, foi de 3,019. Integrando tudo isso em (10.23) temos

$$\hat{dem\text{voto}} = 0,481 - 0,0435 + 0,0544 + 0,0108(3) - 0,0077(3,019) \approx 0,5011.$$

Portanto, com base nas informações conhecidas antes das eleições de novembro, seria possível prever que Clinton teria uma pequena maioria nos votos: cerca de 50,1%. Na verdade, Clinton venceu com mais folga: sua participação nos votos naquela eleição foi de 54,65%.

10.5 TENDÊNCIA E SAZONALIDADE

Caracterização de Séries Temporais com Tendência

Muitas séries temporais econômicas têm uma tendência comum de crescer ao longo do tempo. Temos de saber reconhecer que algumas séries contêm uma **tendência temporal**, com o propósito de obter inferência causal usando dados de séries temporais. Ignorar o fato de que duas seqüências estejam apresentando tendência na mesma direção ou em direções opostas pode nos induzir à conclusão errônea de que alterações em uma variável são de fato causadas por alterações ocorridas em outra variável. Em muitos casos, dois processos de séries temporais parecem ser correlacionados somente porque ambos estão apresentando a mesma tendência ao longo do tempo, devido a razões relacionadas com outros fatores não-observados.

A Figura 10.2 contém um gráfico da produtividade da força de trabalho (produção por hora trabalhada) nos Estados Unidos para os anos de 1947 a 1987. Esta série exibe uma clara tendência ascendente, que reflete o fato de que os trabalhadores se tornaram mais produtivos ao longo do tempo.

Outras séries, pelo menos ao longo de certos períodos, têm clara tendência descendente. Como tendências positivas são mais comuns, elas serão enfatizadas em nossas discussões.

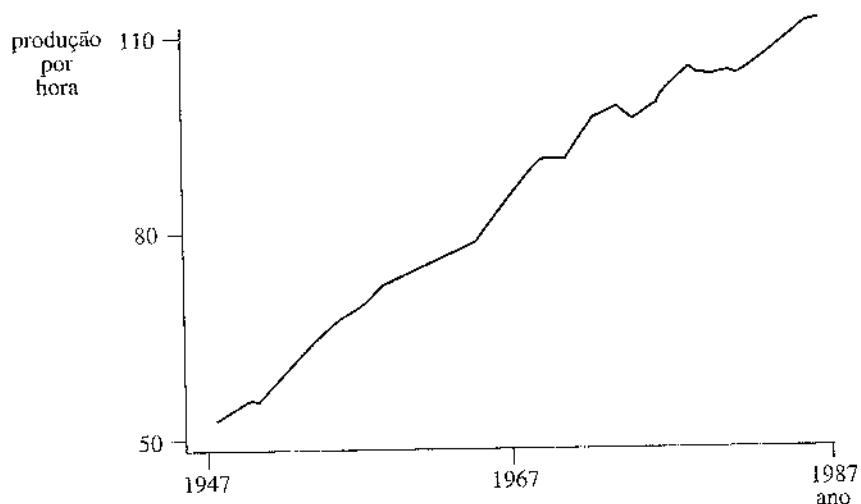
Que tipos de modelos estatísticos capturam adequadamente o comportamento de uma tendência? Uma formulação muito conhecida é escrever as séries $\{y_t\}$ como

$$y_t = \alpha_0 + \alpha_1 t + e_t, t = 1, 2, \dots, \quad (10.24)$$

onde, no caso mais simples, $\{e_t\}$ é uma seqüência independente e identicamente distribuída (i.i.d.) com $E(e_t) = 0$, $\text{Var}(e_t) = \sigma_e^2$. Observe como o parâmetro α_1 multiplica o tempo, t , resultando em uma **tendência temporal linear**. Interpretar α_1 em (10.24) é fácil: mantendo todos os outros fatores fixos (aqueles em e_t), α_1 mede a mudança em y_t de um período para o próximo, motivada pela passagem do tempo; quando $\Delta e_t = 0$,

Figura 10.2

Produção por hora trabalhada nos Estados Unidos durante os anos 1947-1987; 1977 = 100.



$$\Delta y_t = y_t - y_{t-1} = \alpha_1.$$

Outra maneira de pensar em uma seqüência com uma tendência temporal linear é que seu valor médio é uma função linear do tempo:

$$E(y_t) = \alpha_0 + \alpha_1 t. \quad (10.25)$$

Se $\alpha_1 > 0$, então, em média, y_t está crescendo ao longo do tempo e, portanto, tem uma tendência ascendente. Se $\alpha_1 < 0$, então y_t tem uma tendência decrescente. Os valores de y_t não caem exatamente na linha de (10.25) devido à aleatoriedade, mas os valores esperados estão sobre a linha. Diferentemente da média, a variância de y_t é constante ao longo do tempo: $\text{Var}(y_t) = \text{Var}(e_t) = \sigma_e^2$.

No Exemplo 10.4, utilizamos a taxa geral de fertilidade como a variável dependente em um modelo de defasagens distribuídas finitas. De 1950 até meados dos anos 1980, a tgf tem uma clara tendência decrescente. É possível uma tendência linear com $\alpha_1 < 0$ ser realista para todos os períodos de tempo futuros? Explique.

Se $\{e_t\}$ for uma seqüência i.i.d., então $\{y_t\}$ é uma seqüência independentemente, embora não identicamente, distribuída. Uma caracterização mais realista de séries temporais com tendência permite que $\{e_t\}$ seja correlacionada ao longo do tempo, mas isso não altera a forma de uma tendência temporal linear. Aliás, o que é importante para a análise de regressão sob as hipóteses do modelo linear clássico

é que $E(y_t)$ seja linear em t . Quando discutirmos as propriedades de amostras grandes do MQO no Capítulo 11 teremos que discutir quanta correlação serial em $\{e_t\}$ é permitida.

Muitas séries temporais econômicas são mais bem aproximadas por uma **tendência exponencial**, que aparece quando uma série tem a mesma taxa média de crescimento de período para período. A Figura 10.3 reproduz graficamente as importações anuais nominais dos Estados Unidos durante os anos 1948 a 1995 (REP de 1997, Tabela B-101).

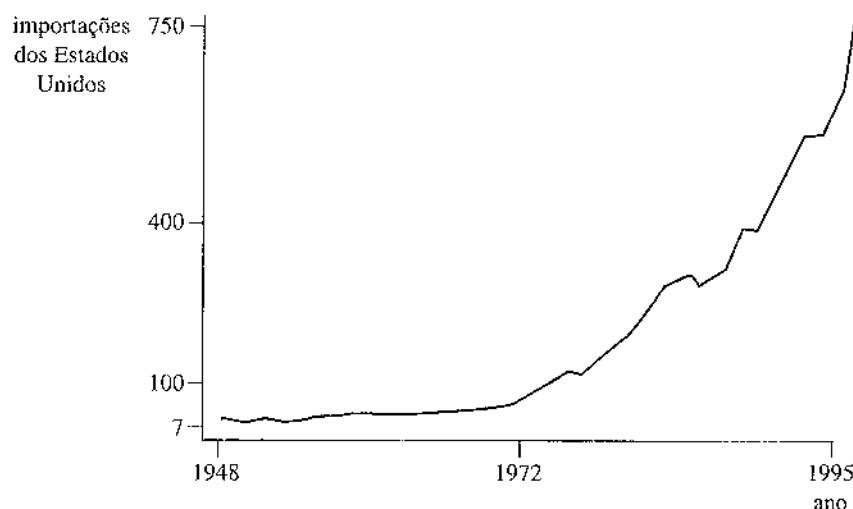
Nos primeiros anos vemos que as mudanças nas importações em cada ano são relativamente pequenas, mas vão crescendo conforme o tempo passa. Isso é coerente com uma *taxa média constante de crescimento*: a mudança percentual é, *grosso modo*, a mesma em cada período.

Na prática, uma tendência exponencial em uma série temporal é captada pela modelagem do logaritmo natural da série como uma tendência linear (assumindo que $y_t > 0$):

$$\log(y_t) = \beta_0 + \beta_1 t + e_t, \quad t = 1, 2, \dots \quad (10.26)$$

Figura 10.3

Importações nominais dos Estados Unidos durante os anos 1948-1995 (em bilhões de dólares).



A exponenciação mostra que y_t tem uma tendência exponencial: $y_t = \exp(\beta_0 + \beta_1 t + e_t)$. Como queremos usar séries de tempo com tendência exponencial em modelos de regressão linear, a equação (10.26) acaba sendo o modo mais conveniente de representar tais séries.

Como interpretamos β_1 em (10.26)? Lembre-se de que, para pequenas mudanças, $\Delta \log(y_t) = \log(y_t) - \log(y_{t-1})$ é aproximadamente a mudança proporcional em y_t :

$$\Delta \log(y_t) \approx (y_t - y_{t-1})/y_{t-1}. \quad (10.27)$$

O lado direito de (10.27) também é chamado de **taxa de crescimento** em y do período $t - 1$ ao período t . Para transformar a taxa de crescimento em percentagem, simplesmente multiplicamos o resultado por 100. Se y_t seguir (10.26), então, levando em conta as mudanças e definindo $\Delta e_t = 0$,

$$\Delta \log(y_t) = \beta_1, \text{ para todos os } t. \quad (10.28)$$

Em outras palavras, β_1 é aproximadamente a taxa média de crescimento por período em y_t . Por exemplo, se t representa ano e $\beta_1 = 0,027$, então y_t cresce cerca de 2,7% por ano, em média.

Embora tendências lineares e exponenciais sejam as mais comuns, as tendências temporais podem ser mais complicadas. Por exemplo, em vez do modelo de tendência linear em (10.24), poderíamos ter uma tendência temporal quadrática:

$$y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + e_t. \quad (10.29)$$

Se α_1 e α_2 são positivos, a inclinação da tendência é crescente, como é facilmente visto calculando-se a inclinação aproximada (mantendo e_t fixo):

$$\frac{\Delta y_t}{\Delta t} \approx \alpha_1 + 2\alpha_2 t. \quad (10.30)$$

[Se você estiver familiarizado com cálculo, reconhecerá o lado direito de (10.30) como a derivada de $\alpha_0 + \alpha_1 t + \alpha_2 t^2$ em relação a t .] Se $\alpha_1 > 0$, mas $\alpha_2 < 0$, a tendência tem um formato em curva. Essa pode não ser uma boa descrição de certas séries com tendência porque isso requer uma tendência crescente seguida, posteriormente, por uma tendência decrescente. Não obstante, ao longo de determinado período de tempo, esta pode ser uma maneira flexível de modelar séries temporais que tenham tendências mais complicadas que as de (10.24) ou (10.26).

Uso de Variáveis com Tendência na Análise de Regressão

Levar em conta as variáveis explicada ou explicativa que apresentam tendência é um assunto bastante objetivo na análise de regressão. Primeiro, não há nada nas variáveis com tendência que necessariamente viole as hipóteses do modelo linear clássico, ST.1 a ST.6. Porém, devemos ser cuidadosos, levando em conta o fato de que fatores que apresentam tendência, não-observados, que afetem y_t , podem também ser correlacionados com as variáveis explicativas. Se ignorarmos essa possibilidade, podemos encontrar uma relação espúria entre y_t e uma ou mais das variáveis explicativas. O fenômeno de descobrir uma relação entre duas ou mais variáveis explicativas com tendência, simplesmente devido ao fato de cada uma delas estar crescendo ao longo do tempo, é um exemplo de **regressão espúria**. Felizmente, a adição de uma tendência temporal elimina esse problema.

Concretamente, considere um modelo em que dois fatores observados, x_{t1} e x_{t2} , afetem y_t . Além disso, existem fatores não-observados que estão sistematicamente crescendo ou decrescendo ao longo do tempo. Um modelo que captura isso é

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 t + u_t. \quad (10.31)$$

O modelo se encaixa na estrutura da regressão linear múltipla com $x_{i3} = t$. Levar em conta a tendência nessa equação é reconhecer explicitamente que y_i pode estar crescendo ($\beta_3 > 0$) ou decrescendo ($\beta_3 < 0$) ao longo do tempo, por razões essencialmente não-relacionadas com x_{i1} e x_{i2} . Se (10.31) satisfizer as hipóteses ST.1, ST.2 e ST.3, então, omitindo t da regressão e computando a regressão de y_i sobre x_{i1} e x_{i2} geralmente obteremos estimadores viesados de β_1 e β_2 : efetivamente, omitimos uma variável importante, t , da regressão. Isso é verdade, especialmente se x_{i1} e x_{i2} apresentarem tendência, pois elas podem ser altamente correlacionadas com t . O próximo exemplo mostra como a omissão de uma tendência temporal pode resultar em regressão espúria.

EXEMPLO 10.7

{Investimento Imobiliário e Preços de Imóveis}

Os dados contidos no arquivo HSEINV.RAW são observações anuais sobre investimento imobiliário e um índice de preços de imóveis nos Estados Unidos de 1947 a 1988. Sejam \hat{invpc} o investimento imobiliário real *per capita* (em milhares de dólares) e *preço* um índice de preço de imóveis (igual a um em 1982). A computação de uma regressão simples em forma de elasticidade constante, que pode ser vista como uma equação de oferta do estoque de imóveis, fornece o seguinte resultado:

$$\begin{aligned}\log(\hat{invpc}) &= -0,550 + 1,241 \log(\text{preço}) \\ &\quad (0,043) \quad (0,382) \\ n &= 42, R^2 = 0,208, \bar{R}^2 = 0,189.\end{aligned}\tag{10.32}$$

A elasticidade do investimento *per capita* em relação ao preço é muito grande e estatisticamente significante; porém, ela não é estatisticamente diferente de um. Devemos ter cuidado com isso. Tanto \hat{invpc} como *preço* apresentam tendência crescente. Particularmente, se fizermos a regressão de $\log(\hat{invpc})$ sobre t , obteremos um coeficiente da tendência igual a 0,0081 (erro-padrão = 0,0018); a regressão de $\log(\text{preço})$ sobre t gera um coeficiente de t igual a 0,0044 (desvio-padrão = 0,0004). Embora os erros-padrão dos coeficientes não sejam necessariamente confiáveis – essas regressões tendem a conter correlação serial substancial – as estimativas dos coeficientes revelam tendência crescente.

Para explicar o comportamento de tendência das variáveis, adicionamos uma tendência temporal ao modelo:

$$\begin{aligned}\log(\hat{invpc}) &= -0,913 - 0,381 \log(\text{preço}) + 0,0098 t \\ &\quad (0,136) \quad (0,679) \quad (0,0035) \\ n &= 42, R^2 = 0,341, \bar{R}^2 = 0,307.\end{aligned}\tag{10.33}$$

A história agora é muito diferente: a elasticidade estimada do preço é negativa e não é estatisticamente diferente de zero. A tendência temporal é estatisticamente significante, e seu coeficiente indica um crescimento de aproximadamente 1% em \hat{invpc} ao ano, em média. Desta análise, não podemos concluir que o investimento imobiliário real *per capita* não seja influenciado pelo preço. Existem outros fatores, captados na tendência temporal, que afetam \hat{invpc} e *preço*, mas nós não os modelamos. Os resultados em (10.32) mostram uma relação espúria entre \hat{invpc} e *preço* devido ao fato de que o preço também tem tendência crescente ao longo do tempo.

Em alguns casos, a adição de uma tendência temporal pode tornar uma variável explicativa importante *mais* significante. Isso pode acontecer se as variáveis dependente e independente tiverem diferentes tipos de tendência (digamos, uma crescente e uma decrescente), mas movimentos na variável independente sobre sua linha de tendência levam a movimentos da variável dependente distantes de sua linha de tendência.

EXEMPLO 10.8

(Equação da Fertilidade)

Se adicionarmos uma tendência temporal linear na equação da fertilidade (10.18), obtemos

$$\hat{tgf}_t = 111,77 + 0,279 ip_t - 35,59 ww2_t + 0,997 pill_t - 1,15 t \quad (10.34)$$

(3,36) (0,040) (6,30) (6,626) (0,19)

$n = 72, R^2 = 0,662, \bar{R}^2 = 0,642.$

O coeficiente de ip é mais do que o triplo da estimativa em (10.18), e é muito mais significante estatisticamente. De maneira curiosa, $pill$ não é significante quando é levada em conta uma tendência linear. Como pode ser visto pela estimativa, tgf está diminuindo, em média, ao longo desse período, outros fatores permanecendo iguais.

Como a taxa geral de fertilidade exibiu tendência tanto crescente como decrescente durante o período de 1913 a 1984, podemos ver o quanto é robusto o efeito estimado de ip quando usamos uma tendência quadrática:

$$\hat{tgf}_t = 124,09 + 0,348 ip_t - 35,88 ww2_t - 10,12 pill_t \quad (10.35)$$

(4,36) (0,040) (5,71) (6,34)

$-2,53 t + 0,0196 t^2$

(0,39) (0,0050)

$n = 72, R^2 = 0,727, \bar{R}^2 = 0,706.$

O coeficiente de ip é ainda maior e mais estatisticamente significante. Agora, $pill$ tem o efeito negativo esperado e é marginalmente significante, e ambos os termos de tendência são estatisticamente significantes. A tendência quadrática é um método flexível para explicar o comportamento incomum da tendência de tgf .

Você deve estar se perguntando: no Exemplo 10.8, por que parar na tendência quadrática? Nada nos impede de adicionar, digamos, t^3 como uma variável independente, e, de fato, isso pode ser justificado (veja Exercício 10.12). Entretanto, temos de ter cuidado para não nos afastarmos da questão quando incluímos um termo de tendência em um modelo. Queremos tendências relativamente simples que capturem amplos movimentos na variável dependente que não sejam explicados pelas variáveis independentes no modelo. Se incluirmos diversos termos polinomiais em t , poderemos rastrear muito bem qualquer série. Contudo, isso será de pouca ajuda para descobrir que variáveis explicativas afetam y .

Interpretação sobre a Retirada da Tendência de Regressões com a Inclusão de uma Tendência Temporal

A inclusão de uma tendência temporal em um modelo de regressão cria uma interpretação interessante em termos da **retirada da tendência** da série de dados originais antes de usá-los na análise de regressão. Concretamente, enfatizamos o modelo (10.31), mas nossas conclusões são de caráter mais geral.

Quando regredimos y_t sobre x_{t1}, x_{t2} e t , obtemos a equação estimada

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{t1} + \hat{\beta}_2 x_{t2} + \hat{\beta}_3 t. \quad (10.36)$$

Podemos expandir os resultados sobre a interpretação parcial do MQO que abordamos no Capítulo 3 para mostrar que $\hat{\beta}_1$ e $\hat{\beta}_2$ podem ser obtidos conforme segue.

(i) Compute a regressão de y_t , x_{t1} e x_{t2} sobre uma constante e a tendência temporal t , e guarde os resíduos, digamos $\ddot{y}_t, \ddot{x}_{t1}, \ddot{x}_{t2}$, $t = 1, 2, \dots, n$. Por exemplo,

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t.$$

Assim, podemos pensar em \ddot{y}_t como uma variável cuja tendência foi excluída linearmente. Ao retirarmos a tendência de y_t , estimamos o modelo

$$y_t = \alpha_0 + \alpha_1 t + e_t$$

pelo método MQO: os resíduos dessa regressão, $\hat{e}_t = \ddot{y}_t$, tiveram removida a tendência temporal (pelo menos na amostra). Uma interpretação semelhante é possível para \ddot{x}_{t1} e \ddot{x}_{t2} .

(ii) Compute a regressão de

$$\ddot{y}_t \text{ sobre } \ddot{x}_{t1} \text{ e } \ddot{x}_{t2}. \quad (10.37)$$

(Não é necessário um intercepto, mas sua inclusão não afeta o resultado: o intercepto será estimado como igual a zero.) Esta regressão produzirá com exatidão $\hat{\beta}_1$ e $\hat{\beta}_2$ a partir de (10.36).

Isso significa que as estimativas de interesse principal, $\hat{\beta}_1$ e $\hat{\beta}_2$, podem ser interpretadas como provenientes de uma regressão *sem* tendência temporal, mas em que primeiro removemos a tendência da variável dependente e de todas as outras variáveis independentes. A mesma conclusão permanece com qualquer número de variáveis independentes e mesmo se a tendência for quadrática ou de algum outro grau polinomial.

Se t for omitida de (10.36), não ocorrerá remoção da tendência, e y_t poderá parecer relacionada com uma ou mais das variáveis x_{ij} simplesmente porque cada uma contém uma tendência; vimos isso no Exemplo 10.7. Se o termo tendência for estatisticamente significante, e os resultados mudarem de maneira importante quando uma tendência temporal for adicionada a uma regressão, os resultados iniciais sem uma tendência devem ser tratados com desconfiança.

A interpretação de $\hat{\beta}_1$ e de $\hat{\beta}_2$ mostra que é uma boa idéia incluir uma tendência na regressão se qualquer variável independente tiver uma tendência, mesmo se a variável y_t não tiver. Se y_t não tem tendência observável, mas, digamos, x_{t1} está crescendo ao longo do tempo, então a exclusão de uma tendência da regressão pode fazer parecer que x_{t1} não tem efeito em y_t , mesmo que movimentos em x_{t1} em torno de sua tendência possa afetar y_t . Isso será captado se t for incluído na regressão.

EXEMPLO 10.9**(Emprego em Porto Rico)**

Quando adicionamos uma tendência linear à equação (10.17), as estimativas são

$$\begin{aligned} \log(\text{prepop}_t) &= -8,70 - 0,169 \log(\text{mincov}_t) + 1,06 \log(\text{uspn}b_t) \\ &\quad (1,30) \quad (0,044) \quad (0,18) \\ &\quad -0,032 t \\ &\quad (0,005) \\ n &= 38, R^2 = 0,847, \bar{R}^2 = 0,834. \end{aligned} \tag{10.38}$$

O coeficiente de $\log(\text{uspn}b)$ mudou dramaticamente: de $-0,012$ e não significante para $1,06$ e bastante significante. O coeficiente do salário mínimo mudou apenas levemente, embora o erro-padrão seja consideravelmente menor, tornando $\log(\text{mincov})$ mais significante que antes.

A variável prepop , não exibe nenhuma tendência clara crescente ou decrescente, mas $\log(\text{uspn}b)$ tem uma tendência crescente, linear. [Uma regressão de $\log(\text{uspn}b)$ sobre t produz uma estimativa de cerca de $0,03$, de modo que $\text{uspn}b$ está crescendo em cerca de 3% por ano ao longo do período.] Podemos pensar na estimativa de $1,06$ como segue: quando $\text{uspn}b$ aumenta em 1% acima de sua tendência de longo prazo, prepop aumenta em cerca de $1,06$.

Cálculo do R-Quadrado quando a Variável Dependente Apresenta Tendência

Os R -quadrados em regressões de séries temporais normalmente são muito altos, especialmente quando comparados com R -quadrados típicos de dados de corte transversal. Será que isso significa que sabemos mais sobre os fatores que afetam y a partir dos dados de séries temporais? Não necessariamente. Por um lado, dados de séries temporais aparecem freqüentemente em forma agregada (como o salário médio por hora na economia dos Estados Unidos), e dados agregados muitas vezes são mais fáceis de explicar do que dados sobre indivíduos, famílias ou empresas, que em geral são característicos de dados de corte transversal. Entretanto, os R -quadrados usuais e ajustados de regressões de séries temporais podem ser artificialmente elevados quando a variável dependente apresentar uma tendência. Lembre-se de que R^2 é uma medida de o quanto é grande a variância do erro em relação à variância de y . A fórmula para o R -quadrado ajustado mostra isso de forma direta:

$$\bar{R}^2 = 1 - (\hat{\sigma}_u^2 / \hat{\sigma}_y^2),$$

onde $\hat{\sigma}_u^2$ é o estimador não-viesado da variância do erro $\hat{\sigma}_y^2 = \text{SQT}/(n - 1)$, e $\text{SQT} = \sum_{t=1}^n (y_t - \bar{y})^2$. Agora, estimar a variância do erro quando y_t apresenta uma tendência não é problema, desde que uma tendência temporal seja incluída na regressão. Contudo, quando $E(y_t)$ segue, digamos, uma tendência temporal linear [veja (10.24)], $\text{SQT}/(n - 1)$ não mais será um estimador não-viesado ou consistente de $\text{Var}(y_t)$. Aliás, $\text{SQT}/(n - 1)$ pode superestimar de forma substancial a variância em y_t , pois ela não explica a tendência em y_t .

Quando a variável dependente satisfaz as tendências linear, quadrática ou qualquer outro polinomial, é fácil computar uma medida de grau-de-ajuste que primeiro remova o efeito de qualquer tendência temporal em y_t . O método mais simples é calcular o R -quadrado usual em uma regressão em que a variável dependente já esteja expurgada de sua tendência. Por exemplo, se o modelo for (10.31), primeiro regredimos y_t sobre t e obtemos os resíduos \hat{y}_t . Depois, regredimos

$$\hat{y}_t \text{ sobre } x_{t1}, x_{t2} \text{ e } t. \quad (10.39)$$

O R -quadrado dessa regressão é

$$1 - \frac{\text{SQR}}{\sum_{t=1}^n \hat{y}_t^2}, \quad (10.40)$$

onde SQR é idêntico à soma dos quadrados dos resíduos de (10.36). Como $\sum_{t=1}^n \hat{y}_t^2 \leq \sum_{t=1}^n (y_t - \bar{y})^2$ (e em geral a desigualdade é estrita), o R -quadrado de (10.40) não será maior, e normalmente será menor, que o R -quadrado de (10.36). (A soma dos resíduos quadrados será idêntica em ambas as regressões.) Quando y_t contiver uma forte tendência temporal linear, (10.40) pode ser muito menor que o R -quadrado usual.

O R -quadrado em (10.40) reflete melhor o quanto x_{t1} e x_{t2} explicam y_t , porque ele remove o efeito da tendência temporal. Afinal de contas, podemos sempre explicar uma variável que esteja crescendo ou decrescendo com algum tipo de tendência, mas isso não significa que tenhamos descoberto quaisquer fatores que causem movimentos em y_t . Um R -quadrado ajustado também pode ser calculado com base em (10.40): divida SQR por $(n - 4)$, pois esse é o total de gl em (10.36) e divida $\sum_{t=1}^n \hat{y}_t^2$ por $(n - 2)$, já que existem dois parâmetros de tendência estimados na remoção da tendência de y_t . Em geral, SQR é dividido por gl na regressão usual (que inclui qualquer tendência temporal), e $\sum_{t=1}^n \hat{y}_t^2$ é dividido por $(n - p)$, onde p é o número de parâmetros de tendência estimados na remoção da tendência de y_t . Veja Wooldridge (1991) para comentários adicionais sobre cálculos de medidas de graus de ajuste com variáveis que apresentam tendência.

EXEMPLO 10.10

[Investimento Imobiliário]

No Exemplo 10.7, vimos que a inclusão de uma tendência temporal linear juntamente com $\log(\text{preço})$ na equação de investimento imobiliário tinha um efeito substancial na elasticidade-preço. Mas o R -quadrado da regressão (10.33), tomado literalmente, diz que estamos “explicando” 34,1% da variação em $\log(\text{invpc})$. Isso é enganoso. Se primeiramente removermos a tendência em $\log(\text{invpc})$ e fizermos a regressão da variável modificada sobre $\log(\text{preço})$ e t , o R -quadrado passa a ser de 0,008, e o R -quadrado ajustado é, na verdade, negativo. Assim, movimentos em $\log(\text{preço})$ em torno de sua tendência não têm virtualmente nenhum poder explicativo dos movimentos em $\log(\text{invpc})$ em torno de sua tendência. Isso é consistente com o fato de que a estatística t de $\log(\text{preço})$ na equação (10.33) é muito pequena.

Antes de deixarmos esta subseção, devemos fazer uma última observação. No cálculo de uma forma R -quadrado de uma estatística F para testar múltiplas hipóteses, apenas usamos os R -quadrados usuais, sem nenhuma remoção de tendência. Lembre-se, a forma R -quadrado da estatística F é apenas um artifício computacional, e portanto a fórmula habitual é sempre apropriada.

Sazonalidade

Se uma série temporal é observada em intervalos mensais ou trimestrais (ou mesmo semanais ou diárias), ela pode exibir **sazonalidade**. Por exemplo, a construção de novas casas no centro-oeste norte-americano é fortemente influenciada pelo clima. Embora as características do clima sejam aleatórias, podemos ter certeza de que o clima durante o mês de janeiro será mais rigoroso que em junho, e portanto o início das construções de novas casas é geralmente maior em junho que em janeiro. Uma maneira de modelar esse fenômeno é permitir que o valor esperado da série, y_t , seja diferente em cada mês. Como outro exemplo, as vendas do varejo no quarto trimestre são tipicamente mais elevadas do que nos três trimestres anteriores devido às festas de Natal. Novamente, isso pode ser captado permitindo que as vendas médias do varejo difiram no transcorrer do ano. Isso é uma adição à possibilidade de se permitir uma média da tendência. Por exemplo, as vendas do varejo nos primeiros trimestres mais recentes foram mais elevadas do que as do quarto trimestre de 30 anos atrás, porque elas vêm crescendo regularmente. Mesmo assim, se compararmos as vendas médias em um ano típico, o fator sazonal das festas natalinas tende a fazer as vendas maiores no quarto trimestre.

Embora muitas séries de dados mensais e trimestrais exibam padrões sazonais, nem todas o fazem. Por exemplo, não existe padrão sazonal observável nas taxas mensais de juros ou de inflação. Além disso, séries que exibem padrões sazonais são freqüentemente **ajustadas sazonalmente** antes de serem informadas para uso público. Uma série sazonalmente ajustada é aquela que, em princípio, teve removidos seus fatores sazonais. Ajustes sazonais podem ser feitos de diversas maneiras, e uma discussão mais cuidadosa sobre esse assunto está além do escopo desta obra. [Veja Harvey (1990) e Hylleberg (1986) para tratamentos detalhados.]

Ajustes sazonais tornaram-se tão comuns que, em muitos casos, não é possível obter dados não ajustados. O PIB trimestral dos Estados Unidos é um exemplo de destaque. No *Economic Report of the President* (Relatório Anual da Presidência – REP) anual, muitos dos conjuntos de dados macroeconômicos descritos em freqüência mensal (pelo menos para os anos mais recentes) e, daqueles que exibem padrões sazonais, todos são sazonalmente ajustados. As principais fontes de séries temporais macroeconômicas, inclusive a *Citibase*, ajustam as séries quanto à sazonalidade. Portanto, o escopo para usarmos nossos próprios procedimentos de ajuste sazonal é muitas vezes limitado.

Algumas vezes trabalhamos com dados não ajustados, e é bom sabermos que existem métodos simples para tratar a sazonalidade em modelos de regressão. Geralmente, podemos incluir um conjunto de **variáveis dummy sazonais** para explicar a sazonalidade na variável dependente, nas variáveis independentes, ou em ambas.

A abordagem é simples. Suponha que temos dados mensais e que entendemos padrões sazonais dentro de um ano como, em linhas gerais, constantes ao longo do tempo. Por exemplo, já que o Natal ocorre sempre na mesma época do ano, podemos esperar que as vendas do varejo sejam, em média, mais altas nos meses do final do ano do que no início do ano. Ou, como os padrões climáticos são amplamente similares ao longo dos anos, o início da construção de novas casas no centro-oeste norte-americano será maior, em média, durante os meses de verão do que nos meses de inverno. Um modelo geral de dados mensais que capta esse fenômeno é

$$y_t = \beta_0 + \delta_1 fev_t + \delta_2 mar_t + \delta_3 abr_t + \dots + \delta_{11} dez_t + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t \quad (10.41)$$

onde $fev_t, mar_t, \dots, dez_t$, são variáveis *dummy* indicando se o período de tempo t corresponde ao mês apropriado. Nesta formulação, janeiro é o mês base, e β_0 é o intercepto de janeiro. Se não houver sazonalidade em y_t , uma vez que x_{it} tenha sido controlada, δ_1 a δ_{11} serão todas zero. Isso é facilmente verificado através de um teste F .

Na equação (10.41), qual é o intercepto de março? Explique por que as variáveis *dummy* sazonais satisfazem a hipótese de exogeneidade estrita.

EXEMPLO 10.11

(Efeitos das Ações Anti-dumping)

No Exemplo 10.5 utilizamos dados mensais que não haviam sido sazonalmente ajustados. Portanto, devemos ter incluído variáveis *dummy* sazonais para termos certeza de que nenhuma das conclusões importantes mudariam. É possível que os meses imediatamente anteriores à impetratura de ações fossem meses em que as importações tenham sido maiores, ou menores, em média, do que nos outros meses. Quando adicionamos as 11 variáveis *dummy* mensais como em (10.41) e testamos sua significância conjunta, obtemos p -valor = 0,59 e, portanto, as *dummies* sazonais são conjuntamente não significantes. Além disso, nada importante muda nas estimativas, uma vez que a significância estatística seja levada em conta. Krupp e Pollard (1996) de fato usaram três variáveis *dummy* para as estações do ano (outono, primavera e verão, considerando o inverno como a estação base), em vez de um conjunto total de *dummies* mensais; o resultado é essencialmente o mesmo.

Se os dados forem trimestrais, então incluiremos variáveis *dummy* para três dos quatro trimestres, com o trimestre omitido sendo o trimestre base. Algumas vezes é útil interagir *dummies* sazonais com algumas das variáveis x_{it} para permitir que o efeito de x_{it} sobre y_t difira ao longo do ano.

Assim como a inclusão de uma tendência temporal em uma regressão tem a interpretação de inicialmente remover a tendência dos dados, a inclusão de variáveis *dummy* sazonais em uma regressão pode ser interpretada como **dessazonalização** dos dados. Concretamente, considere a equação (10.41) com $k = 2$. Os coeficientes de inclinação MQO $\hat{\beta}_1$ e $\hat{\beta}_2$ de x_1 e x_2 podem ser obtidos como segue:

(i) Regreda cada uma das variáveis y_t, x_{t1} e x_{t2} sobre uma constante e sobre as *dummies* mensais, $fev_t, mar_t, \dots, dez_t$, e guarde os resíduos, digamos $\ddot{y}_t, \ddot{x}_{t1}$ e \ddot{x}_{t2} , para todos os $t = 1, 2, \dots, n$). Por exemplo,

$$\ddot{y}_t = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 fev_t - \hat{\alpha}_2 mar_t - \dots - \hat{\alpha}_{11} dez_t.$$

Este é um método para dessazonalizar uma série de tempo mensal. Uma interpretação semelhante se mantém para \ddot{x}_{t1} e \ddot{x}_{t2} .

(ii) Compute a regressão, sem as *dummies* mensais, de \ddot{y}_t sobre \ddot{x}_{t1} e \ddot{x}_{t2} [como em (10.37)]. Isso produz $\hat{\beta}_1$ e $\hat{\beta}_2$.

Em alguns casos, se y_t tiver sazonalidade pronunciada, uma medida melhor do grau-de-ajuste é um R -quadrado baseado na variável y_t dessazonalizada. Isso remove quaisquer efeitos sazonais que não sejam explicados por x_{it} . Ajustes específicos dos graus de liberdade são discutidos em Wooldridge (1991a).

Séries temporais que exibem padrões sazonais também podem ter tendências, em cujo caso, devemos estimar um modelo de regressão com uma tendência temporal e variáveis *dummy* sazonais. As regressões podem, então, ser interpretadas como regressões cujas séries tiveram tanto a tendência como a sazonalidade removidas. Estatísticas de graus de ajuste são discutidas em Wooldridge (1991a); essencialmente, removemos a tendência de y , e dessazonalizamos y , ao regredir y , tanto sobre uma tendência temporal como sobre *dummies* sazonais antes de calcularmos o R -quadrado.

Neste capítulo tratamos da análise de regressão básica com dados de séries temporais. Sob hipóteses paralelas às da análise de corte transversal, o MQO é não-viesado (sob ST.1 a ST.3), o MQO é BLUE (sob ST.1 a ST.5) e os habituais erros-padrão, estatísticas t e estatísticas F do MQO podem ser usados para inferência estatística (sob ST.1 a ST.6). Devido à correlação temporal na maioria dos dados de séries temporais, devemos fazer hipóteses explícitas sobre como os erros estão relacionados com as variáveis explicativas em todos os períodos de tempo e sobre a correlação temporal nos erros propriamente ditos. As hipóteses do modelo linear clássico podem ser bastante restritivas em aplicações de séries temporais, mas são um ponto de partida natural. Nós as aplicamos tanto no modelo de regressão estática como no modelo de defasagens distribuídas finitas.

Logaritmos e variáveis *dummy* são usados regularmente em aplicações de séries temporais e nos estudos de casos. Também discutimos números-índices e séries temporais medidas em termos de valores nominais e reais.

Tendência e sazonalidade podem facilmente ser tratadas em uma estrutura de regressão múltipla com a inclusão de variáveis *dummy* de tempo e de sazonalidade em nossas equações de regressão. Apresentamos problemas com o R -quadrado habitual como uma medida do grau de ajuste e sugerimos algumas alternativas simples bascadas em remoção de tendência e dessazonalização.

10.1 Decida se você concorda ou não com qualquer uma das seguintes declarações e dê uma breve explicação sobre sua decisão:

- (i) Como nas observações de corte transversal podemos assumir que a maioria das observações de séries temporais são independentemente distribuídas.
- (ii) O estimador MQO numa regressão de série temporal é não-viesado sob as três primeiras hipóteses de Gauss-Markov.
- (iii) Uma variável com tendência não pode ser usada como a variável dependente na análise de regressão múltipla.
- (iv) Sazonalidade não é um problema quando usamos observações de séries temporais anuais.

10.2 Sejam $gPIB_t$ a percentagem anual de mudança no produto interno bruto e jur_t uma taxa de juros de curto prazo. Suponha que $gPIB_t$ está relacionado com as taxas de juros por

$$gPIB_t = \alpha_0 + \delta_0 jur_t + \delta_1 jur_{t-1} + u_t$$

onde u_t é não-correlacionado com jur_t , jur_{t-1} e todos os outros valores passados das taxas de juros. Suponha que o Banco Central norte-americano siga a seguinte regra de política monetária:

$$jur_t = \gamma_0 + \gamma_1(gPIB_{t-1} - 3) + v_t,$$

onde $\gamma_1 > 0$. (Quando o crescimento do PIB do ano anterior estiver acima de 3%, o Banco Central norte-americano aumenta as taxas de juros para evitar uma economia “superaquecida”.) Se v_t for não-correlacionado com os valores passados de jur_t e u_t , demonstre que jur_t deve ser correlacionado com u_{t-1} . (*Sugestão:* Faça a defasagem da primeira equação em um período de tempo, substituindo-a em $gPIB_{t-1}$ na segunda equação.) Que hipóteses de Gauss-Markov esse procedimento infringe?

10.3 Suponha que y_t siga um modelo DDF de segunda ordem:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t.$$

Sejam z^* o *valor de equilíbrio* de z_t e y^* o valor de equilíbrio de y_t , de modo que

$$y^* = \alpha_0 + \delta_0 z^* + \delta_1 z^* + \delta_2 z^*.$$

Mostre que a mudança em y^* , devido a uma mudança em z^* , é igual à propensão de longo prazo multiplicada pela mudança em z^* :

$$\Delta y^* = PLP \cdot \Delta z^*.$$

Isso oferece uma maneira alternativa de interpretar a PLP.

10.4 Quando os três indicadores de eventos *befile6*, *affile6* e *afdec6* são eliminados da equação (10.22), obtemos $R^2 = 0,281$ e $\bar{R}^2 = 0,264$. Os indicadores de evento são conjuntamente significantes ao nível de 10%?

10.5 Suponha que você tenha dados trimestrais sobre o início de construção de novas casas, taxas de juros e renda real *per capita*. Especifique um modelo do início de construção de novas casas que explique possíveis tendências e sazonalidade nas variáveis.

10.6 No Exemplo 10.4, vimos que nossas estimativas dos coeficientes individuais defasados em um modelo de defasagens distribuídas eram muito imprecisas. Uma maneira de aliviar o problema da multicolinearidade é assumir que δ_j segue um padrão relativamente simples. De maneira concreta, considere um modelo com quatro defasagens:

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \delta_3 z_{t-3} + \delta_4 z_{t-4} + u_t.$$

Agora, assumamos que δ_j segue uma tendência quadrática na defasagem j :

$$\delta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2,$$

para os parâmetros γ_0 , γ_1 e γ_2 . Isto é um exemplo de um modelo de defasagens distribuídas polinomial (DDP).

- (i) Insira a fórmula de cada δ_j no modelo de defasagens distribuídas e escreva o modelo em termos dos parâmetros γ_h , para $h = 0, 1, 2$.
- (ii) Explique a regressão que você computaria para estimar γ_h .
- (iii) O modelo de defasagens distribuídas polinomial é uma versão restrita do modelo geral. Quantas restrições foram impostas? Como você testaria isso? (*Sugestão:* Pense no teste *F*.)

Questões Adicionais quanto ao Uso do MQO com Dados de Séries Temporais

No Capítulo 10, discutimos as propriedades de amostras finitas do MQO para dados de séries temporais sob conjuntos de hipóteses crescentemente mais fortes. Sob o conjunto total de hipóteses do modelo linear clássico de séries temporais, ST.1 a ST.6, o MQO tem exatamente as mesmas propriedades desejáveis que deduzimos para os dados de corte transversal. Da mesma forma, a inferência estatística é conduzida da mesma maneira, como se fosse para uma análise de corte transversal.

De nossa análise do corte transversal no Capítulo 5, sabemos que há boas razões para estudarmos as propriedades de amostras grandes do MQO. Por exemplo, se os termos erros não forem extraídos de uma distribuição normal, teremos de nos valer do teorema do limite central para justificar as habituais estatísticas de testes e intervalos de confiança do MQO.

A análise de amostras grandes é ainda mais importante no contexto de séries temporais. (O que é de certa forma irônico, dado que amostras grandes de séries temporais podem ser difíceis de se obter; mas, com frequência, não temos outra escolha senão recorrer às aproximações de amostras grandes.) Na Seção 10.3 explicamos como a hipótese de exogeneidade estrita (ST.2) pode ser violada em modelos estáticos e de defasagens distribuídas. Como mostraremos na Seção 11.2, modelos com variáveis dependentes defasadas devem infringir a hipótese ST.2.

Infelizmente, a análise de amostras grandes para problemas de séries temporais é repleta de mais dificuldades do que é o caso da análise de corte transversal. No Capítulo 5, obtivemos as propriedades de MQO de amostras grandes no contexto de amostragem aleatória. O assunto é mais complicado quando permitimos que as observações sejam correlacionadas ao longo do tempo. Contudo, os principais teoremas de limites se sustentam em alguns, embora não em todos, processos de séries temporais. O ponto principal é saber se a correlação entre as variáveis em diferentes períodos de tempo tende para zero com rapidez suficiente. Séries temporais que tenham correlação temporal substancial exigem atenção especial na análise de regressão. Este capítulo chamará a atenção para certos problemas relacionados com esses tipos de séries na análise de regressão.

11.1 SÉRIES TEMPORAIS ESTACIONÁRIAS E FRACAMENTE DEPENDENTES

Nesta seção, apresentaremos os principais conceitos necessários para aplicar as aproximações comuns de amostras grandes na análise de regressão com dados de séries temporais. Os detalhes não são tão importantes quanto o entendimento geral dos problemas.

Séries Temporais Estacionárias e Não-Estacionárias

Historicamente, a idéia de **processo estacionário** tem desempenhado um papel importante na análise de séries temporais. Um processo estacionário de série temporal é aquele em que as distribuições de probabilidades são estáveis no decorrer do tempo no seguinte sentido: se pegarmos qualquer coleção de variáveis aleatórias na seqüência e depois deslocarmos essa seqüência para diante em h períodos de tempo, a distribuição de probabilidade conjunta deve permanecer inalterada. Segue uma definição formal de estacionariedade.

PROCESSO ESTOCÁSTICO ESTACIONÁRIO: O processo estocástico $\{x_t; t = 1, 2, \dots\}$ é estacionário se, para todas as coleções de índices temporais $1 \leq t_1 < t_2 < \dots < t_m$, a distribuição conjunta de $(x_{t_1}, x_{t_2}, \dots, x_{t_m})$ é a mesma que a distribuição conjunta de $(x_{t_1+h}, x_{t_2+h}, \dots, x_{t_m+h})$ para todos os inteiros $h \geq 1$.

Essa definição é um pouco abstrata, mas seu significado é bastante objetivo. Uma implicação (selecionando $m = 1$ e $t_1 = 1$) é que x_t tenha a mesma distribuição de x_1 para todo $t = 2, 3, \dots$. Em outras palavras, a seqüência $\{x_t; t = 1, 2, \dots\}$ é *identicamente distribuída*. A estacionariedade exige ainda mais. Por exemplo, a distribuição conjunta de (x_1, x_2) (os primeiros dois termos da seqüência) deve ser a mesma que a da distribuição conjunta de (x_t, x_{t+1}) para qualquer $t \geq 1$. Mais uma vez, isso não coloca qualquer restrição sobre como x_t e x_{t+1} estão relacionados entre si; aliás, eles podem ser altamente correlacionados. A estacionariedade não exige que a natureza de qualquer correlação entre termos adjacentes seja a mesma para todos os períodos de tempo.

Um processo estocástico que não seja estacionário é chamado de **processo não-estacionário**. Como a estacionariedade é um aspecto do processo estocástico subjacente e não da única realização disponível, pode ser muito difícil determinar se os dados que coletamos foram gerados por um processo estacionário. Porém, é fácil identificar certas seqüências que não sejam estacionárias. Um processo com uma tendência temporal do tipo discutido na Seção 10.5 é claramente não-estacionário: no mínimo, sua média muda no tempo.

Algumas vezes, uma forma mais fraca de estacionariedade é suficiente. Se $\{x_t; t = 1, 2, \dots\}$ tiver um segundo momento finito, isto é, $E(x_t^2) < \infty$ para todo t , então, a seguinte definição se aplica.

PROCESSO DE COVARIÂNCIA ESTACIONÁRIA: Um processo estocástico $\{x_t; t = 1, 2, \dots\}$ com segundo momento finito [$E(x_t^2) < \infty$] tem **covariância estacionária** se (i) $E(x_t)$ é constante; (ii) $\text{Var}(x_t)$ é constante; e (iii) para qualquer $t, h \geq 1$, $\text{Cov}(x_t, x_{t+h})$ depende somente de h e não de t .

Suponha que $\{y_t; t = 1, 2, \dots\}$ é gerado por $y_t = \delta_0 + \delta_1 t + e_t$, onde $\delta_1 \neq 0$, e $\{e_t; t = 1, 2, \dots\}$ é uma seqüência i.d.d. com média zero e variância σ_e^2 . (i) $\{y_t\}$ tem covariância estacionária? (ii) $y_t - E(y_t)$ tem covariância estacionária?

A estacionariedade da covariância enfatiza somente os primeiros dois momentos de um processo estocástico: a média e a variância do processo são constantes no decorrer do tempo e a covariância entre x_t e x_{t+h} depende somente da distância entre os dois termos, h , e não da localização do período de tempo inicial, t . Conclui-se imediatamente que a correlação entre x_t e x_{t+h} também depende somente de h .

Se um processo estacionário tem um segundo momento finito, então, ele deve ter covariância estacionária, mas o inverso, com certeza, não é verdadeiro. Algumas vezes, para enfatizar que a estacionariedade é um requisito mais forte do que a estacionariedade da covariância, a primeira é referida como *estacionariedade estrita*. Como a estacionariedade estrita simplifica o enunciado de algumas de nossas hipóteses subsequentes, “estacionariedade” para nós sempre significará a forma estrita.

Como é utilizada a estacionariedade na econometria de séries temporais? Em nível técnico, a estacionariedade simplifica os enunciados da lei dos grandes números e do teorema do limite central, embora não estejamos preocupados com enunciados formais neste capítulo. Em nível prático, se quisermos entender a relação entre duas ou mais variáveis usando a análise de regressão, teremos que pressupor algum tipo de estabilidade ao longo do tempo. Se permitirmos que a relação entre duas variáveis (digamos, y_t e x_t) mude arbitrariamente em cada período de tempo, não podemos esperar descobrir muito sobre como a mudança em uma variável afeta a outra variável, se somente tivermos acesso a uma única realização da série temporal.

Ao estabelecer um modelo de regressão múltipla para dados de séries temporais, estamos assumindo certa forma de estacionariedade, no sentido de que β_j não muda ao longo do tempo. Além disso, as hipóteses ST.4 e ST.5 implicam que a variância do processo de erro seja constante ao longo do tempo e que a correlação entre os erros em dois períodos adjacentes seja igual a zero, o que claramente é constante ao longo do tempo.

Séries Temporais Fracamente Dependentes

A estacionariedade tem a ver com as distribuições conjuntas de um processo, conforme ele se move ao longo do tempo. Um conceito bastante diferente é o da dependência fraca, que coloca restrições sobre o quanto podem estar fortemente relacionadas as variáveis aleatórias x_t e x_{t+h} , à medida que a distância temporal entre elas, h , fique maior. A noção de dependência fraca é mais facilmente discutida em uma série temporal estacionária: em linhas gerais, um processo estacionário de série temporal $\{x_t; t = 1, 2, \dots\}$ é chamado de **fracamente dependente** se x_t e x_{t+h} forem “quase independentes” enquanto h aumenta sem limites.

Um enunciado semelhante se mantém verdadeiro se a seqüência for não-estacionária, mas, neste caso, temos que supor que o conceito de quase independência não depende do ponto de partida, t .

A descrição de dependência fraca dada no parágrafo anterior é necessariamente vaga. Não podemos definir formalmente a dependência fraca por não existir definição que englobe todos os casos de interesse. Existem muitas formas específicas de dependência fraca que são formalmente definidas, mas estão muito além do escopo deste texto. [Veja White (1984), Hamilton (1994) e Wooldridge (1994b) para abordagens avançadas desses conceitos.]

Para nosso propósito, uma noção intuitiva do significado de dependência fraca é suficiente. Seqüências com covariância estacionária podem ser caracterizadas em termos de correlações: uma série temporal com covariância estacionária é fracamente dependente se a correlação entre x_t e x_{t+h} se aproxima de zero “com rapidez suficiente” quando $h \rightarrow \infty$. (Devido à estacionariedade da covariância, a correlação não depende do ponto de partida, t .) Em outras palavras, à medida que as variáveis se afastam no tempo, a correlação entre elas se torna cada vez menor. Seqüências com covariância estacionária em que $\text{Cov}(x_t, x_{t+h}) \rightarrow 0$ conforme $h \rightarrow \infty$ são chamadas de **assimptoticamente não-correlacionadas**. Intuitivamente, é desta maneira que usualmente caracterizaremos a dependência fraca. Tecnicamente, temos que supor que a correlação converge para zero de maneira suficientemente rápida, mas atenuaremos isso.

Por que a dependência fraca é importante para a análise de regressão? Essencialmente, ela substitui a hipótese de amostragem aleatória indicando que a lei dos grandes números (LGN) e o teorema do limite central (TLC) são válidos. O teorema do limite central mais conhecido para dados de séries

temporais requer estacionariedade e alguma forma de dependência fraca: assim, séries temporais estacionárias e fracamente dependentes são ideais para serem usadas na análise de regressão múltipla. Na Seção 11.2, argumentaremos que o MQO pode ser justificado de maneira bastante geral, recorrendo-se aos enunciados LGN e TLC. Séries temporais que não sejam fracamente dependentes — cujos exemplos vereinos na Seção 11.3 — geralmente não satisfazem o TLC, razão pela qual seu uso na análise de regressão múltipla pode ser delicado.

O exemplo mais simples de uma série temporal fracamente dependente é uma seqüência independente, identicamente distribuída: uma seqüência que seja independente é obviamente fracamente dependente. Um exemplo mais interessante de uma seqüência fracamente dependente é

$$x_t = e_t + \alpha_1 e_{t-1}, t = 1, 2, \dots, \quad (11.1)$$

onde $\{e_t; t = 0, 1, \dots\}$ é uma seqüência i.i.d. com média zero e variância σ_e^2 . O processo $\{x_t\}$ é chamado de **processo de média móvel de ordem um [MA(1)]**: x_t é uma média ponderada de e_t e e_{t-1} ; no período seguinte, eliminamos e_{t-1} e, então, x_{t+1} dependerá de e_{t+1} e e_t . O ajuste do coeficiente de e_t para 1 em (11.1) não produzirá perda de generalidade. [Na equação (11.1), usamos x_t e e_t como rótulos genéricos de processos de séries temporais. Eles não precisam ter nada a ver com as variáveis explicativas ou com os erros em um modelo de regressão de série temporal, embora ambas as variáveis explicativas e os erros possam ser processos MA(1).]

Por que um processo MA(1) é fracamente dependente? Os termos adjacentes na seqüência são correlacionados: como $x_{t+1} = e_{t+1} + \alpha_1 e_t$, $\text{Cov}(x_t, x_{t+1}) = \alpha_1 \text{Var}(e_t) = \alpha_1 \sigma_e^2$. Dado que $\text{Var}(x_t) = (1 + \alpha_1^2) \sigma_e^2$, $\text{Corr}(x_t, x_{t+1}) = \alpha_1 / (1 + \alpha_1^2)$. Por exemplo, se $\alpha_1 = 0,5$, $\text{Corr}(x_t, x_{t+1}) = 0,4$. [A correlação máxima positiva ocorre quando $\alpha_1 = 1$, caso em que $\text{Corr}(x_t, x_{t+1}) = 0,5$.] Porém, quando verificamos as variáveis na seqüência que estejam separadas por dois ou mais períodos de tempo, elas são não-correlacionadas porque são independentes. Por exemplo, $x_{t+2} = e_{t+2} + \alpha_1 e_{t+1}$ é independente de x_t porque $\{e_t\}$ é independente ao longo de t . Devido à hipótese de distribuição idêntica sobre e_t , $\{x_t\}$ em (11.1) é efetivamente estacionário. Assim, uma seqüência MA(1) é estacionária e fracamente dependente, e a lei dos grandes números e o teorema do limite central podem ser aplicados a $\{x_t\}$.

Um exemplo mais conhecido é o processo

$$y_t = \rho_1 y_{t-1} + e_t, t = 1, 2, \dots, \quad (11.2)$$

O ponto de partida na seqüência é y_0 (em $t = 0$), e $\{e_t; t = 1, 2, \dots\}$ é uma seqüência i.i.d. com média zero e variância σ_e^2 . Também assumimos que os e_t são independentes de y_0 e que $E(y_0) = 0$. Esse processo é chamado de **processo auto-regressivo de ordem um [AR(1)]**.

A hipótese crucial para a fraca dependência de um processo AR(1) é a *condição de estabilidade* $|\rho_1| < 1$. Então, dizemos que $\{y_t\}$ é um **processo AR(1) estável**.

Para verificar que um processo AR(1) estável é assintoticamente não-correlacionado, é útil supor que o processo tem covariância estacionária. (De fato, é possível mostrar que geralmente $\{y_t\}$ é estritamente estacionário, mas a prova disso é um tanto técnica.) Então, sabemos que $E(y_t) = E(y_{t-1})$, e de (11.2) com $\rho_1 \neq 1$, isso pode acontecer somente se $E(y_t) = 0$. Tomando a variância de (11.2) e usando o fato de que e_t e y_{t-1} são independentes (e, portanto, não-correlacionados), $\text{Var}(y_t) = \rho_1^2 \text{Var}(y_{t-1}) + \text{Var}(e_t)$, e assim, sob a estacionariedade da covariância, devemos ter $\sigma_y^2 = \rho_1^2 \sigma_y^2 + \sigma_e^2$. Como $\rho_1^2 < 1$ pela condição de estabilidade, podemos encontrar facilmente σ_y^2 :

$$\sigma_y^2 = \sigma_e^2 / (1 - \rho_1^2). \quad (11.3)$$

Agora, podemos encontrar a covariância entre y_t e y_{t+h} para $h \geq 1$. Usando substituições repetidas,

$$\begin{aligned} y_{t+h} &= \rho_1 y_{t+h-1} + e_{t+h} = \rho_1(\rho_1 y_{t+h-2} + e_{t+h-1}) + e_{t+h} \\ &= \rho_1^2 y_{t+h-2} + \rho_1 e_{t+h-1} + e_{t+h} = \dots \\ &= \rho_1^h y_t + \rho_1^{h-1} e_{t+h} + \dots + \rho_1 e_{t+h-1} + e_{t+h} \end{aligned}$$

Como $E(y_t) = 0$ para todos os t , podemos multiplicar esta última equação por y_t , e aplicar o valor esperado para obter $\text{Cov}(y_t, y_{t+h})$. Usando o fato de que e_{t+j} é não-correlacionado com y_t , para todo $j \geq 1$, temos

$$\begin{aligned} \text{Cov}(y_t, y_{t+h}) &= E(y_t y_{t+h}) = \rho_1^h E(y_t^2) + \rho_1^{h-1} E(y_t e_{t+1}) + \dots + E(y_t e_{t+h}) \\ &= \rho_1^h E(y_t^2) = \rho_1^h \sigma_y^2. \end{aligned}$$

Como σ_y é o desvio-padrão de y_t e de y_{t+h} , podemos facilmente encontrar as correlações entre y_t e y_{t+h} para qualquer $h \geq 1$:

$$\text{Corr}(y_t, y_{t+h}) = \text{Cov}(y_t, y_{t+h}) / (\sigma_y \sigma_y) = \rho_1^h. \quad (11.4)$$

Em particular, $\text{Corr}(y_t, y_{t+1}) = \rho_1$, de modo que ρ_1 é o coeficiente de correlação entre dois termos quaisquer adjacentes na seqüência.

A equação (11.4) é importante porque mostra que, embora y_t e y_{t+h} sejam correlacionados para qualquer $h \geq 1$, a correlação se torna muito pequena para h grande: como $|\rho_1| < 1$, $\rho_1^h \rightarrow 0$ à medida que $h \rightarrow \infty$. Mesmo quando ρ_1 é grande — digamos, 0,9, o que implica uma correlação positiva muito grande entre os termos adjacentes — a correlação entre y_t e y_{t+h} tende para zero com bastante rapidez. Por exemplo, $\text{Corr}(y_t, y_{t+5}) = 0,591$, $\text{Corr}(y_t, y_{t+10}) = 0,349$, e $\text{Corr}(y_t, y_{t+20}) = 0,122$. Se t indexar anos, isso significa que a correlação entre o resultado de dois y que estejam separados em 20 anos é de cerca de 0,122. Quando ρ_1 é menor, a correlação extingue-se muito mais rapidamente. (Você pode tentar com $\rho_1 = 0,5$ para verificar isso.)

Esta análise mostra, heuristicamente, que um processo AR(1) estável é fracamente dependente. O modelo AR(1) é especialmente importante na análise de regressão múltipla com dados de séries temporais. Discutiremos aplicações adicionais no Capítulo 12 e seu uso em previsões no Capítulo 18.

Existem muitos outros tipos de séries temporais fracamente dependentes, inclusive híbridos de processos auto-regressivos e de média móvel. Mas os exemplos anteriores servem bem para nosso propósito.

Antes de encerrarmos esta seção, devemos enfatizar um ponto que muitas vezes causa confusão em econometria de séries temporais. Uma série de tendência, embora, com certeza, seja não-estacionária, *pode* ser fracamente dependente. Aliás, no modelo de tendência temporal linear simples, no Capítulo 10 [veja a equação (10.24)], a série $\{y_t\}$ era, efetivamente, independente. Uma série que seja estacionária sobre sua tendência temporal, como também fracamente dependente, é, com freqüência, chamada de **processo de tendência estacionária**. (Observe que o nome não é completamente descriptivo porque assumimos dependência fraça com estacionariedade.) Tais processos podem ser usados em

análise de regressão como fizemos no Capítulo 10, *desde que* tendências temporais apropriadas sejam incluídas no modelo.

11.2 PROPRIEDADES ASSIMPTÓTICAS DO MQO

No Capítulo 10, vimos alguns casos nos quais as hipóteses do modelo linear clássico não são satisfeitas em certos problemas de séries temporais. Em tais casos, temos que recorrer às propriedades de amostras grandes do MQO, da mesma forma que na análise de corte transversal. Nesta seção, estabelecemos as hipóteses e principais resultados que legitimam o MQO de forma mais genérica. As provas dos teoremas neste capítulo, por serem um pouco mais difíceis, são omitidas. Veja Wooldridge (1994b).

H I P Ó T E S E S T . 1 ' (LINEARIDADE E DEPENDÊNCIA FRACA)

Supomos que o modelo seja exatamente como na hipótese ST.1, mas agora adicionamos a hipótese de que $\{(x_t, y_t); t = 1, 2, \dots\}$ é estacionária e fracamente dependente. Em particular, a lei dos grandes números e o teorema do limite central podem ser aplicados às médias amostrais.

O requisito da linearidade nos parâmetros mais uma vez significa que podemos escrever o modelo como

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t, \quad (11.5)$$

onde os β_j são os parâmetros a serem estimados. Diferentemente do Capítulo 10, os x_{ij} podem incluir defasagens da variável dependente. Como sempre, defasagens de variáveis explicativas também são permitidas.

Incluímos a estacionariedade na hipótese ST.1' por conveniência, ao definirmos e interpretarmos as hipóteses. Se estivéssemos trabalhando cuidadosamente com as propriedades assintóticas do MQO, como fazemos no Apêndice E, disponível na página do livro, no site www.thomsonlearning.com.br, a estacionariedade também simplificaria aquelas derivações. Contudo, a estacionariedade não é de maneira nenhuma fundamental para que o MQO tenha propriedades assintóticas padrão. (Como mencionado na Seção 11.1, assumindo que os β_j sejam constantes ao longo do tempo, já estamos assumindo alguma forma de estabilidade nas distribuições ao longo do tempo.) A restrição adicional importante na hipótese ST.1', quando comparada com a hipótese ST.1, é a hipótese de dependência fraca. Na Seção 11.1, dedicamos uma boa parte dela discutindo a dependência fraca porque ela não é de forma alguma uma hipótese inócula. Na próxima seção, apresentaremos processos de séries temporais que violam claramente a hipótese de dependência fraca e também discutiremos o uso de tais processos em modelos de regressão múltipla.

H I P Ó T E S E S T . 2 ' (MÉDIA CONDICIONAL ZERO)

As variáveis explicativas $x_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ são **contemporaneamente exógenas** como na equação (10.10): $E(u_t | x_t) = 0$.

Esta é a hipótese mais natural com referência à relação entre u_t e as variáveis explicativas. Ela é muito mais fraca que a hipótese ST.2, porque não coloca restrições sobre como u_t está relacionada com as variáveis explicativas em outros períodos de tempo. Viremos brevemente exemplos que satis-

fazem ST.2'. Pela estacionariedade, se a exogeneidade contemporânea for válida para um período de tempo, será válida para todos os demais. O relaxamento da estacionariedade simplesmente exigirá que assumamos uma condição válida para todo $t = 1, 2, \dots$.

Para determinados propósitos, é útil saber que o resultado de consistência seguinte apenas exige que u_t tenha média condicional zero e que seja não-correlacionado com cada x_{it} :

$$E(u_t) = 0, \text{Cov}(x_{ij}, u_t) = 0, j = 1, \dots, k. \quad (11.6)$$

Trabalharemos principalmente com a hipótese de média condicional zero, pois ela conduz a mais objetiva análise assintótica.

H I P Ó T E S E S T . 3 ' (COLINEARIDADE IMPERFEITA)

Igual à hipótese ST.3.

T E O R E M A 11.1 (CONSISTÊNCIA DO MQO)

Sob ST.1', ST.2' e ST.3', os estimadores de MQO são consistentes: $\text{plim } \hat{\beta}_j = \beta_j, j = 0, 1, \dots, k$.

Existem algumas importantes diferenças práticas entre os teoremas 10.1 e 11.1. Primeira, no teorema 11.1 concluímos que os estimadores de MQO são consistentes, mas não necessariamente não-viciados. Segunda, no teorema 11.1, enfraquecemos a noção de que as variáveis explicativas devem ser exógenas, mas a dependência fraca é exigida nas séries temporais subjacentes. A dependência fraca também é crucial na obtenção de resultados distribucionais aproximados, sobre os quais falaremos mais adiante.

EXEMPLO 11.1

(O Modelo Estático)

Considere um modelo estático com duas variáveis explicativas:

$$y_t = \beta_0 + \beta_1 z_{t1} + \beta_2 z_{t2} + u_t. \quad (11.7)$$

Sob dependência fraca, a condição suficiente para a consistência do MQO é

$$E(u_t | z_{t1}, z_{t2}) = 0. \quad (11.8)$$

Isso exclui as variáveis omitidas que estejam em u_t e sejam correlacionadas com z_{t1} ou z_{t2} . Tampouco qualquer função de z_{t1} ou z_{t2} pode ser correlacionada com u_t , e assim a hipótese ST.2' exclui formas funcionais mal-especificadas, exatamente como no caso do corte transversal. Outros problemas, como o erro de medida nas variáveis z_{t1} ou z_{t2} , podem fazer com que (11.8) não se sustente.

EXEMPLO 11.1 (continuação)

Fundamentalmente, a hipótese ST.2' não exclui a possibilidade de correlação entre, digamos, u_{t-1} e z_{t1} . Esse tipo de correlação pode surgir se z_{t1} for relacionado com y_{t-1} passado, como, por exemplo

$$z_{t1} = \delta_0 + \delta_1 y_{t-1} + v_t \quad (11.9)$$

Por exemplo, z_{t1} pode ser uma variável de política econômica, como, por exemplo, a mudança percentual mensal da oferta monetária, e essa mudança depende das taxas de inflação do último mês (y_{t-1}). Tal mecanismo geralmente faz com que z_{t1} e u_{t-1} sejam correlacionados (como pode ser verificado facilmente). Esse tipo de retroalimentação é permitido sob a hipótese ST.2'.

EXEMPLO 11.2**(O Modelo de Defasagens Distribuídas Finitas)**

No modelo de defasagens distribuídas finitas,

$$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t \quad (11.10)$$

uma hipótese bastante natural é a de que o valor esperado de u_t seja zero, dados os valores correntes e todos os valores passados de z :

$$E(u_t | z_t, z_{t-1}, z_{t-2}, z_{t-3}, \dots) = 0. \quad (11.11)$$

Isso significa que, uma vez que z_t , z_{t-1} , e z_{t-2} sejam incluídas, nenhuma outra defasagem de z afeta $E(y_t | z_t, z_{t-1}, z_{t-2}, z_{t-3}, \dots)$; se isso não fosse verdadeiro, colocaríamos defasagens adicionais na equação. Por exemplo, y_t poderia ser a mudança percentual anual nos investimentos e z_t uma medida das taxas de juros durante o ano t . Quando definimos $x_t = (z_t, z_{t-1}, z_{t-2})$, a hipótese ST.2' é satisfeita: o MQO será consistente. Como no exemplo anterior, ST.2' não elimina a retroalimentação de y para valores futuros de z .

Os dois exemplos anteriores não exigem, necessariamente, teoria assintótica, pois as variáveis explicativas *poderiam* ser estritamente exógenas. O próximo exemplo viola claramente a hipótese de exogeneidade estrita; portanto, somente podemos recorrer às propriedades de amostras grandes do MQO.

EXEMPLO 11.3**[O Modelo AR(1)]**

Considere o modelo AR(1)

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t \quad (11.12)$$

EXEMPLO 11.3 (continuação)

onde o erro u_t tem um valor esperado de zero, dados todos os valores passados de y :

$$E(u_t | y_{t-1}, y_{t-2}, \dots) = 0 \quad (11.13)$$

Combinadas, essas duas equações implicam que

$$E(y_t | y_{t-1}, y_{t-2}, \dots) = E(y_t | y_{t-1}) = \beta_0 + \beta_1 y_{t-1}. \quad (11.14)$$

Esse resultado é muito importante. Primeiro, ele significa que, uma vez que y defasado em um período seja controlado, nenhuma outra defasagem de y afetará o valor esperado de y_t . (É daí que se origina o termo "primeira ordem".) Segundo, a relação é assumida como linear.

Como x_t contém somente y_{t-1} , a equação (11.13) indica que a hipótese ST.2' é válida. De outro lado, a hipótese de exogeneidade estrita necessária para a inexistência de viés, hipótese ST.2, não se sustenta. Como o conjunto de variáveis explicativas para todos os períodos de tempo inclui todos os valores de y exceto o último, $(y_0, y_1, \dots, y_{n-1})$, a hipótese ST.2 exige que, para todos os t , u_t seja não-correlacionado com cada um dos y_0, y_1, \dots, y_{n-1} . Isso não pode ser verdadeiro. De fato, como u_t é não-correlacionado com y_{t-1} sob (11.13), u_t e y_t devem ser correlacionados. Aliás, pode-se ver facilmente que $\text{Cov}(y_t, u_t) = \text{Var}(u_t) > 0$. Portanto, um modelo com uma variável dependente defasada não pode satisfazer a hipótese ST.2 de exogeneidade estrita.

Para que a condição de dependência fraca seja válida, temos que assumir que $|\beta_1| < 1$, como discutimos na Seção 11.1. Se esta condição se mantiver, então, o teorema 11.1 indica que o estimador MQO da regressão de y_t sobre y_{t-1} produzirá estimadores consistentes de β_0 e de β_1 . Infelizmente, $\hat{\beta}_1$ é viesado, e esse viés pode ser grande se o tamanho da amostra for pequeno ou se β_1 for próximo de um. (Para β_1 próximo de um, $\hat{\beta}_1$ pode ter um severo viés para baixo.) Em amostras de moderadas para grandes, $\hat{\beta}_1$ será um bom estimador de β_1 .

Quando usamos os procedimentos padrão de inferência, temos que impor versões das hipóteses de homoscedasticidade e de inexistência de correlação serial. Elas são menos restritivas do que suas correspondentes do modelo linear clássico do Capítulo 10.

H I P Ó T E S E S T . 4 ' (HOMOSCEDASTICIDADE)

Os erros são contemporaneamente homoscedásticos, isto é, $\text{Var}(u_t | x_t) = \sigma^2$.

H I P Ó T E S E S T . 5 ' (INEXISTÊNCIA DE CORRELAÇÃO SERIAL)

Para todo $t \neq s$, $E(u_t u_s | x_t, x_s) = 0$.

Observe como, em ST.4', a condicionalidade ocorre apenas nas variáveis explicativas no tempo t (compare com ST.4). Na hipótese ST.5', a condicionalidade foi feita apenas nas variáveis explicativas nos

períodos de tempo que coincidem com u_t e u_s . Como mencionado, essa hipótese é um pouco difícil de ser interpretada, mas é a condição correta para estudar as propriedades de amostras grandes do MQO em uma diversidade de regressões de séries temporais. Quando consideramos ST.5', muitas vezes ignoraremos o condicionamento em x_t e x_s , e somente pensaremos se u_t e u_s são não-correlacionadas, para todo $t \neq s$.

A correlação serial freqüentemente é um problema nos modelos de regressão estáticos e de desafagens distribuídas finitas: nada garante que os u_t , não-observáveis, sejam não-correlacionados ao longo do tempo. Fundamentalmente, a hipótese ST.5' é válida no modelo AR(1) especificado nas equações (11.12) e (11.13). Como a variável explicativa no tempo t é y_{t-1} , temos que mostrar que $E(u_t u_s | y_{t-1}, y_{s-1}) = 0$ para todo $t \neq s$. Para verificar isso, suponha que $s < t$. (O outro caso resulta por simetria.) Então, como $u_s = y_s - \beta_0 - \beta_1 y_{s-1}$, u_s é uma função de y datada antes do tempo t . Mas, por (11.13), $E(u_t | u_s, y_{t-1}, y_{s-1}) = 0$, e, portanto, $E(u_t u_s | u_s, y_{t-1}, y_{s-1}) = u_s E(u_t | y_{t-1}, y_{s-1}) = 0$. Pela lei das expectativas iteradas (veja Apêndice B no site da Thomson), $E(u_t u_s | y_{t-1}, y_{s-1}) = 0$. Isso é muito importante: desde que somente uma defasagem pertença a (11.12), os erros devem ser serialmente não-correlacionados. Discutiremos essa característica dos modelos dinâmicos de forma mais geral na Seção 11.4.

Agora obtemos um resultado assintótico praticamente idêntico ao do caso do corte transversal.

TEOREMA 11.2 (NORMALIDADE ASSIMPTÓTICA DO MQO)

Sob as hipóteses ST.1' a ST.5', os estimadores MQO são assintóticos e normalmente distribuídos. Além disso, os usuais erros-padrão, estatísticas t , estatísticas F e estatísticas LM são assintoticamente válidas.

Este teorema fornece justificativa adicional para pelo menos alguns dos exemplos estimados no Capítulo 10: mesmo que as hipóteses do modelo linear clássico não se sustentem, o MQO ainda assim é consistente e os procedimentos habituais de inferência são válidos. Evidentemente, isso depende de que ST.1' a ST.5' sejam verdadeiras. Na próxima seção, discorreremos sobre os modos como a hipótese de dependência fraca pode falhar. Os problemas de correlação serial e heteroscedasticidade são tratados no Capítulo 12.

EXEMPLO 11.4

(A Hipótese de Mercados Eficientes)

Podemos utilizar a análise assintótica para testar uma versão da *hipótese de mercados eficientes* (HME). Seja y_t o retorno percentual semanal (de uma quarta-feira à outra, no encerramento) do índice composto da Bolsa de Valores de Nova York. Uma forma estrita da hipótese de mercados eficientes estabelece que as informações observáveis do mercado anteriores à semana t não devem ajudar a prever o retorno durante a semana t . Se utilizarmos somente informações passadas de y , a HME é especificada como

$$E(y_t | y_{t-1}, y_{t-2}, \dots) = E(y_t). \quad (11.15)$$

Se (11.15) for falsa, poderíamos usar as informações sobre os retornos semanais passados para prever o retorno corrente. A HME presume que tais oportunidades de investimentos serão notadas e desaparecerão quase instantaneamente.

EXEMPLO 11.4 (continuação)

Uma maneira simples de testar (11.15) é especificar o modelo AR(1) em (11.12) como o modelo alternativo. Então, a hipótese nula é facilmente declarada como $H_0: \beta_1 = 0$. Sob a hipótese nula, a hipótese ST.2' é verdadeira por (11.15) e, como discutimos anteriormente, a correlação serial não é um problema. A hipótese de homoscedasticidade é $\text{Var}(y_t | y_{t-1}) = \text{Var}(y_t) = \sigma^2$, a qual assumimos ser verdadeira, por ora. Sob a hipótese nula, os retornos das ações são serialmente não-correlacionados, de modo que podemos assumir com segurança que eles são fracamente dependentes. Então, o teorema 11.2 diz que podemos usar as estatísticas t usuais de MQO de $\hat{\beta}_1$ para testar $H_0: \beta_1 = 0$ contra $H_1: \beta_1 \neq 0$.

Os retornos semanais registrados no arquivo NYSE.RAW são computados utilizando dados de janeiro de 1976 até março de 1989. Nos raros casos em que quarta-feira foi feriado, o fechamento do próximo dia útil foi utilizado. O retorno semanal médio para esse período foi de 0,196, em forma de percentagem, com o maior retorno semanal tendo sido de 8,45% e o menor de -15,32% (durante a crise do mercado de ações de outubro de 1987). A estimativa do modelo AR(1) produz

$$\begin{aligned} \text{returno}_t &= 0,180 + 0,059 \text{returno}_{t-1} \\ &\quad (0,081) \quad (0,038) \\ n &= 689, R^2 = 0,0035, \bar{R}^2 = 0,0020. \end{aligned} \tag{11.16}$$

A estatística t do coeficiente de returno_{t-1} é de cerca de 1,55 e assim $H_0: \beta_1 = 0$ não pode ser rejeitada contra a alternativa bilateral, mesmo ao nível de significância de 10%. A estimativa sugere uma leve correlação positiva no retorno da Bolsa de Valores de Nova York entre uma semana e a seguinte, mas não é suficientemente forte para garantir a rejeição da hipótese de mercados eficientes.

No exemplo anterior, a utilização do modelo AR(1) para testar a HME pode não detectar correlação entre retornos semanais que estejam separados por mais de uma semana. É fácil estimar modelos com mais de uma defasagem. Por exemplo, um *modelo auto-regressivo de ordem dois*, ou AR(2), é

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t \\ E(u_t | y_{t-1}, y_{t-2}, \dots) &= 0. \end{aligned} \tag{11.17}$$

Existem condições de estabilidade para β_1 e β_2 que são necessárias para assegurar que o processo AR(2) seja fracamente dependente, mas aqui isso não chega a ser um problema, porque a hipótese nula estabelece que a HME é válida:

$$H_0: \beta_1 = \beta_2 = 0. \tag{11.18}$$

Se adicionarmos a hipótese de homoscedasticidade $\text{Var}(u_t | y_{t-1}, y_{t-2}) = \sigma^2$, podemos usar uma estatística F padrão para testar (11.18). Se estimarmos um modelo AR(2) para returno_t , obteremos

$$\begin{aligned} \text{returno}_t &= 0,186 + 0,060 \text{returno}_{t-1} - 0,038 \text{returno}_{t-2} \\ &\quad (0,081) \quad (0,038) \quad (0,038) \\ n &= 688, R^2 = 0,0048, \bar{R}^2 = 0,0019 \end{aligned}$$

(onde perdemos uma observação a mais devido à defasagem adicional na equação). As duas defasagens são individualmente não significantes ao nível de 10%. Elas também são conjuntamente não significantes: usando $R^2 = 0,0048$, a estatística F é, aproximadamente, $F = 1,65$; o p -valor para esta estatística F (com 2 e 685 graus de liberdade) é de cerca de 0,193. Assim, não rejeitamos (11.18), mesmo ao nível de significância de 15%.

EXEMPLO 11.5

(A Curva de Phillips de Expectativas Aumentadas)

Uma versão linear da curva de Phillips de expectativas aumentadas pode ser escrita como

$$\text{inf}_t - \text{inf}_t^e = \beta_1(\text{desemp}_t - \mu_0) + e_t,$$

onde μ_0 é a taxa natural de desemprego e inf_t^e é a taxa de inflação esperada formada no ano $t - 1$. Este modelo supõe que a taxa natural é constante, o que é questionado pelos macroeconomistas. A diferença entre o desemprego efetivo e a taxa natural é chamada de *desemprego cíclico*, enquanto a diferença entre a inflação efetiva e a esperada é chamada de *inflação não antecipada*. O termo erro, e_t , é chamado de *choque de oferta* pelos macroeconomistas. Se houver uma relação de substituição entre a inflação não antecipada e o desemprego cíclico, então, $\beta_1 < 0$. [Para uma discussão detalhada sobre a curva de Phillips de expectativas aumentadas, veja Mankiw (1994, Seção 11.2).]

Para completar esse modelo, temos que fazer uma suposição sobre as expectativas inflacionárias. Sob *expectativas adaptativas*, o valor esperado da inflação corrente depende da inflação recentemente observada. Uma formulação particularmente simples é a de que a inflação esperada para este ano é a mesma do ano anterior: $\text{inf}_t^e = \text{inf}_{t-1}$. (Veja a Seção 18.1 para uma formulação alternativa das expectativas adaptativas.) Sob essa hipótese podemos escrever

$$\text{inf}_t - \text{inf}_{t-1} = \beta_0 + \beta_1 \text{desemp}_t + e_t$$

ou

$$\Delta \text{inf}_t = \beta_0 + \beta_1 \text{desemp}_t + e_t,$$

onde $\Delta \text{inf}_t = \text{inf}_t - \text{inf}_{t-1}$ e $\beta_0 = -\beta_1 \mu_0$ (espera-se que β_0 seja positivo, já que $\beta_1 < 0$ e $\mu_0 > 0$). Portanto, sob expectativas adaptativas, a curva de Phillips de expectativas aumentadas relaciona a mudança na inflação ao nível de desemprego e a um choque de oferta, e_t . Se e_t for não-correlacionado com desemp_t , como normalmente é assumido, então, poderemos estimar consistenteamente β_0 e β_1 por MQO. (Não precisamos assumir que, digamos, as taxas de desemprego futuras não serão afetadas pelo choque de oferta corrente.) Assumimos que ST.1' a ST.5' se sustentam. A equação estimada é

EXEMPLO 11.5 (continuação)

$$\begin{aligned}\Delta \hat{\text{inf}}_t &= 3,03 - 0,543 \text{ desemp}_t \\ &\quad (1,38) \quad (0,230) \\ n &= 48, R^2 = 0,108, \bar{R}^2 = 0,088.\end{aligned}\tag{11.19}$$

A relação de substituição entre o desemprego cíclico e a inflação não antecipada é pronunciada na equação (11.19): um aumento de um ponto percentual em *desemp* reduz a inflação não antecipada em mais de meio ponto percentual. O efeito é estatisticamente significante (*p*-valor bilateral $\approx 0,023$). Podemos contrapor isso à curva de Phillips estática no Exemplo 10.1, no qual encontramos uma relação levemente positiva entre a inflação e o desemprego.

Como podemos escrever a taxa natural como $\mu_0 = \beta_0 / (-\beta_1)$, podemos usar (11.19) para obter nossa própria estimativa da taxa natural: $\hat{\mu}_0 = \hat{\beta}_0 / (-\hat{\beta}_1) = 3,03 / 0,543 \approx 5,58$. Assim, estimamos a taxa natural como sendo ao redor de 5,6, o que está dentro da faixa sugerida pelos macroeconomistas: historicamente, 5 a 6% é a faixa geralmente mencionada da taxa natural de desemprego. É difícil obter um erro-padrão desta estimativa, porque temos uma função não linear dos estimadores MQO. Wooldridge (2002, Capítulo 3) contém a teoria das funções não lineares. Na aplicação corrente, o erro-padrão é 0,657, o que leva a um intervalo de confiança assintótico de 95% (baseado na distribuição normal padrão) de cerca de 4,29 a 6,87 para a taxa natural.

Sob as hipóteses ST.1' a ST.5', podemos mostrar que os estimadores MQO são assintoticamente eficientes na classe de estimadores descrita no teorema 5.3, mas substituímos o índice *i* da observação de corte transversal pelo índice *t* das séries temporais. Finalmente, modelos com variáveis explicativas que apresentam tendência podem, efetivamente, satisfazer as hipóteses ST.1' a ST.5', desde que sejam estacionários em torno de sua tendência. Desde que as tendências temporais sejam incluídas nas equações quando necessárias, os procedimentos usuais de inferência são assintoticamente válidos.

Suponha que as expectativas sejam formadas como $\text{inf}_t^e = (1/2)\text{inf}_{t-1} + (1/2)\text{inf}_{t-2}$. Que regressão você deve computar para estimar a curva de Phillips de expectativas aumentadas?

11.3 O USO DE SÉRIES TEMPORAIS ALTAMENTE PERSISTENTES NA ANÁLISE DE REGRESSÃO

A seção anterior mostrou que, desde que a série temporal que usamos seja fracamente dependente, os procedimentos habituais de inferência do MQO são válidos sob hipóteses mais fracas do que aquelas do modelo linear clássico. Infelizmente, muitas das séries temporais econômicas não podem ser caracterizadas pela dependência fraca. O uso de séries temporais com forte dependência em análise de regressão não representa um problema, se as hipóteses MLC do Capítulo 10 se sustentarem. Entretanto, os procedimentos usuais de inferência são muito suscetíveis à violação dessas hipóteses quando os

dados não são fracamente dependentes, porque, então, não podemos recorrer à lei dos grandes números e ao teorema do limite central. Nesta seção, apresentamos alguns exemplos de séries temporais **altamente persistentes** (ou **fortemente dependentes**) e mostramos como elas podem ser transformadas para o uso em análise de regressão.

Séries Temporais Altamente Persistentes

No modelo AR(1) simples (11.2), a hipótese $|\rho_1| < 1$ é crucial para que a série seja fracamente dependente. Constatase que muitas séries temporais econômicas são melhor caracterizadas pelo modelo AR(1) com $\rho_1 = 1$. Neste caso, podemos escrever

$$y_t = y_{t-1} + e_t, t = 1, 2, \dots, \quad (11.20)$$

onde novamente assumimos que $\{e_t; t = 1, 2, \dots\}$ é independente e identicamente distribuída com média zero e variância σ_e^2 . Assumimos que o valor inicial, y_0 , é independente de e_t para todo $t \geq 1$.

O processo em (11.20) é chamado de **passeio aleatório**. O nome vem do fato de que y no tempo t é obtido a partir de seu valor anterior, y_{t-1} , e adicionando uma variável aleatória de média zero que é independente de y_{t-1} . Algumas vezes, um passeio aleatório é definido de maneira diferente, assumindo propriedades diferentes das inovações, e_t , (como a falta de correlação em lugar da independência), mas a definição corrente é suficiente para nosso propósito.

Primeiro, encontramos o valor esperado de y_t . Isso é feito com muita facilidade utilizando-se substituições repetidas, para obter

$$y_t = e_t + e_{t-1} + \dots + e_1 + y_0.$$

Tomando os valores esperados de ambos os lados, obtém-se

$$E(y_t) = E(e_t) + E(e_{t-1}) + \dots + E(e_1) + E(y_0) = E(y_0), \text{ para todo } t \geq 1.$$

Portanto, o valor esperado de um passeio aleatório *não* depende de t . Uma hipótese popular é que $y_0 = 0$ — o processo começa com zero no tempo zero — quando $E(y_t) = 0$ para todos os t .

Em contrapartida, a variância de um passeio aleatório muda com t . Para calcular a variância de um passeio aleatório, por simplicidade assumimos que y_0 é não-aleatório, de forma que $\text{Var}(y_0) = 0$; isso não afeta nenhuma conclusão importante. Então, pela hipótese i.i.d. de $\{e_t\}$,

$$\text{Var}(y_t) = \text{Var}(e_t) + \text{Var}(e_{t-1}) + \dots + \text{Var}(e_1) = \sigma_e^2 t. \quad (11.21)$$

Em outras palavras, a variância de um passeio aleatório aumenta como uma função linear do tempo. Isso mostra que o processo não pode ser estacionário.

Mais importante ainda, um passeio aleatório exibe um comportamento altamente persistente no sentido de que o valor de y hoje é significante para determinar o valor de y em um futuro bem distante. Par ver isso, escreva para h períodos

$$y_{t+h} = e_{t+h} + e_{t+h-1} + \dots + e_{t+1} + y_t.$$

Agora, suponha que queiramos calcular o valor esperado de y_{t+h} no tempo t , dado o valor corrente de y_t . Como o valor esperado de e_{t+j} , dado y_t , é zero para todo $j \geq 1$, teremos

$$E(y_{t+h}|y_t) = y_t, \text{ para todo } h \geq 1. \quad (11.22)$$

Isso significa que, não importa o quanto esteja distante o futuro que focarmos, nossa melhor previsão de y_{t+h} será o valor de hoje, y_t . Podemos contrapor isso com o caso do AR(1) estável, no qual um argumento semelhante pode ser usado para mostrar que

$$E(y_{t+h}|y_t) = \rho_1^h y_t, \text{ para todo } h \geq 1.$$

Sob estabilidade, $|\rho_1| < 1$, e, portanto, $E(y_{t+h}|y_t)$ se aproxima de zero à medida que $h \rightarrow \infty$: o valor de y_t se torna cada vez menos importante, e $E(y_{t+h}|y_t)$ se aproxima mais e mais do valor esperado incondicional, $E(y_t) = 0$.

Quando $h = 1$, a equação (11.22) é remanescente da hipótese de expectativas adaptativas que utilizamos para a taxa de inflação no Exemplo 11.5: se a inflação seguir um passeio aleatório, o valor esperado de \inf_t , dados os valores passados da inflação, será, simplesmente, \inf_{t-1} . Assim, um modelo de passeio aleatório da inflação justifica o uso de expectativas adaptativas.

Também podemos verificar que a correlação entre y_t e y_{t+h} fica próxima de 1 para t grande quando $\{y_t\}$ segue um passeio aleatório. Se $\text{Var}(y_0) = 0$, pode ser mostrado que

$$\text{Corr}(y_t, y_{t+h}) = \sqrt{t/(t+h)}.$$

Portanto, a correlação depende do ponto de partida, t (de forma que $\{y_t\}$ não tem covariância estacionária). Além disso, embora para t fixo a correlação tenda para zero à medida que $h \rightarrow \infty$, isso não acontece com muita rapidez. De fato, quanto maior for t , mais lentamente a correlação tenderá para zero à medida que h fique maior. Se selecionarmos um h grande — digamos, $h = 100$ —, poderemos sempre selecionar um t suficientemente grande, de tal forma que a correlação entre y_t e y_{t+h} seja arbitrariamente próxima de 1. (Se $h = 100$ e queremos que a correlação seja maior que 0,95, então, $t > 1.000$ resolverá o problema.) Portanto, um passeio aleatório não satisfaz os requisitos de uma seqüência assintoticamente não-correlacionada.

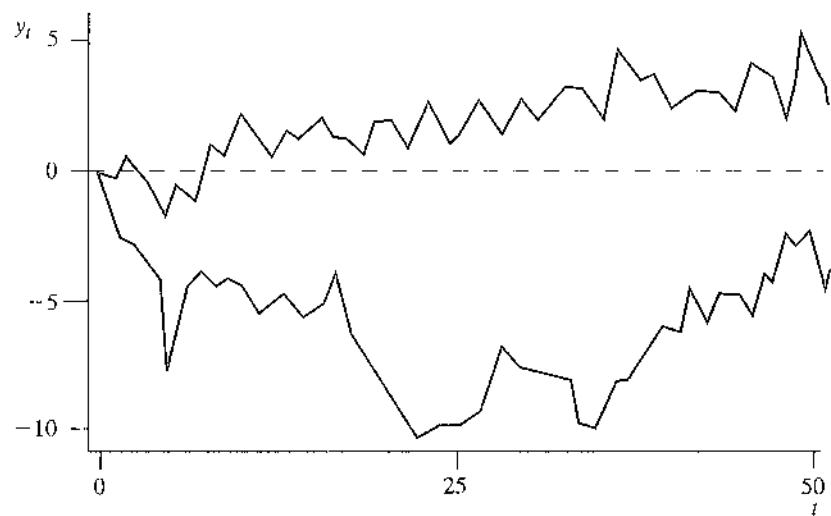
A Figura 11.1 mostra duas realizações de um passeio aleatório com valor inicial $y_0 = 0$ e $e_t \sim \text{Normal}(0,1)$. Geralmente, não é fácil olhar um gráfico de série temporal e determinar se ela é ou não um passeio aleatório. Em seguida, discutiremos um método informal para fazer a distinção entre seqüências fracamente e altamente dependentes; estudaremos testes estatísticos formais no Capítulo 18.

Uma série que geralmente é considerada bem caracterizada como um passeio aleatório é a da taxa de juros das letras do Tesouro norte-americano de três meses. Os dados anuais estão traçados na Figura 11.2 para os anos de 1948 até 1996.

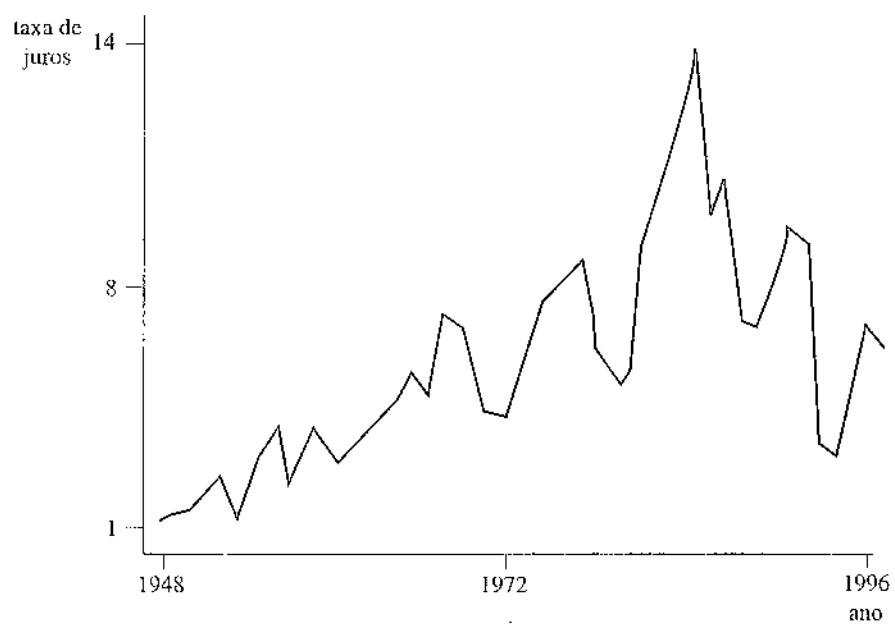
Um passeio aleatório é um caso especial do que é conhecido como **processo de raiz unitária**. O nome vem do fato de que $\rho_1 = 1$ no modelo AR(1). Uma classe mais geral de processos de raiz unitária é gerada como em (11.20), mas agora é permitido que $\{e_t\}$ seja uma série geral, fracamente dependente. [Por exemplo, $\{e_t\}$ poderia seguir um processo MA(1) ou AR(1) estável.] Quando $\{e_t\}$ não é uma seqüência i.i.d., as propriedades do passeio aleatório que derivamos anteriormente não se sustentam mais. Mas a parte essencial de $\{e_t\}$ é preservada: o valor de y hoje é altamente correlacionado com y , mesmo em um futuro distante.

Figura 11.1

Duas realizações do passeio aleatório $y_t = y_{t-1} + e_t$ com $y_0 = 0$, $e_t \sim \text{Normal}[0,1]$, e $n = 50$.

**Figura 11.2**

Taxas de juros das letras do Tesouro dos Estados Unidos de três meses para os anos 1948-1996.

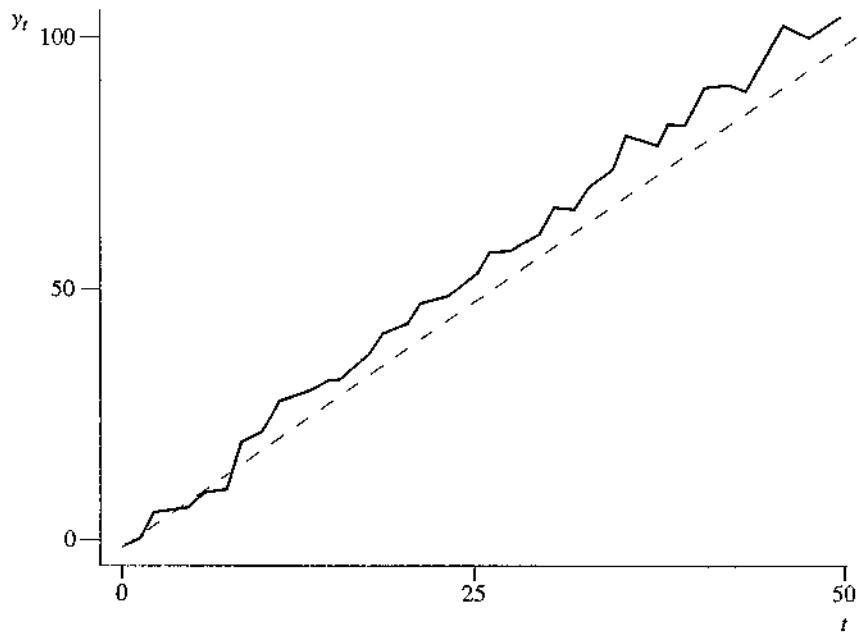


De uma perspectiva de política econômica, muitas vezes é importante saber se uma série temporal econômica é, ou não, altamente persistente. Considere o caso do produto interno bruto dos Estados Unidos. Se o PIB for assimptoticamente não-correlacionado, então, o nível do PIB no ano seguinte será, na melhor das hipóteses, fracamente relacionado com o que tenha sido o PIB, digamos, 30 anos atrás. Isso significa que uma decisão de política econômica que afetou o PIB há muito tempo produz um impacto pouco duradouro. De outro lado, se o PIB for fortemente dependente, o PIB do próximo ano poderá ser altamente correlacionado com o PIB de muitos anos atrás. Então, devemos reconhecer que uma decisão de política econômica que cause uma discreta alteração no PIB pode ter efeitos duradouros.

É extremamente importante não confundir comportamento de tendência com comportamento altamente persistente. Uma série pode apresentar tendência, mas não ser altamente persistente, como vimos no Capítulo 10. Além disso, fatores tais como taxas de juros, taxas de inflação e taxas de desemprego são tidos por muitos como sendo altamente persistentes, mas não mostram tendência óbvia de alta ou de baixa. Porém, muitas vezes acontece de uma série altamente persistente também conter uma clara tendência. Um modelo que leva a esse comportamento é o **passeio aleatório com deslocamento**:

Figura 11.3

Uma realização do passeio aleatório com deslocamento, $y_t = 2 + y_{t-1} + e_t$, com $y_0 = 0$, $e_t \sim \text{Normal}(0,9)$, e $n = 50$. A linha tracejada é o valor esperado de y_t , $E[y_t] = 2t$.



$$y_t = \alpha_0 + y_{t-1} + e_t, t = 1, 2, \dots,$$

(11.23)

onde $\{e_t; t = 1, 2, \dots\}$ e y_0 satisfazem as mesmas propriedades do modelo de passeio aleatório. A novidade é o parâmetro α_0 , que é chamado de *termo de deslocamento*. Basicamente, para gerar y_t , a constante α_0 é adicionada com o ruído aleatório e_t ao valor anterior y_{t-1} . Podemos mostrar que o valor esperado de y_t segue uma tendência temporal linear usando substituições repetidas:

$$y_t = \alpha_0 t + e_t + e_{t-1} + \dots + e_1 + y_0.$$

Portanto, se $y_0 = 0$, $E(y_t) = \alpha_0 t$: o valor esperado de y_t será crescente ao longo do tempo se $\alpha_0 > 0$ e decrescente ao longo do tempo se $\alpha_0 < 0$. Raciocinando da mesma forma que fizemos no caso puro do passeio aleatório, podemos mostrar que $E(y_{t+h}|y_t) = \alpha_0 h + y_t$, e, assim, a melhor previsão de y_{t+h} no tempo t será y_t , mais o deslocamento $\alpha_0 h$. A variância de y_t é a mesma do caso puro de passeio aleatório.

A Figura 11.3 contém uma realização do passeio aleatório com deslocamento, em que $n = 50$, $y_0 = 0$, $\alpha_0 = 2$, e os e_t são variáveis aleatórias com distribuição Normal(0,9). Como pode ser visto neste gráfico, y_t tende a crescer ao longo do tempo, mas a série não retorna de forma regular à linha de tendência.

Um passeio aleatório com deslocamento é outro exemplo de um processo de raiz unitária, pois trata-se do caso especial $\rho_1 = 1$ em um modelo AR(1) com um intercepto:

$$y_t = \alpha_0 + \rho_1 y_{t-1} + e_t$$

Quando $\rho_1 = 1$ e $\{e_t\}$ é qualquer processo fracamente dependente, obtemos uma classe inteira de processos de séries temporais altamente persistentes que também podem ter médias com tendência linear.

Transformações de Séries Temporais Altamente Persistentes

A utilização de séries temporais com forte persistência do tipo exibido por um processo de raiz unitária em uma equação de regressão pode levar a resultados enganosos se as hipóteses do MLC forem violadas. Estudaremos o problema da regressão espúria mais detalhadamente no Capítulo 18, mas, por ora, devemos estar atentos a seus problemas potenciais. Felizmente, existem transformações simples que produzem um processo de raiz unitária fracamente dependente.

Processos fracamente dependentes são chamados de **integrados de ordem zero** ou **I(0)**. Na prática, isso significa que nada precisa ser feito em tais séries antes de utilizá-las na análise de regressão: as médias de tais seqüências já satisfazem os teoremas de limite padrão. Processos de raiz unitária, tais como o de passeio aleatório (com ou sem deslocamento), são chamados de **integrados de ordem um** ou **I(1)**. Isso significa que a **primeira diferença** do processo é fracamente dependente (e geralmente estacionária).

Isso é simples de observar em um passeio aleatório. Com $\{y_t\}$ gerado como em (11.20) para $t = 1, 2, \dots$,

$$\Delta y_t = y_t - y_{t-1} = e_t, t = 2, 3, \dots; \quad (11.24)$$

e, portanto, a série em primeiras diferenças $\{\Delta y_t; t = 2, 3, \dots\}$ é, efetivamente, uma seqüência i.i.d. De maneira mais geral, se $\{y_t\}$ for gerada por (11.24) onde $\{e_t\}$ é qualquer processo fracamente dependente, então, $\{\Delta y_t\}$ será fracamente dependente. Assim, quando suspeitamos que os processos sejam integrados de ordem um, freqüentemente trabalhamos com suas primeiras diferenças para poder usá-los em análise de regressão; veremos alguns exemplos mais tarde.

Muitas séries temporais y_t são tão estritamente positivas que $\log(y_t)$ é integrado de ordem um. Nesse caso, podemos usar a primeira diferença nos logs, $\Delta \log(y_t) = \log(y_t) - \log(y_{t-1})$, na análise de regressão. Alternativamente, como

$$\Delta \log(y_t) \approx (y_t - y_{t-1})/y_{t-1}, \quad (11.25)$$

podemos usar diretamente a mudança proporcional ou percentual em y_t ; foi o que fizemos no Exemplo 11.4, no qual, em vez de especificarmos a hipótese de mercados eficientes em termos do preço das ações, p_t , utilizamos a mudança percentual semanal, $retorno_t = 100[(p_t - p_{t-1})/p_{t-1}]$.

Diferenciar séries de tempo antes de usá-las na análise de regressão tem outro benefício: remove qualquer tendência temporal linear. Isso pode facilmente ser visto escrevendo-se uma variável que apresenta tendência linear como

$$y_t = \gamma_0 + \gamma_1 t + \nu_t,$$

onde ν_t tem média zero. Então, $\Delta y_t = \gamma_1 + \Delta \nu_t$, e assim $E(\Delta y_t) = \gamma_1 + E(\Delta \nu_t) = \gamma_1$. Em outras palavras, $E(\Delta y_t)$ é constante. O mesmo argumento funciona para $\Delta \log(y_t)$ quando $\log(y_t)$ segue uma tendência temporal linear. Portanto, em lugar de incluir uma tendência temporal em uma regressão, podemos, em vez disso, diferenciar as variáveis que exibem tendências óbvias.

A Decisão sobre uma Série de Tempo Ser I(1)

Determinar se a realização de uma série de tempo específica é o resultado de um processo I(1) ou de um processo I(0) pode ser muito difícil. Testes estatísticos podem ser usados para esse propósito, mas trata-se de assunto mais avançado; apresentaremos um tratamento introdutório no Capítulo 18.

Existem métodos informais que fornecem orientações úteis sobre se um processo de série de tempo é, em termos gerais, caracterizado por fraca dependência. Uma ferramenta bastante simples é induzida pelo modelo AR(1): se $|\rho_1| < 1$, então, o processo será I(0), mas será I(1) se $\rho_1 = 1$. Mostramos previamente que, quando o processo AR(1) é estável, $\rho_1 = \text{Corr}(y_t, y_{t-1})$. Portanto, podemos estimar ρ_1 da correlação amostral entre y_t e y_{t-1} . Este coeficiente de correlação amostral é chamado de **autocorrelação de primeira ordem** de $\{y_t\}$; ele é representado por $\hat{\rho}_1$. Aplicando-se a lei dos grandes números, pode-se mostrar que $\hat{\rho}_1$ é um estimador consistente de ρ_1 , desde que $|\rho_1| < 1$. (Porém, $\hat{\rho}_1$ não é um estimador não-viesado de ρ_1 .)

Podemos usar o valor de $\hat{\rho}_1$ para nos ajudar a decidir se o processo é I(1) ou I(0). Infelizmente, como $\hat{\rho}_1$ é uma estimativa, nunca saberemos com certeza se $\rho_1 < 1$. Em condições ideais, podemos calcular um intervalo de confiança de ρ_1 para verificar se ele exclui o valor $\rho_1 = 1$, mas isso acaba sendo muito difícil de se fazer: as distribuições amostrais do estimador $\hat{\rho}_1$ são extremamente diferentes quando ρ_1 é próximo de um e quando ρ_1 é muito menor que um. (De fato, quando ρ_1 é próximo de um, $\hat{\rho}_1$ pode ter um severo viés para baixo.)

No Capítulo 18, mostraremos como testar $H_0: \rho_1 = 1$ contra $H_1: \rho_1 < 1$. Por ora, somente podemos usar $\hat{\rho}_1$ como uma orientação em termos gerais para determinar se uma série necessita ser diferenciada. Não existe uma regra rígida e imediata para fazer essa escolha. A maioria dos economistas acredita que a diferenciação é garantida se $\hat{\rho}_1 > 0,9$; alguns fazem a diferenciação quando $\hat{\rho}_1 > 0,8$.

EXEMPLO 11.6**(Equação da Fertilidade)**

No Exemplo 10.4, explicamos a taxa geral de fertilidade, tgf , em termos do valor da isenção pessoal de impostos, ip . As autocorrelações de primeira ordem dessas séries são bastante grandes: $\hat{\rho}_1 = 0,977$ para tgf e $\hat{\rho}_1 = 0,964$ para ip . Essas autocorrelações são altamente sugestivas de comportamento de raiz unitária e levantam sérias dúvidas sobre nossa utilização das estatísticas normais de MQO neste exemplo no Capítulo 10. Lembre-se, as estatísticas t somente têm distribuições t exatas sob o conjunto total das hipóteses do modelo linear clássico. Para relaxarmos essas hipóteses em qualquer forma e aplicarmos propriedades assimptóticas, geralmente necessitaremos que as séries básicas sejam processos $I(0)$.

Agora estimaremos a equação utilizando primeiras diferenças (e eliminamos a variável *dummy*, por simplicidade):

$$\begin{aligned}\Delta \hat{t}gf &= -0,785 - 0,043 \Delta ip \\ &\quad (0,502) (0,028) \\ n &= 71, R^2 = 0,032, \bar{R}^2 = 0,018.\end{aligned}\tag{11.26}$$

Agora, estima-se que um aumento em ip reduza tgf contemporaneamente, embora a estimativa não seja estatisticamente diferente de zero ao nível de 5%. Isso produz resultados bastante diferentes do que quando estimamos o modelo em níveis, e lança dúvidas sobre nossa análise anterior.

Se adicionarmos duas defasagens de Δip , as coisas melhoram:

$$\begin{aligned}\Delta \hat{t}gf &= -0,964 - 0,036 \Delta ip - 0,014 \Delta ip_{-1} + 0,110 \Delta ip_{-2} \\ &\quad (0,468) (0,027) (0,028) (0,027) \\ n &= 69, R^2 = 0,233, \bar{R}^2 = 0,197.\end{aligned}\tag{11.27}$$

Embora Δip e Δip_{-1} tenham coeficientes negativos, seus coeficientes são pequenos e conjuntamente não significantes (p -valor = 0,28). A segunda defasagem é bastante significante e indica uma relação positiva entre as mudanças em ip e as mudanças subsequentes em tgf a partir de dois anos. Isso faz mais sentido do que ter um efeito contemporâneo. Veja o Exercício 11.12 para análise adicional da equação em primeiras diferenças.

Quando a série em questão tem uma tendência óbvia crescente ou decrescente, faz mais sentido obter a autocorrelação de primeira ordem após a retirada da tendência. Se não forem removidas as tendências dos dados, a correlação auto-regressiva tende a ser superestimada, o que viesará o processo na direção de encontrar uma raiz unitária, nesse processo com tendência.

EXEMPLO 11.7**(Salários e Produtividade)**

A variável $salhr$ é a média de salário por hora na economia dos Estados Unidos, e $prodhr$ é a produção por hora. Uma maneira de estimar a elasticidade do salário por hora em relação à produção por hora é estimar a equação,

EXEMPLO 11.7 {continuação}

$$\log(\text{salhr}_t) = \beta_0 + \beta_1 \log(\text{prodhr}_t) + \beta_2 t + u_t,$$

onde a tendência temporal é incluída porque tanto $\log(\text{salhr}_t)$ quanto $\log(\text{prodhr}_t)$ mostram tendências lineares claras e crescentes. Utilizando os dados contidos no arquivo EARNS.RAW para os anos de 1947 até 1987, obtemos

$$\begin{aligned} \log(\hat{\text{salhr}}_t) &= -5,33 + 1,64 \log(\text{prodhr}_t) - 0,018 t \\ &\quad (0,37) \quad (0,09) \quad (0,002) \\ n &= 41, R^2 = 0,971, \bar{R}^2 = 0,970. \end{aligned} \tag{11.28}$$

(Neste caso, descrevemos as medidas do grau de ajuste; seria melhor apresentar essas informações com base na variável dependente sem a tendência, como na Seção 10.5.) A elasticidade estimada parece ser grande demais: 1% de aumento na produtividade aumenta o salário real em cerca de 1,64%. Como o erro-padrão é muito pequeno, o intervalo de confiança de 95% exclui facilmente uma elasticidade unitária. Os trabalhadores norte-americanos provavelmente teriam problemas em acreditar que seus salários subiriam em mais de 1,5% para cada 1% de aumento na produtividade.

Os resultados da regressão em (11.28) devem ser vistos com cautela. Mesmo depois de retirar linearmente a tendência de $\log(\text{salhr})$, a autocorrelação de primeira ordem é 0,967, e para $\log(\text{prodhr})$, depois da retirada de tendência, $\hat{\rho}_1 = 0,945$. Esses dados sugerem que ambas as séries possuem raízes unitárias, de modo que recalculamos a equação em primeiras diferenças (e não mais necessitamos de uma tendência temporal):

$$\begin{aligned} \Delta \log(\hat{\text{salhr}}_t) &= -0,0036 + 0,809 \Delta \log(\text{prodhr}) \\ &\quad (0,0042) \quad (0,173) \\ n &= 40, R^2 = 0,364, \bar{R}^2 = 0,348. \end{aligned} \tag{11.29}$$

Agora, estima-se que um aumento de 1% na produtividade aumenta os salários reais em cerca de 0,81%, e a estimativa não é, estatisticamente, diferente de um. O R -quadrado ajustado mostra que o crescimento na produtividade explica cerca de 35% do crescimento nos salários reais. Veja o Exercício 11.9 para uma verificação simples de defasagens distribuídas do modelo em primeiras diferenças.

Nos dois exemplos anteriores, tanto as variáveis dependentes quanto as independentes parecem ter raízes unitárias. Em outros casos, podemos ter uma mistura de processos com raízes unitárias com outros que são fracamente dependentes (embora possivelmente apresentando tendência). Um exemplo disso é dado no Exercício 11.8.

11.4 MODELOS DINAMICAMENTE COMPLETOS E A AUSÊNCIA DE CORRELAÇÃO SERIAL

No modelo AR(1) em (11.12), mostramos que, sob a hipótese (11.13), os erros $\{u_t\}$ devem ser **serialmente não-correlacionados**, no sentido de que a hipótese ST.5' seja satisfeita: assumir que não existe

correlação serial é praticamente a mesma coisa que assumir apenas o aparecimento de uma defasagem de y em $E(y_t|y_{t-1}, y_{t-2}, \dots)$.

Podemos fazer uma afirmação semelhante para outros modelos de regressão? A resposta é sim. Considere o modelo de regressão simples estática

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad (11.30)$$

onde y_t e z_t são contemporaneamente datados. Para a consistência de MQO, apenas necessitamos de que $E(u_t|z_t) = 0$. Geralmente, $\{u_t\}$ será serialmente correlacionado. Porém, se *assumirmos* que

$$E(u_t|z_t, y_{t-1}, z_{t-1}, \dots) = 0, \quad (11.31)$$

então (como veremos mais tarde de forma generalizada), a hipótese ST.5' se sustenta. Em particular, os $\{u_t\}$ são serialmente não-correlacionados. Naturalmente, a hipótese (11.31) implica que z_t é contemporaneamente exógeno, ou seja, $E(u_t|z_t) = 0$.

Para obter uma idéia do significado de (11.31), podemos escrever (11.30) e (11.31) equivalentemente como

$$E(y_t|z_t, y_{t-1}, z_{t-1}, \dots) = E(y_t|z_t) = \beta_0 + \beta_1 z_t, \quad (11.32)$$

onde a primeira igualdade é aquela de interesse corrente. Ela diz que, uma vez que z_t esteja sob controle, nenhuma defasagem de y ou de z ajuda a explicar y corrente. Esse é um requisito forte; se ele for falso, então, podemos esperar que os erros serão serialmente correlacionados.

A seguir, considere um modelo de defasagens distribuídas finitas com duas defasagens:

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 z_{t-1} + \beta_3 z_{t-2} + u_t. \quad (11.33)$$

Como esperamos captar os efeitos defasados que z tem sobre y , assumimos naturalmente que (11.33) captura a dinâmica das defasagens distribuídas:

$$E(y_t|z_t, z_{t-1}, z_{t-2}, z_{t-3}, \dots) = E(y_t|z_t, z_{t-1}, z_{t-2}); \quad (11.34)$$

ou seja, no máximo duas defasagens de z importam. Se (11.31) se mantiver, podemos fazer outras afirmações: uma vez que tenhamos o controle de z e de suas duas defasagens, nenhuma defasagem de y ou defasagens adicionais de z afetarão y corrente:

$$E(y_t|z_t, y_{t-1}, z_{t-1}, \dots) = E(y_t|z_t, z_{t-1}, z_{t-2}). \quad (11.35)$$

A equação (11.35) é mais plausível que (11.32), mas ainda exclui y defasado afetando y corrente.

A seguir, considere um modelo com uma defasagem tanto em y quanto em z :

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t$$

Como esse modelo inclui uma variável dependente defasada, (11.31) é uma hipótese natural, já que ela implica que

$$E(y_t|z_t, y_{t-1}, z_{t-1}, y_{t-2}, \dots) = E(y_t|z_t, y_{t-1}, z_{t-1});$$

em outras palavras, uma vez que z_t , y_{t-1} e z_{t-1} tenham sido controladas, nenhuma outra defasagem de y ou de z afetará y corrente.

No modelo geral

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t, \quad (11.36)$$

onde as variáveis explicativas $\mathbf{x}_t = (x_{t1}, \dots, x_{tk})$ podem ou não conter defasagens de y ou z , (11.31) torna-se

$$E(u_t|x_t, y_{t-1}, x_{t-1}, \dots) = 0. \quad (11.37)$$

Escrita em termos de y_t ,

$$E(y_t|x_t, y_{t-1}, x_{t-1}, \dots) = E(y_t|x_t). \quad (11.38)$$

Em outras palavras, seja o que for que esteja em \mathbf{x}_t , defasagens suficientes foram incluídas, de forma que defasagens adicionais de y e das variáveis explicativas não são importantes para explicar y_t . Quando essa condição se mantém, temos um **modelo dinamicamente completo**. Como já vimos, a plenitude dinâmica pode ser uma hipótese muito forte para os modelos estáticos e os de defasagens distribuídas finitas.

Uma vez que começamos a colocar defasagens de y como variáveis explicativas, freqüentemente pensamos que o modelo deve ser dinamicamente completo. Mencionaremos algumas exceções a essa alegação no Capítulo 18.

Como (11.37) é equivalente a

$$E(u_t|x_t, u_{t-1}, x_{t-1}, u_{t-2}, \dots) = 0, \quad (11.39)$$

podemos mostrar que um modelo dinamicamente completo tem que satisfazer a hipótese ST.5'. (Esta derivação não é crucial e pode ser ignorada sem provocar perda de continuidade.) Concretamente, considere $s < t$. Então, pela lei das expectativas iteradas (veja Apêndice B no site da Thomson),

$$\begin{aligned} E(u_t, u_s | \mathbf{x}_t, \mathbf{x}_s) &= E[E(u_t, u_s | \mathbf{x}_t, \mathbf{x}_s, u_s) | \mathbf{x}_t, \mathbf{x}_s] \\ &= E[u_s E(u_t | \mathbf{x}_t, \mathbf{x}_s, u_s) | \mathbf{x}_t, \mathbf{x}_s], \end{aligned}$$

onde a segunda igualdade resulta de $E(u_t u_s | \mathbf{x}_t, \mathbf{x}_s, u_s) = u_s E(u_t | \mathbf{x}_t, \mathbf{x}_s, u_s)$. Agora, como $s < t$, $(\mathbf{x}_t, \mathbf{x}_s, u_s)$ é um subconjunto do conjunto condicional em (11.39). Portanto, (11.39) implica que $E(u_t | \mathbf{x}_t, \mathbf{x}_s, u_s) = 0$, e assim

$$E(u_t u_s | \mathbf{x}_t, \mathbf{x}_s) = E(u_s \cdot 0 | \mathbf{x}_t, \mathbf{x}_s) = 0,$$

o que diz que a hipótese ST.5' é válida.

Dado que especificar um modelo dinamicamente completo significa que não há correlação serial, subentende-se que todos os modelos deveriam ser dinamicamente completos? Como veremos no Capítulo 18, para o propósito de previsão, a resposta é sim. Algumas pessoas acham que todos os modelos deveriam ser dinamicamente completos e que a correlação serial nos erros de um modelo é um sinal de má especificação. Esse posicionamento é rígido demais. Algumas vezes, estamos realmente interessados em um modelo estático (como o da curva de Phillips) ou em um modelo de defasagens distribuídas finitas (como no caso de querermos medir a alteração percentual nos salários a longo prazo, dado um aumento de 1% na produtividade). No próximo capítulo, mostraremos como detectar e corrigir a correlação serial em tais modelos.

Se (11.33) se mantiver, onde $u_t = e_t + \alpha_1 e_{t-1}$, e onde $\{e_t\}$ é uma seqüência i.i.d. com média zero e variância σ_e^2 , pode a equação (11.33) ser dinamicamente completa?

EXEMPLO 11.8

(Equação da Fertilidade)

Na equação (11.27), estimamos um modelo de defasagens distribuídas de Δtgf sobre Δip , permitindo duas defasagens de Δip . Para que esse modelo seja dinamicamente completo, no sentido de (11.38), nem defasagens de Δtgf nem defasagens adicionais de Δip devem aparecer na equação. Podemos facilmente ver que isso é falso se adicionarmos Δtgf_{-1} : o coeficiente estimado é 0,300, e sua estatística t é 2,84. Assim, o modelo não é dinamicamente completo no sentido de (11.38).

O que devemos concluir disso? Deixaremos a interpretação de modelos gerais com variáveis dependentes defasadas para o Capítulo 18. Mas o fato de que (11.27) não é dinamicamente completo sugere que pode haver correlação serial nos erros. Veremos como testar e corrigir isso no Capítulo 12.

11.5 A HIPÓTESE DE HOMOSCEDASTICIDADE PARA MODELOS DE SÉRIES TEMPORAIS

A hipótese de homoscedasticidade para regressões de séries temporais, particularmente ST.4', parece muito semelhante à das regressões de corte transversal. Porém, como x_t pode conter defasagens de y , como também das variáveis explicativas, discutiremos brevemente o significado da hipótese de homoscedasticidade em diferentes regressões de séries temporais.

No modelo estático simples, digamos

$$y_t = \beta_0 + \beta_1 z_t + u_t, \quad (11.40)$$

a hipótese ST.4' exige que

$$\text{Var}(u_t | z_t) = \sigma^2.$$

Portanto, embora $E(y_t | z_t)$ seja uma função linear de z_t , $\text{Var}(y_t | z_t)$ deve ser constante. Isso é bastante claro.

No Exemplo 11.4, vimos que, no modelo AR(1) em (11.2), a hipótese de homoscedasticidade era

$$\text{Var}(u_t | y_{t-1}) = \text{Var}(y_t | y_{t-1}) = \sigma^2;$$

embora $E(y_t | y_{t-1})$ dependa de y_{t-1} , $\text{Var}(y_t | y_{t-1})$ não depende. Assim, a dispersão na distribuição de y_t não pode depender de y_{t-1} .

Esperamos que o padrão esteja claro agora. Se tivermos o modelo

$$y_t = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1} + u_t,$$

a hipótese de homoscedasticidade será

$$\text{Var}(u_t | z_t, y_{t-1}, z_{t-1}) = \text{Var}(y_t | z_t, y_{t-1}, z_{t-1}) = \sigma^2,$$

de forma que a variância de u_t não pode depender de z_t , y_{t-1} ou de z_{t-1} (ou de qualquer outra função do tempo). De forma geral, quaisquer que sejam as variáveis explicativas que apareçam no modelo, devemos assumir que a variância de y_t , dadas essas variáveis explicativas, é constante. Se o modelo contiver defasagens de y ou das variáveis explicativas, estaremos explicitamente excluindo formas dinâmicas de heteroscedasticidade (assunto que estudaremos no Capítulo 12). Contudo, em um modelo estático, estamos apenas preocupados com $\text{Var}(y_t | z_t)$. Na equação (11.40), nenhuma restrição direta foi colocada sobre, digamos, $\text{Var}(y_t | y_{t-1})$.

Neste capítulo, argumentamos que o método MQO pode ser justificado usando análise assintótica, desde que certas condições sejam atendidas. De maneira ideal, os processos de séries temporais são estacionários e fracamente dependentes, embora a estacionariedade não seja crucial. A dependência fraca é necessária para aplicar os resultados padrão de amostras grandes, particularmente o teorema do limite central.

Processos com tendências deterministas que sejam fracamente dependentes podem ser usados diretamente na análise de regressão, desde que tendências temporais sejam incluídas no modelo (como na Seção 10.5). Uma afirmação semelhante é válida para processos com sazonalidade.

Quando as séries de tempo são altamente persistentes (elas possuem raízes unitárias), devemos ter extrema cautela ao usá-las diretamente em modelos de regressão (a menos que estejamos convencidos

de que as hipóteses do MLC do Capítulo 10 sejam válidas). Uma alternativa em relação a usarmos os níveis é usarmos as primeiras diferenças das variáveis. Para a maioria das séries temporais econômicas, a primeira diferença é fracamente dependente. O uso de primeiras diferenças altera a natureza do modelo, mas esse método é, muitas vezes, tão informativo quanto um modelo em níveis. Quando os dados forem altamente persistentes, normalmente teremos mais confiança nos resultados das primeiras diferenças. No Capítulo 18, cobriremos alguns métodos mais recentes e avançados para usarmos variáveis I(1) na análise de regressão múltipla.

Quando os modelos têm dinâmicas completas, no sentido de que nenhuma defasagem adicional de qualquer variável é necessária na equação, vimos que os erros serão serialmente não-correlacionados. Isso é proveitoso porque certos modelos, como os modelos auto-regressivos, são assumidos como tendo dinâmicas completas. Em modelos estáticos e com defasagens distribuídas, a hipótese de eles serem dinamicamente completos quase sempre é falsa, o que geralmente significa que os erros serão serialmente correlacionados. Veremos como avaliar esse problema no Capítulo 12.

11.1 Seja $\{x_t; t = 1, 2, \dots\}$ um processo de covariância estacionária e defina $\gamma_h = \text{Cov}(x_t, x_{t+h})$ para $h \geq 0$. [Portanto, $\gamma_0 = \text{Var}(x_t)$]. Mostre que $\text{Cov}(x_t, x_{t+h}) = \gamma_h/\gamma_0$.

11.2 Seja $\{e_t; t = -1, 0, 1, \dots\}$ uma seqüência de variáveis aleatórias independentes identicamente distribuídas, com média zero e variância um. Defina um processo estocástico por

$$x_t = e_t - (1/2)e_{t-1} + (1/2)e_{t-2}, t = 1, 2, \dots$$

- (i) Encontre $E(x_t)$ e $\text{Var}(x_t)$. Algum deles depende de t ?
- (ii) Mostre que $\text{Corr}(x_t, x_{t+1}) = 1/2$ e que $\text{Corr}(x_t, x_{t+2}) = 1/3$. (*Sugestão:* a maneira mais fácil é usar a fórmula do Problema 11.1.)
- (iii) Qual será a $\text{Corr}(x_t, x_{t+h})$ para $h > 2$?
- (iv) $\{x_t\}$ é um processo assintoticamente não-correlacionado?

11.3 Suponha que um processo de série temporal $\{y_t\}$ seja gerado por $y_t = z + e_t$, para todo $t = 1, 2, \dots$, onde $\{e_t\}$ é uma seqüência i.i.d. com média zero e variância σ_e^2 . A variável aleatória z não muda ao longo do tempo; ela tem média zero e variância σ_z^2 . Assuma que cada e_t seja não-correlacionado com z .

- (i) Encontre o valor esperado e a variância de y_t . Suas respostas dependem de t ?
- (ii) Encontre $\text{Cov}(y_t, y_{t+h})$ para quaisquer t e h . O processo $\{y_t\}$ tem covariância estacionária?
- (iii) Utilize as partes (i) e (ii) para mostrar que $\text{Cov}(y_t, y_{t+h}) = \sigma_z^2 / (\sigma_z^2 + \sigma_e^2)$ para todo t e h .
- (iv) y_t satisfaz o requisito intuitivo de ser assintoticamente não-correlacionada? Explique.

11.4 Suponha que $\{y_t; t = 1, 2, \dots\}$ siga um passeio aleatório, como em (11.20), com $y_0 = 0$. Mostre que $\text{Corr}(y_t, y_{t+h}) = \sqrt{t/(t+h)}$ para $t \geq 1, h > 0$.

11.5 Seja $crpreço$ o crescimento mensal no nível geral de preços e $crsalhr$ o crescimento mensal nos salários por hora para a economia dos Estados Unidos. [Ambos são obtidos como diferenças de logaritmos: $crpreço = \Delta \log(\text{preço})$ e $crsalhr = \Delta \log(\text{salário})$.] Utilizando os dados mensais contidos no arquivo WAGEPRC.RAW, estimamos o seguinte modelo de defasagens distribuídas:

$$\begin{aligned}
crpreço &= -0,00093 + 0,119 crsalhr + 0,097 crsalhr_{-1} + 0,040 crsalhr_{-2} \\
&\quad (0,00057) \quad (0,052) \quad (0,039) \quad (0,039) \\
&+ 0,038 crsalhr_{-3} + 0,081 crsalhr_{-4} + 0,107 crsalhr_{-5} + 0,095 crsalhr_{-6} \\
&\quad (0,039) \quad (0,039) \quad (0,039) \quad (0,039) \\
&+ 0,104 crsalhr_{-7} + 0,103 crsalhr_{-8} + 0,159 crsalhr_{-9} + 0,110 crsalhr_{-10} \\
&\quad (0,039) \quad (0,039) \quad (0,039) \quad (0,039) \\
&+ 0,103 crsalhr_{-11} + 0,016 crsalhr_{-12} \\
&\quad (0,039) \quad (0,052) \\
n &= 273, R^2 = 0,317, \bar{R}^2 = 0,283.
\end{aligned}$$

- (i) Esboce a distribuição de defasagens estimada. Em qual defasagem está o maior efeito de $crsalhr$ sobre $crpreço$? Qual defasagem tem o menor coeficiente?
- (ii) Para quais defasagens as estatísticas t são menores que dois?
- (iii) Qual é a propensão de longo prazo estimada? Ela é muito diferente de um? Explique o que a PLP nos diz nesse exemplo.
- (iv) Que regressão você computaria para obter diretamente o erro-padrão da PLP?
- (v) Como você testaria a significância conjunta de mais seis defasagens de $crsalhr$? Quais seriam os graus de liberdade na distribuição F ? (Cuidado: você perdeu mais seis observações.)

11.6 Seja $hy6_t$ o rendimento da empresa em três meses (em percentagem) por ter comprado letras do Tesouro dos Estados Unidos de seis meses no tempo ($t - 1$) e vendido no tempo t (três meses depois) como letras de três meses. Seja $hy3_{t-1}$ o rendimento da empresa em três meses por ter comprado letras do Tesouro de três meses no tempo ($t - 1$). No tempo ($t - 1$), $hy3_{t-1}$ é conhecido, enquanto que $hy6_t$ é desconhecido, porque $p3_t$ (o preço das letras do Tesouro de três meses) é desconhecido no tempo ($t - 1$). A *hipótese das expectativas* (HE) diz que esses dois diferentes investimentos de três meses deveriam ser os mesmos, em média. Matematicamente, podemos escrever isso como uma expectativa condicional:

$$E(hy6_t | I_{t-1}) = hy3_{t-1},$$

onde I_{t-1} representa todas as informações observáveis até o tempo $t - 1$. Isso sugere que se estime o modelo

$$hy6_t = \beta_0 + \beta_1 hy3_{t-1} + u_t,$$

e que se teste $H_0: \beta_1 = 1$. (Também podemos testar $H_0: \beta_0 = 0$, mas em geral levamos em conta um *termo de prêmio* pela compra de ativos com diferentes datas de vencimento, de forma que $\beta_0 \neq 0$.)

- (i) Estimando a equação anterior por MQO, utilizando os dados contidos no arquivo INTQRT.RAW (espaçado de três em três meses), obtemos

$$\begin{aligned}
\hat{hy6}_t &= -0,058 + 1,104 hy3_{t-1} \\
&\quad (0,070)(0,039) \\
n &= 123, R^2 = 0,866.
\end{aligned}$$

Você rejeita $H_0: \beta_1 = 1$ contra $H_1: \beta_1 \neq 1$ ao nível de significância de 1%? A estimativa parece ser, na prática, diferente de um?

- (ii) Outra implicação da HE é que nenhuma outra variável datada como ($t - 1$), ou com data anterior a esta, ajuda a explicar $\hat{hy\delta}_t$, uma vez que $hy\beta_{t-1}$ tenha sido controlada. A inclusão de uma defasagem de *spread* entre as taxas das letras do Tesouro de seis e de três meses produz

$$\begin{aligned}\hat{hy\delta}_t &= -0,123 + 1,053 hy\beta_{t-1} + 0,480 (r\delta_{t-1} - r\beta_{t-1}) \\ &\quad (0,067) \quad (0,039) \quad (0,109) \\ n &= 123, R^2 = 0,885.\end{aligned}$$

Agora, o coeficiente de $hy\beta_{t-1}$ é estatisticamente diferente de um? O termo defasado de *spread* é significante? De acordo com essa equação, se, no tempo ($t - 1$), $r\delta$ estiver acima de $r\beta$, você deveria investir nas letras do Tesouro de seis meses ou de três meses?

- (iii) A correlação amostral entre $hy\beta_t$ e $hy\beta_{t-1}$ é 0,914. Por que isso deve causar algumas preocupações em relação à análise anterior?
(iv) Como você faria um teste de sazonalidade na equação estimada na parte (ii)?

11.7 Um modelo de ajustamento parcial é

$$\begin{aligned}y_t^* &= \gamma_0 + \gamma_1 x_t + e_t \\ y_t - y_{t-1} &= \lambda(y_t^* - y_{t-1}) + a_t\end{aligned}$$

onde y_t^* é o nível desejável ou ótimo de y , e y_t é o nível efetivo (observado). Por exemplo, y_t^* é o crescimento desejável nos estoques de uma firma, e x_t é o crescimento das vendas da firma. O parâmetro γ_1 mede o efeito de x_t sobre y_t^* . A segunda equação descreve como y observado se ajusta, dependendo da relação entre y desejado no tempo t e y observado no tempo ($t - 1$). O parâmetro λ mede a velocidade de ajustamento e satisfaz $0 < \lambda < 1$.

- (i) Insira a primeira equação para y_t^* na segunda equação e mostre que podemos escrever

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t.$$

Em particular, encontre β_j em termos de γ_j e λ e encontre u_t em termos de e_t e a_t . Portanto, o modelo de ajustamento parcial leva a um modelo com uma variável dependente defasada e um x contemporâneo.

- (ii) Se $E(e_t | x_t, y_{t-1}, x_{t-1}, \dots) = E(a_t | x_t, y_{t-1}, x_{t-1}, \dots) = 0$ e todas as séries são fracamente dependentes, como você estimaria β_j ?
(iii) Se $\hat{\beta}_1 = 0,7$ e $\hat{\beta}_2 = 0,2$, quais são as estimativas de γ_j e λ ?

Correlação Serial e Heteroscedasticidade em Regressões de Séries Temporais

este capítulo, discutimos o problema crítico da correlação serial no termo erro de um modelo de regressão múltipla. Vimos no Capítulo 11 que, quando, em um sentido apropriado, a dinâmica de um modelo tiver sido completamente especificada, os erros não serão serialmente correlacionados. Assim, o teste para se verificar correlação serial pode ser usado para detectar a má especificação dinâmica. Além disso, modelos estáticos e de defasagens distribuídas finitas muitas vezes possuem erros serialmente correlacionados, mesmo se não houver má especificação básica do modelo. Portanto, é importante conhecer as consequências e soluções da correlação serial dessas úteis classes de modelos.

Na Seção 12.1, apresentaremos as propriedades do MQO quando os erros contêm correlação serial. Na Seção 12.2, demonstraremos como fazer o teste para verificar a correlação serial. Discutimos testes que se aplicam a modelos com regressores estritamente exógenos e testes que são assintoticamente válidos com regressores gerais, inclusive com variáveis dependentes defasadas. A Seção 12.3 explicará como corrigir a correlação serial sob a hipótese de variáveis explicativas estritamente exógenas, enquanto a Seção 12.4 mostrará como o uso de dados diferenciados muitas vezes elimina a correlação serial nos erros. A Seção 12.5 traz os avanços mais recentes sobre como ajustar os habituais erros-padrão do MQO e testar estatísticas na presença de correlação serial bastante genérica.

No Capítulo 8, discutimos como testar e corrigir a heteroscedasticidade em aplicações de corte transversal. Na Seção 12.6, mostraremos como métodos utilizados nos casos de corte transversal podem ser ampliados para os casos de séries de tempo. As mecânicas são essencialmente as mesmas, mas existem algumas sutilezas associadas com a correlação temporal nas observações de séries de tempo que devem ser consideradas. Adicionalmente, trataremos de forma breve as consequências de formas dinâmicas de heteroscedasticidade.

12.1 AS PROPRIEDADES DO MQO COM ERROS SERIALMENTE CORRELACIONADOS

Inexistência de Viés e Consistência

No Capítulo 10, provamos a inexistência de viés do estimador do MQO sob as três primeiras hipóteses de Gauss-Markov para regressões de séries temporais (ST.1 a ST.3). Particularmente, o teorema 10.1 nada assumia sobre a correlação serial nos erros. Por conseguinte, desde que as variáveis explicativas sejam estritamente exógenas, os $\hat{\beta}_j$ serão não-viesados, a despeito do grau de correlação nos erros. Isso é análogo à observação de que a heteroscedasticidade nos erros não causa viés em $\hat{\beta}_j$.

No Capítulo 11, relaxamos a hipótese de exogeneidade estrita para $E(u_i|x_i) = 0$ e mostramos que, quando os dados são fracamente dependentes, os $\hat{\beta}_j$ ainda são consistentes (embora não necessariamente não-viesados). Esse resultado não dependia de qualquer hipótese sobre correlação serial nos erros.

Eficiência e Inferência

Como o Teorema de Gauss-Markov (teorema 10.4) exige tanto a homoscedasticidade como os erros sem correlação serial, o MQO não é mais BLUE na presença de correlação serial. Mais importante ainda, os habituais erros-padrão do MQO e testes estatísticos não são válidos, mesmo assintoticamente. Podemos verificar isso calculando a variância do estimador do MQO sob as quatro primeiras hipóteses de Gauss-Markov e o modelo AR(1) para o termo erro. Mais precisamente, assumimos que

$$u_t = \rho u_{t-1} + e_t, t = 1, 2, \dots, n \quad (12.1)$$

$$|\rho| < 1, \quad (12.2)$$

onde os e_t são variáveis aleatórias não-correlacionadas com média zero e variância σ_e^2 ; lembre-se de que, como vimos no Capítulo 11, a hipótese (12.2) é a condição de estabilidade.

Consideramos a variância do estimador de inclinação do MQO no modelo simples de regressão

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

e, apenas para simplificar a fórmula, assumimos que a média amostral de x_t é zero ($\bar{x} = 0$). Então, o estimador de MQO $\hat{\beta}_1$ de β_1 pode ser escrito como

$$\hat{\beta}_1 = \beta_1 + \text{SQT}_x^{-1} \sum_{t=1}^n x_t u_t, \quad (12.3)$$

onde $\text{SQT}_x = \sqrt{\sum_{t=1}^n x_t^2}$. Agora, ao calcular a variância de $\hat{\beta}_1$ (condicional em X), devemos levar em conta a correlação serial em u_t :

$$\begin{aligned} \text{Var}(\hat{\beta}_1) &= \text{SQT}_x^{-2} \text{Var}\left(\sum_{t=1}^n x_t u_t\right) = \text{SQT}_x^{-2} \left(\sum_{t=1}^n x_t^2 \text{Var}(u_t) \right. \\ &\quad \left. + 2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} x_t x_{t+j} E(u_t u_{t+j}) \right) \\ &= \sigma^2 / \text{SQT}_x + 2(\sigma^2 / \text{SQT}_x^2) \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j} \end{aligned} \quad (12.4)$$

onde $\sigma^2 = \text{Var}(u_t)$ e utilizamos o fato de que $E(u_t u_{t+j}) = \text{Cov}(u_t, u_{t+j}) = \rho^j \sigma^2$ [veja equação (11.4)]. O primeiro termo na equação (12.4), σ^2/SQT_x , é a variância de $\hat{\beta}_1$ quando $\rho = 0$, que é a variância familiar do MQO sob as hipóteses de Gauss-Markov. Se ignorarmos a correlação serial e estimarmos a variância da maneira habitual, o estimador da variância será usualmente viesado quando $\rho \neq 0$, pois ele ignora o segundo termo em (12.4). Como veremos mais tarde, em outros exemplos, $\rho > 0$ é mais comum, em cujo caso $\rho^j > 0$ para todos os j . Além disso, as variáveis independentes em modelos de regressão freqüentemente são positivamente correlacionadas ao longo do tempo, de forma que $x_t x_{t+j}$ é positivo para a maioria dos pares t e $t + j$. Portanto, na maioria das aplicações econômicas, o termo $\sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$ é positivo e, assim, a fórmula usual da variância do MQO, σ^2/SQT_x , *subestima* a verdadeira variância do estimador MQO. Se ρ for grande ou x_t tiver um alto grau de correlação serial positiva — um caso comum —, o viés no estimador de variância usual do MQO pode ser substancial. Tendemos a pensar que o estimador de inclinação do MQO seja mais preciso do que na realidade é.

Quando $\rho < 0$, ρ^j é negativo quando j é ímpar, e positivo quando j é par e, portanto, é difícil determinar o sinal de $\sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$. De fato, é possível que a fórmula usual da variância do MQO efetivamente *exagere* a verdadeira variância de $\hat{\beta}_1$. Em qualquer caso, o estimador usual da variância será viesado para $\text{Var}(\hat{\beta}_1)$ na presença de correlação serial.

Como o erro-padrão de $\hat{\beta}_1$ é uma estimativa do desvio-padrão de $\hat{\beta}_1$, o uso do erro-padrão habitual do MQO na presença de correlação serial não é válido. Portanto, as estatísticas t não mais são válidas para testar hipóteses únicas. Como um erro-padrão menor significa uma estatística t maior, as estatísticas t normais muitas vezes serão grandes demais quando $\rho > 0$. As usuais estatísticas F e LM para testar hipóteses múltiplas também não são válidas.

Suponha que, em vez do modelo AR(1), u_t siga o modelo MA(1) $u_t = e_t + \alpha e_{t-1}$. Encontre $\text{Var}(\hat{\beta}_1)$ e demonstre que ela é diferente da fórmula habitual se $\alpha \neq 0$.

O Grau de Ajuste

Algumas vezes, afirma-se que a correlação serial nos erros de um modelo de regressão de séries temporais invalida as medidas habituais do grau de ajuste, R -quadrado e R -quadrado ajustado. Felizmente, esse não é o caso, desde que os dados sejam estacionários e fracamente dependentes. Para verificar porque essas medidas ainda são válidas, lembre-se de que definimos o R -quadrado da população em um contexto de corte transversal como $1 - \sigma_u^2/\sigma_y^2$ (veja a Seção 6.3). Essa definição ainda é apropriada no contexto de regressões de séries temporais com dados estacionários e fracamente dependentes: a variância de ambos, erros e variável dependente, não se altera ao longo do tempo. De acordo com a lei dos grandes números, R^2 e \bar{R}^2 estimam o R -quadrado da população de maneira coerente. O argumento é essencialmente o mesmo do caso do corte transversal, haja ou não heteroscedasticidade (veja Seção 8.1). Como nunca existe um estimador não-viesado do R -quadrado da população, não faz sentido falar em viés no R^2 causado por correlação serial. Tudo o que realmente podemos dizer é que nossas medidas do grau de ajuste ainda são estimadores consistentes do parâmetro da população. Esse argumento não se mantém se $\{y_t\}$ for um processo I(1), pois $\text{Var}(y_t)$ cresce com t ; o grau de ajuste não faz muito sentido nesse caso. Como discutimos na Seção 10.5, tendências ao redor da média de y_t , ou

sazonalidade, podem e devem ser levadas em conta no cálculo do R -quadrado. Outros desvios da estacionariedade não causam dificuldade na interpretação de R^2 e \bar{R}^2 das maneiras habituais.

A Correlação Serial na Presença da Variável Dependente Defasada

Iniciantes em econometria são freqüentemente alertados sobre os perigos de erros serialmente correlacionados na presença da variável dependente defasada. Quase todos os manuais de econometria contêm alguma forma da afirmação “o MQO é inconsistente na presença da variável dependente defasada e de erros serialmente correlacionados”. Infelizmente, como declaração generalizada essa afirmação é falsa. Existe uma versão da afirmação que é correta, mas é importante que seja bem precisa.

Para ilustrar, suponha que o valor esperado de y_t , dado y_{t-1} , seja linear:

$$E(y_t | y_{t-1}) = \beta_0 + \beta_1 y_{t-1}, \quad (12.5)$$

onde assumimos estabilidade, $|\beta_1| < 1$. Sabemos que sempre podemos escrever essa equação com um termo erro como

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t, \quad (12.6)$$

$$E(u_t | y_{t-1}) = 0. \quad (12.7)$$

Por construção, este modelo satisfaz a hipótese crucial ST.3' da consistência do MQO; portanto, os estimadores de MQO, $\hat{\beta}_0$ e $\hat{\beta}_1$, são consistentes. É importante verificar que, sem hipóteses adicionais, os erros $\{u_t\}$ podem ser serialmente correlacionados. A condição (12.7) garante que u_t é não-correlacionado com y_{t-1} , mas u_t e y_{t-2} podem ser correlacionados. Portanto, como $u_{t-1} = y_{t-1} - \beta_0 - \beta_1 y_{t-2}$, a covariância entre u_t e u_{t-1} é $-\beta_1 \text{Cov}(u_t, y_{t-2})$, que não é necessariamente zero. Assim, os erros exibem correlação serial e o modelo contém uma variável dependente defasada, mas o MQO estima consistentemente β_0 e β_1 porque esses são os parâmetros na expectativa condicional (12.5). A correlação serial nos erros fará com que as estatísticas usuais de MQO não sejam válidas para a finalidade de testes, mas não afetará a consistência.

Então, quando o MQO será inconsistente se os erros forem serialmente correlacionados e a regressão contiver uma variável dependente defasada? Isso acontece quando escrevemos o modelo em forma de erro, exatamente como em (12.6), mas assumimos que $\{u_t\}$ segue um modelo AR(1) estável, como em (12.1) e (12.2), onde

$$E(e_t | u_{t-1}, u_{t-2}, \dots) = E(e_t | y_{t-1}, y_{t-2}, \dots) = 0. \quad (12.8)$$

Como e_t é não-correlacionado com y_{t-1} por hipótese, $\text{Cov}(y_{t-1}, e_t) = \rho \text{Cov}(y_{t-1}, u_{t-1})$, que não é zero, a menos que $\rho = 0$. Isso faz com que os estimadores MQO de β_0 e β_1 da regressão de y_t sobre y_{t-1} sejam inconsistentes.

Agora vemos que a estimação MQO de (12.6), quando os erros u_t também seguem um modelo AR(1), leva a estimadores inconsistentes. Porém, a exatidão dessa afirmação não a torna menos equivocada. Temos que perguntar: qual seria a vantagem de estimar os parâmetros em (12.6) quando

os erros seguem um modelo AR(1)? É difícil pensar em casos nos quais isso seria interessante. Pelo menos em (12.5) os parâmetros nos informam o valor esperado de y_t dado y_{t-1} . Quando combinamos (12.6) com (12.1), vemos que y_t de fato segue um modelo auto-regressivo de segunda ordem, ou modelo AR(2). Para verificar isso, escreva $u_{t-1} = y_{t-1} - \beta_0 - \beta_1 y_{t-2}$, inserindo essa expressão em $u_t = \rho u_{t-1} + e_t$. Então, (12.6) pode ser rescrita como

$$\begin{aligned}y_t &= \beta_0 + \beta_1 y_{t-1} + \rho(y_{t-1} - \beta_0 - \beta_1 y_{t-2}) + e_t \\&= \beta_0(1 - \rho) + (\beta_1 + \rho)y_{t-1} - \rho\beta_1 y_{t-2} + e_t \\&= \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + e_t,\end{aligned}$$

onde $\alpha_0 = \beta_0(1 - \rho)$, $\alpha_1 = \beta_1 + \rho$ e $\alpha_2 = -\rho\beta_1$. Dada (12.8), segue que

$$E(y_t | y_{t-1}, y_{t-2}, \dots) = E(y_t | y_{t-1}, y_{t-2}) = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2}. \quad (12.9)$$

Isso significa que o valor esperado de y_t , dados todos os y passados depende de *duas* defasagens de y . É a equação (12.9) que estariamos interessados em usar para qualquer propósito prático, inclusive para fazer previsões, como veremos no Capítulo 18. Estamos especialmente interessados nos parâmetros α_j . Sob as condições apropriadas de estabilidade de um modelo AR(2) — que veremos na Seção 12.3 —, a estimação MQO de (12.9) produz estimadores consistentes e assimptoticamente normais de α_j .

O ponto principal é que você precisa ter uma boa razão para ter tanto uma variável dependente defasada em um modelo como um modelo particular de correlação serial nos erros. Muitas vezes, a correlação serial nos erros de um modelo dinâmico simplesmente indica que a função de regressão dinâmica não foi completamente especificada: no exemplo anterior deveríamos ter adicionado y_{t-2} à equação.

No Capítulo 18, veremos exemplos de modelos com a variável dependente defasada nas quais os erros são serialmente correlacionados e também são correlacionados com y_{t-1} . Entretanto, mesmo nesses casos, os erros não seguem um processo auto-regressivo.

12.2 O TESTE DA CORRELAÇÃO SERIAL

Nesta seção, discutimos vários métodos de testar a correlação serial nos erros, no modelo de regressão linear múltipla

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t.$$

Primeiro, consideraremos o caso em que os regressores são estritamente exógenos. Lembre-se de que isso requer que o erro, u_t , seja não-correlacionado com os regressores em todos os períodos de tempo (veja a Seção 10.3), e, assim, entre outras coisas, ele exclui modelos com a variável dependente defasada.

O Teste t de Correlação Serial AR(1) com Regressores Estritamente Exógenos

Embora haja numerosas maneiras de os erros em um modelo de regressão múltipla poderem ser serialmente correlacionados, o modelo mais popular — e o mais simples de trabalhar — é o modelo AR(1) nas equações (12.1) e (12.2). Na seção anterior, explicamos as implicações de computar o MQO quando os erros são, de forma geral, serialmente correlacionados, e derivamos a variância do estimador de inclinação MQO em um modelo simples de regressão com erros AR(1). Agora mostramos como fazer o teste para verificar a presença de correlação serial AR(1). A hipótese nula é a de que *não* há correlação serial. Portanto, assim como nos testes da heteroscedasticidade, presumimos o melhor e exigimos que os dados forneçam evidência razoavelmente forte de que a hipótese ideal de não existir correlação serial seja violada.

Primeiro, derivamos um teste de amostra grande, sob a hipótese de que as variáveis explicativas são estritamente exógenas: o valor esperado de u_t , dado o histórico completo das variáveis independentes, é zero. Além disso, em (12.1) devemos assumir que

$$\mathbb{E}(e_t | u_{t-1}, u_{t-2}, \dots) = 0 \quad (12.10)$$

e

$$\text{Var}(e_t | u_{t-1}) = \text{Var}(e_t) = \sigma_e^2. \quad (12.11)$$

Essas são hipóteses padrão no modelo AR(1) (que decorrem quando $\{e_t\}$ é uma seqüência i.i.d.) e nos permitem aplicar os resultados de amostras grandes do Capítulo 11 para a regressão dinâmica.

Assim como no teste de heteroscedasticidade, a hipótese nula é a de que a hipótese apropriada de Gauss-Markov é verdadeira. No modelo AR(1), a hipótese nula de que os erros são serialmente correlacionados é

$$H_0: \rho = 0. \quad (12.12)$$

Como podemos testar essa hipótese? Se os u_t fossem observados, sob (12.10) e (12.11), poderíamos imediatamente aplicar os resultados de normalidade assintótica do teorema 11.2 ao modelo de regressão dinâmica.

$$u_t = \rho u_{t-1} + e_t, t = 2, \dots, n. \quad (12.13)$$

(Sob a hipótese nula $\rho = 0$, $\{u_t\}$ é, sem dúvida, fracamente dependente.) Em outras palavras, poderíamos estimar ρ da regressão de u_t sobre u_{t-1} , para todos os $t = 2, \dots, n$, sem um intercepto, e usar a estatística t usual para $\hat{\rho}$. Isso não funciona porque os erros u_t não são observados. Mesmo assim, como no caso do teste de heteroscedasticidade, podemos substituir u_t pelo correspondente resíduo de MQO, \hat{u}_t . Como \hat{u}_t depende dos estimadores MQO, $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$, não é óbvio que o uso de \hat{u}_t em lugar de u_t na regressão não tenha efeito na distribuição da estatística t . Felizmente, constata-se que, devido à hipótese de exogeneidade estrita, a distribuição de amostra grande da estatística t não é afetada pelo uso

dos resíduos de MQO em lugar dos erros. Uma prova disso está muito além do escopo deste texto, mas deriva do trabalho de Wooldridge (1991b).

Podemos resumir o teste assintótico de correlação serial AR(1) de maneira muito simples:

O TESTE DA CORRELAÇÃO SERIAL AR(1) COM REGRESSORES ESTRITAMENTE EXÓGENOS:

(i) Compute a regressão MQO de y_t sobre x_{t1}, \dots, x_{tk} e obtenha os resíduos do MQO, \hat{u}_t , para todo $t = 1, 2, \dots, n$.

(ii) Compute a regressão de

$$\hat{u}_t \text{ sobre } \hat{u}_{t-1}, \text{ para todo } t = 2, \dots, n, \quad (12.14)$$

obtendo o coeficiente $\hat{\rho}$ de \hat{u}_{t-1} e sua estatística t , $t_{\hat{\rho}}$. (Essa regressão pode ou não conter um intercepto; a estatística t de $\hat{\rho}$ será levemente afetada, mas ela é assintoticamente válida de qualquer maneira.)

(iii) Use $t_{\hat{\rho}}$ para testar $H_0: \rho = 0$ contra $H_1: \rho \neq 0$, da maneira habitual. Na realidade, como $\rho > 0$ é freqüentemente esperado *a priori*, a alternativa pode ser $H_1: \rho > 0$. Geralmente, concluímos que a correlação serial é um problema a ser tratado somente se H_0 for rejeitada ao nível de 5%. Como sempre, é melhor descrever o p -valor do teste.

Ao decidir se a correlação serial precisa ser solucionada, devemos nos lembrar da diferença entre significância prática e significância estatística. Com uma amostra de tamanho grande, é possível encontrar correlação serial mesmo que $\hat{\rho}$ seja praticamente pequeno; quando $\hat{\rho}$ é próximo de zero, os procedimentos comuns de inferência de MQO não estarão distantes [veja equação (12.4)]. Tais resultados são relativamente raros em aplicações de séries temporais, porque os conjuntos de dados de séries temporais são, normalmente, pequenos.

EXEMPLO 12.1

[O Teste da Correlação Serial AR(1) na Curva de Phillips]

No Capítulo 10, estimamos uma curva de Phillips estática que explicava a relação de substituição inflação-desemprego nos Estados Unidos (veja Exemplo 10.1). No Capítulo 11, estudamos uma determinada curva de Phillips de expectativas aumentadas, na qual assumimos expectativas adaptativas (veja Exemplo 11.5). Agora, testamos o termo erro em cada equação para verificar a existência de correlação serial. Como a curva de expectativas aumentadas utiliza $\Delta\text{inf}_t = \text{inf}_t - \text{inf}_{t-1}$ como variável dependente, temos algumas observações a fazer.

Para a curva de Phillips estática, a regressão em (12.14) produz $\hat{\rho} = 0,573$, $t = 4,93$, e p -valor = 0,000 (com 48 observações). Isso é uma evidência muito forte de correlação serial positiva, de primeira ordem. Uma consequência disso é que os erros-padrão e as estatísticas t do Capítulo 10 não são válidos. Em contraposição, o teste para verificar correlação serial AR(1) na curva de expectativas aumentadas produz $\hat{\rho} = -0,036$, $t = -0,297$ e p -valor = 0,775 (com 47 observações): não existe evidência de correlação serial AR(1) na curva de Phillips de expectativas aumentadas.

Embora o teste de (12.14) seja derivado do modelo AR(1), o teste pode detectar outros tipos de correlação serial. Lembre-se, $\hat{\rho}$ é um estimador consistente da correlação entre u_t e u_{t-1} . Qualquer correlação serial que faça com que erros adjacentes sejam correlacionados pode ser captada por esse teste.

De outro lado, ele não detecta a correlação serial em que os erros adjacentes sejam não-correlacionados, $\text{Corr}(u_t, u_{t-1}) = 0$. (Por exemplo, u_t e u_{t-2} poderiam ser correlacionados.)

Ao usar a estatística t usual de (12.14), temos que assumir que os erros em (12.13) satisfazem a hipótese de heteroscedasticidade apropriada, (12.14). De fato, é fácil fazer o teste robusto em relação à heteroscedasticidade de e_t ; simplesmente usamos o teste habitual da estatística t robusta em relação à heteroscedasticidade do Capítulo 8. Para a curva de Phillips estática no Exemplo 12.1, a estatística t robusta em relação à heteroscedasticidade é 4,03, que é menor que a estatística t não-robusta, mas ainda muito significante. Na Seção 12.6, discorreremos mais sobre a heteroscedasticidade em regressões de séries temporais, inclusive suas formas dinâmicas.

Como você usaria a regressão (12.14) para construir um intervalo de confiança aproximado de 95% para ρ ?

O Teste de Durbin-Watson sob as Hipóteses Clássicas

Outro teste para verificar a correlação serial AR(1) é o teste de Durbin-Watson. A **estatística de Durbin-Watson (DW)** também é baseada nos resíduos do MQO:

$$DW = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}. \quad (12.15)$$

Um pouco de álgebra mostra que DW e $\hat{\rho}$ de (12.14) estão estreitamente ligados:

$$DW \approx 2(1 - \hat{\rho}). \quad (12.16)$$

Uma razão para essa relação não ser exata é que $\hat{\rho}$ tem $\sum_{t=2}^n \hat{u}_{t-1}^2$ em seu denominador, enquanto a estatística DW tem, em seu denominador, a soma dos quadrados de todos os resíduos do MQO. Mesmo com amostras de tamanhos moderados, a aproximação em (12.16) é freqüentemente muito próxima. Portanto, testes com base na estatística DW e o teste t com base em $\hat{\rho}$ são conceitualmente os mesmos.

Durbin e Watson (1950) derivam a distribuição de DW (condicional em X), algo que requer o conjunto total das hipóteses do modelo linear clássico, inclusive a normalidade dos erros. Infelizmente, essa distribuição depende dos valores das variáveis independentes. (Elas também depende do tamanho da amostra, do número de regressores e de se a regressão contém um intercepto.) Embora alguns programas econometríticos computem valores críticos e p -valores para DW , muitos não o fazem. De qualquer forma, eles dependem do conjunto total das hipóteses do MLC.

Vários textos econometríticos descrevem limites superiores e limites inferiores dos valores críticos que dependem do nível de significância desejável, da hipótese alternativa, do número de observações

e do número de regressores. (Assumimos que um intercepto está incluído no modelo.) Em geral, o teste DW é calculado para a hipótese alternativa

$$H_1: \rho > 0. \quad (12.17)$$

A partir da aproximação em (12.16), $\hat{\rho} \approx 0$ implica que $DW \approx 2$, e $\hat{\rho} > 0$ implica que $DW < 2$. Assim, para rejeitar a hipótese nula (12.12) em favor de (12.17), estamos procurando um valor de DW que seja显著mente menor que dois. Infelizmente, devido aos problemas para a obtenção da distribuição nula de DW, devemos comparar DW com dois conjuntos de valores críticos. Habitualmente, eles são denominados como d_U (para limite superior) e d_L (para limite inferior). Se $DW < d_L$, então, rejeitamos H_0 em favor de (12.17); se $DW > d_U$, não é possível rejeitar H_0 . Se $d_L \leq DW \leq d_U$, o teste não é conclusivo.

Como um exemplo, se selecionarmos um nível de significância de 5% com $n = 45$ e $k = 4$, $d_U = 1,720$ e $d_L = 1,336$ [veja Savin e White (1997)]. Se $DW < 1,336$, rejeitamos a hipótese nula de inexistência de correlação serial ao nível de 5%; se $DW > 1,72$, não é possível rejeitar H_0 ; se $1,336 \leq DW \leq 1,72$, o teste não é conclusivo.

No Exemplo 12.1, da curva de Phillips estática, $DW = 0,80$. Podemos obter o valor crítico mais baixo de 1% de Savin e White (1997) para $k = 1$ e $n = 50$: $d_L = 1,32$. Portanto, rejeitamos a hipótese nula de inexistência de correlação serial, contra a alternativa de correlação serial positiva ao nível de 1%. (Utilizando o teste t anterior, podemos concluir que o p -valor é igual a zero até três casas decimais.) Para a curva de Phillips de expectativas aumentadas, $DW = 1,77$, que está dentro da região de impossibilidade de rejeição até mesmo ao nível de 5% ($d_U = 1,59$).

O fato de que uma distribuição amostral exata de DW pode ser tabelada é a única vantagem que DW tem sobre o teste t de (12.14). Dado que os valores críticos tabulados são válidos somente sob o conjunto total das hipóteses do MLC e que eles podem levar a uma ampla região inconclusiva, as desvantagens práticas da estatística DW são substanciais. A estatística t de (12.14) é fácil de ser calculada e é assintoticamente válida sem erros normalmente distribuídos. A estatística t também é válida na presença de heteroscedasticidade que dependa de x_{ij} ; é fácil torná-la robusta para qualquer forma de heteroscedasticidade.

O Teste da Correlação Serial AR(1) sem Regressores Estritamente Exógenos

Quando as variáveis explicativas não são estritamente exógenas, de forma que uma ou mais x_{ij} são correlacionadas com u_{t-1} , nem o teste t da regressão (12.14) nem a estatística de Durbin-Watson são válidos, mesmo em amostras grandes. O principal caso de regressores não-estritamente exógenos ocorre quando o modelo contém uma variável dependente defasada: y_{t-1} e u_{t-1} são obviamente correlacionados. Durbin (1970) sugeriu duas alternativas para a estatística DW quando o modelo contém uma variável dependente defasada e os outros regressores são não-aleatórios (ou, de forma mais geral, estritamente exógenos). A primeira é chamada *estatística h de Durbin*. Essa estatística tem uma desvantagem prática no fato de que ela nem sempre pode ser calculada, e assim não vamos discuti-la aqui.

A estatística alternativa de Durbin é fácil de ser calculada e é válida quando existe qualquer número de variáveis explicativas não-estritamente exógenas. O teste também funciona se as variáveis explicativas forem estritamente exógenas.

O TESTE DA CORRELAÇÃO SERIAL COM REGRESSORES GERAIS:

- (i) Compute a regressão MQO de y_t sobre x_{t1}, \dots, x_{tk} e obtenha os resíduos MQO, \hat{u}_t , para todo $t = 1, 2, \dots, n$.
(ii) Compute a regressão de

$$\hat{u}_t \text{ sobre } x_{t1}, x_{t2}, \dots, x_{tk}, \hat{u}_{t-1}, \text{ para todo } t = 2, \dots, n \quad (12.18)$$

para obter o coeficiente $\hat{\rho}$ de \hat{u}_{t-1} e sua estatística t , $t_{\hat{\rho}}$.

- (iii) Utilize $t_{\hat{\rho}}$ para testar $H_0: \rho = 0$ contra $H_1: \rho \neq 0$, da maneira habitual (ou use uma alternativa unilateral).

Na equação (12.18), regredimos os resíduos MQO sobre *todas* as variáveis independentes, inclusive um intercepto, e sobre os resíduos defasados. A estatística t dos resíduos defasados é um teste válido de (12.12) no modelo AR(1) (12.13) [quando adicionamos $\text{Var}(u_t | x_t, u_{t-1}) = \sigma^2$ sob H_0]. Qualquer número de variáveis dependentes defasadas pode aparecer entre x_{ij} , e outras variáveis explicativas não-estritamente exógenas também são permitidas.

A inclusão de x_{t1}, \dots, x_{tk} permite explicitamente que cada x_{ij} seja correlacionado com u_{t-1} , e isso assegura que $t_{\hat{\rho}}$ tenha uma distribuição t aproximada em amostras grandes. A estatística t de (12.14) ignora possível correlação entre x_{ij} e u_{t-1} , de modo que ela não é válida sem regressores estritamente exógenos. Em decorrência, como $\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{t1} - \dots - \hat{\beta}_k x_{tk}$, é possível mostrar que a estatística t de \hat{u}_{t-1} é a mesma se y_t for usado em lugar de \hat{u}_t como variável dependente em (12.18).

A estatística t de (12.18) pode facilmente ser transformada em robusta em relação à heteroscedasticidade de forma desconhecida [em particular quando $\text{Var}(u_t | x_t, u_{t-1})$ não for constante]: basta usar a estatística t robusta em relação à heteroscedasticidade de \hat{u}_{t-1} .

EXEMPLO 12.2**[O Teste da Correlação Serial AR(1) na Equação de Salário Mínimo]**

No Capítulo 10 (veja Exemplo 10.9), estimamos o efeito do salário mínimo sobre a taxa de emprego porto-riquenho. Agora verificamos se os erros parecem conter correlação serial, utilizando o teste que não assume exogeneidade estrita nas variáveis do salário mínimo ou do PIB. [Adicionamos o log do PIB real porto-riquenho à equação (10.38), como no Problema 10.9.] Estamos assumindo que os processos estocásticos subjacentes sejam fracamente dependentes, mas permitimos que eles contenham uma tendência temporal linear (ao incluirmos t na regressão).

Fazendo \hat{u}_t representar os resíduos MQO, computamos a regressão de

$$\hat{u}_t \text{ sobre } \log(\text{mincov}_t), \log(\text{prpnbb}_t), \log(\text{uspnbb}_t), t \text{ e } \hat{u}_{t-1},$$

utilizando as 37 observações disponíveis. O coeficiente estimado de \hat{u}_{t-1} é $\hat{\rho} = 0,481$ com $t = 2,89$ (p -valor bilateral = 0,007). Portanto, existe forte evidência de correlação serial AR(1) nos erros, o que significa que as estatísticas t de $\hat{\beta}_j$ que obtivemos anteriormente não são válidas para inferência. Lembre-se, no entanto, de que os $\hat{\beta}_j$ ainda serão consistentes se u_t for contemporaneamente não-correlacionado com cada variável explicativa. Se usarmos a regressão (12.14), obteremos $\hat{\rho} = 0,417$ e $t = 2,63$, de modo que o resultado do teste é semelhante nesse caso.

O Teste da Correlação Serial de Ordem Mais Elevada

O teste a partir de (12.18) é facilmente estendido para ordens mais elevadas de correlação serial. Por exemplo, suponha que queiramos testar

$$H_0: \rho_1 = 0, \rho_2 = 0 \quad (12.19)$$

no modelo AR(2)

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t.$$

O modelo alternativo de correlação serial permite testar *correlação serial de segunda ordem*. Como sempre, estimamos o modelo por MQO e obtemos os resíduos do MQO, \hat{u}_t . Então, podemos computar a regressão de

$$\hat{u}_t \text{ sobre } x_{t1}, x_{t2}, \dots, x_{tk}, \hat{u}_{t-1} \text{ e } \hat{u}_{t-2}, \text{ para todo } t = 3, \dots, n,$$

para obter o teste F da significância conjunta de \hat{u}_{t-1} e \hat{u}_{t-2} . Se essas duas defasagens forem conjuntamente significantes em um nível suficientemente pequeno, digamos 5%, então, rejeitamos (12.19) e concluímos que os erros são serialmente correlacionados.

De maneira mais geral, podemos testar a correlação serial no modelo auto-regressivo de ordem q :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_q u_{t-q} + e_t \quad (12.20)$$

A hipótese nula é

$$H_0: \rho_1 = 0, \rho_2 = 0, \dots, \rho_q = 0. \quad (12.21)$$

O TESTE DA CORRELAÇÃO SERIAL AR(q):

(i) Compute a regressão MQO de y_t sobre x_{t1}, \dots, x_{tk} e obtenha os resíduos do MQO, \hat{u}_t , para todo $t = 1, 2, \dots, n$.

(ii) Compute a regressão de

$$\hat{u}_t \text{ sobre } x_{t1}, x_{t2}, \dots, x_{tk}, \hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-q} \text{ para todo } t = (q+1), \dots, n. \quad (12.22)$$

(iii) Calcule o teste F da significância conjunta de $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-q}$ em (12.12). [A estatística F com y_t como variável dependente em (12.22) também pode ser usada, já que ela fornece uma resposta idêntica.]

Se assumirmos que os x_{ij} são estritamente exógenos, de forma que cada x_{ij} seja não-correlacionado com $u_{t-1}, u_{t-2}, \dots, u_{t-q}$, então, os x_{ij} podem ser omitidos de (12.22). A inclusão de x_{ij} na regressão faz com que o teste seja válido com ou sem a hipótese de exogeneidade estrita. O teste exige a hipótese de homoscedasticidade

$$\text{Var}(u_t | x_t, u_{t-1}, \dots, u_{t-q}) = \sigma^2. \quad (12.23)$$

Uma versão robusta em relação à heteroscedasticidade pode ser calculada, como descrito no Capítulo 8.

Uma alternativa para calcular o teste F é usar a forma da estatística do multiplicador de Lagrange (LM). (Tratamos da estatística LM no teste de restrições de exclusões no Capítulo 5, para a análise de corte transversal.) A estatística LM para testar (12.21) é simplesmente

$$LM = (n - q)R_{\hat{u}}^2, \quad (12.24)$$

onde $R_{\hat{u}}^2$ é o habitual R -quadrado da regressão (12.22). Sob a hipótese nula, $LM \xrightarrow{d} \chi_q^2$. Isso é habitualmente chamado de **teste de Breusch-Godfrey** da correlação serial de AR(q). A estatística LM também requer (12.23), mas ela pode ser tornada robusta em relação à heteroscedasticidade. [Para detalhes, veja Wooldridge (1991b).]

EXEMPLO 12.3

[O Teste da Correlação Serial AR(3)]

No estudo de caso da indústria de cloreto de bário (veja Exemplo 10.5), utilizamos dados mensais, de modo que podemos querer testar a existência de correlação serial de ordens mais altas. Como ilustração, faremos o teste da correlação serial AR(3) dos erros subjacentes da equação (10.22). Utilizando a regressão (12.22), a estatística F para significância conjunta de \hat{u}_{t-1} , \hat{u}_{t-2} e \hat{u}_{t-3} é $F = 5,12$. Originalmente, tínhamos $n = 131$ e perdemos três observações na regressão auxiliar (12.22). Como estimamos dez parâmetros em (12.22) para este exemplo, os g_f na estatística F são três e 118. O p -valor da estatística F é 0,0023, e, portanto, existe forte evidência de existência de correlação serial AR(3).

Usando dados trimestrais ou mensais que não tenham sido ajustados quanto à sazonalidade, algumas vezes desejamos verificar se existem formas sazonais de correlação serial. Por exemplo, com dados trimestrais, podemos postular o modelo auto-regressivo

$$u_t = \rho_4 u_{t-4} + e_t. \quad (12.25)$$

A partir dos testes da correlação serial AR(1), fica bastante claro como prosseguir. Quando os regressores são estritamente exógenos, podemos usar o teste t de \hat{u}_{t-4} na regressão de

\hat{u}_t sobre \hat{u}_{t-4} , para todo $t = 5, \dots, n$.

Também existe uma modificação da estatística de Durbin-Watson [veja Wallis (1972)]. Quando os x_{ij} não forem estritamente exógenos, podemos usar a regressão em (12.18), com \hat{u}_{t-4} substituindo \hat{u}_{t-1} .

No exemplo 12.3, os dados são mensais e não são sazonalmente ajustados. Portanto, faz sentido testar a correlação entre u_t e u_{t-12} . Uma regressão de \hat{u}_t sobre \hat{u}_{t-12} produz $\hat{\beta}_{12} = -0,187$ e p -valor = 0,028, de modo que existe evidência de autocorrelação sazonal *negativa*. (A inclusão dos regressores

altera as coisas apenas modestamente: $\hat{\rho}_{12} = -0,170$ e p -valor = 0,052.) Isso é um pouco incomum e não tem uma explicação óbvia.

Suponha que você tem dados trimestrais e quer verificar a presença de correlação serial de primeira ou de quarta ordem. Usando regressores estritamente exógenos, como você procederia?

12.3 A CORREÇÃO DA CORRELAÇÃO SERIAL COM REGRESSORES ESTRITAMENTE EXÓGENOS

Se detectarmos correlação serial após aplicarmos um dos testes da Seção 12.2, teremos que fazer algo a respeito. Se nossa meta for estimar um modelo com dinâmica completa, precisaremos reespecificar o modelo. Em aplicações nas quais nossa meta não seja estimar um modelo totalmente dinâmico, teremos de encontrar um meio de realizar inferência estatística: como vimos na Seção 12.1, as estatísticas habituais de testes do MQO não mais são válidas. Nesta Seção, começamos com o importante caso de correlação serial AR(1). A abordagem tradicional para esse problema assume regressores fixos. O que na realidade é necessário é usar regressores estritamente exógenos. Portanto, no mínimo, não devemos usar essas correções quando as variáveis explicativas incluírem variáveis dependentes desfasadas.

A Obtenção do Melhor Estimador Linear Não-Viesado no Modelo AR(1)

Assumimos as hipóteses ST.1 a ST.4 de Gauss-Markov, mas relaxamos a hipótese ST.5. Em particular, assumimos que os erros seguem o modelo AR(1)

$$u_t = \rho u_{t-1} + e_t, \text{ para todo } t = 1, 2, \dots \quad (12.26)$$

Lembre-se de que a hipótese ST.2 implica que u_t tenha uma média condicional zero sobre X . Na análise seguinte, permitimos que a condicionalidade sobre X esteja implícita, para simplificar a notação. Assim, escrevemos a variância de u_t como

$$\text{Var}(u_t) = \sigma_e^2 / (1 - \rho^2). \quad (12.27)$$

Por simplicidade, considere o caso com uma única variável explicativa:

$$y_t = \beta_0 + \beta_1 x_t + u_t, \text{ para todo } t = 1, 2, \dots, n.$$

Como o problema nessa equação é a correlação serial em u_t , faz sentido transformar a equação para eliminar a correlação serial. Para $t \geq 2$, escrevemos

$$\begin{aligned} y_{t-1} &= \beta_0 + \beta_1 x_{t-1} + u_{t-1} \\ y_t &= \beta_0 + \beta_1 x_t + u_t \end{aligned}$$

Agora, se multiplicarmos a primeira equação por ρ , subtraindo o resultado da segunda equação, teremos

$$y_t - \rho y_{t-1} = (1 - \rho)\beta_0 + \beta_1(x_t - \rho x_{t-1}) + e_t, t \geq 2,$$

onde usamos o fato de que $e_t = u_t - \rho u_{t-1}$. Podemos escrever isso como

$$\tilde{y}_t = (1 - \rho)\beta_0 + \beta_1\tilde{x}_t + e_t, t \geq 2, \quad (12.28)$$

onde

$$\tilde{y}_t = y_t - \rho y_{t-1}, \tilde{x}_t = x_t - \rho x_{t-1} \quad (12.29)$$

são chamados de **dados quase-diferenciados**. (Se $\rho = 1$, eles serão dados diferenciados, mas lembre-se de que estamos assumindo $|\rho| < 1$.) Os erros em (12.28) são serialmente não-correlacionados; de fato, essa equação satisfaz todas as hipóteses de Gauss-Markov. Isso significa que, se conhecêssemos ρ , poderíamos estimar β_0 e β_1 regredindo \tilde{y}_t sobre \tilde{x}_t , desde que tivéssemos dividido o intercepto estimado por $(1 - \rho)$.

Os estimadores MQO de (12.28) não são exatamente BLUE porque eles não usam o primeiro período de tempo. Isso pode ser facilmente corrigido escrevendo-se a equação para $t = 1$, como

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad (12.30)$$

Como cada e_t é não-correlacionado com u_1 , podemos adicionar (12.30) em (12.28) e ainda assim ter erros serialmente não-correlacionados. Porém, usando (12.27), $\text{Var}(u_1) = \sigma_e^2/(1 - \rho^2) > \sigma_e^2 = \text{Var}(e_t)$. [A equação (12.27) claramente não se sustenta quando $|\rho| \geq 1$, razão pela qual assumimos a condição de estabilidade.] Assim, temos que multiplicar (12.30) por $(1 - \rho^2)^{1/2}$ para obter erros com a mesma variância:

$$(1 - \rho^2)^{1/2} y_1 = (1 - \rho^2)^{1/2} \beta_0 + \beta_1(1 - \rho^2)^{1/2} x_1 + (1 - \rho^2)^{1/2} u_1$$

ou

$$\tilde{y}_1 = (1 - \rho^2)^{1/2} \beta_0 + \beta_1 \tilde{x}_1 + \tilde{u}_1, \quad (12.31)$$

onde $\tilde{u}_1 = (1 - \rho^2)^{1/2} u_1$, $\tilde{y}_1 = (1 - \rho^2)^{1/2} y_1$, e assim por diante. O erro em (12.31) tem variância $\text{Var}(\tilde{u}_1) = (1 - \rho^2) \text{Var}(u_1) = \sigma_e^2$, de modo que podemos usar (12.31) com (12.28) em uma regressão MQO. Isso fornece estimadores BLUE de β_0 e β_1 sob as hipóteses ST.1 a ST.4 e o modelo AR(1) de u_t . Esse é outro exemplo de um estimador de *mínimos quadrados generalizados* (ou MQG). Vimos outros estimadores MQG no contexto da heteroscedasticidade, no Capítulo 8.

A adição de mais regressores muda pouca coisa. Para $t \geq 2$, usamos a equação

$$\tilde{y}_t = (1 - \rho)\beta_0 + \beta_1 \tilde{x}_{t1} + \dots + \beta_k \tilde{x}_{tk} + e_t, \quad (12.32)$$

onde $\tilde{x}_{ij} = x_{ij} - \rho x_{t-1,j}$. Para $t = 1$, temos $\tilde{y}_1 = (1 - \rho^2)^{1/2}y_1$, $\tilde{x}_{1j} = (1 - \rho^2)^{1/2}x_{1j}$ e o intercepto é $(1 - \rho^2)^{1/2}\beta_0$. Para um determinado ρ , é muito fácil transformar os dados e computar o MQO. A menos que $\rho = 0$, o estimador MQG, isto é, MQO sobre os dados transformados, será, em geral, diferente do estimador MQO original. O estimador MQG acaba sendo BLUE, e, como os erros na equação transformada são serialmente não-correlacionados e homoscedásticos, as estatísticas t e F da equação transformada são válidas (pelo menos assimptoticamente e, exatamente, se os erros e_t forem normalmente distribuídos).

A Estimação MQG Factível com Erros AR(1)

O problema com o estimador MQG é que ρ é raramente conhecido na prática. Porém, já sabemos como obter um estimador consistente de ρ : simplesmente regredimos os resíduos de MQO sobre seus equivalentes defasados, exatamente como na equação (12.14). Em seguida, usamos essa estimativa, $\hat{\rho}$, em lugar de ρ para obter as variáveis quase-diferenciadas. Usamos, então, MQO na equação

$$\tilde{y}_t = \beta_0 \tilde{x}_{t0} + \beta_1 \tilde{x}_{t1} + \dots + \beta_k \tilde{x}_{tk} + erro_t, \quad (12.33)$$

onde $\tilde{x}_{t0} = (1 - \hat{\rho})$ para $t \geq 2$ e $\tilde{x}_{10} = (1 - \hat{\rho}^2)^{1/2}$. Isso resulta no **estimador MQG factível (MQGF)** de β_j . O termo erro em (12.33) contém e_t e também os termos que envolvem o erro de estimativa em $\hat{\rho}$. Felizmente, o erro de estimativa em $\hat{\rho}$ não afeta a distribuição assimptótica dos estimadores MQGF.

A ESTIMAÇÃO DO MQG FACTÍVEL NO MODELO AR(1):

- Compute a regressão MQO de y , sobre x_{t1}, \dots, x_{tk} e obtenha os resíduos do MQO, \hat{u}_t , $t = 1, 2, \dots, n$.
- Compute a regressão na equação (12.14) e obtenha $\hat{\rho}$.
- Aplique o MQO à equação (12.33) para estimar $\beta_0, \beta_1, \dots, \beta_k$. Os erros-padrão, estatísticas t e estatísticas F usuais são assimptoticamente válidas.

O custo de usar $\hat{\rho}$ em lugar de ρ é que o estimador MQG factível não tem propriedades de amostra finita passíveis de tratamento. Em particular, ele não é não-viesado, embora seja consistente quando os dados são fracamente dependentes. Além disso, mesmo se e_t , em (12.32) for normalmente distribuído, as estatísticas t e F serão apenas aproximadamente distribuídas como t e F , devido ao erro de estimativa em $\hat{\rho}$. Isso é bom para a maioria dos propósitos, embora devamos ser cuidadosos com amostras de tamanhos pequenos.

Como o estimador MQGF não é não-viesado, certamente não podemos dizer que ele é BLUE. No entanto, ele é assimptoticamente mais eficiente que o estimador MQO quando o modelo AR(1) para correlação serial for válido (e as variáveis explicativas forem estritamente exógenas). Mais uma vez, essa afirmação assume que as séries de tempo são fracamente dependentes.

Existem vários nomes para a estimação MQGF do modelo AR(1) proveniente dos diferentes métodos de estimar ρ e dos diferentes tratamentos da primeira observação. A **estimação de Cochrane-Orcutt (CO)** omite a primeira observação e usa $\hat{\rho}$ de (12.14), enquanto a **estimação de Prais-Winsten**

(PW) usa a primeira observação da maneira anteriormente sugerida. Assimptoticamente, não faz diferença se a primeira observação é usada ou não, mas muitas amostras de séries temporais são pequenas, e assim as diferenças podem ser consideráveis em aplicações.

Na prática, tanto o método de Cochrane-Orcutt como o de Prais-Winsten são usados em um esquema iterativo. Uma vez que o estimador MQGF seja encontrado usando $\hat{\rho}$ de (12.14), podemos calcular um novo conjunto de resíduos, obter um novo estimador de ρ de (12.14), transformar os dados usando a nova estimativa de ρ e estimar (12.33) por MQO. Podemos repetir o processo inteiro muitas vezes, até que a estimativa de ρ mude muito pouco em relação à iteração anterior. Muitos programas de regressão implementam um procedimento iterativo automaticamente e, portanto, não há trabalho adicional para nós. É difícil dizer se mais de uma iteração ajuda. Parece ajudar em alguns casos, mas, teoricamente, as propriedades de amostras grandes do estimador iterado são as mesmas do estimador que usa somente a primeira iteração. Para detalhes sobre esses e outros métodos, veja Davidson e MacKinnon (1993, Capítulo 10).

EXEMPLO 12.4

(Estimação de Cochrane-Orcutt em Estudo de Evento)

Estimamos a equação no Exemplo 10.5 usando a estimação iterada de Cochrane-Orcutt. A título de comparação, também apresentamos os resultados na Tabela 12.1.

Os coeficientes que são estatisticamente significantes na estimação de Cochrane-Orcutt não diferem muito das estimativas MQO [em particular, os coeficientes de $\log(chemp_i)$, $\log(rtwex)$ e $afdec6$]. Não é surpreendente que coeficientes estatisticamente não significantes mudem, talvez destacadamente, entre os diferentes métodos.

Observe como os erros-padrão na segunda coluna são uniformemente mais altos que os da coluna (1). Isso é comum. Os erros-padrão de Cochrane-Orcutt levam em conta a correlação serial; os erros do MQO, não. Como vimos na Seção 12.1, os erros-padrão do MQO habitualmente subestimam a variação efetiva da amostragem nas estimativas MQO e não devemos confiar neles quando a correlação serial significante estiver presente. Portanto, o efeito nas importações chinesas após a decisão da *International Trade Commissions* é agora menos significante estatisticamente do que pensávamos ($t_{afdec6} = -1,68$).

O método Cochrane-Orcutt (CO) registra uma observação a menos que o MQO; isso reflete o fato de que a primeira observação transformada não é usada no método CO. Isso afeta levemente os graus de liberdade que são utilizados em testes de hipóteses.

Finalmente, um R -quadrado é informado na estimação CO, que é bem menor que o R -quadrado da estimação MQO neste caso. Porém, esses R -quadrados não devem ser comparados. No MQO, o R -quadrado, como sempre, é baseado na regressão com as variáveis dependentes e independentes sem transformação. No CO, o R -quadrado provém da regressão final da variável dependente transformada sobre as variáveis independentes transformadas. Não fica claro o que este R^2 está efetivamente medindo; mesmo assim, ele é tradicionalmente informado.

Tabela 12.1Variável Dependente: $\log(chnimp)$

Coeficiente	MQO	Cochrane-Orcutt
$\log(chemp)$	3,12 (0,48)	2,95 (0,65)
$\log(gas)$	0,196 (0,907)	1,05 (0,99)
$\log(rtwex)$	0,983 (0,400)	1,14 (0,51)
$befile6$	0,060 (0,261)	-0,016 (0,321)
$affile6$	-0,032 (0,264)	-0,033 (0,323)
$afdec6$	-0,565 (0,286)	-0,577 (0,343)
<i>intercepto</i>	-17,80 (21,05)	-37,31 (23,22)
$\hat{\rho}$	—	0,293 (0,084)
Observações	131	130
R-quadrado	0,305	0,193

Comparação entre MQO e MQGF

Em algumas aplicações dos métodos de Cochrane-Orcutt ou de Prais-Winsten, as estimativas MQGF diferem, em termos práticos, de maneira importante das estimativas MQO. (Esse não foi o caso no Exemplo 12.4.) Em geral, isso tem sido interpretado como uma verificação da superioridade do MQG factível sobre o MQO. Infelizmente, as coisas não são tão simples. Para entender o motivo, considere o modelo de regressão

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$

onde os processos das séries temporais são estacionários. Agora, assumindo que a lei dos grandes números se sustente, a consistência do MQO para β_1 se mantém se

$$\text{Cov}(x_t, u_t) = 0. \quad (12.34)$$

Anteriormente, afirmamos que o MQGF era consistente sob a hipótese de exogeneidade estrita, que é mais restritiva que (12.34). De fato, pode ser mostrado que a hipótese mais fraca que deve ser válida para que o MQGF seja consistente, *em adição* a (12.34), é que a soma de x_{t-1} com $x_{t+1} = 1$ seja não-correlacionada com u_t :

$$\text{Cov}[(x_{t-1} + x_{t+1}), u_t] = 0. \quad (12.35)$$

Em termos práticos, a consistência do MQGF exige que u_t seja não-correlacionado com x_{t-1} , x_t e x_{t+1} .

Como podemos mostrar que a condição (12.35) é necessária com (12.34)? O argumento será simples se assumirmos que ρ é conhecido e elimina o primeiro período, como no método Cochrane-Orcutt. O argumento quando usamos $\hat{\rho}$ é tecnicamente mais difícil e não produz nenhuma idéia adicional. Como uma observação não pode afetar as propriedades assintóticas de um estimador, sua eliminação não afetará o argumento. Agora, com ρ conhecido, o estimador MQG usa $x_t - \rho x_{t-1}$ como regressor em uma equação, onde $u_t - \rho u_{t-1}$ é o erro. Do teorema 11.1, sabemos que a condição crucial para a consistência do MQO é que o erro e o regressor sejam não-correlacionados. Nesse caso, necessitamos $E[(x_t - \rho x_{t-1})(u_t - \rho u_{t-1})] = 0$. Se expandirmos o valor esperado, obtemos

$$\begin{aligned} E[(x_t - \rho x_{t-1})(u_t - \rho u_{t-1})] &= E(x_t u_t) - \rho E(x_{t-1} u_t) - \rho E(x_t u_{t-1}) + \rho^2 E(x_{t-1} u_{t-1}) \\ &= -\rho [E(x_{t-1} u_t) + E(x_t u_{t-1})] \end{aligned}$$

porque $E(x_t u_t) = E(x_{t-1} u_{t-1}) = 0$ pela hipótese (12.34). Agora, sob estacionariedade, $E(x_t u_{t-1}) = E(x_{t+1} u_t)$, pois estamos apenas movendo o índice temporal um período à frente. Portanto,

$$E(x_{t-1} u_t) + E(x_t u_{t-1}) = E[(x_{t-1} + x_{t+1}) u_t],$$

e o último valor esperado é a covariância na equação (12.35), pois $E(u_t) = 0$. Mostramos que (12.35) é necessária com (12.34), para que o MQG seja consistente para β_1 . [Naturalmente, se $\rho = 0$, não necessitaremos de (12.35), pois estaremos retornando ao MQO.]

Nossa derivação mostra que o MQO e o MQGF podem produzir estimativas significantemente diferentes, porque (12.35) não se sustenta. Nesse caso, o MQO — que ainda é consistente sob (12.34) — é preferível ao MQGF (que é inconsistente). Se x tiver um efeito defasado sobre y , ou se x_{t+1} reage a mudanças em u_t , o MQGF pode produzir resultados equivocados.

Como o MQO e o MQGF são procedimentos diferentes de estimação, nunca esperamos que eles forneçam as mesmas estimativas. Se eles produzirem estimativas semelhantes de β_j , então, o MQGF será o preferido se houver evidência de correlação serial, pois o estimador é mais eficiente e as estatísticas de testes MQGF são pelo menos assintoticamente válidas. Um problema mais difícil surge quando existem diferenças práticas nas estimativas MQO e MQGF: é difícil determinar se tais diferenças são estatisticamente significantes. O método geral proposto por Hausman (1978) pode ser usado, mas está além do escopo desta obra.

A consistência e a normalidade assintótica do MQO e do MQGF apóiam-se fortemente na possibilidade de que os processos das séries temporais y_t e x_{tj} sejam fracamente dependentes. Coisas estranhas podem acontecer se aplicarmos MQO ou MQGF quando alguns processos tiverem raízes unitárias. Discutiremos mais sobre isso no Capítulo 18.

EXEMPLO 12.5

(A Curva de Phillips Estática)

A Tabela 12.2 apresenta estimativas de MQO e de Cochrane-Orcutt iteradas da curva de Phillips estática do Exemplo 10.1

EXEMPLO 12.5 (continuação)**Tabela 12.2**Variável Dependente: *inf*

Coefficiente	MQO	Cochrane-Orcutt
<i>desemp</i>	0,468 (0,289)	-0,665 (0,320)
<i>intercepto</i>	1,424 (1,719)	7,580 (2,379)
$\hat{\rho}$	—	0,774 (0,091)
Observações	49	48
R-Quadrado	0,053	0,086

O coeficiente de interesse é o de *desemp*, e difere acentuadamente entre CO e MQO. Como a estimativa CO é consistente com a relação de substituição inflação-desemprego, nossa tendência é focar nas estimativas CO. De fato, essas estimativas estão muito próximas das que são obtidas com a primeira diferenciação tanto de *inf* como de *desemp* (veja Problema 11.11), o que faz sentido, pois a quase-diferenciação usada em CO com $\hat{\rho} = 0,774$ é semelhante à primeira diferenciação. Pode apenas ser que *inf* e *desemp* não sejam relacionadas em níveis, mas elas têm uma relação negativa nas primeiras diferenças.

A Correção da Correlação Serial para Ordens Mais Elevadas

Também é possível corrigir a correlação serial de ordens mais elevadas. Um tratamento geral é dado por Harvey (1990). Aqui ilustramos a abordagem da correlação serial AR(2):

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + e_t,$$

onde $\{e_t\}$ satisfaz as hipóteses declaradas para o modelo AR(1). As condições de estabilidade agora são mais complicadas. É possível mostrar que elas são [veja Harvey (1990)]

$$\rho_2 > -1, \rho_2 - \rho_1 < 1 \text{ e } \rho_1 + \rho_2 < 1.$$

Por exemplo, o modelo será estável se $\rho_1 = 0,8$ e $\rho_2 = -0,3$; o modelo será instável se $\rho_1 = 0,7$ e $\rho_2 = 0,4$.

Assumindo que as condições de estabilidade se sustentem, podemos obter a transformação que elimina a correlação serial. No modelo de regressão simples, isso é fácil quando $t > 2$:

$$y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} = \beta_0(1 - \rho_1 - \rho_2) + \beta_1(x_t - \rho_1 x_{t-1} - \rho_2 x_{t-2}) + e_t$$

ou

$$\tilde{y}_t = \beta_0(1 - \rho_1 - \rho_2) + \beta_1 \tilde{x}_t + e_t, t = 3, 4, \dots, n. \quad (12.36)$$

Se conhecemos ρ_1 e ρ_2 , podemos facilmente estimar essa equação por MQO após obtermos as variáveis transformadas. Como raramente conhecemos ρ_1 e ρ_2 , temos que estimá-los. Como sempre, podemos usar os resíduos do MQO, \hat{u}_t ; obtemos $\hat{\rho}_1$ e $\hat{\rho}_2$ da regressão de

$$\hat{u}_t \text{ sobre } \hat{u}_{t-1}, \hat{u}_{t-2}, t = 3, \dots, n.$$

[Esta é a mesma regressão usada para testar a correlação serial AR(2) com regressores estritamente exógenos.] Portanto, usamos $\hat{\rho}_1$ e $\hat{\rho}_2$ em lugar de ρ_1 e ρ_2 para obter as variáveis transformadas. Isso fornece uma versão do estimador MQG factível. Se tivermos múltiplas variáveis explicativas, cada uma será transformada como $\tilde{x}_{tj} = x_{tj} - \hat{\rho}_1 x_{t-1,j} - \hat{\rho}_2 x_{t-2,j}$, quando $t > 2$.

O tratamento das primeiras duas observações é um pouco delicado. É possível mostrar que a variável dependente e cada variável independente (inclusive o intercepto) devem ser transformados como

$$\begin{aligned}\tilde{z}_1 &= \{(1 + \rho_2)[1 - \rho_2]^2 - \rho_1^2\}/(1 - \rho_2)^{1/2} z_1 \\ \tilde{z}_2 &= (1 - \rho_2^2)^{1/2} z_2 - [\rho_1(1 - \rho_1^2)^{1/2}/(1 - \rho_2)]z_1,\end{aligned}$$

onde z_1 e z_2 representam ou a variável dependente ou uma variável independente em $t = 1$ e $t = 2$, respectivamente. Não deduziremos essas transformações. Resumidamente, elas eliminam a correlação serial entre as duas primeiras observações e fazem suas variâncias de erros serem iguais a σ_e^2 .

Felizmente, os programas econométricos destinados à análise de séries temporais estimam facilmente modelos com erros AR(q) gerais; raramente necessitamos, nós mesmos, calcular diretamente as variáveis transformadas.

12.4 DIFERENCIACÃO E CORRELAÇÃO SERIAL

No Capítulo 11, apresentamos a diferenciação como uma transformação para tornar um processo integrado fracamente dependente. Existe outra maneira de ver os méritos da diferenciação quando tratamos com dados altamente persistentes. Suponha que iniciemos com um modelo de regressão simples:

$$y_t = \beta_0 + \beta_1 x_t + u_t, t = 1, 2, \dots, \quad (12.37)$$

onde u_t segue o processo AR(1) em (12.26). Como mencionamos na Seção 11.3, e como discutiremos mais completamente no Capítulo 18, os procedimentos de inferência usuais do MQO podem ser muito enganosos quando as variáveis y_t e x_t forem integradas de ordem um, ou I(1). No caso extremo, no qual os erros $\{u_t\}$ em (12.37) seguem um passeio aleatório, a equação não faz sentido porque, entre outras coisas, a variância de u_t cresce com t . É mais lógico diferenciar a equação:

$$\Delta y_t = \beta_1 \Delta x_t + \Delta u_t, t = 2, \dots, n. \quad (12.38)$$

Se u_t seguir um passeio aleatório, então, $e_t \equiv \Delta u_t$, tem média zero e variância constante, e é serialmente não-correlacionado. Assim, assumindo que e_t e Δx_t são não-correlacionados, podemos estimar (12.38) por MQO, em que perdemos a primeira observação.

Mesmo se u_t não seguir um passeio aleatório, mas ρ for positivo e grande, fazer a primeira diferenciação é uma boa idéia: ela eliminará a maior parte da correlação serial. Naturalmente, (12.38) é diferente de (12.37), mas pelo menos podemos ter mais confiança nos erros-padrão e nas estatísticas t do MQO em (12.38). A permissão de múltiplas variáveis explicativas não muda nada.

EXEMPLO 12.6

(A Diferenciação da Equação da Taxa de Juros)

No Exemplo 10.2, estimamos uma equação relacionando a taxa de juros das letras do Tesouro norte-americano de três meses com a inflação e com o déficit federal [veja equação (10.15)]. Se regredirmos os resíduos dessa equação sobre uma única defasagem, obteremos $\hat{\rho} = 0,530$ (0,123), que é, estatisticamente, maior que zero. Se diferenciarmos i_3 , inf e def , e verificarmos se os resíduos têm correlação serial AR(1), obtemos $\hat{\rho} = 0,068$ (0,145), de modo que não existirá evidência de correlação serial. A diferenciação aparentemente eliminou qualquer correlação serial. [Além disso, existe evidência de que i_3 contém uma raiz unitária, e inf também pode ter; portanto, de qualquer forma, a diferenciação pode ser necessária para produzir variáveis l(0).]

Suponha que, após ter estimado um modelo por MQO, você estime ρ da regressão (12.14) e obtenha $\hat{\rho} = 0,92$. O que você faria?

Como explicamos no Capítulo 11, a decisão de diferenciar ou não é difícil de ser tomada. Contudo, essa discussão salienta outro benefício da diferenciação, que é o fato de ela remover a correlação serial. Voltaremos a esse assunto no Capítulo 18.

12.5 INFERÊNCIA ROBUSTA EM RELAÇÃO À CORRELAÇÃO SERIAL APÓS O MQO

Nos últimos anos, tornou-se muito popular estimar modelos por MQO, porém, corrigindo os erros-padrão em relação a formas bastante arbitrárias de correlação serial (e heteroscedasticidade). Embora saibamos que o MQO será ineficiente, existem algumas boas razões para utilizar essa abordagem. Primeiro, as variáveis explicativas podem não ser estritamente exógenas. Nesse caso, o MQGF não será sequer consistente, sem falar em eficiência. Segundo, na maioria das aplicações do MQGF, assume-se que os erros seguem um modelo AR(1). Pode ser melhor calcular erros-padrão das estimativas MQO que sejam robustos em relação a formas mais gerais de correlação serial.

Para entender a idéia, considere a equação (12.4), que é a variância do estimador de inclinação do MQO em um modelo de regressão simples com erros AR(1). Podemos estimar essa variância de maneira bastante simples, inserindo nossos estimadores padrão de ρ e σ^2 . O único problema com isso

é que ele pressupõe o modelo AR(1) e também a homoscedasticidade. É possível relaxar ambas essas hipóteses.

Um tratamento geral de erros-padrão que sejam robustos tanto em relação à heteroscedasticidade quanto em relação à correlação serial é fornecido por Davidson e MacKinnon (1993). É possível apresentar imediatamente um método simples para calcular o erro-padrão robusto de qualquer coeficiente MQO.

Nosso tratamento aqui segue Wooldridge (1989). Considere o modelo padrão de regressão linear múltipla

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t, t = 1, 2, \dots, n, \quad (12.39)$$

que estimamos por MQO. Concretamente, estamos interessados em obter um erro-padrão robusto de $\hat{\beta}_1$, em relação à correlação serial. Isso é feito muito facilmente: escreva x_{t1} como uma função linear das demais variáveis independentes remanescentes e um termo erro,

$$x_{t1} = \delta_0 + \delta_2 x_{t2} + \dots + \delta_k x_{tk} + r_t, \quad (12.40)$$

onde o erro r_t tenha média zero e seja não-correlacionado com $x_{t2}, x_{t3}, \dots, x_{tk}$.

Então, pode ser demonstrado que a variância assintótica do estimador MQO $\hat{\beta}_1$ é

$$\text{Avar}(\hat{\beta}_1) = \left(\sum_{t=1}^n E(r_t^2) \right)^{-2} \text{Var} \left(\sum_{t=1}^n r_t u_t \right).$$

Sob a hipótese de não-correlação serial ST.5', $\{a_t = r_t u_t\}$ é serialmente não-correlacionado, e assim, os erros-padrão habituais do MQO (sob homoscedasticidade) ou os erros-padrão robustos em relação à heteroscedasticidade serão válidos. Mas, se ST.5' não se mantiver, nossa expressão para $\text{Avar}(\hat{\beta}_1)$ deve levar em conta a correlação entre a_t e a_s , quando $t \neq s$. Na prática, é comum supor que, uma vez que os termos estejam mais afastados do que alguns poucos períodos, a correlação será essencialmente zero. Lembre-se de que sob fraca dependência, a correlação deve ser próxima de zero, de modo que esta é uma abordagem razoável.

Seguindo a estrutura geral de Newey e West (1987), Wooldridge (1989) mostra que $\text{Avar}(\hat{\beta}_1)$ pode ser estimado da seguinte maneira. Seja “ep($\hat{\beta}_1$)” o erro-padrão usual (mas incorreto) do MQO e seja $\hat{\sigma}$ o erro-padrão habitual da regressão (ou a raiz quadrada do erro quadrático médio) da estimativa de (12.39) por MQO. Sejam \hat{r}_t os resíduos da regressão auxiliar de

$$x_{t1} \text{ sobre } x_{t2}, x_{t3}, \dots, x_{tk} \quad (12.41)$$

(incluindo uma constante, como habitualmente). Para um dado número inteiro $g > 0$, defina

$$\hat{\nu} = \sum_{t=1}^n \hat{a}_t^2 + 2 \sum_{h=1}^g [1 - h/(g+1)] \left(\sum_{t=h+1}^n \hat{a}_t \hat{a}_{t-h} \right), \quad (12.42)$$

onde

$$\hat{a}_t = \hat{\gamma}_t \hat{u}_t, t = 1, 2, \dots, n.$$

Isso parece um pouco complicado, mas na prática é fácil de ser obtido. O número inteiro g em (12.42) controla o quanto de correlação serial estamos permitindo no cálculo do erro-padrão. Uma vez que tenhamos $\hat{\nu}$, o erro-padrão robusto em relação à correlação serial de $\hat{\beta}_1$ será simplesmente

$$\text{ep}(\hat{\beta}_1) = [\text{ep}(\hat{\beta}_1)/\hat{\sigma}]^2 \sqrt{\hat{\nu}}. \quad (12.43)$$

Em outras palavras, consideramos o erro-padrão usual do MQO de $\hat{\beta}_1$, fazemos sua divisão por $\hat{\sigma}$, elevamos esse quociente ao quadrado, e multiplicamos o resultado pela raiz quadrada de $\hat{\nu}$. Isso pode ser usado para construir intervalos de confiança e estatísticas t de $\hat{\beta}_1$.

É útil verificar o aspecto de $\hat{\nu}$ em alguns casos simples. Quando $g = 1$,

$$\hat{\nu} = \sum_{t=1}^n \hat{a}_t^2 + \sum_{t=2}^n \hat{a}_t \hat{a}_{t-1}, \quad (12.44)$$

e quando $g = 2$,

$$\hat{\nu} = \sum_{t=1}^n \hat{a}_t^2 + (4/3) \left(\sum_{t=2}^n \hat{a}_t \hat{a}_{t-1} \right) + (2/3) \left(\sum_{t=3}^n \hat{a}_t \hat{a}_{t-2} \right). \quad (12.45)$$

Quanto maior for g , maior será o número de termos incluídos para corrigir a correlação serial. O propósito do fator $[1 - h/(g + 1)]$ em (12.42) é para assegurar que $\hat{\nu}$ seja de fato não-negativo. [Newey e West (1987) verificam isso.] Claramente, necessitamos de $\hat{\nu} \geq 0$, pois $\hat{\nu}$ está estimando uma variância e a raiz quadrada de $\hat{\nu}$ aparece em (12.43).

O erro-padrão em (12.43) também acaba sendo robusto em relação à heteroscedasticidade. Aliás, se eliminarmos o segundo termo em (12.42), então (12.43), torna-se o erro-padrão robusto usual em relação à heteroscedasticidade que discutimos no Capítulo 8 (sem o ajuste dos graus de liberdade).

A teoria que fundamenta o erro-padrão em (12.43) é técnica e um pouco sutil. Lembre-se, começamos afirmando não conhecer a forma da correlação serial. Nesse caso, como poderemos selecionar o número inteiro g ? A teoria afirma que (12.43) funciona para formas bastante arbitrárias de correlação serial, desde que g cresça com o tamanho n da amostra. A idéia é que, com tamanho maiores de amostras, podemos ser mais flexíveis sobre o grau de correlação em (12.42). Há muitos trabalhos recentes sobre a relação entre g e n , mas não vamos discutir esse assunto aqui. Para dados anuais, a escolha de um g pequeno, como $g = 1$ ou $g = 2$, muito provavelmente levará em conta a maior parte da correlação serial. Para dados trimestrais ou mensais, g provavelmente terá que ser maior (como $g = 4$ ou 8 para dados trimestrais e $g = 12$ ou 24 para dados mensais), supondo que tenhamos dados suficientes. Newey e West (1987) recomendam considerar g como a parte inteira de $4(n/100)^{2/9}$; outros sugerem que g seja a parte inteira de $n^{1/4}$. A sugestão de Newey-West está implementada pelo programa de econometria Eviews®. Para, digamos, $n = 50$ (que é razoável para dados anuais pós-Segunda Guerra Mundial), $g = 3$. (A parte inteira de $n^{1/4}$ é $g = 2$.)

Façamos um resumo de como obter um erro-padrão de $\hat{\beta}_1$, robusto em relação à correlação serial. Naturalmente, como podemos primeiro listar qualquer variável independente, o procedimento seguinte funciona para calcular um erro-padrão de qualquer coeficiente de inclinação.

ERRO-PADRÃO ROBUSTO DE $\hat{\beta}_1$ EM RELAÇÃO À CORRELAÇÃO SERIAL:

- (i) Estime (12.39) pelo MQO, que produzirá “ $ep(\hat{\beta}_1)$ ”, $\hat{\sigma}$ e os resíduos MQO $\{\hat{u}_t; t = 1, \dots, n\}$.
- (ii) Compute os resíduos $\{\hat{r}_t; t = 1, \dots, n\}$ da regressão auxiliar (12.41). Em seguida, forme $\hat{a}_t = \hat{r}_t \hat{u}_t$ (para cada t).
- (iii) Para sua escolha de g , calcule \hat{v} como em (12.42).
- (iv) Calcule $ep(\hat{\beta}_1)$ de (12.43).

Empiricamente, os erros-padrão robustos em relação à correlação serial são em geral maiores que os erros-padrão usuais do MQO quando existe correlação serial. Isso é devido, na maioria dos casos, ao fato de os erros serem serialmente correlacionados. Porém, é possível ter correlação serial substancial em $\{u_t\}$, mas também ter similaridades entre os erros-padrão habituais e os robustos em relação à correlação serial (CS-robustos) de alguns coeficientes: são as autocorrelações amostrais de $\hat{a}_t = \hat{r}_t \hat{u}_t$, que determinam o erro-padrão robusto de $\hat{\beta}_1$.

O uso de erros-padrão CS-robustos fica em segundo plano relativamente ao uso de erros-padrão robustos somente em relação à heteroscedasticidade, por várias razões. Primeiro, cortes transversais grandes, em que os erros-padrão robustos em relação à heteroscedasticidade têm boas propriedades, são mais comuns que séries temporais grandes. Os erros-padrão CS-robustos podem se comportar de maneira ineficiente quando houver correlação serial substancial e o tamanho da amostra for pequeno (em que até mesmo, digamos, 100, pode ser um número pequeno). Segundo, como temos que escolher o inteiro g na equação (12.42), o cálculo dos erros-padrão CS-robustos não é automático. Como mencionado anteriormente, alguns programas econômétricos fazem a seleção automaticamente, mas você terá que aceitar a escolha.

Outra razão importante pela qual os erros-padrão CS-robustos ainda não são calculados rotineiramente é que na presença de correlação serial severa, o MQO pode ser muito ineficiente, especialmente em amostras de tamanho pequeno. Após executarmos o MQO e corrigirmos os erros-padrão em relação à correlação serial, os coeficientes, com freqüência, são não significantes ou, pelo menos, menos significantes do que eram com os erros-padrão habituais do MQO.

Os erros-padrão CS-robustos após a estimação MQO são muito úteis quando temos dúvidas em saber se algumas das variáveis explicativas são estritamente exógenas, de forma que métodos como o de Cochrane-Orcutt não são sequer consistentes. Também é válido usar os erros-padrão CS-robustos em modelos com a variável dependente defasada, supondo, é claro, que existem boas razões para permitir a correlação serial em tais modelos.

EXEMPLO 12.7

{O Salário Mínimo Porto-Riquenho}

Obtemos um erro-padrão CS-robusto para o efeito do salário mínimo na equação de desemprego porto-riquenho. No Exemplo 12.2, encontramos evidência bastante forte de correlação serial AR(1). Como naquele exemplo, usamos como controles adicionais $\log(uspnrb)$, $\log(prpnrb)$ e uma tendência temporal linear.

A estimativa MQO da elasticidade da taxa de desemprego em relação ao salário mínimo é $\hat{\beta}_1 = -0,2123$, e o erro-padrão MQO usual é “ $ep(\hat{\beta}_1)$ ” = 0,0402. O erro-padrão da regressão é $\hat{\sigma} = 0,0328$. Além disso, utilizando o procedimento anterior com $g = 2$ [veja (12.45)], obtemos $\hat{v} = 0,000805$. Isso produz

EXEMPLO 12.7 (continuação)

o erro-padrão robusto em relação à correlação serial e à heteroscedasticidade $ep(\hat{\beta}_1) = [(0,0402/0,0328)^2]\sqrt{0,000805} \approx 0,0426$. Curiosamente, o erro-padrão robusto é somente pouca coisa maior que o erro-padrão usual do MQO. A estatística t robusta está em torno de $-4,98$, de modo que a elasticidade estimada ainda é, estatisticamente, bastante significante.

A título de comparação, a estimativa CO iterada de β_1 é $-0,1111$, com um erro-padrão de $0,0446$. Portanto, a estimativa MQF está muito mais perto de zero do que a estimativa MQO, e podemos suspeitar de violação da hipótese de exogeneidade estrita. Ou então, a diferença entre as estimativas MQO e MQF poderia ser explicada por erro de amostragem. É muito difícil dizer.

Antes de encerrarmos esta seção, observamos que é possível construir estatísticas do tipo F robustas em relação à correlação serial para testar hipóteses múltiplas, mas esse é um assunto por demais avançado para ser discutido aqui. [Veja abordagens desse assunto em Wooldridge (1991b, 1995) e Davidson e MacKinnon (1993)].

12.6 HETEROSCEDASTICIDADE EM REGRESSÕES DE SÉRIES TEMPORAIS

Discutimos sobre como testar e corrigir a heteroscedasticidade em aplicações de corte transversal no Capítulo 8. A heteroscedasticidade também pode ocorrer em modelos de regressões de séries temporais, e a presença de heteroscedasticidade, embora não cause viés ou inconsistência em $\hat{\beta}_j$, invalida os erros-padrão, as estatísticas t e as estatísticas F usuais. O problema é o mesmo do caso do corte transversal.

Nas aplicações de regressões de séries de tempo, a heteroscedasticidade recebe pouca atenção, quando recebe; o problema dos erros serialmente correlacionados em geral é mais premente. Apesar disso, vale a pena salarmos brevemente sobre alguns dos problemas que surgem na aplicação de testes e de correções de heteroscedasticidade em regressões de séries temporais.

Como as estatísticas habituais de MQO são assimptoticamente válidas sob as hipóteses ST.1' a ST.5', estamos interessados no que acontece quando a hipótese de homoscedasticidade, ST.4', não se sustenta. A hipótese ST.2' elimina más especificações, como a de omissão de variáveis e certos tipos de erros de medida, enquanto a hipótese ST.5' elimina a correlação serial nos erros. É importante lembrar que erros serialmente correlacionados causam problemas cujos testes e ajustes em relação à heteroscedasticidade não são passíveis de tratamento.

Estatísticas Robustas em Relação à Heteroscedasticidade

Ao estudarmos a heteroscedasticidade das regressões de corte transversal, observamos como ela não tem influência na inexistência de viés ou na consistência dos estimadores MQO. Exatamente as mesmas conclusões são válidas no caso de séries temporais, como podemos verificar revendo a hipótese necessária para a inexistência de viés (teorema 10.1) e para a consistência (teorema 11.1).

Na Seção 8.2, discutimos como os erros-padrão, assim como as estatísticas t e F habituais do MQO, podem ser ajustadas para permitir a existência de heteroscedasticidade de forma desconhecida. Esses mesmos ajustes funcionam nas regressões de séries temporais sob as hipóteses ST.1' a ST.5'. Portanto, desde que a única hipótese violada seja a de homoscedasticidade, a inferência válida é facilmente obtida na maioria dos programas econométricos.

O Teste da Heteroscedasticidade

Algumas vezes, desejarcemos testar a presença de heteroscedasticidade nas regressões de séries de tempo, especialmente se estivermos preocupados com o desempenho das estatísticas robustas em relação à heteroscedasticidade, em amostras de tamanhos relativamente pequenos. Os testes que vimos no Capítulo 8 podem ser aplicados diretamente, porém, com algumas limitações. Primeiro, os erros u_t não devem ser serialmente correlacionados; qualquer correlação serial invalidará, de forma geral, um teste da heteroscedasticidade. Assim, faz sentido primeiro testar a presença de correlação serial, usando um teste robusto em relação à heteroscedasticidade se houver suspeita de heteroscedasticidade. Em seguida, após termos feito alguma coisa para corrigir a correlação serial, podemos fazer o teste da heteroscedasticidade.

Segundo, considere a equação utilizada para motivar o teste de Breusch-Pagan da heteroscedasticidade:

$$u_t^2 = \delta_0 + \delta_1 x_{t1} + \dots + \delta_k x_{tk} + \nu_t, \quad (12.46)$$

onde a hipótese nula é $H_0: \delta_1 = \delta_2 = \dots = \delta_k = 0$. Para que a estatística F — com \hat{u}_t^2 substituindo u_t^2 como a variável dependente — seja válida, temos de assumir que os erros $\{\nu_t\}$ sejam, eles próprios, homoscedásticos (como no caso do corte transversal) e serialmente não-correlacionados. Isso é implicitamente assumido no cálculo de todos os testes padrão de heteroscedasticidade, inclusive na versão do teste de White discutido na Seção 8.3. Assumir que $\{\nu_t\}$ seja serialmente não-correlacionado exclui certas formas de heteroscedasticidade dinâmica, assunto que trataremos na próxima subseção.

Se for encontrada heteroscedasticidade em u_t (e os u_t forem serialmente não-correlacionados), poderão ser usadas as estatísticas de testes robustos em relação à heteroscedasticidade. Uma alternativa é usar **mínimos quadrados ponderados**, como na Seção 8.4. A mecânica dos mínimos quadrados ponderados para o caso das séries de tempo é idêntica à do corte transversal.

EXEMPLO 12.8

(A Heteroscedasticidade e a Hipótese de Mercados Eficientes)

No Exemplo 11.4, estimamos o modelo simples

$$\text{returno}_t = \beta_0 + \beta_1 \text{returno}_{t-1} + u_t, \quad (12.47)$$

A HME afirma que $\beta_1 = 0$. Quando testamos essa hipótese utilizando os dados contidos no arquivo NYSE.RAW, obtivemos $t_{\beta_1} = 1,55$ com $n = 689$. Com uma amostra tão grande isso não é muita evidência contra a HME. Embora a HME declare que o retorno esperado, dadas as informações passadas observáveis, deve ser constante, ela nada diz sobre a variância condicional. De fato, o teste de Breusch-Pagan da heteroscedasticidade requer a regressão dos resíduos MQO \hat{u}_t^2 sobre returno_{t-1} :

$$\begin{aligned} \hat{u}_t^2 &= 4,66 - 1,104 \text{returno}_{t-1} + \text{resíduo}_t \\ &\quad (0,43) \quad (0,201) \\ n &= 689, R^2 = 0,042. \end{aligned} \quad (12.48)$$

EXEMPLO 12.8 (continuação)

A estatística t de retorno_{t-1} está em torno de $-5,5$, indicando forte evidência de heteroscedasticidade. Como o coeficiente de retorno_{t-1} é negativo, temos a interessante descoberta de que a volatilidade dos retornos das ações é menor quando o retorno anterior era alto, e vice-versa. Portanto, encontramos o que é comum em muitos estudos financeiros: o valor esperado dos retornos das ações não depende de retornos passados, mas a variância dos retornos depende.

Como você calcularia o teste de White da heteroscedasticidade na equação (12.47)?

A Heteroscedasticidade Condicional Auto-Regressiva

Nos últimos anos, os economistas vêm demonstrando interesse em formas dinâmicas de heteroscedasticidade. Naturalmente, se x_t contiver uma variável dependente defasada, a heteroscedasticidade, como em (12.46), é dinâmica. Entretanto, formas dinâmicas de heteroscedasticidade podem aparecer até mesmo em modelos sem dinâmica na equação de regressão.

Para verificar isso, considere um modelo estático de regressão simples:

$$y_t = \beta_0 + \beta_1 z_t + u_t,$$

e suponha que as hipóteses de Gauss-Markov sejam válidas. Isso significa que os estimadores MQO são BLUE. A hipótese de homoscedasticidade diz que $\text{Var}(u_t | Z)$ é constante, em que Z representa todos os n resultados de z_t . Mesmo se a variância de u_t , dado Z , for constante, existem outras possibilidades de surgir heteroscedasticidade. Engle (1982) sugeriu verificar a variância condicional de u_t , dados os erros passados (nos quais o condicionamento em Z está implícito). Engle sugeriu o que é conhecido como o modelo de **heteroscedasticidade condicional auto-regressiva (ARCH)**. O modelo ARCH de primeira ordem é

$$\mathbb{E}(u_t^2 | u_{t-1}, u_{t-2}, \dots) = \mathbb{E}(u_t^2 | u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2, \quad (12.49)$$

onde deixamos implícito o condicionamento em Z . Essa equação representa a variância condicional de u_t , dados u_t passados, somente se $\mathbb{E}(u_t | u_{t-1}, u_{t-2}, \dots) = 0$, o que significa que os erros são serialmente não-correlacionados. Como variâncias condicionais devem ser positivas, esse modelo somente faz sentido se $\alpha_0 > 0$ e $\alpha_1 \geq 0$; se $\alpha_1 = 0$, não há dinâmica na equação de variância.

É ilustrativo escrever (12.49) como

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \nu_t, \quad (12.50)$$

onde o valor esperado de ν_t (dados u_{t-1}, u_{t-2}, \dots) é zero por definição. (Os ν_t não são independentes de u_t passados, devido à restrição $\nu_t \geq -\alpha_0 - \alpha_1 u_{t-1}^2$.) A equação (12.50) parece um modelo auto-

regressivo em u_t^2 (daí o nome ARCH). A condição de estabilidade dessa equação é $\alpha_1 < 1$, assim como no modelo AR(1) usual. Quando $\alpha_1 > 0$, os erros quadrados contêm correlação serial (positiva), mesmo que os próprios u_t não a contenham.

Que implicações (12.50) tem para o MQO? Como começamos assumindo que as hipóteses de Gauss-Markov são válidas, o MQO é BLUE. Além disso, mesmo se u_t não for normalmente distribuído, sabemos que as estatísticas de teste MQO habituais são assintoticamente válidas sob as hipóteses ST.1' a ST.5', que são satisfeitas por modelos estáticos e de defasagens distribuídas com erros ARCH.

Se o MQO ainda tiver propriedades desejáveis sob ARCH, por que devemos nos importar sobre formas ARCH de heteroscedasticidade em modelos estáticos e de defasagens distribuídas? Devemos nos preocupar por dois motivos. Primeiro, é possível obter estimadores consistentes (mas não-viesados) de β_j que sejam *assimptoticamente* mais eficientes do que os estimadores MQO. Um procedimento de mínimos quadrados ponderados, com base no cálculo de (12.50), resolverá o problema. Um procedimento de máxima verossimilhança também funciona sob a hipótese de que os erros u_t tenham uma distribuição normal condicional. Segundo, economistas de vários campos vêm demonstrando interesse na dinâmica da variância condicional. A aplicação original de Engle enfatizou a variância da inflação do Reino Unido, na qual ele descobriu que uma magnitude maior dos erros no período de tempo anterior (maior u_{t-1}^2) estava associada a uma variância de erro maior no período corrente. Como a variância é freqüentemente usada para medir a volatilidade, e a volatilidade é um elemento crucial nas teorias de determinação de preços de ativos, os modelos ARCH têm se tornado importantes em estudos empíricos da área de finanças.

Os modelos ARCH também se aplicam quando há dinâmica na média condicional. Suponha que tenhamos a variável dependente, y_t , uma variável contemporânea exógena, z_t , e

$$E(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \dots) = \beta_0 + \beta_1 z_t + \beta_2 y_{t-1} + \beta_3 z_{t-1},$$

de forma que no máximo uma defasagem de y e de z aparecem na regressão dinâmica. A abordagem geral é assumir que $\text{Var}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \dots)$ é constante, como discutimos no Capítulo 11. Contudo, essa variância poderia seguir um modelo ARCH:

$$\begin{aligned} \text{Var}(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \dots) &= \text{Var}(u_t | y_{t-1}, z_{t-1}, y_{t-2}, \dots) \\ &= \alpha_0 + \alpha_1 u_{t-1}^2, \end{aligned}$$

onde $u_t = y_t - E(y_t | z_t, y_{t-1}, z_{t-1}, y_{t-2}, \dots)$. Como sabemos do Capítulo 11, a presença do ARCH não afeta a consistência do MQO e os usuais erros-padrão e estatísticas de testes robustos em relação à heteroscedasticidade são válidos. (Lembre-se, eles são válidos para qualquer forma de heteroscedasticidade, e o ARCH é apenas uma forma particular de heteroscedasticidade.)

Se você estiver interessado no modelo ARCH e suas ampliações, veja Bollerslev, Chou e Kroner (1992) e Bollerslev, Engle e Nelson (1994) para análises recentes.

EXEMPLO 12.9

(O ARCH em Retornos de Ações)

No Exemplo 12.8, vimos que havia heteroscedasticidade nos retornos de ações semanais. Essa heteroscedasticidade de fato é melhor caracterizada pelo modelo ARCH em (12.50). Se computarmos os resíduos do MQO de (12.47), elevarmos ao quadrado, e regredirmos o resultado sobre os resíduos quadrados defasados, obteremos

EXEMPLO 12.9 {continuação}

$$\begin{aligned}\hat{u}_t^2 &= 2,95 + 0,337 \hat{u}_{t-1}^2 + \text{resíduo}_t \\ (0,44) &\quad (0,036) \\ n &= 688, R^2 = 0,114.\end{aligned}\tag{12.51}$$

A estatística t de \hat{u}_{t-1}^2 é maior que nove, indicando forte ARCH. Como discutimos anteriormente, um erro maior no tempo $t - 1$ implica uma maior variância nos retornos correntes das ações.

É importante verificar que, embora os resíduos quadrados do MQO sejam autocorrelacionados, os resíduos do MQO não o são (consistente em relação à HME). A regressão de \hat{u}_t sobre \hat{u}_{t-1} produz $\hat{\rho} = 0,0014$ com $t_{\hat{\rho}} \approx 0,038$.

Heteroscedasticidade e Correlação Serial em Modelos de Regressão

Nada impede a possibilidade de que tanto a heteroscedasticidade como a correlação serial estejam presentes em um modelo de regressão. Se estivermos inseguros, podemos sempre utilizar o MQO e calcular erros-padrão totalmente robustos, como descrito na Seção 12.5.

Muito da correlação serial é visto como o problema mais importante, pois usualmente ela tem um impacto maior nos erros-padrão e na eficiência dos estimadores do que a heteroscedasticidade. Como concluímos na Seção 12.2, a obtenção de testes de correlação serial que sejam robustos em relação à heteroscedasticidade arbitrária é bastante direta. Se detectarmos correlação serial usando esse tipo de teste, poderemos empregar a transformação de Cochrane-Orcutt [veja equação (12.32)] e, na equação transformada, usar erros-padrão e estatísticas de testes robustos em relação à heteroscedasticidade. Ou poderemos até mesmo verificar a existência de heteroscedasticidade em (12.32) utilizando os testes de Breusch-Pagan ou de White.

De maneira alternativa, podemos modelar a heteroscedasticidade e a correlação serial e corrigir os dois problemas por meio de um procedimento combinado de AR(1) e mínimos quadrados ponderados. Especificamente, considere o modelo

$$\begin{aligned}y_t &= \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t \\ u_t &= \sqrt{h_t} \nu_t \\ \nu_t &= \rho \nu_{t-1} + e_t, |\rho| < 1,\end{aligned}\tag{12.52}$$

onde as variáveis explicativas X são independentes de e_t para todo t , e h_t é uma função de x_t . O processo $\{e_t\}$ tem média zero e variância constante σ_e^2 e é serialmente não-correlacionado. Portanto, $\{\nu_t\}$ satisfaz um processo AR(1) estável. Suprimindo o condicionamento às variáveis explicativas, temos

$$\text{Var}(u_t) = \sigma_e^2 h_t$$

onde $\sigma_e^2 = \sigma_e^2 / (1 - \rho^2)$. Porém, $\nu_t = u_t / \sqrt{h_t}$ é homoscedástico e segue um modelo AR(1) estável. Portanto, a equação transformada

$$y_t/\sqrt{h_t} = \beta_0(1/\sqrt{h_t}) + \beta_1(x_{t1}/\sqrt{h_t}) + \dots + \beta_k(x_{tk}/\sqrt{h_t}) + v_t \quad (12.53)$$

tem erros AR(1). Agora, se tivermos em mente um tipo particular de heteroscedasticidade — isto é, conhecermos h_t — podemos estimar (12.52) usando os métodos padrão CO ou PW.

Na maioria dos casos, teremos que primeiro estimar h_t . O método seguinte combina o método dos mínimos quadrados ponderados da Seção 8.4 com a correção da correlação serial AR(1) da Seção 12.3.

O MQG FACTÍVEL COM HETEROSCEDASTICIDADE E CORRELAÇÃO SERIAL AR(1):

- (i) Estime (12.52) por MQO e guarde os resíduos \hat{u}_t .
- (ii) Regreda $\log(\hat{u}_t^2)$ sobre x_{t1}, \dots, x_{tk} (ou sobre \hat{y}_t, \hat{y}_t^2) e obtenha os valores estimados, digamos \hat{g}_t .
- (iii) Obtenha as estimativas de h_t ; $\hat{h}_t = \exp(\hat{g}_t)$.
- (iv) Estime a equação transformada

$$\hat{h}_t^{-1/2}y_t = \hat{h}_t^{-1/2}\beta_0 + \beta_1\hat{h}_t^{-1/2}x_{t1} + \dots + \beta_k\hat{h}_t^{-1/2}x_{tk} + erro_t \quad (12.54)$$

pelos métodos padrão de Cochrane-Orcutt ou de Prais-Winsten.

Esses estimadores MQG factíveis são assintoticamente eficientes. Mais importante, todos os erros-padrão e estatísticas de testes dos métodos CO e PW são assintoticamente válidos.

Apresentamos o importante problema da correlação serial nos erros de modelos de regressão múltipla. A correlação positiva entre erros adjacentes é comum, especialmente nos modelos estáticos e nos modelos com defasagens distribuídas finitas. Isso faz com que os usuais erros-padrão e estatísticas MQO sejam enganosos (embora os $\hat{\beta}_j$ possam ainda ser não-viesados, ou pelo menos consistentes). Em geral, os erros-padrão MQO subestimam a verdadeira incerteza nas estimativas dos parâmetros.

O modelo mais popular de correlação serial é o AR(1). Usando-o como ponto de partida, é fácil testar a existência de correlação serial AR(1) utilizando os resíduos do MQO. Uma estatística t assintoticamente válida é obtida regredindo os resíduos do MQO sobre os resíduos defasados, assumindo que os regressores são estritamente exógenos e que a hipótese de homoscedasticidade se sustenta. Tornar o teste robusto quanto à heteroscedasticidade é simples. A estatística de Durbin-Watson está disponível sob as hipóteses do modelo linear clássico, mas ela pode levar a um resultado não conclusivo e tem pouco a oferecer sobre o teste t .

Para modelos com uma variável dependente defasada ou outros regressores não-estritamente exógenos, o teste t padrão de \hat{u}_{t-1} continua válido, desde que todas as variáveis independentes sejam incluídas como regressores com \hat{u}_{t-1} . Podemos usar uma estatística F ou LM para verificar a existência de correlação serial de ordem mais alta.

Em modelos com regressores estritamente exógenos, podemos usar um procedimento MQG factível — Cochrane-Orcutt ou Prais-Winsten — para corrigir a correlação serial AR(1). Isso produz estimativas que são diferentes daquelas do MQO: as estimativas MQGF são obtidas a partir do MQO sobre variáveis quase-diferenciadas. Todas as estatísticas de testes habituais da equação transformada são assintoticamente válidas. Quase todos os programas de regressão possuem recursos embutidos para estimar modelos com erros AR(1).

Outra maneira de trabalhar com a correlação serial, especialmente quando a hipótese de exogeneidade estrita possa não se sustentar, é usar o MQO, mas computar erros-padrão robustos em relação à correlação serial (que também são robustos quanto à heteroscedasticidade). Muitos programas de regressão seguem um método sugerido por Newey e West (1987); também é possível usar programas de regressão padrão para obter um erro-padrão por vez.

Finalmente, discutimos algumas características especiais da heteroscedasticidade em modelos de séries temporais. Como no caso do corte transversal, o tipo mais importante de heteroscedasticidade é o que depende das variáveis explicativas; isso é o que determina se as estatísticas MQO usuais são válidas. Os testes de Breusch-Pagan e de White discutidos no Capítulo 8, podem ser aplicados diretamente, com a advertência de que os erros não devem ser serialmente correlacionados. Nos últimos anos, os economistas — especialmente os que estudam o mercado financeiro — vêm demonstrando interesse em formas dinâmicas de heteroscedasticidade. O modelo ARCH é o principal exemplo nesse sentido.

12.1 Quando os erros em um modelo de regressão têm correlação serial AR(1), por que os erros-padrão de MQO tendem a subestimar a variação amostral em $\hat{\beta}_j$? É sempre verdade que os erros-padrão do MQO são muito pequenos?

12.2 Explique o que está errado na seguinte afirmação: “Os métodos de Cochrane-Orcutt e de Prais-Winsten são ambos usados para a obtenção de erros-padrão válidos das estimativas MQO”.

12.3 No Exemplo 10.6, estimamos uma variante do modelo de Fair para prever os resultados das eleições presidenciais nos Estados Unidos.

- (i) Que comentários podemos fazer sobre o fato de o termo erro dessa equação ser serialmente não-correlacionado? (*Sugestão:* Com que frequência são realizadas eleições presidenciais?)
- (ii) Quando regredimos os resíduos do MQO de (10.23) sobre os resíduos defasados, obtemos $\hat{\rho} = -0,068$ e $ep(\hat{\rho}) = 0,240$. O que você conclui sobre a correlação serial em u_t ?
- (iii) O pequeno tamanho da amostra nesta questão é motivo de preocupação quanto a fazer o teste de correlação serial?

12.4 Verdadeiro ou Falso: “Se os erros em um modelo de regressão contiverem ARCH, eles devem ser serialmente correlacionados”.

12.5 (i) No estudo de caso das zonas industriais no Problema 10.11, uma regressão dos resíduos MQO sobre os resíduos defasados produz $\hat{\rho} = 0,841$ e $ep(\hat{\rho}) = 0,053$. Que implicações isso traz para o MQO?

- (ii) Se você quiser usar o MQO, mas também quiser obter um erro-padrão válido para o coeficiente de EZ, o que você fará?

12.6 No Exemplo 12.8, encontramos evidência de heteroscedasticidade em u_t , na equação (12.47). Portanto, computamos os erros-padrão robustos em relação à heteroscedasticidade (em [-]) com os erros-padrão habituais:

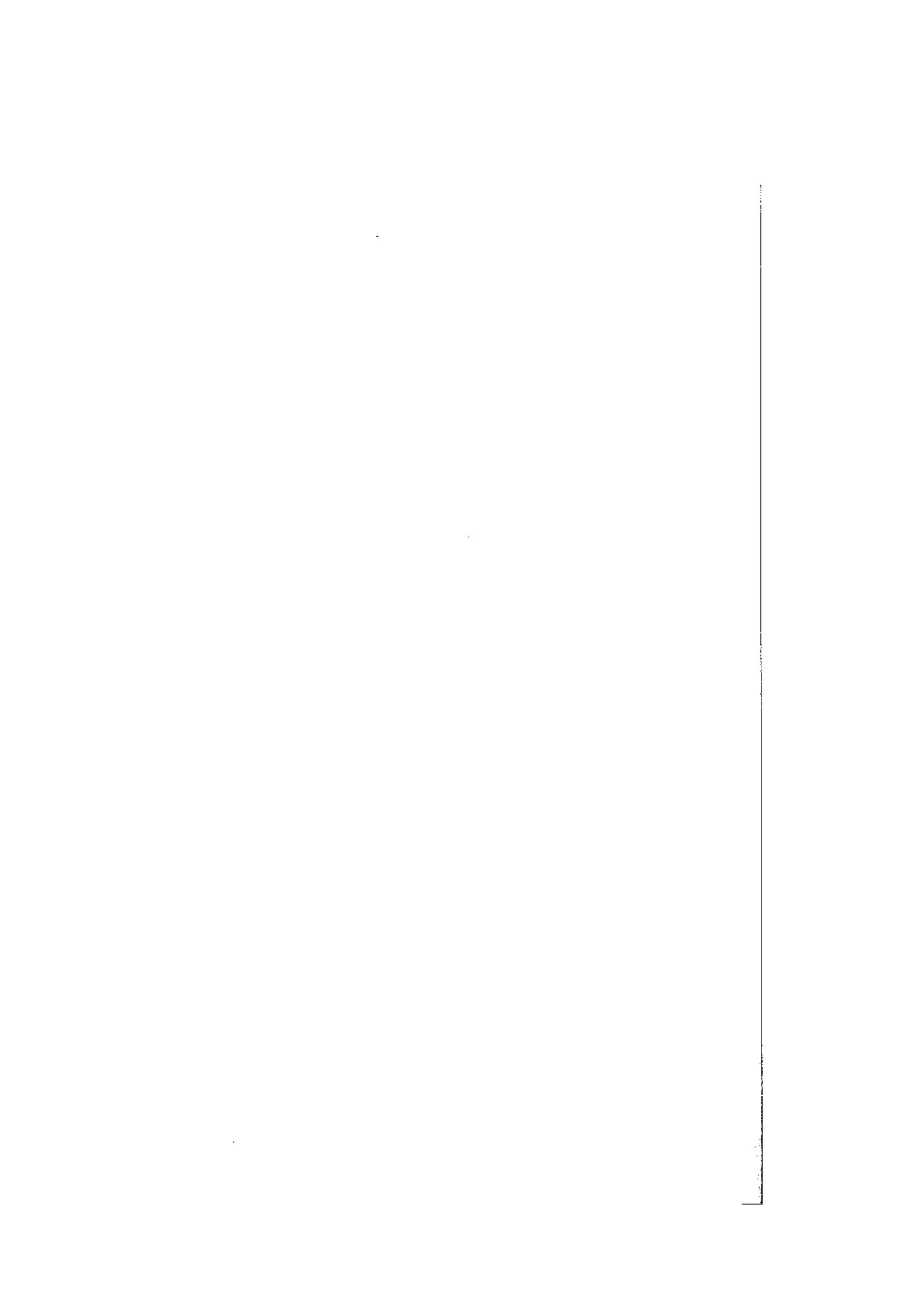
$$\hat{retorno}_t = 0,180 + 0,059 \hat{retorno}_{t-1}$$

(0,081) (0,038)

[0,085] [0,069]

$n = 689, R^2 = 0,0035, \bar{R}^2 = 0,0020.$

O que o uso da estatística t robusta em relação à heteroscedasticidade faz com a significância de $\hat{retorno}_{t-1}$?



Tópicos Avançados

esta parte, nos dedicaremos a alguns tópicos mais especializados que usualmente não são cobertos em cursos introdutórios. Alguns desses tópicos exigem um pouco mais de habilidade matemática que a análise de regressão múltipla das Partes 1 e 2. No Capítulo 13, mostraremos como aplicar a regressão múltipla a agrupamentos independentes de cortes transversais. As questões apresentadas são muito similares à análise padrão do corte transversal, exceto que podemos estudar como as relações mudam ao longo do tempo ao incluirmos variáveis *dummy*. Também ilustramos como dados de painel podem ser analisados com o instrumental da regressão. O Capítulo 14 apresenta métodos de dados de painel mais avançados, utilizados rotineiramente no trabalho aplicado.

Os Capítulos 15 e 16 investigam o problema das variáveis explicativas endógenas. No Capítulo 15, introduzimos o método das variáveis instrumentais como uma maneira de resolver tanto o problema de variáveis omitidas como o de erros de medida. O método de mínimos quadrados de dois estágios é usado frequentemente na economia empírica e é indispensável para estimar modelos de equações simultâneas, tópico que abordaremos no Capítulo 16.

No Capítulo 17, discutem-se alguns tópicos bastante avançados, que são usados na análise de corte transversal, incluindo modelos de variáveis dependentes limitadas e métodos de correção do viés de seleção amostral. O Capítulo 18 volta-se para uma direção diferente, cobrindo alguns avanços recentes da econometria de séries temporais que se vêm mostrando úteis para estinar relações dinâmicas.

O Capítulo 19 tem o objetivo de auxiliar estudantes que devem desenvolver um trabalho de final de curso ou algum artigo no campo das ciências sociais aplicadas. Ao longo do texto serão apresentadas sugestões de como selecionar um tópico, coletar e analisar dados e escrever a versão final do trabalho.

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

té agora, estudamos a análise de regressão múltipla usando dados puros de corte transversal ou de séries temporais. Embora esses dois casos surjam com freqüência 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 de tempo podem, muitas vezes, esclarecer questões importantes de política econômica. Usaremos vários exemplos a esse respeito, 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, em cada ano, podemos extraír uma amostra aleatória de salários por hora, educação, experiência etc., da população de trabalhadores nos Estados Unidos. Ou, em cada ano, podemos extraír 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 *independente*mente. 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 de 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 de 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 de 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 subseqüentes. Isso nos fornecerá dados sobre salários, horas, educação etc. do mesmo grupo de pessoas em anos diferentes.

Conjuntos de dados de 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 de painel; veremos alguns exemplos nas discussões a seguir. Na análise econômica de dados de 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 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 de 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 de 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 de painel mais complexos.

13.1 O AGRUPAMENTO INDEPENDENTE DE CORTES TRANSVERSAIS AO LONGO DO TEMPO

Muitas pesquisas sobre pessoas, famílias e firmas são repetidas a intervalos regulares, muitas vezes a cada ano. Um exemplo é o *Current Population Survey* (Pesquisa Populacional Corrente) — ou CPS —, que de forma aleatória faz pesquisa domiciliar a cada ano. (Veja, por exemplo, o arquivo CPS78_85.RAW, que contém dados das pesquisas de 1978 e 1985.) 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 nesse sentido 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, onde 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 (1994), 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 crianças nascidas 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 em que 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 que 100 mulheres comparáveis em 1972. Como estamos controlando a variável educação, essa queda é separada do declínio da fertilidade devido ao 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 de 46 anos, quando a maior parte das mulheres parou de ter filhos.)

O modelo estimado na Tabela 13.1 assume que o efeito de cada variável explicativa, particularmente a da educação, permanece constante. Isso pode ou não ser verdade; solicitaremos que você explore esse assunto no Problema 13.7.

Finalmente, pode haver heteroscedasticidade no termo 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 dos resíduos quadrados do MQO sobre todas as variáveis independentes da Tabela 13.1, inclusive as *dummies* anuais. (No caso especial da estatística de White, os valores estimados *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).

Ao ler a Tabela 13.1, 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: <i>Kids</i>		
Variáveis Independentes	Coeficientes	Erros-Padrão
<i>educ</i>	-0,128	0,018
<i>idade</i>	0,532	0,138
<i>idade</i> ²	-0,0058	0,0016
<i>negro</i>	1,076	0,174
<i>leste</i>	0,217	0,133
<i>centnorte</i>	0,363	0,121
<i>oeste</i>	0,198	0,167
<i>fazenda</i>	-0,053	0,147
<i>outrural</i>	-0,163	0,175
<i>cidade</i>	0,084	0,124
<i>cidpeq</i>	0,212	0,160
<i>y74</i>	0,268	0,173
<i>y76</i>	-0,097	0,179
<i>y78</i>	-0,069	0,182
<i>y80</i>	-0,071	0,183
<i>y82</i>	-0,522	0,172
<i>y84</i>	-0,545	0,175
<i>constante</i>	-7,742	3,052
<i>n</i> = 1.129		
<i>R</i> ² = 0,1295		
<i>R</i> ² = 0,1162		

EXEMPLO 13.2**{Mudanças no Retorno da Educação e a Diferença Salarial por Gênero}**

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

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

onde 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 um 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 $\log(\text{salárioh})$ 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 $H_1: \delta_5 > 0$. Por simplicidade, assumimos 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 devido à 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 Relatório Econômico Presidencial de 1997, o fator de deflação é $107,6/65,2 \approx 1,65$.) Embora possamos, com facilidade, dividir cada salário de 1985 por 1,65, isso acaba não sendo necessário, dado que uma *dummy* anual de 1985 foi incluída na regressão e $\log(\text{salárioh})$ (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 de 1985 (1,65, se usarmos o IPC). Então, o log do salário real de cada pessoa *i* na amostra de 1985 é

$$\log(\text{salárioh}_i/P85) = \log(\text{salárioh}_i) - \log(P85).$$

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 problema 13.8 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 $\log(\text{salárioh})$ como variável dependente, é importante usar o salário real e incluir uma *dummy* anual.

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

EXEMPLO 13.2 (continuação)

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:

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

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 $0,0185/0,0094 \approx 1,97$, a diferença no retorno da educação é estatisticamente significante ao nível de 5% contra uma alternativa bilateral.

E a diferença salarial por gênero? Em 1978, outros fatores permanecendo iguais, uma mulher ganhava cerca de 31,7% menos que um homem (27,2% é uma estimativa mais precisa). Em 1985, a diferença em $\log(\text{salárioh})$ 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 ao 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 um 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 resíduos quadrados da estimativa 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 (devido à 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 problemas 13.7 e 13.8 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 Problema 7.19. Primeiro, estime o modelo restrito fazendo uma regressão agrupada admitindo diferentes interceptos de tempo; isso produz SQR_w . Depois, compute uma regressão para cada um dos, digamos, T períodos de tempo e obtenha a soma dos resíduos quadrados para cada período de tempo. A soma dos resíduos quadrados sem restrições é obtida como $SQR_{w'} = SQR_1 + SQR_2 + \dots + SQR_T$. 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 $n = n_1 + n_2 + \dots + n_T$ 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: $[(SQR_w - SQR_{w'})/SQR_{w'}][(n - T - Tk)/(T - 1)k]$. Infelizmente, como em qualquer teste F baseado nas somas dos resíduos quadrados 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 DECISÕES GOVERNAMENTAIS COM AGRUPAMENTOS DE CORTES TRANSVERSAIS

Cortes transversais agrupados podem ser muito úteis para a avaliação do impacto de determinado evento ou decisão governamental. O seguinte exemplo 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 razoavelmente 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 pouco tempo após 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). [Nos problemas, 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:

$$rpreço = \gamma_0 + \gamma_1 proxincin + u, \quad (13.3)$$

onde $proxincin$ 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

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

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 aqueles distantes dele sejam os mesmos.

Infelizmente, a equação (13.4) não implica que a localização do incinerador seja 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

$$\begin{aligned} rpreço &= 82.517,23 - 18.824,37 proxincin \\ &\quad (2.653,79) \quad (5.827,71) \\ n &= 179, R^2 = 0,082. \end{aligned} \quad (13.5)$$

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 diferenças-em-diferenças** porque ele pode ser expresso como

$$\hat{\delta}_1 = (\bar{r}preço_{81,pr} - \bar{r}preço_{81,af}) - (\bar{r}preço_{78,pr} - \bar{r}preço_{78,af}), \quad (13.6)$$

onde “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

$$rpreço = \beta_0 + \delta_0 a81 + \beta_1 proxincin + \delta_1 a81 \cdot proxincin + u, \quad (13.7)$$

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 captura 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 devido à 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 devido ao novo incinerador, desde que assumamos 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 13.2Variável Dependente: *rpreço*

Variável Independente	(1)	(2)	(3)
<i>constante</i>	82.517,23 (2.726,91)	89.116,54 (2.406,05)	13.807,67 (11.166,59)
<i>a81</i>	18.790,29 (4.050,07)	21.321,04 (3.443,63)	13.928,48 (2.798,75)
<i>proxincin</i>	-18.824,37 (4.875,32)	9.397,94 (4.812,22)	3.780,34 (4.453,42)

Continua

EXEMPLO 13.3 (continuação)**Tabela 13.2 (conclusão)**

Variável Independente	(1)	(2)	(3)
$a81 \cdot proxincin$	−11.863,90 (7.456,65)	−21.920,27 (6.359,75)	−14.177,93 (4.987,27)
Outros Controles	Não	<i>idade, idade</i> ²	Conjunto Total
Observações <i>R</i> -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; se assim tiver sido, é importante controlar as características que possam ter sido diferentes. Entretanto, tão importante quanto isso, 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 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 $a81 \cdot proxincin$ 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 $a81 \cdot proxincin$ 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) capturam 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 $\log(\text{preço})$ [ou $\log(r\text{preço})$] na análise, de forma a obter um efeito percentual aproximado. O modelo básico torna-se

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

Agora, $100 \cdot \delta_1$ é a redução percentual aproximada nos valores dos imóveis devido ao incinerador. [Assim como no Exemplo 13.2, o uso de $\log(\text{preço})$ em lugar de $\log(r\text{preço})$ apenas afeta o coeficiente de $a81$.] O uso das mesmas 321 observações agrupadas produz

EXEMPLO 13.3 {continuação}

$$\begin{aligned} \log(\hat{\text{preço}}) &= 11,29 + 0,457a8I - 0,340 \text{proxincin} - 0,063 a8I \cdot \text{proxincin} \\ &\quad (0,31) \quad (0,045) \quad (0,055) \quad (0,083) \end{aligned} \quad (13.9)$$

$n = 321, R^2 = 0,409.$

O coeficiente do termo de interação implica que, devido ao 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 $a8I \cdot \text{proxincin}$ 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 — freqüentemente 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

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

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

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

onde a barra significa média, o primeiro subscrito representa o ano e o segundo subscrito representa o grupo. 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 e 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 aqueles que estão sujeitos ao limite antes da mudança da política do governo. Usando amostras aleatórias, tanto do período anterior como do período posterior à mudança, MVD puderam testar se uma remuneração mais generosa faria com que os trabalhadores ficassem mais tempo sem trabalhar (tudo mantido inalterado). Eles iniciaram com uma análise de diferenças-em-diferenças, usando $\log(\text{duração})$ como a variável dependente. Fazendo $apmud$ representar uma variável dummy das observações após a mudança da política e $altrend$, a variável dos trabalhadores de altos rendimentos, a equação estimada, com os erros-padrão entre parênteses, é

$$\begin{aligned} \log(\hat{\text{duração}}) &= 1,126 + 0,0077 \text{ } apmud + 0,256 \text{ } altrend \\ &\quad (0,031) \quad (0,0447) \quad (0,047) \quad \{13.12\} \\ &\quad + 0,191 \text{ } apmud \cdot altrend \\ &\quad (0,069) \\ &n = 5.626, R^2 = 0,021. \end{aligned}$$

Portanto, $\hat{\delta}_1 = 0,191$ ($t = 2,77$), o que implica que o tempo médio em que os trabalhadores de alta renda permaneceram sem trabalhar recebendo indenização trabalhista aumentou cerca de 19% devido ao 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 $\log(\text{duração})$. Isso faz sentido: existem claramente muitos fatores, inclusive a gravidade da lesão, que afetam o tempo 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 Problema 13.10.)

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 vizinhos nos Estados Unidos. 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 houve 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 DE PAINEL DE DOIS PERÍODOS

Retornamos agora à análise do tipo mais simples de dados de 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 computarmos uma regressão simples de txcrim sobre desemp ? Obteremos

$$\begin{aligned} \hat{\text{txcrim}} &= 128,38 - 4,16 \text{desemp} \\ &\quad (20,76) \quad (3,42) \\ n &= 46, R^2 = 0,033. \end{aligned}$$

Se interpretarmos a equação estimada de forma causal, ela implica que um aumento na taxa de desemprego *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 gênero, 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 de 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

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

Na notação y_{it} , i é a pessoa, empresa, cidade etc., e t é o período de tempo. A variável $d2$ 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 dos Estados Unidos 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

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

onde $d87$ 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 nos Estados Unidos, 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 de 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 que o MQO agrupado produza um estimador consistente de β_1 , teremos que assumir o efeito não observado, a_i , como não correlacionado com x_{it} . Podemos ver isso facilmente escrevendo (13.13) como

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

onde $v_{it} = a_i + u_{it}$ é muitas vezes chamado de **erro de composição**. Pelo que conhecemos do MQO, temos que assumir que v_{it} é não correlacionado com x_{it} , onde $t = 1$ ou 2 , para que o MQO estime β_1 (e os outros parâmetros) consistentemente. Isso é verdade independente de usarmos um único corte transversal ou agruparmos os dois cortes transversais. Portanto, mesmo que assumamos que o erro idiossincrático u_{it} seja não-correlacionado com x_{it} , o MQO agrupado será viésado 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.

Suponha que a_i , u_{i1} e u_{i2} tenham média zero e que sejam não-correlacionados dois a dois. Mostre que $\text{Cov}(v_{i1}, v_{i2}) = \text{Var}(a_i)$, de forma que os erros de composição 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:

$$\begin{aligned} \hat{tx\acute{e}rim} &= 93,42 + 7,94 d87 + 0,427 desemp \\ (12,74) &\quad (7,98) \quad (1,188) \end{aligned} \quad (13.16)$$

$n = 92, R^2 = 0,012.$

(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 devido à 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 de painel é levar em conta 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

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

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

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i \quad (13.17)$$

onde “ 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, assumir que Δu_i e Δdesemp_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 Δdesemp_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 *dumy* 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 assumir 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

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

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 que $\Delta \text{desemp} = 0$, espera-se um aumento na taxa de criminalidade (crimes por mil pessoas) de 15,40. Isso reflete um aumento secular 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 de painel de dois anos seja um meio poderoso de controlar efeitos não observados, ele tem um custo. Primeiro, os conjuntos de dados de 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 de painel são mais fáceis de serem obtidos de escolas, cidades, municípios, estados e países.

Mesmo que tenhamos coletado um conjunto de dados de painel, a diferenciação utilizada para eliminar a_i pode reduzir bastante a variação nas variáveis explicativas. Embora x_{it} freqüentemente 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 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 de ano para ano.

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

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

onde 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 $salário_{ih}$ estiver em termos nominais). Como, por definição, a aptidão inata não muda ao longo do tempo, os métodos de dados de painel parecem idealmente apropriados para estimar o retorno da educação. A equação de primeiras diferenças é

$$\Delta \log(salário_{ih}) = \delta_0 + \beta_1 \Delta educ_i + \Delta u_{it}, \quad (13.19)$$

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 primeiras diferenças para estimar o retorno da educação é uma boa idéia, mas ela não funciona muito bem com a maioria dos dados de painel correntemente disponíveis.

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

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + a_i + u_{it}, \quad (13.20)$$

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 de 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, usamos apenas o corte transversal de 1975. O conjunto de dados de painel de 1975 e 1981 tem 239 pessoas, muito menor que o corte transversal de 1975, inclui mais de 700 pessoas. Um modelo de efeitos não observados do total de minutos dormidos por semana é

$$\begin{aligned} \text{dormtot}_{it} = & \beta_0 + \delta_0 d81_t + \beta_1 \text{trabtot}_{it} + \beta_2 \text{educ}_{it} + \beta_3 \text{casado}_{it} \\ & + \beta_4 \text{crianmen}_{it} + \beta_5 \text{boasaude}_{it} + a_i + u_{it}, \quad t = 1,2. \end{aligned}$$

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 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 boasaude 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 . Nossa principal interesse está em β_1 .

A diferenciação ao longo dos dois anos produz a equação estimável

$$\begin{aligned} \Delta \text{dormtot}_i = & \delta_0 + \beta_1 \Delta \text{trabtot}_i + \beta_2 \Delta \text{educ}_i + \beta_3 \Delta \text{casado}_i + \beta_4 \Delta \text{crianmen}_i \\ & + \beta_5 \Delta \text{boasaude}_i + \Delta u_i. \end{aligned}$$

Assumindo 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

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

O coeficiente de $\Delta \text{trabtot}$ 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 ($\sim 6,31$) 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 $\Delta \text{trabtot}$, fornece p -valor = 0,49, o que significa que elas são conjuntamente não significantes a qualquer nível razoável de significância e poderiam ser eliminadas da equação.

O erro-padrão de Δeduc é especialmente grande em relação à estimativa. Esse é o fenômeno descrito anteriormente para a equação do salário. 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 de 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 de 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:

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

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

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

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 de Painel

Quando se usa dados de painel em estudos econômicos, 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 estimativa por diferenciação.

A maioria dos conjuntos de dados de 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 de painel com mais de dois períodos de tempo.

Uma segunda maneira de organizar dois períodos de dados de 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 de 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 de painel com essa estrutura são TRAFFIC1.RAW e VOTE2.RAW. A inconveniência de colocar os dados em um registro é que isso não possibilita 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 de painel com mais de dois períodos de tempo, um caso que consideraremos na Seção 13.5.

13.4 ANÁLISE DE DECISÕES GOVERNAMENTAIS COM DADOS DE PAINEL DE DOIS PERÍODOS

Conjuntos de dados de painel são muito úteis para a análise de decisões governamentais, 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 queremos avaliar o efeito de um programa de treinamento de pessoal de Michigan sobre a produtividade dos trabalhadores de firmas manufatureiras (veja também o Problema 9.8). Façamos ref_{it} representar a taxa de rejeição 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

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

onde $a88$ 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, *grosso modo*, 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 fazer com que os programas de treinamento pareçam eficientes, os administradores 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 mais cedo 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

$$\Delta ref_i = \delta_0 + \beta_1 \Delta subs_i + \Delta u_i. \quad (13.24)$$

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 um 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 estimativa da equação de primeiras diferenças utilizando os dados contidos no arquivo JTRAIN.RAW produz

$$\begin{aligned}\hat{\Delta ref} &= -0,564 \sim 0,739 \Delta subs \\ &\quad (0,405) \quad (0,683) \\ n &= 54, R^2 = 0,022.\end{aligned}$$

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:

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

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 um 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

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

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:

$$\hat{\beta}_1 = \overline{\Delta y}_{it|at} - \overline{\Delta y}_{it|control}. \quad (13.26)$$

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 de painel do estimador de diferenças-em-diferenças na equação (13.11) de dois cortes transversais agrupados. Com dados de 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 devido à 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 Δ_{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 decisão governamental 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 terem 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 seu julgamento. 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 de painel de um período de tempo em que alguns estados tenham adotado novas leis (e alguns estados 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 ($txmорт$). 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.

Usando MQO após a primeira diferenciação produz

$$\begin{aligned} \hat{\Delta}_{txmорт} = & -0,497 - 0,420 \Delta_{abertos} - 0,151 \Delta_{admin} \\ & (0,052) \quad (0,206) \quad (0,117) \end{aligned} \tag{13.27}$$

$n = 51, R^2 = 0,119.$

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%

EXEMPLO 13.7 (continuação)

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.

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

13.5 A DIFERENCIACÃO 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

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

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 idéia 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 $\beta_1, \beta_2, \dots, \beta_k$. 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 viésadas 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 de composição, $v_{it} = a_i + u_{it}$, 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

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

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 por hipótese. Podemos também recorrer a resultados assimptóticos. O requisito importante para que o MQO seja consistente é que Δu_{it} seja 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, $d2$, e $d3$. Para $t = 2$, $\Delta d2_t = 1$ e $\Delta d3_t = 0$; para $t = 3$, $\Delta d2_t = -1$; $\Delta d3_t = 1$. Portanto, (13.30) não contém um intercepto. Isso é inconveniente para certos propósitos, inclusive para o 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 uma única variável *dummy* temporal, normalmente para o terceiro período. Em outras palavras, a equação passa a ser

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

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 para levar em conta as mudanças seculares que não estejam sendo modeladas. Portanto, após fazermos a primeira diferenciação, a equação parecerá com

$$\begin{aligned} \Delta y_{it} = & \alpha_0 + \alpha_3 d3_t + \alpha_4 d4_t + \dots + \alpha_T dT_t + \beta_1 \Delta x_{it1} + \dots \\ & + \beta_k \Delta x_{itk} + \Delta u_{it}, \quad t = 2, 3, \dots, T, \end{aligned} \quad (13.31)$$

onde 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 diferenciação, o arquivo de dados deve consistir de NT registros. Os primeiros T registros são das observações do primeiro corte transversal, ordenados cronologicamente; os segundos T registros são das observações do segundo corte transversal, ordenados cronologicamente, e assim por diante. Depois, computamos

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 falsas 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 assumir 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 assumirmos que os erros idiossincráticos originais, u_{it} , são não-correlacionados ao longo do tempo (uma hipótese que usaremos no Capítulo 14). De fato, se assumirmos 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, computamos uma regressão simples por MQO agrupado de \hat{r}_{it} sobre $\hat{r}_{i,t-1}$, $t = 3, \dots, T$, $i = 1, \dots, N$ 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 iniciarmos 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 um 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 assumindo 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 computadas 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 computados em alguns programas de regressão.

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 computar erros-padrão robustos.

Diferenciar mais de dois anos de dados de painel é muito útil para a análise das decisões de política governamental, como mostrado no exemplo seguinte.

EXEMPLO 13.8

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

Papke (1994) estudou o efeito do programa de instalação de zonas industriais (ZI) no Estado norte-americano de Indiana sobre os pedidos de auxílio-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 da política governamental é

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

onde $uclms_{it}$ é o número de pedidos de auxílio-desemprego registrados durante o ano t na cidade i . O parâmetro θ_i apenas representa um intercepto diferente para cada período de tempo. De forma geral, os pedidos de auxílio-desemprego estavam caindo em âmbito estadual ao longo desse período, e isso deveria estar refletido nos interceptos dos diferentes anos. A variável binária z_{it} é igual a 1 se a cidade i no tempo t era uma zona industrial; estamos interessados em β_1 . O efeito não observado a_i representa fatores fixos que afetam o clima 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 z_{it} e a_i sejam positivamente correlacionados (a_i elevado significa maior número de pedidos de auxílio-desemprego, o que leva a uma maior probabilidade de ser criada uma ZI). Assim, devemos diferenciar a equação para eliminar a_i :

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

A variável dependente nessa equação, a mudança em $\log(uclms_{it})$, é a taxa aproximada de crescimento anual nos pedidos de auxílio-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 auxílio-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). 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_0 , o efeito não observado. (Cornwell e Trumbull utilizam uma transformação diferente, sobre a qual discorremos no Capítulo 14.) Vários fatores, inclusive localização geográfica, atitudes frente à criminalidade, registros históricos e convenções sobre os registros podem estar contidos em a_t . A taxa de criminalidade é o número de crimes por pessoa, $pr pris$ é a probabilidade estimada de prisão, $pr cond$ é a probabilidade de condenação (tendo havido uma prisão), $pr spris$ 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 econométricos sobre a criminalidade, usamos os logs de todas as variáveis com o objetivo de estimar as elasticidades. 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. Os valores entre parênteses são os erros-padrão habituais do MQO; os valores entre colchetes são erros-padrão robustos tanto quanto à correlação serial como quanto à heteroscedasticidade:

$$\begin{aligned}
 \Delta \log(\hat{\lambda} crim) = & 0,008 - 0,100 d83 - 0,048 d84 - 0,005 d85 \\
 & (0,017) (0,024) (0,024) (0,023) \\
 & [0,014] [0,022] [0,020] [0,025] \\
 & + 0,028 d86 + 0,041 d87 - 0,327 \Delta \log(pr pris) \\
 & (0,024) (0,024) (0,030) \\
 & [0,021] [0,024] [0,056] \\
 & - 0,238 \Delta \log(pr cond) - 0,165 \Delta \log(pr spris) \\
 & (0,018) (0,026) \\
 & [0,039] [0,045] \\
 & - 0,022 \Delta \log(sentmed) + 0,398 \Delta \log(polpc) \\
 & (0,022) (0,027) \\
 & [0,025] [0,101]
 \end{aligned} \tag{13.33}$$

$n = 540, R^2 = 0,433, \bar{R}^2 = 0,422.$

As três variáveis de probabilidade — de prisão, de condenação e de cumprir pena prisional — têm o sinal 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 um 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. Primeira, a variável da taxa de criminalidade é calculada com base nos crimes denunciados. Pode

EXEMPLO 13.9 (continuação)

ser que, quando há mais policiais, mais crimes são denunciados. Segunda, 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 produz $F = 75,48$ e $p\text{-valor} = 0,0000$, de modo que existe 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 e são executados por muitos dos programas econometríticos. 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.

Estudamos métodos para analisar conjuntos de cortes transversais agrupados independentemente de dados de 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.) Devido à 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 da política governamental em experimentos naturais.

Os conjuntos de dados de painel estão sendo cada vez mais usados no trabalho aplicado, especialmente na análise da política governamental. Eles são conjuntos de dados nos quais as mesmas unidades de corte transversal são acompanhadas ao longo do tempo. Os conjuntos de dados de 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 assintoticamente 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 devido à diferenciação. Além da homoscedasticidade, devemos assumir que os erros *diferenciados* são serialmente não correlacionados, para podermos aplicar as estatísticas t e F usuais. (O apêndice deste capítulo contém uma lista meticolosa das hipóteses.) Naturalmente, qualquer variável que seja constante ao longo do tempo é eliminada da análise.

EQUAÇÕES DE REGRAÇÃO

13.1 No exemplo 13.1, assuma que a média 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:

$$\log(p\hat{r}e\co) = 11,49 - 0,547 proxincin + 0,394 a81 \cdot proxincin$$

$$(0,26) (0,058) \quad (0,080)$$

$$n = 321, R^2 = 0,220$$

e

$$\log(p\hat{r}e\co) = 11,18 - 0,563 a81 - 0,403 a81 \cdot proxincin$$

$$(0,27) (0,044) \quad (0,067)$$

$$n = 321, R^2 = 0,337.$$

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 de 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 Tabela 3.2.)

13.5 Suponha que queremos estimar o efeito de diversas variáveis sobre a poupança anual e que temos um conjunto de dados de 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 diferenciação, 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 econômétrico você usaria?
-

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).

H I P Ó T E S E P D . 1

Para cada i , o modelo é

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

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

H I P Ó T E S E P D . 2

Temos uma amostra aleatória do corte transversal.

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_{ij} , $t = 1, \dots, T$, $j = 1, \dots, k$.

H I P Ó T E S E P D . 3

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

Quando a hipótese PD.3 se mantém, algumas vezes dizemos que os x_{ij} 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_{ij} e o erro remanescente, u_{it} , para todo s e t . Uma implicação importante da hipótese PD.3 é que $E(\Delta u_{it}|X_i) = 0$, $t = 2, \dots, T$.

H I P Ó T E S E P D . 4

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

Sob essas quatro primeiras hipóteses, os estimadores de primeiras diferenças são não viesados. A hipótese crucial é PD.3, 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 $N \rightarrow \infty$ (e talvez de forma mais genérica).

H I P Ó T E S E P D . 5

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

H I P Ó T E S E P D . 6

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

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).

H I P Ó T E S E P D . 7

Condicional a \mathbf{X}_t , 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.

Métodos Avançados de Dados de Painel

este capítulo, tratamos de dois métodos para estimar modelos de efeitos não observados de dados de painel que são pelo menos tão comuns quanto a primeira diferenciação. Embora esses métodos sejam um pouco mais difíceis de serem descritos e implementados, vários programas econômétricos os suportam.

Na Seção 14.1, discutiremos o estimador de efeitos fixos que, como a primeira diferenciação, usa uma transformação para remover o efeito não observado a_i antes da estimação. Quaisquer variáveis explicativas constantes no tempo são reinvocadas com a_i .

O estimador de efeitos aleatórios é atraente quando pensamos que o efeito não observado é não correlacionado com todas as variáveis explicativas. Se tivermos bons controles em nossa equação, podemos crer que qualquer resto de heterogeneidade que tenha sido negligenciada induz correlação serial somente no termo erro de composição, mas não causa correlação entre os erros de composição e as variáveis explicativas. A estimação de modelos de efeitos aleatórios por mínimos quadrados generalizados é bastante fácil e normalmente feita por muitos programas econômétricos.

Na Seção 14.3, mostraremos como os métodos de dados de painel podem ser aplicados em outras estruturas de dados, inclusive em amostras de pares comparados e de aglomeração.

14.1 ESTIMAÇÃO DE EFEITOS FIXOS

A primeira diferenciação é apenas uma das muitas maneiras de eliminar o efeito fixo, a_i . Um método alternativo que funciona melhor sob certas hipóteses é chamado **transformação de efeitos fixos**. Para verificar o que este método envolve, considere um modelo com uma única variável explicativa; para cada i ,

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2, \dots, T. \quad (14.1)$$

Agora, para cada i , calculamos a média dessa equação ao longo do tempo. Obtemos

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i, \quad (14.2)$$

onde $\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}$, e assim por diante. Como a_i é fixo ao longo do tempo, ele aparece tanto em (14.1) como em (14.2). Se subtrairmos (14.2) de (14.1), para cada t , obtemos

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i, t = 1, 2, \dots, T,$$

ou

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{u}_{it}, t = 1, 2, \dots, T, \quad (14.3)$$

onde $\ddot{y}_{it} = y_{it} - \bar{y}_i$ são os **dados temporais reduzidos** de y e, de maneira análoga, \ddot{x}_{it} e \ddot{u}_{it} . A transformação de efeitos fixos também é chamada de **transformação interna**. O importante sobre a equação (14.3) é que o efeito não observado, a_i , desapareceu. Isso sugere que deveríamos estimar (14.3) pelo MQO agrupado. Um estimador MQO agrupado que seja baseado em variáveis temporais reduzidas é chamado de **estimador de efeitos fixos ou estimador interno**. Este último nome vem do fato de o MQO em (14.3) usar a variação de tempo em y e x *dentro* de cada observação do corte transversal.

O *estimador* que usa a variação temporal *entre* as observações do corte transversal é obtido da mesma forma que o estimador MQO na equação de corte transversal (14.2) (na qual incluímos um intercepto, β_0): utilizamos as médias de tempo tanto de y como de x e depois computamos uma regressão de corte transversal. Não estudaremos em detalhes esse estimador porque ele é viesado quando a_i é correlacionado com \bar{x}_i (veja Problema 14.2). Se entendermos que a_i é não correlacionado com x_{it} , é melhor usarmos o estimador de efeitos aleatórios, que estudamos na Seção 14.2. O estimador que usa a variação entre as observações ignora informações importantes sobre como as variáveis mudam ao longo do tempo.

A adição de mais variáveis explicativas à equação provoca poucas alterações. O modelo original é

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + a_i + u_{it}, t = 1, 2, \dots, T. \quad (14.4)$$

Simplesmente usamos a redução temporal de cada variável explicativa — inclusive *dummies* de períodos de tempo — e, em seguida, fazemos uma regressão pelo MQO agrupado utilizando todas as variáveis que sofreram redução temporal. A equação de redução temporal geral para cada i é

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it1} + \beta_2 \ddot{x}_{it2} + \dots + \beta_k \ddot{x}_{itk} + \ddot{u}_{it}, t = 1, 2, \dots, T, \quad (14.5)$$

que estimamos pelo MQO agrupado.

Sob uma hipótese de exogeneidade estrita das variáveis explicativas, o estimador de efeitos fixos é não viesado: grosso modo, o erro idiossincrático u_{it} deve ser não correlacionado com cada variável explicativa ao longo de *todos* os períodos de tempo. (Veja o apêndice deste capítulo para definições precisas das hipóteses.) O estimador de efeitos fixos leva em conta uma correlação arbitrária entre a_i e as variáveis explicativas em qualquer período de tempo, como na primeira diferenciação. Por esse motivo, qualquer variável explicativa que seja constante ao longo do tempo para todo i é removida pela transformação de efeitos fixos: $\ddot{x}_{it} = 0$ para todo i e t , se x_{it} for constante ao longo de t . Portanto, não podemos incluir variáveis tais como gênero ou distância de uma cidade até um rio.

As outras hipóteses para que uma análise direta do MQO seja válida são as de que os erros u_{it} sejam homoscedásticos e que sejam serialmente não-correlacionados (ao longo de t); veja o apêndice deste capítulo.

Existe um ponto sutil na determinação dos graus de liberdade do estimador de efeitos fixos. Quando estimamos a equação com redução temporal (14.5) por MQO agrupado, temos um total de NT observações e k variáveis independentes. [Observe que não há intercepto em (14.5); ele é eliminado pela transformação de efeitos fixos.] Portanto, deveríamos aparentemente ter $NT - k$ graus de liberdade. Esse cálculo é incorreto. Para cada observação i do corte transversal, perdemos um gl devido à redu-

ção temporal. Em outras palavras, para cada i , os erros reduzidos \hat{u}_{it} resultam em zero quando somados ao longo de t , de modo que perdemos um grau de liberdade. (Não existe tal restrição nos erros idiossincráticos u_{it} originais.) Portanto, os graus de liberdade apropriados são $gl = NT - N - k = N(T - 1) - k$. Felizmente, os programas de regressão modernos que possuem recursos de estimativa de efeitos fixos computam corretamente os gl . Entretanto, se tivermos que fazer, por nós mesmos, a redução temporal e a estimativa pelo MQO agrupado, precisaremos corrigir os erros-padrão e as estatísticas de testes.

- Suponha que em uma equação de poupança familiar, dos anos de 1990, 1991 e 1992, definimos $kids_{it}$ como o número de crianças na família i no ano t . Se o número de crianças for constante ao longo desse período de três anos na maioria das famílias na amostra, que problemas pode isso causar na estimativa do efeito que o número de crianças tem sobre a poupança?

EXEMPLO 14.1

(Efeito do Treinamento de Pessoal sobre as Taxas de Refugos de Produtos das Empresas)

Utilizamos os dados de três anos, 1987, 1988 e 1989, das 54 empresas que informaram suas taxas de refugos em cada ano. Nenhuma das empresas havia recebido subsídio de treinamento antes de 1988; em 1988, 19 empresas receberam subsídios; em 1989, 10 empresas diferentes receberam subsídios. Portanto, também devemos considerar a possibilidade de que o treinamento adicional de pessoal em 1988 tenha tornado os trabalhadores mais produtivos em 1989. Isso é feito com facilidade com a inclusão de um valor defasado do indicador de subsídios. Também incluímos dummies anuais para 1988 e 1989. Os resultados são apresentados na Tabela 14.1:

Tabela 14.1

Estimação de Efeitos Fixos da Equação da Taxa de Refugo

Variável Dependente: $\log(\text{ref})$	
Variáveis Independentes	
$d88$	-0,080 (0,109)
$d89$	-0,247 (0,133)
subs	-0,252 (0,151)
subs_{-1}	-0,422 (0,210)
Observações	162
Graus de Liberdade	104
R -quadrado	0,201

EXEMPLO 14.1 (continuação)

Descrevemos os resultados de uma maneira que enfatiza a necessidade de interpretar as estimativas à luz do modelo de efeitos não observados, (14.4). Estamos controlando explicitamente os efeitos não observados, constantes no tempo, em a_i . A redução temporal nos possibilita estimar β_j , mas (14.5) não é a melhor equação para interpretar as estimativas.

Curiosamente, o efeito defasado estimado do subsídio de treinamento é substancialmente maior do que o efeito contemporâneo: o treinamento de pessoal produz efeito pelo menos um ano mais tarde. Como a variável dependente está na forma logarítmica, prevê-se que a obtenção de um subsídio em 1988 reduz a taxa de refugo da empresa em 1989 em cerca de 34,4% [$\exp(-0,422) - 1 \approx -0,344$]; o coeficiente de $subs_{-1}$ é significante ao nível de 5% contra uma alternativa bilateral. O coeficiente de $subs$ é significante ao nível de 10%, e o tamanho do coeficiente não é nada desprezível. Observe que os g_l são obtidos como $N(T - 1) - k = 54(3 - 1) - 4 = 104$.

O coeficiente de $d89$ indica que a taxa de refugo foi substancialmente menor em 1989 do que no ano-base, 1987, mesmo na ausência de subsídios de treinamento de pessoal. Assim, é importante considerarmos esses efeitos agregados. Se tivéssemos omitido as *dummies* anuais, o aumento secular da produtividade do trabalhador seria atribuído aos subsídios de treinamento de pessoal. A Tabela 14.1 mostra que, mesmo após termos controlado as tendências agregadas na produtividade, os subsídios de treinamento de pessoal tiveram um grande efeito estimado.

Finalmente, é fundamental considerar o efeito defasado no modelo. Se omitirmos $subs_{-1}$, estaremos assumindo que o efeito do treinamento de pessoal não durará até o próximo ano. A estimativa de $subs$, quando eliminamos $subs_{-1}$, é $-0,082$ ($t = -0,65$); esse número é muito menor e estatisticamente não significante.

De acordo com o programa do estado norte-americano de Michigan, se uma empresa recebeu um subsídio em determinado ano, ela não se qualificará para um subsídio no ano seguinte. O que isso sugere sobre a correlação entre $subs$ e $subs_{-1}$?

Ao estimarmos um modelo de efeitos não observados por efeitos fixos, não é claro como devemos computar um indicador do grau de ajuste. O R -quadrado dado na Tabela 14.1 é baseado na transformação interna: ele é o R -quadrado obtido da estimativa de (14.5). Assim, ele é interpretado como o montante da variação temporal em y_{it} , que é explicada pela variação temporal nas variáveis explicativas. São possíveis outras maneiras de calcular o R -quadrado, uma das quais discutiremos mais tarde.

Embora variáveis constantes no tempo não possam ser incluídas por si mesmas em um modelo de efeitos fixos, elas *podem* interagir com variáveis que mudam ao longo do tempo e, particularmente, com variáveis *dummy* anuais. Por exemplo, em uma equação de salários na qual a educação é constante ao longo do tempo para cada indivíduo em nossa amostra, podemos interagir a educação com cada *dummy* anual para verificar como o retorno da educação mudou ao longo do tempo. Porém, não podemos usar efeitos fixos para estimar o retorno da educação no período base — o que significa que não podemos estimar o retorno da educação em qualquer período — somente podemos ver como o retorno da educação em cada ano difere do contido no período base.

Ao incluirmos um conjunto total de *dummies* anuais — isto é, *dummies* anuais para todos os anos, exceto o primeiro —, não podemos estimar o efeito de qualquer variável cuja *mudança* ao longo do tempo seja constante. Um exemplo são os anos de experiência em um conjunto de dados de painel, no

qual cada pessoa trabalha em todos os anos, de forma que a experiência sempre aumenta em uma unidade, a cada ano, para cada pessoa na amostra. A presença de a_i explica as diferenças entre as pessoas em seus anos de experiência no período de tempo inicial. Entretanto, aí não pode o efeito do aumento de um ano de experiência ser distinguido dos efeitos temporais agregados (porque a experiência aumenta na mesma quantidade para todos). Isso também seria verdade se, em lugar de *dummies* anuais separadas, usássemos uma tendência temporal linear: para cada pessoa, a experiência não pode ser distinguida de uma tendência linear.

EXEMPLO 14.2

(O Retorno da Educação Mudou no Transcorrer do Tempo?)

Os dados contidos no arquivo WAGEPAN.RAW são de Vella e Verbeek (1998). Cada um dos 545 homens na amostra trabalhou em todos os anos de 1980 a 1987. Algumas variáveis no conjunto de dados mudam ao longo do tempo: experiência, estado civil e filiação sindical são as três mais importantes. Outras variáveis não mudam: raça e educação são os principais exemplos. Se usarmos efeitos fixos (ou primeira diferenciação), não poderemos incluir raça, educação ou experiência na equação. Porém, podemos incluir interações de *educ* com *dummies* anuais para 1981 a 1987, para testar se o retorno da educação foi constante ao longo desse período de tempo. Usamos *log(salário)* como a variável dependente, um termo quadrático de experiência, variáveis *dummy* para estado civil e filiação sindical, um conjunto completo de *dummies* anuais e os termos de interação *d81·educ*, *d82·educ*, ..., *d87·educ*.

As estimativas desses termos de interação são todas positivas, e geralmente elas ficam maiores para os anos mais recentes. O maior coeficiente (0,030) é o de *d87·educ*, com $t = 2,48$. Em outras palavras, estima-se que o retorno da educação seja cerca de três pontos percentuais maior em 1987 do que no ano-base, 1980. (Não temos uma estimativa do retorno da educação no ano-base pelos motivos apresentados anteriormente.) O outro termo de interação significante é *d86·educ* (coeficiente = 0,027, $t = 2,23$). As estimativas dos primeiros anos são menores e não significantes ao nível de 5% contra uma alternativa bilateral. Se fizermos um teste *F* conjunto, da significância de todos os sete termos de interação, obteremos p -valor = 0,28: isso dá um exemplo de como um conjunto de variáveis é conjuntamente não significante, embora algumas variáveis sejam, individualmente, significantes. [Os *g*'s do teste *F* são 7 e 3.799; o segundo número vem de $N(T - 1) - k = 545(8 - 1) - 16 = 3.799$.] Geralmente, os resultados são consistentes com um aumento no retorno da educação ao longo do período.

A Regressão das Variáveis Dummy

Uma visão tradicional do modelo de efeitos fixos é assumir que o efeito não observado, a_i , é o parâmetro a ser estimado para cada i . Assim, na equação (14.4), a_i é o intercepto para a pessoa i (ou empresa i , cidade i etc.) que tem de ser estimado com β_j . (Claramente, não podemos fazer isso com um único corte transversal: haveria $N + k$ parâmetros a serem estimados com somente N observações. Precisamos, no mínimo, de dois períodos de tempo.) A maneira de estimarmos um intercepto para cada i é introduzir uma variável *dummy* para cada observação do corte transversal, juntamente com as variáveis explicativas (e provavelmente variáveis *dummy* para cada período de tempo). Esse método é habitualmente chamado de **regressão de variáveis dummy**. Mesmo quando N não é muito grande (digamos, $N = 54$, como no exemplo 14.1), isso resultará em muitas variáveis explicativas — na maioria dos casos em quantidade excessiva para explicitamente levar a cabo a regressão.

Dessa forma, o método das variáveis *dummy* não é muito prático para conjuntos de dados de painel com muitas observações de corte transversal.

No entanto, a regressão das variáveis *dummy* tem algumas características interessantes. A mais importante é que ela nos fornece *exatamente* as mesmas estimativas de β_j que obteríamos da regressão dos dados temporais reduzidos, e os erros-padrão, bem como outras estatísticas importantes, são idênticos. Portanto, o estimador de efeitos fixos pode ser obtido com a regressão das variáveis *dummy*. Uma das vantagens dessa regressão é que ela calcula diretamente, e de maneira apropriada, os graus de liberdade. Hoje em dia, essa é uma vantagem menor, já que muitos programas econométricos possuem opções programadas de efeitos fixos.

O R -quadrado da regressão das variáveis *dummy* normalmente é bastante elevado. Isso ocorre devido ao fato de estarmos incluindo uma variável *dummy* para cada unidade de corte transversal, o que explica muito da variação nos dados. Por exemplo, se estimarmos o modelo de efeitos não observados no Exemplo 13.8 por efeitos fixos usando a regressão das variáveis *dummy* (o que é possível com $N = 22$), então, $R^2 = 0,933$. Não devemos nos empolgar com esse grande R -quadrado: não surpreende que possamos explicar muito da variação nos pedidos de auxílio-desemprego usando *dummies* tanto para ano como para cidade. Como no Exemplo 13.8, a estimativa da variável *dummy ZI* é mais importante que o R^2 .

O R -quadrado da regressão das variáveis *dummy* pode ser utilizado para computar os testes F da maneira habitual, assumindo, é claro, que as hipóteses do modelo linear clássico se mantêm (veja o apêndice deste capítulo). Particularmente, podemos testar a significância conjunta de todas as *dummies* do corte transversal ($N - 1$, já que uma unidade é selecionada como grupo base). O R -quadrado restrito é obtido da regressão com todas as *dummies* do corte transversal; o R -quadrado restrito omite essas variáveis. Na vasta maioria das aplicações, as variáveis *dummy* serão conjuntamente significantes.

Ocasionalmente, os interceptos estimados, digamos \hat{a}_i , são de interesse. Esse é o caso se quisermos estudar a distribuição de \hat{a}_i ao longo de i , ou se quisermos selecionar uma empresa ou cidade em particular para verificar se \hat{a}_i está acima ou abaixo do valor médio na amostra. Essas estimativas são disponibilizadas diretamente pela regressão das variáveis *dummy*, mas raramente são descritas pelos programas que possuem rotinas de efeitos fixos (pela razão prática de existirem muitos \hat{a}_i). Após a estimação dos efeitos fixos com N de qualquer tamanho, os \hat{a}_i serão calculados facilmente:

$$\hat{a}_i = \bar{y}_i - \hat{\beta}_1 \bar{x}_{i1} - \dots - \hat{\beta}_k \bar{x}_{ik}, i = 1, \dots, N, \quad (14.6)$$

onde a barra superior refere-se às médias temporais e os $\hat{\beta}_j$ são as estimativas dos efeitos fixos. Por exemplo, se estimarmos um modelo da criminalidade controlando vários fatores de variação temporal, poderemos obter \hat{a}_i para uma cidade, para verificar se os efeitos fixos não observados que contribuem para a criminalidade estão acima ou abaixo da média.

Alguns programas econométricos que suportam a estimação de efeitos fixos registram um “intercepto” que pode causar confusão em vista de nossa afirmação anterior de que a redução temporal elimina todas as variáveis constantes no tempo, inclusive um intercepto global. [Veja a equação (14.5).] A descrição de um intercepto global na estimação de efeitos fixos surge de vermos a_i como um parâmetro a ser estimado. Em geral, o intercepto informado é a média, ao longo de i , de \hat{a}_i . Em outras palavras, o intercepto global é, na realidade, a média dos interceptos individuais específicos. Outra possibilidade seria escolher uma unidade do corte transversal para ser o grupo base, mas tal opção seria geralmente arbitrária e habitualmente não muito interessante.

Na maioria dos estudos, os $\hat{\beta}_j$ são de interesse, e assim as equações de dados temporais reduzidos são usadas para obter essas estimativas. Além disso, usualmente é melhor vermos os a_i como variáveis

omitidas que controlamos por meio da transformação interna. A acepção na qual a_i pode ser estimado geralmente é fraca. De fato, embora \hat{a}_i seja não-viesado (sob as hipóteses EF.1 a EF.4 do apêndice deste capítulo), ele não é consistente com um T fixo e $N \rightarrow \infty$. A razão é que, a cada observação de corte transversal que adicionamos, adicionamos também um novo a_i . Nenhuma informação se acumula em cada a_i quando T é fixo. Com T maior, podemos obter melhores estimativas de a_i , mas a maioria dos conjuntos de dados de painel são da variedade de N grande e T pequeno.

Efeitos Fixos ou Primeira Diferenciação?

Até agora, vimos dois métodos para estimar modelos de efeitos não-observados. Um deles envolve a diferenciação dos dados e o outro a redução temporal. Como saber qual deles usar?

Poderemos eliminar um caso imediatamente: quando $T = 2$, as estimativas EF e PD e todas as estatísticas de teste são *idênticas*, e portanto não importa qual usamos. A primeira diferenciação tem a vantagem de ser feita de maneira direta em praticamente todos os programas econométricos, e é fácil computar estatísticas robustas em relação à heteroscedasticidade na regressão de PD.

Quando $T \geq 3$, os estimadores EF e PD não são os mesmos. Como ambos são não-viesados sob as hipóteses EF.1 a EF.4, não podemos usar a inexistência de viés como um critério. Além disso, ambos são consistentes (com T fixo e $N \rightarrow \infty$) sob EF.1 a EF.4. Para N grande e T pequeno, a escolha entre EF e PD dependerá da eficiência relativa dos estimadores, e isso é determinado pela correlação serial nos erros idiossincráticos, u_{it} . (Assumiremos a homoscedasticidade de u_{it} , visto que comparações de eficiência exigem erros homoscedásticos.)

Quando os u_{it} são serialmente não-correlacionados, os efeitos fixos são mais eficientes que a primeira diferenciação (e os erros-padrão informados pelos efeitos fixos são válidos). Como o modelo de efeitos não-observados é em geral definido (algumas vezes somente de maneira implícita) com erros idiossincráticos serialmente não-correlacionados, o estimador EF é mais usado que o estimador PD. Entretanto, devemos nos lembrar que essa hipótese pode ser falsa. Em muitas aplicações, podemos esperar que os fatores não-observados que se alteram ao longo do tempo sejam serialmente correlacionados. Se u_{it} seguir um passeio aleatório — há uma correlação serial bastante substancial e positiva —, a diferença Δu_{it} será serialmente não-correlacionada, e a primeira diferenciação será melhor. Em muitos casos, os u_{it} exibem alguma correlação serial positiva, mas talvez nem tanto quanto um passeio aleatório. Assim, não podemos comparar facilmente a eficiência dos estimadores EF e PD.

É difícil testar se os u_{it} são serialmente não-correlacionados após a estimativa EF: podemos estimar os erros temporais reduzidos, \tilde{u}_{it} , mas não os u_{it} . Porém, na Seção 13.3, mostramos como verificar se os erros diferenciados, Δu_{it} , são serialmente não-correlacionados. Se parecer ser esse o caso, é possível usar PD. Se houver correlação serial negativa substancial em Δu_{it} , EF provavelmente será melhor. Sempre é bom tentar ambos: tanto melhor se os resultados não forem confiáveis.

Quando T é grande, e especialmente quando N não é muito grande (por exemplo, $N = 20$ e $T = 30$), devemos ter cuidado ao usar o estimador de efeitos fixos. Embora resultados distribucionais exatos permaneçam para qualquer N e T sob as hipóteses de efeitos fixos clássicas, a inferência pode ser bastante sensível a violações das hipóteses quando N é pequeno e T é grande. Particularmente, se estivermos usando processos de raiz unitária — veja o Capítulo 11 — o problema da regressão espúria pode surgir. A primeira diferenciação tem a vantagem de transformar um processo integrado de séries temporais em um processo fracamente dependente. Portanto, se aplicarmos a primeira diferenciação, poderemos recorrer ao teorema do limite central, mesmo nos casos onde T é maior que N . Não é necessária a normalidade nos erros idiossincráticos, e a heteroscedasticidade e a correlação serial podem ser tratadas da forma que mencionamos no Capítulo 13. A inferência com o estimador de efeitos fixos é potencialmente mais sensível à não-normalidade, à heteroscedasticidade e à correlação serial nos erros idiossincráticos.

De outro lado, os efeitos fixos acabam sendo menos sensíveis à violação da hipótese de exogeneidade estrita, especialmente com T grande. Alguns autores até recomendam estimar modelos de efeitos fixos com variáveis dependentes defasadas (o que claramente infringe a hipótese EF.3 no apêndice deste capítulo). Quando os processos são fracamente dependentes ao longo do tempo e T é grande, o viés no estimador de efeitos fixos pode ser pequeno [veja, por exemplo, Wooldridge (2002, Capítulo 11)].

É difícil escolher entre EF e PD quando eles produzem resultados substancialmente diferentes. Faz sentido descrever ambos os conjuntos de resultados e tentar determinar por que eles diferem.

Efeitos Fixos com Painéis Não Equilibrados

Alguns conjuntos de dados de painel, especialmente de pessoas ou empresas, não apresentam alguns anos em pelo menos algumas unidades do corte transversal na amostra. Nesse caso, chamamos o conjunto de dados de **painel não equilibrado**. A mecânica de estimação dos efeitos fixos com um painel não equilibrado não é muito mais difícil que com um painel equilibrado. Se T_i for o número de períodos de tempo da unidade i do corte transversal, simplesmente usamos essas T_i observações para fazer a redução temporal. O número total de observações será, então, $T_1 + T_2 + \dots + T_N$. Como no caso equilibrado, um grau de liberdade será perdido em cada observação de corte transversal devido à redução temporal. Qualquer programa de regressão que faça efeitos fixos faz os ajustes apropriados para essa perda. A regressão das variáveis *dummy* também é feita exatamente da mesma maneira como no caso do painel equilibrado, e os *gl* são apropriadamente obtidos.

É fácil notar que as unidades que possuem somente um único período de tempo não têm participação em uma análise de efeitos fixos. A redução temporal de tais observações resulta em zeros, que não são usados na estimação. (Se T_i for dois, no máximo, para todo i , poderemos usar a primeira diferenciação: se $T_i = 1$ para qualquer i , não teremos dois períodos para diferenciar.)

O problema mais difícil com um painel não equilibrado é determinar a razão de ele não ser equilibrado. Com cidades e estados, por exemplo, algumas vezes os dados de variáveis importantes faltam para certos anos. Desde que a razão da falta de dados de algum i não seja correlacionada com os erros idiossincráticos, u_{it} , o painel não equilibrado não causará problemas. Quando temos dados sobre pessoas, famílias ou empresas, torna-se mais complicado. Imagine, por exemplo, que obtenhamos uma amostra aleatória de indústrias em 1990, e que estamos interessados em verificar como a sindicalização afeta a lucratividade das empresas. De maneira ideal, podemos usar uma análise de dados de painel para controlarmos as características não observadas dos trabalhadores e da administração que afetam a lucratividade e que possam também estar correlacionadas com a fração da força de trabalho da empresa que seja sindicalizada. Se coletarmos os dados em anos subsequentes, algumas empresas podem ser perdidas por terem encerrado suas atividades ou porque foram incorporadas por outras empresas. Se assim for, provavelmente teremos uma amostra não-aleatória nos períodos de tempo subsequentes. A questão é: se aplicarmos efeitos fixos ao painel não equilibrado, quando os estimadores serão não-viesados (ou pelo menos consistentes)?

Se a razão pela qual uma empresa deixa a amostra (ou seja, há uma redução da amostra) for correlacionada com o erro idiossincrático — aqueles fatores não observados que mudam ao longo do tempo e afetam os lucros —, então, o problema resultante dessa redução da amostra (veja o Capítulo 9) pode levar a estimadores viesados. Essa é uma consideração bastante séria nesse exemplo. No entanto, uma característica de grande importância sobre a análise de efeitos fixos é que ela *permite* que a redução da amostra seja correlacionada com a_i , o efeito não observado. A ideia é que, com a amostragem inicial, algumas unidades terão maior probabilidade de serem eliminadas da pesquisa, e isso é capturado por a_i .

EXEMPLO 14.3**(Efeito do Treinamento de Pessoal sobre as Taxas de Refugo de Produtos das Empresas)**

Adicionamos duas variáveis à análise na Tabela 14.1: $\log(vendas_{it})$ e $\log(empreg_{it})$, em que $vendas$ representa as vendas anuais das empresas e $empreg$ é o número de empregados. Três das 54 firmas são inteiramente eliminadas da análise por não possuírem dados sobre vendas ou emprego. Cinco observações adicionais são perdidas devido à falta de dados em uma ou em ambas dessas variáveis para alguns anos, deixando-nos com $n = 148$. O uso de efeitos fixos no painel não equilibrado não altera a situação básica, embora o efeito estimado dos subsídios fique maior: $\hat{\beta}_{subs} = -0,297$, $t_{subs} = -1,89$; $\hat{\beta}_{subs-1} = -0,536$, $t_{subs-1} = -2,389$.

A solução de problemas gerais da redução da amostra em dados de painel é complicada e está além do escopo deste texto. [Veja, por exemplo, Wooldridge (2002, Capítulo 17).]

14.2 MODELOS DE EFEITOS ALEATÓRIOS

Começamos com o mesmo modelo de efeitos não observados, como anteriormente,

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + a_i + u_{it}, \quad (14.7)$$

onde explicitamente incluímos um intercepto de forma a podermos assumir que o efeito não observado, a_i , tem média zero (sem perda de generalidade). Normalmente, consideraremos também *dummies* temporais entre as variáveis explicativas. Ao usar efeitos fixos ou primeira diferenciação, a meta é eliminar a_i , porque ele supostamente estará correlacionado com um ou mais dos x_{ij} . Mas suponha que entendemos a_i como não-correlacionado com cada variável explicativa em todos os períodos de tempo. Nesse caso, o uso de uma transformação para eliminar a_i resultará em estimadores ineficientes.

A equação (14.7) torna-se um **modelo de efeitos aleatórios** quando assumimos que o efeito não observado a_i é não-correlacionado com cada variável explicativa:

$$\text{Cov}(x_{it}, a_i) = 0, t = 1, 2, \dots, T; j = 1, 2, \dots, k. \quad (14.8)$$

De fato, as hipóteses de efeitos aleatórios ideais incluem todas as hipóteses de efeitos fixos mais o requisito adicional de que a_i seja independente de todas as variáveis explicativas, em todos os períodos de tempo. (Veja o apêndice deste capítulo sobre as hipóteses efetivamente usadas.) Se entendermos que o efeito não observado a_i seja correlacionado com qualquer das variáveis explicativas, devaremos usar a primeira diferenciação ou os efeitos fixos.

Sob (14.8) e juntamente com as hipóteses dos efeitos aleatórios, como devemos estimar os β_j ? É importante ver que, se acreditarmos que a_i seja não-correlacionado com as variáveis explicativas, os β_j podem ser consistentemente estimados com o uso de um único corte transversal: não precisamos dos dados de painel. Entretanto, o uso de um único corte transversal desconsidera muitas informações importantes de outros períodos de tempo. Também podemos usar os dados em um procedimento de MQO agrupado: compute o MQO de y_{it} sobre as variáveis explicativas e provavelmente sobre as *dummies* temporais. Isso também produz estimadores consistentes dos β_j sob a hipótese de efeitos aleatórios.

Porém, isso ignora uma característica fundamental do modelo. Se definirmos o **termo erro de composição** como $v_{it} = a_i + u_{it}$, (14.7) pode ser escrita como

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + v_{it}. \quad (14.9)$$

Como a_i é o erro de composição em cada período de tempo, os v_{it} são serialmente correlacionados ao longo do tempo. De fato, sob as hipóteses de efeitos aleatórios,

$$\text{Corr}(v_{it}, v_{is}) = \sigma_a^2 / (\sigma_a^2 + \sigma_u^2), \quad t \neq s,$$

onde $\sigma_a^2 = \text{Var}(a_i)$ e $\sigma_u^2 = \text{Var}(u_{it})$. Essa correlação serial (necessariamente) positiva no termo erro pode ser substancial: como os habituais erros-padrão do MQO agrupado ignoram essa correlação, eles serão incorretos, como também serão incorretas as habituais estatísticas de testes. No Capítulo 12, mostramos como os mínimos quadrados generalizados podem ser usados para estimar modelos com correlação serial auto-regressiva. Também podemos usar os MQG para resolver o problema de correlação serial nesse caso. Para que o procedimento tenha boas propriedades, N deve ser grande e T relativamente pequeno. Assumimos que temos um painel equilibrado, embora o método possa ser estendido para painéis não equilibrados.

A derivação da transformação MQG que elimina a correlação serial nos erros exige álgebra matricial sofisticada [veja, por exemplo, Wooldridge (2002, Capítulo 10)]. Contudo, a transformação em si é simples. Defina

$$\lambda = 1 - [\sigma_u^2 / (\sigma_u^2 + T\sigma_a^2)]^{1/2}, \quad (14.10)$$

que está entre zero e um. Em seguida, a equação transformada resultará em

$$\begin{aligned} y_{it} - \bar{\lambda} \bar{y}_i &= \beta_0(1 - \lambda) + \beta_1(x_{it1} - \lambda \bar{x}_{i1}) + \dots \\ &\quad + \beta_k(x_{itk} - \lambda \bar{x}_{ik}) + (v_{it} - \lambda \bar{v}_i), \end{aligned} \quad (14.11)$$

onde a barra superior novamente representa as médias temporais. Essa é uma equação bastante interessante, por envolver **dados quase-reduzidos** em cada variável. O estimador de efeitos fixos subtrai as médias temporais da variável correspondente. A transformação de efeitos aleatórios subtrai uma fração daquela média temporal, na qual a fração dependerá de σ_u^2 , σ_a^2 e do número de períodos de tempo, T . O estimador MQG é simplesmente o estimador MQO agrupado da equação (14.11). Não é tão óbvio que os erros em (14.11) são serialmente não-correlacionados, mas eles são. (Veja o Problema 14.3.)

A transformação em (14.11) considera variáveis explicativas que sejam constantes ao longo do tempo, e essa é uma vantagem dos efeitos aleatórios (EA) sobre os efeitos fixos ou sobre a primeira diferenciação. Isso é possível porque EA assume que o efeito não observado é não-correlacionado com todas as variáveis explicativas, sejam elas fixas ao longo do tempo ou não. Assim, em uma equação de salários, podemos incluir uma variável como a educação, mesmo que ela não se altere ao longo do tempo. Entretanto, assumimos que educação não se correlaciona com a_i , que contém aptidão e antecedentes familiares. Em muitas aplicações, a principal razão do uso de dados de painel é possibilitar que o efeito não observado seja correlacionado com as variáveis explicativas.

Na prática, o parâmetro λ nunca é conhecido, mas sempre pode ser estimado. Existem maneiras diferentes de fazer isso, que podem se basear no MQO agrupado ou nos efeitos fixos, por exemplo. Em geral, $\hat{\lambda}$ toma a forma $\hat{\lambda} = 1 - \{1/[1 + T(\hat{\sigma}_a^2/\hat{\sigma}_u^2)]\}^{1/2}$, onde $\hat{\sigma}_a^2$ é um estimador consistente de σ_a^2 , e $\hat{\sigma}_u^2$ é um estimador consistente de σ_u^2 . Esses estimadores podem estar baseados nos resíduos do MQO agrupado ou dos efeitos fixos. Uma possibilidade é que $\hat{\sigma}_a^2 = [NT(T-1)/2 - (k+1)]^{-1} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is}$, onde os \hat{v}_{it} são os resíduos de estimar (14.9) pelo MQO agrupado. Em seguida, podemos estimar σ_u^2 usando $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_a^2$, onde $\hat{\sigma}_v^2$ é o quadrado do erro-padrão habitual da regressão pelo MQO agrupado. [Veja Wooldridge (2002, Capítulo 10) para uma discussão adicional sobre esses estimadores.]

Muitos programas econometrícios geram a estimativa de modelos de efeitos aleatórios e automaticamente computam alguma versão de $\hat{\lambda}$. O estimador MQG factível que utiliza $\hat{\lambda}$ em lugar de λ é chamado de **estimador de efeitos aleatórios**. Sob as hipóteses dos efeitos aleatórios no apêndice deste capítulo, o estimador é consistente (não viésado) e distribuído normalmente e assintoticamente conforme N fica maior com T fixo. As propriedades do estimador EA com N pequeno e T grande são ignoradas, embora certamente elas sejam usadas em tais situações.

A equação (14.11) nos permite relacionar o estimador EA tanto ao MQO agrupado como aos efeitos fixos. O MQO agrupado é obtido quando $\lambda = 0$, e o EF quando $\lambda = 1$. Na prática, a estimativa $\hat{\lambda}$ nunca será zero ou um. Contudo, se $\hat{\lambda}$ estiver próximo de zero, as estimativas EA estarão próximas das estimativas do MQO agrupado. Esse é o caso quando o efeito não observado, a_i , é relativamente sem importância (por ter variância pequena em relação a σ_u^2). É mais comum σ_a^2 ser grande em relação a σ_u^2 caso em que $\hat{\lambda}$ estará mais próximo da unidade. Conforme T fica maior, $\hat{\lambda}$ tende para um, e isso faz com que as estimativas EA e EF sejam muito semelhantes.

Podemos ter uma idéia melhor sobre os méritos relativos dos efeitos aleatórios *versus* os efeitos fixos ao escrever o erro quase-reduzido na equação (14.11) como $v_{it} - \lambda \bar{v}_i = (1 - \lambda)a_i + u_{it} - \lambda \bar{u}_i$. Essa expressão simples torna claro que os erros na equação transformada utilizados na estimativa dos efeitos aleatórios ponderam o efeito não observado em $(1 - \lambda)$. Embora a correlação entre a_i e um ou mais dos x_{ij} cause inconsistência na estimativa de efeitos aleatórios, vemos que a correlação é atenuada pelo fator $(1 - \lambda)$. Quando $\lambda \rightarrow 1$, o termo de viés se aproxima de zero, como devido, porque o estimador EA tende ao estimador EF. Se λ estiver próximo de zero, estaremos deixando uma grande fração do efeito não observado no termo erro e, como consequência, o viés assintótico do estimador EA será maior.

EXEMPLO 14.4

(Uma Equação de Salários Usando Dados de Painel)

Utilizamos novamente os dados contidos no arquivo WAGEPAN.RAW para estimar uma equação dos salários dos homens. Usamos três métodos: MQO agrupado, efeitos aleatórios e efeitos fixos. Nos primeiros dois métodos, podemos incluir *educ* e as *dummies* de raça (*negro* e *hispan*), mas elas se afastam da análise dos efeitos ajustados. As variáveis de variação temporal são *exper*, *exper²*, *sindicato* e *casado*. Como detalhado na Seção 14.1, a variável *exper* é abandonada na análise EF (mas *exper²* permanece). Cada regressão também contém um conjunto completo de *dummies* anuais. Os resultados da estimativa estão na Tabela 14.2.

EXEMPLO 14.4 {continuação}**Tabela 14.2**

Três Estimadores Diferentes de uma Equação de Salários

Variável Dependente: $\log(\text{salário})$			
Variáveis Independentes	MQO Agrupado	Efeitos Aleatórios	Efeitos Fixos
<i>educ</i>	0,091 (0,005)	0,092 (0,011)	—
<i>negro</i>	-0,139 (0,024)	-0,139 (0,048)	—
<i>hispan</i>	0,016 (0,021)	0,022 (0,043)	—
<i>exper</i>	0,067 (0,014)	0,106 (0,015)	—
<i>exper</i> ²	-0,0024 (0,0008)	-0,0047 (0,0007)	-0,0052 (0,0007)
<i>casado</i>	0,108 (0,016)	0,064 (0,017)	0,047 (0,018)
<i>sindicato</i>	0,182 (0,017)	0,106 (0,018)	0,080 (0,019)

Os coeficientes de *educ*, *negro* e *hispan* são semelhantes nas estimações por MQO agrupado e por efeitos aleatórios. Os erros-padrão do MQO agrupado são os habituais, mas eles subestimam os verdadeiros erros-padrão porque ignoram a correlação serial positiva; referimo-nos a eles apenas a título de comparação. O perfil da experiência é algo diferente, e os coeficientes tanto de *casado* como de *sindicato* caem de forma considerável na estimativa pelos efeitos aleatórios. Quando eliminamos totalmente o efeito não observado usando efeitos fixos, o coeficiente de *casado* cai para cerca de 4,7%, embora ainda seja estatisticamente significante. A queda do coeficiente de *casado* é consistente com a idéia de que os homens mais capazes — como capturado por um efeito não observado mais alto, a_1 — são, com maior probabilidade, casados. Portanto, na estimativa por MQO agrupado, uma grande parte da magnitude do coeficiente de *casado* reflete o fato de que homens casados ganhariam mais, mesmo que não fossem casados. Para os 4,7% restantes existem, pelo menos, duas possíveis explicações: (1) o casamento realmente torna o homem mais produtivo ou (2) os empregadores pagam mais aos homens casados porque o casamento é um sinal de estabilidade. Não temos condições de fazer a distinção entre essas duas hipóteses.

A estimativa de λ para a estimativa pelos efeitos aleatórios é $\hat{\lambda} = 0,643$, que ajuda a explicar a razão porque, nas variáveis de variação temporal, as estimativas por EA ficam mais próximas das estimativas por EF do que das estimativas por MQO ajustado.

O coeficiente de *sindicato* estimado por efeitos fixos é cerca de dez pontos percentuais mais baixo do que o estimado por MQO. O que isso enfaticamente sugere sobre a correlação entre *sindicato* e o efeito não observado?

Efeitos Aleatórios ou Efeitos Fixos?

Na leitura de trabalhos aplicados, é possível descobrir que os autores decidem entre efeitos fixos ou aleatórios com base em se os a_i (ou qualquer outra notação que o autor use) são melhor entendidos como parâmetros a serem estimados ou como resultados de uma variável aleatória. Quando não podemos considerar as observações como extrações aleatórias de uma grande população — por exemplo, se temos dados de estados ou municípios —, freqüentemente é racional pensar em a_i como parâmetros a estimar, caso em que usamos os métodos dos efeitos fixos. Lembre-se de que usar efeitos fixos é a mesma coisa que permitir um intercepto diferente para cada observação, e podemos estimar esses interceptos pela inclusão de variáveis *dummy* ou por (14.6).

Mesmo se decidirmos tratar os a_i como variáveis aleatórias, temos que decidir se eles são não correlacionados com as variáveis explicativas. Algumas vezes as pessoas acreditam erroneamente que, ao assumir que a_i é aleatório significa automaticamente que o uso de efeitos aleatórios é a estratégia de estimação mais apropriada. Se pudermos assumir que os a_i são não-correlacionados com todos os x_{it} , o método dos efeitos aleatórios será apropriado. Entretanto, se os a_i forem correlacionados com algumas variáveis explicativas, o método dos efeitos fixos (ou a primeira diferenciação) será necessário; se EA for utilizado, os estimadores serão, geralmente, inconsistentes.

A comparação das estimativas EF e EA pode ser um teste para verificar se existe correlação entre a_i e x_{it} , assumindo que os erros idiosincráticos e as variáveis explicativas são não-correlacionados ao longo de todos os períodos de tempo. Hausman (1978) foi o primeiro a sugerir esse teste. Alguns programas econômétricos computam rotineiramente o teste sob as hipóteses ideais dos efeitos aleatórios relacionadas no apêndice deste capítulo. Detalhes sobre essa estatística podem ser encontrados em Wooldridge (2002, Capítulo 10).

14.3 A APLICAÇÃO DE MÉTODOS DE DADOS DE PAINEL A OUTRAS ESTRUTURAS DE DADOS

Métodos de diferenciação, efeitos fixos e efeitos aleatórios podem ser aplicados em estruturas de dados que não envolvam tempo. Por exemplo, em demografia, é comum usar irmãos (algumas vezes gêmeos) para controlar as características familiares e culturais não observadas. A diferenciação entre irmãos ou, de forma mais genérica, usar a transformação interna dentro de uma família, remove efeitos familiares que possam estar correlacionados com as variáveis explicativas.

Como exemplo, Geronimus e Korenman (1992) usaram pares de irmãs para estudar os efeitos do parto na adolescência sobre os resultados econômicos futuros. Quando o resultado é renda em relação às necessidades — algo que depende do número de filhos —, o modelo é

$$\begin{aligned} \log(\text{rendanec}_{fs}) = & \beta_0 + \delta_0 \text{irmã2}_s + \beta_1 \text{partoad}_{fs} \\ & + \beta_2 \text{idade}_{fs} + \text{outros fatores} + a_f + u_{fs}, \end{aligned} \tag{14.12}$$

onde f indexa as famílias e s indexa uma irmã dentro da família. O intercepto da primeira irmã é β_0 , e o intercepto da segunda irmã é $\beta_0 + \delta_0$. A variável de interesse é $partoad_{fs}$, que é uma variável binária igual a um se a irmã s na família f teve um filho na adolescência. A variável $idade_{fs}$ é a idade atual da irmã s na família f . Geronimus e Korenman usaram também alguns outros controles. A variável não observada a_f , que muda apenas na família, é um *efeito familiar não observado* ou um *efeito familiar fixo*. A principal preocupação na análise é quanto ao fato de $partoad$ estar correlacionado com o efeito familiar. Se for assim, uma análise MQO que faça agrupamentos das famílias e irmãs produz um estimador viesado do efeito da maternidade na adolescência sobre os resultados econômicos. A solução desse problema é simples: dentro de cada família, fazemos a diferenciação de (14.12) entre as irmãs para obter

$$\Delta \log(rendanec) = \delta_0 + \beta_1 \Delta partoad + \beta_2 \Delta idade + \dots + \Delta u, \quad (14.13)$$

o que remove o efeito familiar, a_f , e a equação resultante poderá ser estimada por MQO. Observa-se que não há nenhum elemento temporal neste caso: a diferenciação é feita entre irmãs dentro de uma família. Também consideramos diferenças nos interceptos para as irmãs em (14.12), o que leva a um intercepto diferente de zero na equação diferenciada (14.13). Se, ao entrar com os dados, a ordem das irmãs dentro de cada família for essencialmente aleatória, o intercepto estimado deverá estar próximo de zero. Contudo, mesmo em tais casos, nunca é demais incluir um intercepto em (14.13) e fazer com que o intercepto considere o fato de que, digamos, a primeira irmã listada possa sempre ser a mais necessitada.

Ao usar o método da diferenciação, faz sentido a inclusão de variáveis *dummy* para a raça da mãe e do pai em (14.12)? Explique.

Utilizando 129 pares de irmãs do *National Longitudinal Survey of Young Women* (Pesquisa Longitudinal Nacional de Jovens Mulheres) de 1982, Geronimus e Korenman primeiramente estimaram β_1 por MQO agrupado para obter $-0,33$ ou $-0,26$, em que a segunda estimativa foi obtida controlando as variáveis de antecedentes familiares (como a educação dos pais); ambas as estimativas são bastante significantes estatisticamente [veja a Tabela 3 em Geronimus e Korenman (1992)]. Portanto, a maternidade na adolescência produz um grande impacto sobre a renda familiar futura. Entretanto, quando a equação diferenciada é estimada, o coeficiente de $partoad$ é $-0,08$, que é pequeno e estatisticamente não significante. Isso sugere que em grande parte são os antecedentes familiares que afetam a renda familiar futura, mais do que a gravidez e o parto na adolescência.

Geronimus e Korenman também examinaram vários outros resultados e dois outros conjuntos de dados; em alguns casos, as estimativas dentro das famílias eram grandes, econômica e estatisticamente significantes. Eles também mostraram como os efeitos desaparecem por completo quando os níveis de educação das irmãs são controlados.

Ashenfelter e Krueger (1994) usaram a metodologia da diferenciação para estimar o retorno da educação. Eles obtiveram uma amostra de 149 gêmeos idênticos e coletaram informações sobre ganhos, educação e outras variáveis. A razão de terem usado gêmeos idênticos foi a crença de que eles deveriam ter as mesmas aptidões básicas. Isso pode ser diferenciado fazendo a diferenciação entre os gêmeos, em lugar de usar o MQO nos dados agrupados. Como gêmeos idênticos têm a mesma idade,

sexo e raça, todos esses fatores são eliminados na equação diferenciada. Portanto, Ashenfelter e Krueger regrediram a diferença em $\log(ganhos)$ sobre a diferença em educação e estimaram que o retorno da educação estava em torno de 9,2% ($t = 3,83$). Curiosamente, essa estimativa é maior que a estimada por MQO agrupado, de 8,4% (que controla sexo, idade e raça). Ashenfelter e Krueger também estimaram a equação por efeitos aleatórios e obtiveram um retorno da educação de 8,7%. (Veja a Tabela 5 de seu artigo.) A análise de efeitos aleatórios é, invariavelmente, a mesma que a do caso de dados de painel com dois períodos de tempo.

As amostras usadas por Geronimus e Korenman (1992) e Ashenfelter e Krueger (1994) são exemplos de **amostras de pares comparados**. De forma geral, os métodos de efeitos fixos e aleatórios podem ser aplicados a uma **amostra de aglomeração**. Esses são conjuntos de dados de corte transversal, mas cada observação pertence a um agrupamento bem definido. Nos exemplos anteriores, cada família é uma aglomeração. Como outro exemplo, suponha que temos dados sobre a participação em vários planos de pensão, nos quais as empresas oferecem mais de um plano aos seus funcionários. Podemos, então, ver cada empresa como uma aglomeração, e é bastante claro que os efeitos não observados das empresas serão fatores importantes na determinação de taxas de participação nos planos de pensão dentro das empresas.

Dados educacionais sobre estudantes escolhidos aleatoriamente de muitas escolas formam uma amostra de aglomeração, na qual cada escola é uma aglomeração. Como os resultados em uma aglomeração são propensos a serem correlacionados, geralmente é importante considerar um efeito de aglomeração não observado. A estimativa por efeitos fixos é preferível quando pensamos que o **efeito de aglomeração** não observado — do qual um exemplo é a_j em (14.12) — está correlacionado com uma ou mais das variáveis explicativas. Assim, somente podemos incluir variáveis explicativas que variem, pelo menos um pouco, dentro dos agrupamentos. Esses agrupamentos raramente têm os mesmos tamanhos, de modo que geralmente são requeridos métodos de efeitos fixos para painéis não equilibrados.

Os métodos de efeitos aleatórios também podem ser usados com agrupamentos não equilibrados, desde que o efeito de aglomeração seja não-correlacionado com todas as variáveis explicativas. Também podemos usar o MQO agrupado nesse caso, mas os erros-padrão habituais serão incorretos, a menos que não haja correlação entre os agrupamentos. Alguns programas de regressão possuem comandos simples para corrigir os habituais erros-padrão e estatísticas de testes da correlação geral dentro dos agrupamentos (como também a heteroscedasticidade). Essas correções são as mesmas que as utilizadas para o MQO agrupado em conjuntos de dados de painel que descrevemos no Exemplo 13.9. Como exemplo, Papke (1999) estima modelos de probabilidade linear para a continuação de planos de pensão de benefícios definidos, com base no fato de as firmas terem adotado planos de contribuições definidas. Como é provável a existência de um efeito da firma que induz à correlação entre diferentes planos dentro da mesma firma, Papke corrige os habituais erros-padrão do MQO para a amostragem de aglomeração, como também a heteroscedasticidade, no modelo de probabilidade linear.

Estudamos dois métodos comuns para estimar modelos de dados de painel com efeitos não observados. Comparado com a primeira diferenciação, o estimador de efeitos fixos é eficiente quando os erros idiossincráticos são serialmente não-correlacionados (como também homoscedásticos), e não elaboramos nenhuma hipótese sobre a correlação entre o efeito não observado a_i e as variáveis explicativas. Assim como na primeira diferenciação, qualquer variável explicativa constante no tempo é eliminada da análise. Os métodos de efeitos fixos são diretamente aplicados a painéis não equilibrados, mas temos de assumir que os motivos pelos quais alguns períodos de tempo estão faltando não são sistematicamente relacionados aos erros idiossincráticos.

O estimador de efeitos aleatórios é adéquado quando se acredita que o efeito não observado é não-correlacionado com todas as variáveis explicativas. Nesse caso, a_i pode ser deixado no termo erro, e a correlação serial ao longo do tempo pode ser resolvida pela estimação por mínimos quadrados generalizados. Convenientemente, o MQG factível pode ser obtido por uma regressão agrupada sobre dados quase-reduzidos. O valor do parâmetro de transformação estimado, $\hat{\lambda}$, indica se as estimativas estão propensas a se aproximar da estimativa por MQO agrupado ou por efeitos fixos. Se o conjunto completo das hipóteses dos efeitos aleatórios se sustentar, o estimador de efeitos aleatórios é assimptoticamente — conforme N fica maior com T fixo — mais eficiente que aquele do MQO agrupado, da primeira diferenciação ou dos efeitos fixos (que são todos não-viesados, consistentes e assimptoticamente normais).

Finalmente, os métodos de dados de painel estudados nos Capítulos 13 e 14 podem ser usados quando trabalhamos com amostras de pares comparados ou de aglomeração. A diferenciação ou a transformação interna eliminam o efeito de aglomeração. Se o efeito de aglomeração for não-correlacionado com as variáveis explicativas, o MQO agrupado poderá ser usado, mas os erros-padrão e as estatísticas de testes devem ser ajustados quanto à correlação do agrupamento. A estimação dos efeitos aleatórios também é uma possibilidade.

14.1 Suponha que os erros idiosincráticos em (14.4), $\{u_{it}; t = 1, 2, \dots, T\}$ sejam serialmente não-correlacionados com variância constante, σ_u^2 . Mostre que a correlação entre as diferenças adjacentes Δu_{it} e $\Delta u_{i,t+1}$ é $-0,5$. Portanto, sob as hipóteses EF ideais, a primeira diferenciação induz uma correlação serial negativa de valor conhecido.

14.2 Com uma única variável explicativa, a equação usada para obter o estimador interno é

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i,$$

onde a barra superior representa a média ao longo do tempo. Podemos assumir que $E(a_i) = 0$ por termos incluído um intercepto na equação. Suponha que \bar{u}_i seja não-correlacionado com \bar{x}_i , mas que $Cov(x_{it}, a_i) = \sigma_{xa}$ para todo t (e i , devido à amostragem aleatória no corte transversal).

- (i) Definindo $\tilde{\beta}_1$ como o estimador entre as observações, isto é, o estimador MQO usando as médias temporais, mostre que

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \sigma_{xa}/\text{Var}(\bar{x}_i),$$

onde o limite de probabilidade é definido como $N \rightarrow \infty$. [Sugestão: Veja as equações (5.5) e (5.6).]

- (ii) Assuma também que x_{it} , para todo $t = 1, 2, \dots, T$, seja não-correlacionado com a variância constante σ_x^2 . Mostre que $\text{plim } \tilde{\beta}_1 = \beta_1 + T(\sigma_{xa}/\sigma_x^2)$.
- (iii) Se as variáveis explicativas não forem altamente correlacionadas ao longo do tempo, o que a parte (ii) sugere quanto à possibilidade de a inconsistência no estimador entre as observações ser menor quando existem mais períodos de tempo?

14.3 Em um modelo de efeitos aleatórios, defina o erro de composição $v_{it} = a_i + u_{it}$, onde a_i é não-correlacionado com u_{it} e os u_{it} têm variância constante σ_u^2 e são serialmente não-correlacionados. Defina $e_{it} = v_{it} - \lambda \bar{v}_i$, onde λ é dado em (14.10).

- (i) Mostre que $E(e_{it}) = 0$.
- (ii) Mostre que $\text{Var}(e_{it}) = \sigma_u^2$, $t = 1, \dots, T$.
- (iii) Mostre que para $t \neq s$, $\text{Cov}(e_{it}, e_{is}) = 0$.

14.4 Para determinar os efeitos do desempenho atlético universitário dos candidatos, você coleta dados desses candidatos de uma amostra das faculdades da Divisão I dos anos de 1985, 1990 e 1995.

- (i) Que indicadores de êxito atlético você incluiria em uma equação? Quais seriam alguns dos problemas temporais?
- (ii) Que outros fatores você controlaria na equação?
- (iii) Escreva uma equação que possibilite estimar os efeitos do êxito atlético sobre a mudança percentual nas inscrições. Como você estimaria essa equação? Por que você escolheria esse método?

14.5 Suponha que, para um semestre, você possa coletar os seguintes dados em uma amostra aleatória de calouros e veteranos universitários de cada disciplina: nota padronizada de um exame final, percentagem de freqüência às aulas, uma variável *dummy* indicando se a matéria se enquadra na especialidade do aluno, nota média acumulada antes do início do semestre e nota do exame de ingresso no curso superior.

- (i) Por que você classificaria esse conjunto de dados como uma amostra de aglomeração? Quantas observações, aproximadamente, você esperaria para um aluno típico?
- (ii) Escreva um modelo, semelhante à equação (14.12), que explique o desempenho no exame final em termos de freqüência e outras características. Como subscritos, use s para aluno e c para disciplina. Quais variáveis não se alteram para um aluno?
- (iii) Se você agrupar todos os dados e usar MQO, o que você estará assumindo sobre as características não observadas dos alunos que afetam as taxas de desempenho e de freqüência? Que papel, com relação a isso, desempenham a nota de ingresso no curso superior e a nota média acumulada antes do início do semestre?
- (iv) Se você julga que a nota de ingresso no curso superior e a nota média acumulada antes do início do semestre não capturam adequadamente a capacidade dos alunos, como você estima o efeito da freqüência sobre o desempenho no exame final?

Hipóteses dos Efeitos Fixos e Aleatórios

Neste apêndice, apresentamos definições das hipóteses da estimação por efeitos fixos e aleatórios. Também apresentamos uma discussão sobre as propriedades dos estimadores sob diferentes conjuntos de hipóteses. A verificação dessas afirmações não é muito simples, mas pode ser encontrada em Wooldridge (2002, Capítulo 10).

H I P Ó T E S E E F . 1

Para cada i , o modelo é

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

onde os β_j são os parâmetros a serem estimados.

H I P Ó T E S E E F . 2

Temos uma amostra aleatória na dimensão do corte transversal.

H I P Ó T E S E E F . 3

Para cada t , o valor esperado do erro idiossincrático, dadas as variáveis explicativas em todos os períodos de tempo e o efeito não observado, é zero: $E(u_{it}|X_t, a_t) = 0$.

H I P Ó T E S E E F . 4

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

Sob essas quatro primeiras hipóteses — que são idênticas às hipóteses do estimador de primeiras diferenças —, o estimador de efeitos fixos é não-viesado. Novamente, a principal é a hipótese da exogeneidade estrita, EF.3. Sob essas mesmas hipóteses, o estimador EF é consistente com um T fixo conforme $N \rightarrow \infty$.

H I P Ó T E S E E F . 5

$\text{Var}(u_{it}|X_t, a_t) = \text{Var}(u_{it}) = \sigma_u^2$ para todo $t = 1, \dots, T$.

H I P Ó T E S E E F . 6

Para todo $t \neq s$, os erros idiossincráticos são não-correlacionados (condicionais a todas as variáveis explicativas e a_t): $\text{Cov}(u_{it}, u_{is}|X_t, a_t) = 0$.

Sob as hipóteses EF.1 a EF.6, o estimador de efeitos fixos dos β_j é o melhor estimador linear não-viesado. Como o estimador da PD é linear e não-viesado, ele é, necessariamente, pior que o estimador de EF. A hipótese que torna EF melhor que PD é EF.6, implicando que os erros idiosincráticos são serialmente não-correlacionados.

H I P Ó T E S E E F . 7

Condisional em X_i , e a_{it} , os u_{it} são independentes e identicamente distribuídos como $\text{Normal}(0, \sigma_u^2)$.

A hipótese EF.7 implica EF.3, EF.5 e EF.6, mas é mais forte, pois assume uma distribuição normal dos erros idiosincráticos. Se adicionarmos EF.7, o estimador de EF é normalmente distribuído, e as estatísticas t e F têm distribuições t e F exatas. Sem EF.7, podemos recorrer a aproximações assintóticas. Entretanto, sem fazer hipóteses especiais, essas aproximações exigem N grande e T pequeno.

As hipóteses ideais dos efeitos aleatórios são EF.1, EF.2, EF.3, EF.5 e EF.6. Podemos agora considerar variáveis constantes no tempo. (EF.7 poderia ser incluída, mas, na prática, seria de pouca valia.) Porém, precisamos adicionar hipóteses sobre como os a_{it} são relacionados às variáveis explicativas. Assim, a terceira hipótese é reforçada como segue.

H I P Ó T E S E E A . 3

Adicionalmente a EF.3, o valor esperado de a_{it} , dadas todas as variáveis explicativas, é constante: $E(a_{it}|X_i) = \beta_0$.

Essa é a hipótese que elimina a correlação entre o efeito não observado e as variáveis explicativas, e é a principal distinção entre os efeitos fixos e os efeitos aleatórios. Como assumimos que a_{it} não é correlacionado com todos os elementos de x_{it} , podemos incluir variáveis explicativas constantes no tempo. (Tecnicamente, a quase-redução-temporal remove somente uma fração da média temporal, e não a sua totalidade.) Consideraremos uma expectativa diferente de zero de a_{it} na definição da hipótese EA.3, de forma que o modelo sob as hipóteses dos efeitos aleatórios contém um intercepto, β_0 , como na equação (14.7). Lembre-se, em geral incluiríamos, também, um conjunto de interceptos de períodos de tempo, com o primeiro ano agindo como ano-base.

H I P Ó T E S E E A . 4

Não há relações lineares perfeitas entre as variáveis explicativas.

Também necessitamos impor homoscedasticidade em a_{it} , como segue:

H I P Ó T E S E E A . 5

Adicionalmente a EF.5, a variância de a_{it} , dadas todas as variáveis explicativas, é constante: $\text{Var}(a_{it}|X_i) = \sigma_a^2$.

Sob as seis hipóteses dos efeitos aleatórios (EF.1, EF.2, EA.3, EA.4, EA.5 e EF.6), o estimador de efeitos aleatórios é consistente quando N fica maior com T fixo. (Na realidade, somente as quatro primeiras hipóteses são necessárias para a consistência.) O estimador EA não é não-viesado, a menos que conheçamos λ , o que evita termos de estimá-lo. O estimador EA também é aproximadamente normalmente distribuído com N grande, e os habituais erros-padrão e estatísticas t e F , obtidos da regressão quase-reduzida, são válidas com N grande. [Para verificar isso, veja Wooldridge (2002, Capítulo 10).]

Estimação de Variáveis Instrumentais e Mínimos Quadrados de Dois Estágios

este capítulo, estudaremos com mais profundidade o problema das **variáveis explicativas endógenas** em modelos de regressão múltipla. No Capítulo 3, derivamos o viés nos estimadores MQO quando uma variável importante era omitida; no Capítulo 5, mostramos que os estimadores MQO são, em geral, inconsistentes sob **variáveis omitidas**. O Capítulo 9 demonstrou que o viés de variáveis omitidas pode ser eliminado (ou pelo menos suavizado) quando uma variável *proxy* adequada é escolhida para representar uma variável explicativa não observada. Infelizmente, variáveis *proxy* adequadas nem sempre estão disponíveis.

Nos dois capítulos anteriores, explicamos como a estimação por efeitos fixos ou por primeira diferenciação pode ser usada com dados de painel para estimar os efeitos de variáveis independentes que variam no tempo, na presença de variáveis omitidas *constantes no tempo*. Embora tais métodos sejam bastante úteis, nem sempre temos acesso aos dados de painel. Mesmo que possamos obter dados de painel, eles serão de pouca utilidade se estivermos interessados no efeito de uma variável que não se altera ao longo do tempo: a estimação por primeira diferenciação ou por efeitos fixos elimina as variáveis explicativas constantes no tempo. Além disso, os métodos de dados de painel que estudamos até agora não solucionam o problema de as variáveis omitidas que variam no tempo serem correlacionadas com as variáveis explicativas.

Neste capítulo, consideramos uma abordagem diferente do problema da endogeneidade. Você verá como o método das variáveis instrumentais (VI) pode ser usado para solucionar o problema da endogeneidade de uma ou de mais variáveis explicativas. O método de mínimos quadrados de dois estágios (MQ2E ou MQDE) só é superado em popularidade pelo método de mínimos quadrados ordinários usado para estimar equações lineares em econometria aplicada.

Começaremos mostrando como os métodos VI podem ser usados para obter estimadores consistentes na presença de variáveis omitidas. Os métodos VI também podem ser usados para solucionar o problema de **erros nas variáveis**, pelo menos sob certas hipóteses. O próximo capítulo demonstrará como estimar modelos de equações simultâneas usando os métodos VI.

Nossa abordagem da estimação de variáveis instrumentais acompanha de perto nosso desenvolvimento dos mínimos quadrados ordinários na Parte 1, na qual assumimos que tínhamos uma amostra aleatória de uma população básica. Esse é um ponto de partida desejável, pois, além de simplificar a notação, enfatiza que as hipóteses importantes da estimação por VI são definidas em termos da população básica (como acontece com o MQO). Como mostramos na Parte 2, o MQO pode ser aplicado a dados de séries temporais, e o mesmo é verdadeiro para os métodos de variáveis instrumentais. A Seção 15.7 discute alguns problemas especiais que surgem quando os métodos VI são aplicados a dados de séries temporais. Na Seção 15.8, tratamos de aplicações a cortes transversais agrupados e dados de painel.

15.1 MOTIVAÇÃO: VARIÁVEIS OMITIDAS EM UM MODELO DE REGRESSÃO SIMPLES

Quando defrontados com a possibilidade de viés de variáveis omitidas (ou heterogeneidade não observada), até agora examinamos três opções: (1) podemos ignorar o problema e sofrer as consequências de estimadores viesados e inconsistentes; (2) podemos tentar encontrar e usar uma variável *proxy* adequada da variável não observada; ou (3) podemos assumir que a variável omitida não se altera ao longo do tempo e utilizar os métodos de efeitos fixos ou primeira diferenciação vistos nos Capítulos 13 e 14. A primeira opção poderá ser satisfatória se as estimativas estiverem acopladas com a direção dos vieses dos parâmetros-chave. Por exemplo, se pudermos dizer que o estimador de um parâmetro positivo, digamos, o efeito do treinamento de pessoal sobre os salários subsequentes, é viesado para zero e tenhamos constatado uma estimativa positiva estatisticamente significante, ainda assim teremos descoberto alguma coisa: o treinamento de pessoal tem um efeito positivo sobre os salários, e é provável que tenhamos subestimado o efeito. Infelizmente, o caso oposto, no qual nossas estimativas podem ter uma magnitude grande demais, ocorre com freqüência, o que torna muito difícil para se obter qualquer conclusão útil.

A solução da variável *proxy* discutida na Seção 9.2 também pode produzir resultados satisfatórios, mas nem sempre é possível encontrar uma boa *proxy*. Essa abordagem tenta resolver o problema da variável omitida substituindo a variável não observada por uma variável *proxy*.

Outra abordagem deixa a variável não observada no termo erro, mas, em vez de estimar o modelo por MQO, ela usa um método de estimação que reconhece a presença da variável omitida. Isso é o que o método das variáveis instrumentais faz.

A título de ilustração, considere o problema da aptidão não observada em uma equação de salários de trabalhadores adultos. Um modelo simples é

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{aptid} + e,$$

onde e é o termo erro. No Capítulo 9, mostramos como, sob certas hipóteses, uma variável *proxy* como *QI* pode substituir a aptidão, e assim um estimador consistente será obtido a partir da regressão de

$$\log(\text{salário}) \text{ sobre } \text{educ}, \text{ QI}.$$

Suponha, porém, que uma variável *proxy* não esteja disponível (ou não tenha as propriedades necessárias para produzir um estimador consistente de β_1). Então, colocamos *aptid* no termo erro e ficamos com o modelo de regressão simples

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + u, \quad (15.1)$$

onde u contém *aptid*. Naturalmente, se a equação (15.1) for estimada por MQO, o resultado será um estimador viesado e inconsistente de β_1 se *educ* e *aptid* forem correlacionados.

Constata-se ainda que podemos usar a equação (15.1) como a base da estimação, desde que possamos encontrar uma variável instrumental de *educ*. Para descrever essa abordagem, o modelo de regressão simples é escrito como

$$y = \beta_0 + \beta_1 x + u, \quad (15.2)$$

onde acreditamos que x e u são correlacionados:

$$\text{Cov}(x,u) \neq 0. \quad (15.3)$$

O método das variáveis instrumentais funciona sejam x e u correlacionados ou não, mas, por razões que veremos mais tarde, o MQO deverá ser usado se x for não-correlacionado com u .

Para obter estimadores consistentes de β_0 e β_1 quando x e u forem correlacionados, necessitaremos de alguma informação adicional. A informação virá por meio de uma nova variável que satisfaça certas propriedades. Suponha que temos uma variável observável z que satisfaz as seguintes duas hipóteses: (1) z é não-correlacionado com u , isto é,

$$\text{Cov}(z,u) = 0; \quad (15.4)$$

(2) z é correlacionado com x , isto é,

$$\text{Cov}(z,x) \neq 0. \quad (15.5)$$

Então, dizemos que z é uma **variável instrumental** de x .

Algumas vezes, o requisito (15.4) é resumido na frase “ z é exógeno na equação (15.2)”. No contexto das variáveis omitidas, isso significa que z não deve ter efeito parcial sobre y (uma vez que x e as variáveis omitidas em u estejam controladas) e que z não deve ser correlacionado com fatores não observados que afetem y . A equação (15.5) significa que z deve ser relacionado, positiva ou negativamente, com a variável explicativa endógena x .

Existe uma diferença bastante importante entre os dois requisitos de uma variável instrumental. Como (15.4) envolve a covariância entre z e o erro não observado u , não podemos ter a esperança, em geral, de testar essa hipótese: na maioria dos casos, temos que manter $\text{Cov}(z,u) = 0$, recorrendo ao comportamento econômico ou à introspeção. (Em casos menos usuais, é possível que tenhamos uma variável *proxy* observável de algum fator contido em u , caso em que poderemos verificar se z e a variável *proxy* são mais ou menos não-correlacionadas. Evidentemente, se tivermos uma boa *proxy* de um elemento importante de u , poderemos simplesmente adicionar a *proxy* como uma variável explicativa e estimar a equação expandida por mínimos quadrados ordinários. Veja a Seção 9.2.)

Em contraposição, a condição de que z seja correlacionado com x (na população) pode ser testada, dada uma amostra aleatória da população. A maneira mais fácil de fazer isso é estimar uma regressão simples entre x e z . Na população, temos

$$x = \pi_0 + \pi_1 z + v. \quad (15.6)$$

Então, como $\pi_1 = \text{Cov}(z,x)/\text{Var}(z)$, a hipótese (15.5) será válida se, e somente se, $\pi_1 \neq 0$. Assim, devemos ser capazes de rejeitar a hipótese nula

$$H_0: \pi_1 = 0 \quad (15.7)$$

contra a alternativa bilateral $H_1: \pi_1 \neq 0$, a um nível de significância suficientemente pequeno (digamos, 5% ou 1%). Se esse for o caso, podemos ter uma razoável confiança em que (15.5) se mantém.

Para a equação $\log(\text{salário})$ em (15.1), uma variável instrumental z de educ deve ser (1) não-correlacionada com a aptidão (e com quaisquer outros fatores não observáveis que afetem o salário) e (2) correlacionada com a educação. Algo como o último dígito do número da previdência social de um indivíduo, quase que com certeza satisfará o primeiro requisito: ele será não-correlacionado com a aptidão, por ser determinado de forma aleatória. Porém, essa variável não será correlacionada com a educação, e assim será uma variável instrumental muito pobre de educ .

O que chamamos de *variável proxy* da variável omitida transforma-se em uma VI pobre pelo motivo oposto. No exemplo de $\log(\text{salário})$ com a aptidão omitida, uma variável *proxy* de *aptid* deverá ser tão altamente correlacionada quanto possível com *aptid*, por exemplo. Uma variável instrumental deverá ser *não-correlacionada* com *aptid*. Portanto, embora a variável *QI* seja uma boa candidata para ser uma variável *proxy* de *aptid*, não será uma boa variável instrumental de educ .

Os requisitos são menos claros para outras possíveis candidatas a variável instrumental. Em equações de salários, os economistas usam variáveis do perfil familiar como VIs da educação. Por exemplo, a escolaridade da mãe (educm) é positivamente correlacionada com a educação dos filhos, como poderá ser verificado ao se reunir uma amostra de dados sobre pessoas que trabalham e computar uma regressão simples de educ sobre educm . Portanto, educm satisfará a equação (15.5). O problema é que a escolaridade da mãe também poderá estar correlacionada com a aptidão dos filhos (por meio da aptidão da mãe e talvez da qualidade dos cuidados com a educação em tenra idade).

Outra possível VI de educ em (15.1) é o número de irmãos durante o crescimento (irms). Em geral, ter mais irmãos está associado a níveis médios mais baixos de educação. Assim, se o número de irmãos for não-correlacionado com a aptidão, ele pode agir como uma variável instrumental de educ .

Como segundo exemplo, considere o problema de estimar o efeito causal de faltar às aulas sobre as notas do exame final. Em uma estrutura de regressão simples, temos

$$\text{nota} = \beta_0 + \beta_1 \text{faltas} + u, \quad (15.8)$$

onde nota é a nota no exame final e faltas é o número total de faltas às aulas durante o semestre. Certamente devemos estar preocupados com que faltas seja correlacionado com outros fatores em u : alunos melhores devem ter um menor número de faltas. Assim, uma regressão simples de nota sobre faltas pode não produzir uma boa estimativa do efeito causal de faltas às aulas.

Qual poderia ser uma boa VI de faltas ? Necessitamos de algo que não tenha efeito direto sobre nota e que não seja correlacionado com a aptidão do aluno. Ao mesmo tempo, a VI deve ser correlacionada com faltas . Uma opção é usar a distância entre os alojamentos e o campus. Alguns alunos em grandes universidades se deslocam constantemente para o campus, o que pode aumentar a possibilidade de eles faltarem às aulas (devido ao mau tempo, por terem dormido demais etc.). Assim, faltas pode estar positivamente correlacionado com distância ; isso pode ser verificado regredindo faltas sobre distância e fazendo-se um teste t , como descrito anteriormente.

Será distância correlacionado com u ? No modelo de regressão simples (15.8), alguns fatores em u poderão ser correlacionados com distância . Por exemplo, alunos de família de baixa renda provavelmente residem fora do campus; se a renda afetar o desempenho dos alunos, isso pode fazer com que distância seja correlacionado com u . A Seção 15.2 mostra como usar VI no contexto de regressão multipla, de forma que outros fatores que afetem nota possam ser diretamente incluídos no modelo. Assim, distância pode ser uma boa VI de faltas . Uma abordagem VI pode não ser necessária se houver uma boa *proxy* da aptidão do aluno, como a nota média acumulada anterior ao semestre corrente.

Agora demonstramos que a disponibilidade de uma variável instrumental pode ser usada para estimar com consistência os parâmetros na equação (15.2). Particularmente, mostramos que as hipóteses (15.4) e (15.5) [de forma equivalente, as (15.4) e (15.7)] servem para identificar o parâmetro β_1 . **Identificação** de um parâmetro nesse contexto significa que podemos escrever β_1 em termos de momentos populacionais que possam ser estimados usando uma amostra de dados. Para escrever β_1 em termos de covariâncias populacionais, usamos a equação (15.2): a covariância entre z e y é

$$\text{Cov}(z,y) = \beta_1 \text{Cov}(z,x) + \text{Cov}(z,u).$$

Visto que, sob a hipótese (15.4), $\text{Cov}(z,u) = 0$, e sob a hipótese (15.5), $\text{Cov}(z,x) \neq 0$, podemos escrever β_1 como

$$\beta_1 = \frac{\text{Cov}(z,y)}{\text{Cov}(z,x)}. \quad (15.9)$$

[Observe como essa álgebra simples falhará se z e x forem não-correlacionados, isto é, se $\text{Cov}(z,x) = 0$.] A equação (15.9) mostra que β_1 é a covariância populacional entre z e y , dividida pela covariância populacional entre z e x , o que mostra que β_1 é identificada. Dada uma amostra aleatória, estimamos as quantidades populacionais pelos análogos da amostra. Após cancelar os tamanhos das amostras no numerador e no denominador, obtemos o **estimador de variáveis instrumentais (VI)** de β_1 :

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^n (z_i - \bar{z})(x_i - \bar{x})}. \quad (15.10)$$

Dada uma amostra de dados de x , y e z , é simples obter o estimador de VI em (15.10). O estimador de VI de β_0 é simplesmente $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$, que é bastante parecido com o estimador MQO do intercepto, exceto pelo fato de que o estimador de inclinação, $\hat{\beta}_1$, agora é o estimador de VI.

Não é por acaso que quando $z = x$ obtemos o estimador MQO de β_1 . Em outras palavras, quando x é exógeno, ele pode ser usado como seu próprio VI, e o estimador de VI será, então, idêntico ao estimador MQO.

Uma aplicação simples da lei dos grandes números mostra que o estimador VI é consistente para β_1 : $\text{plim}(\hat{\beta}_1) = \beta_1$, desde que as hipóteses (15.4) e (15.5) sejam satisfeitas. Se qualquer uma dessas hipóteses falhar, os estimadores VI não serão consistentes (veremos mais sobre isso adiante). Uma das características do estimador VI é que, quando x e u forem de fato correlacionados — de forma que a estimação por variáveis instrumentais será realmente necessária —, essencialmente ele sempre será viesado. Isso significa que, em amostras pequenas, o estimador VI pode ter um viés substancial, que é uma das razões pela qual amostras grandes são preferidas.

Inferência Estatística com o Estimador de VI

Dadas as estruturas similares dos estimadores de VI e MQO, não surpreende que o estimador de VI tenha uma distribuição aproximadamente normal em amostras de tamanhos grandes. Para fazer inferência sobre β_1 , precisamos de um erro-padrão que possa ser usado para calcular estatísticas t e intervalos de confiança. A abordagem habitual é impor uma hipótese de homoscedasticidade, exatamente

como no caso de MQO. Agora, a hipótese de homoscedasticidade é declarada condicional à variável instrumental, z , e não à variável explicativa endógena, x . Com as hipóteses anteriores sobre u , x e z , adicionamos

$$E(u^2|z) = \sigma^2 = \text{Var}(u). \quad (15.11)$$

Pode ser mostrado que, sob (15.4), (15.5) e (15.11), a variância assintótica de $\hat{\beta}_1$ é

$$\frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2}, \quad (15.12)$$

onde σ_x^2 é a variância populacional de x , σ^2 é a variância populacional de u , e $\rho_{x,z}^2$ é o quadrado da correlação populacional entre x e z . Isso nos informa o quanto x e z são altamente correlacionados na população. Como com o estimador de MQO, a variância assintótica do estimador de VI decrece para zero à taxa de $1/n$, onde n é o tamanho da amostra.

A equação (15.12) é interessante por duas razões. Primeira, ela fornece uma maneira de obter um erro-padrão do estimador de VI. Todas as quantidades em (15.12) podem ser consistentemente estimadas, dada uma amostra aleatória. Para estimar σ_x^2 , simplesmente calculamos a variância amostral de x_i ; para estimar $\rho_{x,z}^2$ podemos computar a regressão de x_i sobre z_i para obter o R -quadrado, digamos, $R_{x,z}^2$. Finalmente, para estimar σ^2 , podemos usar os resíduos de VI,

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad i = 1, 2, \dots, n,$$

onde $\hat{\beta}_0$ e $\hat{\beta}_1$ são as estimativas de VI. Um estimador consistente de σ^2 parece exatamente igual ao estimador de σ^2 de uma regressão simples de MQO.

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2,$$

onde é padrão usar a correção dos graus de liberdade (embora isso tenha pouco efeito conforme o tamanho da amostra cresce).

O erro-padrão (assimptótico) de $\hat{\beta}_1$ é a raiz quadrada da variância assintótica estimada, e esta última é dada por

$$\frac{\hat{\sigma}^2}{\text{SQT}_x \cdot R_{x,z}^2}, \quad (15.13)$$

onde SQT_x é a soma dos quadrados total de x_i . [Lembre-se de que a variância amostral de x_i é SQT_x/n , e assim os tamanhos das amostras são cancelados para nos dar (15.13)]. O erro-padrão resultante pode ser usado para construir estatísticas t de hipóteses que envolvam β_1 ou intervalos de confiança de β_1 . A estimativa $\hat{\beta}_0$ também tem um erro-padrão que não apresentamos aqui. Qualquer programa moderno de econometria calcula o erro-padrão após qualquer estimação por VI.

Antes de dar um exemplo, é útil comparar as variâncias assimptóticas dos estimadores de VI e MQO (quando x e u são não-correlacionados). Sob as hipóteses de Gauss-Markov, a variância do estimador MQO é σ^2/SQT_x , enquanto a fórmula comparável do estimador de VI é $\sigma^2/(SQT_x \cdot R_{x,z}^2)$; elas diferem apenas no fato de $R_{x,z}^2$ aparecer no denominador da variância da VI. Como o R -quadrado é sempre menor que um, a variância de MQ2E é sempre maior que a variância de MQO (quando MQO é válido). Se $R_{x,z}^2$ for pequeno, então, a variância da VI poderá ser muito maior do que a variância do MQO. Lembre-se, $R_{x,z}^2$ mede a intensidade da relação linear entre x e z na amostra. Se x e z forem somente levemente correlacionados, é possível que $R_{x,z}^2$ seja pequeno, e isso poderá ser traduzido em uma variância amostral muito grande do estimador de VI. Quanto mais altamente correlacionado for z com x , mais próximo de um será $R_{x,z}^2$, e menor será a variância do estimador de VI. No caso em que $z = x$, $R_{x,z}^2 = 1$, e obtemos a variância de MQO, como esperado.

A discussão anterior destaca um importante preço a pagar quando computamos uma estimação de VI quando x e u são não-correlacionados: a variância assimptótica do estimador de VI é sempre maior, e algumas vezes muito maior, que a variância assimptótica do estimador de MQO.

EXEMPLO 15.1

(A Estimação do Retorno da Educação para Mulheres Casadas)

Utilizamos os dados sobre mulheres casadas que trabalham contidos no arquivo MROZ.RAW para estimar o retorno da educação no modelo de regressão simples

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + u. \quad (15.14)$$

Para comparação, primeiro obtemos as estimativas por MQO:

$$\begin{aligned} \log(\hat{\text{salário}}) &= -0,185 + 0,109 \text{ educ} \\ &\quad (0,185) \quad (0,014) \\ n &= 428, R^2 = 0,118. \end{aligned} \quad (15.15)$$

A estimativa de β_1 implica um retorno de quase 11% para um ano a mais de estudo.

Em seguida, usamos a educação do pai ($educp$) como uma variável instrumental de $educ$. Temos que sustentar que $educp$ é não-correlacionado com u . O segundo requisito é que $educ$ e $educp$ sejam correlacionados. Podemos verificar isso facilmente, usando uma regressão simples de $educ$ sobre $educp$ (utilizando somente as mulheres que trabalham da amostra):

$$\begin{aligned} \hat{educ} &= 10,24 + 0,269 \text{ educp} \\ &\quad (0,28) \quad (0,029) \\ n &= 428, R^2 = 0,173. \end{aligned} \quad (15.16)$$

A estatística t de $educp$ é 9,28, o que indica que $educ$ e $educp$ têm uma correlação positiva estatisticamente significante. (Aliás, $educp$ explica cerca de 17% da variação em $educ$ na amostra.) A utilização de $educp$ como uma VI de $educ$ produz

EXEMPLO 15.1 (continuação)

$$\begin{aligned} \log(\hat{\text{salário}}) &= -0,441 + 0,059 \text{ educ} \\ &\quad (0,446) (0,035) \end{aligned} \tag{15.17}$$

$n = 428, R^2 = 0,093.$

A estimativa de VI do retorno da educação é 5,9%, que é cerca da metade da estimativa de MQO. Isso sugere que a estimativa de MQO é alta demais e é consistente com o viés de aptidão omitida. Entretanto, devemos lembrar que essas estimativas são de apenas uma amostra; nunca poderemos saber se 0,109 está acima do verdadeiro retorno da educação, ou se 0,059 está mais próximo do verdadeiro retorno da educação. Além disso, o erro-padrão da estimativa de VI é duas vezes e meia maior que o erro-padrão de MQO (isso era esperado, pelas razões dadas anteriormente). O intervalo de confiança de 95% de β_1 , utilizando MQO, é muito mais justo do que utilizando VI; de fato, o intervalo de confiança da VI, na realidade, contém a estimativa de MQO. Portanto, embora as diferenças entre (15.15) e (15.17) sejam grandes na prática, não podemos dizer se a diferença é estatisticamente significante. Mostraremos como testar isso na Seção 15.5.

No exemplo anterior, o retorno estimado da educação, usando VI, foi menor que usando MQO, o que corresponde às nossas expectativas. Contudo, esse poderia não ter sido o caso, como demonstra o exemplo a seguir.

EXEMPLO 15.2**(A Estimação do Retorno da Educação para Homens)**

Agora usamos o arquivo WAGE2.RAW para estimar o retorno da educação para homens. Utilizamos a variável *irms* (número de irmãos) como uma instrumental de *educ*. Elas são negativamente correlacionadas, como podemos verificar com uma regressão simples:

$$\begin{aligned} \hat{\text{educ}} &= 14,14 - 0,228 \text{ } \textit{irms} \\ &\quad (0,11) (0,030) \\ n &= 935, R^2 = 0,057. \end{aligned}$$

Essa equação implica que cada irmão está associado, na média, com cerca de menos 0,23 ano de estudo. Se assumirmos que *irms* é não-correlacionado com o termo erro em (15.14), o estimador de VI será consistente. A estimação da equação (15.14) usando *irms* como uma VI de *educ* produz

$$\begin{aligned} \log(\hat{\text{salário}}) &= 5,13 + 0,122 \text{ educ} \\ &\quad (0,36) (0,026) \\ n &= 935. \end{aligned}$$

EXEMPLO 15.2 {continuação}

(O R -quadrado foi calculado como negativo, de modo que não o descrevemos. Apresentamos a seguir uma discussão sobre o R -quadrado no contexto da estimação de VI.) A título de comparação, a estimativa por MQO de β_1 é 0,059 com um erro-padrão de 0,006. Diferentemente do exemplo anterior, a estimativa de VI agora é muito mais alta que a do MQO. Embora não saibamos se a diferença é estatisticamente significante, isso não interage com o viés da aptidão omitida do MQO. Pode ser que $irms$ também seja correlacionado com a aptidão: mais irmãos significa, em média, menos atenção dos pais, o que pode resultar em menor aptidão. Outra interpretação seria que o estimador MQO é viesado para zero devido a um erro de medida em $educ$. Isso não é inteiramente convincente, pois, como discutimos na Seção 9.3, não é provável que $educ$ satisfaça o modelo clássico de erros nas variáveis.

Nos exemplos anteriores, a variável explicativa endógena ($educ$) e as variáveis instrumentais ($educp$ e $irms$) tinham significados quantitativos. Entretanto, nada impede que a variável explicativa ou a VI sejam variáveis binárias. Angrist e Krueger (1991), em sua análise mais simples, propuseram uma engenhosa variável instrumental binária de $educ$, utilizando dados do censo, sobre homens nos Estados Unidos. Os autores definiram $pritrim$ igual a um se o homem nasceu no primeiro trimestre do ano, e zero, caso contrário. Parece que o termo *erro* em (15.14) — e, particularmente, a aptidão — deveria não ser relacionado com o trimestre de nascimento. Contudo, $pritrim$ também precisa ser correlacionado com $educ$. Acontece que os anos de estudo realmente diferem sistematicamente na população, com base em trimestres de nascimento. Angrist e Krueger argumentaram persuasivamente que isso é devido às leis de estudo obrigatório em vigor em todos os estados. Resumidamente, os alunos nascidos no início do ano, em geral, começam a estudar com mais idade. Portanto, eles atingem o tempo de estudo obrigatório (16 anos, na maioria dos estados) com um pouco menos de educação do que os alunos que começaram a estudar com menos idade. Sobre os alunos que completaram o ensino médio, Angrist e Krueger verificaram não existir relação entre os anos de estudo e o trimestre de nascimento.

Como anos de estudo varia apenas levemente entre os trimestres de nascimento — o que significa que $R^2_{x,z}$ em (15.13) é muito pequeno — Angrist e Krueger precisaram de uma amostra de tamanho muito grande para obter uma estimativa VI razoavelmente precisa. Utilizando 247.199 homens nascidos entre 1920 e 1929, a estimativa por MQO do retorno da educação foi 0,0801 (erro-padrão de 0,0004), e a estimativa VI foi 0,0715 (0,0219); esses resultados estão registrados na Tabela III do trabalho de Angrist e Krueger. Observe como é elevada a estatística t da estimativa por MQO (próxima de 200), enquanto a estatística t da estimativa VI é somente 3,26. Assim, a estimativa VI é estatisticamente diferente de zero, mas seu intervalo de confiança é muito mais amplo do que aquele com base na estimativa por MQO.

Uma constatação interessante feita por Angrist e Krueger é que a estimativa VI não difere muito daquela feita por MQO. De fato, usando os homens nascidos na década seguinte, a estimativa VI é um pouco mais alta que a feita por MQO. Seria possível interpretar isso como uma demonstração de que não existe viés de aptidão omitida quando equações de salários são estimadas por MQO. Porém, o trabalho de Angrist e Krueger foi criticado em seus fundamentos econometríticos. Como discutido por Bound, Jaeger e Baker (1995), não é óbvio que a época de nascimento seja não relacionada com fatores não observados que afetem o salário. Como explicaremos na próxima subseção, mesmo uma pequena correlação entre z e u pode causar sérios problemas para o estimador de VI.

Para análises referentes a decisões de governo, a variável explicativa endógena freqüentemente é binária. Por exemplo, Angrist (1990) estudou o efeito que o fato de ser um veterano da Guerra do Vietnã tinha sobre os ganhos vitalícios. Um modelo simples é o seguinte

$$\log(ganhos) = \beta_0 + \beta_1 \text{veterano} + u, \quad (15.18)$$

onde *veterano* é uma variável binária. A questão de estimar essa equação por MQO é que pode haver um problema de auto-seleção, como mencionamos no Capítulo 7: talvez aqueles que ganham mais decidam se alistar, ou a decisão de se alistar esteja correlacionada com outras características que afetam os ganhos. Isso fará com que *veterano* e *u* sejam correlacionados.

Angrist salientou que o sorteio militar do Vietnã fornecia um **experimento natural** (veja também o Capítulo 13) que criava uma variável instrumental de *veterano*. Foram entregues aos jovens números para sorteio que determinavam se eles seriam chamados para servir no Vietnã. Como os números fornecidos eram aleatoriamente atribuídos, parece possível que os números do sorteio militar fossem não-correlacionados com o termo erro *u*. Entretanto, aqueles que possuíam números baixos teriam que servir no Vietnã, de forma que a probabilidade de ser um veterano estaria correlacionada com os números do sorteio. Se essas duas premissas forem verdadeiras, o número do sorteio militar seria um bom candidato a VI de *veterano*.

Se alguns dos homens que receberam números baixos no sorteio militar obtivessem maior escolaridade para reduzir a probabilidade de serem selecionados, o número do sorteio seria uma boa variável instrumental de *veterano* em (15.18)?

Também é possível ter uma variável explicativa endógena binária e uma variável instrumental binária. Veja o Problema 15.1 como exemplo.

Propriedades da VI com uma Variável Instrumental Pobre

Já vimos que, embora a VI seja consistente quando *z* e *u* são não-correlacionados e *z* e *x* têm qualquer correlação, positiva ou negativa, as estimativas de VI podem ter grandes erros-padrão, especialmente se *z* e *x* forem apenas fracamente correlacionados. A fraca correlação entre *z* e *x* pode ter consequências ainda mais sérias: o estimador VI pode ter um grande viés assimptótico mesmo se *z* e *u* forem só moderadamente correlacionados.

Podemos verificar isso estudando o limite de probabilidade do estimador VI quando *z* e *u* forem possivelmente correlacionados. Isso pode ser derivado em termos de correlações e desvios-padrão populacionais como

$$\text{plim } \hat{\beta}_1 = \beta_1 + \frac{\text{Corr}(z,u)}{\text{Corr}(z,x)} \cdot \frac{\sigma_u}{\sigma_x}, \quad (15.19)$$

onde σ_u e σ_x são, respectivamente, os desvios-padrão de *u* e *x* na população. A parte interessante dessa equação envolve os termos de correlação. Ela mostra que, mesmo se $\text{Corr}(z,u)$ for pequena, a inconsistência no estimador VI pode ser muito grande se $\text{Corr}(z,x)$ também for pequena. Assim, mesmo se nos concentrarmos apenas na consistência, não será necessariamente melhor usar VI em lugar de MQO se a correlação entre *z* e *u* for menor que aquela entre *x* e *u*. Utilizando o fato de que $\text{Corr}(x,u) = \text{Cov}(x,u)/(\sigma_x \sigma_u)$ com a equação (5.3), podemos escrever o plim do estimador MQO — chamando-o de $\tilde{\beta}_1$ — como

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \text{Corr}(x, u) \cdot \frac{\sigma_u}{\sigma_x}. \quad (15.20)$$

A comparação dessas fórmulas mostra que VI é preferível a MQO com base em viés assintótico quando $\text{Corr}(z, u)/\text{Corr}(z, x) < \text{Corr}(x, u)$.

No exemplo de Angrist e Krueger (1991) mencionado anteriormente, no qual x representa anos de escolaridade e z é uma variável binária indicando o trimestre de nascimento, a correlação entre z e x é bastante pequena. Bound, Jaeger e Baker (1995) examinaram as razões de o trimestre de nascimento e u terem alguma correlação. Pela equação (15.19), vemos que isso pode levar a um viés substancial no estimador VI.

Quando z e x não têm nenhuma correlação, as coisas ficam particularmente ruins, seja z correlacionado ou não com u . O exemplo seguinte ilustra porque devemos sempre verificar se a variável explicativa endógena é correlacionada com a candidata a VI.

EXEMPLO 15.3

(A Estimação do Efeito do Hábito de Fumar sobre o Peso de Nascimento)

No Capítulo 6, estimamos o efeito do hábito de fumar sobre o peso dos recém-nascidos. Sem outras variáveis explicativas, o modelo é

$$\log(\text{pesonas}) = \beta_0 + \beta_1 \text{maços} + u, \quad (15.21)$$

onde maços é a quantidade de maços de cigarros fumada pela mãe por dia. Poderíamos suspeitar que maços estivesse correlacionado com outros fatores relativos à saúde ou à existência de um bom procedimento pré-natal, de forma que maços e u pudessem ser correlacionados. Uma possível variável instrumental de maços seria o preço médio dos cigarros no Estado de residência, precig . Assumiremos que precig e u são não-correlacionados (embora o sistema de saúde estadual possa ser correlacionado com os impostos sobre cigarros).

Se cigarros são um típico produto de consumo, a teoria econômica básica sugere que maços e precig são negativamente correlacionados, de forma que precig pode ser usado como uma VI de maços . Para verificar isso, regredimos maços sobre precig , usando os dados contidos no arquivo BWGHT.RAW:

$$\hat{\text{maços}} = 0,067 + 0,0003 \text{ precig}$$

(0,103) (0,0008)

$$n = 1.388, R^2 = 0,0000, \bar{R}^2 = -0,0006.$$

Isso não indica qualquer relação entre o hábito de fumar durante a gravidez e o preço dos cigarros, o que talvez não seja tão surpreendente devido à natureza dependente do hábito de fumar.

Como maços e precig não são correlacionados, não devemos usar precig como uma VI de maços em (15.21). Mas o que acontece se o fizermos? Os resultados da VI seriam

$$\log(\text{pesonas}) = 4,45 + 2,99 \text{ maços}$$

(0,91) (8,70)

$$n = 1.388$$

EXEMPLO 15.3 (continuação)

(o R -quadrado obtido é negativo). O coeficiente de maços é enorme e tem um sinal inesperado. O erro-padrão também é muito grande, de modo que maços não é significante. Entretanto, as estimativas não têm significado, pois *precig* não atende o único requisito que sempre podemos testar: a hipótese (15.5).

O Cálculo do R -Quadrado após a Estimação de VI

A maioria dos programas de regressão calcula o R -quadrado após a estimação de VI, utilizando a fórmula padrão: $R^2 = 1 - \text{SQR}/\text{SQT}$, onde SQR é a soma dos resíduos quadrados da VI e SQT é a soma dos quadrados total de y . Diferentemente do caso MQO, o R -quadrado da estimação VI pode ser negativo, pois a SQR da VI pode, na realidade, ser maior que a SQT. Embora não faça mal algum descrever o R -quadrado da estimação VI, ele não é muito útil. Quando x e u são correlacionados, não podemos decompor a variância de y em $\beta_1^2 \text{Var}(x) + \text{Var}(u)$ e, assim, o R -quadrado não possui interpretação natural. Além disso, como veremos na Seção 15.3, esses R -quadrados *não podem* ser usados da maneira habitual para computar testes F de restrições conjuntas.

Se nossa meta for produzir o maior R -quadrado, sempre usaremos MQO. Os métodos VI são destinados a produzir melhores estimativas do efeito *ceteris paribus* de x sobre y quando x e u forem correlacionados; o grau de ajuste não é um fator importante. Um alto R -quadrado resultante do MQO é de pouca ajuda se não pudermos consistentemente estimar β_1 .

15.2 ESTIMAÇÃO DE VI DO MODELO DE REGRESSÃO MÚLTIPLA

O estimador de VI para o modelo de regressão simples é facilmente estendido para o caso da regressão múltipla. Começamos com o caso no qual somente uma das variáveis explicativas é correlacionada com o erro. Considere um modelo linear padrão com duas variáveis explicativas:

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1. \quad [15.22]$$

Chamamos essa equação de **equação estrutural**, para enfatizar que estamos interessados em β_j , o que simplesmente significa que a equação supostamente mede uma relação causal. Nesse caso, usamos uma nova notação para distinguir as variáveis endógenas das **variáveis exógenas**. A variável dependente y_1 é claramente endógena, já que é correlacionada com u_1 . As variáveis y_2 e z_1 são as variáveis explicativas, e u_1 é o erro. Como sempre, assumimos que o valor esperado de u_1 é zero: $E(u_1) = 0$. Usamos z_1 para indicar que essa variável é exógena em (15.22) (z_1 é não-correlacionado com u_1). Usamos y_2 para indicar que esta variável é suspeita de ser correlacionada com u_1 . Não especificamos porque y_2 e u_1 são correlacionados, mas por enquanto é melhor pensar em u_1 contendo uma variável omitida correlacionada com y_2 . A notação na equação (15.22) tem origem em modelos de equações simultâneas (que trataremos no Capítulo 16), mas a usamos de forma mais genérica para facilmente distinguir variáveis explicativas exógenas de endógenas, em um modelo de regressão múltipla.

Um exemplo de (15.22) é

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + u_1, \quad [15.23]$$

onde $y_1 = \log(salário)$, $y_2 = educ$ e $z_1 = exper$. Em outras palavras, assumimos que $exper$ é exógeno em (15.23), mas permitimos que $educ$ — pelas razões habituais — seja correlacionado com u_1 .

Sabemos que, se (15.22) for estimada por MQO, todos os estimadores serão viesados e inconsistentes. Assim, seguimos a estratégia sugerida na seção anterior e procuramos uma variável instrumental de y_2 . Como assumimos que z_1 é não-correlacionado com u_1 , podemos usar z_1 como uma variável instrumental de y_2 , assumindo que y_2 e z_1 sejam correlacionados? A resposta é não. Como a própria z_1 aparece como uma variável explicativa em (15.22), ela não pode servir como uma variável instrumental de y_2 . Precisamos de outra variável exógena — vamos chamá-la z_2 — que *não* apareça em (15.22). Portanto, as hipóteses fundamentais são que z_1 e z_2 são não-correlacionados com u_1 ; também assumimos que u_1 tem zero como valor esperado, o que não provoca perda de generalidade quando a equação contém um intercepto:

$$E(u_1) = 0, \text{Cov}(z_1, u_1) = 0, \text{e Cov}(z_2, u_1) = 0. \quad (15.24)$$

Devido à hipótese de média zero, as últimas duas hipóteses são equivalentes a $E(z_1, u_1) = E(z_2, u_1) = 0$, e assim a abordagem do método dos momentos sugere a obtenção dos estimadores $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$, resolvendo os correspondentes amostrais de (15.24):

$$\begin{aligned} \sum_{i=1}^n (y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) &= 0 \\ \sum_{i=1}^n (z_{i1}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1})) &= 0 \\ \sum_{i=1}^n z_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) &= 0. \end{aligned} \quad (15.25)$$

Esse é um conjunto de três equações lineares com três incógnitas $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$, e é facilmente resolvido considerando os dados de y_1 , y_2 , z_1 e z_2 . Os estimadores são chamados *estimadores de variáveis instrumentais*. Se entendermos que y_2 é exógeno e escolhermos $z_2 = y_2$, as equações em (15.25) serão exatamente as condições de primeira ordem dos estimadores MQO; veja as equações em (3.13).

Ainda necessitamos que a variável instrumental z_2 seja correlacionada com y_2 , mas o sentido como essas duas variáveis devem ser correlacionadas é complicado pela presença de z_1 na equação (15.22). Agora precisamos estabelecer a hipótese em termos de correlação parcial. A maneira mais fácil de definir a condição é escrever a variável explicativa endógena como uma função linear das variáveis exógenas e um termo erro:

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2, \quad (15.26)$$

onde, por definição,

$$E(v_2) = 0, \text{Cov}(z_1, v_2) = 0 \text{ e } \text{Cov}(z_2, v_2) = 0.$$

e os π_j são parâmetros desconhecidos. A condição de identificação fundamental [com (15.24)] é que

$$\pi_2 \neq 0 \quad (15.27)$$

Suponha que queiramos estimar o efeito do uso de maconha na nota média de graduação. Em uma população de alunos veteranos de uma universidade, defina *diasusados* como o número de dias no mês anterior em que um aluno fumou maconha e considere a equação estrutural

$$nmgrad = \beta_0 + \beta_1 diasusados + \beta_2 sat + u.$$

- (i) Defina *porcem* como a percentagem de uma sala do ensino médio que informou uso regular de maconha. Se ela for uma candidata a VI de *diasusados*, escreva a forma reduzida de *diasusados*. Você acha que é possível (15.27) ser verdadeira?
- (ii) Você acha que *porcem* é verdadeiramente exógena na equação estrutural? Que problemas podem surgir nesse caso?

Em outras palavras, após considerar os efeitos parciais, z_1 , y_1 e z_2 ainda são correlacionados. Essa correlação pode ser positiva ou negativa, mas não pode ser zero. Testar (15.27) é fácil: estimarmos (15.26) por MQO e usamos um teste *t* (possivelmente tornando-o robusto quanto à heteroscedasticidade). Devemos sempre testar essa hipótese. Infelizmente, não podemos testar se z_1 e z_2 são não-correlacionados com u_1 ; é preciso confiar nisso.

A equação (15.26) é um exemplo de uma equação da forma reduzida, o que significa que escrevemos uma variável endógena em termos de variáveis exógenas. Esse nome deriva de modelos de equações simultâneas — que estudaremos no próximo capítulo —, e é um conceito útil sempre que tivermos uma variável explicativa endógena. O nome ajuda a distingui-la da equação estrutural (15.22).

A adição de mais variáveis explicativas exógenas ao modelo é direta. Escreva o modelo estrutural como

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \dots + \beta_k z_{k-1} + u_1, \quad (15.28)$$

onde y_2 é correlacionado com u_1 . Defina z_k como uma variável não pertencente a (15.28) que também seja exógena. Portanto, assumimos que

$$E(u_1) = 0, \text{Cov}(z_j, u_1) = 0, j = 1, \dots, k. \quad (15.29)$$

A forma reduzida de y_2 é

$$y_2 = \pi_0 + \pi_1 z_1 + \dots + \pi_{k-1} z_{k-1} + \pi_k z_k + v_2, \quad (15.30)$$

e necessitamos de alguma correlação parcial entre z_k e y_2 :

$$\pi_k \neq 0. \quad (15.31)$$

Sob (15.29) e (15.31), z_k é uma VI válida de y_2 . (Não nos importamos com os restantes π_j ; alguns ou todos eles podem ser zero.) Faz sentido imaginar que z_1, \dots, z_{k-1} servem como suas próprias VIs; portanto, a lista de variáveis exógenas com frequência é chamada de lista de variáveis instrumentais. Uma hipótese secundária adicional é que não há relações lineares perfeitas entre as variáveis exógenas; isso é análogo à hipótese da não-existência de colinearidade perfeita no contexto de MQO.

Para a inferência estatística padrão, precisamos assumir a homoscedasticidade de u_1 . Faremos uma descrição mais cuidadosa dessas hipóteses, em um cenário mais geral, na Seção 15.3.

EXEMPLO 15.4

(Utilizando a Proximidade da Faculdade como uma VI da Educação)

Card (1995) usou dados sobre salários e educação de uma amostra de homens em 1976 para estimar o retorno da educação. Ele usou uma variável *dummy* para o caso de alguém ter crescido perto de uma faculdade com cursos de graduação de quatro anos (*prox4*) ser uma variável instrumental da educação. Em uma equação $\log(\text{salário})$, ele incluiu outros controles padrão: experiência, uma variável *dummy* para negro, variáveis *dummy* para o caso de a pessoa residir em área metropolitana (EPRM) e residir no sul, e um conjunto completo de variáveis *dummy* regionais e uma *dummy* para o fato de residir em área metropolitana em 1966. Para *prox4* ser uma variável instrumental válida, deve ser não-correlacionada com o termo erro na equação de salários — o que assumimos — e deve ser parcialmente correlacionada com *educ*. Para verificar o último requisito, regredimos *educ* sobre *prox4* e todas as variáveis exógenas que aparecem na equação. (Isto é, estimamos a forma reduzida de *educ*.) Utilizando os dados contidos no arquivo CARD.RAW, obtemos, de forma condensada,

$$\begin{aligned} \hat{\text{educ}} &= 16,64 + 0,320 \text{ prox4} - 0,413 \text{ exper} + \dots \\ &\quad (0,24) \quad (0,088) \quad (0,034) \\ n &= 3.010, R^2 = 0,477. \end{aligned} \tag{15.32}$$

Estamos interessados no coeficiente e na estatística t de *prox4*. O coeficiente implica que em 1976, com os outros fatores fixos (experiência, raça, região etc.), as pessoas que residiam próximas de uma faculdade em 1966 tinham, em média, quatro meses a mais de educação do que os que cresceram em áreas distantes de faculdades. A estatística t de *prox4* é 3,64, o que produz um p -valor que é zero nas primeiras três casas decimais. Portanto, se *prox4* for não-correlacionado com fatores não observados no termo erro, poderemos usar *prox4* como uma VI de *educ*.

As estimativas de MQO e VI são mostradas na Tabela 15.1. Curiosamente, a estimativa de VI do retorno da educação é duas vezes maior que a do MQO, mas o erro-padrão da estimativa de VI é 18 vezes maior que a do MQO. O intervalo de confiança de 95% da estimativa de VI é de 0,024 a 0,239, o que é uma faixa muito ampla. Intervalos de confiança maiores representam um preço que temos de pagar para obter um estimador consistente do retorno da educação quando acreditamos que *educ* é uma variável endógena.

Como discutido anteriormente, não devemos nos importar com o R -quadrado menor na estimação de VI: por definição, o R -quadrado do MQO será sempre maior, pois o MQO minimiza a soma dos resíduos quadrados.

Tabela 15.1Variável dependente: $\log(\text{salário})$

Variáveis Explicativas	MQO	VI
<i>educ</i>	0,075 (0,003)	0,132 (0,055)
<i>exper</i>	0,085 (0,007)	0,108 (0,024)
<i>exper</i> ²	-0,0023 (0,0003)	-0,0023 (0,0003)
<i>negro</i>	-0,199 (0,018)	-0,147 (0,054)
<i>eprm</i>	0,136 (0,020)	0,112 (0,032)
<i>sul</i>	-0,148 (0,026)	-0,145 (0,027)
Observações R-Quadrado	3,010 0,300	3,010 0,238
Outros controles: <i>eprm66</i> , <i>reg662</i> , ..., <i>reg669</i>		

15.3 MÍNIMOS QUADRADOS DE DOIS ESTÁGIOS

Na seção anterior, assumimos que tínhamos uma única variável explicativa endógena (y_2), com uma variável instrumental de y_2 . Acontece com freqüência que temos mais de uma variável exógena que é excluída do modelo estrutural e que pode ser correlacionada com y_2 , o que significa que são VIs válidas de y_2 . Nesta seção, discutiremos como usar variáveis instrumentais múltiplas.

Uma Única Variável Explicativa Endógena

Considere novamente o modelo estrutural (15.22), que tem uma variável explicativa endógena e uma exógena. Suponha agora que temos *duas* variáveis exógenas excluídas de (15.22): z_2 e z_3 . Nossas hipóteses de z_2 e z_3 não aparecerem em (15.22) e por serem não-correlacionadas com o erro u_1 são conhecidas como **restrições de exclusão**.

Se z_2 e z_3 forem ambas correlacionadas com y_2 , poderemos simplesmente usar cada uma delas como uma VI, como na seção anterior. Entretanto, nesse caso, teríamos dois estimadores de VI, e nenhum deles seria, de forma geral, eficiente. Como z_1 , z_2 e z_3 são não-correlacionadas com u_1 , qualquer combinação linear também será não-correlacionada com u_1 , e, portanto, qualquer combinação linear das variáveis exógenas será uma VI válida. Para encontrar a melhor VI, escolhemos a combinação linear que seja mais altamente correlacionada com y_2 . Isso acaba sendo fornecido pela equação de forma reduzida de y_2 . Escreva

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2, \quad (15.33)$$

onde

$$E(v_2) = 0, \text{Cov}(z_1, v_2) = 0, \text{Cov}(z_2, v_2) = 0 \text{ e } \text{Cov}(z_3, v_2) = 0.$$

Portanto, a melhor VI de y_2 (sob as hipóteses dadas no apêndice deste capítulo) é a combinação linear dos z_j em (15.33) que chamamos y_2^* :

$$y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3. \quad (15.34)$$

Para que esta VI não seja perfeitamente correlacionada com z_1 , precisamos que pelo menos um π_2 ou π_3 seja diferente de zero:

$$\pi_2 \neq 0 \text{ ou } \pi_3 \neq 0. \quad (15.35)$$

Essa é a hipótese fundamental de identificação, tão logo assumamos que todos os z_j sejam exógenos. (O valor de π_1 é irrelevante.) A equação estrutural (15.22) não será identificada se $\pi_2 = 0$ e $\pi_3 = 0$. Podemos testar $H_0: \pi_2 = 0$ e $\pi_3 = 0$ contra (15.35), usando uma estatística F .

Uma maneira útil de pensar em (15.33) é que ela divide y_2 em duas partes. A primeira é y_2^* ; esta é a parte de y_2 que é não-correlacionada com o termo erro u_1 . A segunda é v_2 , e esta parte é possivelmente correlacionada com u_1 , razão pela qual y_2 é possivelmente endógeno.

Dados os dados de z_j , podemos calcular y_2^* para cada observação, desde que conheçamos os parâmetros populacionais π_j . Na prática, isso não é real. No entanto, como vimos na seção anterior, sempre podemos estimar a forma reduzida por MQO. Assim, usando a amostra, regredimos y_2 sobre z_1 , z_2 e z_3 e obtemos os valores estimados:

$$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3 \quad (15.36)$$

(isto é, temos \hat{y}_{i2} para cada i). Neste ponto, devemos verificar se z_2 e z_3 são conjuntamente significantes em (15.33) a um nível de significância razoavelmente pequeno (não mais que 5%). Se z_2 e z_3 não forem conjuntamente significantes em (15.33), estaremos perdendo nosso tempo com a estimação de VI.

Uma vez obtido \hat{y}_2 , podemos usá-lo como VI de y_2 . As três equações para estimar β_0 , β_1 e β_2 são as duas primeiras de (15.25) e a terceira substituída por

$$\sum_{i=1}^n \hat{y}_{i2}(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1}) = 0. \quad (15.37)$$

A solução das três equações com três incógnitas nos fornece os estimadores de VI.

Com instrumentos múltiplos, o estimador de VI também é chamado **estimador de mínimos quadrados de dois estágios (MQ2E)**. A razão é simples. Usando a álgebra do MQO, pode ser demonstrado

que, quando usamos \hat{y}_2 como VI de y_2 , as estimativas VI de $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\beta}_2$ são *idênticas* às estimadas por MQO da regressão de

$$y_1 \text{ sobre } \hat{y}_2 \text{ e } z_1. \quad (15.38)$$

Em outras palavras, podemos obter o estimador MQ2E de dois estágios. O primeiro estágio é computar a regressão em (15.36), no qual obteremos os valores estimados \hat{y}_2 . O segundo estágio é a regressão MQO (15.38). Como usamos \hat{y}_2 em lugar de y_2 , as estimativas MQ2E podem diferir substancialmente das estimativas MQO.

Alguns economistas gostam de interpretar a regressão em (15.38) da seguinte maneira: o valor estimado, \hat{y}_2 , é a versão estimada de y_2^* , e y_2^* é não-correlacionado com u_1 . Portanto, o MQ2E primeiro “expurga” y_2 de sua correlação com u_1 antes de fazer a regressão MQO em (15.38). Podemos mostrar isso inserindo $y_2 = y_2^* + v_2$ em (15.22):

$$y_1 = \beta_0 + \beta_1 y_2^* + \beta_2 z_1 + u_1 + \beta_1 v_2. \quad (15.39)$$

Agora o erro de composição $u_1 + \beta_1 v_2$ tem média zero e é não-correlacionado com y_2^* e z_1 , motivo pelo qual a regressão MQO em (15.38) funciona corretamente.

A maioria dos programas econométricos possui comandos especiais de MQ2E, de modo que não há necessidade de fazer explicitamente as duas etapas. Aliás, na maioria dos casos, devemos evitar fazer o segundo estágio manualmente, já que os erros-padrão e as estatísticas de testes obtidos dessa maneira *não* são válidos. [A razão é que o termo erro em (15.39) inclui v_2 , mas o erro-padrão envolve somente a variância de u_1 .] Qualquer programa de regressão que compute MQ2E solicita a variável dependente, a relação das variáveis explicativas (tanto exógenas como endógenas) e a relação total das variáveis instrumentais (isto é, todas as variáveis exógenas). A descrição dos resultados em geral é bastante semelhante à do MQO.

No modelo (15.28) com uma única VI de y_2 , o estimador de VI da Seção 15.2 é idêntico ao estimador MQ2E. Portanto, quando temos uma VI de cada variável explicativa endógena, podemos chamar o método de estimação de VI ou MQ2E.

A adição de mais variáveis exógenas altera muito pouco o processo. Por exemplo, suponha que a equação de salários seja

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{exper}^2 + u_1, \quad (15.40)$$

onde u_1 é não-correlacionado tanto com exper como com exper^2 . Suponha que também entendemos que a escolaridade tanto da mãe como do pai seja não-correlacionada com u_1 . Então, podemos usar as duas variáveis como VIs de educ . A forma reduzida da equação de educ é

$$\text{educ} = \pi_0 + \pi_1 \text{exper} + \pi_2 \text{exper}^2 + \pi_3 \text{educm} + \pi_4 \text{educp} + v_2, \quad (15.41)$$

e a identificação exige que $\pi_3 \neq 0$ ou $\pi_4 \neq 0$ (naturalmente, ou ambas).

EXEMPLO 15.5**(Retorno da Educação para Mulheres que Trabalham)**

Estimamos a equação (15.40) utilizando os dados contidos no arquivo MROZ.RAW. Primeiro, testamos $H_0: \pi_3 = 0, \pi_4 = 0$ em (15.41) usando um teste F . O resultado é $F = 55,40$, e p -valor = 0,0000. Como esperado, $educ$ é (parcialmente) correlacionado com a educação dos pais.

Quando estimamos (15.40) por MQ2E, obtemos:

$$\begin{aligned} \hat{\log}(salário) &= 0,048 + 0,061 educ + 0,044 exper - 0,0009 exper^2 \\ &\quad (0,400) \quad (0,031) \quad (0,013) \quad (0,0004) \\ n &= 428, R^2 = 0,136. \end{aligned}$$

O retorno da educação estimado está em torno de 6,1%, comparado com uma estimativa por MQO de cerca de 10,8%. Devido a seu erro-padrão relativamente grande, a estimativa MQ2E é pouco significante ao nível de 5% contra uma alternativa bilateral.

As hipóteses necessárias para que o MQ2E tenha as desejadas propriedades de amostras grandes são fornecidas no apêndice deste capítulo, mas é útil fazermos aqui um breve resumo delas. Se escrevermos a equação estrutural como em (15.28),

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \dots + \beta_k z_{k-1} + u_1, \quad (15.42)$$

então, assumimos que cada z_j é não-correlacionado com u_1 . Além disso, precisamos de pelo menos uma variável exógena fora de (15.42) que seja parcialmente correlacionada com y_2 . Isso garantirá a consistência. Para que os habituais erros-padrão e estatísticas t do MQ2E sejam assimptoticamente válidos, também necessitamos de uma hipótese de homoscedasticidade: a variância do erro estrutural, u_1 , não pode depender de qualquer das variáveis exógenas. Para aplicações de séries temporais, precisaremos de mais hipóteses, como veremos na Seção 15.7.

Multicolinearidade e MQ2E

No Capítulo 3, apresentamos o problema da multicolinearidade e mostramos como a correlação entre os regressores pode levar a erros-padrão grandes das estimativas MQO. A multicolinearidade pode ser ainda mais grave com o MQ2E. Para verificar a razão disso, a variância (assimptótica) do estimador MQ2E de β_1 pode ser aproximada como

$$\frac{\sigma^2}{SQT_2(1 - \hat{R}_2^2)}, \quad (15.43)$$

onde $\sigma^2 = \text{Var}(u_1)$, SQT_2 é a variação total em \hat{y}_2 , e \hat{R}_2^2 é o R -quadrado de uma regressão de \hat{y}_2 sobre todas as outras variáveis exógenas que aparecem na equação estrutural. Há duas razões para a variância do estimador MQ2E ser maior do que a do MQO. Primeira, \hat{y}_2 , por construção, tem menos variação que y_2 . (Lembre-se: soma dos quadrados total = soma dos quadrados explicada + soma

dos quadrados dos resíduos; a variação em y_2 é a soma dos quadrados total, enquanto a variação em \hat{y}_2 é a soma dos quadrados explicada da primeira etapa da regressão.) Segunda, a correlação entre \hat{y}_2 e as variáveis exógenas em (15.42) freqüentemente é muito mais elevada que a correlação entre y_2 e aquelas variáveis. Esse fato define essencialmente o problema da multicolinearidade no MQ2E.

A título de ilustração, considere o Exemplo 15.4. Quando $educ$ é regredido sobre as variáveis exógenas na Tabela 15.1 (sem a inclusão de $prox4$), R -quadrado = 0,475; isso reflete um grau moderado de multicolinearidade, mas o importante é que o erro-padrão MQO em $\hat{\beta}_{educ}$ é bem pequeno. Quando obtemos os valores estimados do primeiro estágio, \hat{educ} , e regredimos esses valores sobre as variáveis exógenas na Tabela 15.1, R -quadrado = 0,995, o que indica um grau bastante alto de multicolinearidade entre \hat{educ} e as demais variáveis exógenas na tabela. (Esse elevado R -quadrado não surpreende, pois \hat{educ} é uma função de todas as variáveis exógenas da Tabela 15.1, mais $prox4$.) A equação (15.43) mostra que um \hat{R}_2^2 próximo de um pode resultar num erro-padrão bastante grande do estimador MQ2E. Entretanto, como no caso do MQO, uma amostra de tamanho grande pode compensar um grande \hat{R}_2^2 .

Variáveis Explicativas Endógenas Múltiplas

O método de mínimos quadrados de dois estágios também pode ser usado em modelos com mais de uma variável explicativa endógena. Por exemplo, considere o modelo

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + u_1, \quad (15.44)$$

onde $E(u_1) = 0$ e u_1 é não-correlacionado com z_1 , z_2 e z_3 . As variáveis y_2 e y_3 são variáveis explicativas endógenas: cada uma delas pode ser correlacionada com u_1 .

Para estimar (15.44) por MQ2E, precisamos de *pelo menos duas* variáveis exógenas que não apareçam em (15.44), mas que sejam correlacionadas com y_2 e y_3 . Suponha que temos duas variáveis exógenas excluídas, digamos z_4 e z_5 . Então, a partir de nossa análise de uma única variável explicativa endógena, precisaremos que z_4 ou z_5 apareça em cada uma das formas reduzidas de y_2 e y_3 . (Como antes, podemos usar estatísticas F para fazer o teste.) Embora isso seja necessário para a identificação, infelizmente não é suficiente. Suponha que z_4 apareça em cada uma das formas reduzidas, mas que z_5 não apareça em nenhuma. Então, na realidade, não teremos duas variáveis exógenas parcialmente correlacionadas com y_2 e y_3 . O método de mínimos quadrados em dois estágios não produzirá estimadores consistentes dos β_j .

De forma geral, quando temos mais de uma variável explicativa endógena em um modelo de regressão, a identificação pode falhar de várias e complicadas maneiras. Contudo, podemos facilmente estabelecer uma condição necessária para a identificação, que é chamada de **condição de ordem**.

CONDICÃO DE ORDEM PARA IDENTIFICAÇÃO DE UMA EQUAÇÃO: Necessitamos ao menos tantas variáveis exógenas excluídas quantas forem as variáveis explicativas endógenas incluídas na equação estrutural. A condição de ordem é fácil de ser verificada, já que somente consiste em contar as variáveis endógenas e exógenas. A condição suficiente para a identificação é chamada de **condição de classificação**. Já vimos casos especiais da condição de classificação, por exemplo, na discussão em torno da equação (15.35). Uma expressão geral da condição de classificação exige álgebra matricial e está além do escopo deste texto. [Veja Wooldridge (2002, Capítulo 5).]

O modelo seguinte explica a taxa de crimes violentos, ao nível de cidades, em termos de uma variável binária que indica se existem leis de controle de armas e outros controles:

$$\begin{aligned} \text{violento} = & \beta_0 + \beta_1 \text{controledearmas} + \beta_2 \text{desemp} + \beta_3 \text{popul} \\ & + \beta_4 \text{porcnego} + \beta_5 \text{idade18_21} + \dots \end{aligned}$$

Alguns pesquisadores estimaram equações semelhantes usando variáveis como o número de membros da *National Rifle Association* (Associação Nacional do Rifle) na cidade e o número de assinantes de revistas sobre armas como variáveis instrumentais de *controledearmas* [veja, por exemplo, Kleck e Patterson (1993)]. Essas variáveis instrumentais são convincentes?

O Teste de Hipóteses Múltiplas após a Estimação por MQ2E

Deveemos ser cuidadosos ao testar hipóteses múltiplas em um modelo estimado por MQ2E. É tentador usar a soma dos resíduos quadrados ou a forma R -quadrado da estatística F , como aprendemos ao estudar o MQO no Capítulo 4. O fato de o R -quadrado do MQ2E poder ser negativo sugere que a maneira habitual de computar estatísticas F pode não ser apropriada; esse é precisamente o caso. De fato, se usarmos os resíduos do MQ2E para computar os SQRs de ambos os modelos, o restrito e o sem restrições, não haverá garantia de que $\text{SQR}_r \geq \text{SQR}_m$; se o inverso for verdadeiro, a estatística F será negativa.

É possível combinar a soma dos resíduos quadrados da regressão do segundo estágio [tal como (15.38)] com SQR_m para obter uma estatística com uma distribuição aproximadamente F em amostras grandes. Como muitos programas econometríticos incorporam comandos de testes de uso fácil, que podem ser usados para testar hipóteses múltiplas após a estimação por MQ2E, omitimos os detalhes. As publicações de Davidson e MacKinnon (1993) e Wooldridge (2002, Capítulo 5) contêm discussões sobre como computar estatísticas do tipo F do MQ2E.

15.4 SOLUÇÕES DE VI DE PROBLEMAS DE ERROS NAS VARIÁVEIS

Nas seções anteriores, apresentamos o uso de variáveis instrumentais como um meio de solucionar o problema de variáveis omitidas, mas elas também podem ser usadas para tratar o problema de erro de medida. A título de ilustração, considere o modelo

$$y = \beta_0 + \beta_1 x_1^* + \beta_2 x_2 + u, \quad (15.45)$$

onde y e x_2 são observados, mas x_1^* , não. Seja x_1 uma medida observada de x_1^* : $x_1 = x_1^* + e_1$, onde e_1 é o erro de medida. No Capítulo 9, mostramos que a correlação entre x_1 e e_1 faz com que o MQO, onde x_1 é usado em lugar de x_1^* , seja viesado e inconsistente. Podemos verificar isso escrevendo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + (u - \beta_1 e_1). \quad (15.46)$$

Se as hipóteses dos erros clássicos nas variáveis (ECV) se mantiverem, o viés no estimador MQO de β_1 tenderá a zero. Sem hipóteses adicionais, não podemos fazer nada a esse respeito.

Em alguns casos, podemos usar um procedimento de VI para solucionar o problema de erro de medida. Em (15.45), assumimos que u é não-correlacionado com x_1^* , x_1 e x_2 ; no caso de ECV, assumimos que e_1 é não-correlacionado com x_1^* e x_2 . Isso implica que x_2 é exógeno em (15.46), mas que x_1 é correlacionado com e_1 . O que precisamos é de uma VI de x_1 . Tal VI deve ser correlacionada com x_1 , não-correlacionada com u — de forma a poder ser excluída de (15.45) — e não-correlacionada com o erro de estimação, e_1 .

Uma possibilidade é obter uma segunda estimativa de x_1^* , digamos z_1 . Como é x_1^* que afeta y , é natural pressupor que z_1 é não-correlacionado com u . Se escrevermos $z_1 = x_1^* + a_1$, onde a_1 é o erro de medida em z_1 , devemos assumir que a_1 e e_1 são não-correlacionados. Em outras palavras, tanto x_1 como z_1 medem incorretamente x_1^* , mas seus erros de medida são não-correlacionados. Certamente, x_1 e z_1 são correlacionados por suas dependências de x_1^* , de modo que podemos usar z_1 como uma VI de x_1 .

Onde podemos obter duas medidas de uma variável? Algumas vezes, quando um grupo de trabalhadores é questionado sobre seu salário anual, seus empregadores poderão fornecer uma segunda medida. No caso de pessoas casadas, cada cônjuge pode informar independentemente o nível de poupança ou renda familiar. No estudo de Ashenfelter e Krueger (1994) citado na Seção 14.3, foi solicitado a cada gêmeo o número de anos de educação de seu(ua) irmão(â); isso fornece uma segunda medida que pode ser usada como uma VI da educação auto-relatada em uma equação de salários. (Ashenfelter e Krueger combinaram diferenciação e VI para explicar também o problema da aptidão omitida; veja mais sobre esse assunto na Seção 15.8.) De forma geral, no entanto, é raro ter duas medidas de uma variável explicativa.

Uma alternativa é usar outras variáveis exógenas como VIs de uma variável potencialmente mal medida. Por exemplo, nosso uso de $educm$ e $educp$ como VIs de $educ$ no Exemplo 15.5 pode servir a esse propósito. Se pensarmos que $educ = educ^* + e_1$, as estimativas de VI no Exemplo 15.5 não sofrerão do erro de medida se $educm$ e $edup$ forem não-correlacionados com o erro de medida, e_1 . Isso provavelmente é mais razoável que assumir que $educm$ e $edup$ sejam não-correlacionados com a aptidão, o que está contido em u , em (15.45).

Os métodos VI também podem ser adotados quando usamos fatores como notas de testes para controlar características não observadas. Na Seção 9.2, mostramos que, sob certas hipóteses, variáveis proxy podem ser usadas para solucionar o problema de variáveis omitidas. No Exemplo 9.3, usamos o QI como uma variável proxy da aptidão não observada. Isso simplesmente envolve a adição do QI ao modelo e a computação de uma regressão MQO. Entretanto, existe uma alternativa que funciona quando o QI não satisfaz plenamente as hipóteses da variável proxy. Como ilustração, escreva uma equação de salários como

$$\log(salário) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + aptid + u, \quad (15.47)$$

onde novamente temos o problema da aptidão omitida. Contudo, temos duas notas de testes que são indicadores da aptidão. Assumimos que as notas possam ser escritas como

$$teste_1 = \gamma_1 aptid + e_1$$

e

$$teste_2 = \delta_1 aptid + e_2,$$

onde $\gamma_1 > 0$, $\delta_1 > 0$. Como a aptidão é que afeta o salário, podemos assumir que $teste_1$ e $teste_2$ são não-correlacionados com u . Se escrevermos $aptid$ em termos da nota do primeiro teste e inserirmos o resultado em (15.47), obteremos

$$\log(salário) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \alpha_1 teste_1 + (u - \alpha_1 e_1), \quad (15.48)$$

onde $\alpha_1 = 1/\gamma_1$. Agora, se assumirmos que e_1 é não-correlacionado com todas as variáveis explicativas em (15.47), incluindo $aptid$, e_1 e $teste_1$ devem ser correlacionados. [Observe que $educ$ não é endógeno em (15.48); porém, $teste_1$ o é.] Isso significa que estimar (15.48) por MQO produzirá estimadores inconsistentes dos β_j (e α_1). Sob as hipóteses que levantamos, $teste_1$ não satisfaz as hipóteses da variável proxy.

Se assumirmos que e_2 também é não-correlacionado com todas as variáveis explicativas em (15.47) e que e_1 e e_2 são não-correlacionados, e_1 será não-correlacionado com a segunda nota de testes, $teste_2$. Portanto, $teste_2$ pode ser usado como uma VI de $teste_1$.

EXEMPLO 15.6

(O Uso de duas Notas de Testes como Indicadores de Aptidão)

Usamos os dados contidos no arquivo WAGE2.RAW para implementar o procedimento precedente, no qual QI desempenha o papel de primeira nota de teste e KWW (conhecimento do mundo do trabalho) é a segunda nota de teste. As variáveis explicativas são as mesmas do Exemplo 9.3: $educ$, $exper$, $perm$, $casado$, sul , $urban$ e $negro$. Em vez de adicionar QI e computar o MQO, como na coluna (2) da Tabela 9.2, adicionamos QI e usamos KWW como sua variável instrumental. O coeficiente de $educ$ é 0,025 ($ep = 0,017$). Essa é uma estimativa baixa, e não é estatisticamente diferente de zero. Esse é um resultado problemático, sugerindo que uma de nossas hipóteses não se sustenta; talvez e_1 e e_2 sejam correlacionados.

15.5 O TESTE DE ENDOGENEIDADE E O TESTE DE RESTRIÇÕES SOBREIDENTIFICADORAS

Nesta seção, descreveremos dois importantes testes no contexto de estimação de variáveis instrumentais.

O Teste de Endogeneidade

O estimador MQ2E é menos eficiente que o MQO quando as variáveis explicativas são exógenas; como vimos, as estimativas MQ2E podem ter erros-padrão elevados. Portanto, é útil fazer um teste de endogeneidade de uma variável explicativa que mostre se o MQ2E é ainda necessário. Obter tal teste é bastante simples.

Para ilustrar, suponha que temos uma única variável suspeita de ser endógena,

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + u_1, \quad (15.49)$$

onde z_1 e z_2 são exógenos. Temos duas outras variáveis exógenas, z_3 e z_4 , que não aparecem em (15.49). Se y_2 for não-correlacionado com u_1 , devemos estimar (15.49) por MQO. Como podemos testar isso? Hausman (1978) sugeriu fazer uma comparação direta das estimativas MQO e MQ2E e determinar se as diferenças são estatisticamente significantes. Afinal de contas, tanto MQO como MQ2E serão consistentes se todas as variáveis forem exógenas. Se MQ2E e MQO diferirem de forma significante, concluiremos que y_2 deve ser endógeno (supondo que os z_j são exógenos).

É uma boa idéia computar MQO e MQ2E para verificar se as estimativas são diferentes na prática. Para determinar se as diferenças são estatisticamente significantes, é mais fácil usar um teste de regressão. Isso é feito com base na estimativa da forma reduzida de y_2 , que nesse caso é

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + \pi_4 z_4 + v_2. \quad (15.50)$$

Agora, como cada z_j é não-correlacionado com u_1 , y_2 será não-correlacionado com u_1 se, e somente se, v_2 for não-correlacionado com u_1 ; isso é o que queremos testar. Escreva $u_1 = \delta_1 v_2 + e_1$, onde e_1 é não-correlacionado com v_2 e tem média zero. Então, u_1 e v_2 serão não-correlacionados se, e somente se, $\delta_1 = 0$. A maneira mais fácil de verificar esse valor é incluir v_2 como um regressor adicional em (15.49) e fazer um teste t . Só existe um problema com a implementação desse procedimento: v_2 não é observado, porque ele é o termo erro em (15.50). Como podemos estimar a forma reduzida de y_2 por MQO, podemos obter os resíduos da forma reduzida, \hat{v}_2 . Portanto, estimamos

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \beta_3 z_2 + \delta_1 \hat{v}_2 + \text{erro} \quad (15.51)$$

por MQO e testamos $H_0: \delta_1 = 0$, usando uma estatística t . Se rejeitarmos H_0 a um nível pequeno de significância, concluiremos que y_2 é endógeno porque v_2 e u_1 são correlacionados.

EXEMPLO 15.7

(Retorno da Educação para Mulheres que Trabalham)

Podemos testar a endogeneidade de $educ$ em (15.40), obtendo os resíduos \hat{v}_2 da estimativa da forma reduzida (15.41) — usando somente mulheres que trabalham — e incluindo-os em (15.40). Quando fazemos isso, o coeficiente de \hat{v}_2 é $\hat{\delta}_1 = 0,058$, e $t = 1,67$. Essa é uma evidência moderada de correlação positiva entre u_1 e v_2 . Provavelmente é uma boa idéia descrever ambas as estimativas porque a estimativa MQ2E do retorno da educação (6,1%) está bem abaixo da estimativa MQO (10,8%).

O TESTE DE ENDOGENEIDADE DE UMA ÚNICA VARIÁVEL EXPLICATIVA:

- (i) Estime a forma reduzida de y_2 , regredindo y_2 sobre todas as variáveis exógenas (inclusive aquelas da equação estrutural e as VIs adicionais). Obtenha os resíduos, \hat{v}_2 .
- (ii) Adicione \hat{v}_2 à equação estrutural (que inclui y_2) e verifique a significância de \hat{v}_2 , usando uma regressão MQO. Se o coeficiente de \hat{v}_2 for estatisticamente diferente de zero, concluiremos que y_2 é endógeno. Podemos querer usar um teste t robusto em relação à heteroscedasticidade.

Uma característica interessante da regressão da parte (ii) é que as estimativas de todas as variáveis (exceto \hat{v}_2) são idênticas às estimativas MQ2E. Por exemplo, estimar (15.51) por MQO produz $\hat{\beta}_j$, que são idênticos aos da estimativa MQ2E a partir da equação (15.49). Essa é uma maneira fácil de verificar se fizermos a regressão apropriada para o teste de endogeneidade. Ela também fornece uma outra interpretação do MQ2E: a inclusão de \hat{v}_2 na regressão MQO (15.51) melhora a endogeneidade de y_2 .

Também podemos testar a endogeneidade de múltiplas variáveis explicativas. Para cada variável suspeita de ser endógena, obtemos os resíduos da forma reduzida, como na parte (i). Depois, verificamos a significância conjunta desses resíduos na equação estrutural, usando um teste F . A significância conjunta indica que pelo menos uma variável explicativa suspeita é endógena. O número de restrições de exclusão testadas é o número de variáveis explicativas suspeitas de serem endógenas.

O Teste de Restrições Sobreidentificadoras

Quando apresentamos o estimador simples de variáveis instrumentais na Seção 15.1, enfatizamos que uma VI deve satisfazer dois requisitos: ela deve ser não-correlacionada com o erro e correlacionada com a variável explicativa endógena. Vimos em modelos razoavelmente complicados como decidir se o segundo requisito pode ser testado usando um teste t ou F na regressão da forma reduzida. Alegamos que o primeiro requisito não pode ser testado porque ele envolve uma correlação entre a VI e um erro não observado. Porém, se tivermos mais de uma variável instrumental, podemos efetivamente testar se algumas delas são não-correlacionadas com o erro estrutural.

Como exemplo, considere novamente a equação (15.49) com duas variáveis instrumentais adicionais, z_3 e z_4 . Sabemos que podemos estimar (15.49) usando somente z_3 como uma VI de y_2 . Dadas as estimativas de VI, podemos computar os resíduos, $\hat{u}_1 = y_1 - \hat{\beta}_0 - \hat{\beta}_1 y_2 - \hat{\beta}_2 z_1 - \hat{\beta}_3 z_2$. Como z_4 não é usada na estimativa, podemos verificar se z_4 e \hat{u}_1 são correlacionados na amostra. Se forem, z_4 não será uma VI válida de y_2 . Naturalmente, isso não nos informa se z_3 e u_1 são correlacionados; de fato, para que esse teste seja proveitoso, devemos assumir que z_3 e u_1 são não-correlacionados. No entanto, se z_3 e z_4 forem escolhidos usando a mesma lógica —, como a educação da mãe e a educação do pai — constatar que z_4 é correlacionado com u_1 lança dúvidas sobre o uso de z_3 como uma VI.

Como os papéis de z_3 e z_4 podem ser invertidos, também podemos testar se z_3 é correlacionado com u_1 , desde que z_4 e u_1 sejam supostamente não-correlacionados. Qual teste devemos usar? Acontece que não importa qual teste escolhamos. Temos de assumir que pelo menos uma VI é exógena. Então, poderemos testar as **restrições sobreidentificadoras** que são usadas no MQ2E. Para o nosso propósito, o número de restrições sobreidentificadoras é simplesmente o número de variáveis instrumentais extras. Suponha que temos somente uma variável explicativa endógena. Se tivermos somente uma única VI de y_2 , não teremos restrições sobreidentificadoras, e não haverá nada que possa ser testado. Se tivermos duas VIs de y_2 , como no exemplo anterior, teremos uma restrição sobreidentificadora. Se tivermos três VIs, teremos duas restrições sobreidentificadoras, e assim por diante.

Testar as restrições sobreidentificadoras é bastante simples. Devemos obter os resíduos do MQ2E e depois computar uma regressão auxiliar.

O TESTE DE RESTRIÇÕES SOBREIDENTIFICADORAS:

- (i) Estime a equação estrutural por MQ2E e obtenha os resíduos MQ2E, \hat{u}_1 .
- (ii) Regreda \hat{u}_1 sobre todas as variáveis exógenas. Obtenha o R -quadrado, digamos R_1^2 .
- (iii) Sob a hipótese nula de que todas as VIs são não-correlacionadas com u_1 , $nR_1^2 \stackrel{d}{\sim} \chi_q^2$, onde q é o número de variáveis instrumentais fora do modelo menos o número total de variáveis explicativas endógenas. Se nR_1^2 exceder (digamos) o valor crítico de 5% na distribuição χ_q^2 , rejeitamos H_0 e concluímos que pelo menos algumas das VIs não são exógenas.

EXEMPLO 15.8**[Retorno da Educação para Mulheres que Trabalham]**

Quando usamos educm e educp como VIs de educ em (15.40), temos uma única restrição sobreidentificadora. A regressão dos resíduos do MQ2E \hat{u}_i , sobre exper , exper^2 , educm e educp produz $R^2 = 0,0009$. Portanto, $nR^2 = 428(0,0009) = 0,3852$, que é um valor muito pequeno em uma distribuição χ^2_1 (p -valor = 0,535). Portanto, as variáveis da educação dos pais passam no teste de sobreidentificação. Quando adicionamos a educação do marido à lista de VI, obtemos duas restrições sobreidentificadoras, e $nR^2 = 1,11$ (p -valor = 0,574). Portanto, parece razoável adicionarmos educmar à lista de IV, pois isso reduz o erro-padrão da estimativa MQ2E: a estimativa MQ2E de educ usando as três variáveis instrumentais é 0,080 ($ep = 0,022$), de modo que isso torna educ muito mais significante do que quando educmar não é usada como uma VI ($\hat{\beta}_{\text{educ}} = 0,061$, $ep = 0,031$).

No exemplo anterior, aludimos a um fato geral sobre o MQ2E: sob as hipóteses padrões do MQ2E, a adição de variáveis instrumentais à lista melhora a eficiência assimptótica do MQ2E. Entretanto, isso requer que quaisquer novas variáveis instrumentais sejam de fato exógenas — caso contrário, o MQ2E não será sequer consistente —, e isso será apenas um resultado assimptótico. Dados os tamanhos típicos das amostras disponíveis, a adição de variáveis instrumentais em demasia, isto é, o aumento do número de restrições sobreidentificadoras, pode causar vieses severos no MQ2E. Uma discussão detalhada nos desviaria muito do assunto. Uma boa ilustração é dada por Bound, Jaeger e Baker (1995), que argumentam que as estimativas MQ2E do retorno da educação obtidas por Angrist e Krueger (1991), usando muitas variáveis instrumentais, são propensas a ser seriamente viesadas (mesmo com centenas de milhares de observações!).

O teste de sobreidentificação pode ser usado sempre que tivermos mais variáveis instrumentais do que necessitamos. Se tivermos quantidade suficiente de variáveis instrumentais, o modelo é considerado *exatamente identificado*, e o R -quadrado na parte (ii) será zero, identicamente. Como mencionamos anteriormente, não podemos testar a exogeneidade das variáveis instrumentais no caso do modelo exatamente identificado.

O teste pode se tornar robusto quanto à heteroscedasticidade de forma arbitrária; para detalhes, veja Wooldridge (2002, Capítulo 5).

15.6 O MQ2E COM HETEROSCEDASTICIDADE

A heteroscedasticidade no contexto do MQ2E suscita essencialmente os mesmos problemas do MQO. O mais importante é a possibilidade de obter erros-padrão e estatísticas de testes que são (assimptoticamente) robustos quanto à heteroscedasticidade de forma arbitrária e desconhecida. Na verdade, a expressão (8.4) continua a ser válida se os \hat{r}_{ij} forem obtidos como os resíduos da regressão de \hat{x}_{ij} sobre os outros \hat{x}_{ih} , onde o símbolo “ \sim ” representa valores estimados das regressões do primeiro estágio (das variáveis explicativas endógenas). Wooldridge (2002, Capítulo 5) contém mais detalhes. Alguns programas econométricos fazem isso rotineiramente.

Também podemos testar a heteroscedasticidade, usando um teste análogo ao de Breusch-Pagan que apresentamos no Capítulo 8. Sejam \hat{u} os resíduos MQ2E e z_1, z_2, \dots, z_m todas as variáveis exógenas (inclusive as usadas como VIs das variáveis explicativas endógenas). Assim, sob hipóteses razoáveis [explicadas, por exemplo, em Wooldridge (2002, Capítulo 5)], uma estatística assimptoticamente válida será a habitual estatística F da significância conjunta em uma regressão de \hat{u}^2 sobre z_1, z_2, \dots, z_m . A hipótese nula de homoscedasticidade será rejeitada se os z_i forem conjuntamente significantes.

Se aplicarmos esse teste no Exemplo 15.8, usando educm , educp e educmar como variáveis instrumentais de educ , obteremos $F_{5,422} = 2,53$, e p -valor = 0,029. Isso é evidência de heteroscedasticidade ao nível de 5%. Podemos querer computar erros-padrão robustos em relação à heteroscedasticidade para explicar isso.

Se soubermos como a variância do erro depende das variáveis exógenas, poderemos usar um procedimento de MQ2E ponderado, essencialmente o mesmo da Seção 8.4. Após estimar um modelo para $\text{Var}(u|z_1, z_2, \dots, z_m)$, dividimos a variável dependente, as variáveis explicativas e todas as variáveis instrumentais da observação i por $\sqrt{\hat{h}_i}$, onde \hat{h}_i representa a variância estimada. (A constante, que é tanto uma variável explicativa como uma VI, é dividida por $\sqrt{\hat{h}_i}$; veja a Seção 8.4.) Em seguida, aplicamos o MQ2E na equação transformada usando as variáveis instrumentais transformadas.

15.7 A APLICAÇÃO DO MQ2E A EQUAÇÕES DE SÉRIES TEMPORAIS

Quando aplicamos o MQ2E a dados de séries temporais, muitas das considerações que surgiram sobre o MQO nos Capítulos 10, 11 e 12 são pertinentes. Escreva a equação estrutural de cada período de tempo como

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + u_t, \quad (15.52)$$

onde uma ou mais das variáveis explicativas x_{tj} possam ser correlacionadas com u_t . Seja o conjunto de variáveis exógenas representado por z_{t1}, \dots, z_{tm} :

$$\text{E}(u_t) = 0, \text{Cov}(z_{tj}, u_t) = 0, j = 1, \dots, m.$$

Qualquer variável explicativa exógena também é uma z_{tj} . Para a identificação, é necessário que $m \geq k$ (temos tantas variáveis exógenas quanto variáveis explicativas).

A mecânica do MQ2E é idêntica para dados de séries temporais ou de corte transversal, mas para dados de séries temporais as propriedades estatísticas do MQ2E dependem das propriedades de tendência e de correlação das seqüências básicas. Em particular, devemos ser cuidadosos ao decidir incluir uma variável de tendência se tivermos a variável dependente ou variáveis explicativas com tendência. Como uma tendência temporal é exógena, ela pode sempre servir como sua própria variável instrumental. O mesmo é verdade em relação a variáveis *dummy* sazonais, se forem usados dados mensais ou trimestrais.

Um modelo para testar o efeito do crescimento dos gastos governamentais sobre o crescimento da produção é

$$cPIB_t = \beta_0 + \beta_1 cGOV_t + \beta_2 RAZINV_t + \beta_3 cTRAB_t + u_t,$$

onde c indica crescimento, PIB é o produto interno bruto real, GOV é o gasto governamental real, $RAZINV$ é a razão do investimento interno sobre o PIB e $TRAB$ é o tamanho da força de trabalho. [Veja equação (6) em Ram (1996).] Sob quais hipóteses uma variável *dummy* indicando se o presidente no ano $t - 1$ era um Republicano seria uma VI adequada de GOV ?

Séries que possuem forte persistência (têm raízes unitárias) devem ser usadas com cuidado, assim como no MQO. Freqüentemente, diferenciar a equação é necessário antes da estimação, e isso se aplica também às variáveis instrumentais.

Sob hipóteses análogas às do Capítulo 11 para as propriedades assintóticas do MQO, o MQ2E usando dados de séries temporais é consistente e distribuído normal e assintoticamente. Na verdade, se substituirmos as variáveis explicativas pelas variáveis instrumentais ao estabelecer as hipóteses, somente precisaremos adicionar as hipóteses de identificação do MQ2E. Por exemplo, a hipótese de homoscedasticidade é definida como

$$E(u_t^2 | z_{t1}, \dots, z_{tm}) = \sigma^2, \quad (15.53)$$

e a hipótese de ausência de correlação serial é estabelecida como

$$E(u_t u_s | Z_t, Z_s) = 0, \text{ para todo } t \neq s, \quad (15.54)$$

onde z_t representa todas as variáveis exógenas no tempo t . Uma lista completa das hipóteses é apresentada no apêndice deste capítulo. Forneceremos exemplos do MQ2E para problemas de séries temporais no Capítulo 16; veja também o Problema 15.15.

Como no caso do MQO, a hipótese de ausência de correlação serial pode com freqüência ser violada com dados de séries temporais. Felizmente, é bastante fácil testar a existência de correlação serial AR(1). Se escrevermos $u_t = \rho u_{t-1} + e_t$ e inserirmos essa expressão na equação (15.52), obteremos

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + \rho u_{t-1} + e_t, \quad t \geq 2. \quad (15.55)$$

Para testar $H_0: \rho = 0$, devemos substituir u_{t-1} pelos resíduos do MQ2E, \hat{u}_{t-1} . Além disso, se os x_{ij} forem endógenos em (15.52), eles serão endógenos em (15.55), de modo que ainda necessitaremos usar uma VI. Como e_t é não-correlacionado com todos os valores passados de u_t , \hat{u}_{t-1} pode ser usado como sua própria variável instrumental.

O TESTE DA CORRELAÇÃO SERIAL AR(1) APÓS O MQ2E:

- (i) Estime (15.52) por MQ2E e obtenha os resíduos do MQ2E, \hat{u}_t .
- (ii) Estime

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_k x_{tk} + \rho \hat{u}_{t-1} + erro_t, \quad t = 2, \dots, n$$

por MQ2E, usando as mesmas variáveis instrumentais da parte (i), em adição a \hat{u}_{t-1} . Use a estatística t de $\hat{\rho}$ para testar $H_0: \rho = 0$.

Assim como na versão MQO desse teste no Capítulo 12, a estatística t somente tem justificação assintótica, mas na prática ela tende a funcionar bem. Uma versão do teste robusto em relação à heteroscedasticidade pode ser usada para proteção contra a heteroscedasticidade. Além disso, resíduos defasados poderão ser adicionados à equação para testar a existência de formas mais altas de correlação serial, usando um teste F conjunto.

O que acontece se detectarmos correlação serial? Alguns programas econométricos computam erros-padrão robustos em relação a formas razoavelmente gerais de correlação serial e heteroscedasticidade. Esse é um bom e simples caminho a seguir se seu programa de econometria fizer isso. Os cálculos são muito semelhantes aos da Seção 12.5 do MQO. Veja Wooldridge (1995) para fórmulas e outros métodos computacionais.

Uma alternativa é usar o modelo AR(1) e corrigir a correlação serial. O procedimento é semelhante ao do MQO e coloca restrições adicionais sobre as variáveis instrumentais. A equação quase-diferenciada é a mesma que a equação (12.32):

$$\tilde{y}_t = \beta_0(1 - \rho) + \beta_1 \tilde{x}_{t1} + \dots + \beta_k \tilde{x}_{tk} + e_t, t \geq 2, \quad (15.56)$$

onde $\tilde{x}_{ij} = x_{ij} - \rho x_{i-1,j}$. (Podemos usar a observação $t = 1$ como fizemos na Seção 12.3, mas, por simplicidade, omitimos isso aqui.) A questão é: o que podemos usar como variáveis instrumentais? Parece natural usar variáveis instrumentais quase-diferenciadas, $\tilde{z}_{ij} = z_{ij} - \rho z_{i-1,j}$. Isso somente funcionará, porém, se em (15.52) o erro u_t original for não-correlacionado com as variáveis instrumentais nos tempos t , $t - 1$ e $t + 1$. Isto é, as variáveis instrumentais devem ser estritamente exógenas em (15.52). Isso exclui variáveis dependentes defasadas como VIs, por exemplo. Isso também elimina casos em que movimentos futuros nas VIs reagirem a alterações correntes e passadas no erro, u_t .

MQ2E COM ERROS AR(1):

- (i) Estime (15.52) por MQ2E e obtenha os resíduos MQ2E, \hat{u}_t , $t = 1, 2, \dots, n$.
- (ii) Obtenha $\hat{\rho}$ da regressão de \hat{u}_t sobre \hat{u}_{t-1} , $t = 2, \dots, n$ e construa as variáveis quase-diferenciadas $\tilde{y}_t = y_t - \hat{\rho} y_{t-1}$, $\tilde{x}_{ij} = x_{ij} - \hat{\rho} x_{i-1,j}$ e $\tilde{z}_{ij} = z_{ij} - \hat{\rho} z_{i-1,j}$ para $t \geq 2$. (Lembre-se de que, na maioria dos casos, algumas das VIs também serão variáveis explicativas.)
- (iii) Estime (15.56) (onde ρ é substituído por $\hat{\rho}$) por MQ2E, usando as \tilde{z}_{ij} como variáveis instrumentais. Ao assumir que (15.56) satisfaz as hipóteses do MQ2E do apêndice deste capítulo, as estatísticas de testes habituais do MQ2E serão assintoticamente válidas.

Podemos também usar o primeiro período de tempo como na estimação de Prais-Winsten do modelo com variáveis explicativas exógenas. As variáveis transformadas no primeiro período de tempo — a variável dependente, as variáveis explicativas e as variáveis instrumentais — são obtidas simplesmente pela multiplicação de todos os valores do primeiro período por $(1 - \hat{\rho})^{1/2}$. (Veja também a Seção 12.3.)

15.8 A APLICAÇÃO DO MQ2E EM CORTES TRANSVERSAIS AGRUPADOS E EM DADOS DE PAINEL

A aplicação dos métodos de variáveis instrumentais em cortes transversais independentemente agrupados não apresenta novas dificuldades. Como com os modelos estimados por MQO, deveremos freqüentemente incluir variáveis *dummy* temporais para levar em conta os efeitos agregados do tempo. Essas variáveis *dummy* são exógenas —, pois a passagem do tempo é exógena — e assim elas agem como suas próprias variáveis instrumentais.

EXEMPLO 15.9**(Efeito da Educação sobre a Fertilidade)**

No Exemplo 13.1, usamos o corte transversal agrupado contido no arquivo FERTIL1.RAW para estimar o efeito de educação sobre a fertilidade das mulheres, controlando vários fatores. Como em Sander (1992), consideramos a possibilidade de que $educ$ seja endógeno na equação. Como variáveis instrumentais de $educ$, usamos os níveis de educação da mãe e do pai ($educm$, $educp$). A estimativa de MQ2E de β_{educ} é $-0,153$ ($ep = 0,039$), comparada com a estimativa do MQO de $-0,128$ ($ep = 0,018$). A estimativa por MQ2E mostra um efeito da educação sobre a fertilidade de certa forma maior, mas o erro-padrão do MQ2E é mais de duas vezes maior que o do MQO. (Na verdade, o intervalo de confiança de 95% baseado no MQ2E facilmente contém a estimativa MQO.) As estimativas MQO e MQ2E de β_{educ} não são estatisticamente diferentes, como pode ser visto testando a endogeneidade de $educ$, como na Seção 15.5: quando o resíduo da forma reduzida, \hat{v}_2 , é incluído com os outros regressores na Tabela 13.1 (inclusive $educ$), sua estatística t é 0,702, que não é significante em qualquer nível razoável. Portanto, nesse caso, concluímos que a diferença entre o MQ2E e o MQO é devida ao erro de amostragem.

A estimativa de variáveis instrumentais pode ser combinada com métodos de dados de painel, particularmente a primeira diferenciação, para estimar consistentemente parâmetros na presença de efeitos não observados e de endogeneidade em uma ou mais variáveis explicativas com variação temporal. O exemplo simples a seguir ilustra essa combinação de métodos.

EXEMPLO 15.10**(Treinamento de Pessoal e Produtividade de Trabalhadores)**

Suponha que queremos estimar o efeito de uma hora adicional de treinamento sobre a produtividade dos trabalhadores. Para os anos de 1987 e 1988, considere o modelo simples de dados de painel

$$\log(ref_{it}) = \beta_0 + \delta_0 d88_t + \beta_1 hrsemp_{it} + a_i + u_{it}, \quad t = 1, 2,$$

onde ref_{it} é a taxa de refugo dos produtos da firma i no ano t , e $hrsemp_{it}$ representa horas de treinamento por empregado. Como sempre, permitimos interceptos diferentes para os anos e um efeito constante não observado da firma, a_i .

Pelas razões discutidas na Seção 13.2, podemos estar preocupados com a possibilidade de que $hrsemp_{it}$ seja correlacionado com a_i , que contém a aptidão não medida do trabalhador. Como antes, fazemos a diferenciação para remover a_i :

$$\Delta \log(ref_{it}) = \delta_0 + \beta_1 \Delta hrsemp_{it} + \Delta u_{it}. \quad (15.57)$$

Normalmente, estimariamós essa equação por MQO. Entretanto, se Δu_{it} for correlacionado com $\Delta hrsemp_{it}$? Por exemplo, uma empresa pode雇用 trabalhadores mais habilidosos e, simultaneamente, reduzir o nível de treinamento. Nesse caso, necessitamos de uma variável instrumental de $\Delta hrsemp_{it}$. Geralmente, seria difícil encontrar tal VI, mas podemos explorar o fato de que algumas empresas receberam subsídio de treinamento de pessoal em 1988. Se assumirmos que a destinação de subsídios é não-correlacionada com Δu_{it} — possibilidade razoável, pois os subsídios foram concedidos no início de 1988 —, $\Delta subs$ será uma VI

EXEMPLO 15.10 (continuação)

válida, desde que $\Delta hrsemp$ e $\Delta subs$ sejam correlacionados. Utilizando os dados contidos no arquivo JTRAIN.RAW, diferenciados entre 1987 e 1988, a regressão do primeiro estágio será

$$\begin{aligned}\Delta hrsemp &= 0,51 + 27,88 \Delta subs \\ (1,56) &\quad (3,13) \\ n = 45, R^2 &= 0,392.\end{aligned}$$

Isso confirma que a alteração nas horas de treinamento por trabalhador é forte e positivamente relacionada com o recebimento de um subsídio de treinamento de pessoal em 1988. Na verdade, o recebimento de um subsídio de treinamento de pessoal aumenta o treinamento por empregado em quase 28 horas, e a destinação do subsídio foi responsável por quase 40% da variação em $\Delta hrsemp$. A estimação por mínimos quadrados de dois estágios de (15.57) produz

$$\begin{aligned}\Delta \log(\text{ref}) &= -0,033 - 0,014 \Delta hrsemp \\ (0,127) &\quad (0,008) \\ n = 45, R^2 &= 0,016.\end{aligned}$$

Isso significa que dez horas a mais de treinamento por trabalhador reduziria a taxa de refúgio em cerca de 14%. Nas empresas da amostra, a média de horas de treinamento em 1988 foi cerca de 17 horas por trabalhador, com um mínimo de zero e um máximo de 88.

A título de comparação, a estimação por MQO de (15.57) produz $\hat{\beta}_1 = -0,0076$ ($ep = 0,0045$), de modo que a estimativa de β_1 por MQ2E é quase duas vezes maior em magnitude e é levemente mais significante, estatisticamente.

Quando $T \geq 3$, a equação diferenciada pode conter correlação serial. O mesmo teste e a mesma correção da correlação serial AR(1) da Seção 15.7 podem ser usados, onde todas as regressões serão agrupadas ao longo de i e também de t . Como não queremos perder um período de tempo, a transformação de Prais-Winsten deverá ser usada para o período de tempo inicial.

Modelos de efeitos não-observados que contenham variáveis dependentes defasadas também exigem métodos de VI para uma estimação consistente. A razão é que, após fazermos a diferenciação, a $\Delta y_{i,t-1}$ será correlacionado com Δu_{it} , pois $y_{i,t-1}$ e $u_{i,t-1}$ são correlacionados. Podemos usar duas ou mais defasagens de y como VIs de $\Delta y_{i,t-1}$. [Veja Wooldridge (2002, Capítulo 11) para detalhes.]

As variáveis instrumentais após a diferenciação também podem ser usadas em amostras de pares comparados. Ashenfelter e Krueger (1994) diferenciaram a equação de salários entre gêmeos para eliminar a aptidão não observada:

$$\log(\text{salário}_2) - \log(\text{salário}_1) = \delta_0 + \beta_1 (\text{educ}_{2,2} - \text{educ}_{1,1}) + (u_2 - u_1),$$

onde $educ_{1,1}$ são os anos de escolaridade do primeiro gêmeo como relatados por ele, e $educ_{2,2}$ são os anos de escolaridade do segundo gêmeo, como relatado por este. Para considerar a possibilidade de erro de medida nas indicações auto-informadas de escolaridade, Ashenfelter e Krueger usaram $(educ_{2,1} - educ_{1,2})$ como uma VI de $(educ_{2,2} - educ_{1,1})$, onde $educ_{2,1}$ representa os anos de escolaridade do segundo gêmeo como relatado pelo primeiro e $educ_{1,2}$ representa os anos de escolaridade do primeiro gêmeo como relatado pelo segundo gêmeo. A estimativa de VI de β_1 foi 0,167 ($t = 3,88$), comparada com a estimativa de MQO sobre as primeiras diferenças de 0,092 ($t = 3,83$) [veja Ashenfelter e Krueger (1994, Tabela 3)].

No Capítulo 15, apresentamos o método de variáveis instrumentais como uma maneira de estimar consistentemente os parâmetros em um modelo linear quando uma ou mais variáveis explicativas são endógenas. Uma variável instrumental deve ter duas propriedades: (1) ela deve ser exógena, isto é, não-correlacionada com o termo erro da equação estrutural; (2) ela deve ser parcialmente correlacionada com a variável explicativa endógena. Encontrar uma variável com essas duas propriedades em geral é desafiador.

O método dos mínimos quadrados de dois estágios, que possibilita o uso de um maior número de variáveis instrumentais do que o de variáveis explicativas que temos, é usado rotineiramente em ciências sociais empíricas. Quando usado adequadamente, ele pode nos permitir estimar efeitos *ceteris paribus* na presença de variáveis explicativas endógenas. Isso é verdadeiro em aplicações de corte transversal, séries temporais e dados de painel. Mas quando as variáveis instrumentais são pobres — o que significa que elas são correlacionadas com o termo erro, ou somente fracamente correlacionadas com a variável explicativa endógena, ou as duas coisas ao mesmo tempo —, então o MQ2E pode ser pior que o MQO.

Quando temos variáveis instrumentais válidas, podemos testar se uma variável explicativa é endógena, usando o teste da Seção 15.5. Além disso, embora nunca possamos verificar se todas as VIs são exógenas, podemos verificar se pelo menos algumas delas são — assumindo que temos mais variáveis instrumentais do que necessitamos para uma estimativa consistente (isto é, o modelo é sobreespecificado). A heteroscedasticidade e a correlação serial podem ser testadas e tratadas usando métodos semelhantes ao caso de modelos com variáveis explicativas exógenas.

Neste capítulo, usamos variáveis omitidas e erro de medida para ilustrar o método das variáveis instrumentais. Métodos de VI também são indispensáveis nos modelos de equações simultâneas, os quais veremos no Capítulo 16.

15.1 Considere um modelo simples para estimar o efeito da propriedade de um computador pessoal (PC) na nota média de graduação de formandos de uma grande universidade pública:

$$nmgrad = \beta_0 + \beta_1 PC + u,$$

onde PC é uma variável binária indicando a propriedade de um PC.

- (i) Por que a propriedade de um PC pode estar correlacionada com u ?

- (ii) Explique por que PC possivelmente está relacionada à renda anual dos pais. Isso significa que a renda dos pais será uma boa VI de PC ? Por quê?
- (iii) Suponha que, quatro anos atrás, a universidade tenha concedido subvenções para a compra de computadores a aproximadamente metade dos alunos novos, e que os alunos que receberam essas subvenções tenham sido escolhidos aleatoriamente. Explique cuidadosamente como você usaria essa informação para construir uma variável instrumental de PC .

15.2 Suponha que você queira estimar o efeito da frequência às aulas sobre o desempenho dos alunos, como no Exemplo 6.3. Um modelo básico é

$$respad = \beta_0 + \beta_1 taxafreq + \beta_2 nmgradp + \beta_3 tac + u,$$

onde as variáveis foram definidas no Capítulo 6.

- (i) Defina $dist$ como a distância da residência do aluno até o local de estudos. Você considera que $dist$ é não-correlacionado com u ?
- (ii) Assumindo que $dist$ e u sejam não-correlacionados, que outras hipóteses $dist$ deverá satisfazer para ser uma VI válida de $taxafreq$?
- (iii) Suponha, como na equação (6.18), que adicionemos o termo de interação $nmgradp \cdot taxafreq$:

$$respad = \beta_0 + \beta_1 taxafreq + \beta_2 nmgradp + \beta_3 tac + \beta_4 nmgradp \cdot taxafreq + u.$$

Se $taxafreq$ for correlacionado com u , então, em geral, $nmgradp \cdot taxafreq$ também será. O que poderia ser uma boa VI de $nmgradp \cdot taxafreq$? [Sugestão: Se $E(u|nmgradp, tac, dist) = 0$, como acontece quando $nmgradp$, tac e $dist$ são todas variáveis exógenas, então, qualquer função de $nmgradp$ e $dist$ será não-correlacionada com u .]

15.3 Considere o modelo de regressão simples

$$y = \beta_0 + \beta_1 x + u$$

e defina z como uma variável instrumental binária de x . Use (15.10) para mostrar que o estimador de VI $\hat{\beta}_1$ pode ser escrito como

$$\hat{\beta}_1 = (\bar{y}_1 - \bar{y}_0)/(\bar{x}_1 - \bar{x}_0),$$

onde \bar{y}_0 e \bar{x}_0 são as médias amostrais de y_i e x_i da parte da amostra com $z_i = 0$, e onde \bar{y}_1 e \bar{x}_1 são as médias amostrais de y_i e x_i da parte da amostra com $z_i = 1$. Esse estimador, conhecido como estimador agrupado, foi sugerido pela primeira vez por Wald (1940).

15.4 Suponha que você queira usar dados de séries temporais de determinado estado dos Estados Unidos, para estimar o efeito do salário mínimo em nível estadual sobre o emprego de pessoas entre 18 e 25 anos de idade (EMP). Um modelo simples é

$$gEMP_t = \beta_0 + \beta_1 gMIN_t + \beta_2 gPOP_t + \beta_3 gPEB_t + \beta_4 gPIB_t + u_t,$$

onde MIN_t é o salário mínimo, em dólares reais, POP_t é a população com idade entre 18 e 25 anos, PEB_t é o produto estadual bruto e PIB_t é o produto interno bruto norte-americano. O prefixo g indica a taxa de crescimento do ano $t - 1$ ao ano t , que em geral será aproximada pela diferença dos logs.

- (i) Se estivermos preocupados que o estado escolha seu salário mínimo parcialmente baseado em fatores não-observados (por nós) que afetem o emprego dos jovens, qual será o problema da estimação por MQO?
- (ii) Defina $SMAm$, como o salário mínimo dos Estados Unidos, que também é indicado em termos de dólares reais. Você acha que $gSMAm_t$ é não-correlacionado com u_t ?
- (iii) Por lei, qualquer salário mínimo estadual deve ser pelo menos igual ao salário mínimo nacional. Explique por que isso torna $gSMAm_t$ um candidato em potencial para ser uma VI de $gMIN_t$.

15.5 Retorne às equações (15.19) e (15.20). Assuma que $\sigma_u = \sigma_x$, de forma que a variação populacional no termo erro seja a mesma contida em x . Suponha que a variável instrumental, z , seja levemente correlacionada com u : $\text{Corr}(z, u) = 0,1$. Suponha também que z e x tenham uma correlação um pouco maior: $\text{Corr}(z, x) = 0,2$.

- (i) Qual será o viés assimptótico no estimador de VI?
- (ii) Quanta correlação deverá existir entre x e u antes que o MQO tenha mais viés assimptótico que o MQ2E?

15.6 (i) No modelo com uma variável explicativa endógena, uma variável explicativa exógena e uma variável exógena extra, considere a forma reduzida de y_2 , (15.26), inserindo-a na equação estrutural (15.22). Isso produzirá a forma reduzida de y_1 :

$$y_1 = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + v_1.$$

Encontre os α_j em termos de β_j e π_j .

- (ii) Encontre a forma reduzida do erro, v_1 , em termos de u_1 , v_2 e os parâmetros.
- (iii) Como você estimaria consistentemente os α_j ?

15.7 O que segue é um modelo simples para medir o efeito de um programa de escolha de escola sobre o desempenho em um teste padronizado [veja Rouse (1998), para entender a motivação do problema]:

$$\text{nota} = \beta_0 + \beta_1 \text{escolha} + \beta_2 \text{rendfam} + u_1,$$

onde nota é a nota em um teste de âmbito estadual, escolha é uma variável binária indicando se o aluno freqüentou uma escola de sua escolha no último ano e rendfam é a renda familiar.

A VI de escolha é conc , o montante em dólares concedido aos alunos para ser usado como pagamento da anuidade da escola particular de sua escolha. O montante da concessão difere conforme o nível da renda familiar, razão pela qual controlamos rendfam na equação.

- (i) Mesmo com rendfam na equação por que escolha pode ser correlacionado com u_1 ?
- (ii) Se no interior de cada classe de rendimento os montantes de concessão fossem atribuídos aleatoriamente, conc seria não-correlacionado com u_1 ?

- (iii) Escreva a forma reduzida da equação de *escolha*. O que é necessário para *conc* ser parcialmente correlacionado com *escolha*?
- (iv) Escreva a equação da forma reduzida de *nota*. Explique por que isso é importante. (*Sugestão*: Como você interpreta o coeficiente de *conc*?)

15.8 Suponha que você queira testar se meninas que freqüentam um escola de ensino médio só para meninas se saem melhor em matemática que as que freqüentam escolas mistas. Você tem uma amostra aleatória de meninas veteranas de escolas de ensino médio de um estado dos Estados Unidos, e *nota* é a nota de um teste padronizado de matemática. Defina *meninaem* como uma variável *dummy* indicando se uma aluna freqüenta uma escola de ensino médio só para meninas.

- (i) Que outros fatores você controlaria na equação? (Você deve ter condições razoáveis de coletar dados sobre esses fatores.)
- (ii) Escreva uma equação relacionando *nota* com *meninaem* e os outros fatores que você listou na parte (i).
- (iii) Suponha que o suporte e o incentivo dos pais sejam fatores não indicados no termo erro na parte (ii). É possível que eles sejam correlacionados com *meninaem*? Explique.
- (iv) Discuta as hipóteses necessárias para que o número de escolas do ensino médio só para meninas situadas em um raio de 20 milhas (aproximadamente 32 km) da residência de uma menina seja uma VI válida de *meninaem*.

15.9 Suponha que na equação (15.8) você não tenha uma boa candidata a variável instrumental de *faltas*. Entretanto, você tem duas outras informações sobre os alunos: a nota média ponderada de matemática e habilidade verbal do estudante para ingresso em curso superior (*sat*) e a nota média acumulada anterior ao semestre (*nmgradac*). O que você faria em vez da estimação de VI?

15.10 Em um artigo recente, Evans e Schwab (1995) estudaram os efeitos que freqüentar uma escola católica do ensino médio teriam sobre a probabilidade de cursar uma faculdade. Concretamente, defina *faculdade* como uma variável binária igual a um se o aluno estiver na faculdade, e zero caso contrário. Defina *EMCat* como uma variável binária igual a um se o aluno freqüenta uma escola católica do ensino médio. Um modelo de probabilidade linear é

$$\text{faculdade} = \beta_0 + \beta_1 \text{EMCat} + \text{outros fatores} + u,$$

onde, entre os outros fatores, estão sexo, raça, renda familiar e instrução dos pais.

- (i) Por que *EMCat* pode ser correlacionado com *u*?
- (ii) Evans e Schwab tinham dados sobre a nota de um teste padronizado feito quando cada estudante era aluno do 2º ano. O que pode ser feito com essa variável para melhorar a estimativa *ceteris paribus* de freqüentar uma escola católica do ensino médio?
- (iii) Defina *RelCat* como uma variável binária igual a um se o estudante for católico. Detalhe os dois requisitos necessários para que essa seja uma VI válida de *EMCat* na equação precedente. Qual deles pode ser testado?
- (iv) Não surpreendentemente, o fato de ser católico tem um efeito significante sobre freqüentar uma escola católica do ensino médio. Você julga que *RelCat* é uma variável instrumental convincente de *EMCat*?

15.11 Considere um modelo simples de séries temporais no qual a variável explicativa tem erro clássico de medida:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t^* + u_t \\ x_t &= x_t^* + e_t \end{aligned} \quad (15.58)$$

onde u_t tem média zero e é não-correlacionado com x_t^* e e_t . Observamos somente y_t e x_t . Assuma que e_t tem média zero e é não-correlacionado com x_t^* e que x_t^* também tem uma média zero (esta última hipótese é só para simplificar a álgebra).

- (i) Escreva $x_t^* = x_t - e_t$ e insira essa expressão em (15.58). Mostre que o termo erro na nova equação, digamos, v_t , é negativamente correlacionado com x_t se $\beta_1 > 0$. O que isso sugere sobre o estimador MQO de β_1 da regressão de y_t sobre x_t ?
- (ii) Além das hipóteses anteriores, assuma que u_t e e_t são não-correlacionados com todos os valores passados de x_t^* e e_t ; em particular com x_{t-1}^* e e_{t-1} . Mostre que $E(x_{t-1}, v_t) = 0$, onde v_t é o termo erro no modelo da parte (i).
- (iii) É possível que x_t e x_{t-1} sejam correlacionados? Explique.
- (iv) O que as partes (ii) e (iii) sugerem como uma estratégia vantajosa para estimarmos consistentemente β_0 e β_1 ?

Hipóteses do Método de Mínimos Quadrados em Dois Estágios

Este apêndice abrange as hipóteses sob as quais o MQ2E têm propriedades desejáveis de amostra grande. Primeiro, declaramos as hipóteses para as aplicações de corte transversal sob amostragem aleatória. Depois, discutimos o que precisa ser adicionado para que elas se apliquem a séries temporais e dados de painel.

H I P Ó T E S E M Q 2 E . 1 (LINEARIDADE NOS PARÂMETROS)

O modelo na população pode ser escrito como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + u,$$

onde $\beta_0, \beta_1, \dots, \beta_k$ são os parâmetros desconhecidos (constantes) de interesse, e u é um erro aleatório não observável ou termo de扰动 aleatório. As variáveis instrumentais são representadas como z_i .

H I P Ó T E S E M Q 2 E . 2 (AMOSTRAGEM ALEATÓRIA)

Temos uma amostra aleatória de y , x_j e z_j .

H I P Ó T E S E M Q 2 E . 3 (VARIÁVEIS INSTRUMENTAIS EXÓGENAS)

O termo erro u tem média zero, e cada VI é não-correlacionada com u .

Lembre-se de que qualquer x_j que seja não-correlacionado com u também age como uma VI.

H I P Ó T E S E M Q 2 E . 4 (CONDICÃO DE CLASSIFICAÇÃO)

(i) Não há relações lineares perfeitas entre as variáveis instrumentais. (ii) A condição de classificação da identificação se mantém.

Com uma única variável explicativa endógena, como na equação (15.42), a condição de classificação é facilmente descrita. Sejam z_1, \dots, z_m as variáveis exógenas, onde z_k, \dots, z_m não aparecem no modelo estrutural (15.42). A forma reduzida de y_2 é

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \dots + \pi_{k-1} z_{k-1} + \pi_k z_k + \dots + \pi_m z_m + v_2.$$

Então, necessitamos de que pelo menos um dos π_k, \dots, π_m seja diferente de zero. Isso exige pelo menos uma variável exógena que não aparece em (15.42) (a condição de ordem). Declarar a condição de classificação com duas ou mais variáveis explicativas endógenas exige álgebra matricial. [Veja Wooldridge (2002, Capítulo 5).]

T E O R E M A 15 A . 1

Sob as hipóteses MQ2E.1 a MQ2E.4, o estimador MQ2E é consistente.

H I P Ó T E S E M Q 2 E . 4 (HOMOSCEDASTICIDADE)

Seja z a coleção de todas as variáveis instrumentais. Então, $E(u^2|z) = \sigma^2$.

T E O R E M A 15 A . 2

Sob as hipóteses MQ2E.1 a MQ2E.5, os estimadores MQ2E são assintoticamente normalmente distribuídos. Estimadores consistentes da variância assintótica são dados como na equação (15.43), onde σ^2 é substituída por $\hat{\sigma}^2 = (n - k - 1)^2 \sum_{i=1}^n \hat{u}_i^2$ e os \hat{u}_i são os resíduos MQ2E.

O estimador MQ2E também é o melhor estimador de VI sob as cinco hipóteses dadas. Definimos o resultado aqui. Uma prova pode ser encontrada em Wooldridge (2002, Capítulo 5).

T E O R E M A 15 A . 3

Sob as hipóteses MQ2E.1 a MQ2E.5, o estimador MQ2E é assimptoticamente eficiente na classe de estimadores de VI que usa combinações lineares das variáveis exógenas como variáveis instrumentais.

Se a hipótese de homoscedasticidade não se sustentar, os estimadores MQ2E ainda assim serão assimptoticamente normais, mas os erros-padrão (e as estatísticas t e F) precisarão ser ajustados; muitos programas econométricos fazem isso rotineiramente. Além disso, em geral, o estimador MQ2E não mais será o estimador de VI assimptoticamente eficiente. Não estudaremos aqui estimadores mais eficientes. [Veja Wooldridge (2002, Capítulo 8).]

Em aplicações de séries temporais, devemos adicionar algumas hipóteses. Primeira, como no MQO, devemos assumir que todas as séries (inclusive as VIs) são fracamente dependentes: isso garante que a lei dos grandes números e o teorema do limite central sejam válidos. Para que os habituais erros-padrão e estatísticas de testes sejam válidos, como também para a eficiência assimptótica, devemos adicionar uma hipótese de ausência de correlação serial.

H I P Ó T E S E M Q 2 E . 6 (AUSÊNCIA DE CORRELAÇÃO SERIAL)

A equação (15.54) se mantém.

Uma hipótese semelhante de ausência de correlação serial é necessária em aplicações de dados de painel. Testes e correções de correlação serial foram examinados na Seção 15.7.

Modelos de Equações Simultâneas

No capítulo anterior, mostramos como o método das variáveis instrumentais pode solucionar dois tipos de problemas de endogeneidade: variáveis omitidas e erro de medida. Conceitualmente, esses problemas são claros. No caso de variáveis omitidas, existe uma variável (ou mais de uma) que gostaríamos de manter fixa quando estimarmos o efeito *ceteris paribus* de uma ou mais das variáveis explicativas observadas. No caso do erro de medida, gostaríamos de estimar o efeito de certas variáveis explicativas sobre y , mas medimos incorretamente uma ou mais variáveis. Em ambos os casos, poderíamos estimar os parâmetros de interesse por MQO se pudéssemos coletar dados melhores.

Outra forma importante de endogeneidade de variáveis explicativas é a **simultaneidade**. Ela surge quando uma ou mais das variáveis explicativas são *determinadas conjuntamente* com a variável dependente, em geral por meio de um mecanismo de equilíbrio (como veremos mais tarde). Neste capítulo, estudamos métodos de estimar modelos simples de equações simultâneas (MES). Embora um tratamento completo de MES esteja além do escopo desta obra, temos condições de abordar modelos que são amplamente usados.

O principal método para estimar modelos de equações simultâneas é o das variáveis instrumentais. Portanto, a solução dos problemas de simultaneidade é basicamente a mesma que a solução de VIs para os problemas de variáveis omitidas e erro de medida. Porém, elaborar e interpretar MES é um trabalho desafiador. Dessa forma, iniciamos examinando a natureza e o escopo de modelos de equações simultâneas na Seção 16.1. Na Seção 16.2, confirmamos que o MQO aplicado a uma equação em um sistema simultâneo é geralmente viesado e inconsistente.

A Seção 16.3 fornece uma descrição geral sobre identificação e estimação em um sistema de duas equações, enquanto a Seção 16.4 trata resumidamente de modelos com mais de duas equações. Modelos de equações simultâneas são usados para modelar séries temporais agregadas, e na Seção 16.5 incluímos uma discussão sobre alguns problemas especiais que surgem em tais modelos. A Seção 16.6 refere-se a modelos de equações simultâneas com dados de painel.

16.1 A NATUREZA DOS MODELOS DE EQUAÇÕES SIMULTÂNEAS

O ponto mais importante a lembrar no uso de modelos de equações simultâneas é que cada equação no sistema deve ter uma interpretação causal, *ceteris paribus*. Como somente observamos os resultados em equilíbrio, precisamos usar raciocínio contrafactual na construção de equações de um modelo de equações simultâneas. Deveremos pensar em termos de resultados potenciais assim como de resultados efetivos.

O exemplo clássico de um MES é uma equação de oferta e demanda de alguma mercadoria ou de algum insumo na produção (como a mão-de-obra). Concretamente, sejam h_s o total anual de horas supridas por trabalhadores na agricultura, medidas em nível municipal e w a média do salário por hora oferecida a tais trabalhadores. Uma função simples da oferta de mão-de-obra é

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1, \quad (16.1)$$

onde z_1 é alguma variável observada afetando a oferta de mão-de-obra — digamos, a média dos salários industriais no município. O termo erro, u_1 , contém outros fatores que afetam a oferta de mão-de-obra. [Muitos desses fatores são observados e poderiam ser incluídos na equação (16.1); para ilustrar os conceitos básicos, incluímos somente um de tais fatores, z_1 .] A equação (16.1) é um exemplo de uma **equação estrutural**. Esse nome tem origem no fato de ser a função de oferta de mão-de-obra derivável da teoria econômica e tem uma interpretação causal. O coeficiente α_1 indica como a oferta de mão-de-obra muda quando o salário muda; se h_s e w estiverem na forma logarítmica, α_1 será a elasticidade da oferta de mão-de-obra. Em geral, esperamos que α_1 seja positiva (embora a teoria econômica não impeça $\alpha_1 \leq 0$). As elasticidades da oferta de mão-de-obra são importantes na determinação de como os trabalhadores alterarão o número de horas que desejam trabalhar quando os impostos sobre os salários mudam. Se z_1 for o salário industrial, esperamos $\beta_1 \leq 0$: com outros fatores permanecendo iguais, se o salário industrial aumenta, mais trabalhadores irão para a indústria do que para a agricultura.

Quando fazemos o gráfico da oferta de mão-de-obra, descrevemos horas como uma função do salário, com z_1 e u_1 mantidos fixos. Uma alteração em z_1 , assim como uma mudança em u_1 , desloca a função de oferta de mão-de-obra. A diferença é que z_1 é observado, enquanto u_1 não é. Algumas vezes, z_1 é chamado de deslocador observado da oferta, e u_1 é chamado de *deslocador não-observado da oferta*.

Como a equação (16.1) difere das que estudamos anteriormente? A diferença é sutil. Embora a equação (16.1) pretendamente deva ser válida para todos os valores possíveis de salários, não podemos, de forma geral, ver os salários variando exogenamente em um corte transversal de municípios. Se pudéssemos computar um experimento no qual variássemos os níveis salariais industrial e agrícola ao longo de uma amostra de municípios e pesquisar os trabalhadores para obtermos a oferta de mão-de-obra h_s , então, poderíamos estimar (16.1) por MQO. Infelizmente, esse não é um experimento exequível. Em vez disso, temos que coletar dados sobre salários médios nesses dois setores com informações sobre quantas horas-homens foram empregadas na produção agrícola. Ao decidir como analisar esses dados, devemos entender que eles são melhor descritos pela interação entre a oferta e a demanda de mão-de-obra. Sob a hipótese de que os mercados de mão-de-obra compensam-se mutuamente, na verdade, observamos valores de *equilíbrio* de salários e horas trabalhadas.

Para descrever como os salários e horas de equilíbrio são determinados, necessitamos introduzir a demanda por mão-de-obra, que supomos ser dada por

$$h_d = \alpha_2 w + \beta_2 z_2 + u_2, \quad (16.2)$$

onde h_d representa horas demandadas. Como na função de oferta, escrevemos horas demandadas como uma função dos salários, w , mantendo z_2 e u_2 fixos. A variável z_2 , — digamos área agrícola — é um *deslocador observável da demanda*, enquanto u_2 é um *deslocador não-observável da demanda*.

Da mesma forma que na equação da oferta de mão-de-obra, a equação da demanda por mão-de-obra é uma equação estrutural: ela pode ser obtida a partir de considerações sobre a maximização de lucros dos fazendeiros. Se h_d e w estiverem em forma logarítmica, α_2 será a elasticidade da demanda

por mão-de-obra. A teoria econômica nos diz que $\alpha_2 < 0$. Como mão-de-obra e terra são complementares na produção, esperamos $\beta_2 > 0$.

Observe como as equações (16.1) e (16.2) descrevem relações totalmente diferentes. A oferta de mão-de-obra é uma equação comportamental dos trabalhadores, e a demanda por mão-de-obra é uma relação comportamental dos fazendeiros. Cada equação tem uma interpretação *ceteris paribus* e é auto-suficiente. Elas se tornam interligadas em uma análise econométrica somente porque salários e horas observados são determinados pela interseção da oferta e da demanda. Em outras palavras, em cada município i , as horas observadas h_i e os salários observados w_i são determinados pela condição de equilíbrio

$$h_{is} = h_{id}. \quad (16.3)$$

Como somente observamos horas de equilíbrio de cada município i , representamos horas observadas por h_i .

Quando combinamos a condição de equilíbrio em (16.3) com as equações de oferta e de demanda de mão-de-obra, obtemos

$$h_i = \alpha_1 w_i + \beta_1 z_{i1} + u_{i1} \quad (16.4)$$

e

$$h_i = \alpha_2 w_i + \beta_2 z_{i2} + u_{i2}, \quad (16.5)$$

onde explicitamente incluímos o subscrito i para enfatizar que h_i e w_i são os valores de equilíbrio observados de cada município. Essas duas equações constituem um **modelo de equações simultâneas (MES)**, que tem várias características importantes. Primeira, dadas z_{i1} , z_{i2} , u_{i1} e u_{i2} , essas duas equações determinam h_i e w_i . (Na realidade, devemos assumir que $\alpha_1 \neq \alpha_2$, o que significa que as inclinações das funções da oferta e da demanda diferem; veja o Problema 16.1.) Por essa razão, h_i e w_i são as **variáveis endógenas** nesse MES. O que dizer de z_{i1} e de z_{i2} ? Como elas são determinadas fora do modelo, as vemos como **variáveis exógenas**. Do ponto de vista estatístico, a hipótese-chave concernente a z_{i1} e z_{i2} é que ambas são não-correlacionadas com os erros da oferta e da demanda, u_{i1} e u_{i2} , respectivamente. Esses são exemplos de **erros estruturais** porque eles aparecem nas equações estruturais.

Um segundo ponto importante é que, sem a inclusão de z_1 e z_2 no modelo, não existe maneira de dizer qual das equações é a função de oferta e qual é a função de demanda. Quando z_1 representa salários industriais, o raciocínio econômico nos diz que ele é um fator na oferta de mão-de-obra agrícola, pois ele é uma indicação do custo da oportunidade de trabalhar na agricultura; quando z_2 representa a área agrícola, a teoria da produção sugere que ele apareça na função de demanda de mão-de-obra. Portanto, sabemos que (16.4) representa a oferta de mão-de-obra e (16.5) representa a demanda por mão-de-obra. Se z_1 e z_2 forem os mesmos — por exemplo, nível médio de educação dos adultos no município, que pode afetar tanto a oferta como a demanda —, as equações parecerão idênticas, e não há possibilidade de estimar qualquer uma delas.

Resumidamente, isso ilustra o problema de identificação em modelos de equações simultâneas, que examinaremos de forma mais geral na Seção 16.3.

Os exemplos mais convincentes de MES têm as mesmas formas dos exemplos de oferta e demanda. Cada equação deve ter uma interpretação comportamental própria, *ceteris paribus*. Como somente

observamos resultados de equilíbrio, a especificação de um MES exige que façamos perguntas contrafutais como: quanta mão-de-obra os trabalhadores ofereceriam se os salários fossem diferentes de seus valores de equilíbrio? O Exemplo 16.1 oferece outra ilustração de um MES na qual cada equação tem uma interpretação *ceteris paribus*.

EXEMPLO 16.1

(Taxa de Assassinatos e Tamanho da Força Policial)

As municipalidades freqüentemente querem determinar em que proporção a imposição da lei diminuirá suas taxas de assassinatos. Um modelo simples de corte transversal para tratar dessa questão é

$$assasp_{it} = \alpha_1 polpc_{it} + \beta_{10} + \beta_{11} rendapc_{it} + u_{1i} \quad (16.6)$$

onde $assasp_{it}$ representa assassinatos *per capita*, $polpc_{it}$ significa policiais *per capita* e $rendapc_{it}$ é a renda *per capita*. (Deste ponto em diante, não incluiremos um subscrito i .) Consideraremos a renda *per capita* como exógena nessa equação. Na prática, incluiríamos outros fatores, como as distribuições de idade e sexo, níveis de educação, talvez variáveis geográficas, e variáveis que indicassem a severidade da punição. Para organizar o raciocínio, consideramos a equação (16.6).

A questão que esperamos responder é: se uma cidade aumentar exogenousamente sua força policial, esse aumento, em média, reduzirá a taxa de assassinatos? Se pudermos escolher exogenousamente os tamanhos das forças policiais para uma amostra aleatória de cidades, poderíamos estimar (16.6) por MQO. Certamente, não podemos fazer tal experimento. Entretanto, podemos, de qualquer maneira, imaginar o tamanho da força policial como sendo exogenousamente determinada? Provavelmente, não. O gasto de uma cidade com a imposição da lei é pelo menos parcialmente determinado pela taxa de assassinatos antecipada. Para refletir isso, postulamos uma segunda relação:

$$polpc_{it} = \alpha_2 assasp_{it} + \beta_{20} + \text{outros fatores.} \quad (16.7)$$

Esperamos que $\alpha_2 > 0$: outros fatores sendo iguais, as cidades com taxas (esperadas) de homicídios mais elevadas terão mais policiais *per capita*. Assim que especificarmos os outros fatores em (16.7), teremos um modelo de equações simultâneas com duas equações. Na verdade, estamos interessados somente na equação (16.6), mas, como veremos na Seção 16.3, precisamos saber precisamente como a segunda equação é especificada para estimarmos a primeira.

Um ponto importante é que (16.7) descreve o comportamento dos policiais da cidade, enquanto (16.6) descreve as ações dos assassinos em potencial. Isso dá a cada equação uma clara interpretação *ceteris paribus*, o que faz das equações (16.6) e (16.7) um modelo de equações simultâneas apropriado.

A seguir damos um exemplo de uso inapropriado de MES.

EXEMPLO 16.2**{Despesas e Poupança Familiares}**

Suponha que, para uma família escolhida aleatoriamente na população, assumimos que os gastos e poupança familiares anuais sejam conjuntamente determinados por

$$gastofam = \alpha_1 poupfam + \beta_{10} + \beta_{11} renda + \beta_{12} educ + \beta_{13} idade + u_1 \quad (16.8)$$

e

$$poupfam = \alpha_2 gastofam + \beta_{20} + \beta_{21} renda + \beta_{22} educ + \beta_{23} idade + u_2, \quad (16.9)$$

onde *renda* é a renda anual e *educ* e *idade* são indicados em anos. Inicialmente, pode parecer que essas duas equações são uma maneira sensata de verificar como os gastos com habitação e poupança são determinados. Contudo, temos que perguntar: que valor teria uma dessas equações se é a outra? Nenhuma delas tem uma interpretação *ceteris paribus*, pois *gastofam* e *poupfam* são escolhidas pela mesma família. Por exemplo, não faz sentido fazer a pergunta: se a renda anual crescer em 10.000 dólares, como seriam alterados os gastos domésticos, mantendo a poupança fixa? Se a renda familiar aumentar, uma família geralmente alterará a composição ótima de gastos domésticos e poupança. Entretanto, a equação (16.8) faz parecer que queremos saber o efeito da alteração de *renda*, *educ* ou *idade*, mantendo a poupança fixa. Um experimento com esse enfoque não é interessante. Qualquer modelo baseado em princípios econômicos, particularmente a maximização da utilidade, teria a família feito a escolha ótima de *gastofam* e *poupfam* como funções da *renda* e dos preços relativos dos gastos domésticos e poupança. As variáveis *educ* e *idade* afetarão preferências de consumo, poupança e risco. Portanto, *gastofam* e *poupfam* serão cada uma função da renda, educação, idade e outras variáveis que afetem o problema da maximização da utilidade (tais como as diferentes taxas de retorno sobre gastos domiciliares e outras poupanças).

Mesmo que decidimos que os MES em (16.8) e (16.9) têm lógica, não há maneira de estimarmos os parâmetros. (Discutiremos esse problema de forma mais geral na Seção 16.3.) As duas equações são indistintas, a menos que assumamos que renda, educação ou idade apareçam em uma equação, mas não na outra, o que não faria sentido.

Embora esse seja um exemplo pobre do MES, podemos ter interesse em verificar se, com os outros fatores mantidos fixos, existe uma relação de substituição entre os gastos domésticos e a poupança. Contudo, nesse caso, estimariamos somente, digamos, (16.8) por MQO, a menos que haja um problema de variável omitida ou de erro de medida.

O Exemplo 16.2 tem as características de um grande número de aplicações MES. A característica-chave é que ambas as equações representam o comportamento do mesmo agente econômico, e assim nenhuma das equações depende de si mesma. Em contraposição, os exemplos de oferta e demanda e o Exemplo 16.1 têm interpretações *ceteris paribus* naturais. O raciocínio econômico básico, apoiado em alguns casos por modelos econômicos simples, pode nos ajudar a usar os MES de forma inteligente (e saber quando não usar o MES).

Pindyck e Rubinfeld (1992, Seção 11.6) descrevem um modelo de publicidade no qual firmas monopolistas escolhem níveis de preços e gastos com publicidade que maximizam lucros. Isso significa que deveríamos usar um MES para modelar essas variáveis no nível da firma?

16.2 VIÉS DE SIMULTANEIDADE NO MQO

É proveitoso ver, em um modelo simples, que uma variável explicativa que é determinada simultaneamente com a variável dependente geralmente é correlacionada com o termo erro, o que conduz a viés e inconsistência no MQO. Consideremos o modelo estrutural de duas equações

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1 \quad (16.10)$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2 \quad (16.11)$$

e nos concentremos em estimar a primeira equação. As variáveis z_1 e z_2 são exógenas, de forma que cada uma é não-correlacionada com u_1 e u_2 . Por simplicidade, suprimimos o intercepto em cada equação.

Para mostrar que y_2 geralmente é correlacionada com u_1 , solucionamos as duas equações para y_2 em termos das variáveis exógenas e do termo erro. Se inserirmos o lado direito de (16.10) em (16.11), obteremos

$$y_2 = \alpha_2(\alpha_1 y_2 + \beta_1 z_1 + u_1) + \beta_2 z_2 + u_2$$

ou

$$(1 - \alpha_2\alpha_1)y_2 = \alpha_2\beta_1 z_1 + \beta_2 z_2 + \alpha_2 u_1 + u_2. \quad (16.12)$$

Agora, devemos fazer uma hipótese sobre os parâmetros para solucionar a equação para y_2 :

$$\alpha_2\alpha_1 \neq 1. \quad (16.13)$$

Com relação a essa hipótese ser restritiva, depende da aplicação. No Exemplo 16.1, entendemos que $\alpha_1 \leq 0$ e $\alpha_2 \geq 0$, o que implica $\alpha_1\alpha_2 \leq 0$; portanto, (16.13) é bastante razoável para o Exemplo 16.1.

Desde que a condição (16.13) se mantenha, podemos dividir (16.12) por $(1 - \alpha_2\alpha_1)$ e escrever y_2 como

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2, \quad (16.14)$$

onde $\pi_{21} = \alpha_2\beta_1/(1 - \alpha_2\alpha_1)$, $\pi_{22} = \beta_2/(1 - \alpha_2\alpha_1)$, e $v_2 = (\alpha_2u_1 + u_2)/(1 - \alpha_2\alpha_1)$. A equação (16.14), que expressa y_2 em termos das variáveis exógenas e dos termos erros, é a **forma reduzida** de y_2 , um conceito apresentado no Capítulo 15 no contexto da estimativa de variáveis instrumentais. Os parâmetros π_{21} e π_{22} são chamados **parâmetros da forma reduzida**; observe como eles são funções não-lineares dos **parâmetros estruturais**, que aparecem nas equações estruturais, (16.10) e (16.11).

O **erro da forma reduzida**, v_2 , é uma função linear dos termos erros estruturais, u_1 e u_2 . Como u_1 e u_2 são, individualmente, não-correlacionados com z_1 e z_2 , v_2 também é não-correlacionado com z_1 e z_2 . Portanto, podemos consistentemente estimar π_{21} e π_{22} por MQO, algo que é usado para a estimativa por mínimos quadrados em dois estágios (ao qual retornaremos na próxima seção). Além disso, os parâmetros da forma reduzida são algumas vezes de interesse direto, embora estejamos, aqui, nos concentrando em estimar a equação (16.10).

Também existe uma forma reduzida de y_1 sob a hipótese (16.13); a álgebra é semelhante à usada para obter (16.14). Ela tem as mesmas propriedades da forma reduzida da equação de y_2 .

Podemos usar a equação (16.14) para mostrar que, exceto sob hipóteses especiais, a estimativa por MQO da equação (16.10) produzirá estimadores de α_1 e β_1 viesados e inconsistentes na equação (16.10). Como z_1 e u_1 são não-correlacionados por hipótese, o problema está em saber se y_2 e u_1 são não-correlacionados. A partir da forma reduzida em (16.14), vemos que y_2 e u_1 serão correlacionados se, e somente se, v_2 e u_1 forem correlacionados (pois z_1 e z_2 são assumidos como exógenos). Porém, v_2 é uma função linear de u_1 e u_2 , de modo que ele geralmente é correlacionado com u_1 . Na verdade, se assumirmos que u_1 e u_2 são não-correlacionados, v_2 e u_1 devem ser correlacionados sempre que $\alpha_2 \neq 0$. Mesmo se α_2 for igual a zero — o que significa que y_1 não aparece na equação (16.11) —, v_2 e u_1 serão correlacionados se u_1 e u_2 forem correlacionados.

Quando $\alpha_2 = 0$ e u_1 e u_2 forem não-correlacionados, y_2 e u_1 também serão não-correlacionados. Esses são requisitos bastante fortes: se $\alpha_2 = 0$, y_2 não será simultaneamente determinado com y_1 . Se adicionarmos correlação zero entre u_1 e u_2 , isso eliminará variáveis omitidas ou erro de medida em u_1 que sejam correlacionados com y_2 . Não devemos nos surpreender com o fato de que a estimativa por MQO da equação (16.10) funciona nesse caso.

Quando y_2 for correlacionado com u_1 devido à simultaneidade, dizemos que o MQO sofre de **viés de simultaneidade**. A obtenção da direção do viés nos coeficientes é geralmente complicada, como vimos com o viés de variáveis omitidas nos Capítulos 3 e 5. Contudo, em modelos simples, podemos determinar a direção do viés. Por exemplo, suponha que simplifiquemos a equação (16.10) retirando z_1 da equação e assumindo que u_1 e u_2 são não-correlacionados. Então, a covariância entre y_2 e u_1 será

$$\text{Cov}(y_2, u_1) = \text{Cov}(v_2, u_1) = [\alpha_2/(1 - \alpha_2\alpha_1)]E(u_1^2) = [\alpha_2/(1 - \alpha_2\alpha_1)]\sigma_1^2,$$

onde $\sigma_1^2 = \text{Var}(u_1) > 0$. Portanto, o viés assimptótico (ou a inconsistência) no estimador MQO de α_1 terá o mesmo sinal de $\alpha_2/(1 - \alpha_2\alpha_1)$. Se $\alpha_2 > 0$ e $\alpha_2\alpha_1 < 1$, o viés assimptótico será positivo. (Infelizmente, como no caso de nosso cálculo do viés de variáveis omitidas da Seção 3.3, as conclusões não são transportadas para modelos mais gerais. Porém, elas servem como um guia útil.) Por exemplo, no Exemplo 16.1, pensamos que $\alpha_2 > 0$ e $\alpha_2\alpha_1 \leq 0$, o que significa que o estimador MQO de α_1 teria um viés positivo. Se $\alpha_1 = 0$, o MQO estimará, em média, um impacto *positivo* de mais policiais sobre a taxa de assassinatos; geralmente, o estimador de α_1 é viesado para cima. Como esperamos um aumento no tamanho da força policial para reduzir as taxas de criminalidade (*ceteris paribus*), o viés para cima significa que o MQO subestimaré a efetividade de uma força policial maior.

16.3 A IDENTIFICAÇÃO E A ESTIMAÇÃO DE UMA EQUAÇÃO ESTRUTURAL

Como vimos na Seção anterior, o MQO é viesado e inconsistente quando aplicado a uma equação estrutural em um sistema de equações simultâneas. No Capítulo 15, aprendemos que o método dos mínimos quadrados em dois estágios pode ser usado para solucionar o problema de variáveis explicativas endógenas. Agora mostramos como o MQ2E pode ser aplicado a MES.

A mecânica do MQ2E é semelhante à do Capítulo 15. A diferença é que, como especificamos uma equação estrutural para cada variável endógena, podemos imediatamente verificar se existem VIs suficientes para estimar qualquer equação. Iniciamos discutindo o problema da identificação.

A Identificação em um Sistema de Duas Equações

Mencionamos a noção de identificação no Capítulo 15. Quando estimamos um modelo por MQO, a condição crucial de identificação é que cada variável explicativa seja não-correlacionada com o termo erro. Como demonstramos na Seção 16.2, de forma geral, essa condição fundamental não se mantém, para os MES. Porém, se tivermos algumas variáveis instrumentais, poderemos ainda identificar (ou estimar consistentemente) os parâmetros em uma equação MES, da mesma forma que com variáveis omitidas ou erro de medida.

Antes de considerarmos um MES geral de duas equações, é útil adquirirmos conhecimento intuitivo ao considerarmos um exemplo simples de oferta e demanda. Escreva o sistema na forma de equilíbrio (isto é, impondo $q_s = q_d = q$) como

$$q = \alpha_1 p + \beta_1 z_1 + u_1 \quad (16.15)$$

ou

$$q = \alpha_2 p + u_2. \quad (16.16)$$

Concretamente, sejam q o consumo *per capita* de leite em nível municipal, p o preço médio por galão de leite no município e z_1 o preço da alimentação do gado, que assumimos ser exógeno nas equações de oferta e demanda de leite. Isso significa que (16.15) deve ser a função de oferta, já que o preço da alimentação do gado deslocará a oferta ($\beta_1 < 0$), mas não a demanda. A função de demanda não contém deslocadores observados da demanda.

Dada uma amostra aleatória de (q, p, z_1) , qual dessas equações será estimada? Isto é, qual delas é uma equação identificada? É possível constatar que a equação de *demand* (16.16) é identificada, mas a equação da *oferta* não é. Isso é fácil de verificar usando as regras de estimação de VI do Capítulo 15: podemos usar z_1 como uma VI do preço na equação (16.16). Porém, como z_1 aparece na equação (16.15), não temos uma VI do preço na equação de oferta.

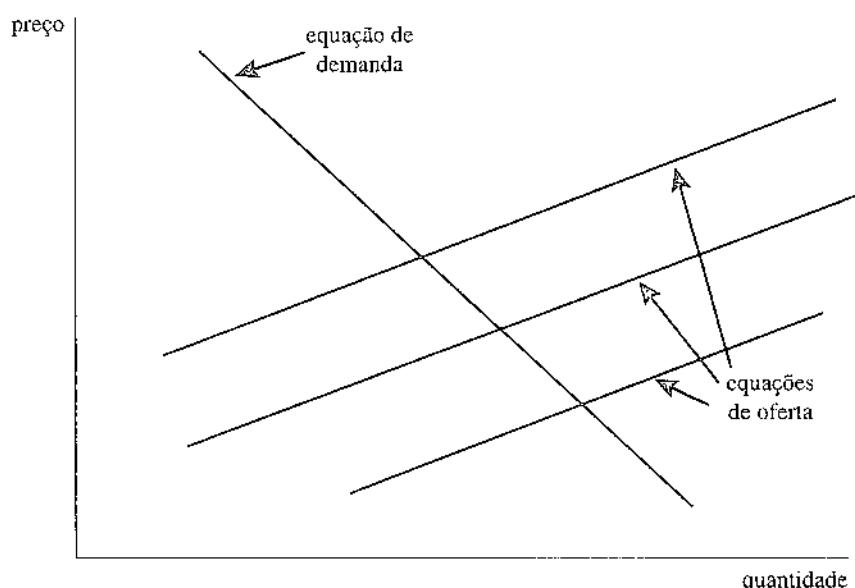
Intuitivamente, o fato de a equação de demanda ser identificada é uma consequência de termos uma variável observada, z_1 , que desloca a equação de oferta sem afetar a equação de demanda. Dada uma variação em z_1 e nenhum erro, podemos desenhar a curva de demanda, como mostrado na Figura 16.1. A presença do deslocador não observado da demanda u_2 faz com que estimemos a equação de demanda com erro, mas os estimadores serão consistentes, desde que z_1 seja não-correlacionado com u_2 .

A equação da oferta não pode ser desenhada porque não existem fatores não observados exógenos deslocando a curva de demanda. Não ajuda o fato de haver fatores não observados deslocando a função de demanda; necessitamos de algo observado. Se, como na função de demanda da mão-de-obra

(16.2), tivéssemos um deslocador observado da demanda exógeno — como a renda na função de demanda do leite —, a função de oferta também poderia ser identificada.

Figura 16.1

O deslocamento nas equações de oferta permite desenhar a equação de demanda. Cada equação de oferta é traçada para um valor diferente da variável exógena, z_1 .



Resumindo: no sistema de (16.15) e (16.16), a presença de uma variável exógena na equação da oferta é que nos possibilita estimar a equação de demanda.

Estender a discussão sobre a identificação a um modelo geral de duas equações não apresenta dificuldades. Escreva as duas equações como

$$y_1 = \beta_{10} + \alpha_1 y_2 + z_1 \beta_1 + u_1 \quad (16.17)$$

e

$$y_2 = \beta_{20} + \alpha_2 y_1 + z_2 \beta_2 + u_2, \quad (16.18)$$

onde y_1 e y_2 são as variáveis endógenas e u_1 e u_2 são os termos erros estruturais. O intercepto na primeira equação é β_{10} , e o intercepto na segunda equação é β_{20} . A variável z_1 representa um conjunto de k_1 variáveis exógenas aparecendo na primeira equação: $z_1 = (z_{11}, z_{12}, \dots, z_{1k_1})$. De forma semelhante, z_2 é o conjunto de k_2 variáveis exógenas na segunda equação: $z_2 = (z_{21}, z_{22}, \dots, z_{2k_2})$. Em muitos casos, z_1 e z_2 se sobreporão. De forma abreviada, usamos a notação

$$z_1\beta_2 = \beta_{11}z_{11} + \beta_{12}z_{12} + \dots + \beta_{1k}z_{1k},$$

ou

$$z_2\beta_2 = \beta_{21}z_{21} + \beta_{22}z_{22} + \dots + \beta_{2k_2}z_{2k_2};$$

isto é, $z_1\beta_2$ representa todas as variáveis exógenas na primeira equação, cada uma multiplicada por um coeficiente, e semelhantemente $z_2\beta_2$. (Alguns autores usam a notação $z'_1\beta_1$ e $z'_2\beta_2$. Se você tiver interesse na abordagem de álgebra matricial em econometria, veja o Apêndice E, disponível na página do livro, no site www.thomsonlearning.com.br.)

O fato de que z_1 e z_2 geralmente contêm variáveis exógenas diferentes significa que impusemos **restrições de exclusão** no modelo. Em outras palavras, assumimos que certas variáveis exógenas não aparecem na primeira equação e outras estão ausentes da segunda equação. Como vimos nos exemplos anteriores de oferta e demanda, isso nos possibilita distinguir entre duas equações estruturais.

Quando podemos solucionar as equações (16.17) e (16.18) para y_1 e y_2 (como funções lineares de todas as variáveis exógenas e dos erros estruturais u_1 e u_2)? A condição é a mesma que em (16.13), ou seja, $\alpha_2\alpha_1 \neq 1$. A prova é virtualmente idêntica à do modelo simples da Seção 16.2. Sob essa hipótese, existirão formas reduzidas para y_1 e y_2 .

A pergunta principal é: sob quais hipóteses podemos estimar os parâmetros em, digamos, (16.17)? Esse é o problema da identificação. A **condição de classificação** para a identificação da equação (16.17) é fácil de estabelecer.

CONDIÇÃO DE CLASSIFICAÇÃO PARA A IDENTIFICAÇÃO DE UMA EQUAÇÃO ESTRUTURAL

A primeira equação em um modelo de equações simultâneas com duas equações será identificada se, e somente se, a *segunda* equação contiver ao menos uma variável exógena (com um coeficiente diferente de zero) que seja excluída da primeira equação.

Essa é a condição necessária e suficiente para que a equação (16.17) seja identificada. A **condição de ordem**, que discutimos no Capítulo 15, é necessária para a condição de classificação. A condição de ordem para a identificação da primeira equação estabelece que pelo menos uma variável exógena seja excluída dessa equação. A condição de ordem é fácil de ser verificada, uma vez que ambas as equações tenham sido especificadas. A condição de classificação exige mais: pelo menos uma das variáveis exógenas excluídas da primeira equação deve ter um coeficiente populacional diferente de zero na segunda equação. Isso garante que pelo menos uma das variáveis exógenas omitidas da primeira equação efetivamente apareça na forma reduzida de y_2 , de forma a podermos usar essas variáveis como instrumentais de y_2 . Podemos verificar isso usando um teste t ou F , como no Capítulo 15; seguem alguns exemplos.

A identificação da segunda equação é, naturalmente, apenas a imagem espelhada da declaração para a primeira equação. Além disso, se escrevermos as equações como no exemplo da oferta e demanda de mão-de-obra da Seção 16.1 — de forma que y_1 apareça no lado esquerdo em *ambas* as equações, com y_2 no lado direito —, as condições de identificação serão idênticas.

EXEMPLO 16.3**(Oferta de Mão-De-Obra de Mulheres Casadas que Trabalham)**

Para ilustrar o problema da identificação, considere a oferta de mão-de-obra de mulheres casadas que já estejam na força de trabalho. Em lugar da função de demanda, escrevemos a oferta de salários como uma função de horas e das variáveis de produtividade habituais. Com a condição de equilíbrio imposta, as duas equações estruturais serão

$$\begin{aligned} \text{horas} = & \alpha_1 \log(\text{salário}) + \beta_{10} + \beta_{11} \text{educ} + \beta_{12} \text{idade} + \beta_{13} \text{crianmed6} \\ & + \beta_{14} \text{nesprend} + u_1 \end{aligned} \quad (16.19)$$

e

$$\text{horas}(\text{salário}) = \alpha_2 \text{horas} + \beta_{20} + \beta_{21} \text{educ} + \beta_{22} \text{exper} + \beta_{23} \text{exper}^2 + u_2 \quad (16.20)$$

A variável *idade* é a idade da mulher, em anos, *crianmed6* é o número de filhos menores de seis anos de idade, *nesprend* é a renda de outra pessoa da família que não a mulher (que inclui os ganhos do marido), e *educ* e *exper* são anos de educação e de experiência anterior, respectivamente. Todas as variáveis, com exceção de *horas* e $\log(\text{salário})$ são assumidas como exógenas. (Essa é uma hipótese fraca, já que *educ* pode ser correlacionado com a aptidão omitida em cada uma das equações. Mas com o propósito de ilustração, ignoramos o problema da aptidão omitida.) A forma funcional nesse sistema — no qual *horas* aparece na forma de nível, mas *salário* está na forma logarítmica — é comum em economia do trabalho. Podemos escrever esse sistema como nas equações (16.17) e (16.18), definindo $y_1 = \text{horas}$ e $y_2 = \log(\text{salário})$.

A primeira equação é a função de oferta. Ela satisfaz a condição de ordem porque duas variáveis exógenas, *exper* e exper^2 , são omitidas da equação de oferta de mão-de-obra. Essas restrições de exclusão são hipóteses cruciais: assumimos que, uma vez que salário, educação, idade, número de filhos pequenos e outras rendas sejam controlados, a experiência passada não tem efeito na oferta corrente de mão-de-obra. Certamente poderia se questionar essa hipótese, mas nós a usamos a título de ilustração.

Dadas as equações (16.19) e (16.20), a condição de classificação para identificar a primeira equação é que pelo menos uma das variáveis *exper* ou exper^2 tenha um coeficiente diferente de zero na equação (16.20). Se $\beta_{22} = 0$ e $\beta_{23} = 0$, não haverá variáveis exógenas aparecendo na segunda equação que também não apareçam na primeira equação (*educ* aparece em ambas). Podemos estabelecer a condição de classificação para a identificação de (16.19) em equivalência com os termos da forma reduzida de $\log(\text{salário})$, que é

$$\begin{aligned} \log(\text{salário}) = & \pi_{20} + \pi_{21} \text{educ} + \pi_{22} \text{idade} + \pi_{23} \text{crianmed6} \\ & + \pi_{24} \text{nesprend} + \pi_{25} \text{exper} + \pi_{26} \text{exper}^2 + v_2. \end{aligned} \quad (16.21)$$

Para a identificação, necessitamos que $\pi_{25} \neq 0$ ou $\pi_{26} \neq 0$, o que podemos testar usando uma estatística *F* padrão, como discutimos no Capítulo 15.

A equação de oferta de salário, (16.20), será identificada se pelo menos uma das variáveis *idade*, *crianmed6* ou *nesprend* tiver um coeficiente diferente de zero na equação (16.19). Isso é a mesma coisa que assumir que a forma reduzida de *horas* — que tem a mesma forma do lado direito de (16.21) — depende

EXEMPLO 16.3 (continuação)

de pelo menos uma das variáveis *idade*, *crianmed6* ou *nesprend*. Na especificação da equação de oferta de salário, estamos assumindo que *idade*, *crianmed6* ou *nesprend* não têm efeito sobre a oferta de salário, uma vez que horas, educação e experiência sejam levadas em conta. Essas serão hipóteses pobres se essas variáveis de alguma maneira tiverem efeitos diretos sobre a produtividade, ou se as mulheres forem discriminadas com base em sua idade ou número de filhos pequenos.

No Exemplo 16.3, consideramos a população de interesse como sendo mulheres casadas que estejam na força de trabalho (de forma que horas em equilíbrio são positivas). Isso exclui o grupo de mulheres casadas que escolheram não trabalhar fora de casa. A inclusão de tais mulheres no modelo provocaria alguns problemas intrincados. Por exemplo, se uma mulher não trabalha, não poderemos observar sua oferta de salário. Abordaremos brevemente esses problemas no Capítulo 17; mas, por enquanto, temos que pensar nas equações (16.19) e (16.20) como válidas somente para mulheres que tenham *horas* > 0.

EXEMPLO 16.4**(Inflação e Abertura da Economia)**

Romer (1993) propõe modelos teóricos de inflação que sugerem que países mais "abertos" devem ter taxas de inflação mais baixas. Sua análise empírica interpreta taxas médias anuais de inflação (desde 1973) em termos da participação média das importações no produto interno (ou nacional) bruto desde 1973 — que é sua medida de abertura da economia. Além de estimar a equação-chave por MQO, ele usa variáveis instrumentais. Embora Romer não especifique ambas as equações em um sistema simultâneo, ele tem em mente um sistema de duas equações:

$$\text{inf} = \beta_{10} + \alpha_1 \text{abertura} + \beta_{11} \log(\text{rendpc}) + u_1 \quad (16.22)$$

$$\text{abertura} = \beta_{20} + \alpha_2 \text{inf} + \beta_{21} \log(\text{rendpc}) + \beta_{22} \log(\text{área}) + u_2, \quad (16.23)$$

onde *rendpc* é a renda *per capita* de 1980, em dólares dos Estados Unidos (assumida como exógena), e *área* é a área do país, em milhas quadradas (também assumida como exógena). A equação (16.22) é a de interesse, com a hipótese de que $\alpha_1 < 0$. (Economias mais abertas têm menores taxas de inflação.) A segunda equação reflete o fato de que o grau da abertura pode depender da taxa de inflação, como também de outros fatores. A variável $\log(\text{rendpc})$ aparece em ambas as equações, mas $\log(\text{área})$ aparece supostamente somente na segunda equação. A idéia é que, *ceteris paribus*, um país menor provavelmente será mais aberto (portanto, $\beta_{22} < 0$).

Usando a regra de identificação que foi declarada anteriormente, a equação (16.22) será identificada, desde que $\beta_{22} \neq 0$. A equação (16.23) não é identificada porque contém ambas as variáveis exógenas. Mas estamos interessados em (16.22).

Se tivermos o crescimento da oferta de moeda desde 1973 de cada país, que assumimos ser exógeno, isso auxiliará a identificar a equação (16.23)?

Estimação por MQ2E

Uma vez que tenhamos determinado que uma equação é identificada, podemos estimá-la por mínimos quadrados em dois estágios. As variáveis instrumentais consistirão das variáveis exógenas que aparecem em cada equação.

EXEMPLO 16.5

(Oferta de Mão-De-Obra de Mulheres Casadas que Trabalham)

Utilizamos os dados sobre mulheres casadas que trabalham contidos no arquivo MROZ.RAW para estimar a equação da oferta de mão-de-obra (16.19) por MQ2E. No conjunto total de variáveis instrumentais estão incluídas *educ*, *idade*, *crianmed6*, *nesprend*, *exper* e *exper²*. A curva da oferta de mão-de-obra é

$$\begin{aligned} \hat{\text{horas}} = & 2.225,66 + 1.639,56 \log(\text{salário}) - 183,75 \text{ educ} \\ & (574,56) \quad (470,58) \quad (59,10) \\ & -7,81 \text{ idade} - 198,15 \text{ crianmed6} - 10,17 \text{ nesprend}, n = 428, \\ & (9,38) \quad (182,93) \quad (6,61) \end{aligned} \tag{16.24}$$

que mostra que a curva da oferta de mão-de-obra tem inclinação para cima. O coeficiente estimado de $\log(\text{salário})$ tem a seguinte interpretação: mantendo fixos os outros fatores, $\Delta\text{horas} \approx 16,4(\% \Delta\text{salário})$. Podemos calcular as elasticidades da oferta de mão-de-obra multiplicando ambos os lados dessa equação por $100/\text{horas}$:

$$100 \cdot (\Delta\text{horas}/\text{horas}) \approx (1.640/\text{horas})(\% \Delta\text{salário})$$

ou

$$\% \Delta\text{horas} \approx (1.640/\text{horas})(\% \Delta\text{salário}),$$

que implica que a elasticidade da oferta de mão-de-obra (com relação a salário) é simplesmente $1.640/\text{horas}$. [A elasticidade não é constante nesse modelo porque *horas*, e não $\log(\text{horas})$, é a variável dependente em (16.24).] Na média de horas trabalhadas, 1.303, a elasticidade estimada é $1.640/1.303 \approx 1,26$, implicando um aumento maior que 1% nas horas trabalhadas, dado um aumento de 1% no salário. Essa é uma grande elasticidade estimada. Com maior número de horas, a elasticidade será menor; com menor número de horas, como *horas* = 800, a elasticidade é maior que dois.

Comparativamente, quando (16.19) é estimada por MQO, o coeficiente de $\log(\text{salário})$ é $-2,05$ ($t_p = 54,88$), o que implica não haver nenhum efeito da oferta de mão-de-obra sobre as horas trabalhadas. Para confirmar que $\log(\text{salário})$ é de fato endógeno em (16.9), podemos aplicar o teste da Seção 15.5. Quando

EXEMPLO 16.5 (continuação)

adicionamos os resíduos da forma reduzida \hat{v}_2 na equação e a estimamos por MQO, a estatística t de \hat{v}_2 é $-6,61$, que é muito significante e, portanto, $\log(\text{salário})$ parece ser endógeno.

A equação da oferta de salário (16.20) também pode ser estimada por MQ2E. O resultado será

$$\begin{aligned} \log(\text{salário}) &= -0,656 + 0,00013 \text{ horas} + 0,110 \text{ educ} \\ &\quad (0,338) \quad (0,00025) \quad (0,016) \\ &+ 0,035 \text{ exper} - 0,00071 \text{ exper}^2, n = 428. \\ &\quad (0,019) \quad (0,00045) \end{aligned} \tag{16.25}$$

Essa equação difere das equações de salários anteriores pelo fato de que *horas* é incluída como uma variável explicativa e o MQ2E são usados para levar em conta a endogeneidade de *horas* (e assumimos que *educ* e *exper* são exógenos). O coeficiente de *horas* é estatisticamente não significante, o que significa que não existe evidência de que a oferta de salário cresce com as horas trabalhadas. Os outros coeficientes são semelhantes aos que obteremos se eliminarmos *horas* e estimarmos a equação por MQO.

A estimativa do efeito da abertura sobre a inflação por variáveis instrumentais também é feita de forma direta.

EXEMPLO 16.6**(Inflação e Abertura da Economia)**

Antes de estimar (16.22) usando os dados contidos no arquivo OPENNESS.RAW, fazemos uma verificação para ver se *abertura* tem correlação parcial suficiente com a VI proposta, $\log(\text{área})$. A regressão da forma reduzida é

$$\begin{aligned} \hat{\text{abertura}} &= 117,08 + 0,546 \log(\text{rendpc}) - 7,57 \log(\text{área}) \\ &\quad (15,85) \quad (1,493) \quad (0,81) \\ &n = 114, R^2 = 0,449. \end{aligned}$$

A estatística t de $\log(\text{área})$ é maior que nove, em valor absoluto, o que verifica a assertiva de Romer, de que países menores são mais abertos economicamente. O fato de que $\log(\text{rendpc})$ seja tão não significante nessa regressão é irrelevante.

A estimativa de (16.22) usando $\log(\text{área})$ como uma VI de *abertura* produz

$$\hat{\text{inf}} = 26,90 - 0,337 \text{ abertura} + 0,376 \log(\text{rendpc}), \quad n = 114. \tag{16.26}$$

(15,40)	(0,144)	(2,015)
---------	---------	---------

O coeficiente de *abertura* é estatisticamente significante ao nível de aproximadamente 1% contra uma alternativa bilateral ($\alpha_1 < 0$). O efeito também é economicamente importante: para cada ponto percentual de aumento na participação das importações no PIB, a inflação anual será de um terço de ponto percentual mais baixa. A título de comparação, a estimativa por MQO é $-0,215$ ($\text{ep} = 0,095$).

Como é possível testar se a diferença entre as estimativas por MQO e por VI de *abertura* é estatisticamente diferente?

16.4 SISTEMAS COM MAIS DE DUAS EQUAÇÕES

Modelos de equações simultâneas podem consistir de mais de duas equações. O estudo da identificação geral desses modelos é difícil e requer álgebra matricial. Uma vez que uma equação em um sistema geral tenha sido identificada, ela poderá ser estimada por MQ2E.

Identificação em Sistemas com três ou mais Equações

Usaremos um sistema de três equações para ilustrar os problemas que surgem na identificação de MES complicados. Com interceptos suprimidos, escreva o modelo como

$$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}z_1 + u_1 \quad (16.27)$$

$$y_2 = \alpha_{21}y_1 + \beta_{21}z_1 + \beta_{22}z_2 + \beta_{23}z_3 + u_2 \quad (16.28)$$

$$y_3 = \alpha_{32}y_2 + \beta_{31}z_1 + \beta_{32}z_2 + \beta_{33}z_3 + \beta_{34}z_4 + u_3, \quad (16.29)$$

onde os y_g são as variáveis endógenas, e os z_j são exógenas. O primeiro subscrito nos parâmetros indica o número da equação, enquanto o segundo indica o número da variável; usamos α para os parâmetros das variáveis endógenas e β para os parâmetros das variáveis exógenas.

Quais dessas equações podem ser estimadas? Mostrar que uma equação em um MES com mais de duas equações é identificada geralmente é difícil, mas é fácil verificar quando certas equações *não* são identificadas. No sistema (16.27) a (16.29), podemos verificar facilmente que (16.29) cai nessa categoria. Como cada variável exógena aparece nessa equação, não temos qualquer VI de y_2 . Portanto, não podemos consistentemente estimar os parâmetros dessa equação. Pelas razões discutidas na Seção 16.2, a estimativa por MQO normalmente não será consistente.

E quanto à equação (16.27)? Aqui o assunto parece promissor, pois as variáveis z_2 , z_3 e z_4 foram todas excluídas da equação — esse é um outro exemplo de *restrições de exclusão*. Embora haja duas variáveis endógenas nessa equação, temos três VIs em potencial para y_2 e y_3 . Portanto, a equação (16.27) passa na condição de ordem. Para finalizar, declararemos a condição de ordem dos MES gerais.

CONDICÃO DE ORDEM PARA A IDENTIFICAÇÃO

Uma equação em qualquer MES satisfaz a condição de ordem para a identificação se o número de variáveis exógenas *excluídas* da equação for pelo menos tão grande quanto o número de variáveis endógenas existentes no lado direito da equação.

A segunda equação, (16.28), também passa na condição de ordem porque existe uma variável exógena excluída, z_4 , e uma variável endógena, y_1 , no lado direito.

Como discutimos no Capítulo 15 e na seção anterior, a condição de ordem é somente necessária, não suficiente, para a identificação. Por exemplo, se $\beta_{34} = 0$, z_4 não aparecerá em nenhum lugar no sistema, o que significa que ela não é correlacionada com y_1 , y_2 ou y_3 . Se $\beta_{34} = 0$, então, a segunda equação não será identificada, porque z_4 será inútil como uma VI de y_1 . Isso novamente ilustra que a identificação de uma equação depende dos valores dos parâmetros (que nunca conhecemos com certeza) nas outras equações.

Existem muitas maneiras sutis de a identificação falhar em MES complicados. Para obter condições suficientes, necessitaremos estender a condição de classificação para a identificação em sistemas de duas equações. Isso é possível, mas requer álgebra matricial [veja, por exemplo, Wooldridge (2002, Capítulo 9)]. Em muitas aplicações, é possível assumir que, a menos que haja falha óbvia de identificação, uma equação que satisfaça a condição de ordem será identificada.

A nomenclatura sobre equações sobreidentificadas e exatamente identificadas do Capítulo 15 tem origem nos MES. Em termos da condição de ordem, (16.27) é uma **equação sobreidentificada** porque precisamos somente de duas VIs (para y_2 e y_3), mas temos três disponíveis (z_2 , z_3 e z_4); existe uma restrição sobreidentificadora nessa equação. Em geral, o número de restrições sobreidentificadoras iguala o número total de variáveis exógenas no sistema, menos o número total de variáveis explicativas na equação. Isso pode ser verificado usando o teste de sobreidentificação da Seção 15.5. A equação (16.28) é uma **equação exatamente identificada**, e a terceira equação é uma **equação não-identificada**.

Estimação

A despeito do número de equações em um MES, cada equação identificada poderá ser estimada por MQ2E. As variáveis instrumentais de uma equação particular consistirão nas variáveis exógenas que aparecerem em qualquer lugar no sistema. Poderão ser obtidos testes de endogeneidade, heteroscedasticidade, correlação serial e de restrições sobreidentificadoras, da mesma forma como demonstrado no Capítulo 15.

Quando qualquer sistema com duas ou mais equações é corretamente especificado e certas hipóteses adicionais são válidas, os *métodos de estimação de sistemas* são geralmente mais eficientes do que estimar cada equação por MQ2E. O método de estimação de sistemas mais comum no contexto dos MES é o dos *mínimos quadrados de três estágios*. Esses métodos, com ou sem variáveis explicativas endógenas, estão além do escopo deste texto. [Veja, por exemplo, Wooldridge (2002, Capítulos 7 e 8).]

16.5 MODELOS DE EQUAÇÕES SIMULTÂNEAS COM SÉRIES TEMPORAIS

Entre as aplicações mais antigas dos MES estavam as estimativas de grandes sistemas de equações simultâneas, que eram usados para descrever a economia de um país. Um modelo simples keynesiano de demanda agregada (que ignora exportações e importações) é

$$C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2r_t + u_{t1} \quad (16.30)$$

$$I_t = \gamma_0 + \gamma_1r_t + u_{t2} \quad (16.31)$$

$$Y_t = C_t + I_t + G_t \quad (16.32)$$

onde C_t é o consumo, Y_t é a renda, T_t é a receita de impostos, r_t é a taxa de juros, I_t são os investimentos e G_t são os gastos governamentais. [Veja, por exemplo, Mankiw (1994, Capítulo 9).] Especificamente, assuma que t representa ano.

A primeira equação é uma função consumo agregado, na qual o consumo depende da renda disponível, da taxa de juros e do erro estrutural não-observado u_{t1} . A segunda equação é uma função investimento bastante simples. A equação (16.32) é uma *identidade* que é um resultado da contabilidade da renda nacional: ela se mantém, por definição, sem erro. Assim, não faz sentido estimar (16.32), mas precisamos dessa equação para completar o modelo.

Como existem três equações no sistema, também deve haver três variáveis endógenas. Dadas as primeiras duas equações, é claro que pretendemos que C_t e I_t sejam endógenas. Além disso, devido à identidade contábil, Y_t é endógena. Assumiremos, pelo menos nesse modelo, que T_t , r_t e G_t são exógenas, de forma que elas são não-correlacionadas com u_{t1} e u_{t2} . (Discutiremos sobre problemas com esse tipo de hipótese mais tarde.)

Se r_t for exógena, então, a estimação da equação (16.31) por MQO será natural. A função consumo, porém, depende da renda disponível, que é endógena porque Y_t é endógena. Temos duas variáveis instrumentais disponíveis sob as hipóteses de exogeneidade: T_t e G_t . Portanto, se seguirmos nossa receita de estimação de equações de corte transversal, estimaremos (16.30) por MQ2E usando as variáveis instrumentais (T_t, G_t, r_t).

Modelos como (16.30) a (16.32) hoje em dia são raramente estimados, por várias e boas razões. Primeiro, é muito difícil justificar, em um nível agregado, a hipótese de que impostos, taxas de juros e gastos governamentais sejam exógenos. Impostos claramente dependem diretamente da renda; por exemplo, com uma única alíquota de imposto de renda τ , no ano t , $T_t = \tau Y_t$. Podemos com facilidade permitir isso substituindo ($Y_t - T_t$) por $(1 - \tau_t)Y_t$ em (16.30) e ainda poderemos estimar a equação por MQ2E se assumirmos que o gasto governamental é exógeno. Poderemos também adicionar a alíquota do imposto na lista das variáveis instrumentais, se ela for exógena. Entretanto, serão os gastos governamentais e as alíquotas de impostos realmente exógenos? Certamente, em princípio, eles poderiam ser, se o governo definisse os gastos e as alíquotas de impostos independentemente do que estivesse acontecendo com a economia. Mas esse é um caso difícil de acontecer na vida real: gastos governamentais geralmente dependem do nível da renda, e com altos níveis de renda, a mesma receita de impostos é arrecadada com alíquotas de impostos menores. Além disso, assumir que as taxas de juros sejam exógenas é extremamente questionável. Poderíamos especificar um modelo mais realista que incluisse a demanda e a oferta de moeda e, então, as taxas de juros poderiam ser determinadas com C_t , I_t e Y_t . Contudo, nesse caso, encontrar suficientes variáveis exógenas para identificar as equações se torna bastante difícil (e os problemas seguintes com esses modelos ainda são pertinentes).

Algumas pessoas têm argumentado que certos componentes dos gastos governamentais, como os gastos com a defesa — veja, por exemplo, Hall (1988) e Ramey (1991) —, são exógenos em uma variedade de aplicações de equações simultâneas. Porém, não há unanimidade sobre isso e, de qualquer forma, os gastos com a defesa nem sempre são apropriadamente correlacionados com as variáveis explicativas endógenas [veja Shea (1993) para uma discussão sobre o assunto e o Problema 16.14 para um exemplo].

Um segundo problema com um modelo como (16.30) a (16.32) é que ele se apresenta completamente estático. Especialmente com dados mensais ou trimestrais, mas até mesmo com dados anuais, freqüentemente esperamos ajustes de desfasagens. (Um argumento a favor dos modelos estáticos do tipo keynesiano é que eles pretendem descrever a dinâmica de longo prazo sem se preocupar com a dinâmica de curto prazo.) Permitir a dinâmica não é muito difícil. Por exemplo, poderíamos adicionar renda desfasada à equação (16.31):

$$I_t = \gamma_0 + \gamma_1 r_t + \gamma_2 Y_{t-1} + u_{t2}. \quad (16.33)$$

Em outras palavras, adicionamos uma variável **endógena defasada** (mas não I_{t-1}) à equação de investimentos. Podemos tratar Y_{t-1} como exógena nessa equação? Sob certas hipóteses sobre u_{t2} , a resposta é sim. Entretanto, em geral, chamamos uma variável endógena defasada em um MES de **variável predeterminada**. Defasagens de variáveis exógenas também são predeterminadas. Se assumirmos que u_{t2} é não-correlacionado com as variáveis exógenas correntes (o que é padrão) e com todas as variáveis endógenas e exógenas *passadas*, então, Y_{t-1} será não-correlacionada com u_{t2} . Dada a exogeneidade de r_t , podemos estimar (16.33) por MQO.

Se adicionarmos consumo defasado à equação (16.30), poderemos tratar C_{t-1} como exógeno nessa equação sob as mesmas hipóteses sobre u_{t1} que fizemos para u_{t2} no parágrafo anterior. A renda disponível corrente ainda será endógena em

$$C_t = \beta_0 + \beta_1(Y_t - T_t) + \beta_2 r_t + \beta_3 C_{t-1} + u_{t1}, \quad (16.34)$$

de modo que poderemos estimar essa equação por MQ2E usando as variáveis instrumentais (T_t, G_t, r_t, C_{t-1}) ; se o investimento for determinado por (16.33), Y_{t-1} deve ser incluído na lista de instrumentais. [Para verificar o motivo, use (16.32), (16.33) e (16.34) para encontrar a forma reduzida de Y_t em termos das variáveis exógenas e predeterminadas T_t, r_t, G_t, C_{t-1} e Y_{t-1} . Como Y_{t-1} aparece nessa forma reduzida, ela deve ser usada como uma VI.]

A presença de dinâmica em MES agregados é, pelo menos para o propósito de previsões, uma clara melhoria sobre os MES estáticos. Contudo, ainda existem alguns problemas importantes na estimação de MES usando dados agregados de séries temporais, alguns dos quais discutimos nos Capítulos 11 e 15. Lembre-se de que a validade dos procedimentos de inferência dos habituais MQO ou MQ2E em aplicações de séries temporais depende da noção de *dependência fraca*. Infelizmente, séries como as de consumo agregado, renda, investimentos e até mesmo de taxas de juros parecem violar os requisitos de dependência fraca. (Na terminologia do Capítulo 11, elas têm *raízes unitárias*.) Essas séries também estão propensas a ter tendências exponenciais, embora isso possa ser parcialmente compensado pelo uso da transformação logarítmica quando assumiu diferentes formas funcionais. Geralmente, mesmo as propriedades de amostras grandes, sem falar das pequenas, do MQO e do MQ2E são complicadas e dependentes de várias hipóteses quando aplicadas a equações com variáveis I(1). Abordaremos levemente esses problemas no Capítulo 18. Uma abordagem geral avançada é dada por Hamilton (1994).

A discussão anterior significa que os MES não são úteis quando aplicados a dados de séries temporais? Muito ao contrário. Os problemas com tendências e alta persistência podem ser evitados especificando sistemas em primeiras diferenças ou em taxas de crescimento. Contudo, devemos reconhecer que esse é um MES diferente de outro especificado em nível. [Por exemplo, se especificarmos o crescimento do consumo como uma função do crescimento da renda disponível e das alterações das taxas de juros, isso será diferente de (16.30).] Além disso, como discutimos anteriormente, a incorporação de dinâmica não é especialmente difícil. Finalmente, o problema de encontrar variáveis verdadeiramente exógenas para serem incluídas nos MES geralmente é mais fácil com dados desagregados. Por exemplo, para indústrias manufatureiras, Shea (1993) descreve como a produção (ou, mais precisamente, o crescimento da produção) em outras indústrias pode ser usada como uma variável instrumental na estimação de funções de oferta. Ramey (1991) também contém uma análise convincente da estimação de funções de custos industriais por variáveis instrumentais utilizando dados de séries temporais.

O próximo exemplo mostra como dados agregados podem ser usados para testar uma importante teoria econômica, a teoria do consumo da renda permanente, habitualmente chamada de *hipótese da renda permanente* (HRP). O método usado nesse exemplo não é, a rigor, baseado em um modelo de equações simultâneas, mas podemos pensar no consumo e no crescimento da renda (como também nas taxas de juros) como sendo determinados conjuntamente.

EXEMPLO 16.7

(O Teste da Hipótese da Renda Permanente)

Campbell e Mankiw (1990) usaram métodos de variáveis instrumentais para testar várias versões da hipótese da renda permanente. Usaremos os dados anuais de 1959 a 1995 do arquivo CONSUMP.RAW para reproduzir uma de suas análises. Campbell e Mankiw usaram dados trimestrais até 1985.

Uma equação estimada por Campbell e Mankiw (usando nossa notação) foi

$$gc_t = \beta_0 + \beta_1 gy_t + \beta_2 r3_t + u_t, \quad (16.35)$$

onde $gc_t = \Delta \log(c_t)$ é o crescimento anual do consumo *per capita* real (excluindo bens duráveis), gy_t é o crescimento da renda disponível real, e $r3_t$ é a (*ex post*) taxa de juros real, medida pelo rendimento da taxa das letras do Tesouro norte-americano de três meses: $r3_t = i3_t - inf_t$, onde a taxa de inflação é baseada no índice de preços ao consumidor. As taxas de crescimento do consumo e da renda disponível não apresentam tendência e são fracamente dependentes; assumiremos também ser esse o caso da $r3_t$, de forma a podermos aplicar a teoria assintótica padrão.

A principal característica da equação (16.35) é que a HRP indica que o termo erro u_t tem uma média zero condicional em todas as informações observadas no momento $t - 1$ ou anterior: $E(u_t | I_{t-1}) = 0$. Porém, u_t não é necessariamente não-correlacionado com gy_t ou com $r3_t$; uma maneira tradicional de se pensar sobre isso é que essas variáveis são conjuntamente determinadas, mas não estamos escrevendo um sistema completo de três equações.

Como u_t é não-correlacionado com todas as variáveis datadas em $t - 1$ ou antes, as variáveis instrumentais válidas para estimar (16.35) são valores defasados de gc , gy e $r3$ (e defasagens de outras variáveis observáveis, mas não as usaremos aqui). Quais são as hipóteses de interesse? A forma pura da HRP tem $\beta_1 = \beta_2 = 0$. Campbell e Mankiw argumentam que β_1 será positivo se alguma fração da população consome renda corrente, em vez de renda permanente. A HRP com uma taxa de juros real não constante implica que $\beta_2 > 0$.

Quando estimamos (16.35) por MQ2E, usando as instrumentais gc_{-1} , gy_{-1} e $r3_{-1}$, obtemos

$$\begin{aligned} \hat{gc}_t &= 0,0081 + 0,586 gy_t - 0,00027 r3_t, \\ &\quad (0,0032) \quad (0,135) \quad (0,00076) \end{aligned} \quad (16.36)$$

$$n = 35, R^2 = 0,678.$$

Portanto, a forma pura da HRP é fortemente rejeitada porque o coeficiente de gy é economicamente grande (um aumento de 1% na renda disponível aumenta em mais de 5% o consumo) e estatisticamente significante ($t = 4,34$). De outro lado, o coeficiente da taxa de juros real é bastante pequeno e estatisticamente não significante. Essas constatações são qualitativamente as mesmas de Campbell e Mankiw.

EXEMPLO 16.7 (continuação)

A HRP também implica que os erros (u_t) são serialmente não-correlacionados. Após a estimação por MQ2E, obtemos os resíduos \hat{u}_t e incluímos \hat{u}_{t-1} como uma variável explicativa adicional em (16.36); ela age como sua própria variável instrumental (veja a Seção 15.7). O coeficiente de \hat{u}_{t-1} é $\hat{\rho} = 0,187$ ($p = 0,133$), de modo que existe alguma evidência de correlação serial positiva, embora não no nível de significância de 5%. Campbell e Mankiw (1990) discutem porque, com os dados trimestrais disponíveis, pode ser encontrada correlação serial nos erros mesmo se a HRP se sustentar; algumas dessas preocupações se estendem aos dados anuais.

Suponha que para uma cidade específica você tenha dados mensais sobre o consumo *per capita* de peixe, renda *per capita*, preços de peixe e preços de frango e carne bovina; a renda e os preços de frango e carne são exógenos. Assuma que não há sazonalidade na função de demanda de peixe, mas que ela existe na função de oferta de peixe. Como você pode usar essa informação para estimar uma equação de demanda de peixe com elasticidade constante? Especifique uma equação e detalhe a identificação. (Sugestão: Você deve ter onze variáveis instrumentais do preço de peixe.)

O uso de taxas de variáveis de crescimento de tendências ou variáveis I(1) em MES é bastante comum em aplicações de séries temporais. Por exemplo, Shea (1993) estima as curvas de ofertas industriais em termos de taxas de crescimento.

Se um modelo estrutural contiver uma tendência temporal — que pode capturar fatores exógenos de tendência que não sejam diretamente modelados —, então, a tendência agirá como sua própria VI.

16.6 MODELOS DE EQUAÇÕES SIMULTÂNEAS COM DADOS DE PAINEL

Modelos de equações simultâneas também surgem no contexto de dados de painel. Por exemplo, podemos imaginar a estimação de equações de oferta de mão-de-obra e de oferta de salários, como no Exemplo 16.3, de um grupo de pessoas que tenha trabalhado por certo período de tempo. Além de permitirnos a determinação simultânea das variáveis dentro de cada período de tempo, podemos admitir efeitos não observados em cada equação. Em uma função de oferta de mão-de-obra, seria útil possibilitar a preferência por lazer não-observado que não se altere ao longo do tempo.

A abordagem básica para estimar MES com dados de painel compreende duas etapas: (1) eliminar os efeitos não-observados das equações de interesse usando a transformação de efeitos fixos ou a primeira diferenciação e (2) encontrar variáveis instrumentais das variáveis endógenas na equação transformada. Isso pode ser bastante desafiador, pois para uma análise convincente, precisaremos encontrar variáveis instrumentais que mudem ao longo do tempo. Para ver o motivo disso, escreva um MES de dados de painel como

$$y_{it1} = \alpha_1 y_{it2} + z_{it1} \beta_1 + \alpha_{i1} + u_{it1} \quad (16.37)$$

$$y_{it2} = \alpha_2 y_{it1} + z_{it2} \beta_2 + \alpha_{i2} + u_{it2}, \quad (16.38)$$

onde i representa o corte transversal, t é o período de tempo, e $z_{it1} \beta_1$ ou $z_{it2} \beta_2$ são funções lineares de um conjunto de variáveis explicativas exógenas em cada equação. A análise mais geral permite que os efeitos não-observados, α_{i1} e α_{i2} , sejam correlacionados com *todas* as variáveis explicativas, mesmo os elementos em z . Porém, assumimos que os erros estruturais idiosincráticos, u_{it1} e u_{it2} , são não-correlacionados com z em ambas as equações e ao longo de todos os períodos de tempo; é nesse sentido que z é exógeno. Exceto sob circunstâncias especiais, y_{it2} será correlacionado com u_{it1} , e y_{it1} será correlacionado com u_{it2} .

Suponha que estamos interessados na equação (16.37). Não podemos estimá-la por MQO, já que o erro de composição $\alpha_{i1} + u_{it1}$ é potencialmente correlacionado com todas as variáveis explicativas. Suponha que façamos a diferenciação ao longo do tempo para remover o efeito não-observado, α_{i1} :

$$\Delta y_{it1} = \alpha_1 \Delta y_{it2} + \Delta z_{it1} \beta_1 + \Delta u_{it1}. \quad (16.39)$$

(Como é usual na diferenciação ou na redução temporal, somente podemos estimar os efeitos de variáveis que se alteram ao longo do tempo de pelo menos algumas unidades de corte transversal.) Agora, o termo erro nessa equação é não-correlacionado com Δz_{it1} por hipótese. Mas Δy_{it2} e Δu_{it1} possivelmente serão correlacionados. Portanto, precisamos de uma VI para Δy_{it2} .

Como no caso de dados de corte transversal puro ou de séries temporais puras, as VIs possíveis virão da *outra* equação: elementos de z_{it2} que não estejam também em z_{it1} . Na prática, precisamos de elementos com *variação temporal* em z_{it2} que não estejam também em z_{it1} . Isso porque necessitamos de uma variável instrumental de Δy_{it2} , e é pouco provável que uma mudança em uma variável de um período para o próximo seja altamente correlacionada com o *nível* das variáveis exógenas. Na verdade, se diferenciarmos (16.38), veremos que as VIs naturais de Δy_{it2} são os elementos Δz_{it2} que não estão também em z_{it1} .

Como exemplo dos problemas que podem surgir, considere uma versão de dados de painel da função de oferta de mão-de-obra no Exemplo 16.3. Após fazer a diferenciação, suponha que temos a equação

$$\Delta horas_{it} = \beta_0 + \alpha_1 \Delta \log(salário_{it}) + \Delta(outros\ fatores_{it}),$$

e queremos usar $\Delta exper_{it}$ como uma variável instrumental de $\Delta \log(salário_{it})$. O problema é que, como estamos examinando pessoas que trabalham em todos os períodos de tempo, $\Delta exper_{it} = 1$ para todos os i e t . (Todas as pessoas adquirem mais um ano de experiência após a passagem de um ano.) Não podemos usar uma VI que tenha o mesmo valor para todos os i e t e, portanto, temos de continuar a procura.

Freqüentemente, a participação em um programa experimental pode ser usada para obter VIs em contextos de dados de painel. No Exemplo 15.10, usamos o recebimento de subsídios de treinamento de pessoal como uma VI da mudança nas horas de treinamento na determinação dos efeitos do treinamento de pessoal sobre a produtividade do trabalhador. De fato, poderíamos ver que, em um contexto de MES, o treinamento de pessoal e a produtividade do trabalhador seriam determinados conjuntamente, mas o recebimento de um subsídio de treinamento de pessoal seria exógeno na equação (15.57).

Pode haver algumas vezes propor variáveis instrumentais engenhosas e convincentes em aplicações de dados de painel, como ilustra o exemplo seguinte.

EXEMPLO 16.8**(Efeito da População Prisional sobre as Taxas de Crimes Violentos)**

Para estimar o efeito causal do aumento da população prisional sobre as taxas de criminalidade em nível estadual, Levitt (1996) usou exemplos de processos judiciais sobre superlotação prisional como variáveis instrumentais do crescimento da população prisional. A equação que Levitt estimou estava em primeiras diferenças; podemos escrever um modelo de efeitos fixos subjacentes como

$$\log(crime_{it}) = \theta_t + \alpha_1 \log(prisão_{it}) + z_{it1}\beta_1 + \alpha_{it1} + u_{it1}, \quad (16.40)$$

onde θ_t representa diferentes interceptos temporais, e *crime* e *prisão* são medidos por 100.000 habitantes. (A variável da população prisional é avaliada no último dia do ano anterior.) O vetor z_{it1} contém outros controles listados no trabalho de Levitt, inclusive medidas de policiais *per capita*, renda *per capita*, taxa de desemprego, raça e as proporções de distribuições metropolitanas e por idade.

A diferenciação de (16.40) produz a equação estimada por Levitt:

$$\Delta \log(crime_{it}) = \xi_t + \alpha_1 \Delta \log(prisão_{it}) + z_{it1}\beta_1 + \Delta u_{it1}. \quad (16.41)$$

A simultaneidade entre taxas criminais e população prisional, ou mais precisamente nas taxas de crescimento, faz com que a estimativa por MQO de (16.41) seja geralmente inconsistente. Usando a taxa de crimes violentos e um subconjunto dos dados de Levitt (no arquivo PRISON.RAW, dos anos de 1980 a 1993, com um total de $51 \cdot 14 = 714$ observações), obtemos a estimativa por MQO agrupado de α_1 , que é $-0,181$ ($ep = 0,048$). Também estimamos (16.41) por MQ2E agrupado, onde as variáveis instrumentais de $\Delta \log(prisão)$ são duas variáveis binárias, uma registrando se uma decisão final sobre o processo judicial da superlotação foi tomada no ano corrente e outra se nos dois anos anteriores. A estimativa por MQ2E agrupado de α_1 é $-1,032$ ($ep = 0,370$). Portanto, o efeito estimado por MQ2E é muito maior; não surpreende que ele seja também muito menos preciso. Levitt (1996) encontrou resultados semelhantes quando usou períodos de tempo mais longos (mas sem observações de períodos mais anteriores para alguns estados) e mais variáveis instrumentais.

O teste da existência de correlação serial AR(1) em $r_{it1} = \Delta u_{it1}$ é fácil. Após a estimativa por MQ2E agrupado, obtenha os resíduos, \hat{r}_{it1} . Depois, inclua uma defasagem desses resíduos na equação original, e estime-a por MQ2E, onde \hat{r}_{it1} age como sua própria variável instrumental. O primeiro ano será perdido devido à defasagem. Então, a habitual estatística t do MQ2E do resíduo defasado será um teste válido para verificar a existência de correlação serial. No Exemplo 16.8, o coeficiente de \hat{r}_{it1} está em torno de somente 0,076 com $t = 1,67$. Com um coeficiente tão pequeno e tão modesta estatística t , podemos, com segurança, assumir independência serial.

Uma abordagem alternativa de estimar MES com dados de painel é usar a transformação de efeitos fixos e depois aplicar uma técnica de VI como o MQ2E agrupado. Um procedimento simples é estimar a equação com redução temporal por MQ2E, que se pareceria com

$$\ddot{y}_{it1} = \alpha_1 \ddot{y}_{it2} = \ddot{z}_{it1}\beta_1 = \ddot{u}_{it1}, \quad t = 1, 2, \dots, T, \quad (16.42)$$

onde \tilde{z}_{it1} e \tilde{z}_{it2} são VIs. Isso é equivalente a usar o MQ2E na formulação de variáveis *dummy*, onde as variáveis *dummy* específicas-unitárias agem como suas próprias variáveis instrumentais. Ayres e Levitt (1998) aplicaram MQ2E em uma equação com redução temporal para estimar o efeito dos dispositivos eletrônicos Lojack contra roubos de automóveis nas cidades. Se (16.42) for estimada diretamente, então, os gl precisarão ser corrigidos para $N(T - 1) - k_1$, em que k_1 será o número total de elementos em α_1 e β_1 . A inclusão de variáveis *dummy* específicas-unitárias e a aplicação do MQ2E agrupado aos dados originais produzirão os gl corretos.

Modelos de equações simultâneas são apropriados quando cada equação no sistema tem uma interpretação *ceteris paribus*. Bons exemplos são quando equações separadas descrevem diferentes ângulos de um mercado ou as relações comportamentais de diferentes agentes econômicos. Exemplos de oferta e demanda são os principais casos, mas existem muitas outras aplicações dos MES em economia e nas ciências sociais.

Uma característica importante dos MES é que, pela completa especificação do sistema, fica claro quais variáveis são assumidas como exógenas e quais delas aparecem em cada equação. Dado um sistema completo, temos condições de determinar quais equações podem ser identificadas (isto é, podem ser estimadas). No importante caso de um sistema com duas equações, é fácil estabelecer a identificação da (digamos) primeira equação: deve haver pelo menos uma variável exógena excluída da primeira equação que apareça com um coeficiente diferente de zero na segunda equação.

Como vimos dos capítulos anteriores, a estimativa por MQO de uma equação que contém uma variável explicativa endógena geralmente produz estimadores viesados e inconsistentes. Diferentemente, o MQ2E pode ser usado para estimar qualquer equação identificada em um sistema. Métodos de sistemas mais avançados estão disponíveis, mas estão além do escopo deste livro.

A distinção entre variáveis omitidas e simultaneidade nas aplicações nem sempre é nítida. Ambos os problemas, sem falar no erro de medida, podem aparecer na mesma equação. Um bom exemplo é a oferta de mão-de-obra de mulheres casadas. Anos de escolaridade (*educ*) aparece tanto na função de oferta de mão-de-obra como na de oferta de salário [veja as equações (16.19) e (16.20)]. Se a aptidão omitida estiver no termo erro da função de oferta de mão-de-obra, então, salário e educação serão ambos endógenos. O fator importante é que uma equação estimada por MQ2E pode basear-se em si própria.

Os MES também podem ser aplicados a dados de séries temporais. Assim como na estimativa por MQO, devemos estar atentos aos processos com tendências e integrados na aplicação do MQ2E. Problemas tais como a correlação serial podem ser tratados como na Seção 15.7. Também apresentamos um exemplo de como estimar um MES usando dados de painel, onde fazemos a primeira diferenciação da equação para remover o efeito não-observado. Depois, podemos estimar a equação diferenciada por MQ2E agrupado, como no Capítulo 15. Alternativamente, em alguns casos, podemos usar a redução temporal de todas as variáveis, inclusive as VIs e, então, aplicar o MQ2E agrupado; isso é o mesmo que incluir *dummies* de cada observação de corte transversal e usar o MQ2E, onde as *dummies* agem com suas próprias variáveis instrumentais. Aplicações de MES com dados de painel são muito poderosas, já que nos possibilitam controlar a heterogeneidade não-observada ao mesmo tempo em que estamos lidando com a simultaneidade. Elas estão se tornando cada vez mais comuns e não são especialmente difíceis de serem estimadas.

16.1 Escreva um sistema de duas equações na forma de “oferta e demanda”, isto é, com a mesma variável y_1 (em geral, “quantidade”) aparecendo do lado esquerdo:

$$\begin{aligned}y_1 &= \alpha_1 y_2 + \beta_1 z_1 + u_1 \\y_1 &= \alpha_2 y_2 + \beta_2 z_2 + u_2.\end{aligned}$$

- (i) Se $\alpha_1 = 0$ ou $\alpha_2 = 0$, explique por que existe uma forma reduzida de y_1 . (Lembre-se, uma forma reduzida expressa y_1 como uma função linear das variáveis exógenas e dos erros estruturais.) Se $\alpha_1 \neq 0$ e $\alpha_2 = 0$, encontre a forma reduzida de y_2 .
- (ii) Se $\alpha_1 \neq 0$, $\alpha_2 \neq 0$ e $\alpha_1 \neq \alpha_2$, encontre a forma reduzida de y_1 . A variável y_2 tem uma forma reduzida nesse caso?
- (iii) A condição $\alpha_1 \neq \alpha_2$ é possível de ser encontrada em exemplos de demanda? Explique.

16.2 Defina *milho* como o consumo *per capita* de milho em toneladas de grãos, no nível de município, *preço* como o preço por tonelada do milho, *renda* como a renda *per capita* do município, e defina *precpluv* como a precipitação pluviométrica em polegadas durante a última safra de plantio de milho. O seguinte modelo de equações simultâneas impõe a condição de equilíbrio em que a oferta se iguala à demanda:

$$\begin{aligned}\text{milho} &= \alpha_1 \text{preço} + \beta_1 \text{renda} + u_1 \\\text{milho} &= \alpha_2 \text{preço} + \beta_2 \text{precpluv} + u_2.\end{aligned}$$

Qual é a equação de oferta e qual é a de demanda? Explique.

16.3 No Problema 3.3 do Capítulo 3, estimamos uma equação para testar uma relação de substituição entre minutos por semana gastos dormindo (*dormir*) e minutos por semana gastos trabalhando (*trabtot*) de uma amostra aleatória de indivíduos. Também incluímos educação e idade na equação. Como *dormir* e *trabtot* são escolhidos conjuntamente por indivíduo, a relação de substituição entre dormir e trabalhar estimada está sujeita a uma crítica de “viés de simultaneidade”? Explique.

16.4 Suponha que os ganhos e o consumo de álcool anuais sejam determinados pelo MÉS

$$\begin{aligned}\log(\text{ganhos}) &= \beta_0 + \beta_1 \text{álcool} + \beta_2 \text{educ} + u_1 \\\text{álcool} &= \gamma_0 + \gamma_1 \log(\text{ganhos}) + \gamma_2 \text{educ} + \gamma_3 \log(\text{preço}) + u_2,\end{aligned}$$

onde *preço* é o índice local de preços do álcool, que inclui impostos locais e estaduais. Assuma que *educ* e *preço* sejam exógenos. Se β_1 , β_2 , γ_1 , γ_2 e γ_3 forem todas diferentes de zero, qual equação será identificada? Como você estimaria essa equação?

16.5 Um modelo simples para determinar a eficácia do uso da camisinha na redução das doenças sexualmente transmissíveis entre alunos do ensino médio sexualmente ativos é

$$\text{taxainf} = \beta_0 + \beta_1 \text{usacamis} + \beta_2 \text{percmasc} + \beta_3 \text{rendfam} + \beta_4 \text{cidade} + u_1,$$

onde *taxainf* é a porcentagem de alunos sexualmente ativos que tenham contraído doença venérea, *usacamis* é a porcentagem de rapazes que afirmam usar camisinha regularmente, *rendfam* é a renda familiar média e *cidade* é uma variável dummy indicando se a escola está em uma cidade; o modelo é construído no nível de escolas.

- (i) Interpretando a equação precedente de uma maneira causal, *ceteris paribus*, qual deverá ser o sinal de β_1 ?
- (ii) Por que *taxainf* e *usacamis* podem ser conjuntamente determinadas?
- (iii) Se o uso de camisinha aumentar com a taxa de doenças venéreas, de forma que $\gamma_1 > 0$ na equação

$$usacamis = \gamma_0 + \gamma_1 taxainf + \text{outros fatores},$$

- qual será o provável viés na estimativa de β_1 por MQO?
- (iv) Defina *disticamis* como uma variável binária igual a um se uma escola tiver um programa de distribuição de camisetas. Explique como isso pode ser usado para estimar β_1 (e os outros betas) por VIs. O que teremos de assumir sobre *disticamis* em cada equação?

16.6 Considere um modelo de probabilidade linear explicando se os empregadores oferecem um plano de pensão com base na porcentagem de trabalhadores que pertençam a um sindicato, bem como outros fatores:

$$\text{pensão} = \beta_0 + \beta_1 pertsind + \beta_2 idademed + \beta_3 educmed + \beta_4 percmasc + \beta_5 perccasad + u_1.$$

- (i) Por que *pertsind* pode ser determinado conjuntamente com *pensão*?
- (ii) Suponha que você possa pesquisar os trabalhadores nas firmas e colher informações sobre as famílias dos trabalhadores. Você consegue pensar em uma informação que poderia ser usada como uma VI de *pertsind*?
- (iii) Como você verificaria se sua variável é pelo menos uma candidata razoável a VI de *pertsind*?

16.7 Suponha que você seja solicitado a estimar a demanda por ingressos de jogos de basquete feminino de uma grande universidade. Você pode coletar dados de séries temporais de mais de dez temporadas, com um total de cerca de 150 observações. Um modelo possível seria

$$lPÚBLICO_t = \beta_0 + \beta_1 lPREÇO_t + \beta_2 PERCVIT_t + \beta_3 RIVAL_t + \beta_4 FIMSEMANA_t + \beta_5 t + u_1,$$

onde *PREÇO*_{*t*} é o preço do ingresso, provavelmente indicado em termos reais — digamos, deflacionado por um índice local de preços ao consumidor ---, *PERCVIT*_{*t*} é a porcentagem de vitórias atuais da equipe, *RIVAL*_{*t*} é uma variável *dummy* indicando um jogo contra um rival e *FIMSEMANA*_{*t*} é uma variável *dummy* indicando se o jogo é realizado durante o fim de semana. O *l* representa o logaritmo natural, de forma que a função de demanda tem uma elasticidade-preço constante.

- (i) Por que é uma boa idéia ter uma tendência temporal na equação?
- (ii) A oferta de ingressos é fixada pela capacidade do estádio; assuma que ela não tenha mudado nos últimos dez anos. Isso significa que a quantidade oferecida não varia com o preço. Significa que o preço será necessariamente exógeno na equação de demanda? (*Sugestão*: A resposta é não.)
- (iii) Suponha que o preço nominal do ingresso se altere lentamente — digamos, no início de cada temporada. O departamento esportivo determina os preços baseando-se parcialmente no público da temporada anterior, como também no sucesso obtido pela equipe na temporada anterior. Sob que hipóteses a porcentagem de vitórias na temporada anterior (*PERVITTEMP*_{*t*-1}) será uma variável instrumental válida de *lPREÇO*_{*t*}?
- (iv) Parece razoável incluir o (log do) preço real dos ingressos dos jogos de basquete masculino na equação? Explique. Que sinal a teoria econômica prevê para seu coeficiente?

Você consegue pensar em outra variável relacionada ao basquetebol masculino que possa pertencer à equação do público nos jogos femininos?

- (v) Se você está preocupado com que algumas das séries, particularmente $IPÚBLICO$ e $IPREÇO$, tenham raízes unitárias, como você poderia alterar a equação estimada?
- (vi) Se alguns jogos tiverem suas lotações esgotadas, que problemas isso causará para a estimativa da função de demanda? (Sugestão: Se um jogo tiver sua lotação esgotada, você necessariamente observará a demanda real?)

16.8 O quanto é grande o efeito dos gastos escolares por aluno sobre os preços de habitação local? Defina $PREÇOC$ como sendo a mediana dos preços de habitação em um distrito escolar e defina $GASTO$ como os gastos por aluno. Usando dados de painel dos anos de 1992, 1994 e 1996, postulamos o modelo

$$IPREÇOC_{it} = \theta_t + \beta_1 lGASTO_{it} + \beta_2 lPOLÍCIA_{it} + \beta_3 lMEDREND_{it} + \beta_4 lIMPPRO_{it} + a_{it} + u_{it},$$

onde $POLÍCIA_{it}$ são os gastos policiais *per capita*, $MEDREND_{it}$ é a mediana da renda e $IMPPRO_{it}$ é a alíquota do imposto sobre a propriedade; l denota logaritmo natural. Gastos e preços de habitação são determinados simultaneamente porque o valor dos imóveis afeta diretamente as receitas disponíveis para financiar as escolas.

Suponha que, em 1994, a maneira pela qual as escolas eram financiadas tenha mudado drasticamente: em vez de serem financiadas pelos impostos locais sobre a propriedade, os financiamentos das escolas tenham sido determinados basicamente em nível estadual. Defina $lALEST_{it}$ como o log da alocação da verba estadual ao distrito i no ano t , que é exógeno na equação precedente, uma vez que controlemos gastos e o efeito fixo de um distrito. Como você estimaria os β_j ?

Modelos com Variáveis Dependentes Limitadas e Correções da Seleção Amostral

No Capítulo 7, estudamos o modelo de probabilidade linear, que simplesmente é uma aplicação do modelo de regressão múltipla a uma variável dependente binária. Uma variável dependente binária é um exemplo de uma **variável dependente limitada (VDL)**. Uma VDL é definida, de modo geral, como uma variável dependente cujo intervalo de valores é substancialmente restrito. Uma variável binária assume somente dois valores, zero e um. Já vimos vários outros exemplos de variáveis dependentes limitadas: a porcentagem de participação em um plano de pensão deve estar entre zero e 100, o número de vezes em que uma pessoa é presa em determinado ano é um inteiro não-negativo, e a nota média da graduação está entre zero e 4,0 na maioria das faculdades dos Estados Unidos.

A maioria das variáveis econômicas que gostaríamos de explicar são de alguma forma limitadas, muitas vezes porque elas devem ser positivas. Por exemplo, o salário por hora, os preços de imóveis e as taxas nominais de juros devem ser maiores que zero. Mas nem todas as variáveis desse tipo precisam de tratamento especial. Se uma variável estritamente positiva assumir vários valores diferentes, raramente será necessário um modelo econométrico especial.

Quando y for discrcta e assumir um pequeno número de valores, não fará sentido tratá-la como uma variável aproximadamente contínua. A descontinuidade de y não significa, em si mesma, que os modelos lineares sejam inadequados. Porém, como vimos no Capítulo 7 sobre a resposta binária, o modelo de probabilidade linear tem certas desvantagens. Na Seção 17.1, discutiremos os modelos logit e probit, que compensam as desvantagens do MPL; a desvantagem é que eles são mais difíceis de ser interpretados.

Outros tipos de variáveis dependentes limitadas surgem na análise econométrica, especialmente quando estamos modelando o comportamento de indivíduos, famílias ou firmas. A otimização de comportamentos freqüentemente leva a uma **resposta de solução de canto** para alguma fração relevante da população. Ou seja, uma quantidade ou valor em dólar zero, por exemplo, é uma escolha ótima. Durante qualquer determinado ano, um número significativo de famílias fará zero contribuições de caridade. Portanto, as contribuições de caridade familiares anuais têm uma distribuição populacional espalhada em uma ampla gama de valores positivos, mas com um acúmulo no valor zero. Embora um modelo linear possa ser apropriado para capturar os valores esperados de contribuições de caridade, muito provavelmente levará a previsões negativas para algumas famílias. Não será possível usar o log natural, pois muitas observações serão zero. O modelo Tobit, que abordaremos na Seção 17.2, é especificamente projetado para modelar variáveis dependentes que tenham soluções de canto.

Outro importante tipo de VDL é uma variável de contagem, que assume valores inteiros não-negativos. A Seção 17.3 ilustra como os modelos de regressão de Poisson são bem apropriados para modelar variáveis de contagem.

Em alguns casos, observamos variáveis dependentes limitadas devido à censura dos dados, um tópico que trataremos na Seção 17.4. O problema geral da seleção amostral, no qual observamos uma amostra não-aleatória da população subjacente, é tratado na Seção 17.5.

Modelos de variáveis dependentes limitadas podem ser usados para séries temporais e dados de painel, mas são aplicados com mais freqüência a dados de corte transversal. Problemas de seleção amostral geralmente estão restritos a dados de corte transversal ou de painel. Neste capítulo, concentraremos-nos em aplicações de corte transversal. Wooldridge (2002) apresenta esses problemas no contexto de modelos de dados de painel e fornece muitos outros detalhes sobre aplicações de corte transversal e dados de painel.

17.1 MODELOS LOGIT E PROBIT DE RESPOSTA BINÁRIA

O modelo de probabilidade linear é fácil de ser estimado e usado, mas tem algumas desvantagens que discutimos na Seção 7.5. As duas desvantagens mais relevantes são que as probabilidades ajustadas podem ser menores que zero ou maiores que um e o efeito parcial de qualquer variável explicativa (aparecendo na forma em nível) é constante. Essas limitações do MPL podem ser compensadas pelo uso de **modelos de resposta binária**.

Em um modelo de resposta binária, o interesse reside, principalmente, na probabilidade de resposta

$$P(y = 1|x) = P(y = 1|x_1, x_2, \dots, x_k), \quad (17.1)$$

onde usamos x para representar o conjunto completo de variáveis explicativas. Por exemplo, quando y for um indicador de emprego, x poderá conter várias características individuais, como educação, idade, estado civil e outros fatores que afetem a situação de emprego, inclusive uma variável binária indicadora da participação em um recente programa de treinamento de pessoal.

A Especificação de Modelos Logit e Probit

No MPL, assumimos que a probabilidade de resposta é linear em um conjunto de parâmetros, β_j ; [veja a equação (7.27)]. Para evitar as limitações do MPL, considere uma classe de modelos de resposta binária da forma

$$P(y = 1|x) = G(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) = G(\beta_0 + x\beta), \quad (17.2)$$

onde G é uma função assumindo valores estritamente entre zero e um: $0 < G(z) < 1$, para todos os números z reais. Isso garante que as probabilidades estimadas de resposta estejam estritamente entre zero e um. Como nos capítulos anteriores, escrevemos $x\beta = \beta_1 x_1 + \dots + \beta_k x_k$.

Várias funções não-lineares têm sido sugeridas para a função G para garantir que as probabilidades estejam entre zero e um. As duas que trataremos aqui são usadas na grande maioria das aplicações (juntamente com o MPL). No **modelo logit**, G é a função logística:

$$G(z) = \exp(z)/[1 + \exp(z)] = \Lambda(z), \quad (17.3)$$

que está entre zero e um para todos os números z reais. Essa é a função de distribuição cumulativa de uma variável aleatória logística padrão. No **modelo probit**, G é a função de distribuição cumulativa (fdc) normal padrão, que é expressa como uma integral:

$$G(z) = \Phi(z) = \int_{-\infty}^z \phi(v)dv, \quad (17.4)$$

onde $\phi(z)$ é a densidade normal padrão

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2/2). \quad (17.5)$$

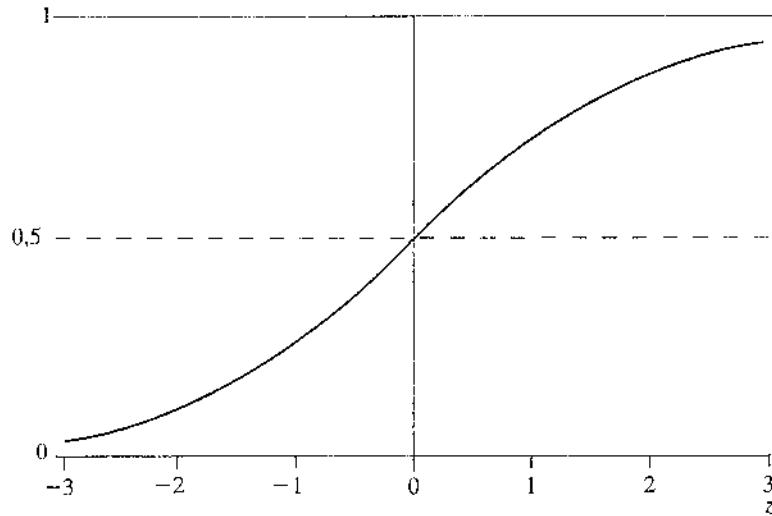
A escolha de G mais uma vez assegura que (17.2) esteja estritamente entre zero e um para todos os valores dos parâmetros e para x_j .

As funções G em (17.3) e (17.4) são ambas funções crescentes. Cada uma delas cresce mais rapidamente com $z = 0$, $G(z) \rightarrow 0$ quando $z \rightarrow -\infty$ e $G(z) \rightarrow 1$ quando $z \rightarrow \infty$. A função logística está representada na Figura 17.1. A fdc normal padrão tem uma forma muito semelhante à da fdc logística.

Figura 17.1

Gráfico da função logística $G(z) = \exp(z)/[1 + \exp(z)]$.

$$G(z) = \exp(z)/[1 + \exp(z)]$$



Os modelos logit e probit podem ser derivados de um **modelo de variável latente** subjacente. Seja y^* uma variável não-observada, ou *latente*, determinada por

$$y^* = \beta_0 + x\beta + e, \quad y = 1[y^* > 0], \quad (17.6)$$

onde introduzimos a notação $1[\cdot]$ para definir um resultado binário. A função $1[\cdot]$ é chamada de *função indicadora*, que assume o valor um se o evento entre colchetes for verdadeiro e zero, caso contrário. Portanto, y será um se $y^* > 0$, e zero se $y^* \leq 0$. Assumimos que e é independente de x e que e tem a distribuição logística padrão ou a distribuição normal padrão. Em qualquer caso, e será simetricamente distribuída ao redor de zero, o que significa que $1 - G(-z) = G(z)$ para todos os números z reais. Os economistas tendem a preferir a hipótese de normalidade de e , razão pela qual o modelo probit é mais popular que o logit em econometria. Além disso, vários problemas de especificação, sobre os quais comentaremos mais tarde, são muito mais facilmente analisados com o uso do probit devido às propriedades da distribuição normal.

Com base em (17.6) e nas hipóteses dadas, podemos derivar a probabilidade de resposta de y :

$$\begin{aligned} P(y = 1|x) &= P(y^* > 0|x) = P[e > -(\beta_0 + x\beta)|x] \\ &= 1 - G[-(\beta_0 + x\beta)] = G(\beta_0 + x\beta), \end{aligned}$$

que é exatamente igual a (17.2).

Na maioria das aplicações de modelos de resposta binária, a meta principal é explicar os efeitos de x_j sobre a probabilidade de resposta $P(y = 1|x)$. A formulação da variável latente tende a dar a impressão de que estamos interessados primeiramente nos efeitos de cada x_j sobre y^* . Como veremos, no logit e no probit, a direção do efeito de x_j sobre $E(y^*|x) = \beta_0 + x\beta$ e $E(y|x) = P(y = 1|x) = G(\beta_0 + x\beta)$ é sempre a mesma. Contudo, a variável latente y^* raramente tem uma unidade de medida bem definida. (Por exemplo, y^* pode ser a diferença, em níveis de utilidade, de duas ações diferentes.) Assim, as magnitudes de cada β_j não são, em si mesmas, de grande valia (ao contrário do que ocorre no modelo de probabilidade linear). Para muitos propósitos, queremos estimar o efeito de x_j sobre a probabilidade de êxito $P(y = 1|x)$, mas isso é complicado devido à natureza não-linear de $G(\cdot)$.

Para encontrarmos o efeito parcial de variáveis, aproximadamente contínuas, temos que confiar no cálculo. Se x_j for aproximadamente contínua, seu efeito parcial sobre $p(x) = P(y = 1|x)$ será obtido da derivada parcial:

$$\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + x\beta)\beta_j, \text{ onde } g(z) = \frac{dG}{dz}(z). \quad (17.7)$$

Como G é a fdc de uma variável aleatória contínua, g é uma função de densidade de probabilidade. Nos casos logit e probit, $G(\cdot)$ será uma fdc estritamente crescente, e assim $g(z) > 0$ para todo z . Portanto, o efeito parcial de x_j sobre $p(x)$ depende de x devido à quantidade positiva $g(\beta_0 + x\beta)$, o que significa que o efeito parcial sempre terá o mesmo sinal de β_j .

A equação (17.7) mostra que os efeitos *relativos* de duas variáveis explicativas contínuas quaisquer não dependem de x : a razão dos efeitos parciais de x_j e x_h é β_j/β_h . No caso típico em que g é uma densidade simétrica ao redor de zero, com uma única moda em zero, o maior efeito ocorre quando $\beta_0 + x\beta = 0$. Por exemplo, no caso probit com $g(z) = \phi(z)$, $g(0) = \phi(0) = 1/\sqrt{2\pi} \approx 0,40$. No caso logit, $g(z) = \exp(z)/[1 + \exp(z)]^2$ e, portanto, $g(0) = 0,25$.

Se, digamos, x_1 for uma variável explicativa binária, o efeito parcial de alterar x_1 de zero para um, mantendo-se todas as outras variáveis fixas, será simplesmente

$$G(\beta_0 + \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k) - G(\beta_0 + \beta_2 x_2 + \dots + \beta_k x_k). \quad (17.8)$$

Mais uma vez, isso depende de todos os valores dos outros x_j . Por exemplo, se y for um indicador de emprego e x_1 for uma variável *dummy* indicando a participação em um programa de treinamento de pessoal, então, (17.8) será a mudança na probabilidade do emprego devido ao programa de treinamento de pessoal; isso dependerá de outras características que afetem a empregabilidade, tais como a educação e a experiência. Observe que o conhecimento do sinal de β_1 será suficiente para determinar se o programa teve um efeito positivo ou negativo. Entretanto, para encontrar a *magnitude* do efeito, teremos que estimar a quantidade em (17.8).

Também podemos usar a diferença em (17.8) para outros tipos de variáveis discretas (como o número de filhos). Se x_k representar essa variável, o efeito sobre a probabilidade de x_k ir de c_k para $c_k + 1$ será simplesmente

$$G[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k (c_k + 1)] - G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k c_k). \quad (17.9)$$

A inclusão de formas funcionais padrão entre as variáveis explicativas é feita de forma direta. Por exemplo, no modelo

$$P(y = 1 | z) = G(\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3),$$

o efeito parcial de z_1 sobre $P(y = 1 | z)$ será $\partial P(y = 1 | z) / \partial z_1 = g(\beta_0 + x\beta)(\beta_1 + 2\beta_2 z_1)$, e o efeito parcial de z_2 sobre a probabilidade de resposta será $\partial P(y = 1 | z) / \partial z_2 = g(\beta_0 + x\beta)(\beta_3/z_2)$, onde $x\beta = \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3$. Portanto, $g(\beta_0 + x\beta)(\beta_3/100)$ será a mudança aproximada na probabilidade de resposta quando z_2 é aumentado em um por cento. Modelos com interações entre as variáveis explicativas, inclusive aqueles entre as variáveis discretas e contínuas, são tratados de forma semelhante. Quando quisermos medir os efeitos de variáveis discretas, deveremos usar (17.9).

Estimação de Máxima Verossimilhança de Modelos Logit e Probit

Como devemos estimar modelos de resposta binária não-linear? Para estimar o MPL, podemos usar mínimos quadrados ordinários (veja a Seção 7.5) ou, em alguns casos, mínimos quadrados ponderados (veja a Seção 8.5). Devido à natureza não-linear de $E(y|x)$, MQO e MQP não são aplicáveis. Poderíamos usar versões não-lineares desses métodos, mas o uso da **estimação de máxima verossimilhança (EMV)** não é mais complicada (veja uma discussão resumida no Apêndice B, disponível em www.thomsonlearning.com.br, na página deste livro). Até agora, precisamos muito pouco da EMV, embora tenhamos notado que, sob as hipóteses do modelo linear clássico, o estimador MQO é o estimador de máxima verossimilhança (condicional nas variáveis explicativas). Para estimar modelos de variáveis dependentes limitadas, os métodos de máxima verossimilhança são indispensáveis. Como a estimação de máxima verossimilhança é baseada na distribuição de y dado x , a heteroscedasticidade em $\text{Var}(y|x)$ é automaticamente considerada.

Assuma que temos uma amostra aleatória de tamanho n . Para obter o estimador de máxima verossimilhança, condicional nas variáveis explicativas, precisamos da densidade y_i dado x_i . Podemos escrever isso como

$$f(y|x_i;\beta) = [G(x_i\beta)]^y [1 - G(x_i\beta)]^{1-y}, \quad y = 0, 1, \quad (17.10)$$

onde, para simplificar, absorvemos o intercepto no vetor x_i . Podemos facilmente ver que quando $y = 1$, obtemos $G(x_i\beta)$ e quando $y = 0$, obtemos $1 - G(x_i\beta)$. A função log-verossimilhança da observação i é uma função dos parâmetros e dos dados (x_i, y_i) e é obtida tomando o log de (17.10):

$$\ell_i(\beta) = y_i \log[G(x_i\beta)] + (1 - y_i) \log[1 - G(x_i\beta)]. \quad (17.11)$$

Como $G(\cdot)$ está estritamente entre zero e um no logit e no probit, $\ell_i(\beta)$ será bem definido para todos os valores de β .

A log-verossimilhança de uma amostra de tamanho n é obtida pela soma de (17.11) para todas as observações: $\mathcal{L}(\beta) = \sum_{i=1}^n \ell_i(\beta)$. A EMV de β , representada por $\hat{\beta}$, maximiza essa log-verossimilhança.

Se $G(\cdot)$ for a fdc logit padrão, $\hat{\beta}$ será o *estimador logit*; se $G(\cdot)$ for a fdc normal padrão, então, $\hat{\beta}$ será o *estimador probit*.

Devido à natureza não-linear do problema de maximização, não podemos escrever fórmulas para as estimativas de máxima verossimilhança logit ou probit. Além de levantar problemas computacionais, isso torna a teoria estatística do logit e do probit muito mais difícil do que o MQO ou mesmo o MQ2E. No entanto, a teoria geral da EMV de amostras aleatórias implica que, sob condições bastante gerais, a EMV é consistente, assimptoticamente normal e assimptoticamente eficiente. [Veja uma discussão geral sobre este tópico em Wooldridge (2002, Capítulo 13).] Aqui, usaremos apenas os resultados: a aplicação de modelos logit e probit é razoavelmente fácil, desde que entendamos o significado das estatísticas.

Cada $\hat{\beta}_j$ vem com um erro-padrão (assimptótico), cuja fórmula é complicada e é apresentada no apêndice deste capítulo. Uma vez que tenhamos os erros-padrão — e eles são descritos com as estimativas dos coeficientes por qualquer programa econométrico que suporte probit e logit —, poderemos construir testes t e intervalos de confiança (assimptóticos), exatamente como nos métodos MQO e MQ2E, e como para os outros estimadores que tenhamos encontrado. Particularmente, para testar $H_0: \beta_j = 0$, formamos a estatística $t = \hat{\beta}_j / \text{ep}(\hat{\beta}_j)$ e conduzimos o teste da maneira habitual, logo que tenhamos decidido sobre uma alternativa unilateral ou bilateral.

Testes de Hipóteses Múltiplas

Também podemos testar restrições múltiplas em modelos logit e probit. Na maioria dos casos, esses serão testes de múltiplas restrições de exclusão, como na Seção 4.5. Aqui nos concentraremos nas restrições de exclusão.

Existem três maneiras de testarmos restrições de exclusão nos modelos logit e probit. O teste do multiplicador de Lagrange ou de escore exige que apenas se estime o modelo sob a hipótese nula, da mesma forma como no caso linear na Seção 5.2; não trataremos aqui do teste de escore, já que raramente é necessário testar as restrições de exclusão. [Veja Wooldridge (2002, Capítulo 15) sobre outros usos do teste de escore em modelos de resposta binária.]

O teste de Wald exige a estimação somente do modelo irrestrito. No caso do modelo linear, a **estatística de Wald**, após uma transformação simples, é essencialmente a estatística F , de modo que não há a necessidade de estudar a estatística de Wald separadamente. A fórmula da estatística de Wald é dada em Wooldridge (2002, Capítulo 15). Essa estatística é calculada por programas econométricos que permitem testar as restrições de exclusão após o modelo irrestrito ter sido estimado. Ela tem uma distribuição qui-quadrada assimptótica, com gl iguais ao número de restrições sendo testadas.

Se tanto o modelo restrito como o irrestrito forem fáceis de se estimar — como normalmente é o caso com restrições de exclusão —, então, o *teste da razão de verossimilhança (RV)* se torna bastante atraente. O teste RV baseia-se no mesmo conceito de o teste F em um modelo linear. O teste F mede o acréscimo na soma dos quadrados dos resíduos quando variáveis são retiradas do modelo. O teste RV baseia-se na diferença das funções log-verossimilhança dos modelos irrestrito e restrito. A idéia é a

seguinte: como a EMV maximiza a função log-verossimilhança, a eliminação de variáveis geralmente conduz a uma log-verossimilhança *menor* — ou, pelo menos, não maior. (Isso é semelhante ao fato de o *R*-quadrado nunca aumentar quando variáveis são eliminadas de uma regressão.) A questão é se a queda na log-verossimilhança será suficientemente grande para concluirmos que as variáveis eliminadas são importantes. Podemos tomar essa decisão logo que tenhamos uma estatística de teste e um conjunto de valores críticos.

A **estatística razão de verossimilhança** é o dobro da diferença nas log-verossimilhanças:

$$LR = 2(\mathcal{L}_{ir} - \mathcal{L}_r), \quad (17.12)$$

onde \mathcal{L}_{ir} é o valor de log-verossimilhança do modelo irrestrito e \mathcal{L}_r é o valor de log-verossimilhança do modelo restrito. Como $\mathcal{L}_{ir} \geq \mathcal{L}_r$, RV será não-negativa e usualmente estritamente positiva. Ao calcular a estatística RV de modelos de resposta binária, é importante saber que a função log-verossimilhança será sempre um número negativo. Esse fato deriva da equação (17.11), porque y_i é ou zero ou um, e ambas as variáveis no interior da função log estão estritamente entre zero e um, o que significa que seus logs naturais são negativos. O fato de as funções log-verossimilhança serem ambas negativas não altera a maneira de calcularmos a estatística RV ; simplesmente preservamos os sinais negativos na equação (17.12).

A multiplicação por dois em (17.12) é necessária para que a RV tenha uma distribuição qui-quadrada aproximada sob H_0 . Se estivermos testando q restrições de exclusão, $LR \stackrel{\text{aprox}}{\sim} X_q^2$. Isso significa que, para testar H_0 ao nível de 5%, usamos como nosso valor crítico o 95º percentil na distribuição X_q^2 . Computar os p -valores é fácil na maioria dos programas econométricos.

Um modelo probit para explicar se uma firma será adquirida por outra durante determinado ano é

$$P(aquisição = 1 | x) = \Phi(\beta_0 + \beta_1 lucrmed + \beta_2 valmerc + \beta_3 dividareceita + \beta_4 perceo + \beta_5 salceo + \beta_6 idadeceo)$$

onde *aquisição* é uma variável de resposta binária, *lucrmed* é a média da margem de lucro da firma de vários anos anteriores, *valmerc* é o valor de mercado da firma, *dividareceita* é a razão dívida/receitas, e *perceo*, *salceo* e *idadeceo* são permanência, salário anual e idade do diretor-executivo, respectivamente. Estabeleça a hipótese nula que, outros fatores permanecendo iguais, as variáveis relacionadas com o diretor-executivo não têm efeito sobre a probabilidade de a firma ser adquirida por terceiros. Quantos gl existem na distribuição qui-quadrada do teste RV ou de Wald?

A Interpretação das Estimativas Logit e Probit

Considerando os modernos computadores, de uma perspectiva prática, o aspecto mais difícil dos modelos logit e probit é apresentar e interpretar os resultados. As estimativas dos coeficientes, seus erros-padrão e o valor da função log-verossimilhança são descritos por todos os programas econométricos que executam logit e probit, e essas informações devem ser descritas em qualquer aplicação. Os coefi-

cientes dão os sinais dos efeitos parciais de cada x_i sobre a probabilidade de resposta, e a significância estatística de x_i é determinada pela condição de podermos rejeitar $H_0: \beta_i = 0$ a um nível de significância suficientemente pequeno.

Uma medida do grau de ajuste que normalmente é apresentada é a chamada **percentagem corretamente prevista**, que é calculada da seguinte maneira: para cada i , calculamos a probabilidade estimada de que y_i assuma o valor um, $G(\hat{\beta}_0 + x_i \hat{\beta})$. Se $G(\hat{\beta}_0 + x_i \hat{\beta}) > 0,5$, a previsão de y_i será um, e se $G(\hat{\beta}_0 + x_i \hat{\beta}) \leq 0,5$, a previsão de y_i será zero. A percentagem de vezes em que y_i previsto se iguala ao y_i verdadeiro (que sabemos ser zero ou um) é a percentagem corretamente prevista. Esse indicador tem a sua utilidade, mas é possível obter percentagens corretamente previstas bastante elevadas sem que o modelo seja de grande utilidade. Por exemplo, suponha que em uma amostra com 200 observações, 180 tenham $y_i = 0$, e 150 tenham sido previstas como sendo zero usando a regra anterior. Mesmo que nenhuma de nossas previsões seja correta quando $y_i = 1$, ainda estaremos prevendo corretamente 75% dos resultados. Devido a exemplos como esse, faz sentido descrever a percentagem corretamente prevista de cada um dos dois resultados. O Problema 17.1 pede para você mostrar que a percentagem corretamente prevista global é uma média ponderada de \hat{q}_0 (a percentagem corretamente prevista de $y_i = 0$) e \hat{q}_1 (a percentagem corretamente prevista de $y_i = 1$), onde os pesos são as frações de zeros e uns na amostra, respectivamente.

Existem também vários indicadores de **pseudos R-quadrados** de resposta binária. McFadden (1974) sugere o indicador $1 - \mathcal{L}_{ii}/\mathcal{L}_0$, onde \mathcal{L}_{ii} é a função log-verossimilhança do modelo estimado e \mathcal{L}_0 é a função log-verossimilhança no modelo com somente um intercepto. Por que esse indicador faz sentido? Lembre-se de que as funções log-verossimilhança são negativas e, portanto, $\mathcal{L}_{ii}/\mathcal{L}_0 = |\mathcal{L}_{ii}|/|\mathcal{L}_0|$. Além disso, $|\mathcal{L}_{ii}| \leq |\mathcal{L}_0|$. Se as covariadas não tiverem poder explicativo, então, $\mathcal{L}_{ii}/\mathcal{L}_0 = 1$, e o pseudo R -quadrado será zero, da mesma forma que o R -quadrado normal é zero em uma regressão linear quando as covariadas não têm poder explicativo. Em geral, $|\mathcal{L}_{ii}| < |\mathcal{L}_0|$, caso em que $1 - \mathcal{L}_{ii}/\mathcal{L}_0 > 0$. Se \mathcal{L}_{ii} fosse zero, o pseudo R -quadrado seria igual à unidade. Aliás, \mathcal{L}_{ii} não pode atingir zero em um modelo probit ou logit, já que isso exigiria que as probabilidades estimadas quando $y_i = 1$ fossem todas a unidade e as probabilidades estimadas quando $y_i = 0$ fossem todas zero.

R -quadrados alternativos do probit e logit estão mais diretamente relacionados com o R -quadrado habitual da estimação por MQO de um modelo de probabilidade linear. Para o probit ou logit, defina $\hat{y}_i = G(\hat{\beta}_0 + x_i \hat{\beta})$ como as probabilidades estimadas. Como essas probabilidades são, também, estimativas de $E(y_i | x_i)$, podemos basear um R -quadrado na proximidade de \hat{y}_i com y_i . Uma possibilidade que é sugerida por uma análise padrão de regressão é computar a correlação quadrada entre y_i e \hat{y}_i . Lembre-se de que, em uma estrutura de regressão linear, essa é uma maneira algebraicamente equivalente de obter o R -quadrado habitual; veja a equação (3.29). Portanto, podemos computar um pseudo R -quadrado do probit e logit que seja diretamente comparável ao habitual R -quadrado da estimação de um modelo de probabilidade linear. Em qualquer caso, o grau de ajuste é menos importante que tentar obter estimativas convincentes dos efeitos *ceteris paribus* das variáveis explicativas.

Freqüentemente, queremos estimar os efeitos das x_i sobre as probabilidades de resposta, $P(y = 1 | x)$. Se x_i for (em linhas gerais) contínuo, então,

$$\Delta \hat{P}(y = 1 | x) \approx [g(\hat{\beta}_0 + x \hat{\beta}) \hat{\beta}_i] \Delta x_i, \quad (17.13)$$

para “pequenas” alterações em x_i . Como $g(\hat{\beta}_0 + x \hat{\beta})$ depende de x , devemos computar a expressão com valores de interesse de x . Muitas vezes, as médias amostrais de x_i são utilizadas para obter $g(\hat{\beta}_0 + \bar{x} \hat{\beta})$. Esse fator pode ser usado para ajustar cada uma das $\hat{\beta}_i$ (ou pelo menos os das variáveis contínuas) para obter o efeito do aumento de uma unidade em x_i . Se x contiver funções não-lineares de algumas variá-

veis explicativas, tais como o log natural ou potências, temos a opção de usar a média na função não-linear ou determinar a média da função não-linear. Para obter o efeito da média unitária na população, faz sentido usar a primeira opção. Se um programa econômétrico escalonar automaticamente os coeficientes por $g(\hat{\beta}_0 + \bar{x}\hat{\beta})$, ele necessariamente calculará a média da função não-linear, já que ele não tem condições de identificar quando uma variável explicativa é uma função não-linear ou uma variável adjacente. A diferença raramente será grande.

Algumas vezes, valores mínimos e máximos, ou quartis inferiores e superiores, de variáveis importantes são usados na obtenção de $g(\hat{\beta}_0 + \bar{x}\hat{\beta})$, de forma que podemos ver como os efeitos parciais se alteram quando alguns elementos de \bar{x} ficam grandes ou pequenos.

A equação (17.13) também sugere como, aproximadamente, comparar magnitudes das estimativas de inclinação do probit e do logit. Como mencionado anteriormente, para o probit, $g(0) \approx 0,4$ e para o logit, $g(0) = 0,25$. Portanto, para tornarmos comparáveis as estimativas de inclinação do probit e do logit, podemos multiplicar as estimativas probit por $0,4/0,25 = 1,6$, ou multiplicar as estimativas logit por 0,625. No modelo de probabilidade linear, $g(0)$ é efetivamente 1, e assim as estimativas de inclinação do logit devem ser divididas por 4 para torná-las aproximadamente comparáveis com as estimativas do MPL; as estimativas de inclinação do probit devem ser divididas por 2,5 para torná-las comparáveis com as estimativas do MPL. Uma comparação mais precisa é multiplicar o coeficiente de inclinação do probit por $\phi(\hat{\beta}_0 + \bar{x}\hat{\beta})$ e o coeficiente de inclinação do logit por $\exp(\hat{\beta}_0 + \bar{x}\hat{\beta})/[1 + \exp(\hat{\beta}_0 + \bar{x}\hat{\beta})]^2$, onde as estimativas são probit ou logit, respectivamente.

Se, digamos, x_k for uma variável binária, pode fazer sentido usar zero ou um em x_k , em vez de \bar{x}_k (que é a fração de uns na amostra). A utilização das médias das variáveis binárias significa que o efeito na realidade não corresponde a um indivíduo em particular. Entretanto, muitas vezes os resultados são semelhantes e a escolha é, realmente, baseada no gosto de cada um.

Se x_k for uma variável discreta, então, poderemos estimar a alteração na probabilidade prevista conforme ela passa de c_k para $c_k + 1$ por

$$\begin{aligned} &G[\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \dots + \hat{\beta}_{k-1}\bar{x}_{k-1} + \hat{\beta}_k(c_k + 1)] \\ &- G(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \dots + \hat{\beta}_{k-1}\bar{x}_{k-1} + \hat{\beta}_kc_k). \end{aligned} \quad (17.14)$$

Em particular, quando x_k é uma variável binária, definimos $c_k = 0$. Naturalmente, temos que decidir o que usar para as outras variáveis explicativas; em geral, usamos médias das variáveis aproximadamente contínuas.

EXEMPLO 17.1

(Participação das Mulheres Casadas na Força de Trabalho)

Agora usamos os dados do arquivo MROZ.RAW para estimar o modelo de participação na força de trabalho do Exemplo 8.8 — veja também a Seção 7.5 — por logit e probit. Também descrevemos as estimativas do modelo de probabilidade linear do Exemplo 8.8, usando os erros-padrão robustos em relação à heteroscedasticidade. Os resultados, com os erros-padrão entre parênteses, são apresentados na Tabela 17.1.

EXEMPLO 17.1 (continuação)**Tabela 17.1**

Estimativas MPL, Logit e Probit da Participação na Força de Trabalho

Variável Dependente: <i>naft</i>			
Variáveis Independentes	MPL	MQO	EMV
<i>nesprend</i>	-0,0034 (0,0015)	-0,021 (0,008)	-0,012 (0,005)
<i>educ</i>	0,038 (0,007)	0,221 (0,043)	0,131 (0,025)
<i>exper</i>	0,039 (0,006)	0,206 (0,032)	0,123 (0,019)
<i>exper</i> ²	-0,0060 (0,00018)	-0,0032 (0,0010)	-0,0019 (0,0006)
<i>idade</i>	-0,016 (0,002)	-0,088 (0,015)	-0,053 (0,008)
<i>crianmed6</i>	-0,262 (0,032)	-1,443 (0,204)	-0,868 (0,119)
<i>crianma6</i>	0,013 (0,013)	0,060 (0,075)	0,036 (0,043)
<i>constante</i>	0,586 (0,151)	0,425 (0,860)	0,270 (0,509)
<i>Percentagem Corretamente Prevista</i>	73,4	73,6	73,4
<i>Valor de Log-Verossimilhança</i>	—	-401,77	-401,30
<i>Pseudo R-Quadrado</i>	0,264	0,220	0,221

As estimativas dos três modelos contam uma história consistente. Os sinais dos coeficientes são os mesmos em todos os modelos, e as mesmas variáveis são estatisticamente significantes em cada modelo. O pseudo *R*-quadrado do MPL é o mesmo *R*-quadrado usual descrito pelo MQO; no logit e probit, o pseudo *R*-quadrado é o indicador baseado nas log-verossimilhanças descritas anteriormente.

Como já enfatizamos, as magnitudes dos coeficientes não são diretamente comparáveis entre os modelos. Usando a regra prática de aproximação discutida anteriormente, podemos dividir as estimativas logit por 4 e as estimativas probit por 2,5 para torná-las comparáveis com as estimativas MPL. Por exemplo, para os coeficientes de *crianmed6*, a estimativa logit escalonada está em torno de -0,361, e a estimativa probit escalonada está em torno de -0,347. Essas são magnitudes maiores do que a estimativa MPL (por razões que explicaremos mais tarde). De forma semelhante, o coeficiente escalonado de *educ* é 0,055 no logit e

EXEMPLO 17.1 (continuação)

0,052 no probit; esses valores também são maiores do que a estimativa MPL de 0,038, mas não diferem muito entre si.

Se avaliarmos a função de densidade de probabilidade normal padrão, $\phi(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k)$, nos valores médios das variáveis independentes na amostra (inclusive a média de $exper^2$), o resultado será aproximadamente 0,391; isso é suficientemente próximo de 0,4 para tornar útil a regra prática de aproximação para escalarizar os coeficientes probit para a obtenção dos efeitos sobre a probabilidade de resposta. Em outras palavras, para estimar a alteração na probabilidade de resposta devida ao aumento de uma unidade em qualquer variável independente, multiplicamos o coeficiente probit correspondente por 0,4. Quando computamos o fator de escalação do caso logístico, obtemos cerca de 0,243; sem surpresa, esse valor está próximo do maior valor possível, 0,25. Assim, podemos com segurança multiplicar os coeficientes logit por 0,25 para torná-los comparáveis com as estimativas de probabilidade linear.

A maior diferença entre o MPL e os modelos logit e probit é que o MPL assume efeitos marginais constantes para $educ$, $crianmed6$, e assim por diante, enquanto os modelos logit e probit implicam magnitudes decrescentes dos efeitos parciais. No MPL, estima-se que uma criança a mais reduz a probabilidade de participação na força de trabalho em aproximadamente 0,262, independentemente de quantos filhos pequenos a mulher já tenha (e independentemente dos níveis das outras variáveis explicativas). Podemos contrastar isso com o efeito marginal estimado pelo probit. Em termos concretos, consideremos uma mulher com $nesprend = 20,13$, $educ = 12,3$, $exper = 10,6$ e $idade = 42,5$ — que são, grosso modo, as médias da amostra — e $crianmed6 = 1$. Qual será a redução estimada na probabilidade de trabalhar quando o número de crianças pequenas aumenta de zero para um? Avaliamos a fdc normal padrão, $\Phi(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k)$, com $crianmed6 = 1$ e $crianmed6 = 0$, e com as outras variáveis independentes definidas com os valores precedentes. Obtemos aproximadamente $0,373 - 0,707 = -0,334$, o que significa que a probabilidade de participação na força de trabalho será cerca de 0,334 menor quando uma mulher tem um filho pequeno. Isso não é muito diferente do coeficiente probit escalonado de $-0,347$. Se uma mulher passa de uma para duas crianças pequenas, a probabilidade cai ainda mais, mas o efeito marginal não é tão grande: $0,117 - 0,373 = -0,256$. Curiosamente, a estimativa do modelo de probabilidade linear, que pretensamente estima o efeito próximo da média, está, na realidade, entre essas duas estimativas.

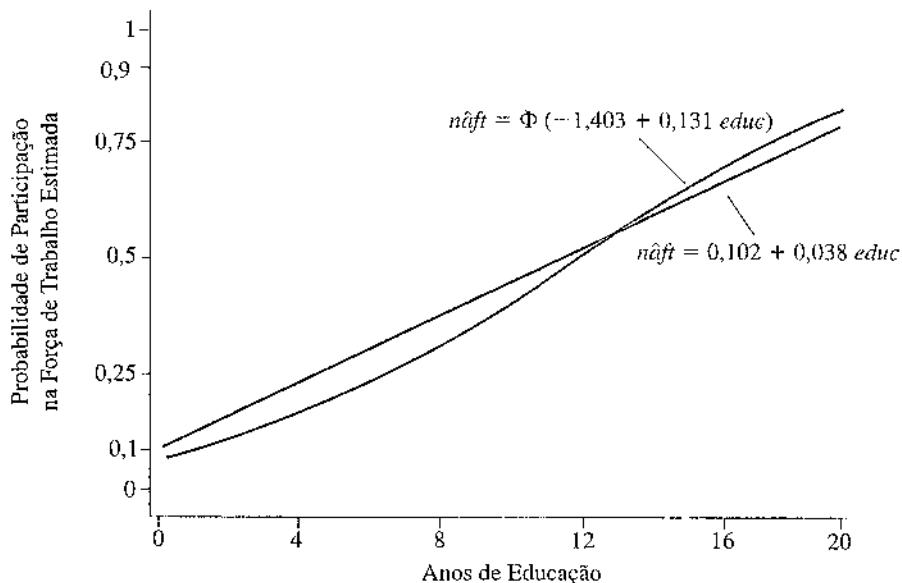
Usando as estimativas probit e a aproximação por cálculo, qual será a alteração aproximada na probabilidade de resposta quando $exper$ aumenta de 10 para 11?

A Figura 17.2 ilustra como as probabilidades de resposta estimadas a partir de modelos não-lineares de respostas binárias podem diferir daquelas do modelo de probabilidade linear. A probabilidade estimada da participação na força de trabalho é traçada em relação aos anos de educação para o modelo de probabilidade linear e o modelo probit. (O gráfico do modelo logit é muito semelhante ao do modelo probit.) Em ambos os casos, as variáveis explicativas, outras que não $educ$, são definidas com o valor de suas médias amostrais. Especificamente, as duas equações traçadas são $\hat{p}_{fit} = 0,102 + 0,038 educ$ para o modelo linear e $\hat{p}_{fit} = \Phi(-1,403 + 0,131 educ)$. Em níveis mais baixos de educação, o modelo de probabilidade linear estima probabilidades de participação na força de trabalho mais altas do que o modelo probit. Por exemplo, com oito anos de educação, o modelo de probabilidade linear estima uma probabilidade de participação na força de trabalho de 0,406, enquanto a estimativa

do modelo probit é de aproximadamente 0,361. As estimativas são as mesmas ao redor de 11 1/3 anos de educação. Em níveis mais altos de educação, o modelo probit prevê probabilidades de participação na força de trabalho mais altas. Na amostra, o menor nível de educação é de cinco anos e o maior é de 17 anos, de modo que não devemos fazer comparações fora dessa faixa.

Figura 17.2

Probabilidade de Resposta Estimada em Relação à Educação para os Modelos de Probabilidade Linear e Probit



Os mesmos problemas relativos às variáveis explicativas endógenas em modelos lineares também aparecem nos modelos logit e probit. Não temos espaço para tratar deles, mas é possível testar e corrigir variáveis explicativas endógenas com o uso de métodos relacionados aos mínimos quadrados em dois estágios. Evans e Schwab (1995) estimaram um modelo probit analisando se um aluno fazia curso superior, onde a variável explicativa principal era uma variável *dummy* que indicava se o aluno freqüentava uma escola católica. Evans e Schwab estimaram um modelo por máxima verossimilhança que permitia que o fato de o aluno freqüentar uma escola católica fosse considerado endógeno. [Veja Wooldridge (2002, Capítulo 15) para uma explicação desses métodos.]

Dois outros problemas receberam atenção no contexto dos modelos probit. O primeiro é a não-normalidade de e no modelo de variável latente (17.6). Naturalmente, se e não tiver uma distribuição normal padrão, a probabilidade de resposta não terá a forma probit. Alguns autores tendem a enfatizar a inconsistência de estimar β_j , mas esse é o foco errado, a menos que estejamos interessados somente na direção dos efeitos. Como a probabilidade de resposta é desconhecida, não podemos estimar a magnitude dos efeitos parciais mesmo que tenhamos estimativas consistentes de β_j .

Um segundo problema de especificação, também definido em termos do modelo de variável latente, é a heteroscedasticidade em e . Se $\text{Var}(e|x)$ depender de x , a probabilidade de resposta não mais terá a forma $G(\beta_0 + x\beta)$; ao contrário, ela dependerá da forma da variância e exigirá estimação mais gené-

rica. Tais modelos, na prática, não são usados com muita freqüência, já que o logit e o probit com formas funcionais flexíveis nas variáveis independentes tendem a funcionar bem.

Modelos de resposta binária aplicam-se com pequenas modificações a cortes transversais agrupados independentemente ou a outros conjuntos de dados nos quais as observações são independentes, mas não necessariamente identicamente distribuídas. Muitas vezes, variáveis *dummy* anuais ou de outro período de tempo são incluídas para avaliar efeitos temporais agregados. Da mesma forma que com os modelos lineares, o logit e o probit podem ser usados para avaliar o impacto de certas decisões no contexto de uma experimentação natural.

O modelo de probabilidade linear pode ser aplicado com dados de painel; em geral, ele será estimado por efeitos fixos (veja o Capítulo 14). Modelos logit e probit com efeitos não-observados vêm se tornando populares. Esses modelos são complicados pela natureza não-linear das probabilidades de resposta, e são difíceis de estimar e interpretar. [Veja Wooldridge (2002, Capítulo 15).]

17.2 O MODELO TOBIT PARA RESPOSTA DE SOLUÇÃO DE CANTO

Conforme mencionado no capítulo introdutório, outro tipo importante de variável dependente limitada é uma resposta de solução de canto. Esse tipo de variável é zero para uma fração não desprezível da população, mas é aproximadamente distribuída de forma contínua ao longo de valores positivos. Um exemplo é o valor gasto por um indivíduo com bebida alcoólica em determinado mês. Na população de pessoas com mais de 21 anos dos Estados Unidos, essa variável assume uma ampla gama de valores. Para certa fração significante, o montante gasto com álcool é zero. A seguinte abordagem omite a verificação de alguns detalhes relacionados ao modelo Tobit. [Eles são dados em Wooldridge (2002, Capítulo 16).]

Seja y uma variável essencialmente contínua ao longo de valores estritamente positivos, mas que assuma zero com probabilidade positiva. Nada nos impede de usar um modelo linear para y . Na verdade, um modelo linear pode ser uma boa aproximação de $E(y|x_1, x_2, \dots, x_k)$, especialmente para x_j próximos dos valores médios. Mas provavelmente obteremos valores estimados negativos, o que conduz a previsões negativas de y ; isso é parecido com os problemas do MPL de resultados binários. Além disso, a hipótese de que uma variável explicativa que apareça em forma de nível tenha um efeito parcial constante sobre $E(y|x)$ pode ser enganosa. Provavelmente, $\text{Var}(y|x)$ será heteroscedástica, embora possamos facilmente lidar com a heteroscedasticidade generalizada computando erros-padrão e estatísticas de testes robustos. Como a distribuição de y se acumula em zero, claramente y não pode ter uma distribuição normal condicional. Portanto, toda a inferência terá somente justificativa assintótica, como acontece com o modelo de probabilidade linear.

Em alguns casos, é importante ter um modelo que implique valores previstos não-negativos de y , e que tenha efeitos parciais sensíveis sobre uma ampla faixa das variáveis explicativas. Mais ainda, algumas vezes queremos estimar características da distribuição de y , dados outros x_1, \dots, x_k além do valor esperado condicional. O **modelo Tobit** é bastante conveniente para esses propósitos. Em geral, o modelo Tobit expressa a resposta observada, y , em termos de uma variável latente subjacente:

$$y^* = \beta_0 + x\beta + u, \quad u|x \sim \text{Normal}(0, \sigma^2) \quad (17.15)$$

$$y = \max(0, y^*). \quad (17.16)$$

A variável latente y^* satisfaz as hipóteses do modelo linear clássico; em particular, ela tem uma distribuição normal, homoscedástica, com uma média condicional linear. A equação (17.16) indica que a

variável observada, y , será igual a y^* quando $y^* \geq 0$, mas $y = 0$ quando $y^* < 0$. Como y^* é normalmente distribuída, y terá uma distribuição contínua sobre valores estritamente positivos. Em particular, a densidade de y , dado x será igual à densidade de y^* dado x para valores positivos. Além disso,

$$\begin{aligned} P(y = 0|x) &= P(y^* < 0|x) = P(u < -x\beta|x) \\ &= P(u/\sigma < -x\beta/\sigma|x) = \Phi(-x\beta/\sigma) = 1 - \Phi(x\beta/\sigma), \end{aligned}$$

devido ao fato de u/σ ter uma distribuição normal padrão e ser independente de x ; absorvemos o intercepto em x por simplicidade de notação. Portanto, se (x_i, y_i) for retirada aleatoriamente da população, a densidade de y_i , dado x_i , será

$$(2\pi\sigma^2)^{-1/2}\exp[-(y - x_i\beta)^2/(2\sigma^2)] = (1/\sigma)\phi[(y - x_i\beta)/\sigma], y > 0 \quad (17.17)$$

$$P(y_i = 0|x_i) = 1 - \Phi(x_i\beta/\sigma), \quad (17.18)$$

onde ϕ é a função densidade normal padrão.

De (17.17) e (17.18), podemos obter a função log-verossimilhança de cada observação i :

$$\begin{aligned} \ell_i(\beta, \sigma) &= 1(y_i = 0)\log[1 - \Phi(x_i\beta/\sigma)] \\ &\quad + 1(y_i > 0)\log\{(1/\sigma)\phi[(y_i - x_i\beta)/\sigma]\}; \end{aligned} \quad (17.19)$$

observe como essa função depende de σ , o desvio-padrão de u , como também de β_j . A log-verossimilhança de uma amostra aleatória de tamanho n é obtida somando (17.19) ao longo de todas as observações i . As estimativas de máxima verossimilhança de β e σ são obtidas pela maximização da log-verossimilhança; isso exigirá métodos numéricos, embora na maioria dos casos isso seja facilmente feito usando uma rotina de um programa econométrico.

Exemplo: casais extraconjugal

Seja y o número de casos extraconjugal de uma mulher casada da população dos Estados Unidos; gostaríamos de explicar essa variável em termos de outras características da mulher — em especial se ela trabalha fora de casa — assim como de seu marido e sua família. Esse exemplo seria um bom candidato para um modelo Tobit?

Como nos casos logit e probit, cada estimativa Tobit é acompanhada de um erro-padrão, e isso pode ser usado para construir estatísticas t de cada $\hat{\beta}_j$; a forma matricial usada para encontrar os erros-padrão é complicada e não será apresentada aqui. [Veja, por exemplo, Wooldridge (2002, Capítulo 16).]

O teste de múltiplas restrições de exclusão é feito facilmente com o uso do teste de Wald ou do teste da razão de verossimilhança. O teste de Wald tem uma forma semelhante às dos casos logit e probit; o teste RV é sempre dado por (17.12), onde, é claro, usamos as funções log-verossimilhança Tobit dos modelos com e sem restrições.

A Interpretação das Estimativas Tobit

Com o uso de computadores modernos, as estimativas de máxima verossimilhança dos modelos Tobit usualmente não são muito mais difíceis de serem obtidas do que as estimativas MQO de um modelo linear. Além disso, os resultados do Tobit e do MQO freqüentemente são semelhantes. Isso torna tentador interpretar os $\hat{\beta}_j$ do Tobit como se fossem estimativas de uma regressão linear. Infelizmente, as coisas não são tão fáceis.

A partir da equação (17.15), vemos que os $\hat{\beta}_j$ medem os efeitos parciais dos x_j sobre $E(y^*|x)$, onde y^* é a variável latente. Algumas vezes, y^* tem um significado econômico interessante, mas, na maioria das vezes, não. A variável que queremos explicar é y , já que ela é o resultado observado (tal como horas trabalhadas ou montante de contribuições de caridade). Por exemplo, por uma questão de critério de decisão, estamos interessados na sensibilidade das horas trabalhadas quanto a mudanças na alíquota de um imposto.

Podemos estimar $P(y = 0|x)$ a partir de (17.18), a qual, naturalmente, permite estimar $P(y > 0|x)$. O que acontecerá se quisermos estimar o valor esperado de y como uma função de x ? Em modelos Tobit, dois valores esperados são de especial interesse: $E(y|y > 0,x)$, que algumas vezes é chamada de “valor esperado condicional” por ser condicional a $y > 0$, e $E(y|x)$, que, infelizmente, é chamado de “valor esperado incondicional”. (Ambos os valores esperados são condicionais às variáveis explicativas.) $E(y|y > 0,x)$ nos informa, para determinados valores de x , o valor esperado de y da subpopulação onde y é positivo. Dado $E(y|y > 0,x)$, podemos facilmente encontrar $E(y|x)$:

$$E(y|x) = P(y > 0|x) \cdot E(y|y > 0,x) = \Phi(x\beta/\sigma) \cdot E(y|y > 0,x). \quad (17.20)$$

Para obter $E(y|y > 0,x)$, usamos um resultado das variáveis aleatórias normalmente distribuídas: se $z \sim \text{Normal}(0,1)$, então, $E(z|z > c) = \phi(c)/[1 - \Phi(c)]$ para qualquer constante c . Mas $E(y|y > 0,x) = x\beta + E(u|u > -x\beta) = x\beta + \sigma E[(u/\sigma)|(u/\sigma) > -x\beta/\sigma] = x\beta + \sigma \phi(x\beta/\sigma)/\Phi(x\beta/\sigma)$, porque $\phi(-c) = \phi(c)$, $1 - \Phi(-c) = \Phi(c)$ e u/σ tem uma distribuição normal padrão independente de x .

Podemos resumir isso como

$$E(y|y > 0,x) = x\beta + \sigma\lambda(x\beta/\sigma), \quad (17.21)$$

onde $\lambda(c) = \phi(c)/\Phi(c)$ é chamado de **razão de Mills inversa**; essa é a razão entre a fdp normal padrão e a fdc normal padrão, cada uma avaliada em c .

A equação (17.21) é importante. Ela mostra que o valor esperado de y condicional a $y > 0$ é igual a $x\beta$, mais um termo estritamente positivo, que é σ vezes a razão de Mills inversa avaliada em $x\beta/\sigma$. Essa equação também mostra porque o uso do MQO somente para observações onde $y_j > 0$ nem sempre estimará β consistentemente; essencialmente, a razão de Mills inversa é uma variável omitida, e geralmente ela é correlacionada com os elementos de x .

A combinação da (17.20) com (17.21) produz

$$E(y|x) = \Phi(x\beta/\sigma)(x\beta + \sigma\lambda(x\beta/\sigma)) = \Phi(x\beta/\sigma)x\beta + \sigma\phi(x\beta/\sigma), \quad (17.22)$$

onde a segunda igualdade decorre porque $\Phi(x\beta/\sigma)\lambda(x\beta/\sigma) = \phi(x\beta/\sigma)$. Essa equação mostra que, quando y segue um modelo Tobit, $E(y|x)$ será uma função não-linear de x e β . Embora não seja óbvio,

pode ser mostrado que o lado direito da equação (17.22) será positivo para quaisquer valores de x e de β . Portanto, logo que tivermos as estimativas de β , podemos ter certeza de que os valores previstos de y — isto é, estimativas de $E(y|x)$ — são positivos. O custo de garantir previsões positivas de y é que a equação (17.22) é mais complicada que um modelo linear de $E(y|x)$. Mais importante ainda, os efeitos parciais de (17.22) são mais complicados do que os de um modelo linear. Como veremos, os efeitos parciais de x_j sobre $E(y|y > 0, x)$ e sobre $E(y|x)$ têm o mesmo sinal do coeficiente, β_j , mas a magnitude dos efeitos depende de todos os valores das variáveis explicativas e dos parâmetros. Como σ aparece em (17.22), não é de surpreender que os efeitos parciais também dependam de σ .

Se x_j for uma variável contínua, poderemos encontrar os efeitos parciais usando cálculo. Primeiro,

$$\frac{\partial E(y|y > 0, x)}{\partial x_j} = \beta_j + \beta_j \cdot \frac{d\lambda}{dc}(x\beta/\sigma),$$

assumindo que x_j não seja funcionalmente relacionado a outros regressores. Fazendo a diferenciação de $\lambda(c) = \phi(c)/\Phi(c)$ e usando $d\Phi/dc = \phi(c)$ e $d\phi/dc = -c\phi(c)$, pode ser demonstrado que $d\lambda/dc = -\lambda(c)[c + \lambda(c)]$. Portanto,

$$\frac{\partial E(y|y > 0, x)}{\partial x_j} = \beta_j[1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]]. \quad (17.23)$$

Isso mostra que o efeito parcial de x_j sobre $E(y|y > 0, x)$ não é determinado apenas por β_j . O fator de ajuste é dado pelo termo entre chaves $\{\cdot\}$, e depende de uma função linear de x , $x\beta/\sigma = (\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)/\sigma$. Pode ser mostrado que o fator de ajuste está estritamente entre zero e um. Na prática, podemos estimar (17.23) inserindo as EMVs de β_j e σ . Como com os modelos logit e probit, devemos inserir valores de x_j , usualmente os valores médios ou outros valores interessantes. A equação (17.23) revela um ponto sutil que algumas vezes é perdido na aplicação do modelo Tobit em respostas de solução de canto: o parâmetro σ aparece diretamente nos efeitos parciais, e assim ter uma estimativa de σ é crucial para estimar os efeitos parciais. Algumas vezes, σ é chamado de um parâmetro “ancilar” (o que significa que ele é auxiliar, ou sem importância). Embora seja verdade que o valor de σ não afeta o sinal dos efeitos parciais, ele afeta as magnitudes, e freqüentemente estamos interessados na importância econômica das variáveis explicativas. Portanto, caracterizar σ como ancilar é equivocado e advém de uma confusão entre o modelo Tobit de aplicações de solução de canto e aplicações de censura de dados. (Veja a Seção 17.4.)

Todas as quantidades econômicas habituais, como as elasticidades, podem ser computadas. Por exemplo, a elasticidade de y em relação a x_1 , condicional a $y > 0$, é

$$\frac{\partial E(y|y > 0, x)}{\partial x_1} \cdot \frac{x_1}{E(y|y > 0, x)}. \quad (17.24)$$

Essa equação pode ser computada quando x_1 aparece em várias formas funcionais, inclusive nas formas em nível, logarítmica e quadrática.

Se x_1 for uma variável binária, o efeito de interesse é obtido como a diferença entre $E(y|y > 0, x)$, com $x_1 = 1$ e $x_1 = 0$. Efeitos parciais que envolvam outras variáveis discretas (como o número de filhos) podem ser tratados de maneira semelhante.

Podemos usar (17.22) para encontrar a derivada parcial de $E(y|\mathbf{x})$ em relação a x_j contínua. Essa derivada leva em conta o fato de que as pessoas que iniciam em $y = 0$ podem escolher $y > 0$ quando x_j muda:

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \frac{\partial P(y > 0|\mathbf{x})}{\partial x_j} \cdot E(y|y > 0, \mathbf{x}) + P(y > 0|\mathbf{x}) \cdot \frac{\partial E(y|y > 0, \mathbf{x})}{\partial x_j}. \quad (17.25)$$

Como $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$,

$$\frac{\partial P(y > 0|\mathbf{x})}{\partial x_j} = (\beta_j/\sigma)\phi(\mathbf{x}\boldsymbol{\beta}/\sigma), \quad (17.26)$$

e, dessa forma, podemos estimar cada termo em (17.25), assim que inserirmos as EMVs de β_j e σ e valores particulares de x_j .

É importante observar que ao inserir (17.23) e (17.26) em (17.25) e usar o fato de que $\Phi(c)\lambda(c) = \phi(c)$ para qualquer c , obtém-se

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \beta_j \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma). \quad (17.27)$$

A equação (17.27) permite fazer comparações aproximadas entre as estimativas MQO e Tobit. [A equação (17.27) também pode ser derivada diretamente da equação (17.22) usando o fato de que $\phi(z)/dz = -z\phi(z)$.] Os coeficientes de inclinação MQO, digamos $\hat{\gamma}_j$, da regressão de y_i sobre $x_{i1}, x_{i2}, \dots, x_{ik}$, $i = 1, \dots, n$ — isto é, usando todos os dados — são estimativas diretas de $\partial E(y|\mathbf{x})/\partial x_j$. Para tornar o coeficiente Tobit, $\hat{\beta}_j$, comparável com $\hat{\gamma}_j$, devemos multiplicar $\hat{\beta}_j$ por um fator de ajuste. Em geral, avaliamos as variáveis explicativas em seus valores médios para obter um fator de ajuste comum. Assim, um coeficiente de inclinação Tobit de uma variável explicativa contínua (aproximadamente) é multiplicado por $\Phi(\bar{x}\hat{\beta}/\hat{\sigma})$, um fator que estará, necessariamente, entre zero e um. Como $P(y > 0|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$, $\Phi(\bar{x}\hat{\beta}/\hat{\sigma})$ será uma estimativa consistente de $P(y > 0|\mathbf{x})$ quando cada variável explicativa assumir seu valor médio. Conforme $P(y > 0|\mathbf{x} = \bar{\mathbf{x}})$ se aproxima de um, os coeficientes Tobit se aproximam dos coeficientes MQO. (No caso extremo onde $y_i > 0$ para todo i , o Tobit e o MQO produzirão estimativas idênticas dos parâmetros.)

Infelizmente, quanto às variáveis explicativas discretas, a comparação das estimativas MQO e Tobit não é tão fácil. Para o Tobit, o efeito parcial de uma variável explicativa discreta, por exemplo, uma variável explicativa binária, deve, na realidade, ser obtido estimando-se $E(y|\mathbf{x})$ a partir da equação (17.22). Por exemplo, se x_1 for binária, primeiro devemos inserir $x_1 = 1$ e depois $x_1 = 0$ ao calcularmos as estimativas de $E(y|\mathbf{x})$, onde as outras variáveis explicativas podem ser definidas pelos seus valores médios ou outros valores de interesse.

EXEMPLO 17.2**(Oferta de Mão-de-Obra Anual de Mulheres Casadas)**

O arquivo MROZ.RAW inclui dados sobre horas trabalhadas de 753 mulheres casadas, 428 das quais trabalharam fora de casa por um salário durante o ano; 325 mulheres trabalharam zero horas.

Tabela 17.2

Estimação MQO e Tobit de Horas Anuais Trabalhadas

Variável Dependente: <i>horas</i>		
Variáveis Independentes	Linear (MQO)	Tobit (EMV)
<i>nesprend</i>	-3,45 (2,54)	-8,81 (4,46)
<i>educ</i>	28,76 (12,95)	80,65 (21,58)
<i>exper</i>	65,67 (9,96)	131,56 (17,28)
<i>exper</i> ²	-0,700 (0,325)	-1,86 (0,54)
<i>idade</i>	-30,51 (4,36)	-54,41 (7,42)
<i>crianmed6</i>	-442,09 (58,85)	-894,02 (111,88)
<i>crianma6</i>	32,78 (23,18)	-16,22 (38,64)
<i>constante</i>	1.330,48 (270,78)	965,31 (446,44)
<i>Valor de Log-Verossimilhança</i>	_____	-3.819,09
<i>R-Quadrado</i>	0,266	0,274
$\hat{\sigma}$	750,18	1.122,02

Para as mulheres que trabalharam horas positivas, a faixa é bastante ampla, de 12 a 4.950. Assim, horas anuais trabalhadas é uma boa candidata a modelo Tobit. Também estimamos um modelo linear (usando todas as 753 observações) por MQO. Os resultados estão na Tabela 17.2

Essa tabela tem várias características às quais devemos atentar. Primeiro, as estimativas dos coeficientes Tobit têm o mesmo sinal das correspondentes estimativas MQO, e a significância estatística das estimativas são semelhantes. (Possíveis exceções são os coeficientes de *nesprend* e *crianma6*, mas as estatísticas t têm magnitudes semelhantes.) Segundo, embora seja tentador comparar as magnitudes das estimativas

EXEMPLO 17.2 (continuação)

MQO e Tobit, isso não será muito informativo. Devemos ter cuidado para não pensar que, como o coeficiente Tobit de *crianma6* é aproximadamente o dobro do coeficiente MQO, o modelo Tobit indica uma resposta muito maior de horas trabalhadas em relação a crianças pequenas.

Podemos multiplicar as estimativas Tobit pelos fatores de ajuste em (17.23) e (17.27) para obter os efeitos parciais das variáveis explicativas sobre os valores esperados condicionais (17.21) e (17.22), respectivamente. O uso do fator em (17.27) conduz a efeitos parciais comparáveis com as estimativas MQO com base na totalidade da amostra (753 observações). Avaliado com os valores amostrais médios de \mathbf{x}_i , o fator em (17.27) será aproximadamente 0,645. Portanto, iniciando-se com os valores médios de todas as variáveis explicativas (inclusive *educ*), estima-se que um ano a mais de educação aumente horas esperadas em cerca de $(0,645)(80,65) \approx 52,0$ horas. Isso é bastante maior que a estimativa MQO, em torno de 28,8. Utilizando a aproximação de mais uma criança, produz uma queda de cerca de $(0,645)(894,02) \approx 576,6$ nas horas esperadas. Obviamente, essa estimativa não tem sentido para uma mulher que já trabalha menos de 576,6 horas. Seria melhor estimar os valores esperados em dois valores diferentes de *crianmed6* (provavelmente um e zero), com as outras variáveis explicativas definidas em seus valores médios, e construir a diferença.

O fator em (17.23) está em torno de 0,451, o que significa que os coeficientes Tobit podem ser multiplicados por 0,451 para obtermos os efeitos parciais aproximados da subpopulação de mulheres casadas com *horas* > 0. Esses coeficientes Tobit escalonados podem ser comparados com os coeficientes MQO obtidos usando somente a amostra das 428 mulheres com *horas* > 0. (Os resultados dessa regressão não são mostrados.) Como exercício, você deve verificar que os efeitos parciais Tobit também são maiores nesse caso.

Descrevemos um *R*-quadrado tanto para o modelo de regressão linear como para o modelo Tobit. O *R*-quadrado do MQO é o habitual. Para o Tobit, o *R*-quadrado é o quadrado do coeficiente de correlação entre y_i e \hat{y}_i , onde $\hat{y}_i = \Phi(\mathbf{x}_i \hat{\beta} / \hat{\sigma}) \mathbf{x}_i \hat{\beta} + \hat{\sigma} \phi(\mathbf{x}_i \hat{\beta} / \hat{\sigma})$ é a estimativa de $E(y_i | \mathbf{x}_i = \mathbf{x}_i)$. Isso é motivado pelo fato de o *R*-quadrado habitual do MQO ser igual à correlação elevada ao quadrado entre y_i e os valores estimados [veja a equação (3.29)]. Em modelos não-lineares como o modelo Tobit, o coeficiente de correlação elevado ao quadrado não é idêntico a um *R*-quadrado baseado na soma dos resíduos quadrados como em (3.28). Isso ocorre porque os valores estimados, como definidos anteriormente, e os resíduos $y_i - \hat{y}_i$ não são não-correlacionados na amostra. Um *R*-quadrado definido como o coeficiente de correlação elevado ao quadrado entre y_i e \hat{y}_i tem a vantagem de sempre estar entre zero e um; um *R*-quadrado baseado na soma dos resíduos quadrados não precisa ter essa característica.

Podemos ver que, com base nas medidas do *R*-quadrado, a função Tobit da média condicional ajusta um pouco os dados sobre horas, mas não de forma substancial. Porém, devemos nos lembrar que as estimativas Tobit não são escolhidas para maximizar um *R*-quadrado — elas maximizam a função log-verossimilhança —, enquanto as estimativas MQO são os valores que efetivamente produzem o mais alto *R*-quadrado, dada a forma funcional linear.

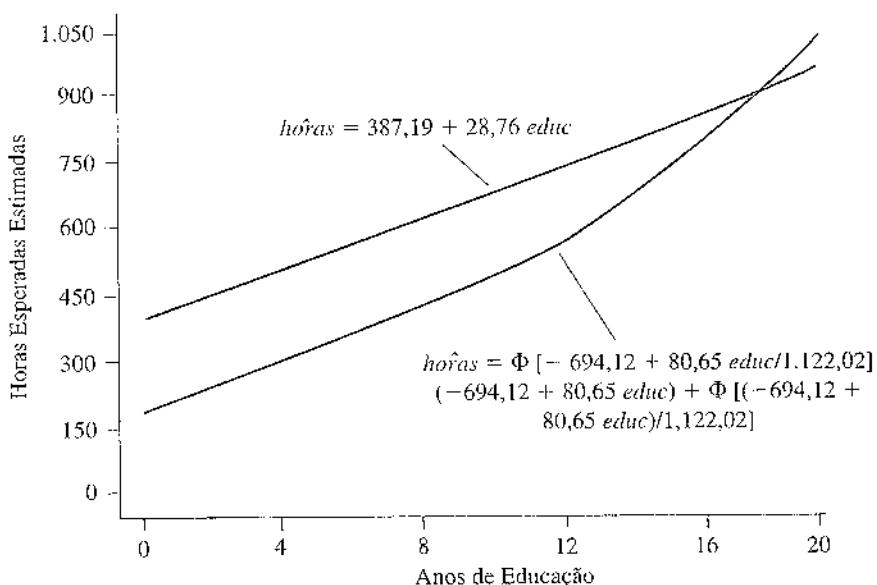
Por construção, todos os valores estimados Tobit de *horas* são positivos. Em contraposição, 39 dos valores estimados MQO são negativos. Embora previsões negativas gerem alguma preocupação, 39 em 753 é apenas um pouco mais de cinco por cento das observações. Não fica totalmente claro como valores estimados negativos do MQO se traduzem em diferenças nos efeitos parciais estimados. A Figura 17.3 traça estimativas de $E(\text{horas} | \mathbf{x})$ como uma função da educação; no modelo Tobit, as outras variáveis explicativas são definidas em seus valores médios. No modelo linear, a equação traçada é $\text{horas} = 387,19 + 28,76 \text{ educ}$. No modelo Tobit, a equação traçada é $\text{horas} = \Phi(-694,12 + 80,65 \text{ educ})/1.122,02 \cdot (-694,12 + 80,65 \text{ educ}) + 1.122,02 \cdot \phi[(-694,12 + 80,65 \text{ educ})/1.122,02]$. Como pode ser visto na figura, o modelo linear produz estimativas notavelmente mais altas das horas trabalhadas esperadas mesmo com altos níveis de educação. Por exemplo, com oito anos de educação, o valor previsto de horas pelo MQO é de aproximadamente 617,5, enquanto a estimativa Tobit está em torno de 423,9. Com 12 anos de educação,

EXEMPLO 17.2 (continuação)

o valor previsto de *horas* é de, aproximadamente, 732,7 e 598,3, respectivamente. As duas linhas de previsão se cruzam após 17 anos de educação, mas nenhuma mulher na amostra tem mais de 17 anos de educação. A inclinação crescente da linha Tobit indica claramente o efeito marginal crescente da educação sobre as horas trabalhadas esperadas.

Figura 17.3

Valores Esperados Estimados de Horas em Relação à Educação para os Modelos Linear e Tobit

**Problemas de Especificação nos Modelos Tobit**

O modelo Tobit, e em particular as fórmulas dos valores esperados em (17.21) e (17.22), dependem essencialmente da normalidade e da homoscedasticidade no modelo subjacente da variável latente. Quando $E(y|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, sabemos, do Capítulo 5, que a normalidade condicional de y não desempenha nenhum papel na inexistência de viés, na consistência ou na inferência de amostras grandes. A heteroscedasticidade não afeta a inexistência de viés ou a consistência do MQO, embora devamos computar erros-padrão e estatísticas de testes robustos para efetuarmos inferência aproximada. Em um modelo Tobit, se qualquer das hipóteses em (17.15) falhar, será difícil saber o que a EMV Tobit estará estimando. No entanto, para desvios moderados das hipóteses, o modelo Tobit pode produzir boas estimativas dos efeitos parciais sobre as médias condicionais. É possível levar em conta mais hipóteses gerais em (17.15), mas tais modelos são muito mais complicados de se estimar e interpretar.

Uma limitação potencialmente importante do modelo Tobit, pelo menos em certas aplicações, é que o valor esperado condicional em $y > 0$ está muito estreitamente ligado com a probabilidade de que $y > 0$. Isso está claro nas equações (17.23) e (17.26). Em particular, o efeito de x_j sobre $P(y > 0|x)$ é propor-

cional a β_j , como também o é o efeito sobre $E(y|y > 0|x)$, no qual ambas as funções que multiplicam β_j são positivas e dependem de x somente por meio de $x\beta/\sigma$. Isso elimina algumas possibilidades interessantes. Por exemplo, considere a relação entre o valor de cobertura de um seguro de vida e a idade da pessoa. Pessoas jovens podem ser menos propensas a fazer seguro de vida, de modo que a probabilidade de $y > 0$ cresce com a idade (pelo menos até determinado ponto). Condicional a ter um seguro de vida, o valor das apólices pode decrescer com a idade, já que seguros de vida se tornam menos importantes na medida em que as pessoas se aproximam do fim de suas vidas. Essa possibilidade não é considerada no modelo Tobit.

Uma maneira de avaliar informalmente se o modelo Tobit é apropriado é estimar um modelo probit no qual o resultado binário, digamos w , será igual a um se $y > 0$, e $w = 0$ se $y = 0$. Assim, de (17.18), w segue um modelo probit, em que o coeficiente de x_j será $\gamma_j = \beta_j/\sigma$. Isso significa que podemos estimar a razão de β_j com σ pelo probit, para cada j . Se o modelo Tobit for válido, a estimativa probit, $\hat{\gamma}_j$, deve ficar “próxima” de $\hat{\beta}_j/\hat{\sigma}$, onde $\hat{\beta}_j$ e $\hat{\sigma}$ são as estimativas Tobit. Elas nunca serão idênticas devido ao erro de amostragem. Mas podemos procurar por certos sinais problemáticos. Por exemplo, se $\hat{\gamma}_j$ for significante e negativo, mas $\hat{\beta}_j$ for positivo, o modelo Tobit poderá não ser apropriado. Ou, se $\hat{\gamma}_j$ e $\hat{\beta}_j$ tiverem o mesmo sinal, mas $|\hat{\beta}_j/\hat{\sigma}|$ for muito maior, ou menor, que $|\hat{\gamma}_j|$, isso também pode indicar problemas. Não devemos nos preocupar muito com as mudanças de sinais ou diferenças em magnitudes nas variáveis explicativas que sejam não significantes em ambos os modelos.

No exemplo de horas anuais trabalhadas, $\hat{\sigma} = 1.122,02$. Quando dividimos o coeficiente Tobit de *nesprend* por $\hat{\sigma}$, obtemos $-8,81/1.122,02 \approx -0,0079$; o coeficiente probit de *nesprend* está em torno de $-0,012$, que é diferente, mas não de forma substancial. Para *criammed*, a estimativa do coeficiente de $\hat{\sigma}$ está em torno de $-0,797$, comparada com a estimativa probit de $-0,868$. Novamente, essa não é uma diferença muito grande, mas indica que o fato de ter filhos pequenos tem um efeito maior sobre a decisão inicial de participar da força de trabalho do que quantas horas uma mulher decide trabalhar, uma vez que ela faça parte da força de trabalho. (O Tobit efetivamente leva em conta a média desses dois efeitos simultaneamente.) Não sabemos se os efeitos são estatisticamente diferentes, mas eles são da mesma ordem de magnitude.

O que acontecerá se concluirmos que o modelo Tobit não é apropriado? Existem modelos, normalmente chamados de modelos de *saltos* ou de *duas partes*, que podem ser usados quando o Tobit parece ser inadequado. Todos eles têm a propriedade de que $P(y > 0|x)$ e $E(y|y > 0|x)$ dependem de parâmetros diferentes, de modo que x_j pode ter efeitos diferentes sobre essas duas funções. [Veja Wooldridge (2002, Capítulo 16) para uma descrição desses modelos.]

17.3 O MODELO DE REGRESSÃO DE POISSON

Outro tipo de variável dependente não-negativa é uma **variável de contagem**, que pode assumir valores inteiros não-negativos: $\{0, 1, 2, \dots\}$. Estamos especialmente interessados em casos nos quais y assume um número relativamente pequeno de valores, inclusive zero. Os exemplos incluem número de filhos de uma mulher, o número de vezes em que alguém foi preso durante o ano, ou o número de patentes solicitadas por uma firma durante um ano. Pelas mesmas razões discutidas quanto às respostas binárias e Tobit, um modelo linear para $E(y|x_1, \dots, x_k)$ pode não fornecer o melhor ajuste para todas as variáveis explicativas. (Mesmo assim, sempre é interessante iniciar com um modelo linear, como fizemos no Exemplo 3.5.)

Assim como um resultado Tobit, não podemos tomar o logaritmo de uma variável de contagem porque ela assume o valor zero. Um método eficaz é modelar o valor esperado como uma função exponencial:

$$E(y|x_1, x_2, \dots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k). \quad (17.28)$$

Como $\exp(\cdot)$ é sempre positivo, (17.28) garante que os valores previstos de y também serão positivos. A função exponencial está traçada na Figura A.5 do Apêndice A, disponível na página do livro, no site www.thomsonlearning.com.br.

Embora (17.28) seja mais complicada que um modelo linear, basicamente já sabemos como interpretar os coeficientes. Tomando o log da equação (17.28):

$$\log[E(y|x_1, x_2, \dots, x_k)] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (17.29)$$

de forma que o log do valor esperado é linear. Portanto, usando as propriedades de aproximação da função log que temos usado nos capítulos anteriores,

$$\% \Delta E(y|x) \approx (100\beta_j)\Delta x_j.$$

Em outras palavras, $100\beta_j$ é aproximadamente a porcentagem de mudança em $E(y|x)$, dado um aumento de uma unidade em x_j . Algumas vezes, necessitamos de uma estimativa mais precisa, e podemos encontrar uma, facilmente, verificando as mudanças discretas no valor esperado. Mantenha fixas todas as variáveis explicativas, exceto x_k , e defina x_k^0 como o valor inicial e x_k^1 como o valor subsequente. Então, a mudança proporcional no valor esperado será

$$[\exp(\beta_0 + x_{k-1}\beta_{k-1} + \beta_k x_k^1)/\exp(\beta_0 + x_{k-1}\beta_{k-1} + \beta_k x_k^0)] - 1 = \exp(\beta_k \Delta x_k) - 1,$$

onde $x_{k-1}\beta_{k-1}$ é a forma abreviada de $\beta_1 x_1 + \dots + \beta_{k-1} x_{k-1}$, e $\Delta x_k = x_k^1 - x_k^0$. Quando $\Delta x_k = 1$ — por exemplo, se x_k for uma variável *dummy* que alteramos de zero para um —, então, a mudança será $\exp(\beta_k) - 1$. Dado $\hat{\beta}_k$, obtemos $\exp(\hat{\beta}_k) - 1$ e o multiplicamos por 100 para transformar a mudança proporcional em uma mudança percentual.

Raciocinando de maneira semelhante ao modelo linear, se β_j multiplica $\log(x_j)$, então, β_j é uma elasticidade. O ponto principal é que, para propósitos práticos, podemos interpretar os coeficientes na equação (17.28) como se tivéssemos um modelo linear, com $\log(y)$ como a variável dependente. Existem algumas diferenças sutis que não precisamos estudar aqui.

Como (17.28) é não-linear em seus parâmetros — lembre-se, $\exp(\cdot)$ é uma função não-linear —, não podemos usar métodos de regressão linear. Poderíamos usar *mínimos quadrados não-lineares*, os quais, como acontece com o MQO, minimizam a soma dos resíduos quadrados. Acontece, porém, que todas as distribuições de dados de contagem padrão exibem heteroscedasticidade, e os mínimos quadrados não-lineares não exploram isso [veja Wooldridge (2002, Capítulo 12)]. Em vez disso, vamos nos valer da máxima verossimilhança e o importante método relacionado da *estimação de quase-máxima verossimilhança*.

No Capítulo 4, introduzimos a normalidade como a hipótese de distribuição padrão da regressão linear. A hipótese de normalidade é razoável para (em linhas gerais) variáveis dependentes contínuas que podem assumir um grande intervalo de valores. Uma variável de contagem não pode ter uma distribuição normal (pois a distribuição normal é de variáveis contínuas que podem assumir todos os valores), e se ela assumir muito poucos valores, a distribuição pode ser muito diferente da normal. Em vez desta, a distribuição normal de dados de contagem é a **distribuição de Poisson**.

Como estamos interessados no efeito das variáveis explicativas sobre y , devemos olhar a distribuição de Poisson condicional em x . A distribuição de Poisson é inteiramente determinada por sua média,

de modo que só precisamos especificar $E(y|x)$. Assumimos que ela tem a mesma forma de (17.28), que escrevemos de forma abreviada como $\exp(x\beta)$. Assim, a probabilidade de que y seja igual ao valor h , condicional em x , será

$$P(y = h|x) = \exp[-\exp(x\beta)][\exp(x\beta)]^h/h!, \quad h = 0, 1, \dots,$$

onde $h!$ significa o fatorial de h (veja o Apêndice B no site da Thomson). Essa distribuição, que é a base do **modelo de regressão de Poisson**, permite encontrar as probabilidades condicionais de qualquer valor das variáveis explicativas. Por exemplo, $P(y = 0|x) = \exp[-\exp(x\beta)]$. Logo que tenhamos as estimativas dos β_j , poderemos inseri-las nas probabilidades de vários valores de x .

Dada uma amostra aleatória $\{(x_i, y_i); i = 1, 2, \dots, n\}$, podemos construir a função log-verossimilhança:

$$\mathcal{L}(\beta) = \sum_{i=1}^n \ell_i(\beta) = \sum_{i=1}^n \{y_i x_i \beta - \exp(x_i \beta)\}, \quad (17.30)$$

onde eliminamos o termo $-\log(y_i!)$ porque ele não depende de β . Essa função log-verossimilhança é simples de maximizar, embora as EMVs de Poisson não sejam obtidas em forma fechada.

Os erros-padrão das estimativas de Poisson $\hat{\beta}_j$ são fáceis de ser obtidos depois de a função log-verossimilhança ter sido maximizada; a fórmula se encontra no apêndice deste capítulo. Elas são descritas com os $\hat{\beta}_j$ por qualquer programa econômétrico.

Assim como nos modelos probit, logit e Tobit, não podemos comparar diretamente as magnitudes das estimativas de Poisson de uma função exponencial com as estimativas MQO de uma função linear. No entanto, é possível fazer uma comparação aproximada, pelo menos para as variáveis explicativas contínuas. Se (17.28) for válida, o efeito parcial de x_j em relação a $E(y|x_1, x_2, \dots, x_k)$ será $\partial E(y|x_1, x_2, \dots, x_k)/\partial x_j = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \cdot \beta_j$. Essa expressão decorre da regra de cadeia do cálculo porque a derivada da função exponencial é a própria função exponencial. Se \hat{y}_i for a estimativa MQO de uma regressão linear de y sobre x_1, \dots, x_k , então, poderemos fazer uma comparação aproximada das magnitudes de \hat{y}_i e de $\exp(\hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \dots + \hat{\beta}_k \bar{x}_k) \cdot \beta_j$, onde \bar{x}_j são as médias amostrais das variáveis explicativas.

Embora a análise da EMV de Poisson seja o primeiro passo natural para dados de contagem, frequentemente ela é restritiva demais. Todas as probabilidades e os momentos de ordem mais alta da distribuição de Poisson são inteiramente determinados pela média. Em particular, a variância é igual à média:

$$\text{Var}(y|x) = E(y|x). \quad (17.31)$$

Isso é restritivo e já foi mostrado que é violado em muitas aplicações. Felizmente, a distribuição de Poisson tem uma propriedade de robustez bastante precisa: independentemente de a distribuição de Poisson ser válida, ainda assim obtemos estimadores dos β_j consistentes e assintoticamente normais. (Isso é análogo ao estimador MQO, que é consistente e assintoticamente normal, independentemente de a hipótese de normalidade ser válida; contudo, o MQO é o EMV sob normalidade.) [Veja Wooldridge (2002, Capítulo 19), para detalhes.]

Quando usamos a EMV de Poisson, mas não assumimos que a distribuição de Poisson seja inteiramente correta, chamamos a análise de **estimação de quase-máxima verossimilhança (EQMV)**. A EQMV de Poisson é bastante prática, pois está programada em vários programas econôméticos.

Porém, a menos que a hipótese de variância de Poisson (17.3) se mantenha, os erros-padrão terão que ser ajustados.

Um ajuste simples dos erros-padrão está disponível quando assumimos que a variância é proporcional à média:

$$\text{Var}(y|x) = \sigma^2 E(y|x), \quad (17.32)$$

onde $\sigma^2 > 0$ é um parâmetro desconhecido. Quando $\sigma^2 = 1$, obtemos a hipótese de variância de Poisson. Quando $\sigma^2 > 1$, a variância será maior que a média para todos os x ; isso é chamado de **super-dispersão**, porque a variância é maior do que no caso Poisson, e é observado em muitas aplicações de regressões de contagem. O caso $\sigma^2 < 1$, chamado de *subdispersion*, é menos comum, mas é permitido em (17.32).

Sob (17.32), é fácil ajustar os erros-padrão habituais da EMV de Poisson. Seja $\hat{\beta}_j$ a EQMV de Poisson e defina os resíduos como $\hat{u}_i = y_i - \hat{y}_i$, onde $\hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})$ é o valor estimado. Como sempre, o resíduo da observação i é a diferença entre y_i e seu valor estimado. Um estimador consistente de σ^2 é $(n - k - 1)^{-1} \sum_{i=1}^n \hat{u}_i^2 / \hat{y}_i$, onde a divisão por \hat{y}_i é o ajuste adequado da heteroscedasticidade, e $n - k - 1$ representa os *gl*, dadas n observações e $k + 1$ estimativas $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$. Definindo $\hat{\sigma}$ como a raiz quadrada de $\hat{\sigma}^2$, multiplicamos os erros-padrão habituais de Poisson por $\hat{\sigma}$. Se $\hat{\sigma}$ for notavelmente maior que 1, os erros-padrão corrigidos podem ser muito maiores que os erros-padrão nominais da EMV de Poisson, geralmente incorretos.

Mesmo (17.32) não é inteiramente geral. Como no modelo linear, podemos também obter erros-padrão da EQMV de Poisson que não restrinjam a variância. [Veja Wooldridge (2002, Capítulo 19), para explicações adicionais.]

Suponha que obtemos $\hat{\sigma}^2 = 2$. Como os erros-padrão ajustados podem ser comparados com os erros-padrão habituais da EMV de Poisson? Como a estatística quase-RV pode ser comparada com a estatística RV habitual?

Sob a hipótese de distribuição de Poisson, podemos usar a estatística razão de verossimilhança para testar restrições de exclusão, as quais, como sempre, têm a forma de (17.12). Se tivermos q restrições de exclusão, a estatística será aproximadamente distribuída como χ_q^2 sob a hipótese nula. Sob a hipótese menos restritiva (17.32), há um ajuste simples (e então chamamos a estatística de **estatística quase-razão de verossimilhança**): dividimos (17.12) por $\hat{\sigma}^2$, onde $\hat{\sigma}^2$ é obtida do modelo sem restrições.

EXEMPLO 17.3

{Regressão de Poisson do Número de Prisões}

Agora aplicamos o modelo de regressão de Poisson aos dados de prisões usados, entre outros locais, no Exemplo 9.1. A variável dependente, *npre86*, é o número de vezes que um homem foi preso em 1986. Essa variável é zero para 1.970 de 2.725 homens na amostra, e somente oito valores de *npre86* são maiores que

EXEMPLO 17.3 (continuação)

cinco. Assim, um modelo de regressão de Poisson é mais apropriado que um modelo de regressão linear. A Tabela 17.3 também apresenta os resultados da estimação por MQO de um modelo de regressão linear.

Tabela 17.3

Determinantes do Número de Prisões de Homens Jovens

Variável Dependente: <i>npre86</i>		
Variáveis Independentes	Linear (MQO)	Exponencial (EQMV de Poisson)
<i>pcond</i>	-0,132 (0,040)	-0,402 (0,085)
<i>sentmed</i>	-0,011 (0,012)	-0,024 (0,020)
<i>temptot</i>	0,012 (0,009)	0,024 (0,015)
<i>ptemp86</i>	-0,041 (0,009)	-0,099 (0,021)
<i>empr86</i>	-0,051 (0,014)	-0,038 (0,029)
<i>rend86</i>	-0,0015 (0,0003)	-0,0081 (0,0010)
<i>negro</i>	0,327 (0,045)	0,661 (0,074)
<i>hispan</i>	0,194 (0,040)	0,500 (0,074)
<i>nasc60</i>	-0,022 (0,033)	-0,051 (0,064)
<i>constante</i>	0,577 (0,038)	-0,600 (0,067)
<i>Valor de Log-Verossimilhança</i>	—	-2.248,76
<i>R-Quadrado</i>	0,073	0,077
$\hat{\sigma}$	0,829	1,232

Os erros-padrão do MQO são os habituais; com certeza, poderíamos tê-los tornado robustos quanto à heteroscedasticidade. Os erros-padrão da regressão de Poisson são os erros-padrão usuais de máxima verossimilhança. Como $\hat{\sigma} = 1,232$, os erros-padrão da regressão de Poisson devem ser corrigidos por esse fator (de forma que cada erro-padrão retificado seja aproximadamente 23% maior.) Por exemplo, um erro-padrão

EXEMPLO 17.3 (continuação)

mais confiável de temptot seria $1,23(0,015) \approx 0,0185$, o que produz uma estatística t de aproximadamente 1,3. O ajustamento dos erros-padrão reduz a significância de todas as variáveis, mas várias delas ainda serão estatisticamente bastante significantes.

Os coeficientes do MQO e de Poisson não são diretamente comparáveis e possuem significados bastante diferentes. Por exemplo, o coeficiente de $pcond$ indica que, se $\Delta pcond = 0,10$, o número esperado de prisões cai em 0,013 ($pcond$ é a proporção de prisões anteriores que levaram a uma condenação). O coeficiente de Poisson indica que $\Delta pcond = 0,10$ reduz as prisões esperadas em cerca de 4% [$0,402(0,10) = 0,0402$, e multiplicamos esse resultado por 100 para obtermos a porcentagem do efeito]. Como regra, isso sugere que podemos reduzir o número total de prisões em cerca de 4% se pudermos aumentar a probabilidade de condenação em 0,1.

O coeficiente de Poisson de *negro* indica que, outros fatores mantidos iguais, o número esperado de prisões de homens negros é cerca de 66% mais alto que o de homens brancos. O coeficiente é altamente significante em termos estatísticos, como também o é o coeficiente de *hispan*.

Como na aplicação Tobit na Tabela 17.2, apresentamos um R -quadrado da regressão de Poisson. Esse é o quadrado do coeficiente de correlação entre y e $\hat{y}_i = \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_k x_{ik})$. A motivação para essa medida de grau de ajuste é a mesma do modelo Tobit. Observamos que o modelo de regressão exponencial, estimado pela EQMV de Poisson, se ajusta um pouco melhor. Lembre-se de que as estimativas MQO são escolhidas para maximizar o R -quadrado, mas as estimativas de Poisson não têm essa finalidade. (Elas são escolhidas para maximizar a função log-verossimilhança.)

Outros modelos de regressão de dados de contagem têm sido propostos e usados em aplicações, que generalizam a distribuição de Poisson de várias maneiras. Se estivermos interessados nos efeitos dos x_j sobre a resposta média, há poucas razões para irmos além da regressão de Poisson: ela é simples, freqüentemente produz bons resultados, e tem a propriedade de robustez discutida anteriormente. Na verdade, podemos aplicar a regressão de Poisson a uma variável y que é um resultado do tipo Tobit, desde que (17.28) se mantenha. Isso pode produzir boas estimativas dos efeitos médios. Extensões da regressão de Poisson são mais úteis quando estamos interessados em estimar probabilidades, tal como em $P(y > 1|x)$. [Veja, por exemplo, Cameron e Trivedi (1998).]

17.4 MODELOS DE REGRESSÃO CENSURADA E TRUNCADA

Os modelos nas Seções 17.1, 17.2 e 17.3 aplicam-se a vários tipos de variáveis dependentes limitadas que freqüentemente surgem em trabalhos aplicados de econometria. Ao usarmos esses métodos, é importante lembrar que a razão de usarmos um modelo probit ou Tobit para uma resposta binária, um modelo Tobit para um resultado de solução de canto, ou um modelo de regressão de Poisson para uma resposta de contagem é porque queremos modelos que avaliem características importantes da distribuição de y . Não existe nenhum problema de observação dos dados. Por exemplo, na aplicação Tobit da participação das mulheres na força de trabalho, no Exemplo 17.2, não existe nenhum problema para se observar horas trabalhadas: é simplesmente o fato de que uma fração não desprezível das mulheres casadas na população escolhem não trabalhar em troca de uma remuneração. Na aplicação da regressão de Poisson nas prisões anuais, observamos a variável dependente para cada jovem na população, mas a variável dependente pode ter valor zero como também outros valores inteiros pequenos. Infelizmente, a distinção entre descontinuidade em uma variável de resposta e dados censurados que envolvam variáveis de respostas pode causar confusão, especialmente na aplicação do modelo Tobit.

Como veremos nesta seção, um modelo com uma estrutura estatística muito semelhante à do modelo Tobit pode ser aplicado para trabalhar com variáveis dependentes que tenham sido censuradas acima ou abaixo de algum valor limite. Em geral, a censura é devida a projeto de pesquisa ou a restrições institucionais. Essencialmente, o problema solucionado por modelos de regressão censurada é um problema de falta de dados, mas é um problema em que temos alguma informação sobre a variável a que os dados se referem, isto é, sabemos se o resultado da variável está acima ou abaixo de um valor limite conhecido.

Um **modelo de regressão truncada** surge quando excluímos, na base de y , um subconjunto da população em nosso esquema de amostragem. Em outras palavras, não temos uma amostra aleatória da população subjacente, mas conhecemos a regra que foi usada para incluir unidades na amostra. Essa regra é determinada pelo critério de y estar acima ou abaixo de certo valor limite. Mais tarde, explicaremos, de maneira mais completa a diferença entre modelos de regressão censurada e truncada.

Modelos de Regressão Censurada

Embora os modelos de regressão censurada possam ser definidos sem hipóteses sobre distribuições, nesta subseção estudaremos o *modelo de regressão normal censurada*. A variável que gostaríamos de explicar, y , segue o modelo linear clássico. Para enfatizar, colocamos um subscrito i em uma extração aleatória da população:

$$y_i = \beta_0 + x_i\beta + u_i, u_i | x_i, c_i \sim \text{Normal}(0, \sigma^2) \quad (17.33)$$

$$w_i = \min(y_i, c_i). \quad (17.34)$$

Em vez de observar y_i , somente a observaremos se ela for inferior a um valor de censura, c_i . Observe que (17.33) inclui a hipótese de que u_i é independente de c_i . (Concretamente, consideramos explicitamente a censura acima, ou *censura à direita*; o problema de fazer a censura abaixo, ou *censura à esquerda*, é tratado de forma semelhante.)

Defina vpm_i como o valor do produto marginal do trabalhador i ; esse é o preço do produto de uma firma multiplicado pelo produto marginal do trabalhador. Assuma que vpm_i é uma função linear de variáveis exógenas, tais como educação, experiência e assim por diante, e um erro não observável. Sob concorrência perfeita e sem restrições institucionais, cada trabalhador recebe seu valor do produto marginal. Defina $salmin_i$ como o salário mínimo do trabalhador i , que varia por estado. Observamos $salário_{it}$, que é o maior dos vpm_{it} e $salmin_i$. Escreva o modelo apropriado para o salário observado.

Um exemplo da censura à direita de dados é a **codificação superior**. Quando uma variável tem codificação superior, conhecemos seu valor somente até certo limite. Para respostas maiores que o limite, somente sabemos que a variável é pelo menos tão grande quanto o limite. Por exemplo, em algumas pesquisas, a riqueza familiar tem codificação superior. Suponha que os entrevistados sejam perguntados sobre sua riqueza, mas as pessoas poderão responder “mais de 500.000 dólares”. Assim, observamos a verdadeira riqueza dos entrevistados cujo valor dos bens seja inferior a 500.000 dólares, mas não daqueles cujo valor dos bens seja superior a 500.000 dólares. Nesse caso, o limite da censura, c_i ,

é o mesmo para todo i . Em muitas situações, o limite da censura muda com as características individuais ou familiares.

Se observarmos uma amostra aleatória de (x, y) , simplesmente estimaremos β por MQO, e a inferência estatística será padrão. (Novamente absorvemos o intercepto em x para simplificar.) A censura causa problemas. Usando argumentos semelhantes aos de um modelo Tobit, uma regressão MQO que use somente observações não-censuradas — isto é, aquelas com $y_i < c_i$ — produz estimadores inconsistentes dos β_j . Uma regressão MQO de w_i sobre x_i , usando todas as observações, não estima consistentemente os β_j , a menos que não haja censura. Isso é semelhante ao caso Tobit, mas o problema é muito diferente. No modelo Tobit, estamos modelando comportamento econômico, que muitas vezes produz resultados iguais a zero; supõe-se que o modelo Tobit reflete isso. Com a regressão censurada, temos um problema de coleção de dados porque, por alguma razão, os dados são censurados.

Sob as hipóteses em (17.33) e (17.34), podemos estimar β (e σ^2) por máxima verossimilhança, dada uma amostra aleatória de (x_i, w_i) . Para isso, precisamos da densidade de w_i , dado (x_i, c_i) . Para observações não censuradas, $w_i = y_i$, e a densidade de w_i será a mesma da y_i : $\text{Normal}(x_i\beta, \sigma^2)$. Para observações censuradas, precisamos da probabilidade de que w_i seja igual ao valor de censura, c_i , dado x_i :

$$P(w_i = c_i | x_i) = P(y_i \geq c_i | x_i) = P(u_i \geq c_i - x_i\beta) = 1 - \Phi[(c_i - x_i\beta)/\sigma].$$

Podemos combinar essas duas partes para obter a densidade de w_i , dados x_i e c_i :

$$f(w | x_i, c_i) = 1 - \Phi[(c_i - x_i\beta)/\sigma], w = c_i, \quad (17.35)$$

$$= (1/\sigma)\phi[(w - x_i\beta)/\sigma], w < c_i. \quad (17.36)$$

O log-verossimilhança da observação i é obtido tomando o log natural da densidade de cada i . Podemos maximizar a soma deles ao longo de i , com relação a β_j e σ para obtermos as EMVs.

É bom saber que podemos interpretar os β_j do mesmo jeito que em um modelo de regressão linear sob amostragem aleatória. Isso é muito diferente das aplicações Tobit, nas quais os valores esperados de interesse são funções não-lineares dos β_j .

Uma aplicação importante dos modelos de regressão censurada é a **análise de duração**. Uma *duração* é uma variável que registra o tempo antes da ocorrência de certo evento. Por exemplo, podemos querer explicar o número de dias antes de um criminoso solto da prisão ser preso novamente. Para alguns criminosos, isso pode nunca acontecer, ou talvez aconteça após um período tão longo que precisaremos censurar a duração para podermos analisar os dados.

Em aplicações de duração de regressão normal censurada, como também na codificação superior, usamos com freqüência o log natural como a variável dependente, o que significa que também tomamos o log do valor de limite em (17.34). Como temos visto ao longo de todo este texto, o uso da transformação logarítmica da variável dependente faz com que os parâmetros sejam interpretados como mudanças percentuais. Além disso, como acontece com muitas variáveis positivas, o log de uma duração em geral tem uma distribuição mais próxima da normal do que a própria duração.

EXEMPLO 17.4**(Duração de Reincidência)**

O arquivo RECID.RAW contém dados sobre o tempo em meses até que um recluso de uma prisão da Carolina do Norte foi preso após ter sido solto; vamos chamá-la *durat*. Alguns dos presidiários participaram de um programa de trabalho durante o tempo em que estiveram na prisão. Também controlamos diversas variáveis demográficas, bem como medidas de prisões e históricos criminais.

De 1.445 reclusos, 893 não foram presos durante o tempo em que foram vigiados; portanto, essas observações foram censuradas. O tempo censurado diferiu entre os reclusos, variando de 70 a 81 meses.

A Tabela 17.4 mostra os resultados da regressão normal censurada de $\log(durat)$. Cada um dos coeficientes, quando multiplicado por 100, informa a mudança percentual estimada na duração esperada, dado um aumento *ceteris paribus* de uma unidade na variável explicativa correspondente.

Vários dos coeficientes na Tabela 17.4 são interessantes. As variáveis *anteriores* (número de condenações anteriores) e *totpris* (total de meses passados na prisão) têm efeitos negativos sobre o tempo até que ocorra a nova prisão. Isso sugere que essas variáveis medem a tendência da atividade criminal, em vez de representar um efeito dissuasor. Por exemplo, um recluso com uma condenação anterior a mais tem uma duração até a próxima prisão que é quase 14% menor. Um ano de reclusão reduz a duração em cerca de $100 \cdot 12(0,019) = 22,8\%$. Uma constatação surpreendente é que um homem que esteja cumprindo pena por delito grave tem uma duração esperada estimada quase 56% ($\exp(0,444) - 1 \approx 0,56$) mais longa que alguém que esteja cumprindo pena por um crime menos grave.

Os que têm um histórico de abuso de drogas ou álcool têm uma duração esperada substancialmente mais curta até a próxima prisão. (As variáveis *álcool* e *drogas* são binárias.) Estima-se que para homens mais velhos e homens que eram casados quando do encarceramento as durações sejam显著mente mais longas até suas próximas prisões. Os negros têm durações substancialmente mais curtas, da ordem de 42% [$\exp(-0,543) - 1 \approx -0,42$].

A variável de decisão crucial, *protrab*, não tem o efeito desejado. A estimativa por ponto é que, outros fatores permanecendo inalterados, homens que tenham participado do programa de trabalho têm durações de reincidência estimadas cerca de 6,3% mais curtas se comparadas aos que não participaram do programa. O coeficiente tem uma estatística *t* pequena, de modo que provavelmente concluiríamos que o programa de trabalho não tem efeito nenhum. Isso pode ser motivado por um problema de auto-seleção, ou pode ser produto da maneira pela qual os homens são alocados para o programa. Naturalmente, é possível que o programa tenha sido ineficiente.

Nesse exemplo, é essencial explicar a censura, especialmente porque quase 62% das durações são censuradas. Se aplicarmos o MQO puro à totalidade da amostra e tratarmos as durações censuradas como se não fossem censuradas, as estimativas dos coeficientes serão notavelmente diferentes. Na verdade, todas elas se reduzem em direção a zero. Por exemplo, o coeficiente em *anteriores* se tornará $-0,059$ ($ep = 0,009$), e o de *álcool* se tornará $-0,262$ ($ep = 0,060$). Embora as direções dos efeitos sejam as mesmas, a importância dessas variáveis é bastante reduzida. As estimativas da regressão censurada são muito mais confiáveis.

Existem outras maneiras de medir os efeitos de cada uma das variáveis explicativas da Tabela 17.4 sobre a duração, em vez de nos concentrarmos apenas na duração esperada. Uma abordagem sobre a análise de duração moderna está além do escopo deste livro. [Para uma introdução ao assunto, veja Wooldridge (2002, Capítulo 20).]

Se qualquer das hipóteses da regressão normal censurada for violada — particularmente se houver heteroscedasticidade ou não normalidade —, as EMVs geralmente serão inconsistentes. Isso mos-

tra que a censura é potencialmente muito onerosa, já que o MQO usando uma amostra não censurada não exige nem normalidade nem heteroscedasticidade para a consistência.

Tabela 17.4

Estimação da Regressão Censurada de Reincidência Criminal

Variável Dependente: $\log(durat)$	
Variáveis Independentes	
<i>protrab</i>	-0,063 (0,120)
<i>anteriores</i>	-0,137 (0,021)
<i>totpris</i>	-0,019 (0,003)
<i>criminoso</i>	0,444 (0,145)
<i>álcool</i>	-0,635 (0,144)
<i>drogas</i>	-0,298 (0,133)
<i>negro</i>	-0,543 (0,117)
<i>casado</i>	0,341 (0,140)
<i>educ</i>	0,023 (0,025)
<i>idade</i>	0,0039 (0,0006)
<i>constante</i>	4,099 (0,348)
<i>Valor de Log-Verossimilhança</i>	-1.597,06
$\hat{\sigma}$	1,810

Há métodos que não exigem que assumamos uma distribuição, mas eles são mais avançados. [Veja Wooldridge (2002, Capítulo 16).]

Modelos de Regressão Truncada

Um modelo de regressão truncada é semelhante a um modelo de regressão censurada, mas difere em um aspecto importante: em um modelo de regressão truncada, não observamos qualquer informação sobre certo segmento da população. Em geral, isso acontece quando uma pesquisa tem por objetivo um subconjunto particular da população e, talvez devido a considerações sobre custos, ignora inteiramente a outra parte da população.

Por exemplo, Hausman e Wise (1977) utilizaram dados de um experimento de imposto de renda negativo para estudar vários determinantes de receitas. Para ser incluída no estudo, uma família tinha que ter renda inferior a 1,5 vezes a linha de pobreza de 1967, em que a linha de pobreza dependia do tamanho da família.

O modelo de regressão truncada normal começa com um modelo de população básica que satisfaça as hipóteses do modelo linear clássico:

$$y = \beta_0 + x\beta + u, u|x \sim \text{Normal}(0, \sigma^2). \quad (17.37)$$

Lembre-se de que esse é um forte conjunto de hipóteses, pois u deve não só ser independente de x , mas também normalmente distribuído. Vamos nos concentrar nesse modelo, pois é difícil relaxar as hipóteses.

Sob (17.37) sabemos que, dada uma amostra aleatória da população, o MQO é o procedimento mais eficiente de estimação. O problema surge porque não observamos uma amostra aleatória da população: a hipótese RLM.2 é violada. Em particular, uma extração aleatória (x_i, y_i) é observada somente se $y_i \leq c_i$, onde c_i é o valor limite do truncamento que pode depender de variáveis exógenas — particularmente de x_i . (No exemplo de Hausman e Wise, c_i depende do tamanho da família). Isso significa que, se $\{(x_i, y_i); i = 1, \dots, n\}$ for nossa amostra observada, y_i será necessariamente menor ou igual a c_i . Isso difere do modelo de regressão censurada, no qual y_i pode ser maior que c_i ; simplesmente não observamos y_i se $y_i > c_i$. Em um modelo de regressão censurada, observamos x_i para qualquer observação extraída aleatoriamente da população; no modelo truncado, somente observamos x_i se $y_i \leq c_i$.

Para estimar os β_j (com σ), necessitamos da distribuição de y_i , dados $y_i \leq c_i$ e x_i . Isso é escrito da seguinte maneira

$$g(y|x_i, c_i) = \frac{f(y|x_i, \beta, \sigma^2)}{F(c_i|x_i, \beta, \sigma^2)}, \quad y \leq c_i, \quad (17.38)$$

onde $f(y|x_i, \beta, \sigma^2)$ representa a densidade normal com média $\beta_0 + x_i\beta$ e variância σ^2 , e $F(c_i|x_i, \beta, \sigma^2)$ é a fdc normal com as mesmas média e variância, avaliadas em c_i . Essa expressão para a densidade, condicional a $y_i \leq c_i$, tem sentido intuitivo: é a densidade da população, dado x_i , dividida pela probabilidade de que y_i seja menor ou igual a c_i (dado x_i), $P(y_i \leq c_i|x_i)$. Na realidade, normalizamos outra vez a densidade dividindo-a pela área sob $f(\cdot|x_i, \beta, \sigma^2)$ que está à esquerda de c_i .

Se tomarmos o log de (17.38), somarmos ao longo de todos os i , e maximizarmos o resultado em relação a β_j e σ^2 , obteremos os estimadores de máxima verossimilhança. Isso conduz a estimadores consistentes e aproximadamente normais. A inferência, inclusive os erros-padrão e estatísticas de log-verossimilhança, é padrão.

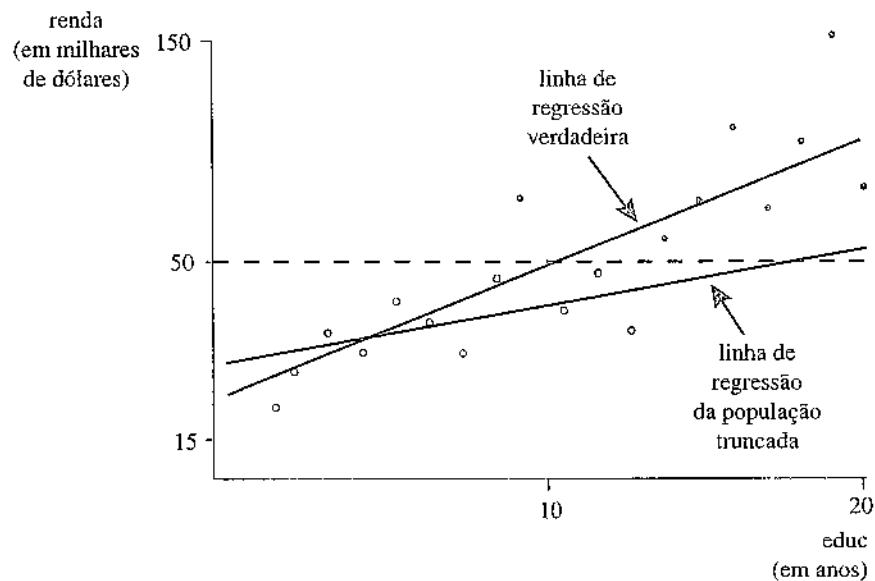
Poderíamos analisar os dados do Exemplo 17.4 como uma amostra truncada se eliminássemos todos os dados de uma observação sempre que ela tivesse sido censurada. Isso nos daria 552 observações de uma distribuição normal truncada, onde o ponto de truncamento difere ao longo das observações i .

Porém, nunca analisaríamos dados de duração (ou de codificação superior) dessa maneira, pois ela elimina informações úteis. O fato de conhecermos um limite inferior de 893 durações, com as variáveis explicativas, é informação útil; a regressão censurada usa essas informações, o que a regressão truncada não faz.

Um exemplo melhor é dado por Hausman e Wise (1977), no qual eles enfatizam que o MQO aplicado a uma amostra truncada geralmente produz estimadores viesados para zero. Intuitivamente, isso faz sentido. Suponha que a relação de interesse seja entre os níveis de renda e educação. Se apenas observarmos pessoas cuja renda esteja abaixo de certo valor, estaremos eliminando a parte superior. Isso tende a nivelar a linha estimada, em relação à verdadeira linha da regressão, na totalidade da população. A Figura 17.4 ilustra o problema quando a renda é truncada acima de 50.000 dólares. Embora observemos os pontos dos dados representados pelos círculos abertos, não observamos os conjuntos de dados representados pelos círculos escurecidos. Uma análise de regressão usando amostra truncada não conduz a estimadores consistentes. A propósito, se a amostra da Figura 17.4 tivesse sido censurada em vez de truncada — isto é, tivéssemos dados com codificação superior —, observaríamos níveis de educação para todos os pontos na Figura 17.4, mas para indivíduos com renda acima de 50.000 dólares não saberíamos o montante exato da renda. Apenas saberíamos que a renda seria de pelo menos 50.000 dólares. Na verdade, todas as observações representadas pelos círculos escurecidos seriam levadas para baixo na linha horizontal de $renda = 50$.

Figura 17.4

Uma linha de regressão verdadeira, ou populacional, e a linha de regressão incorreta da população truncada com renda abaixo de 50.000 dólares



Assim como na regressão censurada, se a hipótese normal homoscedástica subjacente em (17.37) for violada, a EMV normal truncada será viesada e inconsistente. Há métodos que não necessitam dessas hipóteses; veja Wooldridge (2002, Capítulo 17), para discussão e referências.

17.5 CORREÇÕES DA SELEÇÃO AMOSTRAL

A regressão truncada é um caso especial de um problema conhecido como **seleção de amostra não-aleatória**. Entretanto, o projeto da pesquisa não é a única causa da seleção de amostra não-aleatória. Com freqüência, entrevistados não respondem certas perguntas, o que leva a dados ausentes das variáveis dependentes ou independentes. Como não podemos usar essas observações em nossa estimação, devemos imaginar se suas eliminações conduzirão a viés em nossos estimadores.

Outro exemplo genérico é habitualmente chamado de **truncamento ocasional**. Nesse caso, não observamos y em razão do resultado de outra variável. O principal exemplo é estimar a chamada *função de oferta de salário* na área da economia do trabalho. O interesse reside em como vários fatores, como a educação, afetam o salário que um indivíduo poderia ganhar na força de trabalho. Para as pessoas que estão na força de trabalho, observamos a oferta de salário como o salário corrente. Contudo, para aqueles que estejam desempregados, não observamos a oferta de salário. Como trabalhar pode estar sistematicamente correlacionado a fatores não-observáveis que afetam a oferta de salário, usar somente pessoas que estejam trabalhando — como temos feito em todos os exemplos de salários até agora — pode produzir estimadores viesados dos parâmetros na equação de oferta de salário.

A seleção de amostra não-aleatória também pode surgir quando temos dados de painel. No caso mais simples, teremos dois anos de dados, mas, devido às demissões, algumas pessoas saem da amostra. Isso é particularmente um problema na análise de políticas empresariais, nas quais as demissões podem estar relacionadas à eficácia de um programa administrativo.

Quando o MQO é Consistente na Amostra Selecionada?

Na Seção 9.4, apresentamos uma breve explicação dos tipos de seleções amostrais que podem ser ignorados. A distinção crucial é entre seleções amostrais *exógenas* e *endógenas*. No caso Tobit truncado, claramente temos seleção amostral endógena e o MQO é viesado e não consistente. De outro lado, se nossa amostra for determinada somente por uma variável explicativa exógena, teremos seleção amostral exógena. Casos entre esses dois extremos são menos claros, e agora apresentamos cuidadosas definições e hipóteses para eles. O modelo populacional é

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u, E(u|x_1, x_2, \dots, x_k) = 0. \quad (17.39)$$

É útil escrever o modelo populacional de uma extração *aleatória* como

$$y_i = x_i \beta + u_i, \quad (17.40)$$

onde usamos $x_i \beta$ como uma forma abreviada de $\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$. Agora, seja n o tamanho de uma *amostra aleatória* da população. Se pudéssemos observar y_i e cada x_{ij} de todas as observações i , simplesmente usaríamos o MQO. Assuma que, por alguma razão, y_i ou algumas das variáveis independentes não sejam observadas para determinada observação i . Para ao menos algumas observações, observamos o conjunto completo de variáveis. Defina um *indicador de seleção* s_i de cada i como $s_i = 1$ se observarmos todos os (y_i, x_i) , e $s_i = 0$ caso contrário. Assim, $s_i = 1$ indica que usaremos a observação em nossa análise; $s_i = 0$ significa que a observação não será usada. Estamos interessados nas propriedades estatísticas dos estimadores MQO usando a *amostra selecionada*, isto é, usando as observações de cada $s_i = 1$. Portanto, usamos menos que n observações, digamos n_1 .

Ocorre que é fácil obter condições sob as quais o MQO será consistente (e mesmo não-viesado). Efetivamente, em vez de estimar (17.40), podemos somente estimar a equação

$$s_i y_i = s_i x_i \beta + s_i u_i. \quad (17.41)$$

Quando $s_i = 1$, simplesmente teremos (17.40); quando $s_i = 0$, simplesmente teremos $0 = 0 + 0$, o que obviamente não nos diz nada a respeito de β . Fazer a regressão de $s_i y_i$ sobre $s_i x_i$ com $i = 1, 2, \dots, n$ é o mesmo que fazer a regressão de y_i sobre x_i usando as observações para as quais $s_i = 1$. Assim, podemos verificar a consistência de $\hat{\beta}_j$ estudando (17.41) em uma amostra aleatória.

Conforme nossa análise no Capítulo 5, os estimadores MQO de (17.41) serão consistentes se o termo erro tiver média zero e for não-correlacionado com cada variável explicativa. Na população, a hipótese de média zero é $E(su) = 0$, e a hipótese de correlação zero pode ser estabelecida como

$$E[(sx_j)(su)] = E(sx_j u) = 0, \quad (17.42)$$

onde s, x_j e u são variáveis aleatórias representando a população; usamos o fato de $s^2 = s$ porque s é uma variável binária. A condição (17.42) é diferente do que necessitamos se observarmos todas as variáveis de uma amostra aleatória: $E(x_j u) = 0$. Portanto, na população, precisamos que u seja não-correlacionado com sx_j .

A condição mais importante para a inexistência de viés é $E(su|sx_1, \dots, sx_k) = 0$. Como sempre, essa é uma hipótese mais forte do que a necessária para consistência.

Se s for uma função somente das variáveis explicativas, então, sx_j será apenas uma função de x_1, x_2, \dots, x_k ; pela hipótese da média condicional em (17.39), sx_j também será não-correlacionada com u . Na verdade, $E(su|sx_1, \dots, sx_k) = sE(u|sx_1, \dots, sx_k) = 0$, pois $E(u|x_1, \dots, x_k) = 0$. Esse é o caso da **seleção amostral exógena**, na qual $s_i = 1$ é determinado inteiramente por x_{i1}, \dots, x_{ik} . Como um exemplo, ao estimarmos uma equação de salários na qual as variáveis explicativas sejam educação, experiência, permanência, gênero, estado civil, e assim por diante — que são assumidas como exógenas —, poderemos selecionar a amostra com base em qualquer ou todas as variáveis explicativas.

Se a seleção da amostra for inteiramente aleatória no sentido de que s_i é *independente* de (x_i, u_i) , $E(sx_j u) = E(s)E(x_j u) = 0$, pois $E(x_j u) = 0$ sob (17.39). Portanto, se começarmos com uma amostra aleatória e aleatoriamente eliminarmos observações, o MQO ainda será consistente. De fato, o MQO novamente será não-viesado nesse caso, desde que não haja multicolinearidade perfeita na amostra selecionada.

Se s depender das variáveis explicativas e de termos aleatórios adicionais que sejam independentes de x e u , o MQO também será consistente e não-viesado. Por exemplo, suponha que a pontuação do QI seja uma variável explicativa em uma equação de salários, mas que não esteja presente para algumas pessoas. Suponha que pensemos que a seleção possa ser descrita por $s = 1$ se $QI \geq v$, e $s = 0$ se $QI < v$, onde v é uma variável aleatória não observada independente de QI , de u e das outras variáveis explicativas. Isso significa que é mais provável observarmos um QI que seja alto, mas sempre existe alguma probabilidade de não observarmos nenhum QI . Condisional às variáveis explicativas, s será independente de u , o que significa que $E(u|x_1, \dots, x_k, s) = E(u|x_1, \dots, x_k)$, e o último valor esperado será zero por hipótese no modelo populacional. Se adicionarmos a hipótese de homoscedasticidade $E(u^2|x, s) = E(u^2) = \sigma^2$, os habituais erros-padrão e as estatísticas de testes do MQO serão válidos.

Até agora, mostramos várias situações nas quais o MQO na amostra selecionada é não-viesado, ou pelo menos consistente. Quando o MQO na amostra selecionada será inconsistente? Já vimos um

exemplo: a regressão usando uma amostra truncada. Quando o truncamento é acima, $s_i = 1$ se $y_i \leq c_i$, onde c_i é o valor limite do truncamento. De forma equivalente, $s_i = 1$ se $u_i \leq c_i - x_i\beta$. Como s_i depende diretamente de u_i , s_i e u_i não serão não-correlacionados, mesmo condicionais a x_i . Essa é a razão pela qual o MQO na amostra selecionada não estima com consistência os β_j . Existem meios menos óbvios de s e u serem correlacionados; consideraremos isso na próxima subseção.

Os resultados sobre a consistência do MQO se estendem para a estimação de variáveis instrumentais. Se as VIs forem chamadas de z_h na população, a condição crucial para a consistência do MQ2E será $E(sz_h u) = 0$, que será válido se $E(u|z,s) = 0$. Portanto, se a seleção for determinada inteiramente pelas variáveis exógenas z , ou se s depender de outros fatores que sejam independentes de u e de z , então, o MQ2E na amostra selecionada geralmente será consistente. Temos que assumir que as variáveis explicativas e instrumentais são apropriadamente correlacionadas na parte selecionada da população. Wooldridge (2002, Capítulo 17) contém definições precisas dessas hipóteses.

Também pode ser mostrado que, quando a seleção é inteiramente uma função das variáveis exógenas, a estimativa de máxima verossimilhança de um modelo não linear — tal como um modelo logit ou probit — produz estimadores consistentes e assimptoticamente normais, e os habituais erros-padrão e estatísticas de testes são válidos. [Novamente, veja Wooldridge (2002, Capítulo 17)].

Truncamento Ocasional

Como mencionamos anteriormente, uma forma comum de seleção amostral é chamada de truncamento ocasional. Começamos novamente com o modelo populacional em (17.39). Porém, assumimos que sempre observaremos as variáveis explicativas x_j . O problema é que somente observamos y para um subconjunto da população. A regra que determina se observamos y não depende diretamente do resultado de y . Um exemplo importante é quando $y = \log(\text{salário}^\circ)$, onde salário[°] é a *oferta de salário*, ou a remuneração por hora que um indivíduo poderia receber no mercado de trabalho. Se a pessoa estiver trabalhando no momento da pesquisa, observaremos a oferta de salário, porque assumimos que ela é o salário observado. Porém, para as pessoas fora da força de trabalho, não podemos observar o salário[°]. Portanto, o truncamento da oferta salarial é ocasional, pois ele depende de outra variável, ou seja, a participação na força de trabalho. É importante mencionar que geralmente observaremos todas as outras informações sobre um indivíduo, tais como educação, experiência anterior, gênero, estado civil etc.

A abordagem habitual para o truncamento ocasional é adicionar uma equação de seleção explícita ao modelo populacional de interesse:

$$y = x\beta + u, E(u|x) = 0 \quad (17.43)$$

$$s = 1 [z\gamma + v \geq 0], \quad (17.44)$$

onde $s = 1$ se observarmos y , e zero, caso contrário. Assumimos que elementos de x e z são sempre observados, e escrevemos $x\beta = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ e $z\gamma = \gamma_0 + \gamma_1 z_1 + \dots + \gamma_m z_m$.

A equação de maior interesse é (17.43), e é possível estimar β por MQO, dada uma amostra aleatória. A equação de seleção, (17.44), depende das variáveis observadas, z_h , e de um erro não observado, v . Uma hipótese padrão que faremos é que z é exógeno em (17.43):

$$E(u|x,z) = 0$$

Na verdade, para que os métodos seguintes propostos funcionem bem, necessitaremos que x seja um subconjunto estrito de z : qualquer x_i também é um elemento de z , e temos alguns elementos de z que não estão, também, em x . Veremos mais tarde por que isso é essencial.

Assume-se que o termo erro v na equação de seleção amostral é independente de z (e portanto de x). Também assumimos que v tem uma distribuição normal padrão. Podemos facilmente ver que a correlação entre u e v geralmente causa um problema de seleção amostral. Para ver o motivo, assuma que (u, v) seja independente de z . Então, considerando o valor esperado de (17.43), condicional a z e v , e usando o fato de que x é um subconjunto de z produz

$$E(y|z, v) = x\beta + E(u|z, v) = x\beta + E(u|v),$$

onde $E(u|z, v) = E(u|v)$ porque (u, v) é independente de z . Agora, se u e v forem conjuntamente normais (com média zero), $E(u|v) = \rho v$ para algum parâmetro ρ . Portanto,

$$E(y|z, v) = x\beta + \rho v.$$

Não observamos v , mas podemos usar essa equação para computar $E(y|z, s)$ e depois a limitarmos em $s = 1$. Agora temos:

$$E(y|z, s) = x\beta + \rho E(v|z, s).$$

Como s e v são relacionados por (17.43), e v tem uma distribuição normal padrão, podemos mostrar que $E(v|z, s)$ é simplesmente o inverso da razão de Mills, $\lambda(z\gamma)$, quando $s = 1$. Isso leva à importante equação

$$E(y|z, s = 1) = x\beta + \rho\lambda(z\gamma). \quad (17.45)$$

A equação (17.45) mostra que o valor esperado de y , dados z e a observabilidade de y , é igual a $x\beta$, mais um termo adicional que depende do inverso da razão de Mills avaliado em $z\gamma$. Lembre-se de que esperamos estimar β . Essa equação mostra que isso é possível usando somente a amostra selecionada, desde que incluamos o termo $\lambda(z\gamma)$ como um regressor adicional.

Se $\rho = 0$, $\lambda(z\gamma)$ não aparecerá, e o MQO de y sobre x usando a amostra selecionada estima consistemente β . Fora isso, omitimos efetivamente uma variável, $\lambda(z\gamma)$, que geralmente está correlacionada com x . Em que situação $\rho = 0$? A resposta é: quando u e v forem não-correlacionados.

Como γ é desconhecido, não podemos avaliar $\lambda(z\gamma)$ para cada i . Porém, com base nas hipóteses que fizemos, s dado z segue um modelo probit:

$$P(s = 1|z) = \Phi(z\gamma). \quad (17.46)$$

Portanto, podemos estimar γ pelo probit de s_i sobre z_i , usando a amostra inteira. Em uma segunda etapa, poderemos estimar β . Resumimos o procedimento, que recentemente foi apelidado de **método de Heckit** na literatura econométrica depois do trabalho de Heckman (1976).

CORREÇÃO DA SELEÇÃO AMOSTRAL

(i) Usando todas as n observações, estime um modelo probit de s_i sobre \mathbf{z}_i e obtenha as estimativas $\hat{\gamma}_i$. Calcule o inverso da razão de Mills, $\hat{\lambda}_i = \lambda(z_i \hat{y})$ para cada i . (Na realidade, somente necessitaremos desse cálculo para a observação i com $s_i = 1$).

(ii) Usando a amostra selecionada, ou seja, as observações nas quais $s_i = 1$ (digamos, n_1 delas), compute a regressão de

$$y_i \text{ sobre } x_i \hat{\lambda}_i. \quad (17.47)$$

Os $\hat{\beta}_j$ são consistentes e aproximadamente normalmente distribuídos.

Um teste simples do viés de seleção está disponível a partir da regressão (17.47). Em outras palavras, podemos usar a estatística t habitual de $\hat{\lambda}$, como um teste de $H_0: \rho = 0$. Sob H_0 , não há problema de seleção de amostra.

Quando $\rho \neq 0$, os erros-padrão habituais do MQO descritos em (17.47) não serão exatamente corretos. Isso porque eles não explicam a estimativa de γ , que utiliza as mesmas observações na regressão (17.47), e por outros motivos mais. Alguns programas econométricos computam corretamente os erros-padrão. [Infelizmente, isso não é tão simples quanto um ajuste de heteroscedasticidade. Veja Wooldridge (2002, Capítulo 6), para discussões adicionais.] Em muitos casos, os ajustes não levam a diferenças importantes, mas é difícil saber isso antecipadamente (a menos que $\hat{\rho}$ seja pequeno e não significante).

Há pouco mencionamos que \mathbf{x} deveria ser um subconjunto estrito de \mathbf{z} . Isso traz duas implicações. Primeira, qualquer elemento que apareça como uma variável explicativa em (17.43) também deve ser uma variável explicativa na equação de seleção. Embora em raras ocasiões faça sentido excluir elementos da equação de seleção, não custa muito incluir todos os elementos de \mathbf{x} em \mathbf{z} ; a exclusão deles pode levar a inconsistências se eles forem excluídos incorretamente.

Uma segunda implicação importante é que temos pelo menos um elemento de \mathbf{z} que não está também em \mathbf{x} . Isso significa que necessitamos de uma variável que afete a seleção, mas que não tenha um efeito parcial sobre y . Isso não é absolutamente necessário para aplicar o procedimento — de fato, podemos conduzir mecanicamente as duas etapas quando $\mathbf{z} = \mathbf{x}$ —, mas os resultados em geral serão menos que convincentes, a menos que tenhamos uma *restrição de exclusão* em (17.43). A razão para isso é que, embora o inverso da razão de Mills seja uma função não-linear de \mathbf{z} , ele freqüentemente é bem aproximado por uma função linear. Se $\mathbf{z} = \mathbf{x}$, $\hat{\lambda}$, pode ser altamente correlacionado com os elementos de \mathbf{x} . Como sabemos, tal multicolinearidade pode conduzir a erros-padrão muito altos dos $\hat{\beta}_j$. Intuitivamente, se não tivermos uma variável que afeta a seleção, exceto y , será extremamente difícil, se não impossível, distinguir seleção amostral de uma forma funcional mal-especificada em (17.43).

EXEMPLO 17.5

{Equação da Oferta de Salário para Mulheres Casadas}

Aplicamos a correção da seleção amostral aos dados sobre mulheres casadas contidos no arquivo MROZ.RAW. Lembre-se de que das 753 mulheres na amostra, 428 trabalharam por salário durante o ano. A equação da oferta de salário é padrão, com $\log(\text{salário})$ como a variável dependente e educ , exper e exper^2 como as variáveis explicativas. Para testarmos e corrigirmos o viés da seleção amostral — devido à impossibilidade de observar a oferta de salário para as mulheres que não trabalham —, precisamos estimar um modelo probit da participação na força de trabalho. Adicionalmente às variáveis educação e experiência, incluímos os fatores descritos na Tabela 17.1: outra renda, idade, número de filhos pequenos e número de filhos maiores.

EXEMPLO 17.5 (continuação)

velhos. O fato de essas quatro variáveis serem excluídas da equação de oferta de salário é uma *hipótese*: assumimos que, dados os fatores de produtividade, *nesprend*, *idade*, *crianmed6* e *crianma6* não têm efeito sobre a oferta de salário. É evidente, pelos resultados probit na Tabela 17.1, que pelo menos *idade* e *crianmed6* têm um forte efeito sobre a participação na força de trabalho.

A Tabela 17.5 contém os resultados do MQO e de Heckit. [Os erros-padrão dos resultados de Heckit são os mesmos erros-padrão habituais do MQO da regressão (17.47).] Não existe evidência de um problema de seleção amostral na estimativa da equação de oferta de salário. O coeficiente de $\hat{\lambda}$ tem uma estatística t bastante pequena (0,239), e assim não podemos rejeitar $H_0: \rho = 0$. De mesma importância, é o fato de não haver grandes diferenças práticas nos coeficientes de inclinação estimados na Tabela 17.5. Os retornos da educação estimados diferem somente em um décimo de ponto percentual.

Tabela 17.5

Equação da Oferta de Salário para Mulheres Casadas

Variável Dependente: $\log(\text{salário})$		
Variáveis Independentes	MQO	Heckit
<i>educ</i>	0,108 (0,014)	0,109 (0,016)
<i>exper</i>	0,042 (0,012)	0,044 (0,016)
<i>exper</i> ²	-0,00081 (0,00039)	-0,00086 (0,00044)
<i>constante</i>	-0,522 (0,199)	-0,578 (0,307)
$\hat{\lambda}$	—	0,032 (0,134)
<i>Tamanho da Amostra</i>	428	428
<i>R-quadrado</i>	0,157	0,157

Uma alternativa ao método precedente de estimação em duas etapas é a estimação de máxima verossimilhança completa. Ela é mais complicada, já que requer que obtenhamos a distribuição conjunta de y e s . Muitas vezes, faz sentido testar a seleção amostral usando o procedimento anterior; se não houver evidência de seleção amostral, não haverá razão para continuar. Se detectarmos viés de seleção amostral, poderemos ou usar a estimativa em duas etapas ou estimarmos conjuntamente as equações de regressão e seleção por EMV. [Veja Wooldridge (2002, Capítulo 17).]

No Exemplo 17.5, conhecemos mais do que apenas se uma mulher trabalhou durante o ano: sabemos quantas horas cada mulher trabalhou. Acontece que podemos usar essa informação em um procedimento alternativo de seleção amostral. Em lugar do inverso da razão de Mills $\hat{\lambda}_i$, usamos os resíduos

Tobit, digamos, \hat{v}_i , que é calculado como $\hat{v}_i = y_i - x_i \hat{\beta}$ sempre que $y_i > 0$. Pode ser mostrado que a regressão em (17.47) com \hat{v}_i no lugar de $\hat{\lambda}_i$ também produz estimativas consistentes dos β_j , e que a estatística t padrão de \hat{v}_i é um teste válido para o viés de seleção amostral. Essa abordagem tem a vantagem de usar mais informações, mas é menos aplicada. [Veja Wooldridge (2002, Capítulo 17).]

Há mais tópicos envolvendo a questão da seleção amostral. Um digno de ser mencionado é o de modelos com variáveis explicativas endógenas, *em adição ao possível viés de seleção amostral*. Escreva um modelo com uma única variável explicativa endógena como

$$y_1 = \alpha_1 y_2 + z_1 \beta_1 + u_1, \quad (17.48)$$

onde y_1 somente será observado quando $s = 1$, e y_2 poderá ser observado com y_1 . Um exemplo é quando y_1 é a porcentagem de votos recebidos por um candidato, e y_2 é a porcentagem do total de gastos registrado pelo candidato. Para os candidatos que não concorrem, não poderemos observar y_1 ou y_2 . Se tivermos fatores exógenos que afetem a decisão de concorrer e que estejam correlacionados com os gastos de campanha, poderemos estimar consistentemente α_1 e os elementos de β_1 por variáveis instrumentais. Para ser convincente, precisamos de duas variáveis exógenas que não apareçam em (17.48). Efectivamente, uma deve afetar a decisão de seleção, e uma deve ser correlacionada com y_2 [a exigência normal para estimar (17.48) por MQ2E]. Resumidamente, o método é estimar a equação de seleção por probit, no qual *todas* as variáveis exógenas aparecem na equação probit. Depois, adicionamos o inverso da razão de Mills a (17.48) e estimamos a equação por MQ2E. O inverso da razão de Mills age como sua própria instrumental, já que depende somente de variáveis exógenas. Usamos todas as variáveis exógenas como as outras instrumentais. Como antes, podemos usar a estatística t de $\hat{\lambda}_i$ como um teste para o viés de seleção. [Veja Wooldridge (2002, Capítulo 17), para informações adicionais.]

Neste capítulo, estudamos vários métodos avançados que são freqüentemente usados em aplicações, especialmente em microeconomia. Os modelos logit e probit são usados para variáveis de resposta binárias. Esses modelos oferecem algumas vantagens sobre o modelo de probabilidade linear: as probabilidades estimadas estão entre zero e um, e os efeitos parciais decrescem. O principal custo do logit e do probit é que eles são mais difíceis de interpretar.

O modelo Tobit é aplicável a resultados não-negativos que se acumulam em zero, mas que também assumem uma ampla gama de valores positivos. Muitas variáveis de escolha individual, tais como a oferta de mão-de-obra, o valor do seguro de vida e o montante do fundo de pensão investido em ações, possuem essa característica. Assim como no logit e no probit, os valores esperados de y dado x — sejam condicionais a $y > 0$ ou incondicionalmente — dependem de x e de β de maneiras não-linear. Demos as expressões desses valores esperados, como também as fórmulas dos efeitos parciais de cada x_j sobre as expectativas. Elas poderão ser estimadas após o modelo Tobit ter sido estimado por máxima verossimilhança.

Quando a variável dependente é uma variável de contagem — isto é, ela assume valores inteiros não-negativos —, um modelo de regressão de Poisson será apropriado. O valor esperado de y , dados os x_j , tem uma forma exponencial. Isso dá aos parâmetros interpretações como semi-elasticidades ou elasticidades, dependendo se os x_j estão em nível ou na forma logarítmica. Em resumo, podemos interpretar os parâmetros *como se* eles estivessem em um modelo linear com $\log(y)$ como a variável dependente. Os parâmetros podem ser estimados por EMV. Porém, como a distribuição de Poisson impõe

igualdade entre a variância e a média, freqüentemente é necessário computar erros-padrão e estatísticas de testes que admitam superdispersão ou subdispersão. Trata-se de simples ajustes dos habituais erros-padrão e estatísticas da EMV.

Modelos de regressões censurada e truncada resolvem tipos específicos de problemas de ausência de dados. Na regressão censurada, a variável dependente é censurada acima ou abaixo de um valor limite. Podemos usar as informações sobre os resultados censurados porque sempre observamos as variáveis explicativas, como em aplicações de duração ou codificação superior de observações. Um modelo de regressão truncada surge quando uma parte da população é inteiramente excluída: não observamos qualquer informação em unidades que não estejam cobertas pelo esquema de amostragem. Este é um caso especial de problema de seleção amostral.

A Seção 17.5 oferece um tratamento sistemático da seleção amostral não-aleatória. Mostramos que a seleção amostral exógena não afeta a consistência do MQO quando aplicada na subamostra, mas a seleção amostra endógena afeta. Mostramos como testar e corrigir o viés de seleção amostral para o problema geral do truncamento ocasional, no qual observações estão faltando em y devido ao resultado de outra variável (como a participação na força de trabalho). O método de Heckman é relativamente fácil de ser implementado nessas situações.

- 17.1** (i) Para uma resposta binária y , seja \bar{y} a proporção de uns na amostra (que é igual à média amostral de y_i). Sejam \hat{q}_0 a porcentagem corretamente prevista do resultado $y = 0$ e \hat{q}_1 a porcentagem corretamente prevista do resultado $y = 1$. Se $\hat{\rho}$ é a porcentagem global corretamente prevista, mostre que $\hat{\rho}$ é uma média ponderada de \hat{q}_0 e \hat{q}_1 :

$$\hat{\rho} = (1 - \bar{y})\hat{q}_0 + \bar{y}\hat{q}_1.$$

- (ii) Em uma amostra de 300 observações, suponha que $\bar{y} = 0,70$, de forma que existem 210 resultados com $y_i = 1$ e 90 com $y_i = 0$. Suponha que a porcentagem corretamente prevista quando $y = 0$ seja 80, e quando $y = 1$ ela seja 40. Encontre a porcentagem global corretamente prevista.

- 17.2** Defina *grad* como uma variável dummy informando se um estudante-atleta de uma grande universidade se formará em cinco anos. Sejam *nmem* e *sat* a nota média do ensino médio e a nota do exame SAT, respectivamente. Defina *estudo* como o número de horas gastas por semana em uma sala de estudo organizada. Suponha que, usando os dados de 420 estudantes-atletas, obtenha-se o seguinte modelo logit:

$$\hat{P}(\text{grad} = 1 | \text{nmem}, \text{sat}, \text{estudo}) = \Lambda(-1,17 + 0,24 \text{nmem} + 0,00058 \text{sat} + 0,073 \text{estudo}),$$

onde $\Lambda(z) = \exp(z)/[1 + \exp(z)]$ é a função logit. Mantendo fixos *nmem* em 3,0 e *sat* em 1.200, compute a diferença estimada na probabilidade de formatura de alguém que passou dez horas por semana em uma sala de estudo e de alguém que passou cinco horas por semana.

- 17.3** (Exige cálculo.) (i) Suponha no modelo Tobit que $x_1 = \log(z_1)$, e que esse é o único lugar em que z_1 aparece em x . Mostre que

$$\frac{\partial E(y|y > 0, x)}{\partial z_1} = (\beta_1/z_1)\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}, \quad (17.49)$$

onde β_1 é o coeficiente de $\log(z_1)$.

- (ii) Se $x_1 = z_1$, e $x_2 = z_1^2$, mostre que

$$\frac{\partial E(y|y > 0, x)}{\partial z_1} = (\beta_1 + 2\beta_2 z_1)\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\},$$

onde β_1 é o coeficiente de z_1 e β_2 é o coeficiente de z_1^2 .

17.4 Defina vpm_i como o valor do produto marginal do trabalhador i , que é o preço do produto de uma firma multiplicado pelo produto marginal do trabalhador. Assuma que

$$\begin{aligned}\log(vpm_i) &= \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i \\ salário_i &= \max(vpm_i, salmín_i),\end{aligned}$$

onde estão incluídas como variáveis explicativas educação, experiência etc., e $salmín_i$ é o salário mínimo relevante para o indivíduo i . Escreva $\log(salário_i)$ em termos de $\log(vpm_i)$ e $\log(salmín_i)$.

17.5 (Exige cálculo.) Defina $patents$ como o número de patentes requeridas por uma firma durante determinado ano. Assuma que o valor esperado condicional de $patents$, dados $vendas$ e PQ é

$$E(patents|vendas, PQ) = \exp[\beta_0 + \beta_1 \log(vendas) + \beta_2 PQ + \beta_3 PQ^2],$$

onde $vendas$ representa as vendas anuais da firma e PQ é o total de gastos com pesquisa e desenvolvimento nos últimos 10 anos.

- (i) Como você estimaria os β_j ? Justifique sua resposta detalhando a natureza de $patents$.
- (ii) Como você interpreta β_1 ?
- (iii) Encontre o efeito parcial de PQ sobre $E(patents|vendas, PQ)$.

17.6 Considere uma função de poupança familiar para a população de todas as famílias dos Estados Unidos:

$$poup = \beta_0 + \beta_1 renda + \beta_2 tamfam + \beta_3 educ + \beta_4 idade + u,$$

onde $tamfam$ é o tamanho da família, $educ$ são anos de escolaridade do chefe da família e $idade$ é a idade do chefe da família. Assuma que $E(u|renda, tamfam, educ, idade) = 0$.

- (i) Suponha que a amostra inclua apenas famílias cuja idade do seu chefe é superior a 25 anos. Se usarmos o MQO em tal amostra, obteremos estimadores não-viesados dos β_j ? Explique.
- (ii) Agora, suponha que nossa amostra inclua somente casais sem filhos. Podemos estimar todos os parâmetros na equação de poupança? Quais podemos estimar?
- (iii) Suponha que excluirmos de nossa amostra as famílias que pouparam mais de 25.000 dólares por ano. O MQO produzirá estimadores consistentes dos β_j ?

17.7 Suponha que você seja contratado por uma universidade para estudar os fatores que determinam se os alunos admitidos na universidade efetivamente matricularam-se na universidade. Você

recebe uma grande amostra aleatória dos alunos que foram admitidos no ano anterior. Também são disponibilizadas informações sobre se cada aluno decidiu matricular-se, o desempenho no ensino médio, a renda familiar, o auxílio financeiro oferecido, raça e variáveis geográficas. Alguém lhe diz, “Qualquer análise desses dados conduzirá a resultados viesados, pois não se trata de uma amostra aleatória de todos os candidatos às universidades, mas somente daqueles que se candidataram nesta universidade”. Qual sua opinião sobre essa crítica?

Erros-Padrão Assimptóticos em Modelos de Variável Dependente Limitada

As derivações dos erros-padrão assimptóticos dos modelos apresentados neste capítulo estão bem além do escopo deste texto. Não apenas as deduções exigem álgebra matricial como também exigem teoria assimptótica avançada de estimação não-linear. Os fundamentos necessários para uma análise cuidadosa desses métodos e das várias derivações são dados em Wooldridge (2002).

É instrutivo ver as fórmulas de obtenção de erros-padrão assimptóticos de pelo menos alguns dos métodos. Dado um modelo de resposta binária $P(y = 1|x) = G(x\beta)$, onde $G(\cdot)$ é a função logit ou probit, e β é o vetor de parâmetros $k \times 1$, a matriz de variância assimptótica de $\hat{\beta}$ é estimada como

$$\text{Avár}(\hat{\beta}) = \left(\sum_{i=1}^n \frac{[g(x_i\hat{\beta})]^2 x_i' x_i}{G(x_i\hat{\beta})[1 - G(x_i\hat{\beta})]} \right)^{-1}, \quad [17.50]$$

que é uma matriz $k \times k$. (Veja o Apêndice D, no site da Thomson, para um sumário de álgebra matricial.) Sem os termos que envolvem $g(\cdot)$ e $G(\cdot)$, essa fórmula se parece muito com a matriz de variância estimada do estimador MQO, exceto pelo termo $\hat{\sigma}^2$. A expressão em (17.50) leva em consideração a natureza não-linear da probabilidade de resposta — isto é, a natureza não-linear de $G(\cdot)$ — como também a forma particular de heteroscedasticidade em um modelo de resposta binária: $\text{Var}(y|x) = G(x\beta)[1 - G(x\beta)]$.

As raízes quadradas dos elementos diagonais de (17.50) são os erros-padrão assimptóticos dos $\hat{\beta}_j$, e eles são rotineiramente descritos por programas econométricos que suportam análises logit e probit. Dadas essas informações, estatísticas t (assimptóticas) e intervalos de confiança serão obtidos das maneiras habituais.

A matriz em (17.50) também é a base para os testes de Wald de restrições múltiplas de β . [Veja Wooldridge (2002, Capítulo 15).]

A matriz de variância assimptótica do Tobit é mais complicada, mas tem uma estrutura semelhante. Observe que também podemos obter um erro-padrão para $\hat{\sigma}$. A variância assimptótica da regressão de Poisson, considerando $\sigma^2 \neq 1$ em (17.32), tem uma forma muito parecida com (17.50):

$$\text{Avár}(\hat{\beta}) = \hat{\sigma}^2 \left(\sum_{i=1}^n \exp(x_i\hat{\beta}) x_i' x_i \right)^{-1}.$$

As raízes quadradas dos elementos diagonais dessa matriz são os erros-padrão assimptóticos. Se a hipótese de Poisson se mantiver, podemos eliminar $\hat{\sigma}^2$ da fórmula (pois $\sigma^2 = 1$).

Os erros-padrão assimptóticos das regressões censurada e truncada, e da correção da seleção amostral de Heckit são mais complicados, embora compartilhem partes essenciais com as fórmulas anteriores. Veja Wooldridge (2002) para mais detalhes.

Tópicos Avançados sobre Séries Temporais

este capítulo, tratamos de alguns tópicos mais avançados sobre econometria de séries temporais.

Nos capítulos 10, 11 e 12, enfatizamos em várias passagens que o uso de dados de séries temporais na análise de regressão exige alguns cuidados devido à natureza tendencial e persistentes de muitas séries temporais econômicas. Além de estudarmos tópicos como modelos de defasagem distribuída infinita e previsões, também discutimos sobre alguns avanços recentes na análise de processos de séries temporais com raízes unitárias.

Na Seção 18.1, descreveremos os modelos de defasagem distribuída infinita, que permitem uma alteração em uma variável explicativa para afetar todos os valores futuros da variável dependente. Conceitualmente, esses modelos são extensões diretas dos modelos de defasagem distribuída finita estudados no Capítulo 10, mas estimar esses modelos representa alguns desafios interessantes.

Na Seção 18.2, mostraremos como testar formalmente a existência de raízes unitárias em um processo de série temporal. Lembre-se de que no Capítulo 11 excluímos os processos de raiz unitária para aplicarmos a teoria assintótica habitual. Como a presença de uma raiz unitária implica que um choque hoje terá um impacto de longa duração, determinar se um processo tem uma raiz unitária é interessante por sua própria natureza.

Cobriremos, na Seção 18.3, a noção da regressão espúria entre dois processos de séries temporais, em que cada um tem uma raiz unitária. O resultado principal é que, mesmo se séries com duas raízes unitárias forem independentes, é bem provável que a regressão de uma sobre a outra produzirá uma estatística t estatisticamente significante. Isso enfatiza as consequências potencialmente sérias do uso de inferência padrão quando as variáveis dependentes e independentes são processos integrados.

A noção de co-integração se aplica quando duas séries são $I(1)$, mas uma combinação linear delas é $I(0)$; nesse caso, a regressão de uma sobre a outra não será espúria, mas, ao contrário, nos informará alguma coisa sobre a relação de longo prazo entre elas. A co-integração entre duas séries também implica um tipo particular de modelo, chamado de modelo de correção de erro, para a dinâmica de curto prazo. Tratamos desses modelos na Seção 18.4.

Na Seção 18.5, damos uma visão geral da previsão e reunimos todas as ferramentas deste capítulo e dos anteriores para mostrar como os métodos de regressão podem ser usados para prever resultados futuros de uma série temporal. A literatura sobre previsões é vasta, de modo que nos concentramos nos métodos baseados em regressão mais comuns. Também nos referimos ao tópico relacionado da causalidade de Granger.

18.1 MODELOS DE DEFASAGEM DISTRIBUÍDA INFINITA

Seja $\{(y_t, z_t); t = \dots, -2, -1, 0, 1, 2, \dots\}$ um processo de série temporal bivariado (que é observado somente parcialmente). Um **modelo de defasagem distribuída infinita (DDI)** relacionando y_t com o valor atual e todos os valores passados de z é

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \dots + u_t, \quad (18.1)$$

onde a soma dos z defasados estende-se até o passado indefinido. Esse modelo é apenas uma aproximação da realidade, já que nenhum processo econômico começou em um passado infinito. Comparado a um modelo de defasagem distribuída finita, um modelo DDI não requer que trunquemos a defasagem em um valor particular.

Para que o modelo (18.1) tenha sentido, os coeficientes da defasagem, δ_j , devem tender a zero quando $j \rightarrow \infty$. Isso não quer dizer que δ_2 seja menor em magnitude que δ_1 ; isso somente significa que o impacto de z_{t-j} sobre y_t deve finalmente se tornar pequeno conforme j se torna grande. Na maioria das aplicações, isso também tem lógica econômica: o passado distante de z deve ser menos importante para explicar y do que seu passado recente.

Mesmo que decidamos que (18.1) é um modelo útil, não poderemos estimá-la sem algumas restrições. Além disso, somente observamos uma história finita dos dados. A equação (18.1) envolve um número infinito de parâmetros, $\delta_0, \delta_1, \delta_2, \dots$, que não podem ser estimados sem restrições. Mais tarde, colocaremos restrições sobre δ_j que permitirão estimar (18.1).

Como nos modelos de defasagem distribuída finita (DDF), a propensão de impacto em (18.1) é simplesmente δ_0 (veja o Capítulo 10). De forma geral, os δ_h têm a mesma interpretação que em uma DDF. Suponha que $z_s = 0$ para todo $s < 0$ e que $z_0 = 1$ e $z_s = 0$ para todo $s > 1$; em outras palavras, no momento $t = 0$, z aumenta temporariamente em uma unidade e depois reverte-se ao seu nível inicial de zero. Para qualquer $h \geq 0$, teremos $y_h = \alpha + \delta_h + u_h$ de todas $h \geq 0$ e, portanto,

$$E(y_h) = \alpha + \delta_h, \quad (18.2)$$

onde usamos a hipótese padrão que u_h tem média zero. Segue-se que δ_h é a alteração em $E(y_h)$, dada uma mudança temporária de uma unidade em z no momento zero. Acabamos de dizer que δ_h deve tender a zero quando h se torna grande para que a DDI faça sentido. Isso significa que uma alteração temporária em z não terá *efeito de longo prazo* sobre o y esperado: $E(y_h) = \alpha + \delta_h \rightarrow \alpha$ quando $h \rightarrow \infty$.

Assumimos que o processo z começa em $z_s = 0$ e que o aumento de uma unidade ocorreu em $t = 0$. Isso foi feito apenas com fim ilustrativo. De forma mais geral, se z aumentar temporariamente em uma unidade (a partir de qualquer nível inicial) no momento t , então, δ_h medirá a alteração no valor esperado de y após h períodos. A distribuição defasada, que é δ_h traçada como uma função de h , mostra a trajetória esperada que o futuro y seguirá, dado o aumento temporário de uma unidade em z .

A propensão de longo prazo no modelo (18.1) é a soma de todos os coeficientes de defasagens:

$$\text{PLP} = \delta_0 + \delta_1 + \delta_2 + \delta_3 + \dots, \quad (18.3)$$

onde assumimos que a soma infinita está bem definida. Como os δ_j devem convergir para zero, a PLP pode freqüentemente ser bem aproximada por uma soma finita da forma $\delta_0 + \delta_1 + \dots + \delta_p$ para p suficiente.

cientemente grande. Para interpretar a PLP, suponha que o processo z_t seja estável em $z_s = 0$ para $s < 0$. Em $t = 0$, o processo aumenta permanentemente em uma unidade. Por exemplo, se z_t for a mudança percentual na oferta monetária e y_t for a taxa de inflação, então, estaremos interessados nos efeitos de um aumento permanente de um ponto percentual no crescimento da oferta de moeda. Então, fazendo $z_s = 0$ quando $s < 0$ e $z_t = 1$ quando $t \geq 0$, teremos

$$y_h = \alpha + \delta_0 + \delta_1 + \dots + \delta_h + u_h,$$

onde $h \geq 0$ é qualquer horizonte. Como u_t tem média zero para todo t , temos

$$E(y_h) = \alpha + \delta_0 + \delta_1 + \dots + \delta_h. \quad (18.4)$$

[É interessante comparar (18.2) e (18.4).] Quando o horizonte aumenta, isto é, quando $h \rightarrow \infty$, o lado direito de (18.4) será, por definição, a propensão de longo prazo, mais 2. Assim, a PLP indica a alteração de longo prazo no valor esperado de y , dado um aumento permanente de uma unidade em z .

Suponha que $z_s = 0$ para $s < 0$ e que $z_0 = 1$, $z_1 = 1$ e $z_s = 0$ para $s > 1$. Encontre $E(y_{-1})$, $E(y_0)$ e $E(y_h)$ para $h \geq 1$. O que acontece quando $h \rightarrow \infty$?

A derivação anterior da PLP e a interpretação dos δ_j usaram o fato de que os erros têm média zero; como usual, isso não é muito mais que uma hipótese, desde que um intercepto seja incluído no modelo. Um exame mais rigoroso de nosso raciocínio mostra que assumimos que a alteração de z durante qualquer período de tempo não teve efeito sobre o valor esperado de u_t . Essa é a versão de defasagem distribuída infinita da hipótese de *exogeneidade estrita* que apresentamos no Capítulo 10 (em particular, a hipótese ST.2). Formalmente,

$$E(u_t | z_{t-2}, z_{t-1}, z_t, z_{t+1}, \dots) = 0, \quad (18.5)$$

de maneira que o valor esperado de u_t não depende de z em *qualquer* período de tempo. Embora (18.5) seja natural para algumas aplicações, ela elimina outras possibilidades importantes. Na verdade, (18.5) não permite a realimentação de y_t para o z futuro porque z_{t+h} deve ser não correlacionado com u_t para $h > 0$. No exemplo inflação/crescimento da oferta monetária, onde y_t é a inflação e z_t é o crescimento da base monetária, (18.5) elimina as alterações futuras no crescimento da oferta monetária que estão vinculadas às taxas de inflação presente. Considerando que a política monetária freqüentemente tenta manter as taxas de juros e de inflação em certos níveis, isso pode não ser realista.

Uma abordagem para estimar os δ_j , que trataremos na próxima subseção, requer uma hipótese de exogeneidade estrita para produzir estimadores consistentes dos δ_j . Uma hipótese mais fraca é

$$E(u_t | z_t, z_{t-1}, \dots) = 0. \quad (18.6)$$

Sob (18.6), o erro é não-correlacionado com os z atual e passados, mas pode estar correlacionado com z futuro; isso possibilita que z , seja uma variável que segue as regras de decisão que dependam do y passado. Algumas vezes, (18.6) é suficiente para estimar os δ_j ; explicaremos isso na próxima subseção.

Algo que temos que lembrar é que nem (18.5) nem (18.6) dão qualquer informação sobre as propriedades de correlação serial de $\{u_t\}$. (Isso é o mesmo que ocorre nos modelos de defasagem distribuída finita.) No mínimo, podemos esperar que $\{u_t\}$ seja serialmente correlacionado, pois (18.1) não o é, de forma geral, dinamicamente completa no sentido discutido na Seção 11.4. Estudaremos o problema da correlação serial mais tarde.

Como interpretamos os coeficientes da defasagem e da PLP se (18.6) for válida, mas (18.5) não? A resposta é: da mesma maneira que antes. Ainda podemos fazer a experimentação anteriormente imaginada (ou contrafactual), mesmo que os dados que observamos sejam gerados por alguma retroalimentação entre y_t e futuros z . Por exemplo, podemos com certeza perguntar sobre o efeito de longo prazo de um aumento permanente no crescimento da oferta monetária sobre a inflação, mesmo que os dados sobre o crescimento da oferta de moeda não possam ser caracterizados como estritamente exógenos.

A Defasagem Distribuída Geométrica (ou de Koyck)

Como geralmente há um número infinito de δ_j , não podemos estimá-los consistentemente sem algumas restrições. A versão mais simples de (18.1), que ainda faz com que o modelo seja dependente de um número infinito de defasagens, é a **defasagem distribuída geométrica (ou de Koyck)**. Neste modelo, os δ_j dependem somente de dois parâmetros:

$$\delta_j = \gamma p^j, |\rho| < 1, j = 0, 1, 2, \dots . \quad (18.7)$$

Os parâmetros γ e ρ podem ser positivos ou negativos, mas ρ deve ser menor que um em valor absoluto. Isso garante que $\delta_j \rightarrow 0$ quando $j \rightarrow \infty$. De fato, essa convergência acontece a uma velocidade muito rápida. (Por exemplo, com $\rho = 0,5$ e $j = 10$, $\rho^j = 1/1024 < 0,001$.)

A propensão de impacto (PI) na DDI é simplesmente $\delta_0 = \gamma$ e, assim, o sinal da PI é determinado pelo sinal de γ . Se $\gamma > 0$, digamos, e $\rho > 0$, então, os coeficientes de todas as defasagens serão positivos. Se $\rho < 0$, os coeficientes das defasagens terão sinais alternados (ρ^j será negativa quando j for ímpar). A propensão de longo prazo é mais difícil de ser obtida, mas podemos usar um resultado padrão da soma de uma série geométrica: para $|\rho| < 1$, $1 + \rho + \rho^2 + \dots + \rho^j + \dots = 1/(1 - \rho)$ e, portanto

$$PLP = \gamma/(1 - \rho).$$

A PLP tem o mesmo sinal de γ .

Se inserirmos (18.7) em (18.1), ainda teremos um modelo que depende de z retroativa ao passado indefinido. No entanto, uma simples subtração produz um modelo estimável. Escreva a DDI nos momentos t e $t - 1$ como:

$$y_t = \alpha + \gamma z_t + \gamma \rho z_{t-1} + \gamma \rho^2 z_{t-2} + \dots + u_t \quad (18.8)$$

e

$$y_{t-1} = \alpha + \gamma z_{t-1} + \gamma \rho z_{t-2} + \gamma \rho^2 z_{t-3} + \dots + u_{t-1}. \quad (18.9)$$

Se multiplicarmos a segunda equação por ρ e a subtraímos da primeira, todos os termos, com exceção de uns poucos, serão cancelados:

$$y_t - \rho y_{t-1} = (1 - \rho)\alpha + \gamma z_t + u_t - \rho u_{t-1},$$

e podemos escrever a seguinte equação:

$$y_t = \alpha_0 + \gamma z_t + \rho y_{t-1} + u_t - \rho u_{t-1}, \quad (18.10)$$

onde $\alpha_0 = (1 - \rho)\alpha$. Essa equação se parece com um modelo padrão com uma variável dependente defasada, onde z_t aparece contemporaneamente. Como γ é o coeficiente de z_t e ρ é o coeficiente de y_{t-1} , parece que podemos estimar esses parâmetros. [Se, por alguma razão, estivermos interessados em α , poderemos sempre obter $\hat{\alpha} = \hat{\alpha}_0/(1 - \hat{\rho})$ após termos estimado ρ e α_0 .]

A simplicidade de (18.10) é um pouco enganosa. O termo erro nessa equação, $u_t - \rho u_{t-1}$, geralmente é correlacionado com y_{t-1} . De (18.9), fica bastante claro que u_{t-1} e y_{t-1} são correlacionados. Portanto, se escrevermos (18.10) como

$$y_t = \alpha_0 + \gamma z_t + \rho y_{t-1} + v_t, \quad (18.11)$$

onde $v_t = u_t - \rho u_{t-1}$, então, geralmente teremos correlação entre v_t e y_{t-1} . Sem hipóteses adicionais, a estimativa MQO de (18.11) produzirá estimativas inconsistentes de γ e ρ .

Um caso em que v_t deve ser correlacionado com y_{t-1} ocorre quando u_t for independente de z_t e de todos os valores passados de z e y . Nesse caso, (18.8) será dinamicamente completa e, assim, u_t será não-correlacionado com y_{t-1} . De (18.9), a covariância entre v_t e y_{t-1} é $-\rho \text{Var}(u_{t-1}) = -\rho \sigma_u^2$, que será zero somente se $\rho = 0$. Podemos facilmente verificar que v_t é serialmente correlacionado, pois $\{v_t\}$ é serialmente não-correlacionado, $E(v_t v_{t-1}) = E(u_t u_{t-1}) - \rho E(u_{t-1}^2) - \rho E(u_t u_{t-2}) + \rho^2 E(u_{t-1} u_{t-2}) = -\rho \sigma_u^2$. Para $j > 1$, $E(v_t v_{t-j}) = 0$. Assim, $\{v_t\}$ é um processo de média móvel de ordem um (veja a Seção 11.1). Isso e a equação (18.11) dão exemplo de um modelo — que é derivado do modelo de interesse original — que tem uma variável dependente defasada e um tipo particular de correlação serial.

Se fizermos a hipótese de exogeneidade estrita (18.5), z_t será não-correlacionado com u_t e u_{t-1} , e portanto, com v_t . Assim, se pudermos encontrar uma variável instrumental adequada para y_{t-1} , então, poderemos estimar (18.11) por VI. Qual será uma boa candidata a variável instrumental de y_{t-1} ? Por hipótese, u_t e u_{t-1} são ambos não-correlacionados com z_{t-1} e, portanto, z_t é não-correlacionado com z_{t-1} . Se $\gamma \neq 0$, z_{t-1} e y_{t-1} serão correlacionados, mesmo após os efeitos parciais de z_t . Portanto, podemos usar as instrumentais (z_t, z_{t-1}) para estimar (18.11). De forma geral, os erros-padrão precisarão ser ajustados quanto à correlação serial em $\{v_t\}$, como discutimos na Seção 15.7.

Uma alternativa à estimação VI explora o fato de que $\{u_t\}$ pode conter um tipo específico de correlação serial. Particularmente, em adição a (18.6), suponha que $\{u_t\}$ obedeça ao modelo AR(1)

$$u_t = \rho u_{t-1} + e_t \quad (18.12)$$

$$E(e_t | z_t, y_{t-1}, z_{t-1}, \dots) = 0. \quad (18.13)$$

É importante observar que o ρ que aparece em (18.12) é o mesmo parâmetro que multiplica y_{t-1} em (18.11). Se (18.12) e (18.13) se sustentarem, podemos escrever a equação (18.10) como

$$y_t = \alpha_0 + \gamma z_t + \rho y_{t-1} + e_t, \quad (18.14)$$

que é um modelo dinamicamente completo sob (18.13). Conforme visto no Capítulo 11, podemos obter estimadores consistentes e assimptoticamente normais dos parâmetros MQO. Isso é bastante conveniente, já que não há a necessidade de lidarmos com correlação serial nos erros. Se e_t satisfizer a hipótese de homoscedasticidade $\text{Var}(e_t | z_n, y_{t-1}) = \sigma_e^2$, a inferência habitual se aplicará. Assim que tenhamos estimado γ e ρ , poderemos estimar PLP: $\hat{PLP} = \hat{\gamma}/(1 - \hat{\rho})$.

A simplicidade desse procedimento vale-se da hipótese potencialmente forte de que $\{u_t\}$ segue um processo AR(1) com o mesmo ρ que aparece em (18.7). Normalmente, isso não é pior que assumir que os $\{u_t\}$ são serialmente não-correlacionados. Contudo, como a consistência dos estimadores depende fortemente dessa hipótese, uma boa idéia é testá-la. Um teste simples começa pela especificação de $\{u_t\}$ como um processo AR(1) com um parâmetro *diferente*, digamos $u_t = \lambda u_{t-1} + e_t$. McClain e Wooldridge (1995) desenvolveram um teste simples do multiplicador de Lagrange de H_0 : $\lambda = \rho$ que pode ser computado após a estimação MQO de (18.14).

O modelo de defasagem distribuída geométrica se estende para múltiplas variáveis explicativas — de forma a ter uma DD infinita para cada variável explicativa —, mas devemos ser capazes de escrever o coeficiente de $z_{t-j,h}$ como $\gamma_h \rho^j$. Em outras palavras, embora γ_h seja diferente para cada variável explicativa, ρ é o mesmo. Assim, podemos escrever

$$y_t = \alpha_0 + \gamma_1 z_{t1} + \dots + \gamma_k z_{tk} + \rho y_{t-1} + v_t. \quad (18.15)$$

Os mesmos problemas que surgiram no caso com um z surgem no caso com muitos z . Sob a extensão natural de (18.12) e (18.13) — apenas substitua z_t por $z_t = (z_{t1}, \dots, z_{tk})$ —, os estimadores MQO serão consistentes e assimptoticamente normais. Ou, pode ser usado um método VI.

Modelos de Defasagem Distribuída Racional

A DD geométrica implica uma distribuição de defasagem razoavelmente restritiva. Quando $\gamma > 0$ e $\rho > 0$, os δ_j são positivos e monotonicamente declinantes para zero. É possível ter modelos de defasagem distribuída infinita mais gerais. A DDG é um caso especial do que é geralmente chamado de **modelo de defasagem distribuída racional (DDR)**. Uma abordagem geral está além de nosso escopo — Harvey (1990) é uma boa referência —, mas podemos tratar de uma extensão simples e útil.

Um modelo desse tipo é mais facilmente descrito se incluirmos uma defasagem de z na equação (18.11):

$$y_t = \alpha_0 + \gamma_0 z_t + \rho y_{t-1} + \gamma_1 z_{t-1} + v_t, \quad (18.16)$$

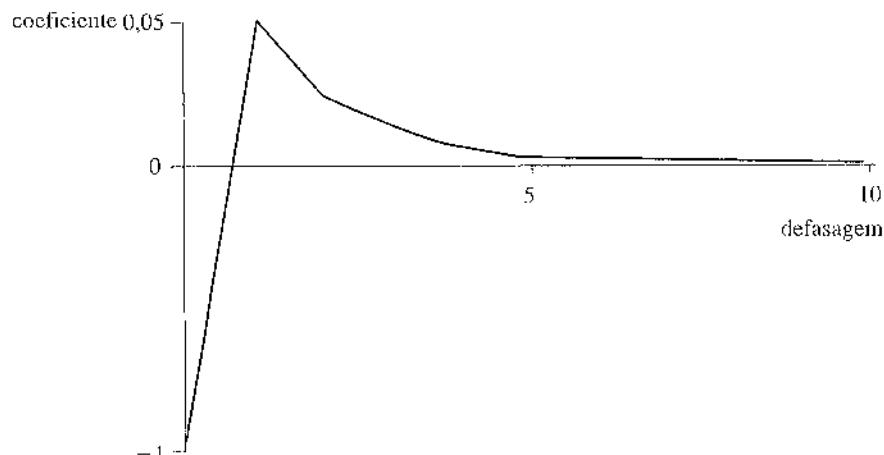
onde $v_t = u_t - \rho u_{t-1}$, como antes. Por substituições repetidas, pode ser mostrado que (18.16) é equivalente ao modelo de defasagem distribuída infinita

$$\begin{aligned}
 y_t &= \alpha + \gamma_0(z_t + \rho z_{t-1} + \rho^2 z_{t-2} + \dots) \\
 &\quad + \gamma_1(z_{t-1} + \rho z_{t-2} + \rho^2 z_{t-3} + \dots) + u_t \\
 &= \alpha + \gamma_0 z_t + (\rho \gamma_0 + \gamma_1) z_{t-1} + \rho(\rho \gamma_0 + \gamma_1) z_{t-2} \\
 &\quad + \rho^2(\rho \gamma_0 + \gamma_1) z_{t-3} + \dots + u_t,
 \end{aligned}$$

onde novamente necessitamos da hipótese $|\rho| < 1$. Dessa última equação, podemos compreender a distribuição da defasagem. Em particular, a propensão de impacto é γ_0 , enquanto o coeficiente de z_{t-h} é $\rho^{h-1}(\rho \gamma_0 + \gamma_1)$ para $h \geq 1$. Portanto, esse modelo permite que a propensão de impacto tenha sinal diferente dos coeficientes das outras defasagens, mesmo se $\rho > 0$. Porém, se $\rho > 0$, δ_h terá o mesmo sinal de $(\rho \gamma_0 + \gamma_1)$ para todo $h \geq 1$. A distribuição da defasagem está traçada na Figura 18.1 para $\rho = 0,5$, $\gamma_0 = -1$ e $\gamma_1 = 1$.

Figura 18.1

Distribuição de defasagem para a defasagem distribuída racional [18.16] com $\rho = 0,5$, $\gamma_0 = -1$ e $\gamma_1 = 1$.



A maneira mais fácil de computar a propensão de longo prazo é definir y e z em seus valores de longo prazo para todos os t , digamos y^* e z^* , e depois encontrar as alterações em y^* em relação a z^* (veja também o Problema 10.3). Temos $y^* = \alpha_0 + \gamma_0 z^* + \rho y^* + \gamma_1 z^*$, cuja solução produz $y^* = \alpha_0/(1 - \rho) + (\gamma_0 + \gamma_1)/(1 - \rho)z^*$. Agora, usamos o fato de que $PLP = \Delta y^*/\Delta z^*$:

$$PLP = (\gamma_0 + \gamma_1)/(1 - \rho).$$

Como $|\rho| < 1$, a PLP terá o mesmo sinal de $\gamma_0 + \gamma_1$, e a PLP será zero se, e somente se, $\gamma_0 + \gamma_1 = 0$, como na Figura 18.1.

EXEMPLO 18.1**(Investimento Imobiliário e Inflação nos Preços Residenciais)**

Estimamos ambos os modelos de defasagem distribuída, o geométrico básico e o racional, pela aplicação do MQO em (18.14) e (18.16), respectivamente. A variável dependente é $\log(invpc)$ após ter sido removida uma tendência temporal linear [isso é, nós retiramos linearmente a tendência de $\log(invpc)$]. Para z_t , usamos o crescimento no índice de preços. Isto nos possibilita estimar de que maneira a inflação nos preços residenciais afeta os movimentos do investimento imobiliário em torno de sua tendência. Os resultados da estimação, utilizando os dados contidos no arquivo HSEINV.RAW, são dados na Tabela 18.1.

Tabela 18.1

Modelos de Defasagem Distribuída para o Investimento Imobiliário

Variável Dependente $\log(invpc)$, sem tendência		
Variáveis Independentes	DD geométrica	DD racional
$crpreço$	3,108 (0,933)	3,256 (0,970)
y_{-1}	0,340 (0,132)	0,547 (0,152)
$cpreço_{-1}$	—	-2,936 (0,973)
constante	-0,010 (0,018)	-0,578 (0,307)
Propensão de Longo Prazo	4,688	0,706
Tamanho da Amostra R -Quadrado Ajustado	41 0,375	40 0,504

O modelo de defasagem distribuída geométrica é claramente rejeitado pelos dados, já que $crpreço_{-1}$ é muito significante. Os R -quadrados ajustados também mostram que o modelo DDR se ajusta muito melhor.

Os dois modelos produzem estimativas bastante diferentes da propensão de longo prazo. Se usarmos incorretamente a DDG, a PLP estimada será quase cinco: um aumento permanente de um ponto percentual na inflação do preço residencial aumenta o investimento imobiliário de longo prazo em 4,7% (acima de seu valor de tendência). Economicamente, isso parece implausível. A PLP estimada pelo modelo de defasagem distribuída racional é menor que um. De fato, não podemos rejeitar a hipótese nula $H_0: \gamma_0 + \gamma_1 = 0$ em qualquer nível razoável de significância (p -valor = 0,83), de modo que não existe qualquer evidência de que a PLP seja diferente de zero. Esse é um bom exemplo de como a má especificação da dinâmica de um modelo omitindo defasagens relevantes pode levar a conclusões errôneas.

18.2 O TESTE DE RAÍZES UNITÁRIAS

Agora voltamos ao importante problema do teste de **raízes unitárias**. No Capítulo 11, demos algumas orientações vagas e necessariamente informais para concluirmos se uma série era I(1) ou não. Em muitos casos, é interessante fazer um teste formal para verificar a existência de raiz unitária. Como veremos, tais testes devem ser aplicados com cautela.

O método mais simples para testar a existência de raiz unitária começa com um modelo AR(1):

$$y_t = \alpha + \rho y_{t-1} + e_t, \quad t = 1, 2, \dots, \quad (18.17)$$

onde y_0 é o valor inicial observado. Em toda esta seção, definimos que $\{e_t\}$ representa um processo que tem média zero, dados os y passados observados:

$$E(e_t | y_{t-1}, y_{t-2}, \dots, y_0) = 0. \quad (18.18)$$

[Sob (18.18), $\{e_t\}$ é chamado de **sequência da diferença de um martingale** em relação a $\{y_{t-1}, y_{t-2}, \dots\}$. Se assumirmos que $\{e_t\}$ é i.i.d. com média zero e é independente de y_0 , então, ele também satisfará (18.18).]

Se $\{y_t\}$ seguir (18.17), ele terá uma raiz unitária se, e somente se, $\rho = 1$. Se $\alpha = 0$ e $\rho = 1$, $\{y_t\}$ seguirá um passeio aleatório sem tendência [com as inovações e_t satisfazendo (18.18)]. Se $\alpha \neq 0$ e $\rho = 1$, $\{y_t\}$ será um passeio aleatório com tendência, o que significa que $E(y_t)$ é uma função linear de t . Um processo de raiz unitária com tendência se comporta de maneira bastante diferente de outro sem tendência. Contudo, é comum deixar α sem especificação sob a hipótese nula, e essa é a abordagem que usamos. Portanto, a hipótese nula é que $\{y_t\}$ tem uma raiz unitária:

$$H_0: \rho = 1. \quad (18.19)$$

Em quase todos os casos, estamos interessados na alternativa unilateral

$$H_1: \rho < 1. \quad (18.20)$$

(Na prática, isso significa que $0 < \rho < 1$, pois será muito raro ter $\rho < 0$ para uma série que suspeitamos ter uma raiz unitária). A alternativa $H_1: \rho > 1$ usualmente não é considerada, pois ela implica que y_t é explosivo. De fato, se $\alpha > 0$, y_t terá uma tendência exponencial em sua média quando $\rho > 1$.

Quando $|\rho| < 1$, $\{y_t\}$ é um processo AR(1) estável, o que significa que ele é fracamente dependente ou assimptoticamente não-correlacionado. Lembre-se do Capítulo 11 que $\text{Corr}(y_t, y_{t+h}) = \rho^h \rightarrow 0$ quando $|\rho| < 1$. Portanto, testar (18.19) no modelo (18.17), com a hipótese alternativa dada por (18.20), é, na realidade, um teste para verificar se $\{y_t\}$ é I(1) contra a hipótese alternativa de que $\{y_t\}$ é I(0). [A razão pela qual não consideramos a hipótese nula ser I(0) nessa estrutura é que $\{y_t\}$ é I(0) para qualquer valor de ρ estritamente entre -1 e 1, algo que os testes de hipóteses clássicos não resolvem com facilidade. Existem testes nos quais a hipótese nula é I(0) contra a hipótese alternativa I(1), mas eles utilizam métodos diferentes. Veja, por exemplo, Kwiatkowski, Phillips, Schmidt e Shin (1992).]

Uma equação conveniente para realizar o teste de raiz unitária é subtrair y_{t-1} de ambos os lados de (18.17) e definir $\theta = \rho - 1$:

$$\Delta y_t = \alpha + \theta y_{t-1} + e_t \quad (18.21)$$

Sob (18.18), esse será um modelo dinamicamente completo, e assim ele parece ser adequado para testar $H_0: \theta = 0$ contra $H_1: \theta < 0$. O problema é que sob H_0 , y_{t-1} é $I(1)$ e, assim, o teorema do limite central usual que fundamenta a distribuição normal padrão assumptótica para a estatística t não é aplicável: a estatística t não possui uma distribuição normal padrão aproximada mesmo em amostras de tamanhos grandes. A distribuição assumptótica da estatística t sob H_0 tornou-se conhecida como a **distribuição de Dickey-Fuller**, nome advindo de Dickey-Fuller (1979).

Embora não possamos usar os valores críticos habituais, podemos usar a estatística t usual de $\hat{\theta}$ em (18.21), pelo menos logo que os valores críticos apropriados tenham sido tabulados. O teste resultante é conhecido como **teste de Dickey-Fuller (DF)** de uma raiz unitária. A teoria usada para obter os valores críticos assumptóticos é bastante complicada e é tratada em textos avançados de econometria de séries temporais. [Veja, por exemplo, Banerjee, Dolado, Galbraith e Hendry (1993), ou BDGH, para um breve panorama.] Em contraposição, a utilização desses resultados é bastante fácil. Os valores críticos da estatística t foram tabulados por vários autores, iniciando com o trabalho original de Dickey-Fuller (1979). A Tabela 18.2 contém os valores críticos de amostras grandes de vários níveis de significância, extraídas de BDGH (1993, Tabela 4.2). (Valores críticos ajustados para amostras de tamanhos pequenos estão disponíveis no trabalho de BDGH.)

Tabela 18.2Valores Críticos Assumptóticos do Teste t de Raiz Unitária: Sem Tendência Temporal

Nível de significância	1%	2,5%	5%	10%
Valor Crítico	-3,43	-3,12	-2,86	-2,57

Rejeitamos a hipótese nula $H_0: \theta = 0$ contra $H_1: \theta < 0$ se $t_{\hat{\theta}} < c$, onde c é um dos valores negativos da Tabela 18.2. Por exemplo, para realizar o teste no nível de significância de 5%, rejeitaremos H_0 se $t_{\hat{\theta}} < -2,86$. Isso requer uma estatística t com uma magnitude muito maior do que se usássemos o valor crítico normal padrão, que seria $-1,65$. Se usarmos o valor crítico normal padrão para testar a existência de uma raiz unitária, rejeitaremos H_0 com mais freqüência que 5% das vezes quando H_0 for verdadeira.

EXEMPLO 18.2**(Teste de Raiz Unitária das Taxas das Letras do Tesouro de Três Meses)**

Usamos os dados trimestrais do arquivo INTQRT.RAW para testar a existência de uma raiz unitária nas taxas das letras do Tesouro norte-americano de três meses. Quando estimamos (18.20), obtemos

$$\begin{aligned} \Delta \hat{r}_3 &= 0,625 - 0,091 r_{3,t-1} \\ &\quad (0,261) \quad (0,037) \quad (18.22) \\ n &= 123, R^2 = 0,048, \end{aligned}$$

EXEMPLO 18.2 (continuação)

onde mantemos a convenção de descrever os erros-padrão entre parênteses abaixo das estimativas. Devemos nos lembrar que esses erros-padrão não podem ser usados para a construção dos intervalos de confiança habituais ou para a realização dos testes t tradicionais, pois eles não se comportam das maneiras habituais quando existe uma raiz unitária. O coeficiente de $r\beta_{t-1}$ mostra que a estimativa de ρ é $\hat{\rho} = 1 + \hat{\theta} = 0,909$. Embora isso seja menor que a unidade, não sabemos se é estatisticamente menor que um. A estatística t de $r\beta_{t-1}$ é $-0,091/0,037 = -2,46$. Pela Tabela 18.2, o valor crítico a 10% é $-2,57$; portanto, não é possível rejeitar $H_0: \rho = 1$ contra $H_1: \rho < 1$ ao nível de 10%.

Como acontece com outros testes de hipóteses, quando não é possível rejeitar H_0 , isso não quer dizer que aceitamos H_0 . Por quê? Suponha que testemos $H_0: \rho = 0,9$ no exemplo anterior usando um teste t padrão que é assimptoticamente válido, pois y_t é I(0) sob H_0 . Então, obteremos $t = 0,001/0,037$, que é muito pequeno e não oferece qualquer evidência contra $\rho = 0,9$. Porém, não faz sentido aceitarmos $\rho = 1$ e $\rho = 0,9$.

Quando não é possível rejeitar uma raiz unitária, como no exemplo anterior, devemos apenas concluir que os dados não fornecem forte evidência contra H_0 . No exemplo, o teste fornece certa evidência contra H_0 , pois a estatística t está próxima do valor crítico a 10%. (Idealmente, computaríamos um p -valor, mas isso exige um programa especial devido à distribuição não normal.) Além disso, embora $\hat{\rho} \approx 0,91$ implique uma quantidade razoável de persistência em $\{r\beta_t\}$, a correlação entre observações que estejam separadas por dez períodos de tempo em um modelo AR(1) com $\rho = 0,9$ é de aproximadamente 0,35, em vez de quase um se $\rho = 1$.

O que acontece se agora quisermos usar $r\beta_t$ como uma variável explicativa em uma análise de regressão? O resultado do teste de raiz unitária implica que devemos ser extremamente cautelosos: se $r\beta_t$ tiver uma raiz unitária, as aproximações assimptóticas não terão que ser válidas (como discutimos no Capítulo 11). Uma solução é usar a primeira diferença de $r\beta_t$ em qualquer análise. Como veremos na Seção 18.4, essa não é a única possibilidade.

Também precisamos testar a existência de raízes unitárias em modelos com dinâmicas mais complicadas. Se $\{y_t\}$ seguir (18.17) com $\rho = 1$, então, Δy_t será serialmente não-correlacionado. Podemos facilmente permitir que $\{\Delta y_t\}$ siga um modelo AR, aumentando a equação (18.21) com defasagens adicionais. Por exemplo,

$$\Delta y_t = \alpha + \theta y_{t-1} + \gamma_1 \Delta y_{t-1} + e_t \quad (18.23)$$

onde $|\gamma_1| < 1$. Isso garante que, sob $H_0: \theta = 0$, $\{\Delta y_t\}$ siga um modelo AR(1) estável. Sob a hipótese alternativa $H_1: \theta < 0$, pode ser mostrado que $\{y_t\}$ segue um modelo AR(2) estável.

De forma mais genérica, podemos adicionar p defasagens de Δy_t na equação para levar em conta a dinâmica no processo. A maneira como testamos a hipótese nula de uma raiz unitária é bastante semelhante: computamos a regressão de

$$\Delta y_t \text{ sobre } y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p} \quad (18.24)$$

e realizamos o teste t de $\hat{\theta}$, o coeficiente de y_{t-1} , exatamente como antes. Essa versão estendida do teste de Dickey-Fuller é normalmente chamada de **teste de Dickey-Fuller aumentado**, pois a regressão foi aumentada com as alterações defasadas Δy_{t-h} . Os valores críticos e a regra de rejeição são os mesmos

de antes. A inclusão das variações defasadas em (18.24) tem por objetivo retirar qualquer correlação serial em Δy_t . Quanto mais defasagens incluirmos em (18.24), mais observações iniciais perderemos. Se incluirmos defasagens demasia, o poder de amostra pequena do teste sofrerá. Mas se incluirmos muito poucas defasagens, o tamanho do teste será incorreto, mesmo assintoticamente, pois a validade dos valores críticos na Tabela 18.2 depende de a dinâmica ter sido completamente modelada. Freqüentemente, a extensão da defasagem é ditada pela freqüência dos dados (como também pelo tamanho da amostra). Para dados anuais, uma ou duas defasagens em geral são suficientes. Para dados mensais, devemos incluir 12 defasagens. Mas não existem regras rígidas a serem seguidas em qualquer caso.

Interessante observar como as estatísticas t das variações defasadas têm distribuições t aproximadas. As estatísticas F da significância conjunta de qualquer grupo de termos Δy_{t-h} são também assintoticamente válidas. (Elas mantêm a hipótese de homoscedasticidade discutida na Seção 11.5.) Portanto, podemos usar testes padrão para determinar se temos suficientes variações defasadas em (18.24).

EXEMPLO 18.3

(Teste de Raiz Unitária da Inflação Norte-Americana Anual)

Utilizamos os dados anuais da inflação norte-americana, com base no IPC, para testar uma raiz unitária na inflação (veja o arquivo PHILLIPS.RAW). A série se estende de 1948 a 1996. A permissão de uma defasagem de Δinf_t na regressão de Dickey-Fuller produz

$$\begin{aligned}\Delta \hat{\text{inf}}_t &= 1,36 - 0,310 \text{ inf}_{t-1} + 0,138 \Delta \text{inf}_{t-1} \\ &\quad (0,517) (0,103) \quad (0,126) \\ n &= 47, R^2 = 0,172.\end{aligned}$$

A estatística t do teste de raiz unitária é $-0,310/0,103 = -3,01$. Como o valor crítico a 5% é $-2,86$, rejeitamos a hipótese de raiz unitária ao nível de 5%. A estimativa de ρ está em torno de 0,690. Juntas, é uma evidência razoavelmente forte contra a existência de uma raiz unitária na inflação. A defasagem Δinf_{t-1} tem uma estatística t aproximada de 1,10, de modo que não precisamos incluí-la, mas não tínhamos como saber disso antecipadamente. Se eliminarmos Δinf_{t-1} , a evidência contra uma raiz unitária é um pouco mais forte: $\hat{\theta} = -0,335 (\hat{\rho} = 0,665)$ e $t_{\hat{\theta}} = -3,13$.

Para séries que claramente possuem tendências temporais, precisamos modificar o teste de raízes unitárias. Um processo estacionário na tendência — que possui uma tendência linear em sua média, mas é I(0) em relação à sua tendência — pode ser confundido com um processo de raiz unitária se não controlarmos uma tendência temporal na regressão de Dickey-Fuller. Em outras palavras, se conduzirmos o teste DF habitual ou o teste DF aumentado em uma série com tendência, mas que é I(0), provavelmente teremos pouco poder para rejeitar uma raiz unitária.

Para permitir séries com tendência temporal, alteramos a equação básica para

$$\Delta y_t = \alpha + \delta t + \theta y_{t-1} + e_t, \quad (18.25)$$

onde novamente a hipótese nula é $H_0: \theta = 0$ e a alternativa é $H_1: \theta < 0$. De acordo com a hipótese alternativa, $\{y_t\}$ é um processo estacionário em torno de sua tendência. Se y_t tiver uma raiz unitária, então, $\Delta y_t = \alpha + \delta t + e_t$, e, assim, a mudança em y_t terá uma média linear em t a menos que $\delta = 0$. [Pode ser mostrado que $E(y_t)$ é na realidade um valor quadrático em t .] É pouco comum a primeira diferença de uma série econômica ter uma tendência linear e, assim, uma hipótese nula mais apropriada seria $H_0: \theta = 0, \delta = 0$. Embora seja possível testar essa hipótese conjunta usando um teste F — mas com valores críticos modificados — é comum testar-se apenas $H_0: \theta = 0$ usando um teste t . Aqui seguimos esse método. [Veja BDGH (1993, Seção 4.4), para mais detalhes sobre o teste conjunto.]

Quando incluímos uma tendência temporal na regressão, os valores críticos do teste mudam. Intuitivamente, isso acontece porque a retirada da tendência de um processo de raiz unitária tende a fazer com que ele se pareça mais com um processo I(0). Portanto, requeremos uma magnitude maior da estatística t para rejeitar H_0 . Os valores críticos de Dickey-Fuller para o teste t que incluem uma tendência temporal são apresentados na Tabela 18.3; eles foram tirados de BDGH (1993, Tabela 4.2).

Tabela 18.3

Valores críticos assintóticos do teste t de raiz unitária: Tendência Temporal Linear

Nível de Significância	1%	2,5%	5%	10%
Valor Crítico	-3,96	-3,66	-3,41	-3,12

Por exemplo, para rejeitar uma raiz unitária ao nível de 5%, precisamos que a estatística t de $\hat{\theta}$ seja menor que -3,41, em comparação com -2,86 sem uma tendência temporal.

Podemos aumentar a equação (18.25) com defasagens da Δy_t para levar em conta a correlação serial, da mesma forma que no caso sem uma tendência. Isso não altera a maneira de conduzir o teste.

EXEMPLO 18.4

{Raiz Unitária no Log do Produto Interno Bruto Real dos Estados Unidos}

Podemos aplicar o teste de raiz unitária com uma tendência temporal aos dados do PIB dos Estados Unidos contidos no arquivo INVEN.RAW. Esses dados anuais referem-se ao anos de 1959 a 1995. Testamos se $\log(PIB_t)$ tem uma raiz unitária. Essa série tem uma tendência pronunciada que parece, em termos gerais, linear. Incluímos uma única defasagem de $\Delta \log(PIB_t)$, que é simplesmente o crescimento no PIB (em forma decimal) para considerar a dinâmica:

$$\begin{aligned} g\hat{PIB}_t &= 1,65 + 0,0059 t - 0,210 \log(PIB_{t-1}) + 0,264 g\log(PIB_{t-1}) \\ &\quad (0,67) \quad (0,0027) \quad (0,087) \quad (0,165) \\ &\quad n = 35, R^2 = 0,268. \end{aligned} \tag{18.26}$$

Dessa equação, obtemos $\hat{\rho} = 1 - 0,21 = 0,79$, que claramente é menor que um. Mas não podemos rejeitar uma raiz unitária no log do PIB: a estatística t de $\log(PIB_{t-1})$ é $-0,210/0,087 = -2,41$, que está bem acima do valor crítico de -3,12 ao nível de 10%. A estatística t de $g\log(PIB_{t-1})$ é 1,60, que é quase significante ao nível de 10% contra uma alternativa bilateral.

Qual deve ser nossa conclusão quanto a uma raiz unitária? Novamente, não podemos rejeitar uma raiz unitária, mas a estimativa por pontos de ρ não está especialmente próxima de um. Quando temos uma amostra de tamanho pequeno — e $n = 35$ é considerado muito pequeno —, é muito difícil rejeitar a hipótese nula de uma raiz unitária se o processo tiver alguma coisa próxima de uma raiz unitária. Utilizando mais

EXEMPLO 18.4

dados em períodos de tempo mais longos, muitos pesquisadores concluíram haver pouca evidência contra a hipótese de raiz unitária de $\log(PIB)$. Isso tem levado muitos deles a assumir que o crescimento do PIB é I(0), o que significa que $\log(PIB)$ é I(1). Infelizmente, considerando os tamanhos de amostras atualmente disponíveis, não podemos ter muita confiança nessa conclusão.

Se omitirmos a tendência temporal, haverá muito menos evidência contra H_0 , pois $\hat{\theta} = -0,023$ e $t_{\hat{\theta}} = -1,92$. Nesse caso, a estimativa de ρ está muito mais próxima de um, mas isso é enganoso devido a omissão da tendência temporal.

É tentador comparar a estatística t na tendência temporal em (18.26) com o valor crítico de uma distribuição padrão normal ou distribuição t para verificar se a tendência temporal é significante. Infelizmente, a estatística t da tendência não tem uma distribuição normal padrão assimptótica (a menos que $|\rho| < 1$). A distribuição assimptótica dessa estatística t é conhecida, mas é raramente usada. Em geral, usamos a intuição (ou gráficos da série temporal) para tomar a decisão de incluir uma tendência no teste DF.

Existem muitas outras variantes dos testes de raiz unitária. Em uma versão que somente é aplicável em séries que claramente não têm tendências, o intercepto é omitido da regressão; isto é, α é definido como 0 em (18.21). Essa variante do teste de Dickey-Fuller é raramente usada devido ao viés induzido se $\alpha \neq 0$. Além disso, podemos permitir séries temporais mais complicadas, tal como uma quadrática. Novamente, isso raramente é usado.

Uma outra categoria de testes tenta levar em conta a correlação serial em Δy , de uma maneira diferente da inclusão de defasagens em (18.21) ou (18.25). O método está relacionado com os erros-padrão robustos em relação à correlação serial dos estimadores MQO que discutimos na Seção 12.5. A idéia é ser tão agnóstico quanto possível sobre a correlação serial em Δy . Na prática, o teste de Dickey-Fuller (aumentado) tem se sustentado bem. [Veja BDGH (1993, Seção 4.3) para explicações sobre outros testes.]

18.3 REGRESSÃO ESPÚRIA

Em um ambiente de corte transversal, usamos a frase “correlação espúria” para descrever uma situação na qual duas variáveis são relacionadas pelas suas correlações com uma terceira variável. Em particular, se regredirmos y sobre x , encontraremos uma relação significante. Entretanto, quando controlarmos outra variável, digamos z , o efeito parcial de x sobre y torna-se zero. Naturalmente, isso também pode acontecer em contextos de séries temporais com variáveis I(0).

Como discutimos na Seção 10.5, é possível encontrar uma relação espúria entre séries temporais que tenham tendência crescente ou decrescente. Desde que as séries sejam fracamente dependentes de suas tendências temporais, o problema será solucionado de maneira eficaz pela inclusão de uma tendência temporal no modelo de regressão.

Quando lidamos com processos integrados de ordem um, existe uma complicação adicional. Mesmo que as duas séries tenham médias sem tendência, uma regressão simples envolvendo duas séries I(1) *independentes* freqüentemente resultará em uma estatística t significante.

Para sermos mais precisos, definimos $\{x_t\}$ e $\{y_t\}$ como passeios aleatórios gerados por

$$x_t = x_{t-1} + a_t, \quad t = 1, 2, \dots, \tag{18.27}$$

e

$$y_t = y_{t-1} + e_t, t = 1, 2, \dots, \quad (18.28)$$

onde $\{a_t\}$ e $\{e_t\}$ são inovações independentes e identicamente distribuídas, com média zero e variâncias σ^2_a e σ^2_e , respectivamente. Concretamente, tome os valores iniciais como $x_0 = y_0 = 0$. Assuma ainda que $\{a_t\}$ e $\{e_t\}$ sejam processos independentes. Isso implica que $\{x_t\}$ e $\{y_t\}$ também são independentes. Mas se executarmos a regressão simples

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t, \quad (18.29)$$

e obtivermos a estatística t habitual de $\hat{\beta}_1$ e o R -quadrado habitual? Como y_t e x_t são independentes, esperaríamos que $\text{plim } \hat{\beta}_1 = 0$. De forma ainda mais importante, se testarmos $H_0: \beta_1 = 0$ contra $H_1: \beta_1 \neq 0$ ao nível de 5%, esperamos que a estatística t de $\hat{\beta}_1$ seja não significante em 95% das vezes. Por meio de uma simulação, Granger e Newbold (1974) demonstraram *não* ser esse o caso: embora y_t e x_t sejam *independentes*, a regressão de y_t sobre x_t produz uma estatística t estatisticamente significante em uma grande porcentagem das vezes, muito maior que o nível de significância nominal. Granger e Newbold chamaram isso de **problema da regressão espúria**: não tem sentido y e x serem relacionadas, mas uma regressão MQO usando as estatísticas t normais freqüentemente indica uma relação entre ambos.

Sob a definição precedente, na qual $\{x_t\}$ e $\{y_t\}$ são geradas por (18.27) e (18.28) e $\{e_t\}$ e $\{a_t\}$ são seqüências i.i.d., qual é o plim do coeficiente de inclinação, digamos $\hat{\gamma}_1$, da regressão de Δy_t sobre Δx_t ? Descreva o comportamento da estatística t de $\hat{\gamma}_1$.

Novos resultados de simulações são fornecidos por Davidson e MacKinnon (1993, Tabela 19.1), onde a_t e e_t são gerados como variáveis aleatórias normais, independentes e identicamente distribuídas, e 10.000 amostras diferentes são geradas. Em uma amostra de tamanho $n = 50$ ao nível de significância de 5%, a estatística t padrão de $H_0: \beta_1 = 0$ contra a alternativa bilateral rejeita H_0 cerca de 66,2% das vezes sob H_0 , em vez de 5% das vezes. Conforme o tamanho da amostra aumenta, a situação *piora*: com $n = 250$, a hipótese nula é rejeitada 84,7% das vezes!

Eis uma maneira de sabermos o que está acontecendo quando regredimos o nível de y sobre o nível de x . Escreva o modelo subjacente (18.29) como

$$y_t = \beta_0 + \beta_1 x_t + u_t. \quad (18.30)$$

Para que a estatística t de $\hat{\beta}_1$ tenha uma distribuição normal padrão aproximada em amostras grandes, no mínimo, $\{u_t\}$ deve ser um processo de média zero, serialmente não-correlacionado. Mas sob $H_0: \beta_1 = 0$, $y_t = \beta_0 + u_t$, e como $\{y_t\}$ é um passeio alcatório iniciando em $y_0 = 0$, a equação (18.30) será válida sob H_0 somente se $\beta_0 = 0$ e, mais importante, se $\bar{u}_t = y_t = \sum_{j=1}^t e_j$. Em outras palavras, $\{u_t\}$ é um

passeio aleatório sob H_0 . Isso claramente viola até mesmo a versão assintótica das hipóteses de Gauss-Markov do Capítulo 11.

A inclusão de uma tendência temporal não altera muito a conclusão. Se y_t ou x_t for um passeio aleatório com tendência e uma tendência temporal não for incluída, o problema da regressão espúria será ainda pior. As mesmas conclusões qualitativas serão válidas se $\{a_t\}$ e $\{e_t\}$ forem processos gerais I(0), em vez de seqüências i.i.d.

Além de a estatística t habitual não ter uma distribuição normal padrão limitante — aliás, ela aumenta para o infinito conforme $n \rightarrow \infty$ —, o comportamento do R -quadrado não é padrão. Em contextos de cortes transversais ou em regressões com variáveis I(0) de séries temporais, o R -quadrado converge em probabilidade para o R -quadrado da população: $1 - \sigma_u^2 / \sigma_y^2$. Esse não é o caso em regressões espúrias com processos I(1). Em vez de o R -quadrado ter um plim bem definido, ele na verdade converge para uma variável aleatória. A formalização dessa noção está além do escopo deste texto. [Uma explicação das propriedades assintóticas da estatística t e do R -quadrado pode ser encontrada em BDGH (Seção 3.1).] A implicação é que o R -quadrado é grande com probabilidade alta, mesmo que $\{y_t\}$ e $\{x_t\}$ sejam processos de séries temporais independentes.

As mesmas considerações surgem com variáveis independentes múltiplas, cada uma delas podendo ser I(1) ou algumas delas podendo ser I(0). Se $\{y_t\}$ for I(1) e pelo menos algumas das variáveis explicativas forem I(1), os resultados da regressão podem ser espúrios.

A possibilidade de regressão espúria com variáveis I(1) é bastante importante e tem levado os economistas a reexaminarem muitas regressões de séries temporais agregadas cujas estatísticas t eram muito significantes e cujos R -quadrados eram extremamente altos. Na próxima seção, mostraremos que a regressão de uma variável dependente I(1) sobre uma variável independente I(1) *pode* ser informativa, mas somente se essas variáveis forem relacionadas em um sentido preciso.

18.4 CO-INTEGRAÇÃO E MODELOS DE CORREÇÃO DE ERRO

A discussão sobre regressão espúria na seção anterior certamente faz com que sejamos cuidadosos quanto ao uso dos níveis de variáveis I(1) em análise de regressão. Em capítulos anteriores, sugerimos que variáveis I(1) deveriam ser diferenciadas antes de serem usadas em modelos de regressão linear, fossem elas estimadas por MQO ou por variáveis instrumentais. Certamente esse é um caminho seguro a ser seguido, e é o método usado em muitas regressões de séries temporais a partir do trabalho original de Granger e Newbold sobre o problema da regressão espúria. Infelizmente, diferenciar sempre as variáveis I(1) limita a extensão das questões que podemos responder.

Co-Integração

A noção de **co-integração**, que recebeu um tratamento formal em Engle e Granger (1987), torna as regressões que envolvam variáveis I(1) potencialmente significativas. Um tratamento completo da co-integração envolve matemática, mas podemos descrever os problemas e métodos básicos que são usados em muitas aplicações.

Se $\{y_t; t = 0, 1, \dots\}$ e $\{x_t; t = 0, 1, \dots\}$ forem dois processos I(1), então, de forma geral, $y_t - \beta x_t$ será um processo I(1) para qualquer número β . No entanto, é possível que para alguns $\beta \neq 0$, $y_t - \beta x_t$ seja um processo I(0), o que significará que ele tem média constante, variância constante, autocorrelações que dependem somente da distância temporal entre duas variáveis quaisquer na série, e é assintoticamente não-correlacionado. Se existir tal β , dizemos que y e x são *co-integrados*, e chamamos β de parâmetro de co-integração. [Alternativamente, poderíamos examinar $x_t - \gamma y_t$ para $\gamma \neq 0$: se $y_t - \beta x_t$ for I(0), então, $x_t - (1/\beta)y_t$ será I(0). Portanto, a combinação linear de y e x , não é única, mas se fixar-

mos o coeficiente de y_t na unidade, então, β será único. Veja o Problema 18.3. Em termos concretos, consideramos combinações lineares da forma $y_t - \beta x_t$.]

[Exercício]

Seja $\{(y_t, x_t); t = 1, 2, \dots\}$ como sendo uma série temporal bivariada na qual cada série é I(1) sem tendência. Explique por que, se y_t e x_t forem co-integradas, y_t e x_{t-1} também serão co-integradas.

A título de ilustração, considere $\beta = 1$, suponha que $y_0 = x_0 = 0$ e escreva $y_t = y_{t-1} + r_t, x_t = x_{t-1} + v_t$, onde $\{r_t\}$ e $\{v_t\}$ são dois processos I(0) com médias zero. Nesse caso, y_t e x_t têm uma tendência de oscilar e não retornar ao valor zero inicial com qualquer regularidade. Em contraposição, se $y_t - x_t$ for I(0), terá média zero e retornará a zero com alguma regularidade.

Como um exemplo específico, defina $r6_t$ como a taxa de juros anualizada das letras do Tesouro norte-americano de seis meses (no final do trimestre t) e defina $r3_t$ como a taxa de juros anualizada das letras do Tesouro de três meses. (Esses dados são conhecidos como rendimentos equivalentes de títulos, e são divulgados nas páginas financeiras dos jornais.) No Exemplo 18.2, utilizando os dados contidos no arquivo INTQRT.RAW, encontramos pouca evidência contra a hipótese de que $r3_t$ tenha uma raiz unitária; o mesmo ocorre com $r6_t$. Defina o *spread* entre as taxas de seis e três meses como $spr_t = r6_t - r3_t$. Então, usando a equação (18.21), a estatística t de Dickey-Fuller de spr_t será $-7,71$ (com $\hat{\theta} = -0,67$ ou $\hat{\rho} = 0,33$). Portanto, rejeitamos firmemente uma raiz unitária para spr_t em favor de I(0). A conclusão é que, embora $r6_t$ e $r3_t$ pareçam ser processos de raízes unitárias, a diferença entre elas é um processo I(0). Em outras palavras, $r6_t$ e $r3_t$ são co-integradas.

A co-integração nesse exemplo, como em muitos outros casos, tem uma interpretação econômica. Se $r6_t$ e $r3_t$ não fossem co-integradas, a diferença entre as taxas de juros poderia se tornar muito grande, sem tendência de elas voltarem a se juntar. Com base em um argumento simples de arbitragem, isso não parece ser razoável. Suponha que o *spread* spr_t continue a crescer durante vários períodos de tempo, tornando as letras de seis meses um investimento muito mais desejável. Nesse caso, os investidores sairiam das letras de três meses indo para as de seis meses, elevando os preços desses papéis e baixando os de três meses. Como juros são inversamente relacionados a preços, isso diminuiria $r6_t$ e aumentaria $r3_t$ até que o *spread* entre eles fosse reduzido. Portanto, não se espera que grandes desvios entre $r6_t$ e $r3_t$ continuem: o *spread* tem uma tendência de retornar ao seu valor médio. (O *spread*, na realidade, tem uma média ligeiramente positiva, pois os investidores de longo prazo são mais bem recompensados em relação aos de curto prazo.)

Existe outra maneira de caracterizar o fato de que spr_t não se desviará por longos períodos de seu valor médio: $r6_t$ e $r3_t$ têm uma relação de *longo prazo*. Para descrever o que queremos dizer com isso, defina $\mu = E(spr_t)$ como o valor esperado do *spread*. Então, podemos escrever

$$r6_t = r3_t + \mu + e_t$$

onde $\{e_t\}$ é um processo I(0) de média zero. O equilíbrio, ou relação de longo prazo ocorre quando $e_t = 0$, ou $r6_t^* = r3_t^* + \mu$. Em qualquer período de tempo pode haver desvios do equilíbrio, mas eles serão temporários: existem forças econômicas que levam $r6_t$ e $r3_t$ de volta à relação de equilíbrio.

No exemplo da taxa de juros, usamos o raciocínio econômico para nos informar o valor de β se y_t e x_t forem co-integrados. Se tivermos um valor hipotético de β , é fácil *testar* se as duas séries são co-integradas: simplesmente definimos uma nova variável, $s_t = y_t - \beta x_t$, e aplicamos o teste DF, usual ou

aumentado, em $\{s_t\}$. Se *rejeitarmos* uma raiz unitária em $\{s_t\}$ em favor da alternativa $I(0)$, constataremos que y_t e x_t são co-integrados. Em outras palavras, a hipótese nula é que y_t e x_t não são co-integrados.

Testar a existência de co-integração é mais difícil quando o parâmetro (potencial) de co-integração β é desconhecido. Em vez de testar a existência de uma raiz unitária em $\{s_t\}$, primeiro precisamos estimar β . Se y_t e x_t forem co-integrados, constataremos que o estimador MQO $\hat{\beta}$ da regressão

$$y_t = \hat{\alpha} + \hat{\beta}x_t \quad (18.31)$$

será consistente para β . O problema é que a hipótese nula específica que as duas séries são *não* co-integradas, o que significa que, de acordo com a H_0 , estaremos computando uma regressão espúria. Felizmente, é possível tabular valores críticos mesmo quando β é estimado, onde aplicamos o teste usual, ou aumentado de Dickey-Fuller nos resíduos, digamos $\hat{u}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$, a partir de (18.31). A única diferença é que os valores críticos são responsáveis pela estimativa de β . Os valores críticos assintóticos são dados na Tabela 18.4. Eles foram extraídos de Davidson e MacKinnon (1993, Tabela 20.2).

Tabela 18.4

Valores Críticos Assimptóticos do Teste de Co-integração: Sem Tendência Temporal

Nível de Significância	1%	2,5%	5%	10%
Valor Crítico	-3,90	-3,59	-3,34	-3,04

No teste básico, computamos a regressão de $\Delta\hat{u}_t$ sobre \hat{u}_{t-1} e compararmos a estatística t de \hat{u}_{t-1} ao valor crítico desejado na Tabela 18.4. Se a estatística t estiver abaixo do valor crítico, teremos evidência de que $y_t - \beta x_t$ é $I(0)$ para alguns β ; isto é, y_t e x_t são co-integrados. Podemos adicionar defasagens de $\Delta\hat{u}_t$ para avaliarmos a correlação serial. Se compararmos os valores críticos da Tabela 18.4 com os da Tabela 18.2, deveremos obter uma estatística t muito maior em magnitude para encontrar a co-integração do que se usássemos os valores críticos DF usuais. Isso ocorre porque o método MQO, que minimiza a soma dos resíduos quadrados, tende a produzir resíduos que se parecem com uma seqüência $I(0)$, mesmo se y_t e x_t forem não co-integrados.

Se y_t e x_t forem não co-integrados, uma regressão de y_t sobre x_t será espúria e não nos informará nada importante: não existe relação de longo prazo entre y e x . Podemos ainda computar uma regressão envolvendo as primeiras diferenças, Δy_t e Δx_t , incluindo defasagens. Entretanto, devemos interpretar essas regressões pelo que elas são: elas explicam a diferença em y em termos da diferença em x e não têm, necessariamente, nada a ver com a relação em níveis.

Se y_t e x_t forem co-integrados, poderemos usar isso para especificar modelos dinâmicos mais gerais, como veremos na próxima subseção.

A discussão anterior assume que nem y_t nem x_t têm uma tendência. Isso é razoável quando tratamos de taxas de juros, mas não para outras séries temporais. Se y_t e x_t contiverem termos de tendências, $E(y_t)$ e $E(x_t)$ serão funções lineares (normalmente crescentes) do tempo. A definição estrita de co-integração exige que $y_t - \beta x_t$ seja $I(0)$ *sem* uma tendência. Para ver o que isso acarreta, escreva $y_t = \delta t + g_t$ e $x_t = \lambda t + h_t$, onde $\{g_t\}$ e $\{h_t\}$ são processos $I(1)$, δ é a tendência em y_t [$\delta = E(\Delta y_t)$], e λ é a tendência em x_t [$\lambda = E(\Delta x_t)$]. Agora, se y_t e x_t forem co-integrados, deve existir β de forma que $g_t - \beta h_t$ seja $I(0)$. Mas, então,

$$y_t - \beta x_t = (\delta - \beta \lambda)t + (g_t - \beta h_t),$$

que geralmente é um processo *estacionário em torno de sua tendência*. A forma estrita de co-integração exige que não haja uma tendência, o que significa $\delta = \beta\lambda$. Em processos I(1) com tendência, é possível que as partes estocásticas — isto é, g_t e h_t — sejam co-integradas, mas que o parâmetro β que faz com que $g_t - \beta h_t$ seja I(0) não elimine a tendência temporal linear.

Podemos testar a co-integração entre g_t e h_t , sem fazer uma parada na parte da tendência, computando a regressão

$$\hat{y}_t = \hat{\alpha} + \hat{\eta}t + \hat{\beta}x_t \quad (18.32)$$

e aplicar o teste DF, usual ou aumentado, aos resíduos, \hat{u}_t . Os valores críticos assimptóticos estão na Tabela 18.5 [extraídos de Davidson e MacKinnon (1993, Tabela 20.2)].

Tabela 18.5

Valores Críticos Assimptóticos do Teste de Co-Integração: Tendência Temporal Linear

Nível de Significância	1%	2,5%	5%	10%
Valor Crítico	-4,32	-4,03	-3,78	-3,50

Encontrar co-integração nesse caso abre a possibilidade de que $y_t - \beta x_t$ tenha uma tendência linear. Mas pelo menos ela não será I(1).

EXEMPLO 18.5

(Co-Integração entre Fertilidade e Isenção de Impostos)

Nos capítulos 10 e 11, estudamos vários modelos para estimar a relação entre a taxa geral de fertilidade (tgf) e o valor real da isenção pessoal de impostos (ip) nos Estados Unidos. Os resultados da regressão estática em níveis e em primeiras diferenças são notavelmente diferentes. A regressão em níveis, com a inclusão de uma tendência temporal, produz um coeficiente MQO de ip igual a 0,187 ($ep = 0,035$) e $R^2 = 0,500$. Em primeiras diferenças (sem uma tendência), o coeficiente de Δip é -0,043 ($ep = 0,028$), e $R^2 = 0,032$. Embora haja outras razões para essas diferenças — como dinâmica de defasagens distribuídas mal-especificadas —, a discrepância entre as regressões em níveis e em alterações sugere que devemos testar a existência de co-integração. Naturalmente, isso presume que tgf e ip sejam processos I(1). Esse parece ser o caso: os testes DF aumentados, com uma única alteração defasada e uma tendência temporal linear, produzem ambos estatísticas t em torno de -1,47, e os coeficientes AR(1) estimados são próximos de um.

Quando obtemos os resíduos da regressão de tgf sobre t e ip e aplicarmos o teste DF aumentado com uma defasagem, obtemos uma estatística t de \hat{u}_{t-1} , de -2,43, que não está nem de longe próxima do valor crítico de 10%, -3,50. Portanto, devemos concluir que existe pouca evidência de co-integração entre tgf e ip , mesmo considerando tendências separadas. É muito provável que os resultados da regressão anterior que obtivemos em níveis sofra do problema de regressão espúria.

A boa notícia é que, quando usamos primeiras diferenças e permitimos duas defasagens — veja a equação (11.27) —, encontramos um efeito geral positivo e significante de longo prazo de Δip sobre Δtgf .

Se considerarmos que duas séries são co-integradas, muitas vezes queremos testar hipóteses sobre o parâmetro de co-integração. Por exemplo, uma teoria pode estabelecer que o parâmetro de co-integração é um. Idealmente, poderíamos usar uma estatística t para testar essa hipótese.

Tratamos explicitamente do caso sem tendências temporais, embora a extensão para o caso de tendência linear seja imediata. Quando y_t e x_t são I(1) e co-integrados, podemos escrever

$$y_t = \alpha + \beta x_t + u_t \quad (18.33)$$

onde u_t é um processo I(0) de média zero. Geralmente, $\{u_t\}$ contém correlação serial, mas sabemos, pelo Capítulo 11, que isso não afeta a consistência do MQO. Como mencionado anteriormente, o MQO aplicado a (18.33) estima consistentemente β (e α). Infelizmente, como x_t é I(1), os procedimentos habituais de inferência não serão necessariamente aplicáveis: o MQO não é assintoticamente normalmente distribuído, e a estatística t de $\hat{\beta}$ não tem, necessariamente, uma distribuição t aproximada. Sabemos do Capítulo 10 que, se $\{x_t\}$ for estritamente exógena — veja a hipótese ST.2 — e os erros forem homoscedásticos, serialmente não-correlacionados e normalmente distribuídos, o estimador MQO será também normalmente distribuído (condicional nas variáveis explicativas) e a estatística t terá uma distribuição t exata. Infelizmente, essas hipóteses são fortes demais para serem aplicadas na maioria das situações. A noção de co-integração não implica nada sobre a relação entre $\{x_t\}$ e $\{u_t\}$ — realmente, eles podem ser arbitrariamente correlacionados. Além disso, exceto pelo requisito de que $\{u_t\}$ seja I(0), a co-integração entre y_t e x_t não restringe a dependência serial em $\{u_t\}$.

Felizmente, o recurso de (18.33) que torna a inferência a mais difícil — a falta de exogeneidade estrita de $\{x_t\}$ — pode ser corrigido. Como x_t é I(1), a própria noção de exogeneidade estrita é que u_t é não-correlacionado com Δx_s , para qualquer t e s . Sempre podemos organizar isso para um novo conjunto de erros, pelo menos aproximadamente, definindo u_t como uma função de Δx_s para todo s próximo de t . Por exemplo,

$$\begin{aligned} u_t = & \eta + \phi_0 \Delta x_t + \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} \\ & + \gamma_1 \Delta x_{t+1} + \gamma_2 \Delta x_{t+2} + e_t, \end{aligned} \quad (18.34)$$

onde, por construção, e_t é não-correlacionado com cada Δx_s que aparece na equação. A esperança é que e_t seja não-correlacionado com outros atrasos e adiantamentos de Δx_s . Sabemos que, conforme $|s - t|$ cresce, a correlação entre e_t e Δx_s se aproxima de zero, pois eles são processos I(0). Agora, se inserirmos (18.34) em (18.33), obteremos

$$\begin{aligned} y_t = & \alpha_0 + \beta x_t + \phi_0 \Delta x_t + \phi_1 \Delta x_{t-1} + \phi_2 \Delta x_{t-2} \\ & + \gamma_1 \Delta x_{t+1} + \gamma_2 \Delta x_{t+2} + e_t. \end{aligned} \quad (18.35)$$

Essa equação parece um pouco estranha, pois os Δx_s futuros aparecem tanto com o Δx_t corrente como o defasado. O importante é que o coeficiente de x_t ainda é β , e, por construção, x_t agora é estritamente exógeno na equação. A hipótese de exogeneidade estrita é a importante condição necessária para obter uma estatística t aproximadamente normal para $\hat{\beta}$. Se u_t for não-correlacionado com todos os Δx_s , $s \neq t$, então, poderemos eliminar os adiantamentos e atrasos das mudanças e simplesmente incluir a alteração contemporânea, Δx_t . Então, a equação que estimamos parece mais padronizada, mas ainda inclui a primeira diferença de x_t , juntamente com seu nível: $y_t = \alpha_0 + \beta x_t + \phi_0 \Delta x_t + e_t$. Com efeito, a adição de Δx_t resolve qualquer endogeneidade contemporânea entre x_t e u_t . (Lembre-se de que qualquer endogeneidade não causa inconsistência. Mas estamos tentando obter uma estatística t assintoticamente normal).

Se necessitamos incluir adiantamentos e atrasos das alterações, e quantos, é realmente um problema empírico. Cada vez que acrescentamos um adiantamento ou atraso adicional, perdemos uma observação, e isso pode sair caro, a menos que tenhamos um grande conjunto de dados.

O estimador MQO de β , a partir de (18.35), é chamado o **estimador de adiantamentos e atrasos de β** devido à maneira como ele emprega Δx . [Veja, por exemplo, Stock e Watson (1993).] O único problema com o qual devemos nos preocupar em (18.35) é a possibilidade de correlação serial em $\{e_t\}$. Isso pode ser resolvido computando-se um erro-padrão robusto em relação à correlação serial para $\hat{\beta}$ (como descrito na Seção 12.5) ou usando-se uma correção AR(1) padrão (como a de Cochrane-Orcutt).

EXEMPLO 18.6

(Parâmetro de Co-Integração das Taxas de Juros)

Anteriormente, testamos a existência de co-integração entre $r6$ e $r3$ — taxas das letras do Tesouro norte-americano de seis e três meses —, assumindo que o parâmetro de co-integração era igual a um. Isso nos levou a encontrar co-integração e, naturalmente, a concluir que o parâmetro de co-integração é igual à unidade. Contudo, vamos estimar o parâmetro de co-integração diretamente e testar $H_0: \beta = 1$. Aplicamos o estimador de adiantamentos e atrasos com dois adiantamentos e dois atrasos de $\Delta r3$, como também a alteração contemporânea. A estimativa de β é $\hat{\beta} = 1,038$, e o erro-padrão MQO habitual é 0,0081. Portanto, a estatística t para $H_0: \beta = 1$ é $(1,038 - 1)/0,0081 \approx 4,69$, que indica uma forte rejeição estatística de H_0 . (Naturalmente, saber se 1,038 é economicamente diferente de um é uma consideração relevante.) Existe pouca evidência de correlação serial nos resíduos, e assim podemos usar essa estatística t como se ela tivesse uma distribuição normal aproximada. (A título de comparação, a estimativa MQO de β sem os termos de adiantamentos, atrasos ou o contemporâneo de $\Delta r3$ — e usando cinco ou mais observações — é 1,026 ($ep = 0,0077$). Mas a estatística t de (18.33) não será necessariamente válida.)

Existem muitos outros estimadores de parâmetros de co-integração, e essa continua sendo uma área bastante ativa de pesquisa. A noção de co-integração se aplica a mais de dois processos, mas os procedimentos de interpretação, teste e estimação são muito mais complicados. Um problema é que, mesmo depois de termos normalizado um coeficiente para que ele seja igual a um, pode haver muitas relações co-integradas. BDGH oferece algumas explicações e várias referências.

Modelos de Correção de Erro

Além de informar sobre uma potencial relação de longo prazo entre duas séries, o conceito de co-integração enriquece os tipos de modelos dinâmicos à nossa disposição. Se y_t e x_t forem processos I(1) e não forem co-integrados, poderemos estimar um modelo dinâmico em primeiras diferenças. Como um exemplo, considere a equação

$$\Delta y_t = \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + u_t, \quad (18.36)$$

onde u_t tem média zero, dados Δx_t , Δy_{t-1} , Δx_{t-1} e defasagens adicionais. Essencialmente, isso é a equação (18.16), mas em primeiras diferenças, em vez de em níveis. Se entendermos isso como um modelo de defasagem distribuída racional, poderemos encontrar a propensão de impacto, a propensão de longo prazo e a distribuição de defasagens de Δy como uma defasagem distribuída em Δx .

Se y_t e x_t forem co-integrados com parâmetro β , então, há variáveis I(0) adicionais que podemos incluir em (18.36). Defina $s_t = y_t - \beta x_t$, de forma que s_t seja I(0), e assuma, para simplificar, que s_t tem média zero. Agora, podemos incluir defasagens da s_t na equação. No caso mais simples, incluímos uma defasagem de s_t :

$$\begin{aligned}\Delta y_t &= \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \delta s_{t-1} + u_t, \\ &= \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_0 \Delta x_t + \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + u_t,\end{aligned}\quad (18.37)$$

onde $E(u_t | I_{t-1}) = 0$ e I_{t-1} contém informações sobre Δx_t e todos os valores passados de x e y . O termo $\delta(y_{t-1} - \beta x_{t-1})$ é chamado de *termo de correção de erro* e (18.37) é um exemplo de um **modelo de correção de erro**. (Em alguns modelos de correção de erro, a alteração contemporânea em x_t , Δx_t , é omitida. Se ela será incluída ou não, vai depender parcialmente do propósito da equação. Em previsão, Δx_t raramente é incluída, pelos motivos que veremos na Seção 18.5.)

Um modelo de correção de erro nos possibilita estudar a dinâmica de curto prazo na relação entre y e x . Para simplificar, considere o modelo sem defasagens de Δy_t e Δx_t :

$$\Delta y_t = \alpha_0 + \gamma_0 \Delta x_t + \delta(y_{t-1} - \beta x_{t-1}) + u_t, \quad (18.38)$$

onde $\delta < 0$. Se $y_{t-1} > \beta x_{t-1}$, então, y no momento anterior ultrapassou o equilíbrio; como $\delta < 0$, o termo de correção de erro empurra y para trás, em direção ao equilíbrio. De forma semelhante, se $y_{t-1} < \beta x_{t-1}$, o termo de correção de erro induzirá uma alteração positiva em y para trás, em direção ao equilíbrio.

Como estimamos os parâmetros de um modelo de correção de erro? Se conhecermos β , isso será fácil. Por exemplo, em (18.38), simplesmente fazemos a regressão de Δy_t sobre Δx_t e s_{t-1} , onde $s_{t-1} = (y_{t-1} - \beta x_{t-1})$.

EXEMPLO 18.7

{Modelo de Correção de Erro dos Rendimentos de Títulos}

No Problema 11.6, regredimos $hy6_t$, o rendimento de três meses (em percentagem) da compra de letras do Tesouro de seis meses no momento $t - 1$ e venda das mesmas no momento t como se fossem letras do Tesouro de três meses, sobre $hy3_{t-1}$, o rendimento de três meses da compra de letras do Tesouro de três meses no momento $t - 1$. A hipótese das expectativas sugere que o coeficiente de inclinação não deve ser estatisticamente diferente de um. Constatata-se que existe evidência de uma raiz unitária em $\{hy3_t\}$, o que levanta a questão da análise de regressão padrão. Assumiremos que os dois rendimentos são processos I(1).

A hipótese das expectativas implica que, no mínimo, $hy6_t$ e $hy3_{t-1}$ são co-integrados com β igual a um, que parece ser o caso (veja o Exercício 18.14). Sob essa hipótese, um modelo de correção de erro será

$$\Delta hy6_t = \alpha_0 + \gamma_0 hy3_{t-1} + \delta(hy6_{t-1} - hy3_{t-2}) + u_t,$$

onde u_t tem média zero, dados todos os $hy3$ e $hy6$ datados de $t - 1$ e de momentos anteriores. As defasagens nas variáveis no modelo de correção de erro são ditadas pelas hipóteses das expectativas.

EXEMPLO 18.7 (continuação)

Utilizando os dados contidos no arquivo INTQRT.RAW, teremos

$$\begin{aligned} \Delta \hat{h}_t^6 &= 0,090 + 1,218 \Delta h_3, -1 - 0,840 (h_6, -1 - h_3, -2) \\ &\quad (0,043) \quad (0,264) \quad (0,244) \end{aligned} \tag{18.39}$$

$n = 122, R^2 = 0,790.$

O coeficiente de correção de erro é negativo e muito significante. Por exemplo, se o rendimento das letras do Tesouro de seis meses for superior ao de três meses em um ponto percentual, h_6 cairá em 0,84 pontos, em média, no trimestre seguinte. Curiosamente, $\hat{\delta} = -0,84$ não é estatisticamente diferente de -1 , como pode ser verificado computando-se o intervalo de confiança de 95%.

Como você testaria $H_0: \gamma_0 = 1, \delta = -1$ no modelo de correção de erro dos rendimentos de títulos?

Em muitos outros exemplos, o parâmetro de co-integração deve ser estimado. Depois, substituímos s_{t-1} por $\hat{s}_{t-1} = y_{t-1} - \hat{\beta}x_{t-1}$, onde $\hat{\beta}$ pode ser vários estimadores de β . Já tratamos do estimador MQO padrão como também do estimador de adiantamentos e atrasos. Isso levanta o problema de como a variação amostral em $\hat{\beta}$ afeta a inferência sobre os outros parâmetros no modelo de correção de erro. Felizmente, como mostrado por Engle e Granger (1987), podemos ignorar a estimativa preliminar de β (assimptoticamente). Essa propriedade é muito conveniente e implica que a eficiência assintótica dos estimadores dos parâmetros no modelo de correção de erro não será afetada quer usemos o estimador MQO ou o estimador de adiantamentos e atrasos para $\hat{\beta}$. Naturalmente, a escolha de $\hat{\beta}$ terá, geralmente, um efeito sobre os parâmetros de correção de erro estimados em qualquer amostra, mas não temos nenhuma maneira sistemática de decidir qual estimador preliminar de β usar. O procedimento de substituição de β por $\hat{\beta}$ é chamado **procedimento de dois estágios de Engle-Granger**.

18.5 PREVISÃO

A previsão de séries temporais econômicas é muito importante em certos ramos da economia, e é uma área que continua sendo ativamente estudada. Nesta seção, concentraremos-nos em métodos de previsões baseados em regressão. Diebold (2001) oferece uma introdução abrangente da previsão, inclusive dos recentes desenvolvimentos.

Assumimos nesta seção que o foco principal é prever valores futuros de um processo de séries temporais e não necessariamente estimar modelos econômicos causais ou estruturais.

É interessante primeiro entender alguns fundamentos da previsão que não dependem de um modelo específico. Suponha que no momento t queremos estimar o resultado de y no momento $t+1$, ou y_{t+1} . O período de tempo pode corresponder a um ano, um trimestre, um mês, uma semana, ou mesmo um dia. Seja I_t a informação que podemos observar no momento t . Esse **conjunto de informações** inclui y_t , valores anteriores de y e freqüentemente outras variáveis datadas no momento t ou de momentos

anteriores. Podemos combinar essas informações em inumeráveis maneiras para fazermos a previsão de y_{t+1} . Existirá alguma maneira melhor?

A resposta é sim, desde que especifiquemos a *perda* associada com o erro de previsão. Seja p_t a previsão de y_{t+1} feita no momento t . Chamamos p_t previsão com um passo à frente. O erro de previsão é $e_{t+1} = y_{t+1} - p_t$, que observamos logo que o resultado de y_{t+1} seja observado. A medida mais comum de perda é a mesma que conduz a uma estimativa de mínimos quadrados ordinários de um modelo de regressão linear múltipla: o erro quadrático, e_{t+1}^2 . O erro de previsão quadrático trata de erros de previsão positivos e negativos simetricamente, e erros de previsão grandes recebem relativamente mais peso. Por exemplo, erros de +2 e -2 produzem a mesma perda, e a perda é quatro vezes maior que os erros de previsão de +1 ou -1. O erro de previsão quadrático é um exemplo de uma função da perda. Outra função da perda popular é o valor absoluto do erro de previsão, $|e_{t+1}|$. Por razões que veremos em breve, nos concentraremos por ora na perda do erro quadrático.

Dada a função da perda do erro quadrático, podemos determinar como usar melhor a informação do momento t para fazer a previsão de y_{t+1} . Mas teremos que reconhecer que no momento t , não conhecemos e_{t+1} : ele é uma variável aleatória, pois y_{t+1} é uma variável aleatória. Portanto, qualquer critério útil para a seleção de p_t deve ser baseado no que conhecemos no momento t . É natural preferir a previsão que minimize o erro de previsão quadrático esperado, dado I_t :

$$E(e_{t+1}^2 | I_t) = E[(y_{t+1} - p_t)^2 | I_t]. \quad (18.40)$$

Um fato básico da probabilidade (veja propriedade EC.6 no Apêndice B, disponível na página do livro, no site www.thomsonlearning.com.br) é que a expectativa condicional, $E(y_{t+1} | I_t)$, minimiza (18.40). Em outras palavras, se quisermos minimizar o erro de previsão quadrático esperado dada a informação no momento t , nossa previsão deverá ser o valor esperado de y_{t+1} , dadas as variáveis que conhecemos no momento t .

Em muitos processos populares de séries temporais, é fácil obter a expectativa condicional. Suponha que $\{y_t; t = 0, 1, \dots\}$ seja uma seqüência da diferença de um *martingale* (SDM) e I_t seja $\{y_0, y_1, \dots, y_t\}$, o passado observado de y . Por definição, $E(y_{t+1} | I_t) = 0$ para todo t ; a melhor previsão de y_{t+1} no momento t é sempre zero! Lembre-se da Seção 18.2 que uma seqüência i.i.d. com média zero é uma seqüência da diferença de um *martingale*.

Uma seqüência da diferença de um *martingale* é aquela na qual o passado não é útil para prevermos o futuro. Os retornos de ações são largamente aceitos como uma aproximação de uma SDM ou, talvez, com uma média positiva. O importante é que $E(y_{t+1} | y_t, y_{t-1}, \dots) = E(y_{t+1})$: a média condicional é igual à média incondicional, caso em que o y passado não auxilia na previsão do y futuro.

Um processo $\{y_t\}$ será um *martingale* se $E(y_{t+1} | y_t, y_{t-1}, \dots, y_0) = y_t$ para todo $t \geq 0$. [Se $\{y_t\}$ for um *martingale*, $\{\Delta y_t\}$ será uma seqüência da diferença de um *martingale*, que é de onde vem o nome desta última.] O valor previsto de y no próximo período será sempre o valor de y neste período.

Um exemplo mais complicado é

$$E(y_{t+1} | I_t) = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \dots + \alpha(1 - \alpha)y_0, \quad (18.41)$$

onde $0 < \alpha < 1$ é um parâmetro que devemos escolher. Esse método de previsão é chamado suavização exponencial porque os pesos dos y defasados declinam para zero exponencialmente.

A razão para escrevermos a esperança na forma de (18.41) é que ela conduz a uma relação de recorrência muito simples. Defina $p_0 = y_0$. Em seguida, para $t \geq 1$, as previsões poderão ser obtidas como

$$p_t = \alpha y_t + (1 - \alpha) p_{t-1}.$$

Em outras palavras, a previsão de y_{t+1} é uma média ponderada de y_t , e da previsão de y_t , feita no momento $t - 1$. A suavização exponencial é apropriada somente para séries temporais muito específicas e exige que se escolha α . Métodos de regressão, que abordaremos a seguir, são mais flexíveis.

A discussão anterior concentrou-se na previsão de y somente um período adiante. Os problemas gerais que surgem na previsão de y_{t+h} no momento t , no qual h é qualquer inteiro positivo, são semelhantes. Em particular, se usarmos o erro de previsão quadrático esperado como nossa medida de perda, o melhor previsor será $E(y_{t+h}|I_t)$. Quando estamos lidando com uma **previsão com múltiplos passos à frente**, usamos a notação $p_{t,h}$ para indicar a previsão de y_{t+h} feita no momento t .

Tipos de Modelos de Regressão Utilizados na Previsão

Existem muitos modelos de regressão diferentes que podemos usar para fazer a previsão de valores futuros de uma série temporal. O primeiro modelo de regressão de dados de séries temporais do Capítulo 10 foi o modelo estático. Para ver como podemos fazer previsão com esse modelo, assuma que temos uma única variável explicativa:

$$y_t = \beta_0 + \beta_1 z_t + u_t. \quad (18.42)$$

Suponha, por enquanto, que os parâmetros β_0 e β_1 sejam conhecidos. Escreva essa equação no momento $t + 1$ como $y_{t+1} = \beta_0 + \beta_1 z_{t+1} + u_{t+1}$. Agora, se z_{t+1} for conhecida no momento t , de forma que ele seja um elemento de I_t , e $E(u_{t+1}|I_t) = 0$, então,

$$E(y_{t+1}|I_t) = \beta_0 + \beta_1 z_{t+1},$$

onde I_t contém $z_{t+1}, y_t, z_t, \dots, y_1, z_1$. O lado direito dessa equação é a previsão de y_{t+1} no momento t . Esse tipo de previsão é habitualmente chamado de **previsão condicional** porque ele é condicional a conhecer o valor de z no momento $t + 1$.

Infelizmente, em qualquer momento, raramente conhecemos o valor das variáveis explicativas em períodos de tempo futuros. As exceções são as tendências temporais e as variáveis *dummy* sazonais, que estudaremos explicitamente mais adiante, mas fora isso o conhecimento de z_{t+1} no momento t é raro. Algumas vezes, queremos gerar previsões condicionais de vários valores de z_{t+1} .

Outro problema com (18.42) como um modelo de previsão é que $E(u_{t+1}|I_t) = 0$ significa que $\{u_t\}$ não pode conter correlação serial, algo que vimos ser falso na maioria dos modelos de regressão estatística. [O Problema 18.8 pede que você derive a previsão em um modelo simples de defasagem distribuída com erros AR(1).]

Se z_{t+1} não for conhecido no momento t , não poderemos incluí-lo em I_t . Então, teremos

$$E(y_{t+1}|I_t) = \beta_0 + \beta_1 E(z_{t+1}|I_t).$$

Isso significa que, para fazermos a previsão de y_{t+1} , precisamos primeiro fazer a previsão de z_{t+1} , com base no mesmo conjunto de informações. Isso é normalmente chamado de **previsão incondicional**, porque não assumimos o conhecimento de z_{t+1} no momento t . Infelizmente, essa é uma denominação imprópria, já que nossa previsão ainda é condicional à informação em I_t . Mas o nome está arraigado na literatura sobre previsão.

Para fazer previsões, a menos que estejamos ligados ao modelo estático de (18.42) por outras razões, faz mais sentido especificar um modelo que dependa somente de valores defasados de y e z . Isso nos poupará uma etapa extra de ter de fazer a previsão de uma variável do lado direito da equação antes de fazer a previsão de y . O tipo de modelo que temos em mente é

$$\begin{aligned} y_t &= \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + u_t \\ E(u_t | I_{t-1}) &= 0, \end{aligned} \quad (18.43)$$

onde I_{t-1} contém y e z datadas no momento $t-1$ e em momentos anteriores. Agora, a previsão de y_{t+1} no momento t é $\hat{y}_t = \delta_0 + \alpha_1 y_t + \gamma_1 z_t$; se conhecermos os parâmetros, podemos simplesmente utilizar os valores de y_t e z_t .

Se apenas quisermos usar o y passado para prever o y futuro, então, poderemos eliminar z_{t-1} de (18.43). Naturalmente, podemos adicionar mais defasagens de y ou z e defasagens de outras variáveis. Especialmente para a previsão um passo à frente, tais modelos podem ser muito úteis.

Previsão um Passo à Frente

Obter uma previsão um período após o término da amostra é relativamente simples com o uso de modelos como (18.43). Como sempre, definimos n como o tamanho da amostra. A previsão de y_{n+1} será

$$\hat{p}_n = \hat{\delta}_0 + \hat{\alpha}_1 y_n + \hat{\gamma}_1 z_n, \quad (18.44)$$

onde assumimos que os parâmetros tenham sido estimados por MQO. Usamos um sinal de acento círcunflexo em p_n para enfatizar que os parâmetros foram estimados no modelo de regressão. (Se conhecêssemos os parâmetros, não haveria erros de estimação na previsão.) O erro de previsão — que não conheceremos até o momento $n+1$ — será

$$\hat{e}_{n+1} = y_{n+1} - \hat{p}_n. \quad (18.45)$$

Se adicionarmos mais defasagens de y ou z à equação de previsão, simplesmente perderemos mais observações no início da amostra.

A previsão \hat{p}_n de y_{n+1} é normalmente chamada de **previsão por ponto**. Também podemos obter um **intervalo de previsão**. Um intervalo de previsão é essencialmente o mesmo que aquele que estudamos na Seção 6.4. Lá, mostramos como, sob as hipóteses do modelo linear clássico, obter um intervalo de previsão exato de 95%. Aqui, um intervalo de previsão é obtido exatamente da mesma maneira. Se o modelo não satisfizer as hipóteses do modelo linear clássico — por exemplo, se ele contiver variáveis dependentes defasadas, como em (18.44) —, o intervalo de previsão ainda será aproximadamente válido, desde que u_t , dado I_{t-1} , seja normalmente distribuído com média zero e variância constante. (Isso garante que os estimadores MQO serão aproximadamente normalmente distribuídos com as variâncias habituais do MQO e que u_{n+1} será independente dos estimadores MQO com média zero e variância σ^2 .) Seja $ep(\hat{p}_n)$ o erro-padrão da previsão e $\hat{\sigma}$ o erro-padrão da regressão. [Como vimos na Seção 6.4, podemos obter \hat{p}_n e $ep(\hat{p}_n)$ como o intercepto e seu erro-padrão a partir da regressão de y_t sobre $(y_{t-1} - y_n)$ e $(z_{t-1} - z_n)$, $t = 1, 2, \dots, n$; ou seja, subtraímos o valor do momento n de y e de cada y defasado, e de forma semelhante de z , antes de fazermos a regressão.] Então,

$$\text{ep}(\hat{e}_{n+1}) = \{[\text{ep}(\hat{p}_n)]^2 + \hat{\sigma}^2\}^{1/2}, \quad (18.46)$$

e o intervalo de previsão (aproximado) de 95% será

$$\hat{p}_n \pm 1,96 \text{ ep}(\hat{e}_{n+1}). \quad (18.47)$$

Como $\text{ep}(\hat{p}_n)$ é aproximadamente proporcional a $1/\sqrt{n}$, $\text{ep}(\hat{p}_n)$ será em geral pequeno relativo à incerteza no erro u_{n+1} , como medido por $\hat{\sigma}$. [Alguns programas econômétricos computam intervalos de previsão rotineiramente, mas outros requerem algumas manipulações simples para obter (18.47).]

EXEMPLO 18.8

(Previsão da Taxa de Desemprego Norte-Americana)

Utilizamos os dados contidos no arquivo PHILLIPS.RAW, correspondente ao período 1948 a 1996, para fazermos a previsão da taxa de desemprego civil norte-americana para 1997. Usamos dois modelos. O primeiro é um modelo AR(1) simples de *desemp*:

$$\begin{aligned} \hat{de\hat{s}emp}_t &= 1,572 + 0,732 \hat{desemp}_{t-1} \\ &\quad (0,577) \quad (0,097) \\ n &= 48, \bar{R}^2 = 0,544, \hat{\sigma} = 1,049. \end{aligned} \quad (18.48)$$

Em um segundo modelo, adicionamos a inflação com uma defasagem de um ano:

$$\begin{aligned} \hat{de\hat{s}emp}_t &= 1,304 + 0,647 \hat{desemp}_{t-1} + 0,184 \hat{inf}_{t-1} \\ &\quad (0,490) \quad (0,084) \quad (0,041) \\ n &= 48, \bar{R}^2 = 0,677, \hat{\sigma} = 0,883. \end{aligned} \quad (18.49)$$

A taxa de inflação defasada é bastante significante em (18.49) ($t \approx 4,5$), e o R -quadrado ajustado da segunda equação é muito mais alto do que o da primeira. No entanto, isso não significa, necessariamente, que a segunda equação produzirá uma previsão melhor para 1997. Tudo que podemos dizer até agora é que, usando os dados até 1996, uma defasagem da inflação ajuda a explicar a variação na taxa de desemprego.

Para obter a previsão para 1997, precisamos conhecer *desemp* e *inf* de 1996. Elas foram 5,4 e 3,0, respectivamente. Portanto, a previsão de \hat{desemp}_{1997} da equação (18.48) é $1,572 + 0,732(5,4)$, ou cerca de 5,52. A previsão da equação (18.49) é $1,304 + 0,647(5,4) + 0,184(3,0)$, ou cerca de 5,35. A taxa real de desemprego civil em 1997 foi de 4,9, e assim ambas as equações superestimaram a taxa real. A segunda equação produziu uma previsão um pouco melhor.

Podemos facilmente obter um intervalo de previsão de 95%. Quando fazemos a regressão de \hat{desemp}_t sobre $(\hat{desemp}_{t-1} - 5,4)$ e $(\hat{inf}_{t-1} - 3,0)$, obtemos 5,35 como o intercepto — que já havíamos computado como a previsão — e $\text{ep}(\hat{p}_n) = 0,137$. Portanto, como $\hat{\sigma} = 0,883$, temos $\text{ep}(\hat{e}_{n+1}) = \{[0,137]^2 + (0,883)^2\}^{1/2} \approx 0,894$. O intervalo de previsão de 95% de (18.47) é $5,35 \pm 1,96(0,894)$, ou aproximadamente $[3,6; 7,1]$. Esse é um intervalo bem amplo e o valor efetivo de 1997, 4,9, está bem dentro do intervalo. Como esperado, o erro-padrão de u_{n+1} , que é 0,883, é uma fração muito grande de $\text{ep}(\hat{e}_{n+1})$.

Um profissional que faça previsões deve, normalmente, fazer uma previsão a cada período de tempo. Por exemplo, no momento n , ele faz a previsão de y_{n+1} . Depois, quando y_{n+1} e z_{n+1} se tornarem disponíveis, ele deve fazer a previsão de y_{n+2} . Mesmo que a pessoa que esteja fazendo a previsão tenha se decidido pelo modelo (18.43), haverá duas possibilidades para fazer a previsão de y_{n+2} . A primeira é usar $\hat{\delta}_0 + \hat{\alpha}_1 y_{n+1} + \hat{\gamma}_1 z_{n+1}$, onde os parâmetros são estimados utilizando as primeiras n observações. A segunda possibilidade é *reestimar* os parâmetros utilizando todas as $n+1$ observações e em seguida usar a mesma fórmula para fazer a previsão de y_{n+2} . Para fazer a previsão de períodos de tempo subsequentes, podemos utilizar, de forma geral, as estimativas dos parâmetros obtidos das n observações iniciais, ou podemos atualizar os parâmetros da regressão cada vez que obtivermos um novo ponto de dados. Embora o último método exija mais computação, o trabalho extra é relativamente pequeno, e pode (embora não necessariamente) funcionar melhor, pois os coeficientes da regressão se ajustam, pelo menos um pouco, aos novos pontos de dados.

Como um exemplo específico, suponha que queremos fazer a previsão da taxa de desemprego para 1998, usando o modelo com uma única defasagem de *desemp* e de *inf*. A primeira possibilidade é simplesmente inserir os valores do desemprego e da inflação de 1997 no lado direito da equação (18.49). Com $desemp = 4,9$ e $inf = 2,3$ em 1997, teremos uma previsão de $desemp_{1998}$ de aproximadamente 4,9. (É apenas uma coincidência esse valor ser o mesmo da taxa de desemprego de 1997). A segunda possibilidade é recalcular a equação adicionando as observações de 1997 e depois usar essa nova equação (veja o Exercício 18.15).

O modelo na equação (18.43) é uma equação do que é conhecido como **modelo auto-regressivo vetorial (VAR)**. Sabemos o que é um modelo auto-regressivo do Capítulo 11: modelamos uma única série, $\{y_t\}$, em termos de seu próprio passado. Em modelos auto-regressivos vetoriais, modelamos várias séries — que, se você estiver familiarizado com álgebra linear, é de onde advém o termo “vetor” — em termos de seus próprios passados. Se tivermos duas séries, y_t e z_t , uma auto-regressão vetorial consistirá de equações parecidas com

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + \alpha_2 y_{t-2} + \gamma_2 z_{t-2} + \dots \quad (18.50)$$

e

$$z_t = \eta_0 + \beta_1 y_{t-1} + \rho_1 z_{t-1} + \beta_2 y_{t-2} + \rho_2 z_{t-2} + \dots,$$

onde cada equação contém um erro com valor esperado zero, dadas as informações passadas de y e z . Na equação (18.43) — e no exemplo estimado em (18.49) —, assumimos que uma defasagem de cada variável capturou toda a dinâmica. (Um teste F da significância conjunta de $desemp_{t-2}$ e inf_{t-2} confirmará que é necessária somente uma defasagem de cada variável.)

Como o Exemplo 18.8 ilustra, modelos VAR podem ser úteis para previsões. Em muitos casos, estamos interessados em fazer a previsão de somente uma variável, y , em cujo caso somente precisamos estimar e analisar a equação de y . Nada nos impede de adicionar outras variáveis defasadas, digamos, w_{t-1} , w_{t-2} , ..., à equação (18.50). Tais equações são eficientemente estimadas por MQO, desde que tenhamos incluído defasagens suficientes de todas as variáveis e a equação satisfaça a hipótese de homoscedasticidade das regressões de séries temporais.

Equações como (18.50) nos possibilitam testar se, após termos controlado os y passados, o passado de z ajudará a fazer a previsão de y_t . De forma geral, dizemos que z *Granger-causa* y se

$$E(y_t | I_{t-1}) \neq E(y_t | J_{t-1}), \quad (18.51)$$

onde I_{t-1} contém informações passadas sobre y e z , e J_{t-1} contém apenas informações passadas de y . Quando (18.51) se mantém, os z passados são úteis, *adicionalmente aos y passados*, para fazermos a previsão de y_t . O termo “causa” na expressão “Granger-causa” deve ser interpretado com cuidado. O único sentido em que z “causa” é dado em (18.51). Em particular, ele não tem nada a dizer sobre causalidade contemporânea entre y e z e, portanto, não nos permite determinar se z_t é uma variável exógena ou endógena em uma equação relacionando y , com z_t . (Essa também é a razão de a noção de **causalidade de Granger** não se aplicar em contextos de cortes transversais puros.)

Assim que assumirmos um modelo linear e decidirmos quantas defasagens de y devem ser incluídas em $E(y_t | y_{t-1}, y_{t-2}, \dots)$, podemos facilmente testar a hipótese nula de que z não Granger-causa y . Para sermos mais específicos, suponha que $E(y_t | y_{t-1}, y_{t-2}, \dots)$ dependa de apenas três defasagens:

$$\begin{aligned} y_t &= \delta_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + u_t, \\ E(u_t | y_{t-1}, y_{t-2}, \dots) &= 0. \end{aligned}$$

Agora, sob a hipótese nula de que z não Granger-causa y , *quaisquer* defasagens de z que adicionarmos à equação deve ter zero coeficientes populacionais. Se adicionarmos z_{t-1} , então, poderemos fazer um teste t simples de z_{t-1} . Se adicionarmos duas defasagens de z , poderemos fazer um teste F da significância conjunta de z_{t-1} e z_{t-2} na equação

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 y_{t-3} + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + u_t.$$

(Se houver heteroscedasticidade, poderemos usar uma forma robusta do teste. Não pode haver correlação serial sob H_0 , pois o modelo é dinamicamente completo.)

Do ponto de vista prático, como decidimos quantas defasagens de y e z devemos incluir? Primeiro, começamos estimando um modelo auto-regressivo de y e fazendo testes t e F para determinar quantas defasagens de y devem aparecer. Com dados anuais, o número de defasagens é em geral pequeno, digamos, uma ou duas. Com dados trimestrais ou mensais, normalmente existem muito mais defasagens. Assim que um modelo auto-regressivo de y tiver sido escolhido, poderemos testar as defasagens de z . A escolha das defasagens de z é menos importante, pois quando z não Granger-causa y , nenhum conjunto de z 's defasados deve ser significante. Com dados anuais, uma ou duas defasagens são geralmente usadas; com dados trimestrais, usualmente quatro ou oito; com dados mensais, talvez seis, 12 ou talvez até mesmo 24, desde que haja dados suficientes.

Já demos um exemplo do teste de causalidade de Granger na equação (18.49). O modelo auto-regressivo que melhor se adapta ao desemprego é um AR(1). Na equação (18.49), adicionamos uma única defasagem da inflação, e ela era muito significante. Portanto, a inflação Granger-causa o desemprego.

Existe uma definição ampliada da causalidade de Granger que muitas vezes é útil. Seja $\{w_t\}$ uma terceira série (ou ela poderia representar várias séries adicionais). Então, z Granger-causa y condicional em w se (18.45) se mantiver, mas agora I_{t-1} contém informações passadas sobre y , z e w , enquanto J_{t-1} contém informações passadas sobre y e w . Certamente, é possível que z Granger-cause y , mas z não Granger-causa y condicional em w . Um teste da hipótese nula de que z não Granger-causa y condicional em w é obtido testando a significância de z defasados em um modelo de y que também dependa de y e w defasados. Por exemplo, para verificar se o crescimento da oferta monetária Granger-causa

crescimento do PIB real, condicional às alterações nas taxas de juros, faríamos a regressão de $gPIB$, sobre defasagens de $gPIB$, Δ_{jur} e gM e faríamos testes de significância das defasagens de gM . [Veja, por exemplo, Stock e Watson (1989).]

A Comparação de Previsões um Passo à Frente

Em quase todos os problemas de previsão, existem vários métodos competitivos de previsão. Mesmo quando nos restringimos a modelos de regressão, existem muitas possibilidades. Quais variáveis devem ser incluídas e com quantas defasagens? Devemos usar logs, níveis ou primeiras diferenças das variáveis?

Para decidir sobre um método de previsão, precisamos de um meio para saber qual o mais apropriado. De forma ampla, podemos distinguir entre **critérios dentro da amostra** e **critérios fora da amostra**. Em um contexto de regressão, os critérios dentro da amostra incluem o R -quadrado e o R -quadrado especialmente ajustado. Existem muitas outras *estatísticas de seleção de modelos*, mas não trataremos delas aqui [veja, por exemplo, Ramanathan (1995, Capítulo 4)].

Para fazer previsões, é melhor usarmos os critérios fora da amostra, pois previsão é essencialmente um problema que está fora da amostra. Um modelo pode fornecer um bom ajuste para y na amostra usada para estimar os parâmetros. Entretanto, isso não se traduzirá, necessariamente, em um bom desempenho da previsão. Uma comparação fora da amostra envolve o uso da primeira parte de uma amostra para estimar os parâmetros do modelo, reservando a última parte da amostra para avaliar sua capacidade de previsão. Isso simula o que teríamos que fazer na prática se ainda não conhecêssemos os valores futuros das variáveis.

Suponha que tenhamos $n + m$ observações, nas quais usamos as primeiras n observações para estimar os parâmetros em nosso modelo e reservamos as últimas m observações para fazer a previsão. Seja \hat{p}_{n+h} a previsão um passo à frente de y_{n+h+1} para $h = 0, 1, \dots, m-1$. O m erros de previsão são $\hat{e}_{n+h+1} = y_{n+h+1} - \hat{p}_{n+h}$. Como devemos avaliar o quanto nosso modelo faz bem a previsão de y quando ele estiver fora da amostra? Duas medidas são as mais comuns. A primeira é a raiz do erro quadrático médio (REQM):

$$REQM = \left(m^{-1} \sum_{h=0}^{m-1} \hat{e}_{n+h+1}^2 \right)^{1/2}. \quad (18.52)$$

Isso é essencialmente o desvio-padrão amostral dos erros de previsão (sem qualquer ajuste dos graus de liberdade). Se computarmos o REQM de dois ou mais métodos de previsão, preferiremos o método com o menor REQM fora da amostra.

Uma segunda medida comum é o erro absoluto médio (EAM), que é a média dos erros absolutos da previsão:

$$EAM = m^{-1} \sum_{h=0}^{m-1} |\hat{e}_{n+h+1}|. \quad (18.53)$$

Novamente, preferiremos o menor EAM. Outros possíveis critérios incluem a minimização do maior dos valores absolutos dos erros de previsão.

EXEMPLO 18.9**(Comparações Fora da Amostra de Previsões do Desemprego)**

No Exemplo 18.8, verificamos que a equação (18.49) se ajustava melhor em nossa amostra que a equação (18.48) e, pelo menos para fazer a previsão de 1997, o modelo com inflação defasada funcionava melhor. Agora, estimamos ambos os modelos utilizando os dados até 1989, reservando 1990 a 1996 para comparações fora da amostra. Isso deixa sete observações fora da amostra ($n = 41$ e $m = 7$, para sermos precisos). Para o modelo AR(1), REQM = 0,632 e EAM = 0,515. Para o modelo que adiciona inflação defasada, REQM = 0,550 e EAM = 0,362. Assim, por qualquer indicador, o modelo que inclui \inf_{t-1} produz melhores previsões fora da amostra para os anos 1990. Nesse caso, tanto o critério dentro da amostra como o fora da amostra usam o mesmo modelo.

Em vez de usar somente as primeiras n observações para estimar os parâmetros do modelo, podemos reestimar os modelos cada vez que adicionarmos uma nova observação e usar o novo modelo para fazer a previsão do próximo período de tempo.

Previsão com Múltiplos Passos à Frente

Fazer previsão de mais de um período à frente é geralmente mais difícil do que prever um período à frente. Podemos formalizar isso da seguinte maneira. Suponha que consideremos fazer a previsão de y_{t+1} no momento t e em um período anterior s (de forma que $s < t$). Então, $\text{Var}[y_{t+1} - E(y_{t+1}|I_t)] \leq \text{Var}[y_{t+1} - E(y_{t+1}|I_s)]$, onde em geral a desigualdade é estrita. Não provaremos esse resultado de forma geral, mas, intuitivamente, ele faz sentido: a variância do erro de previsão ao fazermos a previsão de y_{t+1} é maior do que quando fazemos a previsão com base em menos informação.

Se y_t seguir um modelo AR(1) (que inclui um passeio aleatório, possivelmente com tendência), podemos facilmente mostrar que a variância do erro aumenta com o horizonte da previsão. O modelo é

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

$$E(u_t|I_{t-1}) = 0, I_{t-1} = \{y_{t-1}, y_{t-2}, \dots\},$$

e $\{u_t\}$ tem variância constante σ^2 condicional a I_{t-1} . No momento $t+h-1$, nossa previsão de y_{t+h} será $\alpha + \rho y_{t+h-1}$, e o erro de previsão será simplesmente u_{t+h} . Portanto, a variância da previsão com um passo à frente é simplesmente σ^2 . Para encontrar previsões com múltiplos passos à frente, teremos, por substituições repetidas,

$$y_{t+h} = (1 + \rho + \dots + \rho^{h-1})\alpha + \rho^h y_t + \rho^{h-1}u_{t+1} + \rho^{h-2}u_{t+2} + \dots + u_{t+h}.$$

No momento t , o valor esperado de u_{t+j} , para todo $j \geq 1$, será zero. Assim,

$$E(y_{t+h}|I_t) = (1 + \rho + \dots + \rho^{h-1})\alpha + \rho^h y_t, \quad (18.54)$$

e o erro de previsão será $e_{t,h} = \rho^{h-1}u_{t+1} + \rho^{h-2}u_{t+2} + \dots + u_{t+h}$. Isso é a soma de variáveis aleatórias não-correlacionadas e, assim, a variância da soma é a soma das variâncias: $\text{Var}(e_{t,h}) = \sigma^2[\rho^{2(h-1)} + \rho^{2(h-2)} + \dots + \rho^2 + 1]$. Como $\rho^2 > 0$, cada termo que multiplica σ^2 será positivo, e a variância do

erro de previsão aumenta com h . Quando $\rho^2 < 1$, conforme h fica maior, a variância de previsão converge para $\sigma^2/(1 - \rho^2)$, que é exatamente a variância incondicional de y_t . No caso de um passeio aleatório ($\rho = 1$), $p_{t,h} = \alpha h + y_t$, e $\text{Var}(e_{t,h}) = \sigma^2 h$: a variância da previsão cresce sem limites conforme o horizonte h aumenta. Isso demonstra que é muito difícil fazer a previsão de um passeio aleatório, com ou sem tendência, para um futuro longínquo. Por exemplo, previsões de taxas de juros para um futuro distante se tornam drasticamente menos precisas.

A equação (18.54) mostra que o uso do modelo AR(1) para fazer previsão com múltiplos passos à frente é fácil, uma vez que tenhamos estimado ρ por MQO. A previsão de y_{n+h} no momento n é

$$\hat{p}_{n,h} = (1 + \hat{\rho} + \dots + \hat{\rho}^{h-1})\hat{\alpha} + \hat{\rho}^h y_n. \quad (18.55)$$

A obtenção de intervalos de previsão é mais difícil, a menos que $h = 1$, pois é difícil obter o erro-padrão de $\hat{p}_{n,h}$. Contudo, o erro-padrão de $\hat{p}_{n,h}$ habitualmente é pequeno, comparado com o desvio-padrão do termo erro, e este último pode ser estimado como $\hat{\sigma}[\hat{\rho}^{2(h-1)} + \hat{\rho}^{2(h-2)} + \dots + \hat{\rho}^2 + 1]^{1/2}$, onde $\hat{\sigma}$ é o erro-padrão da regressão a partir da estimativa AR(1). Podemos usar isso para obter um intervalo de confiança aproximado. Por exemplo, quando $h = 2$, um intervalo de confiança aproximado de 95% (para n grande) será

$$\hat{P}_{n,2} \pm 1,96 \hat{\sigma} (1 + \hat{\rho}^2)^{1/2}. \quad (18.56)$$

Como estamos subestimando o desvio-padrão de y_{n+h} , esse intervalo será muito estreito, mas talvez não em demasia, especialmente se n for grande.

Uma abordagem menos tradicional, mas útil, é estimar um modelo diferente para cada horizonte de previsão. Por exemplo, suponha que queremos fazer a previsão de y dois períodos adiante. Se I_t depender somente de y até o momento t , podemos assumir que $E(y_{t+2}|I_t) = \alpha_0 + \gamma_1 y_t$, [que, como vimos anteriormente, se manterá se $\{y_t\}$ seguir um modelo AR(1)]. Podemos estimar α_0 e γ_1 fazendo a regressão de y_t sobre um intercepto e sobre y_{t-2} . Embora os erros nessa equação contenham correlação serial — erros em períodos adjacentes são correlacionados — podemos obter estimadores consistentes e aproximadamente normais de α_0 e γ_1 . A previsão de y_{n+2} no momento n será simplesmente $\hat{p}_{n,2} = \hat{\alpha}_0 + \hat{\gamma}_1 y_n$. Além disso, e de forma importante, o erro-padrão da regressão será exatamente o que precisamos para computar um intervalo de confiança da previsão. Infelizmente, para obter o erro-padrão de $\hat{p}_{n,2}$ usando o truque da previsão com um passo à frente exige que obtenhamos um erro-padrão robusto em relação à correlação serial do tipo descrito na Seção 12.5. Esse erro-padrão tende a zero conforme n vai ficando grande, enquanto a variância do erro é constante. Portanto, podemos obter um intervalo aproximado usando (18.56) e colocando o EPR da regressão de y_t sobre y_{t-2} em lugar de $\hat{\sigma}(1 + \hat{\rho}^2)^{1/2}$. Contudo, devemos nos lembrar que isso ignora o erro de estimação em $\hat{\alpha}_0$ e $\hat{\gamma}_1$.

Também podemos computar previsões com múltiplos passos à frente com modelos auto-regressivos mais complicados. Por exemplo, suponha que $\{y_t\}$ siga um modelo AR(2) e que no momento n queremos fazer a previsão de y_{n+2} . Agora, $y_{n+2} = \alpha + \rho_1 y_{n+1} + \rho_2 y_n + u_{n+2}$ e, assim,

$$E(y_{n+2}|I_n) = \alpha + \rho_1 E(y_{n+1}|I_n) + \rho_2 y_n.$$

Podemos escrever essa equação como

$$p_{n,2} = \alpha + \rho_1 p_{n,1} + \rho_2 y_n,$$

de modo que a previsão com dois passos à frente no momento n poderá ser obtida logo que tenhamos a previsão com um passo à frente. Se os parâmetros do modelo AR(2) tiverem sido estimados por MQO, então, operacionalizaremos isso como

$$\hat{p}_{n,2} = \hat{\alpha} + \hat{\rho}_1 \hat{p}_{n,1} + \hat{\rho}_2 y_n. \quad (18.57)$$

Agora, $\hat{p}_{n,1} = \hat{\alpha} + \hat{\rho}_1 y_n + \hat{\rho}_2 y_{n-1}$, que podemos computar no momento n . Em seguida, inserimos essa expressão em (18.57), juntamente com y_n , para obter $\hat{p}_{n,2}$. Para qualquer $h > 2$, a obtenção de qualquer previsão com h passos à frente para um modelo AR(2) é fácil de encontrar em uma maneira recursiva: $\hat{p}_{n,h} = \hat{\alpha} + \hat{\rho}_1 \hat{p}_{n,h-1} + \hat{\rho}_2 \hat{p}_{n,h-2}$.

Raciocínio semelhante pode ser usado para obter previsões com múltiplos passos à frente para modelos VAR. Para ilustrar, suponha que tenhamos

$$y_t = \delta_0 + \alpha_1 y_{t-1} + \gamma_1 z_{t-1} + u_t \quad (18.58)$$

e

$$z_t = \eta_0 + \beta_1 y_{t-1} + \rho_1 z_{t-1} + v_t.$$

Agora, se quisermos fazer a previsão de y_{t+1} no momento t , simplesmente usamos $\hat{p}_{t,1} = \hat{\delta}_0 + \hat{\alpha}_1 y_t + \hat{\gamma}_1 z_t$. Igualmente, a previsão de z_{t+1} no momento t será (digamos) $\hat{g}_{t,1} = \hat{\eta}_0 + \hat{\beta}_1 y_t + \hat{\rho}_1 z_t$. Agora, suponha que queremos obter uma previsão com dois passos à frente de y no momento t . De (18.58), temos

$$E(y_{t+2}|I_t) = \delta_0 + \alpha_1 E(y_{t+1}|I_t) + \gamma_1 E(z_{t+1}|I_t)$$

[porque $E(u_{t+2}|I_t) = 0$] e, assim, podemos escrever a previsão como

$$\hat{p}_{t,2} = \hat{\delta}_0 + \hat{\alpha}_1 \hat{p}_{t,1} + \hat{\gamma}_1 \hat{g}_{t,1}. \quad (18.59)$$

Essa equação mostra que a previsão com dois passos à frente de y depende da previsão com um passo à frente de y e z . De forma geral, podemos construir previsões com múltiplos passos à frente de y com o uso da fórmula recursiva

$$\hat{p}_{t,h} = \hat{\delta}_0 + \hat{\alpha}_1 \hat{p}_{t,h-1} + \hat{\gamma}_1 \hat{g}_{t,h-1}, h \geq 2.$$

EXEMPLO 18.10**(Previsão dois Anos à Frente da Taxa de Desemprego)**

Para usar a equação (18.49) para fazermos a previsão do desemprego dois anos à frente — digamos, a taxa de 1998 usando os dados até 1996 —, precisamos de um modelo da inflação. O melhor modelo de \inf em termos de defasagens de $desemp$ e \inf parece ser um simples modelo AR(1) ($desemp_{t-1}$, não é significante quando adicionado à regressão):

$$\begin{aligned}\hat{\inf}_t &= 1,277 + 0,665 \inf_{t-1} \\ &\quad (0,558) (0,107) \\ n &= 48, R^2 = 0,457, \bar{R}^2 = 0,445.\end{aligned}$$

Se inserirmos o valor de \inf de 1996 nessa equação, obteremos a previsão de \inf para 1997: $\hat{\inf}_{1997} = 3,27$. Agora, podemos agregar esse resultado, com $desemp_{1997} = 5,35$ (que obtivemos anteriormente), em (18.59) para fazer a previsão de $desemp_{1998}$:

$$desemp_{1998} = 1,304 + 0,647(5,35) + 0,184(3,27) \approx 5,37.$$

Lembre-se, essa previsão utiliza informações somente até 1996. A previsão com um passo à frente de $desemp_{1998}$ obtida pela agregação dos valores de 1997 de $desemp$ e \inf em (18.48), foi aproximadamente 4,90. Pode-se obter a taxa efetiva de desemprego de 1998 em um recente *Economic Report of the President*. Você verificará que a previsão com um passo à frente produz um resultado mais próximo do efetivo que a previsão com dois passos à frente.

Como acontece na previsão com um passo à frente, uma raiz do erro quadrático médio fora da amostra ou um erro absoluto médio pode ser usado para fazermos a escolha entre os métodos de previsão com múltiplos passos à frente.

A Previsão de Processos com Tendência, Sazonais e Integrados

Agora, voltamos para a previsão de séries que exibem tendências, possuem sazonalidade ou têm raízes unitárias. Lembre-se dos capítulos 10 e 11 que uma abordagem para lidar com variáveis dependentes ou independentes com tendência em modelos de regressão é a inclusão de tendências temporais, das quais a mais popular é a tendência linear. Tendências também podem ser incluídas em equações de previsões, embora devam ser usadas com cuidado.

No caso mais simples, suponha que $\{y_t\}$ tenha uma tendência linear, mas seja imprevisível em torno dessa tendência. Nesse caso, podemos escrever

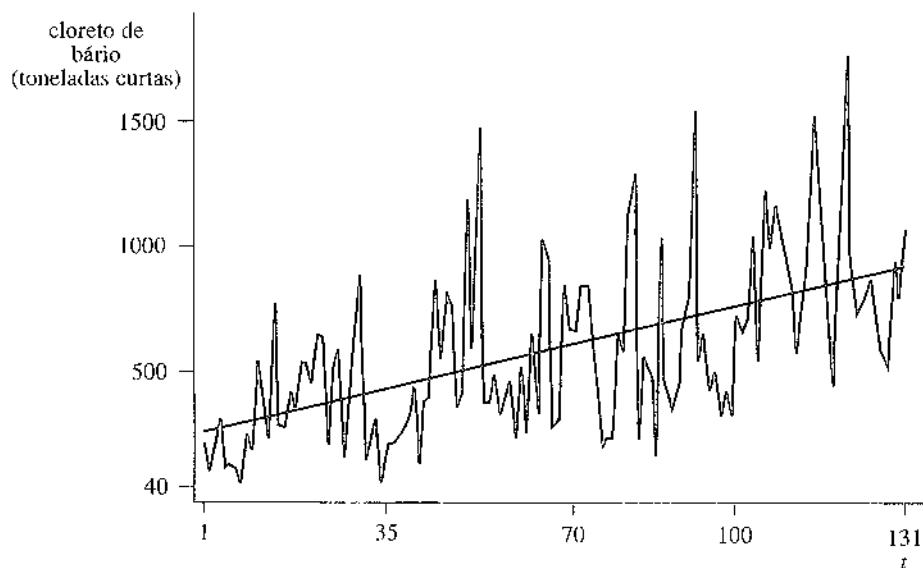
$$y_t = \alpha + \beta t + u_t, E(u_t | I_{t-1}) = 0, t = 1, 2, \dots, \quad (18.60)$$

onde, como sempre, I_{t-1} contém informação observada até o momento $t - 1$ (que inclui pelo menos y passados). Como fazemos a previsão de y_{n+h} no momento n para qualquer $h \geq 1$? Isso é simples, pois $E(y_{n+h} | I_n) = \alpha + \beta(n + h)$. A variância do erro de previsão será simplesmente $\sigma^2 = \text{Var}(u_t)$ (assumindo uma variância constante ao longo do tempo). Se estimarmos α e β por MQO usando as primeiras

n observações, então, nossa previsão de y_{n+h} no momento n será $\hat{y}_{n,h} = \hat{\alpha} + \hat{\beta}(n + h)$. Em outras palavras, simplesmente agregamos o período de tempo correspondente a y na função de tendência estimada. Por exemplo, se usarmos as $n = 131$ observações do arquivo BARIUM.RAW para fazermos a previsão das importações mensais de bário feitas pelos Estados Unidos da China, obteremos $\hat{\alpha} = 249,56$ e $\hat{\beta} = 5,15$. O período da amostra termina em dezembro de 1988, de modo que a importação chinesa seis meses depois será $249,56 + 5,15(137) = 955,11$, medida em toneladas curtas¹. A título de comparação, o valor de dezembro de 1988 foi 1.087,81, maior, portanto, que o valor previsto para seis meses depois. A série e sua linha de tendência estimada são mostradas na Figura 18.2.

Figura 18.2

Importações de cloreto de bário chinês feitas pelos Estados Unidos (em toneladas curtas) e sua linha de tendência linear estimada, $249,56 + 5,15t$.



Como discutido no Capítulo 10, a maioria das séries temporais econômicas é mais bem caracterizada como tendo, pelo menos aproximadamente, um taxa constante de crescimento, o que sugere que $\log(y)$ segue uma tendência temporal linear. Suponha que usemos n observações para obter a equação

$$\hat{\log}(y_t) = \hat{\alpha} + \hat{\beta}t, t = 1, 2, \dots, n. \quad (18.61)$$

Então, para prever $\log(y)$ em qualquer período de tempo futuro $n + h$, simplesmente agregamos $n + h$ na equação de tendência, como antes. Entretanto, isso não nos dá condições de fazer a previsão de y , que normalmente é o que queremos. É tentador simplesmente exponenciar $\hat{\alpha} + \hat{\beta}(n + h)$ para obter a previsão de y_{n+h} , mas isso não é muito correto, pelas mesmas razões que apresentamos na Seção 6.4. Temos

¹ NRT: 1 tonelada curta = 2.000 libras = 0,907,2 kg.

que considerar de forma apropriada o erro implícito em (18.61). A maneira mais fácil de fazer isso é usar as n observações para fazer a regressão de y_t sobre $\exp(\log y_t)$ sem um intercepto. Seja $\hat{\gamma}$ o coeficiente de inclinação de $\exp(\log y_t)$. Então, a previsão de y no período $n + h$ será simplesmente

$$\hat{p}_{n+h} = \hat{\gamma} \exp[\hat{\alpha} + \hat{\beta}(n + h)]. \quad (18.62)$$

Suponha que você modele $\{y_t; t = 1, 2, \dots, 46\}$ como uma tendência temporal linear, na qual os dados são anuais, começando em 1950 e terminando em 1995. Defina a variável ano_t como indo de 50 quando $t = 1$ a 95 quando $t = 46$. Se você estimar a equação $\hat{y}_t = \hat{\gamma} + \hat{\delta}\text{ano}_t$, como $\hat{\gamma}$ e $\hat{\delta}$ se compararão com $\hat{\alpha}$ e $\hat{\beta}$ em $\hat{y}_t = \hat{\alpha} + \hat{\beta}t$? Como se compararão as previsões das duas equações?

Como um exemplo, se usarmos as primeiras 687 semanas dos dados do índice da bolsa de valores de Nova Iorque contidos no arquivo NYSE.RAW, obteremos $\hat{\alpha} = 3,782$ e $\hat{\beta} = 0,0019$ [regredindo $\log(\text{preço}_t)$ sobre uma tendência temporal linear]; isso mostra que o índice cresce cerca de 0,2% por semana, em média. Quando regredimos preço_t sobre os valores estimados exponenciados, obtemos $\hat{y} = 1,018$. Agora, fazemos a previsão de preço_t quatro semanas à frente, que é a última semana na amostra, usando (18.62): $1,018 \cdot \exp(3,782 + 0,0019(691)) \approx 166,12$. O valor real foi 164,25, de modo que nossa previsão foi um pouco superestimada. Mas esse resultado é muito melhor do que se estimássemos uma tendência temporal linear das primeiras 687 semanas: o valor previsto da semana 691 será 152,23, que é uma previsão substancialmente subestimada.

Embora modelos de tendência possam ser úteis para fazermos previsões, eles devem ser usados com cuidado, especialmente em previsões de futuros longínquos de séries integradas que tenham tendências. O problema potencial pode ser visto considerando-se um passeio aleatório com tendência. No momento $t + h$, podemos escrever y_{t+h} como

$$y_{t+h} = \beta h + y_t + u_{t+1} + \dots + u_{t+h},$$

onde β é o termo de tendência (normalmente $\beta > 0$), e cada u_{t+j} tem média zero dado I_t e variância constante σ^2 . Como vimos anteriormente, a previsão de y_{t+h} no momento t é $E(y_{t+h}|I_t) = \beta h + y_t$, e a variância do erro de previsão é $\sigma^2 h$. O que acontece se usarmos um modelo de tendência linear? Seja y_0 o valor inicial do processo no momento zero, que assumimos como não-aleatório. Então, podemos também escrever

$$\begin{aligned} y_{t+h} &= y_0 + \beta(t + h) + u_1 + u_2 + \dots + u_{t+h} \\ &= y_0 + \beta(t + h) + v_{t+h}. \end{aligned}$$

Isso se parece com um modelo de tendência linear com o intercepto $\alpha = y_0$. Mas o erro v_{t+h} , embora tenha média zero, tem variância $\sigma^2(t + h)$. Portanto, se usarmos a tendência linear $y_0 + \beta(t + h)$ para fazer a previsão de y_{t+h} no momento t , a variância do erro de previsão será $\sigma^2(t + h)$ comparado com $\sigma^2 h$ quando usamos $\beta h + y_t$. A razão das variâncias de previsão será $(t + h)/h$, que pode ser grande para t grande. O ponto principal é que não deveremos usar uma tendência linear para fazer a previsão de um passeio aleatório com tendência. (O problema 18.17 pede que você compare previsões de uma

linha de tendência cúbica e as de um modelo simples de passeio aleatório para a taxa geral de fertilidade nos Estados Unidos.)

Tendências determinísticas também podem produzir previsões pobres se os parâmetros de tendência forem estimados usando dados velhos e o processo tiver um deslocamento subsequente na linha de tendência. Algumas vezes, choques exógenos — como o da crise do petróleo nos anos 1970 — podem mudar a trajetória das variáveis de tendência. Se uma linha de tendência antiga for usada para fazer a previsão em um futuro distante, as previsões poderão ficar muito longe da realidade. Esse problema pode ser mitigado com o uso de dados mais recentes disponíveis para obter os parâmetros da linha de tendência.

Nada nos impede de combinar tendências com outros modelos para fazer previsões. Por exemplo, podemos adicionar uma tendência linear a um modelo AR(1), que pode funcionar bem na previsão de séries com tendências lineares, mas que também são processos AR estáveis em torno da tendência.

Também é simples fazer previsões de processos com sazonalidade determinista (séries mensais ou trimestrais). Por exemplo, o arquivo BARIUM.RAW contém a produção mensal de gasolina nos Estados Unidos de 1978 a 1988. Essa série não tem uma tendência óbvia, mas tem um forte padrão sazonal. (A produção de gasolina é mais alta nos meses de verão e em dezembro.) No modelo mais simples, faríamos a regressão de gas (medida em galões) sobre onze *dummies* mensais, digamos de fevereiro a dezembro. Em seguida, a previsão para qualquer mês futuro é simplesmente o intercepto mais o coeficiente da *dummy* mensal apropriada. (Para janeiro, a previsão será exatamente o intercepto na regressão.) Também podemos adicionar defasagens de variáveis e tendências temporais para levar em conta séries gerais com sazonalidade.

Fazer previsões de processos com raízes unitárias também merece atenção especial. Anteriormente, obtivemos o valor esperado de um passeio aleatório condicional às informações até o momento n . Para prever um passeio aleatório, com possível tendência α , h períodos no futuro no momento n , usamos $\hat{p}_{n,h} = \hat{\alpha}h + y_n$, onde $\hat{\alpha}$ é a média amostral dos Δy , até $t = n$. (Se não houver tendência, definimos $\hat{\alpha} = 0$.) Essa abordagem impõe a raiz unitária. Uma alternativa seria estimar um modelo AR(1) para $\{y_t\}$ e usar a fórmula da previsão (18.55). Essa abordagem não impõe uma raiz unitária, mas se houver uma, \hat{p} convergirá em probabilidade para um conforme n fica maior. No entanto, \hat{p} pode ser substancialmente diferente de um, especialmente se o tamanho da amostra não for muito grande. A questão sobre qual método produz melhor previsão fora da amostra é um problema empírico. Se no modelo AR(1) ρ for menor que um, mesmo que levemente, o modelo AR(1) tenderá a produzir melhores previsões de longo prazo.

Em termos gerais, há duas abordagens para produzir previsões para processos I(1). O primeiro é impor uma raiz unitária. Para uma previsão com um passo à frente, obtemos um modelo para prever a mudança em y , Δy_{t+1} , dadas as informações até o momento t . Então, como $y_{t+1} = \Delta y_{t+1} + y_t$, $E(y_{t+1}|I_t) = E(\Delta y_{t+1}|I_t) + y_t$. Portanto, nossa previsão de y_{t+1} no momento n será exatamente

$$\hat{p}_n = \hat{g}_n + y_n,$$

onde \hat{g}_n é a previsão de Δy_{n+1} no momento n . Em geral, um modelo AR (que é necessariamente estável) ou uma auto-regressão vetorial é usado para Δy_t .

Isso pode ser estendido para previsões com múltiplos passos à frente escrevendo y_{n+h} como

$$y_{n+h} = (y_{n+h} - y_{n+h-1}) + (y_{n+h-1} - y_{n+h-2}) + \dots + (y_{n+1} - y_n) + y_n,$$

ou

$$y_{n+h} = \Delta y_{n+h} + \Delta y_{n+h-1} + \dots + \Delta y_{n+1} + y_n.$$

Portanto, a previsão de y_{n+h} no momento n será

$$\hat{p}_{n,h} = \hat{g}_{n,h} + \hat{g}_{n,h-1} + \dots + \hat{g}_{n,1} + y_n, \quad (16.63)$$

onde $\hat{g}_{n,j}$ é a previsão Δy_{n+j} no momento n . Por exemplo, podemos modelar Δy_t como um AR(1) estável, obter a previsão com múltiplos passos à frente a partir de (18.55) (mas com $\hat{\alpha}$ e $\hat{\rho}$ obtidos de Δy_t , sobre Δy_{t-1} , e y_t substituída por Δy_t), e depois agregar isso em (18.63).

A segunda abordagem para prever variáveis I(1) é usar um modelo geral AR ou VAR para $\{y_t\}$. Isso não impõe a raiz unitária. Por exemplo, se usarmos um modelo AR(2),

$$y_t = \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + u_t, \quad (18.64)$$

então, $\rho_1 + \rho_2 = 1$. Se agregarmos $\rho_1 = 1 - \rho_2$ e reorganizarmos, obteremos $\Delta y_t = \alpha - \rho_2 \Delta y_{t-1} + u_t$, que é um modelo AR(1) estável na diferença que nos leva de volta à primeira abordagem descrita anteriormente. Nada nos impede de estimar (18.64) diretamente por MQO. Algo interessante sobre essa regressão é que podemos usar a estatística t habitual de $\hat{\rho}_2$ para determinar se y_{t-2} é significante. (Isso assume que a hipótese de homoscedasticidade é mantida; se ela não for mantida, podemos usar a forma robusta em relação à heteroscedasticidade.) Não mostraremos isso formalmente, pois isso consiste em reescrever a equação como $y_t = \alpha + \gamma y_{t-1} - \rho_2 \Delta y_{t-1} + u_t$, onde $\gamma = \rho_1 + \rho_2$. Mesmo se $\gamma = 1$, ρ_2 será o coeficiente de um processo $\{\Delta y_{t-1}\}$ estacionário e fracamente dependente, com sinal negativo. Como os resultados da regressão serão idênticos a (18.64), podemos usar (18.64) diretamente.

Como um exemplo, vamos estimar um modelo AR(2) da taxa geral de fertilidade usando os dados do arquivo FERTIL3.RAW, até 1979. (No Exercício 18.17, você será solicitado a usar esse modelo para fazer a previsão, razão pela qual guardamos algumas observações do final da amostra.)

$$\begin{aligned} \hat{tgc}_t &= 3,22 + 1,272 tgc_{t-1} - 0,311 tgc_{t-2} \\ (2,92) \quad (0,120) &\quad (0,121) \\ n = 65, R^2 &= 0,949, \bar{R}^2 = 0,947 \end{aligned} \quad (18.65)$$

A estatística t da segunda defasagem está em torno de $-2,57$, que estatisticamente é diferente de zero em torno do nível de 1%. (A primeira defasagem também tem uma estatística t bastante significante, que tem uma distribuição t aproximada pelo mesmo raciocínio usado para $\hat{\rho}_2$.) O R -quadrado, ajustado ou não, não é especialmente informativo como um indicador do grau de ajuste, pois tgc aparentemente contém uma raiz unitária, e não faz muito sentido questionar quanto da variância de tgc estamos explicando.

Os coeficientes das duas defasagens em (18.65) somam 0,961, o que é próximo e estatisticamente não diferentes de um (como pode ser verificado pela aplicação do teste de Dickey-Fuller aumentando à equação $\Delta tgc_t = \alpha + \theta tgc_{t-1} + \delta_1 \Delta tgc_{t-1} + u_t$). Embora não tenhamos imposto a restrição de raiz unitária, ainda podemos usar (18.65) para previsão, como discutimos anteriormente.

Antes de terminar esta seção, salientamos uma melhora potencial na previsão no contexto de modelos vetoriais auto-regressivos com variáveis I(1). Suponha que $\{y_t\}$ e $\{z_t\}$ sejam cada um processos I(1). Uma abordagem para obter previsões de y é estimar uma auto-regressão bivariada nas variáveis Δy_t e Δz_t , e depois usar (18.63) para gerar previsões com um ou múltiplos passos à frente; essencialmente, essa é a primeira abordagem que descrevemos anteriormente. Porém, se y_t e z_t forem co-integrados,

temos mais variáveis estacionárias e estáveis no conjunto de informações que podem ser usadas na previsão de Δy_t : ou seja, defasagens de $y_t - \beta z_t$, onde β é o parâmetro de co-integração. Um modelo simples de correção de erro é

$$\begin{aligned}\Delta y_t &= \alpha_0 + \alpha_1 \Delta y_{t-1} + \gamma_1 \Delta z_{t-1} + \delta_1 (y_{t-1} - \beta z_{t-1}) + e_t \\ \mathbb{E}(e_t | I_{t-1}) &= 0.\end{aligned}\quad (18.66)$$

Para fazer a previsão de y_{n+1} , usamos as observações até n para estimar o parâmetro de co-integração, β e, então, estimar os parâmetros do modelo de correção de erro por MQO, como descrito na Seção 18.4. A previsão de Δy_{n+1} é fácil: simplesmente agregamos Δy_n , Δz_n e $y_n - \hat{\beta} z_n$ na equação estimada. Tendo obtido a previsão de Δy_{n+1} , nós a adicionamos a y_n .

Reorganizando o modelo de correção de erro, podemos escrever

$$y_t = \alpha_0 + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t, \quad (18.67)$$

onde $\rho_1 = 1 + \alpha_1 + \delta$, $\rho_2 = -\alpha_1$, e assim por diante, que é a primeira equação de um modelo VAR para y_t e z_t . Observe que isso depende de cinco parâmetros, a mesma quantidade do modelo de correção de erro. A questão é que, para o propósito de fazer previsão, o modelo VAR nos níveis e o modelo de correção de erro são essencialmente os mesmos. Esse não será o caso em modelos mais gerais de correção de erro. Por exemplo, suponha que $\alpha_1 = \gamma_1 = 0$ em (18.66), mas que temos um segundo termo de correção de erro, $\delta_2(y_{t-2} - \beta z_{t-2})$. Nesse caso, o modelo de correção de erro envolve somente quatro parâmetros, enquanto (18.67) — que tem a mesma ordem de defasagens para y e z — contém cinco parâmetros. Assim, modelos de correção de erro podem economizar nos parâmetros; isto é, eles são, de forma geral, mais *parcimoniosos* do que os modelos VAR em níveis.

Se y_t e z_t forem $I(1)$, mas não co-integrados, o modelo apropriado será (18.66), sem o termo de correção de erro. Isso pode ser usado para fazer a previsão de Δy_{n+1} , e podemos adicioná-lo a y_n para fazer a previsão de y_{n+1} .

Os tópicos sobre séries temporais tratados neste capítulo são usados rotineiramente em macroeconomia empírica, finanças empíricas, e em uma variedade de outros campos aplicados. Iniciamos mostrando como modelos de defasagem distribuída infinita podem ser interpretados e estimados. Eles podem proporcionar distribuições flexíveis de defasagens com menor número de parâmetros do que um modelo de defasagem distribuída finita. Os modelos de defasagem distribuída geométrica e, de forma mais geral, de defasagem distribuída racional, são os mais populares. Eles podem ser estimados usando procedimentos econometríticos padrão em equações dinâmicas simples.

O teste da existência de uma raiz unitária tem se tornado muito comum em séries temporais econométricas. Se uma série tiver uma raiz unitária, então, em muitos casos, as habituais aproximações normais de amostras grandes não mais serão válidas. Além disso, um processo de raiz unitária tem a propriedade de que uma inovação tem um efeito permanente, que é de interesse por si só. Embora existam muitos testes de raízes unitárias, o teste t de Dickey-Fuller — e sua extensão, o teste de Dickey-Fuller aumentado — é provavelmente o mais popular e mais fácil de ser implementado. Podemos levar

em conta uma tendência linear durante o teste de raízes unitárias adicionando uma tendência na regressão de Dickey-Fuller.

Quando é feita a regressão de uma série I(1), y_t , sobre outra série I(1), x_t , existe séria preocupação com respeito à regressão espúria, mesmo que as séries não contenham tendências óbvias. Esse assunto tem sido estudado com profundidade no caso de um passeio aleatório: mesmo se os dois passeios aleatórios forem independentes, o teste t habitual da significância do coeficiente de inclinação, baseado nos habituais valores críticos, rejeitará muito mais que o tamanho nominal do teste. Além disso, o R^2 tende a uma variável aleatória, em vez de a zero (como seria o caso se fizéssemos a regressão da diferença de y , sobre a diferença de x).

Em um caso importante, uma regressão que envolve variáveis I(1) não será espúria, e isso ocorre quando as séries são co-integradas. Isso significa que uma função linear das duas variáveis I(1) será I(0). Se y_t e x_t forem I(1), mas $y_t - \beta x_t$ for I(0), y_t e x_t não poderão tender arbitrariamente para direções muito distintas. Existem testes simples da hipótese nula de não co-integração contra a alternativa de co-integração, um dos quais é baseado na aplicação do teste de raiz unitária de Dickey-Fuller aos resíduos de uma regressão estática. Também existem estimadores simples do parâmetro de co-integração que produzem estatísticas t com distribuições normais padrão aproximadas (e intervalos de confiança assintoticamente válidos). Tratamos do estimador de adiantamentos e atrasos na Seção 18.4.

A co-integração entre y_t e x_t implica que os termos de correção de erro podem aparecer em um modelo relacionando Δy_t a Δx_t ; os termos de correção de erro são defasagens em $y_t - \beta x_t$, onde β é o parâmetro de co-integração. Um procedimento simples de estimação em duas etapas está disponível para estimar modelos de correção de erro. Primeiro, β é estimado usando uma regressão estática (ou a regressão de adiantamentos e atrasos). Depois, o MQO é usado para estimar um modelo dinâmico simples em primeiras diferenças que inclui os termos de correção de erro.

A Seção 18.5 apresentou uma introdução à previsão, com ênfase em métodos com base em regressões. Modelos estáticos ou, de forma mais genérica, modelos que contenham variáveis explicativas datadas contemporaneamente com a variável dependente, são limitados, pois há a necessidade de fazer a previsão das variáveis explicativas. Se agregarmos valores hipotéticos de variáveis explicativas cujo futuro é desconhecido, obteremos uma previsão condicional. Previsões incondicionais são semelhantes a simplesmente modelar y_t como uma função da informação *passada* que tenhamos observado no momento para o qual necessitamos a previsão. Modelos de regressão dinâmica, inclusive as auto-regressões e as auto-regressões vetoriais, são usados rotineiramente. Além da obtenção de previsões por ponto com um passo à frente, também discutimos a construção de intervalos de previsão.

Vários critérios são usados para se fazer a seleção entre os métodos de previsão. Os indicadores de desempenho mais comuns são a raiz do erro quadrático médio e o erro absoluto médio. Ambos estimam o tamanho do erro de previsão médio. É bastante informativo computar esses indicadores usando previsões fora da amostra.

As previsões com múltiplos passos à frente apresentam novos desafios e estão sujeitas a grandes variâncias do erro de previsão. Contudo, para modelos como os de auto-regressões e de auto-regressões vetoriais, previsões com múltiplos passos à frente podem ser computadas, e intervalos de previsão aproximados podem ser obtidos.

A previsão de séries com tendência e I(1) requer cuidado especial. Processos com tendências determinísticas podem ser previstos com a inclusão de tendências temporais em modelos de regressão, possivelmente com defasagens de variáveis. Uma desvantagem potencial é que tendências determinísticas podem produzir previsões pobres para previsões de horizonte longínquo: uma vez estimada, uma tendência linear continua a aumentar ou a diminuir. A abordagem típica para fazer a previsão de um processo I(1) é fazer a previsão da diferença no processo e adicionar o nível da variável àquela diferença prevista. Alternativamente, modelos vetoriais auto-regressivos podem ser usados nos níveis das séries. Se as séries forem co-integradas, poderemos usar modelos de correção de erro.

18.1 Considere a equação (18.15) com $k = 2$. Usando o método VI para estimar γ_h e ρ , o que você usaria como instrumentais de y_{t-1} ?

18.2 Um modelo econômico interessante que leva a um modelo econométrico com uma variável dependente defasada relaciona y_t ao *valor esperado* de x_t , digamos x_t^* , onde a esperança é baseada em todas as informações observadas no momento $t - 1$:

$$y_t = \alpha_0 + \alpha_1 x_t^* + u_t. \quad (18.68)$$

Uma hipótese natural sobre $\{u_t\}$ é que $E(u_t | I_{t-1}) = 0$, onde I_{t-1} representa todas as informações de y e x observadas no momento $t - 1$; isso significa que $E(y_t | I_{t-1}) = \alpha_0 + \alpha_1 x_t^*$. Para completar esse modelo, necessitamos de uma hipótese sobre como a esperança x_t^* é formada. Vimos um exemplo simples de expectativas adaptativas na Seção 11.2, onde $x_t^* = x_{t-1}$. Um esquema de expectativas adaptativas mais complicado é

$$x_t^* - x_{t-1}^* = \lambda(x_{t-1} - x_{t-1}^*), \quad (18.69)$$

onde $0 < \lambda < 1$. Essa equação implica que a mudança nas expectativas reage se o valor realizado do último período estava acima ou abaixo de sua expectativa. A hipótese $0 < \lambda < 1$ implica que a mudança nas expectativas é uma fração do erro do último período.

(i) Mostre que as duas equações implicam que

$$y_t = \lambda \alpha_0 + (1 - \lambda)y_{t-1} + \lambda \alpha_1 x_{t-1} + u_t - (1 - \lambda)u_{t-1}.$$

[*Sugestão:* Faça a defasagem da equação (18.68) em um período, multiplique-a por $(1 - \lambda)$, e subtraia esse resultado de (18.68). Depois, use (18.69).]

(ii) Se $E(u_t | I_{t-1}) = 0$, $\{u_t\}$ é serialmente não-correlacionado. O que isso implica sobre os novos erros, $v_t = u_t - (1 - \lambda)u_{t-1}$?

(iii) Se escrevermos a equação da parte (i) como

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + v_t,$$

como você estimaria os β_j consistentemente?

(iv) Dados estimadores consistentes dos β_j , como você estimaria λ e α_1 consistentemente?

18.3 Suponha que $\{y_t\}$ e $\{z_t\}$ sejam séries I(1), mas $y_t - \beta z_t$ seja I(0) para algum $\beta \neq 0$. Mostre que para qualquer $\delta \neq \beta$, $y_t - \delta z_t$ deve ser I(1).

18.4 Considere o modelo de correção de erro na equação (18.37). Mostre que se você adicionar mais uma defasagem do termo de correção de erro, $y_{t-2} - \beta x_{t-2}$, a equação sofrerá de colinearidade perfeita. [Sugestão: mostre que $y_{t-2} - \beta x_{t-2}$ é uma função linear perfeita de $y_{t-1} - \beta x_{t-1}$, Δx_{t-1} e Δy_{t-1} .]

18.5 Suponha que o processo $\{(x_t, y_t); t = 0, 1, 2, \dots\}$ satisfaz as equações

$$y_t = \beta x_t + u_t$$

e

$$\Delta x_t = \gamma \Delta x_{t-1} + \nu_t,$$

onde $E(u_t | I_{t-1}) = E(\nu_t | I_{t-1}) = 0$, I_{t-1} contém informações sobre x e y datadas no momento $t-1$ e nos anteriores, $\beta \neq 0$, e $|\gamma| < 1$ [de forma que x_t e, portanto, y_t , são $I(1)$]. Mostre que essas duas equações implicam um modelo de correção de erro da forma

$$\Delta y_t = \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + e_t,$$

onde $\gamma_1 = \beta\gamma$, $\delta = -1$ e $e_t = u_t + \beta\nu_t$. (Sugestão: Primeiro, subtraia y_{t-1} de ambos os lados da primeira equação. Depois, adicione e subtraia βx_{t-1} do lado direito e reorganize. Por fim, use a segunda equação para obter o modelo de correção de erro que contém Δx_{t-1} .)

18.6 Usando os dados mensais contidos no arquivo VOLAT.RAW, foi estimado o seguinte modelo:

$$\hat{pcip} = 1,54 + 0,344 pcip_{-1} + 0,074 pcip_{-2} + 0,073 pcip_{-3} + 0,031 pcsp_{-1}$$

(0,56)	(0,042)	(0,045)	(0,042)	(0,013)
--------	---------	---------	---------	---------

$$n = 554, R^2 = 0,174, \bar{R}^2 = 0,168,$$

onde $pcip$ é a mudança percentual na produção industrial, a uma taxa anualizada, e $pcsp$ é a mudança percentual no Índice Standard & Poors 500, também a uma taxa anualizada.

- (i) Se os três meses anteriores de $pcip$ forem zero e $pcsp_{-1} = 0$, qual será o crescimento previsto na produção industrial para este mês? Será ele, estatisticamente, diferente de zero?
- (ii) Se os três meses anteriores de $pcip$ forem zero, mas $pcsp_{-1} = 10$, qual será o crescimento previsto da produção industrial?
- (iii) Qual sua conclusão sobre o efeito do mercado de ações sobre a atividade econômica real?

18.7 Seja cM_t o crescimento anual da oferta de moeda e $desemp$, a taxa de desemprego. Assumindo que $desemp$, segue um processo AR(1) estável, explique em detalhes como você verificaria se cM Granger-causa $desemp$.

18.8 Suponha que y_t segue o modelo

$$y_t = \alpha + \delta_1 z_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

$$E(e_t | I_{t-1}) = 0,$$

onde I_{t-1} contém y e z datadas em $t-1$ e em momentos anteriores.

- (i) Mostre que $E(y_{t+1}|I_t) = (1 - \rho)\alpha + \rho y_t + \delta_1 z_t - \rho\delta_1 z_{t-1}$. (Sugestão: Escreva $u_{t-1} = y_{t-1} - \alpha - \delta_1 z_{t-2}$ e agregue isso na segunda equação, depois, agregue o resultado na primeira equação e tire a esperança condicional.)
- (ii) Suponha que você use n observações para estimar α , δ_1 e ρ . Escreva a equação para fazer a previsão de y_{n+1} .
- (iii) Explique o motivo de o modelo com uma defasagem z e correção serial AR(1) ser um caso especial do modelo

$$y_t = \alpha_0 + \rho y_{t-1} + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + e_t$$

- (iv) O que a parte (iii) sugere sobre o uso de modelos com correção serial AR(1) para fazer previsões?

18.9 Seja $\{y_t\}$ uma sequência I(1). Suponha que \hat{g}_n seja a previsão com um passo à frente de Δy_{n+1} e seja $\hat{p}_n = \hat{g}_n + y_n$ a previsão com um passo à frente de y_{n+1} . Explique por que os erros de previsão de Δy_{n+1} e y_{n+1} são idênticos.

A Montagem de um Projeto na Prática

este capítulo, discorremos sobre os ingredientes de uma análise empírica bem-sucedida, com ênfase na montagem de um projeto de trabalho de final de curso. Além de relembrar sobre importantes problemas que surgiram ao longo deste texto, enfatizamos temas recorrentes que são importantes para a pesquisa aplicada. Também fazemos sugestões de temas como uma maneira de estimular sua imaginação. São dadas, como referências, várias fontes de pesquisas econômicas e de dados.

19.1 A FORMULAÇÃO DE UMA PERGUNTA

A importância de propor uma pergunta bem específica não pode ser considerada exagerada. Se não fôrmos explícitos sobre a meta de nossa análise, não teremos condições de saber nem mesmo por onde começar. A ampla disponibilidade de fartos conjuntos de dados faz com que seja tentador lançar uma busca de acervos de dados com base em idéias mal planejadas, o que freqüentemente será contraproducente. É provável que, sem cuidadosa formulação de suas hipóteses e definição do tipo de modelo que será necessário para fazer a estimativa, você se esqueça de coletar informações sobre variáveis importantes, obtenha uma amostra da população errada ou colete dados do período errado.

Isso não quer dizer que você deve formular sua pergunta no vazio. Especialmente para projetos de cursos de um período, não é possível ser muito ambicioso. Portanto, ao escolher um assunto, deve-se estar razoavelmente seguro de que existe o conjunto de dados que possibilitará responder à pergunta formulada no tempo previsto.

É necessário decidir qual área da economia ou de outras ciências sociais lhe interessa, quando estiver escolhendo um tema. Por exemplo, se você fez um curso sobre economia do trabalho, muito provavelmente viu teorias que podem ser testadas empiricamente ou relações que possuam alguma relevância para a tomada de decisões. Os economistas especializados na área do trabalho estão continuamente propondo novas variáveis que possam explicar diferenciais salariais. Entre os exemplos, encontramos a qualidade do ensino médio [Card e Krueger (1992) e Betts (1995)], o grau de matemática e ciência ministrado no ensino médio [Levine e Zimmerman (1995)] e a aparência física [Hamermesh e Biddle (1994), Averett e Korenman (1996) e Biddle e Hamermesh (1998)]. Pesquisadores de finanças públicas estaduais e locais estudam de que forma a atividade econômica local depende das variáveis de política econômica, tais como os impostos sobre a propriedade, os impostos sobre vendas, o nível e a qualidade dos serviços públicos (tais como escolas, corpo de bombeiros e policiamento), e assim por diante. [Veja, por exemplo, White (1986), Papke (1987), Bartik (1991) e Netzer (1992).]

Os economistas que estudam problemas da educação estão interessados em conhecer como as despesas públicas afetam o desempenho [Hanushek (1986)], ou se freqüentar certos tipos de escolas melhora o desempenho [por exemplo, Evans e Schwab (1995)], assim como na determinação dos fatores que afetam a decisão das escolas particulares quanto à escolha do local onde devem se localizar [Downes e Greenstein (1996)].

Os macroeconomistas estão interessados nas relações de várias séries temporais agregadas, tais como o elo entre o crescimento do produto interno bruto (PIB) e o crescimento no investimento fixo ou em maquinaria [veja De Long e Summers (1991)] ou o efeito dos impostos sobre as taxas de juros [por exemplo, Peek (1982)].

Certamente, existem razões para se estimar modelos que sejam predominantemente descritivos. Por exemplo, consultores de impostos sobre a propriedade usam modelos (chamados *modelos de preços hedônicos* — veja Exemplo 4.8) para estimar valores dos imóveis que não tenham sido negociados recentemente. Isso envolve um modelo de regressão relacionando o preço de um imóvel com suas características (tamanho, número de quartos, número de banheiros etc.). Como tema para um trabalho, isso não é muito estimulante: provavelmente não descobriremos nada muito surpreendente e análises desse tipo não têm implicações óbvias sobre decisões. A adição da taxa de criminalidade das redondezas como uma variável explicativa nos possibilitaria determinar a importância do fator criminalidade sobre os preços dos imóveis, algo que seria útil na estimativa dos custos da criminalidade.

Diversas relações têm sido estimadas utilizando dados macroeconômicos que são predominantemente descritivos. Por exemplo, uma função poupança agregada pode ser usada para estimar a propensão marginal agregada a poupar, como também a resposta da poupança aos retornos de ativos (como as taxas de juros). Uma análise desse tipo poderia ficar mais interessante usando dados de séries temporais de um país que tenha uma história de reviravoltas políticas e determinando se as taxas de poupança declinam durante períodos de incerteza política.

Após a decisão sobre uma área de pesquisa, existe uma variedade de maneiras para localizar publicações específicas sobre o tema. O *Journal of Economic Literature* (JEL) tem um sistema detalhado de classificação, de forma que a cada documento é atribuído um conjunto de códigos de identificação que o classifica dentro de certos campos da economia. O JEL também contém uma lista de artigos publicados em uma ampla variedade de periódicos, organizada por temas, e contém até mesmo pequenos resumos de alguns artigos.

Especialmente conveniente para encontrar documentos sobre os mais diversos assuntos são os serviços da **Internet**, tais como o *EconLit*, que é assinado por muitas universidades. O *EconLit* oferece aos usuários a possibilidade de fazer uma pesquisa extensa de quase todas as publicações sobre economia por autor, assunto, palavra no título, e assim por diante. O *Social Science Citation Index* é útil para encontrar artigos sobre uma ampla gama de assuntos no campo das ciências sociais, inclusive de artigos populares que tenham sido citados com freqüência em outros trabalhos publicados.

Ao pensar sobre um tema, temos que ter alguns pontos em mente. Primeiro, para uma pergunta ser interessante, ela não precisa ter uma base extensa de implicações sobre regras de decisão; de certa forma, ela pode ser de interesse local. Por exemplo, podemos estar interessados em saber se o fato de estudantes morarem em uma comunidade da universidade faz com que eles tenham maiores ou menores notas. Isso pode ou não ser de interesse de pessoas de fora da universidade, mas provavelmente será de interesse de algumas pessoas dentro da universidade. De outro lado, podemos estudar um problema que começa como de interesse local, mas acaba tendo um interesse mais amplo, como a determinação de quais fatores afetam, e quais políticas da universidade podem deter o abuso de álcool nos *campi* das universidades.

Segundo, é muito difícil, especialmente em um projeto trimestral ou semestral, fazer uma pesquisa verdadeiramente original, utilizando os agregados macroeconômicos padrão da economia dos Estados Unidos.

Por exemplo, a questão sobre se o crescimento monetário, o crescimento nos gastos governamentais, e assim por diante, afeta o crescimento econômico tem sido e continua sendo estudada por macroeconomistas profissionais. A questão de saber se os retornos das ações ou outros ativos podem ser sistematicamente previstos utilizando as informações conhecidas tem, por razões óbvias, sido estudada muitometiculosamente. Isso não significa que devemos evitar a estimativa de modelos empíricos macroeconômicos ou financeiros, mesmo porque o uso de dados mais recentes pode ser um acréscimo construtivo a um debate. Além disso, podemos, algumas vezes, encontrar uma nova variável que tenha um efeito importante sobre agregados econômicos ou sobre retornos financeiros; tal descoberta pode ser estimulante.

A questão é que exercícios como o uso de alguns anos adicionais para estimar uma curva de Phillips padrão ou uma função de consumo agregado da economia norte-americana, ou de outra economia importante, dificilmente produzirão novas descobertas, embora possam ser instrutivas para os estudantes. Em vez disso, podemos usar dados de um país menor para estimar uma curva de Phillips estática ou dinâmica, ou para testar a hipótese de mercados eficientes, e assim por diante.

No nível fora da macroeconomia também existem várias questões que têm sido muito estudadas. Por exemplo, os economistas especializados na área do trabalho têm publicado muitos trabalhos sobre a estimativa do retorno da educação. Essa questão continua sendo estudada por sua importância, e novos conjuntos de dados, como também novas abordagens econométricas, continuam sendo desenvolvidos. Por exemplo, como vimos no Capítulo 9, certos conjuntos de dados possuem melhores variáveis que substituem a aptidão do que outros. (Compare os arquivos WAGE1.RAW e WAGE2.RAW.) Em outros casos, podemos obter dados de painel ou dados de uma experimentação natural — veja o Capítulo 13 — que nos possibilitam abordar uma questão antiga a partir de uma perspectiva diferente.

Como outro exemplo, os criminologistas estão interessados em estudar os efeitos de diversas leis sobre a criminalidade. A questão que investiga se a pena capital tem um efeito dissuasor vem sendo discutida há muito tempo. De forma semelhante, os economistas vêm investigando se os impostos sobre fumo e álcool reduzem o consumo (como sempre, em um sentido *ceteris paribus*). Conforme mais dados em nível estadual se tornam disponíveis, um conjunto de dados de painel mais farto pode ser criado, e isso pode ajudar a responder melhor as principais questões sobre decisões políticas. E mais, existem descobertas recentes no combate ao crime — como o advento do policiamento comunitário — cuja efetividade pode ser avaliada empiricamente.

Quando estiver formulando uma pergunta, é interessante discutir suas idéias com seus colegas, com seu professor e com seus amigos. Você deve ser capaz de convencer as pessoas de que a resposta à sua pergunta é de alguma forma interessante. (Naturalmente, se você pode responder de forma persuasiva sua pergunta é outra coisa, mas você precisa começar com uma pergunta interessante.) Se alguém lhe perguntar sobre seu ensaio e você responder com “estou fazendo meu ensaio sobre a criminalidade” ou “estou fazendo meu ensaio sobre taxas de juros” é bem provável que você apenas tenha se decidido sobre uma área geral sem ainda ter formulado uma verdadeira pergunta. Você deveria ter condições de dizer algo mais ou menos como “estou estudando os efeitos do policiamento comunitário sobre os índices de criminalidade das cidades dos Estados Unidos” ou “estou verificando como a volatilidade da inflação afeta as taxas de curto prazo no Brasil”.

19.2 A REVISÃO DA LITERATURA

Todos os artigos, mesmo que relativamente curtos, devem conter uma revisão sobre a literatura relevante. É raro alguém empreender um projeto para o qual não haja um precedente já publicado.

Se você pesquisar em periódicos ou utilizar serviços de busca *on-line*, como o EconLit, para encontrar um assunto, já estará bem avançado em seu caminho para uma revisão da literatura. Se você

selecionar um tema por conta própria — como estudar os efeitos do uso de drogas sobre o desempenho dos alunos da sua universidade —, então, muito provavelmente você terá um trabalho mais árduo. Mas os serviços de busca *on-line* tornam esse trabalho muito mais fácil, já que você pode fazer a pesquisa por palavra-chave, por palavras incluídas no título, por autor etc. Você poderá, então, ler resumos dos documentos para verificar o quanto eles são relevantes para seu próprio trabalho.

Quando estiver fazendo sua pesquisa sobre a literatura, você deve pensar em temas relacionados que possam não ser exibidos em uma pesquisa feita com o uso de um punhado de palavras-chave. Por exemplo, se você estiver estudando os efeitos do uso de drogas sobre os salários ou sobre as notas de graduação, você provavelmente deve olhar a literatura sobre como o uso de álcool afeta tais fatores. O saber como fazer uma pesquisa meticulosa da literatura é uma habilidade que se adquire, mas você pode ir longe se pensar antes de fazer a pesquisa.

Os pesquisadores discordam sobre como uma revisão da literatura deve ser incorporada em um trabalho. Alguns gostam de ter uma seção separada chamada “revisão da literatura”, enquanto outros preferem incluí-la como parte da introdução. Isso é, em grande parte, uma questão de gosto, embora uma revisão da literatura extensa provavelmente mereça sua própria seção. Se o trabalho de conclusão for o ponto principal do curso — digamos, em um seminário de alunos mais experientes ou em um curso avançado de econometria —, sua revisão da literatura provavelmente será extensa. Trabalhos de conclusão de curso de um só período são normalmente mais curtos e as revisões da literatura são mais resumidas.

19.3 A COMPILAÇÃO DOS DADOS

A Decisão sobre o Conjunto de Dados Apropriado

A coleta de dados para um trabalho de conclusão de curso pode ser educativa, estimulante e, algumas vezes, até mesmo frustrante. Primeiro, você deve decidir sobre o tipo de dados necessários para responder à questão que você propôs. Como discutimos na introdução e tratamos em toda a extensão deste livro, os conjuntos de dados são apresentados em grande variedade de formas. Os tipos mais comuns são os de corte transversal, séries temporais, cortes transversais agrupados e dados de painel.

Muitas questões podem ser respondidas usando qualquer uma das estruturas de dados que descrevemos. Por exemplo, para estudar se uma maior imposição da lei reduz a criminalidade, poderíamos usar um corte transversal de cidades, uma série temporal de determinada cidade, ou um conjunto de dados de painel de cidades — que consiste de dados das mesmas cidades ao longo de dois ou mais anos.

A decisão sobre qual tipo de dados coletar muitas vezes depende da natureza da análise. Para responder a perguntas em nível individual ou familiar, na maioria das vezes, temos acesso a somente um corte transversal; em geral, esses dados são obtidos por meio de uma pesquisa. Depois, devemos indagar se poderemos obter dados suficientes para fazermos uma análise *ceteris paribus* convincente. Por exemplo, suponha que queremos saber se as famílias que fizeram planos de previdência privada (PPP) — que oferecem certas vantagens tributárias — investiram menos em outros tipos de aplicações. Em outras palavras, os investimentos em planos de previdência reduzem as aplicações em outros tipos de aplicações? Existem conjuntos de dados, tais como o *Survey of Consumer Finances*, que contêm informações sobre vários tipos de aplicações para uma diferente amostra de famílias a cada ano. Vários problemas aparecem com o uso de tal conjunto de dados. Talvez o mais importante seja saber se existem controles suficientes — entre os quais se incluiriam a renda, dados demográficos e variáveis substitutas das preferências quanto a aplicações financeiras — para fazer uma análise *ceteris paribus* razoável. Se esses forem os únicos tipos de dados disponíveis, devemos fazer o que for possível com eles.

As mesmas questões surgem com dados de corte transversal sobre empresas, cidades, estados etc. Na maioria dos casos, não é óbvio que teremos condições de fazer uma análise *ceteris paribus* com um único corte transversal. Por exemplo, qualquer estudo dos efeitos da imposição da lei sobre a criminalidade deve reconhecer a endogeneidade dos gastos com a imposição da lei. Quando usamos métodos padrão de regressão, pode ser muito difícil completar uma análise *ceteris paribus* convincente, independentemente de quantos controles tenhamos. (Veja Seção 19.4 para discussão adicional.)

Se você leu os capítulos avançados sobre métodos de dados de painel, você sabe que ter as mesmas unidades de corte transversal em dois ou mais pontos de tempo diferentes pode nos possibilitar o controle de efeitos não-observados constantes no tempo que normalmente confundiriam a regressão sobre um único corte transversal. Conjuntos de dados de painel sobre indivíduos ou famílias são relativamente difíceis de serem obtidos — embora existam alguns importantes, tal como o *Panel Study of Income Dynamics* —, mas podem ser usados de maneiras bastante convincentes. Conjuntos de dados de painel sobre empresas também existem. Por exemplo, a *CompuStat* e o *Center for Research on Securities Prices (CRSP)* administram conjuntos bastante grandes de dados de painel de informações financeiras de empresas. Mais fáceis de serem obtidos são conjuntos de dados de painel de unidades maiores, como escolas, cidades, municípios e estados, já que esses dados tendem a não desaparecer ao longo do tempo, e as agências governamentais são responsáveis pela coleta de informações das mesmas variáveis a cada ano. Por exemplo, o *Federal Bureau of Investigation* coleta e publica informações detalhadas sobre as taxas de criminalidade no nível das cidades. Fontes de dados estão listadas ao final deste capítulo.

Os dados são fornecidos em uma variedade de formas. Alguns conjuntos de dados, especialmente os históricos, estão disponíveis somente na forma impressa. Se o conjunto de dados for pequeno, é conveniente você mesmo digitar os dados retirados da fonte impressa. Algumas vezes, são publicados artigos com pequenos conjuntos de dados — especialmente aplicações de séries temporais. Eles podem ser usados em um estudo empírico, talvez com a suplementação dos dados de anos mais recentes.

Muitos conjuntos de dados estão disponíveis em disquetes ou fitas magnéticas. É muito fácil trabalhar com os primeiros. Hoje em dia, conjuntos de dados bastante grandes podem ser colocados em pequenos disquetes. Várias agências governamentais vendem disquetes de dados, como também empresas particulares. Autores de trabalhos muitas vezes se prontificam a fornecer seus dados na forma de disquetes ou como anexos de e-mail.

Cada vez mais, conjuntos de dados são disponibilizados na Internet. A Rede Mundial é uma vasta fonte de **bases de dados on-line**. Foram criados, recentemente, inúmeros sites na rede contendo links para conjuntos de dados. Vários outros sites contêm links para conjuntos de dados que são de interesse dos economistas; alguns deles estão listados ao final deste capítulo. De maneira geral, fazer pesquisa na Internet sobre conjuntos de dados é razoavelmente fácil e se tornará ainda mais conveniente no futuro.

A Entrada e o Armazenamento de Seus Dados

Após ter se decidido por um tipo de dado e ter localizado a fonte desses dados, você deve converter os dados em formato utilizável. Se os dados estão em disquetes, eles já estarão em algum tipo de formato, esperançosamente em um de amplo uso. A maneira mais flexível de se obter dados em disquetes é como um **arquivo de texto (ASCII)** padrão. Todos os programas de estatística e econometria permitem que dados puros sejam armazenados nessa forma. Em geral, é simples ler um arquivo de texto diretamente em um programa de econometria, desde que o arquivo tenha sido adequadamente estruturado. Os arquivos de dados que temos usado em todo este livro fornecem vários exemplos de como conjuntos de dados de corte transversal, séries temporais, cortes transversais agrupados e painéis são habitualmente armazenados. Como regra geral, os dados devem ter uma forma tabular, com cada observação

representando uma linha diferente; as colunas no conjunto de dados representam diferentes variáveis. Ocasionalmente, é possível encontrar conjuntos de dados armazenados com cada coluna representando uma observação e cada linha uma variável diferente. Isso não é o ideal, mas muitos programas possibilitam que os dados sejam lidos nessa forma e depois reformulados. Naturalmente, é crucial saber como os dados estão organizados antes de transferi-los para seu programa de econometria.

Quanto aos conjuntos de dados de séries temporais, existe somente uma maneira sensata de entrar e armazenar os dados: a saber, cronologicamente, com o período de tempo mais antigo listado como a primeira observação, e o mais recente como a última observação. Muitas vezes é útil incluir variáveis indicando os anos e, se relevantes, os trimestres ou os meses. Isso facilitará a estimativa de uma variedade de modelos mais tarde, inclusive levando em conta a sazonalidade e intervalos em diferentes períodos. Para cortes transversais agrupados ao longo do tempo, normalmente é melhor ter o corte transversal do ano mais antigo preenchendo o primeiro bloco de observações, seguido pelo corte transversal do segundo ano, e assim sucessivamente. (Veja FERTIL1.RAW como um exemplo.) Essa disposição não é fundamental, mas é importante ter uma variável declarando o ano atribuído a cada observação.

Com referência aos dados de painel, como examinamos na Seção 13.5, ele será melhor se todos os anos de cada observação do corte transversal forem adjacentes e estiverem em ordem cronológica. Com esse ordenamento, poderemos usar todos os métodos de dados de painel dos capítulos 13 e 14. Com dados de painel, é importante incluir um identificador exclusivo de cada unidade de corte transversal, juntamente com uma variável anual.

Se seus dados estão na forma impressa, você tem várias opções para colocá-los no computador. Primeiro, você pode criar um arquivo de texto usando um **editor de texto** padrão. (Foi assim que vários dos conjuntos de dados no formato RAW foram inicialmente criados.) De modo geral, é preciso que cada linha inicie uma nova observação, que cada linha contenha o mesmo ordenamento das variáveis — particularmente, cada linha deve ter o mesmo número de entradas — e que os valores sejam separados ao menos um espaço. Algumas vezes, um separador diferente, como a vírgula, será melhor, mas isso dependerá do programa que você estiver usando. Se houver observações faltantes em algumas variáveis, você deve decidir como representará esse fato; simplesmente deixar um espaço em branco geralmente não funciona. Muitos programas de regressão aceitam um ponto como um símbolo de valor ausente. Algumas pessoas preferem usar um número — presumivelmente um valor impossível para a variável de interesse — para representar valores ausentes. Se você não tiver cuidado, isso poderá ser perigoso; falaremos sobre isso mais adiante.

Se você tiver dados não numéricos — por exemplo, você quer incluir os nomes em uma amostra de universidades ou os nomes de cidades —, então, você terá que verificar o programa de econometria que será utilizado para ver a melhor forma de inserir tais variáveis (muitas vezes chamadas de *strings*). Em geral, *strings* são colocadas entre aspas simples ou duplas. Ou o arquivo de texto pode obedecer a uma formatação rígida, que normalmente exige um pequeno programa para ler o arquivo de texto. Mas você precisa verificar seu programa de econometria para encontrar mais detalhes.

Outra opção geralmente disponível é usar uma **planilha** para entrar seus dados, como uma de Excel. Ela oferece duas vantagens sobre o arquivo de texto. Primeira, como cada observação de cada variável será uma célula, é menos provável que os números se juntem (como aconteceria se você esquecesse de colocar um espaço em um arquivo de texto). Segunda, as planilhas permitem manipulação de dados, como classificação, cálculo de médias etc. Essa segunda vantagem é menos importante se você utilizar um programa econométrico que permita tratamento sofisticado dos dados; muitos pacotes de programas, inclusive o Eviews e o Stata, estão nessa categoria.

Se você usar uma planilha como a primeira entrada de dados, muitas vezes você terá que exportar os dados em um formato que poderá ser lido por seu programa econométrico. Em geral, isso é bastante simples, já que as planilhas exportam arquivos de texto usando uma variedade de formatos.

Uma terceira alternativa é entrar os dados diretamente em seu programa econômétrico. Embora essa alternativa elimine a necessidade de um editor de texto ou uma planilha, ela pode ser mais complicada se você não puder livremente mover-se entre as diferentes observações para fazer correções ou adições.

Os dados baixados da Internet podem vir em uma diversidade de formas. Freqüentemente, os dados vêm como arquivos de texto, mas convenções diferentes são usadas na separação das variáveis; nos conjuntos de dados de painel, as convenções sobre como ordenar os dados podem diferir. Alguns conjuntos de dados da Internet vêm em arquivos de planilhas, em cujo caso você terá que usar uma planilha apropriada para ler os dados.

Inspeção, Limpeza e Sumário de Seus Dados

É extremamente importante familiarizar-se com qualquer conjunto de dados que você usar em uma análise empírica. Se você mesmo fizer a entrada dos dados, você será forçado a saber tudo sobre eles. Porém, se os dados forem obtidos de uma fonte externa, ainda assim você terá que dedicar algum tempo para entender sua estrutura e suas convenções. Mesmo conjuntos de dados que são amplamente usados e fartamente documentados podem conter pequenas falhas. Se você estiver usando um conjunto de dados obtidos do autor de um artigo, você deve estar ciente que regras usadas na construção de conjuntos de dados podem ter sido esquecidas.

Anteriormente, examinamos as maneiras padronizadas de armazenar vários conjuntos de dados. Você também precisa saber como os valores ausentes são codificados. Preferencialmente, valores ausentes são indicados por um caractere não numérico, como um ponto. Se um número for usado como um código de valor ausente, como “999” ou “-1”, você precisará ser muito cuidadoso quando usar essas observações na computação de qualquer estatística. Seu pacote econômétrico provavelmente não reconhecerá que certo número na realidade representa um valor ausente; é provável que tais observações serão usadas como se fossem válidas, e isso poderá produzir resultados bastante equivocados. O melhor método é definir quaisquer códigos numéricos para valores ausentes como algum outro caractere (um ponto, por exemplo) que não possa ser confundido com dados reais.

Você também deve conhecer a natureza das variáveis no conjunto de dados. Quais são as variáveis binárias? Quais são as variáveis ordinais (como as de classificação de risco de crédito)? Quais são as unidades de medida das variáveis? Por exemplo, os valores monetários estão expressos em dólares, milhares de dólares, milhões de dólares, ou de alguma outra forma? As variáveis que representam uma taxa — tais como taxas de evasão escolar, taxas de inflação, taxas de sindicalização ou taxas de juros — são medidas como uma percentagem ou como uma proporção?

Especialmente em dados de séries temporais, é crucial saber se os valores monetários estão expressos em valores nominais (da época) ou reais (constantes). Se os valores estiverem em termos reais, qual é o ano, ou período base?

Se você receber um conjunto de dados de um autor, algumas variáveis já podem ter sido transformadas. Por exemplo, algumas vezes somente o log de uma variável (como a remuneração ou o salário) é descrito no conjunto de dados.

É necessário detectar os erros nos dados para que seja preservada a integridade de qualquer análise dos dados. Sempre é útil encontrar os mínimos, os máximos, as médias e os desvios-padrão de todas as variáveis da análise, ou pelo menos das mais importantes. Por exemplo, se você verificar que o valor mínimo da educação em sua amostra é -99, você sabe que pelo menos uma entrada da variável educação deve ser definida como valor ausente. Se, depois de mais verificações, você descobrir que várias observações têm -99 como o nível de educação, você pode confiar que se deparou com um código de valor ausente da educação. Como um outro exemplo, se você verificar que uma taxa média de condenações por assassinatos ao longo de uma amostra de cidades é 0,632, você saberá que ela é

medida como uma proporção, e não como uma percentagem. Depois, se o valor máximo estiver acima de um, provavelmente será um erro tipográfico. (Não é raro encontrar um conjunto de dados no qual a maioria das entradas de uma variável de taxa foi feita como uma percentagem, mas algumas delas estão expressas como uma proporção, e vice-versa. Tais erros de codificação de dados são difíceis de serem detectados, mas é importante tentar identificá-los.)

Também devemos ter cuidado ao usar dados de séries temporais. Se estivermos usando dados mensais ou trimestrais, precisamos saber quais variáveis, se houver, foram ajustadas quanto à sazonalidade. A transformação de dados também exige grande cuidado. Suponha que temos um conjunto de dados mensais e queremos criar a mudança em uma variável de um mês para o outro. Para fazer isso, precisamos ter certeza de que os dados estejam ordenados cronologicamente, do período mais antigo para o mais recente. Se por alguma razão esse não for o caso, o resultado da diferenciação será ininteligível. Para ter certeza de que os dados estão adequadamente ordenados, é interessante termos um indicador de período de tempo. Com dados anuais, é suficiente saber o ano, mas precisamos saber se o ano entrou com quatro ou dois dígitos (por exemplo, 1998 versus 98). Com dados mensais ou trimestrais, é também útil ter uma variável, ou variáveis, indicando o mês ou o trimestre. Com dados mensais, podemos ter um conjunto de variáveis *dummy* (11 ou 12) ou uma variável indicando o mês (1 a 12 ou uma variável *string*, tal como jan., fev., e assim por diante).

Com ou sem indicadores anuais, mensais ou trimestrais, podemos facilmente construir tendências temporais em todos os programas econômétricos. A criação de variáveis *dummy* sazonais é fácil se o mês ou o trimestre for indicado; no mínimo, devemos saber o mês ou o trimestre da primeira observação.

A manipulação de dados de painel pode ser ainda mais instigante. No Capítulo 13, tratamos do MQO agrupado nos dados diferenciados como um método geral de controlar efeitos não observados. Ao construir os dados diferenciados, devemos ter cuidado para não criar observações fantasmas. Suponha que temos um painel equilibrado de cidades de 1992 a 1997. Mesmo que os dados estejam ordenados cronologicamente no interior de cada unidade de corte transversal — algo que deve ser feito antes de prosseguir —, uma diferenciação descuidada criará uma observação de 1992 de todas as cidades, exceto a primeira da amostra. Essa observação será o valor de 1992 da cidade i , menos o valor de 1997 da cidade $i - 1$; isso é claramente um absurdo. Portanto, devemos nos certificar que 1992 seja ausente em todas as variáveis diferenciadas.

19.4 A ANÁLISE ECONOMÉTRICA

Este livro se concentrou na análise econômética, e não pretendemos fazer uma revisão dos métodos econôméticos nesta seção. No entanto, podemos dar algumas orientações gerais sobre os tipos de problemas que precisam ser considerados em uma análise empírica.

Como explicamos anteriormente, após termos nos decidido por um tópico, devemos coletar um conjunto de dados apropriado. Assumindo que isso também tenha sido feito, precisamos, então, decidir sobre os métodos econôméticos apropriados.

Se seu curso concentrou-se na estimação por mínimos quadrados ordinários de um modelo de regressão linear múltipla, usando dados de corte transversal ou séries temporais, a abordagem econômética praticamente já foi decidida para você. Isso não é necessariamente uma fraqueza, já que o MQO ainda é o método econômético mais amplamente usado. Naturalmente, você ainda terá de decidir se alguma das variantes do MQO — tais como o método de mínimos quadrados ponderados ou a correção da correlação serial em uma regressão de séries temporais — são necessárias.

Para justificar o MQO, você também deve defender de maneira convincente que as hipóteses-chave do MQO são satisfeitas em seu modelo. Como discutimos com alguns detalhes, o primeiro

problema é saber se o termo erro é não-correlacionado com as variáveis explicativas. Idealmente, você terá tido condições de controlar um número suficiente de outros fatores para considerar que aqueles que foram deixados no erro são não relacionados com os regressores. Especialmente quando lidamos com dados de corte transversal em nível de indivíduos, famílias ou empresas, o problema de auto-seleção — que estudamos nos capítulos 7 e 15 — é muitas vezes relevante. No caso do exemplo do PPP da Seção 19.3, pode ser que as famílias com preferência a poupar não observada sejam também as que fazem PPP. Você também deve ter condições de sustentar que as outras fontes potenciais de endogeneidade — especialmente erro de medida e simultaneidade — não são um problema sério.

Ao especificar seu modelo, você também deve tomar decisões sobre a forma funcional. Algumas variáveis devem aparecer na forma logarítmica? (Em aplicações econométricas, a resposta é quase sempre sim.) Algumas variáveis devem ser incluídas em nível e elevadas ao quadrado para possivelmente capturar um efeito decrescente? Como os fatores qualitativos devem aparecer? Será suficiente incluir apenas variáveis binárias para atributos ou grupos diferentes? Ou elas precisarão interagir com variáveis quantitativas? (Veja o Capítulo 7 para mais detalhes.)

Um engano comum, especialmente entre iniciantes, é a inclusão incorreta de variáveis explicativas em um modelo de regressão que estão listadas como valores numéricos, mas não possuem significado quantitativo. Por exemplo, em um conjunto de dados ao nível de indivíduos que contém informação sobre salários, educação, experiência e outras variáveis, uma variável da “ocupação” poderá ser incluída. Caracteristicamente, tratam-se de códigos arbitrários que são atribuídos às diferentes ocupações; o fato que a um professor do ensino de primeiro grau é atribuído, digamos, o valor 453 enquanto a um técnico de computador é atribuído o valor, digamos, 751, é relevante somente porque nos possibilita fazer a distinção entre as duas ocupações. Não faz sentido incluir a variável ocupacional em um modelo de regressão. (Qual seria o sentido de medir o efeito do aumento da *ocupação* em uma unidade quando o aumento de uma unidade não tem significado quantitativo?) Em vez disso, variáveis *dummy* diferentes devem ser definidas para diferentes ocupações (ou grupos de ocupações, se houver muitas ocupações). Em seguida, as variáveis *dummy* poderão ser incluídas no modelo de regressão. Um problema menos grave ocorre quando uma variável qualitativa ordenada é incluída como uma variável explicativa. Suponha que em um conjunto de dados de salários uma variável é incluída para medir a “satisfação no trabalho”, definida em uma escala de 1 a 7, com 7 representando o máximo de satisfação. Desde que tenhamos dados suficientes, gostaríamos de definir um conjunto de seis variáveis *dummy* para, digamos, níveis de satisfação no trabalho de 2 até 7, deixando o nível 1 como o grupo base. Com a inclusão de seis *dummies* da satisfação no trabalho na regressão, possibilitamos um relacionamento totalmente flexível entre a variável de resposta e a satisfação no trabalho. Considerar a variável satisfação no trabalho implicitamente assume que um aumento de uma unidade na variável ordinal tem significado quantitativo. Embora a direção do efeito geralmente seja estimado de maneira apropriada, a interpretação do coeficiente de uma variável ordinal é difícil. Se uma variável ordinal assume muitos valores, então, podemos definir um conjunto de variáveis *dummy* para faixas de valores. Veja um exemplo na Seção 7.3.

Algumas vezes, queremos explicar uma variável que é uma resposta ordinal. Por exemplo, poderíamos pensar em usar uma variável satisfação no trabalho do tipo descrito anteriormente como a variável dependente em um modelo de regressão com as características tanto do trabalhador como do empregador entre as variáveis independentes. Infelizmente, com a variável satisfação no trabalho em sua forma original, os coeficientes no modelo são difíceis de serem interpretados: cada um mede a mudança na satisfação no trabalho face a um aumento unitário na variável independente. Certos modelos — *probit ordenado* e *logit ordenado* — são os mais comuns — são bastante apropriados para respostas ordenadas. Esses modelos essencialmente ampliam os modelos probit e logit binários que examinamos no Capítulo 17. [Veja Wooldridge (2002, capítulo 15), para um tratamento dos modelos de resposta ordenada.] Uma solução simples é transformar qualquer resposta ordenada em uma resposta binária.

Por exemplo, poderíamos definir uma variável igual a um se a satisfação no trabalho for de pelo menos quatro, e zero, caso contrário. Infelizmente, a criação de uma variável binária joga fora informações e exige que usemos um corte algo arbitrário.

Em relação à análise de corte transversal, um problema secundário, porém importante é se houver heteroscedasticidade. No Capítulo 8, explicamos como isso pode ser tratado. A maneira mais simples é computar estatísticas robustas em relação à heteroscedasticidade.

Como enfatizamos nos capítulos 10, 11 e 12, as aplicações de séries temporais exigem atenção adicional. A equação deve ser estimada em níveis? Se níveis forem usados, serão necessárias tendências temporais? A diferenciação dos dados será mais apropriada? Se os dados forem mensais ou trimestrais, a sazonalidade deve ser considerada? Se você estiver possibilitando uma dinâmica — por exemplo, a dinâmica de defasagem distribuída —, quantas defasagens deverão ser incluídas? Você deve iniciar com algumas defasagens com base na intuição ou no bom senso, mas, no final das contas, trate-se de uma questão empírica.

Se seu modelo tiver alguma má especificação potencial, tal como variáveis omitidas, e você usar o MQO, você deve fazer alguma **análise de má especificação** dos tipos que discutimos nos capítulos 3 e 5. Você pode determinar, com base em hipóteses razoáveis, a direção de qualquer viés nos estimadores?

Se você estudou o método de variáveis instrumentais, você sabe que ele pode ser usado para solucionar várias formas de endogeneidade, inclusive variáveis omitidas (Capítulo 15), erros nas variáveis (Capítulo 15) e simultaneidade (Capítulo 16). Naturalmente, você terá que pensar muito para concluir se as variáveis instrumentais que você estiver considerando têm possibilidades de serem válidas.

Bons trabalhos nas ciências sociais empíricas contêm **análise de sensibilidade**. De um modo geral, isso significa que você estima seu modelo original e depois o modifica nas formas que pareçam razoáveis. Espera-se que as conclusões importantes não se alterem. Por exemplo, se você usar como uma variável explicativa um indicador do consumo de álcool (digamos, em uma equação de nota média de um curso de graduação), você obterá resultados qualitativos semelhantes se substituir o indicador de consumo por uma variável *dummy* indicando o consumo de álcool? Se a variável binária de consumo for significante, mas a variável da quantidade de álcool não for, pode ser que o consumo reflete alguns atributos não observados que afetam a nota média e que também sejam correlacionados com o consumo de álcool. Mas isso deve ser considerado caso a caso.

Se algumas observações forem muito diferentes da massa da amostra — digamos, você tem algumas empresas em uma amostra que são muito maiores que as outras —, seus resultados serão muito alterados se você excluir essas observações da estimação? Se assim for, você poderá ter que alterar formas funcionais para levar em conta essas observações ou supor que elas seguem um modelo totalmente diferente. O problema das observações extremas foi explicado no Capítulo 9.

A utilização de dados de painel suscita alguns problemas econometrícicos adicionais. Suponha que você tenha coletado dois períodos. Existem pelo menos quatro maneiras de usar dois períodos de dados de painel sem recorrer às variáveis instrumentais. Você pode reunir os dois anos em uma análise padrão MQO, como explicado no Capítulo 13. Embora isso possa aumentar o tamanho da amostra, comparado com um único corte transversal, isso não controla as variáveis não-observáveis constantes no tempo. Além disso, os erros em tais equações são quase sempre serialmente correlacionados devido a um efeito não observado. A estimação de efeitos aleatórios corrige o problema da correlação serial e produz estimadores assintoticamente eficientes, desde que o efeito não observado tenha média zero, dados os valores das variáveis explicativas em todos os períodos de tempo.

Outra possibilidade é incluir na equação uma variável dependente defasada para o segundo ano. No Capítulo 9, apresentamos essa possibilidade como uma maneira de pelo menos mitigar o problema das variáveis omitidas, já que estamos em qualquer evento mantendo fixo o resultado inicial da variável

dependente. Isso freqüentemente conduz a resultados semelhantes aos da diferenciação dos dados, como tratamos no Capítulo 13.

Com mais anos de dados de painel, temos as mesmas opções, mais uma opção adicional. Podemos usar a transformação de efeitos fixos para eliminar o efeito não observado. (Com dois anos de dados, isso é o mesmo que fazer a diferenciação.) No Capítulo 15, mostramos como as técnicas das variáveis instrumentais podem ser combinadas com as transformações de dados de painel para relaxar ainda mais as hipóteses de exogeneidade. Como regra geral, é uma boa idéia aplicar vários métodos econométricos razoáveis e comparar os resultados. Isso muitas vezes nos possibilita determinar quais de nossas hipóteses pode ser falsa.

Mesmo que você tenha bastante cuidado no planejamento de seu tema, na postulação de seu modelo, na coleta de seus dados e na execução econômética, é bastante possível que você obtenha resultados incompreensíveis — pelo menos algumas vezes. Quando isso acontece, a inclinação natural é tentar modelos diferentes, técnicas diferentes de estimação, ou talvez diferentes subconjuntos de dados, até que os resultados correspondam mais acuradamente com o que era esperado. Virtualmente todos os pesquisadores examinam vários modelos antes de encontrar o “melhor” deles. Infelizmente, a prática da **exploração da base de dados** viola as hipóteses que fizemos em nossa análise econômética. Os resultados sobre a inexistência de viés do MQO e de outros estimadores, como também sobre as distribuições t e F que derivamos para os testes de hipóteses, pressupõem que observamos uma amostra que segue o modelo populacional e que o tenhamos estimado uma vez. A estimativa de modelos que sejam variações de nosso modelo original viola essa hipótese, pois estaremos usando o mesmo conjunto de dados em uma *pesquisa de especificação*. Na verdade, usamos os resultados dos testes com o uso dos dados para reespecificar nosso modelo. As estimativas e testes de diferentes especificações de modelos não são independentes entre si.

Algumas pesquisas de especificação são embutidas em programas econôméticos. Uma bastante comum é conhecida como *regressão stepwise*, na qual diferentes combinações de variáveis explicativas são usadas em análises de regressão múltipla na tentativa de encontrar o melhor modelo. Há várias maneiras de se usar a regressão *stepwise*, e não é nossa intenção examiná-las aqui. A idéia geral é ou começar com um modelo grande e manter as variáveis cujos p -valores estejam abaixo de um certo nível de significância ou começar com um modelo simples e adicionar variáveis que tenham p -valores significantes. Algumas vezes grupos de variáveis são testados com um teste F . Infelizmente, o modelo final freqüentemente depende da ordem em que as variáveis foram retiradas ou adicionadas. [Para saber mais sobre a regressão *stepwise*, veja Draper e Smith (1981).] Além disso, essa é uma forma severa de exploração de base de dados, e é difícil interpretar as estatísticas t e F no modelo final. É possível argumentar que a regressão *stepwise* simplesmente automatiza o que os pesquisadores fazem de qualquer jeito ao pesquisar vários modelos. Porém, na maioria das aplicações, uma ou duas variáveis explicativas são de interesse primordial e, assim, o objetivo é verificar quanto são robustos os coeficientes dessas variáveis no tocante à adição ou eliminação de outras variáveis ou quanto à alteração da forma funcional.

Em princípio, é possível incorporar os efeitos da exploração de base de dados em nossa inferência estatística; na prática, isso é muito difícil e raramente é feito, especialmente em trabalhos empíricos sofisticados. [Veja Leamer (1983), para uma discussão envolvente sobre esse problema.] Entretanto, podemos tentar minimizar a exploração de base de dados fazendo a pesquisa em um número reduzido de modelos ou métodos de estimação, até que um resultado significante seja encontrado e, então, descrevendo somente esse resultado. Se uma variável for estatisticamente significante em somente uma pequena fração dos modelos estimados, é bem provável que a variável não tenha efeito na população.

19.5 A REDAÇÃO DE UM ENSAIO EMPÍRICO

Escrever um ensaio que utiliza análise econométrica é bastante desafiador, mas também pode ser gratificante. Um ensaio bem-sucedido combina uma análise dos dados cuidadosa e convincente com boas explicações e boas interpretações. Portanto, você deve ter um bom domínio de seu tema, boa compreensão dos métodos econômicos, e sólida capacidade de redação. Não se desencoraje se você achar difícil escrever um ensaio empírico; a maioria dos pesquisadores profissionais gastaram muitos anos no aprendizado de como elaborar uma análise empírica e escrever os resultados de uma forma convincente.

Embora os estilos de redação variem, muitos trabalhos seguem as mesmas linhas gerais. Os parágrafos seguintes são idéias de títulos de seções e explicações sobre o que cada seção deve conter. São apenas sugestões e não há a necessidade de segui-las de forma estrita. No trabalho final, a cada seção será dado um número, normalmente começando-se com o número um para a introdução.

Introdução

A introdução expõe os objetivos básicos do estudo e explica os motivos de ele ser importante. Ela geralmente inclui uma revisão da literatura, indicando o que foi feito e como trabalhos anteriores podem ser melhorados. (Como foi discutido na Seção 19.2, uma revisão bibliográfica extensa pode ser colocada em uma seção separada.) A apresentação de estatísticas ou gráficos simples que revelam uma relação aparentemente paradoxal é uma maneira útil de fazer a introdução do tema do trabalho. Por exemplo, suponha que você esteja escrevendo um artigo sobre os fatores que afetam a fertilidade em um país em desenvolvimento, com foco nos níveis de educação das mulheres. Uma maneira interessante de introduzir o tema seria produzir uma tabela ou um gráfico mostrando que a fertilidade vem caindo (digamos) ao longo do tempo e uma breve explicação de como você espera examinar os fatores que contribuem para o declínio. Nesse ponto, você pode já estar sabendo que, *ceteris paribus*, mulheres com níveis mais altos de educação têm menor número de filhos e os níveis médios de educação aumentaram ao longo do tempo.

A maioria dos pesquisadores gosta de resumir os resultados de seus trabalhos na introdução. Isso pode ser um instrumento útil para prender a atenção dos leitores. Por exemplo, você pode declarar que sua melhor estimativa do efeito de perder dez horas de aula durante um curso de 30 horas é de cerca de meio ponto na nota de avaliação de um curso de graduação. Mas o resumo não deve ser muito envolvente, pois nem os métodos nem os dados usados para obter as estimativas foram apresentados ainda.

Estrutura Conceitual (ou Teórica)

Esta é a seção na qual você descreve o método geral para responder à pergunta que você fez. Ela pode ser teoria econômica formal, mas, em muitos casos, é uma discussão intuitiva sobre que problemas conceituais surgem ao responder à sua pergunta.

Como exemplo, suponha que você esteja estudando os efeitos das oportunidades econômicas e a severidade da punição no comportamento criminoso. Um método para explicar a participação na criminalidade é especificar um problema de maximização da utilidade, na qual o indivíduo escolhe o montante de tempo gasto em atividades legais e ilegais, dadas as taxas de remuneração de ambos os tipos de atividade, assim como variáveis indicando a probabilidade e a severidade da punição da atividade criminosa. A utilidade de tal exercício é que ele sugere quais variáveis devem ser incluídas na análise empírica; ele fornece uma orientação (mas raramente detalhes) sobre como as variáveis devem aparecer no modelo econômico.

Freqüentemente, não há necessidade de discorrer sobre uma teoria econômica. Para o plano da análise econométrica, o bom senso normalmente basta para especificar um modelo. Por exemplo, suponha que você esteja interessado em estimar os efeitos da participação na *Aid for Families with Dependent Children* (AFDC), uma campanha de ajuda a famílias com crianças dependentes, sobre os efeitos do desempenho escolar das crianças. A AFCD oferece renda suplementar, mas a participação também torna mais fácil receber assistência médica gratuita e outros benefícios. A parte difícil de tal análise é definir o conjunto de variáveis que devem ser controladas. Nesse exemplo, poderíamos controlar a renda familiar (incluindo a AFCD e qualquer outra renda de assistência social), a educação da mãe, se a família vive em uma área urbana, e outras variáveis. Então, a inclusão de um indicador de participação na AFCD medirá (supostamente) os benefícios não financeiros da participação na AFCD. Uma discussão de quais fatores devem ser controlados e dos mecanismos pelos quais a participação na AFCD pode melhorar o desempenho escolar substitui a teoria econômica formal.

Modelos Econométricos e Métodos de Estimação

É de grande importância ter uma seção que contenha algumas equações daquilo que você estima e apresenta na seção de resultados do trabalho. Isso lhe possibilita fixar as idéias sobre o que são as variáveis explicativas e quais outros fatores você controlará. Escrever equações que contenham termos errôneos permite discutir se um método como o MQO será apropriado.

A distinção entre um *modelo* e um método de estimação deve ser feita nesta seção. Um modelo representa uma relação populacional (definido genericamente para levar em conta equações de séries temporais). Por exemplo, poderíamos escrever

$$nmgrad = \beta_0 + \beta_1 álcool + \beta_2 nmem + \beta_3 sat + \beta_4 feminino + u \quad (19.1)$$

para descrever a relação entre a nota média de graduação e o consumo de álcool, com alguns outros controles na equação. Presumivelmente, essa equação representa uma população, tal como a de todos os alunos de um curso de graduação em uma universidade. Não existem “chapéus” (^) nos β_j ou em $nmgrad$, pois esse é um modelo, e não uma equação estimada. Não colocamos números dos β_j , pois não conhecemos (e nunca conheceremos) esses números. Mais tarde, eles serão estimados. Nesta seção, não antecipe a apresentação de seus resultados empíricos. Em outras palavras, não comece com um modelo geral e depois diga que você omitiu certas variáveis devido ao fato de elas terem se revelado não significantes. Tais explicações devem ser deixadas para a seção de resultados.

Um modelo de série temporal para relacionar roubos de carros em nível de cidades à taxa de desemprego (e outros controles) poderia parecer com

$$roubos_t = \beta_0 + \beta_1 desemp_t + \beta_2 desemp_{t-1} + \beta_3 carros_t + \beta_4 txcond_t + \beta_5 txcond_{t-1} + u_t \quad (19.2)$$

onde o subscrito t é útil para enfatizar qualquer dinâmica na equação (nesse caso, possibilitando que o desemprego e a taxa de condenação por roubo de automóveis tenham efeitos defasados).

Após termos especificado um ou mais modelos, é apropriado explicar os métodos de estimação. Na maioria dos casos, ele será o MQO, mas, por exemplo, em uma equação de séries temporais, você poderá usar o MQG para fazer uma correção da correlação serial (como no Capítulo 12). Porém, o método para estimar um modelo é bastante distinto do modelo em si. Não é significativo, por exemplo,

falar sobre “um modelo MQO”. O método de mínimos quadrados ordinários é um método de estimação, como também o é o método de mínimos quadrados ponderados, o de Cochrane-Orcutt, e assim por diante. Normalmente, existem muitos meios para estimar quaisquer modelos. Você deve explicar por que o método que você escolheu é seguro.

Quaisquer hipóteses que sejam usadas na obtenção de um modelo econômico estimável a partir de um modelo econômico subjacente deve ser discutido com clareza. No exemplo da qualidade das escolas de ensino médio mencionado na Seção 19.1, a questão de como medir a qualidade das escolas é central na análise. Ela deve ser baseada nas notas de avaliação, na percentagem de formados que estão freqüentando a universidade, na razão aluno-professor, no nível médio de educação dos professores, em alguma combinação desses itens, ou possivelmente em outros indicadores?

Sempre temos que fazer hipóteses sobre a forma funcional, tenha ou não sido apresentado um modelo teórico. Como você sabe, modelos de elasticidade constante e de semi-elasticidade constante são atraentes porque os coeficientes são fáceis de serem interpretados (como efeitos percentuais). Não existem regras rígidas quanto à escolha da forma funcional, mas as linhas gerais discutidas na Seção 6.2 parecem funcionar bem na prática. Não há a necessidade de dar uma extensa explicação sobre a forma funcional, mas é interessante mencionar se você estimará elasticidades ou semi-elasticidades. Por exemplo, se você estiver estimando o efeito de alguma variável sobre a remuneração ou o salário, a variável dependente quase com certeza estará na forma logarítmica, e você também poderá incluí-la em qualquer equação desde o início. Você não precisa apresentar todas, nem mesmo a maioria das variações da forma funcional que você descreverá mais tarde na seção de resultados.

Freqüentemente, os dados usados em economia aplicada estão em níveis de cidades ou municípios. Por exemplo, suponha que, para a população de pequenas ou médias cidades, você queira testar a hipótese de que o fato de existir uma equipe de beisebol na liga amadora faz com que a cidade tenha uma taxa menor de divórcios. Nesse caso, você deve considerar o fato de que cidades maiores terão mais divórcios. Uma maneira de levar em conta o tamanho das cidades é escalar os divórcios pela população das cidades ou pela população adulta. Assim, um modelo razoável seria

$$\log(div/pop) = \beta_0 + \beta_1 lbeis + \beta_2 porcat + \beta_3 \log(renda/pop) + outrosfatores, \quad (19.3)$$

onde *lbeis* é uma variável *dummy* igual a um se a cidade tiver uma equipe de beisebol da liga amadora e *porcat* é a percentagem católica da população (portanto, é um número, digamos 34,6, significando 34,6%). Observe que *div/pop* é uma taxa de divórcio, que, de forma geral, é mais fácil de ser interpretada do que o número absoluto de divórcios.

Outra maneira de controlar a população é estimar o modelo

$$\log(div) = \gamma_0 + \gamma_1 lbeis + \gamma_2 porcat + \gamma_3 \log(renda) + \gamma_4 (pop) + outrosfatores. \quad (19.4)$$

O parâmetro de interesse, γ_1 , quando multiplicado por 100, fornece a diferença percentual entre as taxas de divórcio, mantendo constantes a população, a percentagem católica, a renda e o que mais estiver em “outros fatores”. Na equação 19.3, β_1 mede o efeito percentual da liga amadora de beisebol sobre *div/pop*, que pode mudar devido ao número de divórcios ou às mudanças na população. Usando o fato de que $\log(div/pop) = \log(div) - \log(pop)$ e $\log(renda/pop) = \log(renda) - \log(pop)$, podemos reescrever (19.3) como

$$\log(div) = \beta_0 + \beta_1 lbeis + \beta_2 porcat + \beta_3 \log(renda) + (1 - \beta_3) \log(pop) + outros fatores,$$

que mostra que (19.3) é um caso especial de (19.4) com $\gamma_4 = (1 - \beta_3)$ e $\gamma_j = \beta_j$, $j = 0, 1, 2, 3$. Alternativamente, (19.4) é equivalente a adicionar $\log(pop)$ como uma variável explicativa adicional em (19.3). Isso torna fácil testar um efeito populacional separado sobre a taxa de divórcio.

Se você estiver usando um método de estimação mais avançado, como os mínimos quadrados de dois estágios, você precisará fornecer algumas razões sobre a razão de estar fazendo isso. Se usar o MQ2E, você deverá fornecer uma explicação detalhada sobre as razões de suas VIs escolhidas das variáveis explicativas endógenas serem válidas. Como mencionamos no Capítulo 15, existem dois requisitos para que uma variável seja considerada uma boa VI. Primeiro, ela deve ser omitida e exógena na equação de interesse (equação estrutural). Isso é algo que temos que assumir. Segundo, ela deve ter alguma correlação parcial com a variável explicativa endógena. Isso nós podemos verificar. Por exemplo, na equação (19.1), você pode usar uma variável binária para o caso de um aluno morar em um dormitório (*dorm*) como uma VI do consumo de álcool. Isso exige que a condição de moradia não tenha impacto direto sobre *nmgrad* — portanto, ela será omitida de 19.1 — e que seja não-correlacionada com fatores não-observados em *u* que tenham um efeito sobre *nmgrad*. Também teríamos que testar se *dorm* é parcialmente correlacionado com *álcool* fazendo a regressão de *álcool* sobre *dorm*, *nnem*, *sat* e *feminino*. (Veja o Capítulo 15 para mais detalhes.)

Você deve levar em conta o problema de variável omitida (ou heterogeneidade omitida) usando dados de painel. Mais uma vez, isso será facilmente descrito escrevendo uma ou duas equações. Aliás, é útil mostrar como diferenciar as equações ao longo do tempo, para reinover fatores não observáveis constantes no tempo; isso produz uma equação que pode ser estimada por MQO. Ou, se você estiver usando a estimação de efeitos fixos, você simplesmente faz essa declaração.

Como exemplo simples, suponha que você esteja testando se impostos municipais mais altos reduzem a atividade econômica, medida pela produção industrial *per capita*. Suponha que, para os anos de 1982, 1987 e 1992, o modelo seja

$$\log(manuf_{it}) = \beta_0 + \delta_1 d87_t + \delta_2 d92_t + \beta_1 imposto_{it} + \dots + a_i + u_{it},$$

onde *d87*, e *d92*, são variáveis dummy anuais e *imposto_{it}* a alíquota do imposto do município *i* no momento *t* (em percentagem). Teríamos na equação outras variáveis que se alteram ao longo do tempo, incluindo indicadores de custos comerciais (tal como salários médios), indicadores da produtividade do trabalhador (medida pela educação média), e assim por diante. O termo *a_i* é o efeito fixo, contendo todos os fatores que não variam ao longo do tempo e *u_{it}* é o termo erro idiossincrático. Para remover *a_i*, podemos fazer a diferenciação ao longo dos anos ou usar redução temporal (a transformação de efeitos fixos).

Os Dados

Você sempre deve ter uma seção que descreva detalhadamente os dados usados na estimação empírica. Isso é particularmente importante se seus dados não forem padrão ou não tiverem sido amplamente usados por outros pesquisadores. Deve ser apresentada informação bastante para que os leitores possam, em princípio, obter os dados e refazer sua análise. Em particular, todas as fontes públicas de dados aplicáveis devem ser incluídas nas referências, e pequenos conjuntos de dados podem ser listados em um apêndice. Se você fez sua própria pesquisa para obter os dados, uma cópia do questionário deve ser apresentada no apêndice.

Com uma explicação das fontes dos dados, certifique-se de detalhar as unidades de cada uma das variáveis (por exemplo, a renda é medida em centenas ou milhares de dólares?). A inclusão de uma

tabela das definições das variáveis será bastante útil ao leitor. Os nomes na tabela deverão corresponder aos nomes usados na descrição dos resultados econométricos na seção seguinte.

Também será bastante informativo apresentar uma tabela de estatísticas resumidas, tais como valores mínimos e máximos, médias e desvios-padrão de cada variável. A existência de tal tabela torna mais fácil interpretar os coeficientes das estimativas na próxima seção, enfatizando as unidades de medida das variáveis. Para variáveis binárias, o único resumo estatístico necessário é a fração de unidades na amostra (que é igual à média da amostra). Para variáveis com tendência, itens como as médias são menos interessantes. Muitas vezes é útil computar a taxa média de crescimento de uma variável de sua amostra ao longo dos anos.

Você deve sempre deixar claro quantas observações você tem. Para conjuntos de dados de séries temporais, identifique os anos que você esteja usando na análise, incluindo uma descrição de qualquer período especial na história (como a Segunda Guerra Mundial, por exemplo, e se for o caso). Se você usar um conjunto de dados de corte transversal agrupados ou dados de painel, assegure-se de descrever quantas unidades de corte transversal (pessoas, cidades, e assim por diante) você tem de cada ano.

Resultados

A seção de resultados deve incluir suas estimativas de quaisquer modelos formulados na seção de modelos. Você pode começar com uma análise bastante simples. Por exemplo, suponha que a percentagem de alunos freqüentando um curso universitário no último período (*peruniv*) seja usada como um indicador da qualidade da escola de ensino médio que uma pessoa freqüentou. Assim, uma equação a ser estimada seria

$$\log(\text{salário}) = \beta_0 + \beta_1 \text{peruniv} + u.$$

Naturalmente, essa equação não controla vários outros fatores que possam determinar as remunerações e que possam estar correlacionados com *peruniv*. Mas uma análise simples pode levar o leitor até a análise mais sofisticada e revelar a importância de se controlar outros fatores.

Se apenas umas poucas equações forem estimadas, você pode apresentar os resultados em forma de equação com os erros-padrão entre parênteses abaixo dos coeficientes estimados. Se seu modelo tiver diversas variáveis explicativas e você estiver apresentando diversas variações do modelo geral, será melhor descrever os resultados em forma tabular do que na forma de equação. A maioria de vocês deverá ter pelo menos uma tabela, que sempre deverá conter pelo menos o *R*-quadrado e o número de observações de cada equação. Outras estatísticas, como os *R*-quadrados ajustados, também poderão ser listadas.

O mais importante é discutir a interpretação e a solidez de seus resultados empíricos. Os coeficientes têm os sinais esperados? Eles são estatisticamente significantes? Se um coeficiente for estatisticamente significante, mas tiver um sinal não esperado, por que isso pode ser verdadeiro? Ele pode estar revelando um problema com os dados ou com o método econométrico (por exemplo, o MQO pode ser inadequado devido a problemas de variáveis omitidas).

Certifique-se de descrever as *magnitudes* dos coeficientes das principais variáveis explicativas. Muitas vezes, uma ou duas variáveis são fundamentais ao estudo. Seus sinais, magnitude e significância estatística devem ser tratados em detalhe. Lembre-se de fazer a distinção entre significância econômica e estatística. Se uma estatística *t* for pequena, é porque o coeficiente é pequeno na prática, ou porque o erro-padrão é grande?

Além de explicar as estimativas do modelo mais geral, você pode produzir casos especiais interessantes, especialmente aqueles que precisam testar certas hipóteses múltiplas. Por exemplo, em um estudo para determinar os diferenciais salariais entre indústrias, você pode apresentar a equação sem as *dummies* das indústrias; isso permite que o leitor facilmente verifique se os diferenciais das indústrias são estatisticamente significantes (usando a forma *R*-quadrado do teste *F*). Não se preocupe muito em eliminar várias variáveis para encontrar a “melhor” combinação de variáveis explicativas. Como mencionamos anteriormente, essa é uma tarefa difícil e ainda não muito bem definida. Somente se a eliminação de um conjunto de variáveis alterar substancialmente a magnitude e/ou a significância dos coeficientes de interesse, isso será importante. A eliminação de um grupo de variáveis para simplificar o modelo —, tais como termos quadráticos ou de interação — pode ser justificada por meio de um teste *F*.

Se você usou pelo menos dois métodos diferentes — como, por exemplo, o MQO e o MQ2E, ou níveis e diferenciação de uma série temporal, o MQO agrupado *versus* diferenciação com um conjunto de dados de painel —, você deve comentar sobre quaisquer diferenças críticas. Em particular, se o MQO produzir resultados pouco intuitivos, o uso do MQ2E ou de métodos de dados de painel melhora as estimativas?

Conclusões

Esta pode ser uma seção curta que resume o que você descobriu. Por exemplo, você pode querer apresentar a magnitude de um coeficiente no qual você tinha um interesse particular. A conclusão deve também explicar as condições das conclusões apresentadas, e pode até mesmo sugerir orientação para estudos adicionais. É importante imaginar o leitor lendo primeiro as conclusões para então decidir se lerá o resto do documento.

Sugestões de Estilo

Você deve dar ao seu trabalho um título que reflita seu tema. Os trabalhos devem ser datilografados (ou digitados) com espaço duplo. Todas as equações devem começar em uma nova linha e devem ser centralizadas e numeradas em seqüência, isto é, (1), (2), (3), e assim por diante. Gráficos e tabelas grandes podem ser incluídos depois do corpo principal. No texto, refira-se aos trabalhos por autor e data, por exemplo, White (1980). A seção de referências no final do documento deve ser feita em formato padrão. Vários exemplos são dados nas referências, na parte final deste texto.

Quando você apresentar uma equação na seção de modelos econométricos, você deve descrever as variáveis importantes: a variável dependente e a variável ou variáveis independentes cruciais. Para se concentrar em uma única variável independente, você poderia escrever uma equação do tipo

$$nmgrad = \beta_0 + \beta_1 álcool + x\delta + u$$

ou

$$\log(salário) = \beta_0 + \beta_1 educ + x\delta + u,$$

onde a notação $x\delta$ é a forma abreviada de várias outras variáveis explicativas. Nesse momento, você precisa apenas descrever as variáveis genericamente; elas podem ser descritas especificamente em uma tabela, na seção de dados. Por exemplo, em um estudo dos fatores que afetam os salários dos principais diretores executivos de empresas, você poderia incluir a tabela seguinte na seção de dados:

Tabela 1

Descrições das Variáveis

<i>salário</i>	salário anual (inclusive bônus) em 1990 (em milhares)
<i>vendas</i>	vendas da empresa em 1990 (em milhões)
<i>rma</i>	retorno médio da ação de 1988-1990 (em percentagem)
<i>pcsal</i>	percentagem de alteração nos salários de 1988-1990
<i>pcrma</i>	percentagem de alteração no rma de 1988-1990
<i>indust</i>	= 1 se empresa industrial, 0 caso contrário
<i>financeira</i>	= 1 se empresa financeira, 0 caso contrário
<i>prodcons</i>	= 1 se empresa de produtos de consumo, 0 caso contrário
<i>serv</i>	= 1 se empresa de serviços públicos
<i>permceo</i>	número de anos como diretor executivo da empresa

Uma tabela de resumo de estatísticas usando o conjunto de dados do arquivo 401K.RAW, que usamos para estudar os fatores que afetam a participação nos planos de pensão dos Estados Unidos, pode ser montada como segue:

Tabela 2

Resumo das Estatísticas

Variável	Média	Desvio-Padrão	Mínimo	Máximo
<i>taxap</i>	0,869	0,167	0,023	1
<i>taxcont</i>	0,746	0,844	0,011	5
<i>empreg</i>	4.621,01	16.299,64	53	443.040
<i>idade</i>	13,14	9,63	4	76
<i>único</i>	0,415	0,493	0	1
Número de Observações = 3.784				

Na seção de resultados, você pode escrever as estimativas na forma de equação, como temos frequentemente feito, ou em uma tabela. Especialmente quando vários modelos foram estimados com diferentes conjuntos de variáveis explicativas, as tabelas são bastante úteis. Se você escrever as estimativas como uma equação, por exemplo,

$$\begin{aligned} \log(\text{salário}) &= 2,45 + 0,236 \log(\text{vendas}) + 0,008 \text{ rma} + 0,061 \text{ permceo} \\ &\quad (0,93) \quad (0,115) \quad (0,003) \quad (0,028) \\ n &= 204, R^2 = 0,351, \end{aligned}$$

certifique-se de escrever próximo da primeira equação que os erros-padrão estão entre parênteses. É aceitável descrever estatísticas t para testar $H_0: \beta_j = 0$, ou seus valores absolutos, mas é da maior importância explicar o que você está fazendo.

Se você descrever seus resultados na forma tabular, assegure-se de que as variáveis dependentes e independentes estejam claramente indicadas. Novamente, explique se os erros-padrão ou as estatísticas t estão abaixo dos coeficientes (sendo preferido os primeiros). Alguns autores gostam de usar asteriscos para indicar significância estatística para diferentes níveis de significância (por exemplo, uma estrela significa significante a 5%, duas estrelas significa significante a 10%, mas não a 5%, e assim por diante). Isso não será necessário se você explicar cuidadosamente a significância das variáveis explicativas no texto.

Segue uma amostra de tabela de resultados:

Tabela 3: Resultados do MQO

Variável dependente: taxap (percentual de trabalhadores com uma conta ativa em plano de pensão dos Estados Unidos)

Variáveis Independentes			
<i>taxcont</i>	0,156 (0,012)	0,239 (0,042)	0,218 (0,342)
<i>taxcont</i> ²	—	-0,087 (0,043)	-0,096 (0,073)
<i>log(emp)</i>	-0,112 (0,014)	-0,112 (0,014)	-0,098 (0,111)
<i>log(emp)</i> ²	0,0057 (0,0009)	0,0057 (0,0009)	0,0052 (0,0007)
<i>idade</i>	0,0060 (0,0010)	0,0059 (0,0010)	0,0050 (0,0021)
<i>idade</i> ²	-0,00007 (0,00002)	-0,00007 (0,00002)	-0,00006 (0,00002)
<i>único</i>	-0,0001 (0,0058)	0,0008 (0,0058)	0,0006 (0,0061)
<i>constante</i>	1,213 (0,051)	0,198 (0,052)	0,085 (0,041)
Dummies industriais?	Não	Não	Sim
Observações	3.784	3.784	3.784
R-quadrado	0,143	0,152	0,162

Nota: Os números entre parênteses abaixo das estimativas são os erros-padrão.

Seus resultados serão mais fáceis de serem lidos e interpretados se você definir as unidades tanto de suas variáveis dependentes como de suas variáveis independentes, de forma que os coeficientes não

sejam demasiado grandes ou pequenos. Você nunca deve descrever números como $1,051e - 7$ ou $3,524e + 6$ de seus coeficientes ou erros-padrão, e nunca deve usar notação científica. Se os coeficientes forem extremamente pequenos ou extremamente grandes, redimensione as variáveis dependentes ou independentes, como explicamos no Capítulo 6. Você deve limitar o número de dígitos após o ponto decimal. Por exemplo, se seu programa de regressão estimar um coeficiente como 0.54821059, escreva-o como 0,548, ou mesmo 0,55, no trabalho.

Como regra geral, os comandos que seu programa de econometria usa para produzir resultados não devem aparecer no trabalho; somente os resultados são importantes. Se algum comando especial foi usado para conduzir certo método de estimação, isso poderá ser informado em um apêndice, que também é um bom lugar para incluir resultados extras que dêem suporte à sua análise, mas que não sejam fundamentais para ela.

Neste capítulo, discutimos os ingredientes de um estudo empírico bem-sucedido e oferecemos sugestões que podem melhorar a qualidade de uma análise. Em última análise, o sucesso de qualquer estudo depende crucialmente do cuidado e do empenho dedicado a ele.

Em todo este livro, vimos exemplos de análises econométricas que vieram de trabalhos publicados ou foram motivadas por eles. É nossa esperança que eles tenham dado uma boa idéia sobre o escopo da análise empírica. Incluímos a lista seguinte como exemplos adicionais de questões que outras pessoas levantaram ou provavelmente julgaram interessantes. Eles visam estimular sua imaginação; nenhuma tentativa é feita para fornecer todos os detalhes de modelos específicos, requisitos dos dados ou métodos de estimação alternativos. Deve ser possível terminar esses projetos ao longo de um período letivo.

1. Faça sua própria pesquisa no *campus* para responder a uma questão de interesse em sua universidade. Por exemplo: Qual é o efeito de trabalhar sobre a avaliação dos alunos? Você pode perguntar aos estudantes sobre como foi sua avaliação no ensino médio, suas avaliações no curso superior, suas notas de ingresso no curso superior, horas trabalhadas por semana, participação em esportes, horas de estudo, sexo, raça, e assim por diante. Depois, utilize essas variáveis para criar um modelo que explique a avaliação dos alunos em seu curso. Qual o efeito, se houver, de mais uma hora de trabalho por semana sobre essa avaliação? Uma questão de interesse é que horas trabalhadas pode ser uma variável endógena: elas podem estar correlacionadas com fatores não-observados que afetam a avaliação, ou avaliações mais baixas podem fazer com que os alunos trabalhem mais.

Um método melhor seria coletar notas médias acumuladas anteriores ao semestre em curso e depois obter a avaliação do semestre mais recente, com o montante de horas trabalhadas durante esse semestre e as demais variáveis. Agora, a nota média acumulada pode ser usada como um controle (variável explicativa) na equação.

2. Existem muitas variações do tópico precedente. Você poderá estudar os efeitos do uso de drogas ou de álcool, ou de se morar em uma república de estudantes, sobre a nota média no curso de graduação. Você vai querer controlar muitas variáveis da formação familiar, como também variáveis dos desempenhos anteriores.

3. As leis de controle de armas de fogo no nível de cidades reduzem os crimes violentos? Esse tipo de questão pode ser difícil de responder com um único corte transversal porque as leis municipais e estaduais freqüentemente são endógenas. [Veja Kleck e Patterson (1993) para um exemplo. Eles usaram dados de corte transversal e métodos de variáveis instrumentais, mas suas VIs são questionáveis.] Dados de painel podem ser muito úteis para inferir causalidade nesses contextos. No mínimo, você poderia controlar a taxa de crimes violentos de um ano anterior.
 4. Low e McPheters (1983) usaram dados de corte transversal de cidades sobre taxas de remuneração e estimativas do risco de morte de policiais, com outros controles. A idéia era determinar se os policiais eram recompensados por trabalhar em cidades com riscos mais altos de morte ou de se ferir no horário de trabalho.
 5. As leis de consentimento dos pais aumentam as taxas de natalidade na adolescência? Você pode usar dados de nível estadual para este caso: uma série temporal de determinado estado ou, ainda melhor, um conjunto de dados de painel com vários estados. As mesmas leis reduzem as taxas de aborto entre as adolescentes? O *Statistical Abstract of the United States* contém todos os tipos de dados no nível de estados. Levine, Trainor e Zimmerman (1996) estudaram os efeitos das restrições do financiamento a abortos sobre os mesmos resultados. Outros fatores, como o acesso ao aborto, podem afetar a natalidade na adolescência e as taxas de aborto.
 6. As mudanças nas leis de trânsito afetam as fatalidades no tráfego? McCarthy (1994) apresenta uma análise de dados de séries temporais para o estado norte-americano da Califórnia. Um conjunto de variáveis *dummy* pode ser usado para indicar os meses nos quais certas leis entraram em vigor. O arquivo TRAFFIC2.RAW contém os dados usados por McCarthy. Uma alternativa é obter um conjunto de dados de painel dos estados dos Estados Unidos, no qual você poderá explorar variações nas leis entre estados, como também ao longo do tempo. (Veja o arquivo TRAFFIC2.RAW.)
- Mullahy e Sindelar (1994) usaram dados em nível individual compatíveis com leis estaduais e impostos sobre o consumo de álcool para estimar os efeitos das leis e dos impostos sobre a probabilidade de dirigir embriagado.
7. Os negros são discriminados no mercado financeiro? Hunter e Walker (1996) examinaram essa questão; aliás, usamos os dados deles nos Exercícios 7.16 e 17.9.
 8. Existe uma recompensa em ser casado, para os atletas profissionais? Korenman e Neumark (1991) encontraram uma recompensa salarial significante para os homens casados após terem utilizado uma variedade de métodos econométricos. Os atletas profissionais — como os jogadores da *National Basketball Association*, os jogadores da liga principal de beisebol e os golfistas profissionais — formam um grupo interessante para estudar a recompensa por ser casado, pois podemos observar vários indicadores de produtividade. Com jogadores de esportes individuais, como o golfe ou o tênis, os ganhos refletem diretamente a produtividade. Em esportes praticados em equipes, o salário pode não refletir inteiramente a produtividade — por exemplo, o número de anos na liga pode ser significativo. Assim, podemos incluir um indicador de casamento em uma equação com algo como pontos conquistados como a variável dependente, assim como em uma regressão na qual $\log(salário)$ será a variável dependente e vários controles da produtividade estarão entre as variáveis independentes.
 9. Responda a questão: Os fumantes são menos produtivos? Uma variante dessa questão seria: os trabalhadores fumantes faltam mais por doenças (tudo o mais sendo igual)? Mullahy e Portney (1990) usaram dados em nível individual para avaliar essa questão. Você poderia usar dados, digamos, em nível metropolitano. Algo como a produtividade média na produção industrial poderia ser relacionado à percentagem de trabalhadores fumantes. Outras

variáveis, tais como a educação média do trabalhador, o capital por trabalhador e o tamanho da cidade (você pode pensar em outras) deveriam ser controladas.

10. Os salários mínimos reduzem a pobreza? Você poderá usar dados municipais ou estaduais para responder essa pergunta. A idéia é que o salário mínimo varia entre os estados porque alguns estados têm salários mínimos mais altos que o federal. Além disso, ocorrem alterações ao longo do tempo no salário mínimo nominal dentro de um estado, algumas devido a alterações no nível federal e algumas por causa de alterações no nível estadual. Neumark e Wascher (1995) usaram um conjunto de dados de painel de estados para estimar os efeitos do salário mínimo sobre as taxas de emprego de trabalhadores jovens, como também sobre as taxas de matrículas nas escolas.
11. Quais fatores afetam o desempenho dos alunos das escolas públicas? É razoavelmente fácil obter dados no nível das escolas, ou pelo menos no nível dos distritos, na maioria dos estados. O gasto público por aluno é importante? As razões alunos-professores têm algum efeito? É difícil estimar efeitos *ceteris paribus* porque o gasto está relacionado com outros fatores, como, por exemplo, as rendas familiares ou as taxas de pobreza. O conjunto de dados MEAP93.RAW, para as escolas de ensino médio do estado norte-americano de Michigan, contém um indicador das taxas de pobreza. Outra possibilidade é usar dados de painel, ou pelo menos controlar um indicador de desempenho de um ano anterior (como, por exemplo, a nota média ou a percentagem de alunos que foram aprovados em determinado exame).

Você pode examinar fatores menos óbvios que afetam o desempenho do aluno. Por exemplo, após ter a renda controlada, a estrutura familiar importa? Talvez famílias com os dois genitores, mas somente um deles trabalhando com remuneração, tenham um efeito positivo sobre o desempenho. (Poderia haver pelo menos dois canais: os pais passam mais tempo com os filhos, e eles também podem ser voluntários na escola.) E que tal o efeito de lares com só um genitor, com a renda e outros fatores controlados? Você pode combinar dados de censo de um ou dois anos com os dados do distrito escolar.

As escolas públicas com maior número de escolas particulares nas redondezas educam melhor seus alunos devido à concorrência? Existe uma questão ardilosa de simultaneidade nesse caso, pois provavelmente as escolas particulares se localizam em áreas onde haja falta de escolas públicas. Hoxby (1994) usou um método de variáveis instrumentais, no qual as proporções populacionais de várias religiões eram VIs do número de escolas particulares.

Rouse (1998) estudou uma questão diferente: os estudantes que tinham condições de freqüentar uma escola particular devido ao programa de bolsas de estudos de Milwaukee tinham melhor desempenho que os outros? Ela usou dados de painel e teve condições de controlar um efeito estudantil não-observado.

12. O excesso de retorno de uma ação, ou de um índice de ações, pode ser previsto pela razão preço/dividendo defasada? Ou pelas taxas de juros ou mudanças semanais na política monetária, com alguma defasagem? Seria interessante usar um índice de ações estrangeiro, ou um dos menos conhecidos dos Estados Unidos. O trabalho de Cochrane (1997) contém uma boa pesquisa de teorias e resultados empíricos recentes para explicar os excessos de retorno de ações.
13. Existe discriminação racial no mercado de cartões com as figuras dos jogadores de beisebol? Isso exigirá relacionar os preços dos cartões com fatores que possam afetar seus preços, tais como as estatísticas das carreiras dos jogadores, se o jogador está no "hall" da fama, e assim por diante. Mantendo fixos todos os outros fatores, os cartões de jogadores negros ou hispânicos são vendidos por preços mais baixos?
14. Você pode testar se o mercado de apostas esportivas é eficiente. Por exemplo, as cotações dos jogos de futebol ou de basquete contêm todas as informações que podem ser utilizadas

para as apostas? O conjunto de dados PNTSPRD.RAW contém informações sobre jogos de equipes universitárias masculinas de basquetebol. A variável de resultado é binária. As cotações das casas de apostas foram cobertas ou não? Em seguida, você pode tentar encontrar informações que eram conhecidas antes da realização de cada jogo para prever se tais cotações seriam cobertas. (Boa sorte!)

15. Que efeito, se houver, o sucesso no atletismo universitário tem sobre outros aspectos da universidade (candidaturas a cursos, qualidade dos alunos, qualidade dos departamentos não ligados ao atletismo)? McCormick e Tinsley (1987) examinaram os efeitos do sucesso no atletismo nas principais universidades sobre as mudanças nas notas do vestibular dos calouros. A escolha do momento é importante nesse caso: presumivelmente, é o sucesso no passado recente que afeta as candidaturas e a qualidade dos alunos no presente. Deve-se controlar muitos outros fatores — tais como o valor da anuidade escolar e indicadores da qualidade das escolas — para fazer com que a análise seja convincente, pois, sem o controle de outros fatores, haverá uma correlação negativa entre os desempenhos acadêmico e atlético.

Uma variante é combinar rivais naturais no futebol ou no basquetebol masculino e entender as diferenças entre as escolas como uma função de qual escola venceu o jogo de futebol ou um ou mais dos jogos de basquetebol. Os arquivos ATHLET1.RAW e ATHLET2.RAW são pequenos conjuntos de dados que poderiam ser ampliados e atualizados.

16. Reúna dados sobre taxas de criminalidade de uma amostra de cidades ou municípios (digamos, dos relatórios padronizados sobre ocorrências criminais dos órgãos de segurança pública) de dois anos. Modele o último ano de tal forma que seja fácil obter variáveis econômicas e demográficas do *County and City Data Book*. Você pode obter no *Statistical Abstract of the United States* o número de pessoas que estão no corredor da morte, mas as execuções dos anos intervenientes nos níveis dos estados. Se os anos forem 1990 e 1985, você poderá estimar

$$txhom_{i90} = \beta_0 + \beta_1 txhom_{i85} + \beta_2 execuções + outros fatores,$$

onde o interesse estará no coeficiente de *execuções*. A taxa de criminalidade defasada e outros fatores servem de controle.

Outros fatores podem também agir como dissuasores da criminalidade. Por exemplo, Cloninger (1991) apresentou uma análise de corte transversal dos efeitos da resposta letal da polícia sobre as taxas de criminalidade.

Como uma guinada diferente, que fatores afetam a taxa de criminalidade nos *campi* universitários? As frações de estudantes que moram em alojamentos da universidade têm algum efeito? O tamanho da força policial ou o tipo de policiamento usado importam? (Tenha cuidado ao inferir causalidade nesse caso.) A existência de um programa de escolta ajuda a reduzir a criminalidade? E as taxas de criminalidade nas comunidades vizinhas? Recentemente, foi exigido que as instituições de ensino superior passassem a reportar estatísticas criminais. Antigamente isso era feito voluntariamente.

17. Que fatores afetam a produtividade industrial no nível estadual? Além dos níveis de capital e da educação dos trabalhadores, você poderia investigar o grau de sindicalização. Uma análise de dados de painel seria bastante convincente nesse caso, usando dois anos de censos (digamos, 1980 e 1990). Clark (1990) fornece uma análise de como a sindicalização afeta o desempenho e a produtividade das empresas. Que outras variáveis poderiam explicar a produtividade?

Dados no nível das empresas podem ser obtidos do *Compustat*. Por exemplo, mantendo-se fixos outros fatores, as alterações na sindicalização afetam os preços das ações de uma empresa?

18. Utilize dados em níveis estaduais ou municipais ou, se possível, de distritos escolares, para verificar os fatores que afetam o gasto público educacional por aluno. Uma questão interessante seria: outros fatores sendo iguais (tais como renda e níveis de educação dos residentes), os distritos com uma percentagem maior de pessoas mais velhas gastam menos com escolas? Dados de censos podem ser comparados com os dados dos gastos dos distritos para obter um corte transversal bastante grande. O Departamento de Educação dos Estados Unidos compila tais dados.
19. Quais são os efeitos das leis estaduais, tais como as leis sobre o uso obrigatório de capacetes, sobre as fatalidades com motocicletas? Ou as diferenças nas leis de navegação — como, por exemplo, idade mínima para operar uma embarcação — ajudam a explicar os índices de acidentes com barcos? O Departamento de Transportes dos Estados Unidos compila tais informações. Esses dados podem ser mesclados com os do *Statistical Abstract of the United States*. Uma análise de dados de painel parece ser justificada nesse caso.
20. Que fatores afetam o crescimento da produção? Dois fatores de interesse são inflação e investimentos [por exemplo, Blomström, Lipsey e Zejan (1996)]. Você poderá usar dados de séries temporais de um país que julgue interessante. Ou poderia usar um corte transversal de países, como fizeram De Long e Summers (1991). Friedman e Kuttner (1992) encontraram evidências de que, pelo menos nos anos 1980, a diferença entre as taxas dos *commercial papers* e das letras do Tesouro norte-americano afetavam a produção real.
21. Qual é o comportamento das fusões na economia norte-americana (ou em alguma outra economia)? Shughart e Tollison (1984) caracterizaram (o log das) fusões anuais na economia norte-americana como um passeio aleatório ao mostrar que a diferença nos logs — grosso modo, a taxa de crescimento — é imprevisível considerando as taxas de crescimento passadas. Isso ainda é válido? Isso se mantém entre várias empresas? Quais indicadores passados da atividade econômica poderão ser usados para fazer previsões de fusões?
22. Que fatores podem explicar as diferenças raciais e de gênero no emprego e nos salários? Por exemplo, Holzer (1991) analisou a “hipótese da divergência espacial” para explicar as diferenças nos índices de emprego entre negros e brancos. Korenman e Neumark (1992) examinaram os efeitos da gravidez nos salários das mulheres, enquanto Hersch e Stratton (1997) examinaram os efeitos das responsabilidades domésticas sobre os salários dos homens e das mulheres.
23. Obtenha dados mensais ou trimestrais dos índices do trabalho de adolescentes, o salário mínimo e fatores que afetem o trabalho de adolescentes para estimar os efeitos do salário mínimo sobre o emprego de adolescentes. Solon (1985) usou dados trimestrais dos Estados Unidos, enquanto Castillo-Freeman e Freeman (1992) usaram dados anuais de Porto Rico. Pode ser útil analisar dados de séries temporais de um estado com baixos salários dos Estados Unidos, onde as alterações no salário mínimo provavelmente terão os maiores efeitos.
24. Em nível de cidades, estime um modelo de série temporal para a criminalidade. Um exemplo é o trabalho de Cloninger e Sartorius (1979). Como uma variação, você poderia estimar os efeitos do policiamento comunitário ou dos programas de basquetebol noturnos, inovações relativamente recentes no combate ao crime. Inferir causalidade será complicado. A inclusão de uma variável dependente defasada pode ajudar. Como você estará usando dados de séries temporais, você deverá estar prevenido quanto ao problema da regressão espúria. Grogger (1990) utilizou dados sobre a contagem diária de homicídios para estimar o efeito dissuasor da pena capital. Pode haver outros fatores — como, por exemplo, as notícias sobre as respostas letais da polícia — que tenham efeito sobre a contagem diária de homicídios?

25. Há efeitos de produtividade agregada decorrentes do uso do computador? Você teria que obter dados de séries temporais, talvez em nível nacional, sobre a produtividade, percentagem de empregados que usam computadores e outros fatores. E quanto ao dispêndio (provavelmente como uma fração do total das vendas) em pesquisa e desenvolvimento? Que fatores sociológicos podem afetar a produtividade? O uso de álcool? Os índices de divórcios?
26. Que fatores afetam os salários dos diretores executivos de grandes empresas? Os arquivos CEOSAL1.RAW e CEOSAL2.RAW são conjuntos de dados que contêm vários indicadores de desempenhos de empresas, e também informações tais como tempo de permanência no emprego e educação. Você certamente poderá atualizar esses dados e procurar outros fatores interessantes. Rose e Shepard (1997) consideraram a diversificação das empresas como um determinante importante da remuneração dos diretores executivos.
27. As diferenças nos códigos de impostos entre os estados afetam o montante de investimentos estrangeiros diretos? Hines (1996) estudou os efeitos dos impostos corporativos estaduais, com a capacidade de reduzir impostos externos, sobre os investimentos provenientes de fora dos Estados Unidos.
28. Que fatores afetam os resultados das eleições? Os gastos com campanha importam? Os votos devidos a assuntos específicos importam? O estado da economia local importa? Veja, por exemplo, Levitt (1994) e os conjuntos de dados VOTE1.RAW e VOTE2.RAW. Fair (1996) fez uma análise de série temporal sobre as eleições presidenciais dos Estados Unidos.
29. Teste se lojas ou restaurantes praticam discriminação de preços com base na raça ou etnia. Graddy (1997) utilizou dados sobre restaurantes de atendimento expresso em Nova Jersey e Pensilvânia, com as características em nível de códigos de endereçamento postal, para verificar se os preços variavam de acordo com as características da população local. Ela verificou que os preços dos produtos padrão, como, por exemplo, refrigerantes, aumentavam quando a fração de residentes negros aumentava. (Seus dados estão no arquivo DISCRIM.RAW.) Você poderá compilar dados semelhantes em sua área local pesquisando os preços de itens comuns em lojas e restaurantes e comparando-os com dados de censos recentes. Veja o trabalho de Graddy para detalhes da análise.
30. Faça seu próprio estudo de “auditoria” para testar a discriminação de raça ou gênero nas contratações trabalhistas. (Um estudo como esse está descrito no exemplo C.3 do Apêndice 3, no site da Thomson.) Faça com que pares de amigos igualmente qualificados, digamos um homem e uma mulher, se candidatem a vagas de empregos em bares ou restaurantes locais. Você poderá munir-los com currículos falsos que dêem a cada um deles a mesma experiência e formação nas quais a única diferença será o sexo (ou a raça). Em seguida, você poderá acompanhar o processo para saber quem foi chamado para uma entrevista ou recebeu uma oferta de trabalho. Neumark (1996) descreve um trabalho como esse conduzido na Filadélfia. Uma variante seria verificar se a boa aparência, ou uma característica específica, tal como a obesidade ou ter tatuagens ou *piercing* visíveis, tem participação nas decisões de empregar pessoas. Nesse caso, você usaria pares igualados do mesmo sexo, e pode não ser fácil obter voluntários para tal tipo de estudo.

O que segue é uma lista parcial de periódicos populares que contêm pesquisas empíricas comerciais, econômicas e de outras ciências sociais. Um conjunto completo de publicações pode ser encontrado na Internet.

American Economic Review
American Journal of Agricultural Economics
American Political Science Review
Applied Economics
Brookings Papers on Economic Activity
Canadian Journal of Economics
Demography
Economic Development and Cultural Change
Economic Enquiry
Economica
Economics Letters
Empirical Economics
Federal Reserve Bulletin
International Economic Review
International Tax and Public Finance
Journal of Applied Econometrics
Journal of Business and Economics Statistics
Journal of Development Economics
Journal of Economic Education
Journal of Empirical Finance
Journal of Environmental Economics and Management
Journal of Finance
Journal of Health Economics
Journal of Human Resources
Journal of Industrial Economics
Journal of International Economics
Journal of Labor Economics
Journal of Monetary Economics
Journal of Money, Credit, and Banking
Journal of Political Economy
Journal of Public Economics
Journal of Quantitative Criminology
Journal of Urban Economics
National Bureau of Economic Research Working Paper Series
National Tax Journal
Public Finance Quarterly
Quarterly Journal of Economics
Regional Science & Urban Economics
Review of Economic Studies
Review of Economics and Statistics

Numerosas fontes de dados estão disponíveis no mundo todo. Os governos da maioria dos países compilam uma profusão de dados; algumas fontes de dados gerais e de fácil acesso dos Estados Unidos, tais como o *Economic Report of the President*, o *Statistical Abstract of the United States* e o *County*

and City Data Book, já foram mencionadas. Dados financeiros internacionais sobre muitos países são publicados anualmente no *International Financial Statistics*. Várias revistas, como a *Business Week* e a *U.S. News and World Report*, freqüentemente publicam estatísticas — tais como os salários dos diretores executivos e desempenhos de empresas, ou classificação de programas acadêmicos — que são originais e podem ser usados em uma análise econométrica.

Em vez de tentar produzir uma lista aqui, preferimos fornecer alguns endereços na Internet que são fontes abrangentes para os economistas. Um portal bastante útil para economistas, chamado *Resources for Economists on the Internet*, é mantido por Bill Goffe na SUNY, Oswego. O endereço é

<http://www.rfe.org>.

Esse portal fornece links para publicações, fontes de dados e listas de economistas profissionais e acadêmicos. Ele é bastante fácil de usar.

A seção *Business and Economic Statistics* da *American Statistical Association* contém uma lista extremamente detalhada de fontes de dados e fornece links para elas. O endereço é

<http://www.econ-datalinks.org>.

Além disso o *Journal of Applied Econometrics* e o *Journal of Business and Economics Statistics* possuem arquivos de dados que contêm conjuntos de dados usados na maioria dos trabalhos publicados nessas publicações ao longo de vários anos. Se você quer encontrar um conjunto de dados de seu interesse, esse é um bom lugar para ir, pois muito da limpeza e formatação dos dados já foram feitas. A inconveniência é que alguns desses conjuntos de dados são usados em análises econométricas mais avançadas do que as que aprendemos neste livro. De outro lado, muitas vezes é útil estimar modelos mais simples usando métodos econométricos padrão a título de comparação.

Muitas universidades, como a Universidade da Califórnia, em Berkeley, a Universidade de Michigan e a Universidade de Maryland, mantêm conjuntos de dados bastante abrangentes como também links para uma variedade de conjuntos de dados. Sua própria biblioteca possivelmente contém um amplo conjunto de links para bases de dados comerciais, econômicos e de outras ciências sociais. Os bancos centrais regionais dos Estados Unidos, como o de St. Louis, administram uma diversidade de dados. O *National Bureau of Economic Research* publica conjuntos de dados usados por alguns de seus pesquisadores. Os governos federal e estaduais agora publicam uma profusão de dados que podem ser acessados pela Internet. Dados dos censos estão publicamente disponíveis no Departamento do Censo dos Estados Unidos. (Duas publicações bastante úteis são o *Census of Manufacturing* publicado nos anos terminados em 2 e 7, e o *Census of the Population*, publicado no início de cada década.) Outras agências, como o Departamento de Justiça, também colocam dados à disposição do público.

Apêndice G*

Tabelas Estatísticas

TABELA G.1

Áreas Acumuladas sob a Distribuição Normal Padrão

z	0	1	2	3	4	5	6	7	8	9
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451

Continua...

* Os apêndices A a F estão disponíveis na página deste livro no site da Thomson (www.thomsonlearning.com.br). (NE)

TABELA G.1 (continuação)

Áreas Acumuladas Sob a Distribuição Normal Padrão

Z	0	1	2	3	4	5	6	7	8	9
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7474	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

Exemplos: Se $Z \sim \text{Normal}(0,1)$, então, $P(Z \leq -1,32) = 0,0934$ e $P(Z \leq -1,84) = 0,9671$.*Fonte:* Esta tabela foi gerada usando a função normd do Stata[®].

TABELA G.2Valores Críticos da Distribuição t

		Nível de Significância				
Monocaudal:		0,10	0,05	0,025	0,01	0,005
Bicaudal:		0,20	0,10	0,05	0,02	0,01
G r a u s d e L i b e r d a d e	1	3,078	6,314	12,706	31,821	63,657
	2	1,886	2,920	4,303	6,965	9,925
	3	1,638	2,353	3,182	4,541	5,841
	4	1,533	2,132	2,776	3,747	4,604
	5	1,476	2,015	2,571	3,365	4,032
	6	1,440	1,943	2,447	3,143	3,707
	7	1,415	1,895	2,365	2,998	3,499
	8	1,397	1,860	2,306	2,896	3,355
	9	1,383	1,833	2,262	2,821	3,250
	10	1,372	1,812	2,228	2,764	3,169
	11	1,363	1,796	2,201	2,718	3,106
	12	1,356	1,782	2,179	2,681	3,055
	13	1,350	1,771	2,160	2,650	3,012
	14	1,345	1,761	2,145	2,624	2,977
	15	1,341	1,753	2,131	2,602	2,947
	16	1,337	1,746	2,120	2,583	2,921
	17	1,333	1,740	2,110	2,567	2,898
	18	1,330	1,734	2,101	2,552	2,878
	19	1,328	1,729	2,093	2,539	2,861
	20	1,325	1,725	2,086	2,528	2,845
	21	1,323	1,721	2,080	2,518	2,831
	22	1,321	1,717	2,074	2,508	2,819
	23	1,319	1,714	2,069	2,500	2,807
	24	1,318	1,711	2,064	2,492	2,797
	25	1,316	1,708	2,060	2,485	2,787
	26	1,315	1,706	2,056	2,479	2,779
	27	1,314	1,703	2,052	2,473	2,771
	28	1,313	1,701	2,048	2,467	2,763
	29	1,311	1,699	2,045	2,462	2,756
	30	1,310	1,697	2,042	2,457	2,750
	40	1,303	1,684	2,021	2,423	2,704
	60	1,296	1,671	2,000	2,390	2,660
	90	1,291	1,662	1,987	2,368	2,632
	120	1,289	1,658	1,980	2,358	2,617
	∞	1,282	1,645	1,960	2,326	2,576

Exemplos: O valor crítico de 1% de um teste monocaudal com 25 g/ é 2,485. O valor crítico de 5% de um teste bicaudal com g/ grande (> 120) é 1,96.

Fonte: Esta tabela foi gerada usando a função invt do Stata®.

TABELA G.3aValores Críticos de 10% da Distribuição F

		Graus de Liberdade do Numerador									
		1	2	3	4	5	6	7	8	9	10
Graus de Liberdade do Numerador	10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32
	11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25
	12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
	13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
	14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
	15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
	16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
	17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
	18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
	19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
Denominador	20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
	21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92
	22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90
	23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89
	24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88
	25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87
	26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86
	27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85
	28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84
	29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83
Tabela G.3a	30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82
	40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76
	60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71
	90	2,76	2,36	2,15	2,01	1,91	1,84	1,78	1,74	1,70	1,67
	120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65
	∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60

Exemplos: O valor crítico de 10% dos g_1 do numerador = 2 e g_1 do denominador = 40 é 2,44.

Fonte: Esta tabela foi gerada usando a função invFprob do Stata®.

TABELA G.3bValores Críticos de 5% da Distribuição *F*

		Graus de Liberdade do Numerador									
		1	2	3	4	5	6	7	8	9	10
Graus de Liberdade do Numerador	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
Graus de Liberdade do Denominador	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
	27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
	29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
Graus de Liberdade do Denominador	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
	90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94
	120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91
	∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83

Exemplos: O valor crítico de 5% dos gl do numerador = 4 e gl do denominador grande (∞) é 2,37.

Fonte: Esta tabela foi gerada usando a função invFprob do Stata®.

TABELA G.3cValores Críticos de 1% da Distribuição *F*

		Graus de Liberdade do Numerador									
		1	2	3	4	5	6	7	8	9	10
Graus de Liberdade do Numerador	10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
	11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
	12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
	13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
	14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
	15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
	16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
	17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
	18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
	19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
Denominador	20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37
	21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31
	22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26
	23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21
	24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17
	25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13
	26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09
	27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06
	28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03
	29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00
Observações	30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98
	40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80
	60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63
	90	6,93	4,85	4,01	3,54	3,23	3,01	2,84	2,72	2,61	2,52
	120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47
	∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32

Exemplos: O valor crítico de 1% dos gl do numerador = 3 e gl do denominador = 60 é 4,13.Fonte: Esta tabela foi gerada usando a função `invFprob` do Stata®.

TABELA G.4

Valores Críticos da Distribuição Qui-Quadrado

		Nível de Significância		
		0,10	0,05	0,01
G r a u s d e	1	2,71	3,84	6,63
	2	4,61	5,99	9,21
	3	6,25	7,81	11,34
	4	7,78	9,49	13,28
	5	9,24	11,07	15,09
	6	10,64	12,59	16,81
	7	12,02	14,07	18,48
	8	13,36	15,51	20,09
	9	14,68	16,92	21,67
	10	15,99	18,31	23,21
L i b e r d a d e	11	17,28	19,68	24,72
	12	18,55	21,03	26,22
	13	19,81	22,36	27,69
	14	21,06	23,68	29,14
	15	22,31	25,00	30,58
	16	23,54	26,30	32,00
	17	24,77	27,59	33,41
	18	25,99	28,87	34,81
	19	27,20	30,14	36,19
	20	28,41	31,41	37,57
	21	29,62	32,67	38,93
	22	30,81	33,92	40,29
	23	32,01	35,17	41,64
	24	33,20	36,42	42,98
	25	34,38	37,65	44,31
	26	35,56	38,89	45,64
	27	36,74	40,11	46,96
	28	37,92	41,34	48,28
	29	39,09	42,56	49,59
	30	40,26	43,77	50,89

Exemplo: O valor crítico de 5% com $gl = 8$ é 15,51.

Fonte: Esta tabela foi gerada usando a função invchi do Stata®.

Referências Bibliográficas

- ANGRIST, J. D. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review*, v. 80, p. 313-336, 1990.
- ANGRIST, J. D.; KRUEGER, A. B. Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, v. 106, p. 979-1014, 1991.
- ASHENFELTER, O.; KRUEGER, A. B. Estimates of the Economic Return to Schooling from a New Sample of Twins. *American Economic Review*, v. 84, p. 1157-1173, 1994.
- AVERETT, S.; KORENMAN, S. The Economic Reality of the Beauty Myth. *Journal of Human Resources*, v. 31, p. 304-330, 1996.
- AYERS, I.; LEVITT, S. D. Measuring Positive Externalities from Unobservable Victim Precaution: An Empirical Analysis of Lojack. *Quarterly Journal of Economics*, v. 108, p. 43-77, 1998.
- BANERJEE, A. J. et al. *Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press, 1993.
- BARTIK, T. J. The Effects of Property Taxes and Other Local Public Policies on the Intrametropolitan Pattern of Business Location. In: HERZOG, H. W.; SCHLOTTMANN, A. M. (Ed.) *Industry Location and Public Policy*. Knoxville: University of Tennessee Press, 1991. p. 57-80.
- BECKER, G. S. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, v. 76, p. 169-217, 1968.
- BELSLEY, D.; KUH, E.; WELSCH, R. E. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Nova York: Wiley, 1980.
- BERK, R. A. A Primer on Robust Regression. In: FOX, J.; LONG, J. S. (Ed.) *Modern Methods of Data Analysis*. Newbury Park, CA: Sage Publications, p. 292-324, 1990.
- BETTS, J. R. Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth. *Review of Economics and Statistics*, v. 77, p. 231-250, 1995.
- BIDDLE, J. E.; HAMERMESH, D. S. Sleep and the Allocation of Time. *Journal of Political Economy*, v. 98, 1990. p. 922-943.
- _____. Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre. *Journal of Labor Economics*, v. 16, p. 172-201, 1998.
- BLACKBURN, M.; KORENMAN, S. The Declining Marital-Status Earnings Differential. *Journal of Population Economics*, v. 7, p. 247-270, 1994.

- BLACKBURN, M.; NEUMARK, D. Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials. *Quarterly Journal of Economics*, v. 107, p. 1421-1436, 1992.
- BLÖMSTROM, M.; LIPSEY, R. E.; ZEJAN, M. Is Fixed Investment the Key to Economic Growth? *Quarterly Journal of Economics*, v. 111, p. 269-276, 1996.
- BOLLERSLEV, T.; CHOU, R. Y.; KRONER, K. F. ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics*, v. 52, p. 5-59, 1992.
- BOLLERSLEV, T. et al. Models. In: ENGLE, R. F.; McFADDEN, D. L. (Ed.) *Handbook of Econometrics*. Amsterdã: North-Holland, 1994. v. 4, cap. 49, p. 2959-3038.
- BOUND, J.; JAEGER, D. A.; BAKER, R. M. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and Endogenous Explanatory Variables is Weak. *Journal of the American Statistical Association*, v. 90, p. 443-450, 1995.
- BREUSCH, T. S.; PAGAN, A. R. A Simple Test for Heteroskedasticity and Random Coefficient Variation. *Econometrica*, v. 50, p. 987-1007, 1979.
- CAMERON, A. C.; TRIVEDI, P. K. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, 1998.
- CAMPBELL, J. Y.; MANKIW, N. G. Permanent Income, Current Income, and Consumption. *Journal of Business and Economic Statistics*, v. 8, p. 265-279, 1990.
- CARD, D. Using Geographic Variation in College Proximity to Estimate the Return to Schooling. In: CHRISTOPHIDES, L. N.; GRANT, E. K.; SWIDINSKY, R. (Ed.) *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press, 1995. p. 201-222.
- CARD, D.; KRUEGER, A. Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States. *Journal of Political Economy*, v. 100, p. 1-40, 1992.
- CASTILLO-FREEMAN, A. J.; FREEMAN, R. B. When the Minimum Wage Really Bites: The Effect of the U.S.-Level Minimum on Puerto Rico. In: BORJAS, G. J.; FREEMAN, R. B. (Ed.) *Immigration and the Work Force*. Chicago: University of Chicago Press, p. 1992. 177-211.
- CLARK, K. B. Unionization and Firm Performance: The Impact of Profits, Growth, and Productivity. *American Economic Review*, v. 74, p. 893-919, 1984.
- CLONINGER, D. O. Lethal Police Response as a Crime Deterrent: 57-City Study Suggests a Decrease in Certain Crimes. *American Journal of Economics and Sociology*, v. 50, p. 59-69, 1991.
- CLONINGER, D. O.; SARTORIUS, L. C. Crime Rates, Clearance Rates and Enforcement Effort: The Case of Houston, Texas. *American Journal of Economics and Sociology*, v. 38, p. 389-402, 1979.
- COCHRANE, J. H. Where is the Market Going? Uncertain Facts and Novel Theories. *Economic Perspectives*, Federal Reserve Bank of Chicago, v. 21, p. 3-37, 1997.
- CORNWELL, C.; TRUMBULL, W. N. Estimating the Economic Model of Crime Using Panel Data. *Review of Economics and Statistics*, v. 76, p. 360-366, 1994.
- CURRIE, J. *Welfare and the Well-Being of Children*. Chur, Suíça: Harwood Academic Publishers, 1995.
- CURRIE, J.; COLE, N. Welfare and Child Health: The Link Between AFDC Participation and Birth Weight. *American Economic Review*, v. 83, p. 971-983, 1993.

- CURRIE, J.; THOMAS D. Does Head Start Make a Difference? *American Economic Review*, v. 85, p. 341-364, 1995.
- DAVIDSON, R.; MacKINNON, J. G. *Several Tests of Model Specification in the Presence of Alternative Hypotheses*. *Econometrica*, v. 49, p. 781-793, 1981.
- _____. *Estimation and Inference in Econometrics*. Nova York: Oxford University Press, 1993.
- DE LONG, J. B.; SUMMERS, L. H. Equipment Investment and Economic Growth. *Quarterly Journal of Economics*, v. 106, p. 445-502, 1991.
- DICKEY, D. A.; FULLER, W. A. Distributions of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, v. 74, p. 427-431, 1979.
- DIEBOLD, F. X. *Elements of Forecasting*. 2. ed. Cincinnati, OH: South-Western, 2001.
- DOWNES, T. A.; GREENSTEIN, S. M. Understanding the Supply Decisions of Nonprofits: Modeling the Location of Private Schools. *Rand Journal of Economics*, v. 27, p. 365-390, 1996.
- DRAPER, N.; SMITH, H. *Applied Regression Analysis*. 2. ed. Nova York: Wiley, 1981.
- DURBIN, J. Testing for Serial Correlation in Least Squares Regressions When Some of the Regressors are Lagged Dependent Variables. *Econometrica*, v. 38, p. 410-421, 1970.
- DURBIN, J.; WATSON, G. S. Testing for Serial Correlation in Least Squares Regressions I. *Biometrika*, v. 37, p. 409-428, 1950.
- EICKER, F. Limit Theorems for Regressions with Unequal and Dependent Errors. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, v. 1, p. 59-82. Berkeley: University of California Press, 1967.
- EIDE, E. *Economics of Crime: Deterrence and the Rational Offender*. Amsterdã: North-Holland, 1994.
- ENGLE, R. F. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, v. 50, p. 987-1007, 1982.
- ENGLE, R. F.; GRANGER, C. W. J. Co-integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, v. 55, p. 251-276, 1987.
- EVANS, W. N.; SCHWAB, R. M. Finishing High School and Starting College: Do Catholic Schools Make a Difference? *Quarterly Journal of Economics*, v. 110, p. 941-974, 1995.
- FAIR, R. C. Econometrics and Presidential Elections. *Journal of Economic Perspectives*, v. 10, p. 89-102, 1996.
- FRIEDMAN, B. M.; KUTTNER, K. N. Money, Income, Prices, and Interest Rates. *American Economic Review*, v. 82, p. 472-492, 1992.
- GAREN, J. E. Executive Compensation and Principal-Agent Theory. *Journal of Political Economy*, v. 102, p. 1175-1199, 1994.
- GERONIMUS, A. T.; KORENMAN, S. The Socioeconomic Consequences of Teen Childbearing Reconsidered. *Quarterly Journal of Economics*, v. 107, p. 1187-1214, 1992.
- GOLDBERGER, A. S. *A Course in Econometrics*. Cambridge, MA: Harvard University Press, 1991.
- GRADDY, K. Testing for Imperfect Competition at the Fulton Fish Market. *Rand Journal of Economics*, v. 26, p. 75-92, 1995.
- _____. Do Fast-Food Chains Price Discriminate on the Race and Income Characteristics of an Area? *Journal of Business and Economic Statistics*, v. 15, p. 391-401, 1997.

- GRANGER, C. W. J.; NEWBOLD, P. Spurious Regressions in Econometrics. *Journal of Econometrics*, v. 2, p. 111-120, 1974.
- GREENE, W. *Econometric Analysis*. 3. ed. Nova York: MacMillan, 1997.
- GRILICHES, Z. Specification Bias in Estimates of Production Functions. *Journal of Farm Economics*, v. 39, p. 8-20, 1957.
- GROGGER, J. The Deterrent Effect of Capital Punishment: An Analysis of Daily Homicide Counts. *Journal of the American Statistical Association*, v. 410, p. 295-303, 1990.
- _____. Certainty vs. Severity of Punishment. *Economic Inquiry*, v. 29, p. 297-309, 1991.
- HALL, R. E. The Relation Between Price and Marginal Cost in U. S. Industry. *Journal of Political Economy*, v. 96, p. 921-948, 1998.
- HAMERMESH, D. S.; BIDDLE, J. E. Beauty and the Labor Market. *American Economic Review*, v. 84, p. 1174-1194, 1994.
- HAMILTON, J. D. *Time Series Analysis*. Princeton, NJ: Princeton University Press, 1994.
- HANUSHEK, E. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, p. 1141-1177, 1986.
- HARVEY, A. *The Econometric Analysis of Economic Time Series*. 2. ed. Cambridge, MA: MIT Press, 1990.
- HAUSMAN, J. A. Specification Tests in Econometrics. *Econometrica*, v. 46, p. 1251-1271, 1978.
- HAUSMAN, J. A.; WISE, D. A. Social Experimentation, Truncated Distributions, and Efficient Estimation. *Econometrica*, v. 45, p. 319-339, 1977.
- HERRNSTEIN, R. J.; MURRAY, C. *The Bell Curve: Intelligence and Class Structure in American Life*. Nova York: Free Press, 1994.
- HERSCH, J.; STRATTON, L. S. Housework, Fixed Effects, and Wages of Married Workers. *Journal of Human Resources*, v. 32, p. 285-307, 1997.
- HINES, J. R. Altered States: Taxes and the Location of Foreign Direct Investment in America. *American Economic Review*, v. 86, p. 1076-1094, 1996.
- HOLZER, H. The Spatial Mismatch Hypothesis: What Has the Evidence Shown? *Urban Studies*, v. 28, p. 105-122, 1991.
- HOLZER, H. et al. Are Training Subsidies Effective? The Michigan Experience. *Industrial and Labor Relations Review*, v. 46, p. 626-636, 1993.
- HOXBY, C. M. Do Private Schools Provide Competition for Public Schools? *National Bureau of Economic Research Working Paper* n. 4978, 1994.
- HUBER, P. J. The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press, 1967. v. 1, p. 221-233.
- HUNTER, W. C., WALKER, M. B. The Cultural Affinity Hypothesis and Mortgage Lending Decisions. *Journal of Real Estate Finance and Economics*, v. 13, p. 57-70, 1996.
- HYLLEBERG, S. *Modelling Seasonality*. Oxford: Oxford University Press, 1991.
- KANE, T. J.; ROUSE, C. E. Labor-Market Returns to Two and Four-Year Colleges. *American Economic Review*, v. 85, p. 600-614, 1995.

- KIEL, K. A.; McCLAIN K. T. House Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation. *Journal of Environmental Economics and Management*, v. 28, p. 241-255, 1995.
- KLECK, G.; PATTERSON, E. B. The Impact of Gun Control and Gun Ownership Levels on Violence Rates. *Journal of Quantitative Criminology*, v. 9, p. 249-287, 1993.
- KOENKER, R. A Note on Studentizing a Test for Heteroskedasticity. *Journal of Econometrics*, v. 17, p. 107-112, 1981.
- KORENMAN, S.; NEUMARK, D. Does Marriage Really Make Men More Productive? *Journal of Human Resources*, v. 26, p. 282-307, 1991.
- KORENMAN, S.; NEUMARK, D. Marriage, Motherhood, and Wages. *Journal of Human Resources*, v. 27, p. 233-255, 1992.
- KRUEGER, A. B. How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989. *Quarterly Journal of Economics*, v. 108, p. 33-60, 1993.
- KRUPP, C. M.; POLLARD, P. S. Market Responses to Antidumping Laws: Some Evidence from the U.S. Chemical Industry. *Canadian Journal of Economics*, v. 29, p. 199-227, 1996.
- KWIATKOWSKI, D. et al. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We that Economic Time Series Have a Unit Root? *Journal of Econometrics*, v. 54, p. 159-178, 1992.
- LARSEN, R. J.; MARX, M. L. *An Introduction to Mathematical Statistics and Its Applications*. 2. ed. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- LEAMER, E. E. Let's Take the Con Out of Econometrics. *American Economic Review*, v. 73, p. 31-43, 1983.
- LEVINE, P. B.; TRAINOR, A. B.; ZIMMERMAN, D. J. The Effect of Medicaid Abortion Funding Restrictions on Abortions, Pregnancies, and Births. *Journal of Health Economics*, v. 15, p. 555-578, 1996.
- LEVINE, P. B.; ZIMMERMAN, D. J. The Benefit of Additional High-School Math and Science Classes for Young Men and Women. *Journal of Business and Economics Statistics*, v. 13, p. 137-149, 1995.
- LEVITT, S. D. Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House. *Journal of Political Economy*, v. 102, p. 777-798, 1994.
- _____. The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Legislation. *Quarterly Journal of Economics*, v. 111, p. 319-351, 1996.
- LOW, S. A.; MCPHETERS, L. R. Wage Differentials and the Risk of Death: An Empirical Analysis. *Economic Inquiry*, v. 21, p. 271-280, 1983.
- LYNCH, L. M. Private Sector Training and the Earnings of Young Workers. *American Economic Review*, v. 82, p. 299-312, 1992.
- MACKINNON, J. G.; WHITE, H. Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics*, v. 29, p. 305-325, 1985.
- MADDALA, G. S. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.
- MALONEY, M. T.; McCORMICK, R. E. An Examination of the Role that Intercollegiate Athletic Participation Plays in Academic Achievement: Athletes' Feats in the Classroom. *Journal of Human Resources*, v. 28, p. 555-570, 1993.

- MANKIW, N. G. *Macroeconomics*. 2. ed. Nova York: Worth, 1994.
- McCARTHY, P. S. Relaxed Speed Limits and Highway Safety: New Evidence from California. *Economics Letters*, v. 46, p. 173-179, 1994.
- McCLAIN, K. T.; WOOLDRIDGE, J. M. A Simple Test for the Consistency of Dynamic Linear Regression in Rational Distributed Lag Models. *Economics Letters*, v. 48, p. 235-240, 1995.
- McCORMICK, R. E.; TINSLEY, M. Athletics vs. Academics: Evidence from SAT Scores. *Journal of Political Economy*, v. 95, p. 1103-1116, 1987.
- McFADDEN, D. L. Conditional Logit Analysis of Qualitative Choice Behavior. In: ZAREMBKA, P. (Ed.) *Frontiers in Econometrics*. Nova York: Academic Press, 1974. p. 105-142.
- MEYER, B. D. Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics*, v. 13, p. 151-161, 1995.
- MEYER, B. D.; VISCUSI, W. K.; DURBIN, D. L. Workers' Compensation and Injury Duration: Evidence from a Natural Experiment. *American Economic Review*, v. 85, p. 322-340, 1995.
- MIZON, G. E.; RICHARD, J. F. The Encompassing Principle and Its Application to Testing Nonnested Hypotheses. *Econometrica*, v. 54, p. 657-678, 1986.
- MROZ, T. A. The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica*, v. 55, p. 765-799, 1987.
- MULLAHY, J.; PORTNEY, P. R. Air Pollution, Cigarette Smoking, and the Production of Respiratory Health. *Journal of Health Economics*, v. 9, p. 193-205, 1990.
- MULLAHY, J.; SINDELAR, J. L. Do Drinkers Know When to Say When? An Empirical Analysis of Drunk Driving. *Economic Inquiry*, v. 32, p. 383-394, 1994.
- NETZER, D. Differences in Reliance on User Charges by American State and Local Governments. *Public Finance Quarterly*, v. 20, p. 499-511, 1992.
- NEUMARK, D. Sex Discrimination in Restaurant Hiring: An Audit Study. *Quarterly Journal of Economics*, v. 111, p. 915-941, 1996.
- NEUMARK, D.; WASCHER, W. Minimum Wage Effects on Employment and School Enrollment. *Journal of Business and Economic Statistics*, v. 13, p. 199-206, 1995.
- NEWHEY, W. K.; WEST, K. D. A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, v. 55, p. 703-708, 1987.
- PAPKE, L. E. Subnational Taxation and Capital Mobility: Estimates of Tax-Price Elasticities. *National Tax Journal*, v. 40, p. 191-203, 1987.
- _____. Tax Policy and Urban Development: Evidence from the Indiana Enterprise Zone Program. *Journal of Public Economics*, v. 54, p. 37-49, 1994.
- _____. Participation in and Contributions to 401(k) Pension Plans: Evidence from Plan Data. *Journal of Human Resources*, v. 30, p. 311-325, 1995.
- _____. Are 401(k) Plans Replacing Other Employer-Provided Pensions? Evidence from Panel Data. *Journal of Human Resources*, v. 34, p. 346-368, 1999.
- PARK, R. Estimation with Heteroskedastic Error Terms. *Econometrica*, v. 34, p. 888, 1966.
- PAVLIK, E. L.; BELKAOUI, A. *Determinants of Executive Compensation*. Nova York: Quorum Books, 1991.

- PEEK, J. Interest Rates, Income Taxes, and Anticipated Inflation. *American Economic Review*, v. 72, p. 980-997, 1982.
- PINDYCK, R. S.; RUBINFELD, D. L. *Microeconomics*. 2. ed. Nova York: MacMillan, 1992.
- RAM, R. Government Size and Economic Growth: A New Framework and Some Evidence from Cross-Section and Time-Series Data. *American Economics Review*, v. 76, p. 191-203, 1986.
- RAMANATHAN, R. *Introductory Econometrics with Applications*. 3. ed. Fort Worth: Dryden Press, 1995.
- RAMEY, V. Nonconvex Costs and the Behavior of Inventories. *Journal of Political Economy*, v. 99, p. 306-334, 1991.
- RAMSEY, J. B. Tests for Specification Errors in Classical Linear Least-Squares Analysis. *Journal of the Royal Statistical Association, Série B*, v. 71, p. 350-371, 1969.
- ROMER, D. Openness and Inflation: Theory and Evidence. *Quarterly Journal of Economics*, v. 108, p. 869-903, 1993.
- ROSE, N. L. The Incidence of Regulatory Rents in the Motor Carrier Industry. *Rand Journal of Economics*, v. 16, p. 299-318, 1985.
- ROSE, N. L.; SHEPARD, A. Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment? *Rand Journal of Economics*, v. 28, p. 489-514, 1997.
- ROUSE, C. E. Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program. *Quarterly Journal of Economics*, v. 113, p. 553-602, 1998.
- SANDER, W. The Effect of Women's Schooling on Fertility. *Economic Letters*, v. 40, p. 229-233, 1992.
- SAVIN, N. E.; WHITE, K. J. [1977] The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors. *Econometrica*, v. 45, 1989-1996.
- SHEA, J. The Input-Output Approach to Instrument Selection. *Journal of Business and Economic Statistics*, v. 11, p. 145-155, 1993.
- SHUGART, W. F.; TOLLISON, R. D. The Random Character of Merger Activity. *Rand Journal of Economics*, v. 15, p. 500-509, 1984.
- SOLON, G. The Minimum Wage and Teenage Employment: A Re-analysis with Attention to Serial Correlation and Seasonality. *Journal of Human Resources*, v. 20, p. 292-297, 1985.
- STIGLER, S. M. *The History of Statistics*. Cambridge, MA: Harvard University Press, 1986.
- STOCK, J. H.; WATSON, M. W. Interpreting the Evidence on Money-Income Causality. *Journal of Econometrics*, v. 40, p. 161-181, 1989.
- _____. A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems. *Econometrica*, v. 61, p. 783-820, 1993.
- VELLA, F.; VERBEEK, M. Whose Wages Do Unions Raise? A Dynamic Model of Unionism and Wage Rate Determination for Young Men. *Journal of Applied Econometrics*, v. 13, p. 163-183, 1998.
- WALD, A. The Fitting of Straight Lines if Both Variables Are Subject to Error. *Annals of Mathematical Statistics*, v. 11, p. 284-300, 1940.
- WALLIS, K. F. Testing for Fourth-Order Autocorrelation in Quarterly Regression Equations. *Econometrica*, v. 40, p. 617-636, 1972.
- WHITE, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, v. 48, p. 817-838, 1980.

WHITE, H. *Asymptotic Theory for Econometricians*. Orlando: Academic Press, 1984.

WHITE, M. J. Property Taxes and Firm Location: Evidence from Proposition 13. In: ROSEN, H. S. (Ed.) *Studies in State and Local Public Finance*. Chicago: University of Chicago Press, 1986. p. 83-112.

WHITTINGTON, L. A.; ALM, J.; PETERS, H. E. Fertility and the Personal Exemption: Implicit Pronatalist Policy in the United States. *American Economic Review*, v. 80, p. 545-556, 1990.

WOOLDRIDGE, J. M. A Computationally Simple Heteroskedasticity and Serial Correlation-Robust Standard Error for the Linear Regression Model. *Economics Letters*, v. 31, p. 239-243, 1989.

WOOLDRIDGE, J. M. A Note on Computing R^2 -Squared and Adjusted R^2 -Squared for Trending and Seasonal Data. *Economics Letters*, v. 36, p. 49-54, 1991a.

WOOLDRIDGE, J. M. On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances. *Journal of Econometrics*, v. 47, p. 5-46, 1991b.

_____. A Simple Specification Test for the Predictive Ability of Transformation Models. *Review of Economics and Statistics*, v. 76, p. 59-65, 1994a.

_____. Estimation and Inference for Dependent Processes. In: ENGLE, R. F.; McFADDEN, D. L. (Ed.) *Handbook of Econometrics*. Amsterdã: North-Holland, 1994b. v. 4, cap. 45, p. 2639-2738.

_____. Score Diagnostics for Linear Models Estimated by Two Stage Least Squares. In: MADDALA, G. S.; PHILLIPS, P. C. B.; SRINIVASAN, T. N. (Ed.) *Advances in Econometrics and Quantitative Economics*. Oxford: Blackwell, 1995. p. 66-87.

_____. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

Glossário

A

Ajustada Sazonalmente: Dados de séries temporais mensais ou trimestrais nos quais algum procedimento estatístico – possivelmente regressão em variáveis *dummy* sazonais – foi usado para remover o componente sazonal.

Alternativa Bilateral: Uma alternativa na qual o parâmetro populacional pode ser menor ou maior que o valor estabelecido sob a hipótese nula.

Alternativa Unilateral: Uma hipótese alternativa que declara que o parâmetro é maior (ou menor) que o valor hipotético sob a hipótese nula.

Amostra de Aglomeração: Uma amostra de aglomerações ou grupos naturais que habitualmente é composto por pessoas.

Amostra de Pares Comparados: Uma amostra na qual cada observação é comparada com outra, como em uma amostra consistindo de marido e mulher ou um conjunto de dois irmãos.

Amostra Selecionada: Um amostra de dados obtida não pela amostragem aleatória, mas pela seleção com base em alguma característica observada ou não observada.

Amostragem Aleatória: Um esquema de amostragem pelo qual cada observação é extraída aleatoriamente da população. Em particular, nenhuma unidade é mais provável de ser selecionada que qualquer outra, e cada extração é independente de todas as outras.

Amostragem Estratificada: Um esquema de amostragem não-aleatória pelo qual a população primeiro é dividida em diversos estratos completos não-sobrepostos e depois amostras aleatórias são tiradas de cada estrato.

Análise de Decisão Política: Uma análise empírica que utiliza métodos econometrícios para avaliar os efeitos de determinada decisão política.

Análise de Duração: Uma aplicação do modelo de regressão censurada, em que a variável dependente é defasada no tempo até que certo evento ocorra, como por exemplo, o tempo até que uma pessoa desempregada se reempregue.

Análise de Má Especificação: O processo de determinação de prováveis vieses que podem aparecer devido a variáveis omitidas, erro de medida, simultaneidade e outros tipos de má especificação do modelo.

Análise de Regressão Múltipla: Um tipo de análise que é usada para descrever estimação e inferência no modelo de regressão linear múltipla.

Análise de Resíduo: Um tipo de análise que estuda o sinal e o tamanho dos resíduos para observações particulares depois de um modelo de regressão múltipla ter sido estimado.

Análise de Sensibilidade: O processo de verificar se os efeitos estimados e a significância estatística de variáveis explicativas cruciais são sensíveis à inclusão de outras variáveis explicativas, à forma funcional, à eliminação de observações

potencialmente atípicas, ou a diferentes métodos de estimação.

Análise Empírica: Um estudo que usa dados em uma análise econométrica formal para testar uma teoria, estimar uma relação, ou determinar a efetividade de uma determinada decisão.

Armadilha da Variável Dummy: O erro de incluir muitas variáveis *dummy* entre as variáveis independentes; ocorre quando existe no modelo um intercepto global e uma variável *dummy* é incluída para cada grupo.

Arquivo de Texto (ASCII): Um formato de arquivo universal que pode ser transportado para diversas plataformas de computadores.

Assimptoticamente Eficiente: Para estimadores consistentes com distribuições assimptoticamente normais, o estimador com a menor variância assimptótica.

Assimptoticamente Não-Correlacionado: Um processo de série temporal no qual a correlação entre variáveis aleatórias em dois pontos no tempo tende a zero conforme o intervalo temporal entre eles aumenta. (*Ver também* fracamente dependente.)

Autocorrelação de Primeira Ordem: Para um processo de série temporal ordenado cronologicamente, o coeficiente de correlação entre pares de observações adjacentes.

Autocorrelação: *Ver* correlação serial.

Avaliação de Programa: Uma análise de um determinado programa público ou particular usando métodos econométricos para obter o efeito causal do programa.

B

Bancos de Dados On-Line: Bancos de dados que podem ser acessados por meio de uma rede de computadores.

BLUE: *Ver* melhor estimador linear não-viesado.

C

Causalidade de Granger: Uma noção limitada de causalidade na qual valores passados de uma série (x_t) são úteis para prever valores futuros de outra série (y_t), após terem sido controlados os valores passados de y_t .

Censura de Dados: Uma situação que surge quando nem sempre observamos o resultado sobre a variável dependente porque em um limite superior (ou inferior) somente sabemos que o resultado estava acima (ou abaixo) do limite. (*Ver também* modelo de regressão censurada.)

Ceteris Paribus: Todos os outros fatores relevantes são mantidos fixos.

Codificação Superior: Uma forma de censura de dados na qual o valor de uma variável não é registrado quando ele estiver acima de determinado limite; somente saberemos que ele é pelo menos tão grande quanto o limite.

Coeficiente de Correlação: Uma medida da dependência linear entre duas variáveis aleatórias que não depende das unidades de medida e é limitado entre -1 e 1.

Coeficiente de Determinação: *Ver* R-quadrado.

Coeficientes Beta: *Ver* coeficientes padronizados.

Coeficientes Padronizados: Coeficientes de regressão que medem a alteração do desvio-padrão na variável dependente dado um aumento de um desvio padrão em uma variável independente.

Cointegração: A noção de que uma combinação linear de duas séries, cada uma delas integrada de ordem um, é integrada de ordem zero.

Colinearidade Perfeita: Em regressão múltipla, uma variável independente é uma função linear exata de uma ou mais variáveis independentes.

Condição de Classificação: Uma condição suficiente para a identificação de um modelo com uma ou mais variáveis explicativas endógenas.

Condição de Ordem: Uma condição necessária para a identificação dos parâmetros em um mode-

lo com uma ou mais variáveis explicativas endógenas; o número total de variáveis exógenas deve ser pelo menos tão grande quanto o número total de variáveis explicativas.

Condições de Primeira Ordem: O conjunto de equações lineares usado para encontrar as estimativas MQO.

Conjunta e Estatisticamente Significante: A hipótese nula de que duas ou mais variáveis explicativas têm coeficientes populacionais nulos é rejeitada ao nível de significância escolhido.

Conjunto de Dados de Corte Transversal: Um conjunto de dados coletado de uma população em um momento determinado.

Conjunto de Informações: Em projeção, o conjunto de variáveis que podem ser observadas antes de construirmos nossa projeção.

Contemporaneamente Exógeno: Em aplicações de séries temporais ou de dados de painel, um regressor é contemporaneamente exógeno se ele for não-correlacionado com o termo erro no mesmo período de tempo, embora ele possa ser correlacionado com os erros em outros períodos de tempo.

Correlação Amostral: Para resultados de duas variáveis aleatórias, a covariância amostral dividida pelo produto dos desvios-padrão amostrais.

Correlação Espúria: Uma correlação entre duas variáveis que não é devida à causalidade, mas possivelmente à dependência das duas variáveis em relação a outro fator não observado.

Correlação Serial AR(1): Os erros em um modelo de regressão de série temporal seguem um modelo AR(1).

Correlação Serial: Em um modelo de série temporal ou de dados de painel, a correlação entre os erros em diferentes períodos de tempo.

Corte Transversal Agrupado: Uma configuração de dados na qual cortes transversais independentes, normalmente coletados em diferentes pontos do tempo, são combinados para produzir um único conjunto de dados.

Corte Transversal Independentemente Agrupado: Um conjunto de dados obtido pelo agrupamento de amostras aleatórias independentes a partir de diferentes pontos no tempo.

Covariada: Ver variável explicativa.

Covariância Amostral: Um estimador não-viesado da covariância populacional entre duas variáveis aleatórias.

Covariância Estacionária: Um processo de série temporal com média e variância constantes no qual a covariância entre quaisquer duas variáveis aleatórias na sequência depende somente da distância entre elas.

Covariância: Uma medida da dependência linear entre duas variáveis aleatórias.

Critério Dentro da Amostra: Critério para escolher modelos de previsão que são baseados em graus de ajuste no interior da amostra usada para a obtenção das estimativas dos parâmetros.

Critério Fora da Amostra: Critério usado para a escolha de modelos de previsão que são baseados em uma parte da amostra que não foi usada na obtenção das estimativas dos parâmetros.

D

Dados Ausentes: Um problema de dados que ocorre quando não observamos valores de algumas variáveis para certas observações (indivíduos, cidades, períodos de tempo etc.) na amostra.

Dados de Painel: Um conjunto de dados construído de cortes transversais repetidos ao longo do tempo. Com um painel *equilibrado*, as mesmas unidades aparecem em cada período de tempo. Com um painel *não-equilibrado*, algumas unidades não aparecem em cada período de tempo, freqüentemente devido a problemas de redução da amostra.

Dados de Séries Temporais: Dados coletados ao longo do tempo para uma ou mais variáveis.

Dados Experimentais: Dados que foram obtidos pela condução de um experimento controlado.

Dados Longitudinais: Ver dados de painel.

Dados Não-Experimentais: Dados que não foram obtidos por meio de um experimento controlado.

Dados Observacionais: Ver dados não-experimentais.

Dados Quase-Diferenciados: Na estimativa de um modelo de regressão com correlação serial AR(1), é a diferença entre o período de tempo corrente e um múltiplo do período de tempo anterior, na qual o múltiplo é o parâmetro no modelo AR(1).

Dados Quase-Reduzidos: Na estimação de efeitos aleatórios de dados de painel, são os dados originais em cada período de tempo menos uma fração da média temporal; esses cálculos são feitos para cada observação de corte transversal.

Dados Temporais Reduzidos: Dados de painel nos quais, para cada unidade de corte transversal, a média ao longo do tempo é subtraída dos dados em cada período de tempo.

Defasagem Distribuída Geométrica (ou de Koyek): Um modelo de defasagens distribuídas infinitas no qual os coeficientes de defasagem diminuem a uma razão geométrica.

Deslocamento de Intercepto: O intercepto em um modelo de regressão difere por grupo ou período de tempo.

Dessazonalização: A remoção dos componentes sazonais de uma série temporal mensal ou trimestral.

Desvio-Padrão: Uma medida comum de extensão na distribuição de uma variável aleatória.

Desvio-Padrão Amostral: Um estimador consistente do desvio padrão populacional.

Desvio-Padrão de $\hat{\beta}_j$: Uma medida comum de extensão na distribuição amostral de $\hat{\beta}_j$.

Diferença nas Inclinações: Uma descrição de um modelo na qual alguns parâmetros de inclinação podem diferir por grupos ou períodos de tempo.

Distribuição Amostral: A distribuição de probabilidades de um estimador sobre todos os possíveis resultados amostrais.

Distribuição Binomial: A distribuição de probabilidades do número de êxitos em n ensaios de Bernoulli independentes, na qual cada ensaio tem a mesma probabilidade de êxito.

Distribuição Condicional: A distribuição de probabilidades de uma variável aleatória, dados os valores de uma ou mais variáveis aleatórias.

Distribuição Conjunta: A distribuição de probabilidades que determina as probabilidades dos resultados envolvendo duas ou mais variáveis aleatórias.

Distribuição de Defasagens: Em um modelo de defasagens distribuídas finitas ou infinitas, os coeficientes de defasagem desenhados como uma função do tamanho da defasagem.

Distribuição de Dickey-Fuller: A distribuição limite da estatística t no teste da hipótese nula de uma raiz unitária.

Distribuição de Poisson: Uma distribuição de probabilidades de variáveis de contagem.

Distribuição F: A distribuição de probabilidades obtida pela formação da razão de duas variáveis aleatórias qui-quadradas independentes, na qual cada uma foi dividida por seus graus de liberdade.

Distribuição Normal Padrão: A distribuição normal com média zero e variância um.

Distribuição Normal: Uma distribuição de probabilidades comumente usada em estatística e econometria para modelar uma população. Sua função de distribuição de probabilidades tem a forma de um sino.

Distribuição Qui-Quadrado: Uma distribuição de probabilidades obtida pela adição dos quadrados das variáveis aleatórias normais padronizadas independentes. O número de termos na soma iguala o número de graus de liberdade na distribuição.

Distribuição Simétrica: Uma distribuição de probabilidades caracterizada por uma função de densidade de probabilidade que é simétrica em torno de seu valor mediano, que também deve ser o valor médio (sempre que a média existir).

Distribuição t: A distribuição da razão de uma variável aleatória normal padrão e da raiz quadra-

da de uma variável aleatória independente qui-quadrada, na qual a variável aleatória qui-quadrada é primeiro dividida por seus gl .

E

Editor de Texto: Programa de computador que pode ser usado para editar arquivos de texto.

Efeito Causal: Uma alteração *ceteris paribus* em uma variável tem um efeito sobre outra variável.

Efeito de Aglomeração: Um efeito não observado que é comum a todas as unidades, normalmente pessoas, no grupo.

Efeito de Interação: Em regressão múltipla, o efeito parcial de uma variável explicativa depende do valor de uma variável explicativa diferente.

Efeito Marginal Decrescente: O efeito marginal de uma variável explicativa torna-se menor conforme o valor da variável explicativa cresce.

Efeito Marginal: O efeito na variável dependente que resulta de uma pequena alteração em uma variável independente.

Efeito Não Observado: Em um modelo de dados de painel, uma variável não observada no termo erro que não se altera ao longo do tempo. Para amostras de aglomeração, uma variável não observada que é comum a todas as unidades na aglomeração.

Efeito Parcial: O efeito de uma variável explicativa sobre a variável dependente, mantendo-se fixos os outros fatores no modelo de regressão.

Efeitos Fixos: Ver efeitos não observados.

Elasticidade: A alteração percentual em uma variável, dado um aumento *ceteris paribus* de 1% em outra variável.

Elasticidade de Curto Prazo: A propensão de impacto em um modelo de defasagens distribuídas quando as variáveis dependente e independentes estão na forma logarítmica.

Elasticidade de Impacto: Em um modelo de defasagens distribuídas, a alteração percentual

imediata na variável dependente dado um aumento de 1% na variável independente.

Elasticidade de Longo Prazo: A propensão de longo prazo em um modelo de defasagens distribuídas com as variáveis dependente e independentes na forma logarítmica; portanto, a elasticidade de longo prazo é o aumento percentual final na variável explicada, dado um aumento permanente de 1% na variável explicativa.

Endogeneidade: Um termo usado para descrever a presença de uma variável explicativa endógena.

Equação da Forma Reduzida: Uma equação linear na qual uma variável endógena é uma função de variáveis exógenas e erros não observados.

Equação de Primeiras Diferenças: Em modelos de séries temporais ou de dados de painel, uma equação na qual as variáveis dependente e independentes foram todas diferenciadas.

Equação Estrutural: Uma equação derivada de teoria econômica ou de raciocínio econômico menos formal.

Equação Identificada: Uma equação cujos parâmetros podem ser consistentemente estimados, especialmente em modelos com variáveis explicativas endógenas.

Equação Justamente Identificada: Para modelos com variáveis explicativas endógenas, uma equação que é identificada, mas não seria identificada com uma variável instrumental a menos.

Equação Não-Identificada: Uma equação com uma ou mais variáveis explicativas endógenas na qual não existem suficientes variáveis instrumentais para identificar os parâmetros.

Equação Sobreidentificada: Em modelos com variáveis explicativas endógenas, uma equação na qual o número de variáveis instrumentais é estritamente maior que o número de variáveis explicativas endógenas.

Erro Absoluto Médio (EAM): Uma medida de desempenho em previsão calculada como a média dos valores absolutos dos erros de previsão.

Erro da Forma Reduzida: O termo erro que aparece em uma equação da forma reduzida.

Erro de Composição: Em um modelo de dados de painel, a soma do efeito não observado constante no tempo com o erro idiossincrático.

Erro de Medida Multiplicativo: Erro de medida no qual a variável observada é o produto da verdadeira variável não observada e um erro de medida positivo.

Erro de Medida: A diferença entre uma variável observada e a variável que pertence a uma equação de regressão múltipla.

Erro de Previsão: A diferença entre o resultado efetivo e uma previsão desse resultado.

Erro de Projeção: A diferença entre o resultado efetivo e sua previsão.

Erro Estrutural: O termo erro em uma equação estrutural, que pode ser uma equação em um modelo de equações simultâneas.

Erro Idiossincrático: Em modelos de dados de painel, o erro que se altera ao longo do tempo como também ao longo das unidades (digamos, indivíduos, firmas ou cidades).

Erro Quadrático Médio: A distância quadrada esperada que um estimador esteja do valor populacional; ele iguala a variância mais o quadrado de qualquer viés.

Erro Tipo I: Uma rejeição da hipótese nula quando ela é verdadeira.

Erro Tipo II: A falha em rejeitar a hipótese nula quando ela é falsa.

Erro-Padrão Assimptótico: Um erro-padrão que é válido em amostras grandes.

Erro-Padrão da Estimativa: Ver erro-padrão da regressão.

Erro-Padrão da Regressão (EPR): Em análise de regressão múltipla, a estimativa do desvio-padrão do erro populacional, obtida como a raiz quadrada da soma dos resíduos quadrados sobre os graus de liberdade.

Erro-Padrão de $\hat{\beta}_j$: Uma estimativa do desvio-padrão na distribuição amostral de $\hat{\beta}_j$.

Erro-Padrão Robusto em Relação à Correlação Serial: Um erro-padrão de um estimador que é (assimptoticamente) válido sejam, ou não, os erros no modelo serialmente correlacionados.

Erro-Padrão Robusto em Relação à Heteroscedasticidade: Um erro-padrão que é (assimptoticamente) robusto quanto à heteroscedasticidade de forma desconhecida.

Erros Clássicos nas Variáveis (ECV): Um modelo de erro de medida no qual a medida observada iguala a variável efetiva mais um erro de medida independente, ou pelo menos não-correlacionado.

Erros nas Variáveis: Uma situação na qual a variável dependente ou algumas variáveis independentes são medidas com erros.

Estatística de Chow: Uma estatística F para testar a igualdade de parâmetros de regressão entre diferentes grupos (digamos, homens e mulheres) ou períodos de tempo (digamos, antes e depois da alteração de uma regra de decisão).

Estatística de Durbin-Watson (DW): Uma estatística usada para testar correlação serial de primeira ordem nos erros de um modelo de regressão de série temporal sob as hipóteses do modelo linear clássico.

Estatística de Escore: Ver estatística multiplicador de Lagrange.

Estatística de Wald: Um teste estatístico geral para testar hipóteses em uma variedade de cenários econôméticos; em geral, a estatística de Wald tem uma distribuição qui-quadrado assimptótica.

Estatística Descritiva: Uma estatística usada para resumir um conjunto de números; a média amostral, a mediana amostral e o desvio-padrão amostral são os mais comuns.

Estatística F Robusta em Relação à Heteroscedasticidade: Uma estatística do tipo F que é (assimptoticamente) robusta quanto à heteroscedasticidade de forma desconhecida.

Estatística F: Uma estatística usada para testar hipóteses múltiplas sobre os parâmetros em um modelo de regressão múltipla.

Estatística LM Robusta em Relação à Heteroscedasticidade: Uma estatística LM que é robusta quanto à heteroscedasticidade de forma desconhecida.

Estatística Multiplicador de Lagrange: Um teste estatístico com justificação de amostra grande que pode ser usado para testar variáveis omitidas, heteroscedasticidade e correlação serial, entre outros problemas de especificação de modelos.

Estatística Razão de Quase-Verossimilhança: Uma modificação da estatística razão de verossimilhança que leva em conta uma possível má especificação distributiva, como em um modelo de regressão de Poisson.

Estatística Razão de Verossimilhança: Uma estatística que pode ser usada para testar hipóteses únicas ou múltiplas quando os modelos restritos e irrestritos tenham sido estimados por máxima verossimilhança. A estatística será duas vezes a diferença entre as log-verossimilhanças restrita e irrestrita.

Estatística R-n-quadrado: Ver estatística multiplicador de Lagrange.

Estatística t Assimptótica: Uma estatística t que tem uma distribuição normal padrão aproximada em amostras grandes.

Estatística t Robusta em Relação à Heteroscedasticidade: Uma estatística t que é (assimptoticamente) robusta quanto à heteroscedasticidade de forma desconhecida.

Estatística t: A estatística usada para testar uma única hipótese sobre os parâmetros em um modelo econômico.

Estatisticamente Diferente de Zero: Ver estatisticamente significante.

Estatisticamente Não significante: Falha em rejeitar a hipótese nula de que um parâmetro populacional é igual a zero, ao nível de significância selecionado.

Estatisticamente Significante: Rejeição da hipótese nula de que um parâmetro é igual a zero contra a alternativa especificada, ao nível de significância selecionado.

Estimação Agrupada MQO: Estimação por MQO com cortes transversais, dados de painel ou amostras de aglomerações, independentemente agrupados, nas quais as observações são agrupadas ao longo do tempo (ou dos grupos) como também entre as unidades de cortes transversais.

Estimação de Cochrane-Orcutt (CO): Um método de estimar um modelo de regressão linear múltipla com erros AR(1) e variáveis explicativas estritamente exógenas; diferentemente da estimação de Prais-Winsten, a estimação de Cochrane-Orcutt não usa a equação para o primeiro período de tempo.

Estimação de Máxima Verossimilhança (EMV): Um método de estimação largamente aplicável no qual as estimativas dos parâmetros são escolhidas para maximizar a função log-verossimilhança.

Estimação de Prais-Winsten (PW): Um método para estimar um modelo de regressão linear múltipla com erros AR(1) e variáveis explicativas estritamente exógenas; diferentemente da estimação de Cochrane-Orcutt, a estimação de Prais-Winsten usa a equação do primeiro período de tempo na estimação.

Estimação de Quase-Máxima Verossimilhança: Estimação de máxima verossimilhança, mas na qual a função log-verossimilhança pode não corresponder à distribuição condicional efetiva da variável dependente.

Estimador Consistente: Um estimador que converge em probabilidade para o parâmetro populacional, conforme o tamanho da amostra cresce ilimitadamente.

Estimador de Adiantamentos e Atrasos: Um estimador de um parâmetro de cointegração em uma regressão com variáveis I(1) no qual as primeiras diferenças nas variáveis explicativas correntes, algumas passadas e algumas futuras, são incluídas como regressores.

Estimador de Diferenças-em-Diferenças: Um estimador que surge em análises com dados de dois períodos de tempo. Uma versão do estimador aplica-se a cortes transversais agrupados independentes e outra, a conjuntos de dados de painel.

Estimador de Efeitos Aleatórios: Um estimador MQG factível no modelo de efeitos não observados, no qual o efeito não observado é assumido como não-correlacionado com as variáveis explicativas em cada período de tempo.

Estimador de Efeitos Fixos: Para o modelo de dados de painel de efeitos não observados, o estimador obtido pela aplicação dos MQO agrupados a uma equação temporal reduzida.

Estimador de Mínimos Quadrados Generalizados (MQG): Um estimador que leva em conta uma estrutura conhecida da variância dos erros (heteroscedasticidade), do padrão de correlação serial nos erros, ou ambos, por meio de uma transformação do modelo original.

Estimador de Primeiras Diferenças: Em uma estrutura de dados de painel, o estimador MQO agrupado aplicado às primeiras diferenças dos dados ao longo do tempo.

Estimador de Variáveis Instrumentais (VI): Um estimador em um modelo linear usado quando variáveis instrumentais estão disponíveis para uma ou mais variáveis explicativas endógenas.

Estimador do Método dos Momentos: Um estimador obtido com o uso de amostra análoga dos momentos populacionais; mínimos quadrados ordinários e mínimos quadrados em dois estágios são ambos estimadores do método dos momentos.

Estimador dos Mínimos Quadrados em Dois Estágios (MQ2E): Um estimador de variáveis instrumentais no qual a VI de uma variável explicativa endógena é obtida como o valor estimado de uma regressão da variável explicativa endógena sobre todas as variáveis exógenas.

Estimador Interno: Ver estimador de efeitos fixos.

Estimador Linear Não-Viesado: Em análise de regressão múltipla, um estimador não-viesado

que é uma função linear dos resultados da variável dependente.

Estimador MQG Factível (MQGF): Um procedimento MQG no qual os parâmetros de variância ou de correlação são desconhecidos e, portanto, devem primeiramente ser estimados. (Ver também estimador de mínimos quadrados generalizados.)

Estimador Não-Viesado de Variância Mínima: Um estimador com a menor variância na classe de todos os estimadores não-viesados.

Estimador Não-Viesado: Um estimador cujo valor esperado (ou a média de sua distribuição amostral) é igual ao valor populacional (independente do valor populacional).

Estimador por Intervalo: Uma regra que usa dados para obter os limites superior e inferior de um parâmetro populacional. (Ver também intervalo de confiança.)

Estimador Viesado: Um estimador cujo valor esperado, ou média amostral, é diferente do valor populacional que ele supostamente está estimando.

Estimador: Uma regra de combinar dados para produzir um valor numérico de um parâmetro populacional; a forma da regra não depende da amostra particular obtida.

Estimador de Mínimos Quadrados Ponderados (MQP): Um estimador usado para ajustar uma forma conhecida de heteroscedasticidade, em que cada resíduo quadrado é ponderado pelo inverso da variância (estimada) do erro.

Estimativa de Inclinação MQO: Uma inclinação em uma reta de regressão MQO.

Estimativa do Intercepto MQO: O intercepto em uma reta de regressão MQO.

Estimativa: O valor numérico assumido por um estimador para uma amostra particular de dados.

Estritamente Exógena: Uma característica das variáveis explicativas em um modelo de série temporal ou de dados de painel no qual o termo erro em qualquer período de tempo tem expectativa zero, condicional nas variáveis explicativas em todos os períodos de tempo; uma versão

menos restritiva é definida em termos de correlações nulas.

Estudo de Evento: Uma análise econométrica dos efeitos de um evento, tais como uma alteração na regulamentação governamental ou em políticas econômicas, sobre uma variável de resultado.

Exclusão de uma Variável Relevante: Em análise de regressão múltipla, excluir uma variável que tem um efeito parcial diferente de zero sobre a variável dependente.

Exogeneidade Estrita: Uma hipótese que se mantém em um modelo de série temporal ou de dados de painel quando as variáveis explicativas são estritamente exógenas.

Expectativa Condicional: O valor esperado ou médio de uma variável aleatória, chamada de variável dependente ou explicada, que depende dos valores de uma ou mais variáveis, chamadas de variáveis independentes ou explicativas.

Experimento Natural: Uma situação na qual o ambiente econômico – algumas vezes resumido por uma variável explicativa – é alterado de forma exógena, talvez inadvertidamente, devido a uma alteração política ou institucional.

Experimento: Em probabilidade, um termo geral usado para representar um evento cujo resultado é impreciso. Em análise econômica, ele representa uma situação na qual os dados são coletados atribuindo aleatoriamente indivíduos a grupos de controle e de tratamento.

Exploração da Base de Dados: A prática de se usar o mesmo conjunto de dados para estimar numerosos modelos em uma busca para encontrar o “melhor” modelo.

F

Forma R -quadrada da Estatística F : A estatística F para testar restrições de exclusões expressas em termos dos R -quadrados dos modelos restrito e irrestrito.

Fortemente Dependente: Ver processo altamente persistente.

Fracamente Dependente: Um termo que descreve um processo de série temporal no qual algum grau de dependência entre variáveis aleatórias em dois pontos do tempo – como a correlação – diminui conforme o intervalo entre os dois pontos no tempo aumenta.

Freqüência dos Dados: O intervalo no qual dados de séries temporais são coletados. Anuais, trimestrais e mensais são as freqüências de dados mais comuns.

Função de Densidade de Probabilidade (fdp): Uma função que, para variáveis aleatórias discretas, fornece a probabilidade de que a variável aleatória assuma para cada valor; para variáveis aleatórias contínuas, a área sob a fdp fornece a probabilidade de vários eventos.

Função de Distribuição Cumulativa (fdc): Uma função que fornece a probabilidade de uma variável aleatória ser menor que ou igual a qualquer número real especificado.

Função de Perda: Uma função que mede a perda quando uma previsão difere do resultado efetivo; os exemplos mais comuns são as perdas de valor absoluto e as perdas de quadrados.

Função de Regressão Amostral: Ver reta de regressão MQO.

Função de Regressão Populacional: Ver expectativa condicional.

Função Exponencial: Um função matemática definida para todos os valores que tenham uma inclinação crescente, mas uma mudança proporcional constante.

Função Linear: Uma função na qual a alteração na variável dependente, dado uma alteração de uma unidade em uma variável independente, é constante.

Função Logarítmica: Uma função matemática definida por argumentos positivos, que tem uma inclinação positiva, mas decrescente.

Função Log-Verossimilhança: A soma das log-verossimilhanças, na qual a log-verossimilhança de cada observação é o log da densidade da variável dependente dadas as variáveis independentes;

a função log-verossimilhança é vista como uma função dos parâmetros a serem estimados.

Função Não-Linear: Uma função cuja inclinação não é constante.

Funções Quadráticas: Funções que contêm quadrados de uma ou mais variáveis explicativas; elas capturam os efeitos decrescentes ou crescentes sobre a variável dependente.

G

Graus de Liberdade (gl): Em análise de regressão múltipla, o número de observações menos o número de parâmetros estimados.

Graus de Liberdade do Denominador: Em um teste F , os graus de liberdade no modelo irrestrito.

Graus de Liberdade do Numerador: Em um teste F , o número de restrições sendo testadas.

Grupo Base: O grupo representado pelo intercepto global em um modelo de regressão múltipla que inclui variáveis *dummy* explicativas.

Grupo de Controle: Em avaliação de programas (econômicos ou sociais), o grupo que não participa do programa.

Grupo de Tratamento: Em avaliação de programas, o grupo que participa do programa. (Ver também grupo experimental.)

Grupo Experimental: Ver grupo de tratamento.

Grupo Referencial: Ver grupo base.

H

Heterogeneidade Não Observada: Ver efeito não observado.

Heteroscedasticidade Condicional Auto-Regressiva (ARCH): Um modelo de heteroscedasticidade dinâmica no qual a variância do termo erro, em decorrência de informações passadas, depende linearmente dos erros quadrados passados.

Heteroscedasticidade de Forma Desconhecida: Heteroscedasticidade que pode depender das variáveis explicativas de uma maneira arbitrária e desconhecida.

Heteroscedasticidade: A variância do termo erro, dadas as variáveis explicativas, não é constante.

Hipótese Alternativa: A hipótese contra a qual é testada a hipótese nula.

Hipótese de Média Condisional Zero: Uma hipótese-chave usada em análise de regressão múltipla que estabelece que, dados quaisquer valores das variáveis explicativas, o valor esperado do erro será igual a zero. (Ver as hipóteses RLM.3, ST.2, e ST.2'.)

Hipótese de Normalidade: A hipótese do modelo linear clássico que afirma que o erro (ou a variável dependente) tem uma distribuição normal, condicional nas variáveis explicativas.

Hipótese Nula: No teste de hipótese clássico, tornamos esta hipótese como verdadeira e exigimos que os dados forneçam evidência substancial contra ela.

Hipóteses de Gauss-Markov: O conjunto de hipóteses (hipóteses RLM.1 a RLM.5 ou ST.1 a ST.5) sob as quais os MQO são BLUE.

Hipóteses do Modelo Linear Clássico (MLC): O conjunto ideal de hipóteses da análise de regressão múltipla: para análise de corte transversal, hipóteses RLM.1 até RLM.6 e para análise de séries temporais hipóteses ST.1 até ST.6. As hipóteses englobam linearidade nos parâmetros, ausência de colinearidade perfeita, a hipótese de média condicional zero, homoscedasticidade, ausência de correlação serial e normalidade dos erros.

Homoscedasticidade Contemporânea: Em aplicações de séries temporais ou de dados de painel, a variância do termo erro, condicional nos regressores no mesmo período de tempo, é constante.

Homoscedasticidade: Os erros em um modelo de regressão têm variância constante, condicional nas variáveis explicativas.

I

Inclusão de uma Variável Irrelevante: A inclusão de uma variável explicativa em um modelo de regressão que tem um parâmetro populacional zero ao estimar uma equação por MQO.

Inconsistência: A diferença entre o limite de probabilidade de um estimador e o valor do parâmetro.

Inferência Estatística: O ato de testar hipóteses sobre os parâmetros populacionais.

Integrado de Ordem Um [I(1)]: Um processo de série temporal que precisa passar por uma primeira diferenciação para produzir um processo I(0).

Integrado de Ordem Zero [I(0)]: Um processo estacionário de série temporal fracamente dependente que, quando usado em análise de regressão, satisfaz a lei dos grandes números e o teorema do limite central.

Internet: Uma rede mundial de computadores que pode ser usada para acessar informações e baixar bases de dados.

Intervalo de Confiança (IC): Uma regra usada para construir um intervalo aleatório, de forma que certa percentagem de todos os conjuntos de dados, determinada pelo nível de confiança, produz um intervalo que contém o valor populacional.

Intervalo de Confiança Assimptótico: Um intervalo de confiança que é aproximadamente válido em amostras de tamanhos grandes.

Intervalo de Previsão: Um intervalo de confiança para um resultado desconhecido sobre uma variável dependente em um modelo de regressão múltipla.

Intervalo de Projeção: Em projeção, um intervalo de confiança de um valor futuro ainda não realizado de uma variável de série temporal. (*Ver também* intervalo de previsão.)

Inverso da Razão de Mills: Um termo que pode ser adicionado a um modelo de regressão múltipla para remover o viés de seleção amostral.

L

Lei das Expectativas Iteradas: Um resultado da probabilidade que relaciona expectativas incondicionais e condicionais.

Lei dos Grandes Números (LGN): Um teorema que diz que a média de uma amostra aleatória converge em probabilidade para a média populacional; a LGN também é válida para séries temporais estacionária e fracamente dependentes.

Limite de Probabilidade: O valor para o qual um estimador converge conforme o tamanho da amostra cresce sem limites.

Logaritmo Natural: Ver função logarítmica.

M

Má Especificação da Forma Funcional: Um problema que ocorre quando um modelo tem funções omitidas das variáveis explicativas (como por exemplo termos quadráticos) ou usa as funções erradas da variável dependente ou de algumas das variáveis explicativas.

Martingale: Um processo de série temporal cujo valor esperado, dados todos os resultados passados da série, simplesmente iguala o valor mais recente.

Matriz: Um arranjo de números.

Média Amostral: A soma de n números dividida por n ; uma medida de tendência central.

Média: A soma de n números dividida por n .

Média: Ver valor esperado.

Mediana: Em uma distribuição de probabilidades, ela é o valor no qual existe 50% de possibilidade de estar abaixo do valor e 50% de possibilidade de estar acima dele. Em uma amostra de números é o valor do meio após os números terem sido ordenados.

Medida do Grau de Ajuste: Uma estatística que resume o quanto um conjunto de variáveis expli-

cativas explica uma variável dependente ou de resposta.

Melhor Estimador Linear Não-Viesado: Entre todos os estimadores lineares não-viesados, o estimador com menor variância. O MQO é o melhor estimador linear não-viesado, condicional aos valores amostrais das variáveis explicativas sob as hipóteses de Gauss-Markov.

Método de Heckit: Um procedimento econômico usado para corrigir o viés de seleção amostral devido a truncamento incidental ou alguma outra forma de dados ausentes não-aleatoriamente.

Micronumerosidade: Um termo apresentado por Arthur Goldberger para descrever propriedades de estimadores econômicos com amostras de tamanho pequeno.

Mínimos Desvios Absolutos: Um método para se estimar os parâmetros de um modelo de regressão múltipla com base na minimização da soma dos valores absolutos dos resíduos.

Mínimos Quadrados Ordinários (MQO): Um método para estimar os parâmetros de um modelo de regressão múltipla. As estimativas dos mínimos quadrados ordinários são obtidas pela minimização da soma dos quadrados dos resíduos.

Modelo Auto-Regressivo Vetorial (VAR): Um modelo de duas ou mais séries temporais no qual cada variável é modelada como uma função linear dos valores passados de todas as variáveis, mais perturbações que têm média zero, dados todos os valores passados das variáveis observadas.

Modelo de Correção de Erros: Um modelo de série temporal em primeiras diferenças que também contém um termo de correção de erro, que serve para introduzir duas séries I(1) no equilíbrio de longo prazo.

Modelo de Defasagem Distribuída Racional (DDR): Um tipo de modelo de defasagens distribuídas infinitas no qual a distribuição defasada depende de relativamente poucos parâmetros.

Modelo de Defasagens Distribuídas Finitas (DDF): Um modelo dinâmico no qual uma ou mais variáveis explicativas podem ter efeitos defasados sobre a variável dependente.

Modelo de Defasagens Distribuídas Infinitas (DDI): Um modelo de defasagens distribuídas no qual uma alteração na variável explicativa pode ter um impacto sobre a variável dependente no futuro indefinido.

Modelo de Defasagens Distribuídas: Um modelo de série temporal que relaciona a variável dependente a valores correntes e passados de uma variável explicativa.

Modelo de Efeitos Aleatórios: O modelo de dados de painel de efeitos não observados no qual o efeito não observado é assumido como não-correlacionado com as variáveis explicativas em cada período de tempo.

Modelo de Efeitos Não Observados: Um modelo de dados de painel ou de amostras de aglomeração cujo termo erro contém um efeito não observado.

Modelo de Elasticidade Constante: Um modelo no qual a elasticidade da variável dependente, em relação a uma variável explicativa, é constante; em regressão múltipla, ambas as variáveis aparecem na forma logarítmica.

Modelo de Equações Simultâneas (MES): Um modelo que determina conjuntamente duas ou mais variáveis endógenas, no qual cada variável endógena pode ser uma função de outras variáveis endógenas como também de variáveis exógenas e de um termo erro.

Modelo de Probabilidade Linear (MPL): Um modelo de resposta binária no qual a probabilidade de resposta é linear em seus parâmetros.

Modelo de Regressão Bivariada: Ver modelo de regressão linear simples.

Modelo de Regressão Censurada: Um modelo de regressão múltipla no qual a variável dependente foi censurada acima ou abaixo de algum valor limite conhecido.

Modelo de Regressão de Poisson: Um modelo de uma variável dependente de contagem no qual a variável dependente, condicional nas variáveis explicativas, é nominalmente assumida como tendo uma distribuição de Poisson.

Modelo de Regressão Linear Múltipla (RLM): Um modelo linear em seus parâmetros, no qual a variável dependente é uma função de variáveis independentes mais um termo erro.

Modelo de Regressão Linear Simples: Um modelo no qual a variável dependente é uma função linear de uma única variável independente, mais um termo erro.

Modelo de Regressão Truncada: Um modelo clássico de regressão linear de dados de corte transversal no qual o esquema de amostragem exclui inteiramente, com base nos resultados da variável dependente, parte da população.

Modelo de Resposta Binária: Um modelo de uma variável dependente binária (*dummy*).

Modelo de Variável Latente: Um modelo no qual a variável dependente observada é suposta como sendo uma função de uma variável subjacente latente, ou não observada.

Modelo Dinamicamente Completo: Um modelo de série temporal no qual nenhuma defasagem adicional da variável dependente ou das variáveis explicativas ajudará a explicar a variável dependente.

Modelo Econométrico: Uma equação relacionando a variável dependente a um conjunto de variáveis explicativas e distúrbios não observados, quando parâmetros populacionais desconhecidos determinam o efeito *ceteris paribus* de cada variável explicativa.

Modelo Econômico: Uma relação derivada da teoria econômica ou de um raciocínio econômico menos formal.

Modelo Estático: Um modelo de série temporal no qual somente variáveis explicativas contemporâneas afetam a variável dependente.

Modelo Irrestrito: Em teste de hipóteses, o modelo no qual não foi colocada nenhuma restrição em seus parâmetros.

Modelo Linear Clássico: O modelo de regressão linear múltipla sob o conjunto completo das hipóteses do modelo linear clássico.

Modelo Logit: Um modelo de resposta binária no qual a probabilidade de resposta é a função logit avaliada em uma função linear das variáveis explicativas.

Modelo Log-Log: Um modelo de regressão no qual a variável dependente e as variáveis independentes (pelo menos algumas) estão na forma logarítmica.

Modelo Log-Nível: Um modelo de regressão no qual a variável dependente está na forma logarítmica e as variáveis independentes estão na forma em nível (ou original).

Modelo Nível-Log: Um modelo de regressão no qual a variável dependente está na forma em nível e (pelo menos algumas das) variáveis independentes estão na forma logarítmica.

Modelo Nível-Nível: Um modelo de regressão no qual a variável dependente e as variáveis independentes estão na forma em nível (ou original).

Modelo Parcimonioso: Um modelo com o menor número possível de parâmetros necessários para capturar quaisquer características almejadas.

Modelo Populacional: Um modelo, especialmente um modelo de regressão linear múltipla, que descreve uma população.

Modelo Probit: Um modelo de resposta binária no qual a probabilidade de resposta é a fdc normal padrão avaliada em uma função linear das variáveis explicativas.

Modelo Restrito: Em teste de hipóteses, o modelo obtido após a imposição de todas as restrições requeridas sob a hipótese nula.

Modelo Tobit: Um modelo de uma variável dependente que assume o valor zero com probabilidade positiva, mas, de forma aproximada é continuamente distribuída sobre valores estritamente positivos. (Ver também resposta de solução de canto.)

Modelo Verdadeiro: O modelo populacional efetivo relacionando a variável dependente às variáveis independentes relevantes, mais uma perturbação, em que a hipótese da média condicional zero se mantém.

Modelos Não-Aninhados: Dois (ou mais) modelos nos quais nenhum deles pode ser escrito como um caso especial do outro pela imposição de restrições nos parâmetros.

MQO: Ver mínimos quadrados ordinários.

Mudança em Pontos Percentuais: A mudança em uma variável que é medida como um percentual.

Mudança Percentual: A mudança proporcional em uma variável, multiplicada por 100.

Mudança Proporcional: A mudança em uma variável relativa ao seu valor inicial; matematicamente, a mudança dividida por seu valor inicial.

Mudança Relativa: Ver mudança proporcional.

Multicolinearidade: Um termo que se refere à correlação entre as variáveis independentes em um modelo de regressão múltipla; ela é habitualmente invocada quando alguma correlação é “grande”, mas uma magnitude efetiva não está bem definida.

Multiplicador de Impacto: Ver propensão de impacto.

Multiplicador de Longo Prazo: Ver propensão de longo prazo.

N

Nível de Confiança: A percentagem de amostras na qual queremos que nosso intervalo de confiança contenha o valor populacional; 95% é o intervalo de confiança mais comum, mas os de 90% e 99% também são usados.

Nível de Significância: A probabilidade de erro Tipo I em testes de hipóteses.

Normalidade Assimptótica: A distribuição amostral de um estimador adequadamente normalizado converge para a distribuição normal padrão.

Notação Matricial: Uma conveniente notação matemática, fundamentada na álgebra matricial, para expressar e manipular o modelo de regressão múltipla.

Número Índice: Uma estatística que agrupa informações sobre a atividade econômica, tal como produção ou preços.

O

Observações Atípicas: Ver observações extremas.

Observações Extremas: Observações em um conjunto de dados que são substancialmente diferentes da massa de dados, possivelmente devido a erros ou devido a alguns dados terem sido gerados por um modelo diferente daquele que gerou a maioria dos outros dados.

Operador Somatório: Uma notação, representada por Σ , usada para definir a soma de um conjunto de números.

P

Painel Desequilibrado: Um conjunto de dados de painel em que dados de certos anos (ou períodos) estão faltando para algumas unidades de corte transversal.

Painel Equilibrado: Um conjunto de dados de painel no qual todos os anos (ou períodos) de dados estão disponíveis para todas as unidades de corte transversal.

Parâmetro de Inclinação: O coeficiente de uma variável independente em um modelo de regressão múltipla.

Parâmetro de Intercepto: O parâmetro em uma regressão linear múltipla que fornece o valor esperado da variável dependente quando todas as variáveis independentes são iguais a zero.

Parâmetro: Um valor desconhecido que descreve uma relação populacional.

Parâmetros da Forma Reduzida: Os parâmetros que aparecem em uma equação da forma reduzida.

Parâmetros Estruturais: Os parâmetros que aparecem em uma equação estrutural.

Passeio Aleatório com Tendência: Um passeio aleatório que tem uma constante (ou tendência) adicionada em cada período.

Passeio Aleatório: Um processo de série temporal no qual o valor do próximo período é obtido como o valor deste período mais um termo erro independente (ou pelo menos não-correlacionado).

Percentagem Corretamente Prevista: Em um modelo de resposta binária, o percentual de vezes em que a previsão de zero ou um coincide com o resultado efetivo.

Período Base: Para números índices, tais como os índices de preços ou de produção, o período contra o qual todos os outros períodos de tempo são comparados.

Perturbação: Ver termo erro.

Planilha: Programa de computador usado para entrada e manipulação de dados.

Poder de um Teste: A probabilidade de rejeitar a hipótese nula quando ela for falsa; o poder depende dos valores dos parâmetros populacionais sob a hipótese alternativa.

População: Um grupo bem definido (de pessoas, firmas, cidades etc.) que é o foco de uma análise estatística ou econômica.

Previsão com Múltiplos Passos à Frente: Uma previsão de série temporal de mais de um período de tempo no futuro.

Previsão com um Passo à Frente: Uma previsão de série temporal de um período de tempo no futuro.

Previsão Condicional: Uma previsão que assume serem os valores futuros de algumas variáveis explicativas conhecidos com certeza.

Previsão Incondicional: Uma previsão que não depende de conhecer, ou assumir, os valores futuros das variáveis explicativas.

Previsão por Ponto: O valor previsto de um resultado futuro.

Previsão: A estimativa de um resultado obtida pela agregação de valores específicos das variá-

veis explicativas em um modelo estimado, em geral um modelo de regressão múltipla.

Primeira Diferença: Uma transformação em uma série temporal levando em conta a diferença de períodos de tempo adjacentes, em que o período de tempo anterior é subtraído do mais recente.

Probabilidade de Resposta: Em um modelo de resposta binária, a probabilidade de que a variável dependente assuma o valor um, condicional nas variáveis explicativas.

Problema da Regressão Espúria: Um problema que surge quando a análise de regressão indica uma relação entre dois ou mais processos de séries temporais não relacionadas simplesmente porque cada um tem uma tendência, é uma série temporal integrada (como por exemplo um passeio aleatório), ou ambos.

Procedimento de Duas Etapas de Engle-Granger: Um método em duas etapas para estimar modelos de correção de erros por meio dos quais o parâmetro de cointegração é estimado na primeira etapa, e os parâmetros de correção de erros são estimados na segunda etapa.

Processo Altamente Persistente: Um processo de série temporal no qual os resultados no futuro distante são altamente correlacionados com os resultados atuais.

Processo AR(1) Estável: Um processo AR(1) no qual o parâmetro da defasagem é menor que um em valor absoluto. A correlação entre duas variáveis aleatórias na sequência declina para zero a uma taxa geométrica conforme a distância entre as variáveis aleatórias aumenta, e assim um processo AR(1) estável é fracamente dependente.

Processo Auto-Regressivo de Ordem um [AR(1)]: Um modelo de série temporal cujo valor atual depende linearmente de seu valor mais recente mais um distúrbio imprevisível.

Processo de Média Móvel de Ordem um [MA(1)]: Um processo de série temporal gerado como uma função linear do valor atual e do valor com uma defasagem de um processo estocástico não-correlacionado com variância constante e média zero.

Processo de Raiz Unitária: Um processo de série temporal altamente persistente no qual o valor corrente é igual ao valor do período anterior, mais uma perturbação fracamente dependente.

Processo de Séries Temporais: Ver processo estocástico.

Processo de Tendência Estacionária: Um processo que é estacionário uma vez que uma tendência temporal tenha sido removida; normalmente fica implícito que a série modificada é fracamente dependente.

Processo de Tendência: Um processo de série temporal cujo valor esperado é uma função crescente ou decrescente do tempo.

Processo Estacionário: Um processo de série temporal no qual a distribuição marginal e todas as distribuições conjuntas são invariantes ao longo do tempo.

Processo Estocástico: Uma seqüência de variáveis aleatórias indexada pelo tempo.

Processo Não-Estacionário: Um processo de série temporal cujas distribuições conjuntas não são constantes ao longo de épocas diferentes.

Propensão de Impacto: Em um modelo de defasagens distribuídas, a alteração imediata na variável dependente dado um aumento de uma unidade na variável independente.

Propensão de Longo Prazo: Em um modelo de defasagens distribuídas, a alteração final na variável dependente, dado um aumento permanente de uma unidade na variável independente.

Propriedades Assimptóticas: Propriedades de estimadores e estatísticas de testes que se aplicam quando o tamanho da amostra cresce sem limites.

Propriedades de Amostras Grandes: Ver propriedades assíntoticas.

p-valor: O menor nível de significância ao qual a hipótese nula pode ser rejeitada. De forma equivalente, o maior nível de significância ao qual a hipótese nula não pode ser rejeitada.

Q

Quase-Experimento: Ver experimento natural.

R

Raiz do Erro Quadrático Médio (REQM): Outro nome do erro-padrão da regressão em análise de regressão múltipla.

Razão t: Ver estatística t .

R-Barra Quadrado: Ver R -quadrado ajustado.

Região de Rejeição: O conjunto de valores de uma estatística de teste que leva à rejeição da hipótese nula.

Regra de Rejeição: Em testes de hipóteses, a regra que determina quando a hipótese nula será rejeitada em favor da hipótese alternativa.

Regressando: Ver variável dependente.

Regressão através da Origem: Análise de regressão na qual o intercepto é definido como zero; as inclinações são obtidas minimizando a soma dos resíduos quadrados, como sempre.

Regressão Auxiliar: Uma regressão utilizada para calcular uma estatística de teste – tal como a estatística de teste da heteroscedasticidade ou da correlação serial – ou qualquer outra regressão que não estima o modelo de interesse principal.

Regressão de Variável Dummy: Em uma estrutura de dados de painel, a regressão que inclui uma variável *dummy* para cada unidade de corte transversal, juntamente com as demais variáveis explicativas. Ela produz o estimador de efeitos fixos.

Regressor: Ver variável explicativa.

Resíduo: A diferença entre o valor efetivo e o valor estimado (ou previsto); existe um resíduo para cada observação na amostra usada para obter uma reta de regressão MQO.

Resposta de Solução de Canto: Uma variável dependente não-negativa que é em termos gerais contínua com valores estritamente positivos, mas que assume o valor zero com alguma regularidade.

Restrições de Exclusão: Restrições que especificam que certas variáveis são excluídas do modelo (ou têm coeficientes populacionais nulos).

Restrições Múltiplas: Mais de uma restrição nos parâmetros, em um modelo econométrico.

Restrições Sobreidentificadoras: A condição de momento extra que surge por haver mais variáveis instrumentais do que variáveis explicativas endógenas em um modelo linear.

Reta de Regressão MQO: A equação que relaciona o valor previsto da variável dependente às variáveis independentes, em que as estimativas dos parâmetros tenham sido obtidas por MQO.

Retirada de Tendência: A prática de remover a tendência de uma série temporal.

R-Quadrado Ajustado: Uma medida do grau de ajuste na análise de regressão múltipla que penaliza variáveis explicativas adicionais ao utilizar um ajustamento dos graus de liberdade na estimativa da variância dos erros.

R-Quadrado Populacional: Na população, a fração da variação na variável dependente que é explicada pelas variáveis explicativas.

R-Quadrado: Em um modelo de regressão múltipla, a proporção da variação amostral total na variável dependente que é explicada pela variável independente.

S

Sazonalidade: Uma característica de séries temporais mensais ou trimestrais, na qual o valor médio difere sistematicamente por época do ano.

Seleção Amostral Endógena: Seleção amostral não-aleatória ocorre quando a seleção está relacionada com a variável dependente, seja diretamente ou por meio do termo erro da equação.

Seleção Amostral Exógena: Seleção amostral que depende de variáveis explicativas exógenas ou é independente do termo erro na equação de interesse.

Seleção de Amostra Não-Aleatória: Um processo de seleção amostral que não pode ser caracterizado pela extração aleatória da população de interesse.

Semi-Elasticidade: A alteração percentual na variável dependente dado um aumento de uma unidade em uma variável independente.

Seqüência da Diferença de um Martingale: A primeira diferença de um *martingale*. Ela é imprevisível (ou tem uma média zero), dados os valores passados da seqüência.

Serialmente Não-Correlacionados: Os erros em um modelo de série temporal ou de dados de painel são não-correlacionados dois a dois ao longo do tempo.

Serviços de Busca On-Line: Programa de computador que possibilita que a Internet ou bancos de dados na Internet sejam pesquisados por tópico, nome, título ou palavras-chave.

Significância Econômica: Ver significância prática.

Significância Geral de uma Regressão: Um teste da significância conjunta de todas as variáveis explicativas que aparecem em uma equação de regressão múltipla.

Significância Prática: A importância prática ou econômica de uma estimativa, que é indicada por seu sinal e magnitude, em oposição à sua significância estatística.

Simultaneidade: Um termo que significa que pelo menos uma variável explicativa em um modelo de regressão linear múltipla é determinada conjuntamente com a variável dependente.

Solução Integrada do Problema de Variáveis Omitidas: Uma variável *proxy* é substituída por uma variável omitida não observada em uma regressão por MQO.

Soma dos Quadrados dos Resíduos: Ver soma dos resíduos quadrados.

Soma dos Quadrados Explicada (SQE): A variação amostral total dos valores estimados em um modelo de regressão múltipla.

Soma dos Quadrados Total (SQT): A variação amostral total em uma variável dependente sobre sua média amostral.

Soma dos Resíduos Quadrados: Em análise de regressão múltipla, a soma dos quadrados dos resíduos MQO ao longo de todas as observações.

Suavização Exponencial: Um método simples de fazer a previsão de uma variável que envolve uma ponderação de todos os resultados anteriores dessa variável.

Subespecificando um Modelo: Ver excluindo uma variável relevante.

Superdispersão: Na modelagem de uma variável de contagem, a variância é maior que a média.

Superespecificando um Modelo: Ver inclusão de uma variável irrelevante.

T

Taxa de Crescimento: A mudança proporcional em uma série temporal, a partir do período anterior. Ela poderá ser aproximada como a diferença em logs ou descrita na forma percentual.

Tendência Exponencial: Uma tendência com uma taxa de crescimento constante.

Tendência Temporal Linear: Uma tendência que é uma função linear do tempo.

Tendência Temporal: Uma função do tempo que é o valor esperado de um processo de série temporal com tendência.

Teorema de Gauss-Markov: O teorema que afirma que sob as cinco hipóteses de Gauss-Markov (em modelos de cortes transversais ou de séries temporais) o estimador MQO é BLUE (condicional nos valores amostrais das variáveis explicativas).

Teorema do Limite Central: Um resultado básico da teoria da probabilidade implica que a soma das variáveis aleatórias independentes, ou mesmo de variáveis aleatórias fracamente dependentes, quando padronizadas por seu desvio-padrão tem uma distribuição que tende a normal padrão quando o tamanho da amostra cresce.

Termo de Interação: Uma variável independente em um modelo de regressão que é o produto de duas variáveis explicativas.

Termo Erro: A variável em uma equação de regressão simples ou múltipla que contém fatores não observados que afetam a variável dependente. O termo erro também pode incluir erros de medida nas variáveis dependente ou independentes observadas.

Teste Bicaudal: Um teste contra uma alternativa bilateral.

Teste Consistente: Um teste no qual, sob a hipótese alternativa, a probabilidade de rejeitar a hipótese nula converge para um conforme o tamanho da amostra cresce ilimitadamente.

Teste de Breusch-Godfrey: Um teste assintoticamente justificado de correlação serial AR(p), com o AR(1) sendo o mais popular; o teste leva em conta variáveis dependentes defasadas como também outros regressores que não sejam estritamente exógenos.

Teste de Breusch-Pagan: Um teste de heteroscedasticidade no qual os resíduos quadrados MQO são regredidos sobre as variáveis explicativas no modelo.

Teste de Davidson-MacKinnon: Um teste que é usado para testar um modelo contra uma alternativa não-aninhada; ele pode ser implementado como um teste t sobre os valores ajustados do modelo concorrente.

Teste de Dickey-Fuller (DF): Um teste t da hipótese nula de raiz unitária em um modelo AR(1). (Ver também teste de Dickey-Fuller aumentado.)

Teste de Dickey-Fuller Aumentado: Um teste de uma raiz unitária que inclui variações defasadas da variável como regressores.

Teste de Hipótese: Um teste estatístico da hipótese nula, ou mantida, contra uma hipótese alternativa.

Teste de Hipóteses Conjuntas: Um teste envolvendo mais de uma restrição sobre os parâmetros em um modelo.

Teste de Hipóteses Múltiplas: Um teste de uma hipótese nula envolvendo mais de uma restrição nos parâmetros.

Teste de White: Um teste de heteroscedasticidade que envolve fazer a regressão dos resíduos quadrados MQO sobre os valores estimados MQO e sobre os quadrados dos valores estimados; em sua forma mais generalizada, os resíduos quadrados MQO são regredidos sobre as variáveis explicativas, os quadrados das variáveis explicativas e todos os produtos cruzados não-redundantes das variáveis explicativas.

Teste do Erro de Especificação da Regressão (RESET): Um teste geral da forma funcional em um modelo de regressão múltipla; ele é um teste F da significância conjunta dos quadrados, cubos e talvez de potências mais altas dos valores estimados a partir da estimação inicial por MQO.

Teste Estatístico: Uma regra usada para testar hipóteses na qual cada resultado amostral produz um valor numérico.

Teste Monocaudal: Um teste de hipótese contra uma alternativa unilateral.

Transformação de Efeitos Fixos: Para dados de painel, os dados temporais reduzidos.

Transformação Interna: Ver transformação de efeitos fixos.

Truncamento Incidental: Um problema de seleção amostral pelo qual uma variável, habitualmente a variável dependente, somente é observada para certos resultados de outra variável.

V

Valor Base: O valor atribuído ao período base na construção de um índice numérico; habitualmente, o valor base é um ou 100.

Valor Crítico: No teste de hipóteses, o valor contra o qual um teste estatístico é comparado para determinar se a hipótese nula é rejeitada ou não.

Valor Esperado: Uma medida de tendência central na distribuição de uma variável aleatória, inclusive um estimador.

Valores Estimados: Os valores estimados da variável dependente quando os valores das variáveis independentes de cada observação são agregados na reta de regressão MQO.

Variância Amostral: A variância na distribuição amostral de um estimador, que mede a extensão na distribuição amostral.

Variância Assimptótica: O quadrado do valor pelo qual devemos dividir um estimador para obter uma distribuição normal padrão assimptótica.

Variância Condicional: A variância de uma variável aleatória, dada uma ou mais variáveis aleatórias.

Variância da Amostra: Um estimador consistente, não-viesado, da variância populacional.

Variância do Erro de Previsão: A variância no erro que surge ao prever um valor futuro da variável dependente com base em uma equação de regressão múltipla estimada.

Variância do Erro: A variância do termo erro em um modelo de regressão múltipla.

Variância: Uma medida de extensão na distribuição de uma variável aleatória.

Variáveis Aleatórias Independentes: Variáveis aleatórias cuja distribuição conjunta é o produto das distribuições marginais.

Variáveis Aleatórias Não-Correlacionadas Duas a Duas: Um conjunto de duas ou mais variáveis aleatórias no qual cada par é não-correlacionado.

Variáveis Aleatórias Não-Correlacionadas: Variáveis aleatórias que não são linearmente relacionadas.

Variáveis Dummy Anuais: Para conjuntos de dados com um componente de série temporal as variáveis *dummy* (binárias) serão iguais a um no ano relevante e zero em todos os outros anos.

Variáveis Dummy Sazonais: Um conjunto de variáveis *dummy* usadas para representar os trimestres ou meses do ano.

Variáveis Endógenas: Em modelos de equações simultâneas, variáveis que são determinadas pelas equações do sistema.

Variáveis Omitidas: Uma ou mais variáveis, que gostaríamos de controlar, foram omitidas ao estimar um modelo de regressão.

Variável Aleatória Contínua: Uma variável aleatória que assume qualquer valor específico com probabilidade zero.

Variável Aleatória de Bernoulli: Uma variável aleatória que assume os valores zero ou um.

Variável Aleatória Discreta: Uma variável aleatória que assume no máximo um número finito ou infinito contável de valores.

Variável Aleatória Padronizada: Uma variável aleatória transformada pela subtração de seu valor esperado e tendo o resultado sido dividido por seu desvio-padrão; a nova variável aleatória terá média zero e desvio-padrão um.

Variável Aleatória: Uma variável cujo resultado é incerto.

Variável Binária: Ver variável *dummy*.

Variável de Contagem: Uma variável que assume valores inteiros não-negativos.

Variável de Controle: Ver variável explicativa.

Variável de Resposta: Ver variável dependente.

Variável Dependente Defasada: Uma variável explicativa que é igual à variável dependente de um período de tempo anterior.

Variável Dependente Limitada: Uma variável dependente ou de resposta cujo alcance está limitado de alguma maneira importante.

Variável Dependente: A variável a ser explicada em um modelo de regressão múltipla (e em uma variedade de outros modelos).

Variável Dummy Dependente: Ver modelo de resposta binária.

Variável Dummy: Uma variável que assume os valores zero ou um.

Variável Endógena Defasada: Em um modelo de equações simultâneas, um valor defasado de uma das variáveis endógenas.

Variável Exógena: Qualquer variável que seja não-correlacionada com o termo erro no modelo de interesse.

Variável Explicada: Ver variável dependente.

Variável Explicativa Endógena: Uma variável explicativa em um modelo de regressão múltipla que é correlacionada com o termo erro, devido a uma variável omitida, um erro de medida ou simultaneidade.

Variável Explicativa Exógena: Uma variável explicativa que é não-correlacionada com o termo erro.

Variável Explicativa: Em análise de regressão, uma variável que é usada para explicar a variação na variável dependente.

Variável Independente: Ver variável explicativa.

Variável Instrumental (VI): Em uma equação com uma variável explicativa endógena, uma VI é uma variável que não aparece na equação, é não-correlacionada com o erro na equação, e é (parcialmente) correlacionada com a variável explicativa endógena.

Variável Nominal: Uma variável medida em unidades monetárias nominais ou correntes.

Variável Ordinal: Uma variável na qual a ordenação dos valores transmite informações, mas a magnitude dos valores não.

Variável Predeterminada: Em um modelo de equações simultâneas, uma variável endógena defasada ou uma variável exógena defasada.

Variável Previsora: Ver variável explicativa.

Variável Prevista: Ver variável dependente.

Variável Proxy: Uma variável observada que é relacionada, mas não idêntica a uma variável explicativa não observada em análise de regressão múltipla.

Variável Qualitativa: Uma variável que descreve uma característica não quantitativa de um indivíduo, uma firma, uma cidade etc.

Variável Real: Uma variável monetária indicada em termos de um período base.

Viés Assimptótico: Ver inconsistência.

Viés de Atenuação: Viés em um estimador que sempre tende a zero; assim, o valor esperado de um estimador com viés de atenuação é menor em magnitude que o valor absoluto do parâmetro.

Viés de Heterogeneidade: O viés no MQO devido a heterogeneidade omitida (ou variáveis omitidas).

Viés de Seleção Amostral: Viés no estimador MQO que é induzido pelo uso de dados surgidos da seleção amostral endógena.

Viés de Simultaneidade: O viés que surge do uso do MQO para estimar uma equação em um modelo de equações simultâneas.

Viés de Variável Omitida: O viés que surge nos estimadores MQO quando uma variável relevante é omitida da regressão.

Viés para Baixo: O valor esperado de um estimador é menor que o valor populacional do parâmetro.

Viés para Cima: O valor esperado de um estimador é maior que o valor do parâmetro populacional.

Viés: A diferença entre o valor esperado de um estimador e o valor populacional que o estimador supostamente está estimando.

Viésado Para Zero: Uma descrição de um estimador cujo valor esperado em termos absolutos é menor que o valor absoluto do parâmetro populacional.

Índice Remissivo

A

abertura política, inflação e
Exemplo 16.4, 502
Exemplo 16.6, 504
ações anti-dumping e importações de produtos químicos,
Exemplo 10.5, 325
efeitos das, Exemplo 10.11,
337
agrupamento independente de cortes transversais ao longo do tempo, 403
agrupamento independentemente de cortes transversais , 402
altamente persistente, 353
alternativa bilateral, 121
alternativas não-aninhadas, testes contra, 277
alternativas unilaterais, 116
amostra de aglomeração, 447
amostra não-aleatória, 292, 293
seleção de, 549
amostra selecionada, 549
amostragem aleatória, 6
hipótese RLM.2, 81
hipótese RLS.2, 46
hipótese MQ2E.2, 489
amostragem estratificada, 294
amostras de pares comparados, 447
análise de duração, 544

análise de má especificação, 86, 611
análise de políticas públicas, 213
e avaliação de programas, 236
com agrupamentos de cortes transversais, 408
com dados de painel de dois períodos, 421
análise de regressão
controlando grande número de fatores na, 193
utilizando séries temporais altamente persistentes na, 353
uso de variáveis com tendência na, 330
análise de resíduos, 199
previsão e, 195
análise de sensibilidade, 611
análise econometrítica, 609
causalidade e a noção de *ceteris paribus* na, 12
análise empírica, 2
análise empírica, passos na análise, 2
aptidão, o uso de duas notas de testes como indicadores de, Exemplo 15.6, 475
ARCH, 394
em retornos de ações, Exemplo 12.9, 395

arquivo de texto (ASCII), 606
assimptoticamente não-correlacionados, 342
ausência de correlação serial hipótese ST.5, 316
hipótese ST.5', 368
hipótese MQ2E.6, 90
ausência de dados, 292
autoconrelação, 316
autocorrelação de primeira ordem, 358
auto-seleção, 237
avaliação de programas, 213
análise de decisões políticas e, 236

B

bases de dados *on-line*, 606

C

causalidade de Granger, 587
causalidade e a noção de *ceteris paribus* na análise econometrítica, 12
ceteris paribus, 13, 71
classificação das faculdades de direito sobre os salários iniciais, efeitos da, Exemplo 7.8, 220
codificação superior, 543
coeficiente de determinação, 38
coeficientes beta, 177, 178

- coeficientes de variáveis *dummy*
 explicativas quando a variável
 dependente é $\log(y)$,
 interpretação dos, 214
 coeficientes padronizados, 178
 co-integração, 574
 e modelos de correção de erro, 579
 entre fertilidade e isenção de impostos, Exemplo 18.5, 577
 colinearidade não perfeita
 hipótese RLM.4, 82
 hipótese ST.3, 314
 hipótese ST.3', 346
 colinearidade perfeita, 82
 combinação linear única dos parâmetros, teste das hipóteses sobre, 134
 comparação fora da amostra da previsão de desemprego, Exemplo 18.9, 589
 componentes das variâncias
 MQO: multicolinearidade, 92
 condição de classificação, 472, 500
 hipótese MQ2E.4, 489
 para identificação de uma equação estrutural, 504
 condição de ordem, 472, 500
 de uma equação, 472
 para identificação, 505
 condições de primeira ordem, 29, 70
 conjuntamente não significantes, 142
 conjunto de dados de corte transversal, 5
 sobre salários e outras características individuais, Tabela 1.1, 7
 conjunto de dados de painel
 sobre estatísticas de crime nas cidades para dois anos, Tabela 1.5, 11
 conjunto de dados
 de corte transversal, 5
 escolhendo o mais apropriado, 605
 conjunto de informações, 581
 consistência do MQO
 Teorema 5.1, 160
 Teorema 11.1, 346
 consistência, 158, 159
 inexistência de viés e, 368
 contemporaneamente exógenas, 313, 345
 contemporaneamente
 homoscedásticos, 348
 controle de muitos fatores na análise de regressão, 193
 convenção sobre o índice temporal, 311
 correções da seleção amostral, 549, 553
 correlação serial AR(1)
 com regressores estritamente exógenos, teste t , 373
 com regressores estritamente exógenos, teste, 374
 MQG factível com
 heteroscedasticidade e, 397
 na curva de Phillips, teste, Exemplo 12.1, 374
 na equação de salário mínimo, teste, Exemplo 12.2, 377
 sem regressores estritamente exógenos, teste, 376
 teste após MQ2E, 480
 correlação serial AR(3), teste, Exemplo 12.3, 379
 correlação serial AR(q), teste, 378
 correlação serial, 316
 com regressores gerais, teste da, 377
 correções para ordens mais elevadas, 386
 diferenciação e, 387
 em modelos de regressão, heteroscedasticidade e, 396
 modelos dinamicamente completos e a ausência de, 360
 na presença da variável dependente defasada, 371
 teste da ordem mais elevada na, 378
 teste da, 371
 correlação, não-serial, hipótese TS.5', 348
 cortes transversais agrupados, 9
 análise de decisões governamentais com, 408
 dois anos de preços de moradias, Tabela 1.4, 10
 em dados de painel, aplicação do MQ2E na, 481
 covariância estacionária, 341
 covariável, 21
 crime, modelo econômico do, Exemplo 1.1, 3
 Exemplo 5.3, 168-169
 Exemplo 9.1, 274
 crimes no *campus* e matrículas, Exemplo 4.4, 124
 critérios dentro da amostra, 588
 critérios fora da amostra, 588
 cumprimento da lei sobre os níveis de criminalidade das cidades, efeito do, Exemplo 1.5, 15
 curva de Phillips de expectativas aumentadas, Exemplo 11.5, 351
 curva de Phillips estática
 Exemplo 10.1, 319
 Exemplo 12.5, 385
 curva de Phillips, expectativas aumentadas, Exemplo 11.5, 351

D
 dados de painel (ou longitudinais), 10, 402

- aplicação do MQ2E em cortes transversais agrupados e, 481
- equação de salários usando, Exemplo 14.4, 443
- métodos para outras estruturas de dados, aplicação de, 445
- modelos de equações simultâneas com, 510
- organização dos, 420
- dados de painel de dois períodos análise, 414
- análise de decisões governamentais com, 421
- dados, 616
- dados econômicos, estrutura dos, 5
- dados experimentais, 2
- dados longitudinais, 402
- dados não-experimentais, 1
- dados observacionais, 2
- dados quase-diferenciados, 381
- dados quase-reduzidos, 442
- dados temporais reduzidos, 434
- bases de, *on-line*, 606
- compilação de, 605
- conjunto de, de painel sobre estatísticas de crime nas cidades, para dois anos, Tabela 1.5, 11
- conjunto de, sobre taxas de crescimento econômico e características de países, Tabela 1.2, 8
- de painel, 10, 402
- de séries de tempo, 8
- entrada e armazenamento, 606
- estrutura dos dados econômicos, 5
- experimentais, 2
- exploração da base de, 612
- frequência dos, 8
- inspeção, limpeza e sumário, 608
- longitudinais, 402
- na lista parcial dos dados do arquivo WAGE1.RAW, Tabela 7.1, 208
- não-experimentais, 2
- observacionais, 2
- para Porto Rico (salário mínimo, desemprego e dados relacionados), Tabela 1.3, 9
- sobre as taxas de inflação e de desemprego dos Estados Unidos, lista parcial dos, Tabela 10.1, 307
- defasagem distribuída geométrica (ou de Koyck), 562
- defasagem distribuída racional (18.16), distribuição de defasagens da, Figura 18.1, 565
- defasagens distribuídas da taxa de criminalidade sobre a taxa de esclarecimento de crimes, Exemplo 13.6, 420
- demanda de cigarros, Exemplo 8.7, 264
- derivação da inconsistência no método MQO, 161
- derivação da equação (3.22), 105
- derivação das condições de primeira ordem, da equação (3.13), 105
- desempenho de empresas, remuneração dos diretores executivos e, Exemplo 6.4, 193
- desempenho de estudantes e tamanho de escolas, Exemplo 4.2, 119
- desempenho econômico, resultados das eleições e, Exemplo 10.6, 326
- desempenho em matemática de estudantes e o programa de merenda escolar, Exemplo 2.12, 50
- desemprego
- efeito do salário mínimo sobre o, Exemplo 1.6, 16
- pedidos de auxílio-, efeito das zonas industriais sobre os, Exemplo 13.8, 427
- previsões de, comparações fora da amostra de, Exemplo 18.9, 589
- deslocamentos nas equações de oferta permitindo desenhar a equação de demanda, Figura 16.1, 499
- despesas e poupança familiares, Exemplo 16.2, 495
- dessazonalização, 337
- desvio-padrão de $\hat{\beta}_j$, 98
- determinantes da fertilidade feminina, Tabela 13.1, 405
- determinantes da *nmgrad*
- Exemplo 3.1, 71
- Exemplo 3.4, 78
- Exemplo 4.3, 122
- determinantes da propriedade de computadores pessoais, Exemplo 8.9, 268
- determinantes do número de prisões de homens jovens, Tabela 17.3, 541
- diferença salarial por gênero, mudanças no retorno da educação, Exemplo 13.2, 406
- diferença salarial, mudanças no retorno da educação por gênero, Exemplo 13.2, 406
- diferenças nas inclinações, 223
- diferenciação
- com mais de dois períodos de tempo, 424
- e correlação serial, 387
- equação da taxa de juros, Exemplo 12.6, 388
- distribuição de defasagens, 309

- com duas defasagens diferentes de zero, Figura 10.1, 310
- da defasagem distribuída racional (18.16), Figura 18.1, 565
- distribuição de Dickey-Fuller, 568
- distribuição de Poisson, 538
- distribuição normal
- homoscedástica com uma única variável explicativa, Figura 4.1, 111
- distribuição para os estimadores padronizados, Teorema 4.2, 114
- distribuições amostrais normais
- Teorema 4.1, 112
 - Teorema 10.5, 318
- dormir *versus* trabalhar, Exemplo 13.5, 419
- duração da reincidência, Exemplo 17.4, 545
- E**
- econometria, 1
- editor de texto, 607
- educação
- dos países em uma equação do peso de nascimentos, Exemplo 4.9, 145
 - e a diferença salarial por gênero, mudanças no retorno da, Exemplo 13.2, 406
 - gráfico da dispersão de salário e, Figura 2.3, 28
 - medindo o retorno da, Exemplo 1.4, 14
 - para homens, a estimiação do retorno da, Exemplo 15.2, 460
 - para mulheres casadas, a estimiação do retorno da, Exemplo 15.1, 459
 - para mulheres que trabalham, retorno da, Exemplo 15.5, 471
- Exemplo 15.7, 476
- Exemplo 15.8, 478
- relação entre a probabilidade de estar na força de trabalho e anos de, Figura 7.3, 232
- retorno da, mudou no transcorrer do tempo, Exemplo 14.2, 437
- salário e, sobre a fertilidade, efeito da, Exemplo 15.9, 482
- utilizando a proximidade da faculdade como uma VI da, Exemplo 15.4, 467
- Exemplo 2.4, 32
 - Exemplo 2.7, 36
- efeito causal, 12
- efeito de aglomeração, 447
- efeito de interação, 187
- efeito não observado, 415
- efeito parcial, 71
- efeitos aleatórios
- de efeitos fixos, 445
 - estimador, 443
 - modelo, 441
- efeitos fixos e aleatórios
- hipóteses EA.3- EA.5, 451
 - hipóteses EF.1- EF.7, 450-451
- efeitos fixos, 415
- com painéis não equilibrados, 440
 - efeitos aleatórios ou, 445
 - estimador, 434
 - modelo, 415
 - ou primeira diferenciação, 439
 - transformação, 433
- eficiência assintótica de MQO, 169
- Teorema 5.3, 171
- eficiência de MQO: o Teorema de Gauss-Markov, 99
- e inferência, 369
- elasticidade de curto prazo, 320
- elasticidade de longo prazo, 309
- elasticidade, 44
- emprego porto-riquenho e
- salário mínimo, Exemplo 10.3, 320
 - Exemplo 10.9, 334
- emprego
- em Porto Rico, Exemplo 10.9, 334
 - e o salário mínimo em Porto Rico, Exemplo 10.3, 320
- endogeneidade
- de uma única variável explicativa, e o teste de, 476
 - e o teste de restrições sobre identificadoras, e o teste de, 475
 - teste de, 475
- equação da fertilidade
- Exemplo 10.8, 332
 - Exemplo 11.6, 359
 - Exemplo 11.8, 363
- equação da oferta salarial de mulheres casadas
- Exemplo 17.5, 553,554
 - Tabela 17.5, 554
- equação de *ungrad* com círo de medida, Exemplo 9.7, 291
- equação de poupança familiar, Exemplo 8.6, 259
- equação de primeiras diferenças, 417
- equação de salários
- heteroscedasticidade em uma, Exemplo 2.13, 53
 - três estimadores diferentes da, Tabela 14.2, 444
 - usando dados de painel, Exemplo 14.4, 443
- equação do logaritmo dos salários com erros-padrões robustos
- em relação à heteroscedasticidade, Exemplo 8.1, 247
- Exemplo 2.10, 42

- equação do peso de nascimentos, educação dos pais em uma, Exemplo 4.9, 145
- erros-padrão em uma, Exemplo 5.2, 166
- equação do salário hora
 - Exemplo 3.2, 72
 - Exemplo 3.6, 89
 - Exemplo 4.1, 117
 - Exemplo 7.1, 211
- equação dos preços de imóveis
 - Exemplo 9.2, 276
- heteroscedasticidade na, Exemplo 8.4, 253
- equação estrutural, 464, 492
- identificação e estimação, 498
- equação exatamente identificada, 506
- equação identificada, 498
- equação log dos salários por hora
 - Exemplo 7.5, 215
 - Exemplo 7.6, 216
 - Exemplo 7.10, 225
- equação não-identificada, 506
- equação simples dos salários, Exemplo 2.2, 22
- equação sobreidentificada, 506
- equações, sistemas com mais de duas, 505
- erro absoluto médio (EAM), 588
- erro de composição, 415
 - termo, 442
- erro de medida multiplicativo, 287
- erro de medida, 272
 - em uma variável explicativa, 287
 - equação da *nmgrad* com, Exemplo 9.7, 291
 - função de poupança com, 286
 - na variável dependente, 285
 - nas taxas de rejeição, Exemplo 9.6, 287
- erros AR(1)
 - estimação MQG factível com, 382
 - MQ2E com, 481
- erros clássico nas variáveis (ECV), 289
- erros estruturais, 493
- erros nas variáveis, 453
 - problemas de, solução VI de, 473
- erros serialmente correlacionados, propriedades do MQO com, 368
- estatística de Chow, 229
- estatística de Durbin-Watson (DW), 375
- estatística de escore, 167
- estatística *F* robusta em relação à heteroscedasticidade, 248
 - Exemplo 8.2, 248
- estatística *F*, 140
 - para a significância geral de uma regressão, 147
- propriedades do método MQO sob, 285
- erro idiosincrático, 415
- erro-padrão assumptótico, 166
 - em modelos de variável dependente limitada, 558
- erro-padrão robusto em relação à correlação serial, 390
 - de $\hat{\beta}_1$, 390
- erro-padrão robusto em relação à heteroscedasticidade, 247
 - equação do log dos salários com, Exemplo 8.1, 247
- erro-padrão
 - em uma equação do peso de nascimentos, Exemplo 5.2, 166
 - de $\hat{\beta}_1$, 57
 - de $\hat{\beta}_j$, 98
 - da regressão (EPR), 57, 98
 - dos estimadores de MQO, estimação de σ^2 , 97
- erros AR(1)
 - estimação MQG factível com, 382
 - MQ2E com, 481
- erros clássico nas variáveis (ECV), 289
- erros estruturais, 493
- erros nas variáveis, 453
 - problemas de, solução VI de, 473
- erros serialmente correlacionados, propriedades do MQO com, 368
- estatística de Chow, 229
- estatística de Durbin-Watson (DW), 375
- estatística de escore, 167
- estatística *F* robusta em relação à heteroscedasticidade, 248
 - Exemplo 8.2, 248
- estatística *F*, 140
 - para a significância geral de uma regressão, 147
- estatística LM robusta em relação à heteroscedasticidade, 249
 - Exemplo 8.3, 250
- estatística multiplicador de Lagrange (*LM*), 166
- outros testes de amostras grandes, 166
 - para q restrições de exclusão, 168
- estatística quase-razão de verossimilhança (EQMV), 540
- estatística razão de quase-verossimilhança, 540
- estatística razão de verossimilhança, 523
- estatística *R-n*-quadrado, 168
- estatística *t* robusta em relação à heteroscedasticidade, 246
- estatística *t*, 115
- estatisticamente não significante, 122
- estatisticamente significante, 122
- estatisticamente significantes e conjuntamente, 142
- estatísticas *F* e *t*, relação entre as, 143
- estatísticas MQO
 - efeitos da dimensão dos dados nas, 174
 - efeitos da mudança das unidades de medida nas, 40
 - propriedades algébricas das, 36
- estatísticas robustas em relação à heteroscedasticidade, 392
- estatísticas *t* assumptóticas, 166
- estimação da regressão censurada de reincidência criminal, Tabela 17.4, 546
- estimação de Cochrane-Orcutt (CO), 382
 - em estudo de evento, Exemplo 12.4, 383
- estimação de efeitos fixos, 433

- da equação da taxa de refugo, Tabela 14.1, 435
- estimação de Prais-Winsten (PW), 382-383
- estimação de variáveis instrumentais cálculo do R -quadrado após a, 464 do modelo de regressão múltipla, 464
- estimação MQO inferência robusta em relação à heteroscedasticidade, após a, 244
- estimação não-viesada de σ^2 Teorema 2.3, 56 Teorema 3.3, 98 Teorema 10.3, 317
- estimação Tobit de horas anuais trabalhadas, MQO e, Tabela 17.2, 534
- estimação, 506 por MQ2E, 503
- estimador de adiantamentos e atrasos, 579
- estimador de diferenças-em-diferenças, 410
- estimador de primeiras diferenças, 417
- estimador de variáveis instrumentais, 457 inferência estatística com, 457
- estimador interno, 434
- estimador linear não-viesado no modelo AR(1), obtenção do o melhor, 380
- estimadores de mínimos quadrados generalizados (MQG), 258
- estimadores de uma equação de salários, três diferentes, Tabela 14.2, 444
- estimadores MQO
- distribuições amostrais dos, 110 e o Teorema de Gauss-Markov, variâncias dos, 315 estimando σ^2 : os erros-padrão dos, 97 e variâncias dos, 45 valor esperado dos, 80 variâncias amostrais em, Teorema 2.2, 54 variâncias dos, 51, 91
- estimativas das regressões simples e múltipla, comparação das, 75
- estimativas de Logit e Probit, interpretação das, 523
- estimativas Probit, interpretando Logit e, 523
- estimativas Tobit, interpretação, 531
- estritamente exógenas, 313
- estrutura conceitual (ou teórica), 613
- estruturas de dados, um comentário sobre, 12
- estudo de evento, 322
- estimação de Cochrane-Orcutt em, Exemplo 12.4, 383
- excluir uma variável relevante, 86
- exogeneidade estrita, 417
- experimento natural, 412, 462 explicando os registros de prisões, Exemplo 3.5, 78
- F**
- fatalidades no trânsito, efeitos da legislação a respeito da condução de veículos sob embriaguez na, Exemplo 13.7, 423 fazendo previsões com um passo à frente, 585
- fertilidade e isenção de impostos, co-integração entre, Exemplo 18.5, 577
- fertilidade feminina ao longo do tempo, Exemplo 13.1, 404
- determinantes da, Tabela 13.1, 405
- fertilizantes efeitos dos, sobre a produção agrícola, Exemplo 1.3, 13 produção de soja e, Exemplo 2.1, 21
- força policial, taxa de assassinatos e tamanho da, Exemplo 16.1, 494
- forma especial do teste de White na equação do log dos preços de imóveis, Exemplo 8.5, 255
- forma funcional, 179, 320 envolvendo logaritmos, resumo das, Tabela 2.3, 44 unidades de medida e, 39
- forma reduzida, 497
- equação da, 466 erro da, 497 parâmetros da, 497
- formas funcionais logarítmicas, 179
- fortemente dependentes, 353
- fracamente dependentes, 342 séries temporais, 342
- freqüência, efeitos sobre o desempenho em exames finais, Exemplo 6.3, 188
- função da perda, 582
- função de poupança com erro de medida, 286
- função de regressão amostral (FRA), 30, 70
- função linear Figura 2.1, 24
- função log-verossimilhança, 521

- função
 de regressão amostral (FRA), 30
 de regressão populacional (FRP), 24
 funções quadráticas, 182
- G**
- gastos de campanha, resultados eleitorais e, Exemplo 2.5, 33
 Exemplo 2.9, 39
 gráfico
 da intensidade de pesquisa e desenvolvimento contra as vendas das firmas, Figura 9.1, 297
 de poupança e renda, e a regressão populacional, Figura 2.2, 26
 de salários e educação, Figura 2.3, 28
 gráficos
 de crime, Figura 4.5, 125
 da equação (7.16), Figura 7.2, 224
 de salários, Figura 7.1, 210
 função logística, Figura 17.1, 519
 y_i , Figura 2.7, 47
 grau de ajuste, 38, 77, 189, 370
 e seleção de regressores
 graus de liberdade (gl), 56, 97
 graus de liberdade do denominador, 141
 graus de liberdade do numerador, 141
 grupo base, 209
 grupo de controle, 213
 grupo de referência, 210
 grupo de tratamento, 213
 grupo experimental, 213
- H**
- heterogeneidade não observada, 415
 heteroscedasticidade condicional auto-regressiva, ver ARCH
 heteroscedasticidade, 53
 a necessidade de estimar a função de: MQG factível, 262
 de forma desconhecida, 244
 e a hipótese de mercados eficientes, Exemplo 12.8, 393
 e correlação serial em modelos de regressão, 396
 em equações de preços de imóveis, Exemplo 8.4, 253
 em regressões de séries temporais, 392
 em uma equação de salários, Exemplo 2.13, 53
 é percebida como uma constante multiplicativa, 256
 forma especial do teste de White de, 255
 no MQO, consequências da, 243
 procedimento MQG factível
 para corrigir a, 263
 teste de, 251, 393
 teste de Breusch-Pagan de, 252
 teste de White da, 254
 hipótese da renda permanente, teste da, Exemplo 16.7, 509
 hipótese de média condicional zero
 hipótese RLM.3, 81
 hipótese RLS.3, 46
 hipótese ST.2, 313
 hipótese ST.2', 345
 hipótese dos mercados eficientes Exemplo 11.4, 349
 heteroscedasticidade e a, Exemplo 12.8, 393
 hipótese nula, 114
 hipóteses alternativas, 116
 hipóteses das expectativas, Exemplo 1.7, 17
 hipóteses de Gauss-Markov, 91
 hipóteses do modelo linear clássico (MLC), 111
 inferência sob as, 318
 hipóteses múltiplas
 após a estimativa por MQ2E, teste de, 473
 testando, 522
 teste de, 138
 hipóteses
 média condicional zero, 23
 de efeitos fixos e aleatórios (hipóteses EF.1-EF.7)
 (hipóteses EA.3-EA.5), 449-452
 de mínimos quadrados em dois estágios (hipóteses MQ2E.1-MQ2E.6), 522-524
 de MQO agrupados usando primeiras diferenças (hipóteses PD.1-PD.7), 431-432
 hipóteses
 expectativas das, Exemplo 1.7, 17
 mercados eficientes, Exemplo 11.4, 349-350
 sobre $\hat{\beta}_j$, teste de outras, 123
 teste de múltiplas, 522
 histograma de *taxap*, Figura 5.2, 164
 homoscedasticidade, 51
 hipótese de modelos de séries temporais, 363
 hipótese MQ2E.5, 489
 hipótese RLM.5, 91
 hipótese RLS.5, 51
 hipótese ST.4, 315
 hipótese ST.4', 315
 hipótese ST.4', 348
 modelo de regressão simples sob, Figura 2.8, 52

- I**
- identificação, 457
 - de uma equação, condição de ordem para, 472
 - de uma equação estrutural, condição de classificação para a, 500
 - em sistemas com três ou mais equações, 505
 - em um sistema de duas equações, 498
 - importações de produtos químicos, ações anti-*dumping* e, Exemplo 10.5, 325
 - importações nominais dos Estados Unidos, Figura 10.3, 329
 - inclinações
 - considerando diferentes, 223
 - inclusão de variáveis irrelevantes, 85
 - inconsistência, 161
 - Índice Remissivo
 - índices temporais, 311
 - inexistência de viés em MQO, 45, 311
 - Teorema 2.1, 49
 - Teorema 3.1, 84
 - Teorema 10.1, 315
 - inexistência de viés e consistência, 368
 - inferência robusta em relação à correlação serial após o MQO, 388
 - inferência robusta em relação à heteroscedasticidade após a estimação MQO, 244
 - inferência
 - eficiência e, 369
 - sob as hipóteses do modelo linear clássico, 318
 - inflação
 - e abertura política
 - Exemplo 16.4, 502
 - Exemplo 16.6, 504
 - e déficits sobre as taxas de juros, efeitos da, Exemplo 10.2, 319
 - e taxas de desemprego, lista parcial dos dados dos Estados Unidos, Tabela 10.1, 307
 - inflação dos preços residenciais, investimento imobiliário e, Exemplo 18.1, 566
 - inflação norte-americana anual, teste de raiz unitária da, Exemplo 18.3, 570
 - informação ordinal usando variáveis *dummy*, incorporando, 218
 - informações qualitativas, 207
 - integrado de ordem um [I(1)], 357
 - integrado de ordem zero [I(0)], 357
 - intensidade de pesquisa e desenvolvimento contra as vendas das empresas, gráfico da, Figura 9.1, 297 e o tamanho das empresas, Exemplo 9.8, 296
 - Exemplo 9.9, 297
 - intercepto, 67
 - deslocamento, 209
 - parâmetro dc, 21
 - Internet, 603
 - intervalo de confiança (IC), 131
 - da *nmgrad* futura, Exemplo 6.6, 199
 - da *nmgrad* prevista, Exemplo 6.5, 197
 - das previsões, 195
 - intervalo de confiança assimptótico, 166
 - inversa
 - da razão de Mills, 531
 - investimento imobiliário e inflação nos preços residenciais, Exemplo 18.1, 566
 - e preços, Exemplo 10.7, 331
 - Exemplo 10.10, 335
 - isenção pessoal de impostos sobre a taxa de fertilidade, efeitos da, Exemplo 10.4, 321

L

 - legislação a respeito da condução de veículos sob embriaguez sobre as fatalidades no trânsito, efeitos da, Exemplo 13.7, 423
 - leis de indenizações trabalhistas sobre os prazos de afastamento dos trabalhadores, Exemplo 13.4, 413
 - letras do Tesouro norte-americano, taxas das
 - Figura 11.2, 355
 - teste de raiz unitária das, de três meses, Exemplo 18.2, 568
 - linear nos parâmetros
 - Hipótese RLM.1, 80
 - Hipótese RLS, 45
 - Hipótese ST.1, 312
 - Hipótese MQ2E.1, 488
 - linearidade e fraca dependência, hipótese ST.1', 345
 - linha de tendência estimada, Figura 18.2, 593
 - lista parcial de dados sobre taxas de inflação e de desemprego nos Estados Unidos, Tabela 10.1, 307
 - log do produto interno bruto real dos Estados Unidos, raiz unitária no, Exemplo 18.4, 571
 - log(*preço*) como uma função quadrática, Figura 6.2, 185

M

má especificação da forma funcional, 272
 RESET como um teste geral da, 275
martingale, 582
 seqüência da diferença, 567
 matrículas, crimes no *campus* e, Exemplo 4.4, 124
 máxima verossimilhança estimação (EMV), 521
 de modelos logit e probit, 521
 mecânica e interpretação dos mínimos quadrados ordinários, 69
 média zero e correlação zero, hipótese RLM.3', 161
 medida
 e forma funcional, unidades de, 39
 sobre as estatísticas MQO, efeitos da mudança das unidades da, 40
 medindo o retorno da educação, Exemplo 1.4, 14
 melhor estimador linear não-viesado (BLUE), 99
 método de Heckit, 552
 micronumerosidade, 94
 estimador, 111
 mínimos desvios absolutos (MDA), 299
 mínimos quadrados de dois estágios (MQ2E ou MQDE), 453, 468
 a equações de séries temporais, aplicação do, 479
 com erros AR(1), 481
 com heteroscedasticidade, 478
 em cortes transversais agrupados e dados de painel, aplicando, 481
 estimação por, 503

cstimação, teste de hipóteses múltiplas após a, 473
 estimador, 469
 hipóteses do, 488-490
 multicolinearidade e, 471
 teste da correlação serial AR(1) após, 480
 mínimos quadrados ordinários, ver MQO
 mínimos quadrados ponderados, 393
 estimadores (MQP), 258
 estimando o modelo de probabilidade linear por, 268
 modelo AR(1)
 estimação do MQG factível no, 382
 Exemplo 11.3, 347
 obtendo o melhor estimador linear não-viesado no, 382
 modelo de correção de erro, 579, 580
 co-integração e, 574
 dos rendimentos de títulos, Exemplo 18.7, 580
 modelos de defasagem distribuída infinita (DDI), 560
 modelo de defasagem distribuída racional (DDR), 565
 modelo de efeitos não observados, 415
 modelo de elasticidade constante, 43
 modelo de preço hedônico de imóveis, Exemplo 4.8, 133
 modelo de probabilidade linear (MPL), 231
 uma variável dependente binária, 230
 modelo de regressão censurada, 543
 modelo de regressão de Poisson, 537, 539
 modelo de regressão linear bivariada, 20
 modelo de regressão linear bivariada, 20
 modelo de regressão linear múltipla, 67
 modelo de regressão linear simples, 20
 modelo de regressão simples, 20
 sob homoscedasticidade, Figura 2.8, 52
 variáveis omitidas em um, 454
 modelo de variável latente, 519
 modelo dinamicamente completo, 361
 e a ausência de correlação serial, 360
 modelo econométrico, 4
 e métodos de estimação, 614
 modelo econômico, 2
 do crime
 Exemplo 1.1, 3
 Exemplo 5.3, 168-169
 Exemplo 9.1, 274
 modelo estático, 308
 Exemplo 11.1, 346
 modelo irrestrito, 139
 modelo linear clássico, 111
 modelo linear e Tobit, valores esperados estimados de horas em relação à educação para, Figura 17.3, 536
 modelo Logit, 518
 modelo Probit, 519
 probabilidades de respostas estimadas em respeito à educação, Figura 17.2, 528
 modelo restrito, 139
 modelo Tobit, 529
 para resposta de solução de canto, 529
 problemas de especificação no, 536

- valores esperados estimados de horas em relação à educação, Figura 17.3, 536
- modelo verdadeiro, 81
- modelo AR(1), Exemplo 11.3, 347
- auto-regressivo vetorial (VAR), 586
- com duas variáveis independentes, 64
- com funções quadráticas, 182
- com k variáveis independentes, 67
- com termos de interação, 187
- correção de erro, 579, 580
- defasagem distribuída racional (DDR), 564
- defasagens distribuídas finitas (DDF), 308 Exemplo 11.2, 347
- defasagens distribuídas infinitas (DDI), 560
- dinamicamente completo, 362
- duas variáveis, regressão linear de, 20
- econômico, 4
- econômico, 2
- efeito não observado, 415
- efeitos aleatórios, 441
- efeitos fixos, 415
- elasticidade constante, 43
- estático, 308 Exemplo 11.1, 346
- irrestrito, 139
- linear clássico, 111 *logit*, 518
- não-aninhado, 191
- populacional, 81
- probabilidade linear (MLP), 231
- Probit, 519
- regressão censurada, 543
- regressão de Poisson, 537, 539
- regressão linear bivariada, 20
- regressão linear múltipla, 67
- regressão linear simples, 20
- regressão simples, 20
- restrito, 139
- subespecificar o, 86
- superespecificação do, 85
- variável latente, 519
- verdadeiro, 81
- modelos de defasagem distribuída para o investimento imobiliário, Tabela 18.1, 566
- modelos de defasagens distribuídas finitas (DDF), 308 Exemplo 11.2, 347
- modelos de equações simultâneas (MES), 493 com dados de painel, 510 com séries temporais, 506 natureza dos, 491
- modelos de regressão censurada e truncada, 542
- modelos de regressão exemplos de séries temporais, 307 inclusão de variáveis irrelevantes nos, 85 utilizados para fazer previsões, tipos de, 583
- modelos de resposta binária, 518
- modelos Logit e Probit de resposta binária, 518 estimativa da máxima verossimilhança dos, 521 especificando, 518
- modelos mal-especificados, variâncias, 96
- modelos não-aninhados, 191, 277
- MPL, Logit e Probit, estimativas da participação da força de trabalho, Tabela 17.1, 526
- MQG factível (MQGF), 382 a função de heteroscedasticidade, necessidade de estimar, 262
- com heteroscedasticidade e correlação serial AR(1), 396
- Comparação entre MQO e, 384
- estimação com erros AR(1), 382
- estimação no modelo AR(1), 382
- estimador, 262
- procedimento para corrigir a heteroscedasticidade, 263
- MQO (mínimos quadrados ordinários), 28, 69
- com erros serialmente correlacionados, propriedades do, 368
- conseqüências da heteroscedasticidade, 243
- derivação da inconsistência no método, 161
- e estimação Tobit de horas anuais trabalhadas, Tabela 17.2, 534
- e MQGF, comparação entre, 384
- equação de regressão, interpretação da, 70
- estimadores de inclinação, variância amostral dos, Teorema 3.2, 92
- estimativa de intercepto, 70
- estimativas de inclinação, 70
- estimativas, 25, 69
- inferência robusta em relação à correlação serial, após o, 388
- mecânica e interpretação dos, 34, 69
- na amostra constante selecionada, 549
- propriedades assintóticas dos, 345
- resultados, Tabela 3, 620
- sob as hipóteses clássicas, propriedades de amostra finita dos, 311

- sob erro de medida,
propriedades dos, 285
- Teorema de Gauss-Markov,
eficiência de, 99
- usando primeiras diferenças,
hipóteses das agrupadas,
(hipóteses PD.1-PD.7), 431-
432
- valores estimados e resíduos, 74
- variâncias amostrais de,
Teorema 10.2, 317
- variâncias: multicolinearidade,
componentes das, 92
- víés de simultaneidade nos,
496
- MQO, consistência do
Teorema 5.1, 160
Teorema 11.1, 346
- MQO, eficiência assintótica
de, 169
Teorema 5.2, 164
Teorema 5.3, 171
Teorema 11.2, 349
- MQO, inexistência de víés em,
45, 311
Teorema 2.1, 49
Teorema 3.1, 84
Teorema 10.1, 315
- mulheres casadas, oferta de
mão-de-obra, Exemplo 17.2,
534
- mulheres casadas, participação
na força de trabalho, Exemplo
17.1, 525
- multicolinearidade, 93
e MQ2E, 471
- componentes das variâncias de
MQO, 92
- multiplicador de impacto, 309
- multiplicador de longo prazo, 309
- N**
- não-linearidades na regressão
simples, incorporação de, 41
- níveis de criminalidade das
cidades, o efeito do
cumprimento da lei sobre os,
Exemplo 1.5, 15
- nmgrad*
determinantes do
Exemplo 3.1, 71
Exemplo 3.4, 78
Exemplo 4.3, 122-123
- efeitos de se possuir
computadores sobre,
Exemplo 7.2, 272
- intervalo de confiança de
notas médias futuras,
Exemplo 6.6, 199
- intervalo de confiança do
previsto, Exemplo 6.5, 197
- normalidade assintótica, 163
e inferência em amostras
grandes, 163
- do MQO
Teorema 11.2, 349
Teorema 5.2, 164
- normalidade
hipótese RLM.6, 110
hipótese ST.6, 318
hipótese, 110
- notas de testes como indicadores
de aptidão, uso de duas,
Exemplo 15.6, 475
- número de prisões
de jovens, determinantes do,
Tabela 17.3, 541
- regressão de Poisson do,
Exemplo 17.3, 540
- número-índice, 320,323
- O**
- observações atípicas, 292, 295
- observações extremas, 295
- oferta de mão-de-obra
anual de mulheres casadas,
Exemplo 17.2, 534
- de mulheres casadas que
trabalham
Exemplo 16.3, 501
Exemplo 16.5, 503
- P**
- painel equilibrado, 425
- painel não equilibrado, 440
- parâmetro de co-integração das
taxas de juros, 579
- parâmetro,
intercepto, 21
inclinação, 21
- parâmetros de inclinação, 21, 67
- parâmetros estruturais, 497
- participação na força de trabalho
das mulheres casadas
Exemplo 8.8, 267
Exemplo 17.1, 525
- estimativas MPL, logit e probit
da, Tabela 17.1, 526
- participação nos planos de pensão
dos Estados Unidos,
Exemplo 3.3, 76
- passeio aleatório, 353
com tendência, 356
- Figura 11.1, 355
Figura 11.3, 356
- percentagem corretamente
prevista, 524
- pergunta, propondo, 602
- período base, 323
- perturbação, 21, 67
- peso de nascimentos, a estimação
do efeito do hábito de fumar
sobre o, Exemplo 15.3, 463
- planilhas, 607
- planos de pensão nos Estados
Unidos
participação nos, Exemplo 3.3,
76
- taxas de participação nos,
Exemplo 4.6, 129

- poluição do ar, preços de casas e, Exemplo 4.5, 126
- população prisional sobre as taxas de crimes violentos, efeito da, Exemplo 16.8, 512
- populacional de prisões, Exemplo 7.2, 234
- função de regressão (FRP), 24
- modelo, 81
- por mínimos quadrados ponderados, estimando, 268
- probabilidade de resposta estimada em relação à educação, Figura 17.2, 528
- revisitado, 266
- R*-quadrado da, 190
- poupança, despesas familiares e, Exemplo 16.2, 495
- preços de casas
- efeitos da poluição nos Exemplo 6.1, 179
 - Exemplo 6.2, 184
 - e distância de um incinerador Exemplo 5.1, 162
 - Exemplo 13.3, 408
 - e poluição do ar, Exemplo 4.5, 126
- preços, investimento imobiliário e, Exemplo 10.7, 331
- previsão com múltiplos passos à frente, 583, 589
- previsão com um passo à frente, comparando, 588
- previsão condicional, 583
- Previsão de y quando a variável dependente é $\log(y)$, 201, 202
- previsão incondicional, 583
- previsão por ponto, 584
- previsão
- comparações fora da amostra do desemprego, Exemplo 18.9, 589
 - comparando um passo à frente, 588
- condicional e incondicional, 583
- da taxa de desemprego, dois anos à frente, Exemplo 18.10, 592
- erro de, 582
- intervalo, 584
- múltiplos passos à frente, 589
- por ponto, 584
- previsão, 195
- e análise de resíduos, 195
 - erro de, 198
 - intervalo, 198
 - intervalos de confiança da, 195
- previsões, 582
- com um passo à frente, 584
 - da taxa de desemprego norte-americana, Exemplo 18.8, 585
- processos de tendências, sazonais e integrados, 592
- tipos de modelos de regressão utilizados nas, 584
- primeira diferença, 357
- primeira diferenciação, efeitos fixos ou, 439
- probabilidade de resposta, 235
- problemas de especificação nos modelos Tobit, 536
- procedimento de dois estágios de Engle-Granger, 581
- processo AR(1) estável, 343
- processo auto-regressivo de ordem um [AR(1)], 343
- processo de covariância estacionária, 341
- processo de média móvel de ordem um [MA(1)], 343
- processo de tendência estacionária, 344
- processo estacionário, 341
- processo estocástico estacionário, 341
- processo estocástico, 307
- processo não-estacionário, 341
- produção agrícola, efeitos dos fertilizantes sobre a, Exemplo 1.3, 13
- produção de soja e fertilizantes, Exemplo 2.1, 21
- produção por hora trabalhada nos Estados Unidos, Figura 10.2, 328
- produtividade dos trabalhadores e subsídios de treinamento de pessoal
- Exemplo 1.2, 4
 - Exemplo 15.10, 482
- produtividade, salários e, Exemplo 11.7, 359-360
- prognosticando salários dos diretores executivos
- Exemplo 6.7, 202
 - Exemplo 6.8, 203
- programa de merenda escolar, desempenho em matemática de estudantes e o, Exemplo 2.12, 50
- propensão de impacto, 309
- propensão de longo prazo (PLP), 309
- propriedade de computadores pessoais, determinantes da, Exemplo 8.9, 268
- propriedades algébricas das estatísticas MQO, 36
- propriedades assimptóticas, 158
- do MQO, 345
- propriedades da amostra finita do MQO, sob hipóteses clássicas, 311
- propriedades de amostras grandes, 158
- prova de teoremas
- 3.1, 106
 - 3.2 e 3.4, 108

proximidade da faculdade como uma VI da educação, utilizando a, Exemplo 15.4, 467
 pseudos R -quadrados, 524
 p -valor, 127
 Figura 4.6, 127
 dos testes F , calculando, 146
 dos testes t , cálculo dos, 126

Q

QI como uma *proxy* de aptidão, Exemplo 9.3, 280
 quase-experimento, 412

R

raiz do erro quadrático médio (REQM), 588
 raízes unitárias, 567
 no log do produto interno bruto real dos Estados Unidos, Exemplo 18.4, 571
 processo de, 354
 teste das taxas das letras do Tesouro norte-americano para três meses, Exemplo 18.2, 568
 teste de, 567
 teste de, da inflação norte-americana anual, Exemplo 18.3, 570
 razão t , 115
 razão F , 140
 R -barra quadrado, 190
 redimensionamento de dados efeitos do, Tabela 6.1, 175 nas estatísticas MQO, efeitos do, 174
 registros de prisões, explicando os, Exemplo 3.5, 78
 regra de rejeição a 5% Figura 4.2, 117
 Figura 4.3, 119
 Figura 4.4, 122

regra de rejeição, 116
 regressando, 20
 regressão auxiliar, 167
 regressão de Poisson do número de prisões, Exemplo 17.3, 540
 regressão dos preços de imóveis, Exemplo 7.4, 215
 regressão espúria 330, 572
 auto-regressivo vetorial (VAR), 586
 com duas variáveis independentes, 64
 com funções quadráticas, 182
 com k variáveis independentes, 67
 com termos de interação, 187
 duas variáveis, regressão linear de, 20
 efeito não observado, 415
 estático, 308
 irrestrito, 139
 problema da, 573
 regressão linear simples, 20
 regressão simples, 20
 Exemplo 11.1, 346
 subespecificar o, 86
 verdadeiro, 81
 regressão linear, 44
 regressão múltipla estimativas, comparação das regressão simples e, 75
 funcionalidade da, 64
 modelo de, estimativa de VI do, 464
 sobre o significado de “manter outros fatores fixos” na, 73
 terminologia para a, Tabela 3.1, 68
 uma interpretação parcial da análise, 64
 regressão simples incorporação de não-linearidades, 41

terminologia para, Tabela 2.1, 21
 regressão truncada, modelos de, 542
 censurada e, 542, 543
 regressão através da origem, 58, 79
 com uma tendência temporal, uma interpretação da remoção da tendência da, 320
 funções de regressão entre grupos, verificação de diferenças na, 227
 linha de, Figura 17.4, 548
 resultados da, descrição dos, 149
 regressor, 20
 regressores estritamente exógenos, 373
 testes de hipóteses sobre um único parâmetro populacional, 113
 relação quadrática, Figura 6.1, 184
 relação salários-benefícios de professores, Exemplo 4.10, 150-151
 relação salários-benefícios, teste da, Tabela 4.1, 151
 relações lineares entre as variáveis independentes, 93
 remoção da tendência, 334
 interpretação de regressões com uma tendência temporal, 334
 remuneração de diretores executivos e desempenho de empresas, Exemplo 6.4, 193
 rendimentos e títulos, modelo de correção de erro dos, Exemplo 18.7, 550
 RESET (teste de erro de especificação de regressão), 275

- como um teste geral da má especificação da forma funcional, 275
- resíduo, 29, 74
valores estimados e, 34
Figura 2.4, 29
Tabela 2.2, 35
valores estimados e, MQO, 74
- resposta de solução de canto, 517
- restrições de exclusão, 137, 468, 500
teste de, 137
- restrições de sobreidentificação, 477
teste de, 477
- restrições lineares gerais, teste de, 148
- restrições múltiplas, 137
- restrições sobreidentificadoras, teste de a endogeneidade e teste de, 475
- resultado de eleições e desempenho econômico, Exemplo 10.6, 326
- resultados eleitorais e gastos de campanha
Exemplo 2.5, 33
Exemplo 2.9, 39
- resumo das formas funcionais envolvendo logaritmos, Tabela 2.3, 44
- resumo de estatísticas, Tabela 2, 619
- reta de regressão de MQO, 30, 70 e a função de regressão populacional, Figura 2.5, 32
- retorno da educação e a diferença salarial por gênero, mudanças no, Exemplo 13.2, 406
mudou no transcorrer do tempo, Exemplo 14.2, 437
- para mulheres casadas, estimativa do, Exemplo 15.1, 459
- para homens, estimativa do, Exemplo 15.2, 460
- para mulheres que trabalham Exemplo 15.5, 471
Exemplo 15.7, 476
Exemplo 15.8, 478
medindo o, Exemplo 1.4, 14
- retorno de ações, salários dos diretores-executivos e Exemplo 2.3, 31
Exemplo 2.6, 35
Exemplo 2.8, 39
- retornos de ações, ARCH nos, Exemplo 12.9, 395
- revisão da literatura, 604
- R*-quadrado ajustado, 190
escolha entre modelos não-aninhados, 191
- R*-quadrado, 38
cálculo após a estimativa da VI, 464
correlação serial e, 371
forma da estatística *F*, 144, 145
heteroscedasticidade e, 243
quando a variável dependente segue uma tendência, calculando, 334
- S**
- salário de diretores executivos e retornos de ações
Exemplo 2.3, 31
Exemplo 2.6, 35
Exemplo 2.8, 39
- salário de diretores executivos e vendas das empresas, Exemplo 2.11, 43
- salário mínimo porto-riquenho, Exemplo 12.7, 391
- salário mínimo
desemprego, e dados relacionados, para Porto Rico, Tabela 1.3, 9
efeito do, sobre o desemprego, Exemplo 1.6, 16
emprego em Porto Rico e, Exemplo 10.3, 320
porto-riquenho, Exemplo 12.7, 391
- salários de diretores executivos, prognosticando Exemplo 6.7, 202
Exemplo 6.8, 203
- salários e educação Exemplo 2.4, 32
Exemplo 2.7, 36
- salários e educação, gráfico de, Figura 2.3, 28
- efeitos da atratividade física sobre os, Exemplo 7.7, 219
- efeitos da utilização de computadores nos, Exemplo 7.9, 222
- e outras características individuais, conjunto de dados de corte transversal sobre, Tabela 1.1, 7
- e produtividade, Exemplo 11.7, 359-360
Figura 2.6, 42
- salários, efeitos da raça sobre, dos jogadores de beisebol, Exemplo 7.11, 226
- sazonalidade, 336
ajustada, 336
tendências e, 327
- seleção amostral endógena, 293
seleção amostral exógena, 293, 550
- semi-elasticidade, 44

- serialmente não-correlacionados, 360
 séries temporais altamente persistentes, 352
 na análise de regressão, 352
 transformações nas, 357
 séries temporais estacionárias e fracamente dependentes, 340
 séries temporais estacionárias e não-estacionárias, 341
 séries temporais
 caracterizando tendências de, 327
 dados, 8, 306
 equações, aplicação do MQ2E, 479
 estacionárias e fracamente dependentes, 340
 estacionárias e não-estacionárias, 341
 ϵ I(1), decidindo se, 358
 fracamente dependentes, 342
 modelos de equações simultâneas com, 506
 modelos, hipótese de homoscedasticidade em, 363
 modelos de regressão, exemplos de, 307
 processo de, 307
 regressões, heteroscedasticidade nas, 393
 serviços de busca *on-line*, 604-605
 significância econômica, ou prática, *versus* estatística, 129
 nível de, 116
 significância econômica, 129
 significância estatística econômica ou prática, *versus*, 129
 significância global de uma regressão, 147
 significância prática, 129
 simultaneidade, 491
 sistema de duas equações, identificação em um, 498
 sistemas com mais de duas equações, 505
 sistemas com três ou mais equações, identificação em, 505
 solução plugada do problema de variáveis omitidas, 279
 soma dos quadrados dos resíduos (SQR), 29, 77
 minimizando, 62
 soma dos quadrados dos resíduos (SQR), 37, 77
 minimizando, 62-63
 soma dos quadrados explicada (SQE), 37, 77
 soma dos quadrados total (SQT), 37, 77
 suavização exponencial, 582
 subespecificando o modelo, 86
 subsídios de treinamento de pessoal sobre as taxas de rejeição das empresas, efeitos dos, Exemplo 4.7, 130
 subsídios de treinamento nas horas de treinamento, efeito da, Exemplo 7.3, 213
 Sumário do viés em $\hat{\beta}_j$, Tabela 3.2, 88
 superdispersão, 540
 superespecificar o modelo, 86
- T**
- tamanho das empresas, intensidade de pesquisa e desenvolvimento e, Exemplo 9.8, 296
- tamanho de escolas, desempenho de estudantes e, Exemplo 4.2, 119
 taxa de assassinatos e tamanho da força policial, Exemplo 16.1, 494
 taxa de crescimento, 329
 taxa de criminalidade sobre a taxa de esclarecimento de crimes, defasagens distribuídas da, Exemplo 13.6, 420
 taxa de desemprego
 lista parcial de dados, sobre taxas de inflação e de, dos Estados Unidos, Tabela 10.1, 307
 previsão da norte-americana, Exemplo 18.8, 585
 previsão para dois anos à frente da, Exemplo 18.10, 592
 taxas de crescimento econômico e características de países, conjunto de dados sobre, Tabela 1.2, 8
 taxas de criminalidade municipal na Carolina do Norte, Exemplo 13.9, 428
 taxas de criminalidade nas cidades, Exemplo 9.4, 284
 taxa de esclarecimento de crimes, defasagens distribuídas da taxa de criminalidade sobre a, Exemplo 13.6, 420
 taxas de juros
 efeitos da inflação e dos déficits sobre as, Exemplo 10.2, 319
 parâmetro de co-integração das, 579
 taxas de participação nos planos de pensão dos Estados Unidos, Exemplo 4.6, 129

- t**
 taxas de rejeição
 efeitos das subvenções da treinamento de pessoal nas firmas, Exemplo 4.7, 130
 efeito do treinamento de pessoal nas firmas Exemplo 14.1, 435
 Exemplo 14.3, 441
 erro de medida nas, Exemplo 9.6, 287
 taxas estaduais de mortalidade infantil, Exemplo 9.10, 298
 tendência exponencial, 329
 tendência temporal linear, 327
 tendência temporal, 327
 tendências e sazonalidade, 327
 Teorema 10.4, 317
 variâncias dos estimadores MQO e o, 315
 Teorema 15.A.1-15.A.3, 489-490
 Teorema de Gauss-Markov, 99
 Teorema 3.4, 100
 terminologia
 para a regressão múltipla, Tabela 3.1, 68
 para a regressão simples, Tabela 2.1, 21
 nota sobre, 34
 termo de interação, 222
 termo erro (perturbação), 21, 67
 teste contra, 116
 teste bi-caudal, 121
 teste de Breusch-Godfrey, 379
 teste de Breusch-Pagan da heteroscedasticidade (teste BP), 253
 teste de Chow de diferença estrutural ao longo do tempo, 407
 teste de co-integração sem tendência temporal, valores críticos assimptóticos do, Tabela 18.4, 576
 tendência temporal linear, valores críticos
 assimptóticos do, Tabela 18.5, 577
 valores críticos
 assimptóticos do, Tabela 18.5, 577
- d**
 dados, 616
 escrever, 613
 estrutura conceitual (ou teórica), 613
 introdução, 613
 modelos econométricos e métodos de estimação, 614
 resultados, 617
 sugestões de estilo, 618
 transformação interna, 434
 treinamento de pessoal e produtividade dos trabalhadores Exemplo 1.2, 4
 Exemplo 15.10, 482
 sobre as taxas de rejeição das empresas, efcitos dos Exemplo 14.1, 435
 Exemplo 14.3, 441
 truncamento ocasional, 549, 551
- U**
 única variável explicativa endógena, 468
 única variável explicativa, teste de endogeneidade da, 476
- V**
 valor base, 323
 valor crítico de 5% e a região de rejeição, Figura 4.7, 142
 valor crítico, 116
 valor estimado, 28
 valores críticos assimptóticos do teste t de raiz unitária sem tendência temporal, Tabela 18.2, 568
 tendência temporal linear, Tabela 18.3, 571
 valores críticos assimptóticos para teste de co-integração sem tendência temporal, Tabela 18.4, 576
 tendência temporal linear, Tabela 18.5, 577

- valores estimados e resíduos, 34
 Figura 2.4, 29
 Tabela 2.2, 35
 $\text{Var}(\hat{\beta}_1)$ como uma função de R^2_1 ,
 Figura 3.1, 94
 $\text{Var}(\text{salário}|\text{educ})$ crescendo com
 educ , Figura 2.9, 53
 variação amostral na variável
 independente, hipótese RLS.4,
 47
 de $\hat{\beta}_1$, Figura 5.1, 159
 dos estimadores de MQO, 110
 dos estimadores MQO de
 inclinação, Teorema 3.2, 92
 em MQO, Teorema 2.2, 54
 variação amostral total, 93
 variância assimptótica, 164
 variância do erro, 52, 92
 adição de regressores para
 reduzir a, 194
 estimação da, 55
 variância
 amostragem MQO, Teorema
 10.2, 317
 do erro, 52, 92
 estimação da, 55
 cm modelos mal-especificados,
 96
 dos estimadores de MQO, 51,
 91
 do erro de previsão, 198
 e o Teorema de Gauss-
 Markov, 315
 variáveis de tendência na análise
 de regressão, utilizando, 330
 variáveis *dummy* anuais, 403
 variáveis *dummy* sazonais, 336
 variáveis endógenas, 493
 variáveis exógenas, 464, 493
 variáveis explicativas *dummy*
 quando a variável dependente
 é $\log(y)$, interpretando os
 coeficientes de, 214
- variáveis explicativas endógenas
 múltiplas, 472
 variáveis explicativas exógenas,
 82, 466
 variáveis explicativas, 20
 erro de medida nas, 287
 na equação (10.3), do exemplo
 de X, Tabela 10.2, 312
 utilizando variáveis *proxy* para
 variáveis não-observadas,
 278
 variáveis instrumentais (VI),
 453, 455
 com uma variável instrumental
 pobre, propriedades das,
 462
 da educação, utilizando a pro-
 ximidade da faculdade
 como, Exemplo 15.4, 467
 soluções de problemas de
 erros nas variáveis, 473
 variáveis instrumentais exógenas,
 hipótese MQ2E.3, 489
 variáveis omitidas, 453
 em um modelo de regressão
 simples, 454
 variáveis ordinais, 218
 variável binária, 207
 variável de contagem, 537
 variável de controle, 20
 variável de previsora, 20
 variável de resposta, 20
 variável dependente binária: o
 modelo de probabilidade
 linear, 230
 variável dependente defasada, 283
 correlação serial na presença
 da, 371
 variável dependente é $\log(y)$
 interpretação dos coeficientes
 de variáveis *dummy*
 explicativas quando, 214
 previsão de y quando, 201, 202
- variável dependente limitada
 (VDL), 517
 modelos, erro-padrão
 assimptótico em, 166
 variável dependente, 21
 erro de medida na, 285
 \inf , Tabela 12.2, 386
 $\log(\text{chepi})$, Tabela 12.1, 384
 $\log(\text{salário})$
 Tabela 15.1, 468
 Tabela 9.2, 281
 n_{pris86} , Tabela 9.1, 274
 p_{oup} , Tabela 8.1, 260
 $r_{\text{preço}}$, Tabela 13.2, 410-411
 Tabela 9.3, 284
 variável *dummy* independente
 única, 209
 variável *dummy*, 207, 320
 armadilha, 209
 de inúmeras categorias, 216
 envolvendo interações, 221
 incorporando informação
 ordinal usando, 218
 regressão, 437
 variável endógena defasada, 508
 variável explicada, 20
 variável explicativa endógena,
 82, 272, 453
 variável independente, 20
 relações lineares entre, 93
 variando mais de uma
 simultaneamente, 73
 variável irrelevante, inclusão de
 uma, 85
 variável predeterminada, 508
 variável prevista, 20
 variável *proxy*, 278
 utilizando variáveis
 dependentes defasadas
 como, 283
 variável relevante, excluir, 86
 variável zero-um, 207
 variável
 binária, 207

- de controle, 20
- de múltiplas categorias, utilizando *dummys*, 216
- dependente, 20
- de previsão, 20
- de resposta, 20
- descrições da, Tabela 1, 619
- dummys*, 207
- explicada, 20
- explicativa, 20
- explicativa endógena, 82, 272
- explicativa exógena, 82
- inclusão de, irrelevantes, 85
- independente, 20
- independente *dummy* única, 209
- modelo com duas independentes, 64
- modelo com k independentes, 67
- ordinal, 218
- prevista, 20
- zero-um, 207
- vendas das empresas, salário dos diretores executivos e, Exemplo 2.11, 43
- vetor
- modelo auto-regressivo (VAR), 586
- viés assintótico, 161
- viés de atenuação, 290
- viés de heterogeneidade, 415
- viés de simultaneidade, 497 no MQO, 496
- viés de variável omitida, 88 casos mais gerais, 90
- geral, 107
- o caso simples, 86
- viés para baixo, 89
- viés para cima, 89
- viés
- em $\tilde{\beta}_1$ sumário do, Tabela 3.2, 88
- viesado para zero, 89
- W**
- Wald, estatística de, 522
- Z**
- zonas industriais sobre os pedidos de auxílio-desemprego, efcito das, Exemplo 13.8, 427