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

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

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

1. Econometria II. Título.

10-11298 CDD-330.015195

**Índices para catálogo sistemático:**

1. Econometria 330.015195

# INTRODUÇÃO À ECONOMETRIA
## Uma Abordagem Moderna

**Jeffrey M. Wooldridge**
Michigan State University

Tradução da quarta edição norte-americana

**Tradução**
José Antônio Ferreira

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

CENGAGE Learning

Austrália Brasil Canadá Cingapura Espanha Estados Unidos México Reino Unido

# 14

# Métodos Avançados de Dados em Painel

Neste capítulo, tratamos de dois métodos para estimar modelos de efeitos não observados de dados em painel que são pelo menos tão comuns quanto a primeira diferença. Embora esses métodos sejam um pouco mais difíceis de serem descritos e implementados, vários programas econométricos os suportam.

Na Seção 14.1, discutiremos o estimador de efeitos fixos que, como a primeira diferença, usa transformação para remover o efeito não observado $a_i$ antes da estimação. Quaisquer variáveis explicativas constantes no tempo são removidas com $a_i$.

O estimador de efeitos aleatórios na seção 14.2 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 de erro composto, mas não causa correlação entre os erros compostos e as variáveis explicativas. A estimação de modelos de efeitos aleatórios por mínimos quadrados generalizados é bastante fácil e normalmente feita por muitos programas econométricos.

Na Seção 14.3, mostraremos como os métodos de dados em painel podem ser aplicados em outras estruturas de dados, inclusive em amostras pareadas e por agrupamentos.

## 14.1 ESTIMAÇÃO DE EFEITOS FIXOS

A primeira diferença 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 esse método envolve, considere um modelo com uma única variável explicativa: para cada $i$,

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \quad t = 1, 2, \ldots, T. \tag{14.1}$$

Agora, para cada $i$, calculamos a média dessa equação ao longo do tempo. Obtemos

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i, \tag{14.2}$$

em que $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$, 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 $t$, ficamos com

$$\ddot{y}_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i, \qquad t = 1, 2, ..., T,$$

ou

$$\ddot{y}_{it} = \beta_1\ddot{x}_{it} + \ddot{u}_{it}, \qquad t = 1, 2, ..., T, \tag{14.3}$$

em que $\ddot{y}_{it} = y_{it} - \bar{y}_i$ são os **dados centrados na média** de $y$ e, de maneira análoga, $\ddot{x}_{it}$ e $\ddot{u}_{it}$. A transformação de efeitos fixos também é chamada de **transformação intragrupo**. 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 intragrupo**. Este último nome vem do fato de o MQO em (14.3) usar a variação temporal 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, $\beta_0$): utilizamos as médias de tempo tanto de $y$ como de $x$ e depois executamos uma regressão de corte transversal. Não estudaremos em detalhes o estimador porque ele é viesado quando $a_i$ é correlacionado com $\bar{x}_i$ (veja Problema 14.2 no final deste capítulo). 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 de efeitos não observados** original é

$$y_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + ... + \beta_k x_{itk} + a_i + u_{it}, \qquad t = 1, 2, ..., T. \tag{14.4}$$

Simplesmente usamos a centralização na média 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 centralização na média. A equação de centralização na média geral para cada $i$ é

$$\ddot{y}_{it} = \beta_1\ddot{x}_{it1} + \beta_2\ddot{x}_{it2} + ... + \beta_k\ddot{x}_{itk} + \ddot{u}_{it}, \qquad t = 1, 2, ..., T, \tag{14.5}$$

que estimamos pelo MQO agrupado.

Sob uma hipótese de exogeneidade estrita das variáveis explicativas, o estimador de efeitos fixos é não viesado: de certa forma, 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 diferença. 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 sexo 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 de centralização na média (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, deveremos aparentemente ter $NT - k$ graus de liberdade. Esse cálculo é incorreto. Para cada observação $i$ do corte transversal, perdemos um $gl$ em razão da centralização na média. Em outras palavras, para cada $i$, os erros centrados $\ddot{u}_{it}$ 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 estimação de efeitos fixos calculam corretamente os $gl$. Entretanto, se tivermos que fazer, por nós mesmos, a centralização na média e a estimação pelo MQO agrupado, precisaremos corrigir os erros-padrão e as estatísticas de testes.



**QUESTÃO 14.1**

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 isso pode 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, outras 10 empresas 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(*ref*)** | |
|---|---|
| **Variáveis independentes** | **Coeficiente (erro-padrão)** |
| *d88* | −0,080<br>(0,109) |
| *d89* | −0,247<br>(0,133) |
| *subs* | −0,252<br>(0,151) |
| $subs_{-1}$ | −0,422<br>(0,210) |
| Observações<br>Graus de Liberdade<br>*R*-quadrado | 162<br>104<br>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 centralização na média nos possibilita estimar $\beta_j$, 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 ≈ −0,344]; o coeficiente de $subs_{-1}$ é significante no nível de 5% contra uma alternativa bilateral. O coeficiente de *subs* é significante no nível de 10%, e o tamanho do coeficiente não é nada desprezível. Observe que os *gl* 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 duradouro 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 presumindo 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.

**QUESTÃO 14.2**

De acordo com o programa do estado norte-americano de Michigan, se uma empresa recebeu 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 calcular um indicador de qualidade de ajuste. O *R*-quadrado dado na Tabela 14.1 é baseado na transformação intragrupo: 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, incluindo 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 em 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 diferença), 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(*salárioh*) como a variável dependente, 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 no 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 *gl* 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 maneira tradicional de ver o método de efeitos ajustados é presumir 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 $\beta_j$. (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 em 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 $\beta_j$ que obteríamos da regressão dos dados centrados na média, 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 da regressão dessas variáveis *dummy* 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 em razão do 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 do Capítulo 13 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 seguro-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 calcular os testes $F$ da maneira habitual, presumindo, 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 irrestrito 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:

$$\hat{a}_i = \bar{y}_i - \hat{\beta}_1 \bar{x}_{i1} - \ldots - \hat{\beta}_k \bar{x}_{ik},\ i = 1, \ldots, N, \tag{14.6}$$

em que 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 centralização na média 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 (EF) 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 que é um estimador não viesado e consistente de $\alpha = E(a_i)$.

Na maioria dos estudos, os $\hat{\beta}_j$ são de interesse, e assim as equações de dados centrados na média 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 em painel são da espécie $N$ grande e $T$ pequeno.



## Efeitos Fixos ou Primeira Diferença?

Até agora, sem considerar o MQO agrupado, vimos dois métodos para estimar modelos de efeitos não observados. Um deles envolve a diferenciação dos dados e o outro a centralização na média. Como saber qual deles usar?

Podemos eliminar um caso imediatamente: quando $T = 2$, as estimativas EF e PD, como também todos os testes estatísticos são *idênticos*, e portanto não importa qual usamos. Claro, a equivalência entre as estimativas EF e PD exige que estimemos o mesmo modelo em cada caso. Em particular, como discutimos no Capítulo 13, é natural incluirmos um intercepto na equação PD: este intercepto será, na verdade, o intercepto do segundo período de tempo no modelo original escrito para dois períodos temporais. Portanto, a estimação EF deve incluir uma variável *dummy* para o segundo período de tempo com o propósito de ser idêntica às estimativas PD que incluem um intercepto.

Com $T = 2$, as PD têm a vantagem de serem simples de serem implementados em qualquer pacote de programa econométrico ou estatístico que suporte manipulação básica de dados, e é fácil calcular a estatística de heteroscedasticidade robusta após a estimação PD (pois quando $T = 2$, a estimação PD é apenas uma regressão do corte transversal).

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, $u_{it}$. (Consideraremos 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 diferença (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 —, então a diferença $\Delta u_{it}$ será serialmente não correlacionada, e a primeira diferença 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 estimação EF: podemos estimar os erros centrados na média, $\ddot{u}_{it}$, mas não os $u_{it}$. Porém, na Seção 13.3 do Capítulo 13, mostramos como verificar se os erros diferenciados, $\Delta u_{it}$, são serialmente não correlacionados. Se parecer ser esse o caso, é possível usar PD. Se houver correlação serial negativa substancial em $\Delta u_{it}$, EF provavelmente será melhor. Sempre é bom tentar ambos: pouco importa 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 diferença tem a vantagem de transformar um processo integrado de séries temporais em um processo fracamente dependente. Portanto, se aplicarmos a primeira diferença, 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.

Do mesmo modo que o estimador de primeiras diferenças, o estimador de efeitos ajustados pode ser bastante sensível ao erro de medição clássico em uma ou mais variáveis explicativas. Porém, se cada $x_{itj}$ for não correlacionada com $u_{it}$, mas a hipótese de exogeneidade estrita for de alguma forma infringida — por exemplo, uma variável dependente defasada é incluída entre os regressores ou existe retroalimentação entre as $u_{it}$ e os futuros resultados da variável explicativa — então o estimador EF provavelmente terá substancialmente menos viés do que o estimador PD (a menos que $T = 2$). O fato teórico importante é que o viés no estimador PD não depende de $T$, enquanto o estimador EF tende à zero na razão de $1/T$. Para detalhes, veja Wooldridge (2002, Seção 11.1).

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 em painel, especialmente de pessoas ou empresas, estão ausentes certos anos em pelo menos algumas unidades do corte transversal na amostra. Nesse caso, chamamos o conjunto de dados em **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 centralização na média. O número total de observações será, então, $T_1 + T_2 + \ldots + T_N$. Como no caso equilibrado, um grau de liberdade será perdido em cada observação de corte transversal em razão da centralização na média. Qualquer programa de regressão que faça efeitos fixos leva 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 centralização na média 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 diferença; 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. Idealmente, podemos usar uma análise de dados em 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 (conhecido como atrito) 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 o atrito da amostra seja correlacionado 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 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 da empresa 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 em razão da 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 do atrito da amostra em dados em 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,

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}, \quad \textbf{(14.7)}$$

em que explicitamente incluímos um intercepto de maneira que possamos presumir 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 diferença, a meta é eliminar $a_i$, porque ele supostamente estará correlacionado com um ou mais dos $x_{itj}$. 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 presumimos que o efeito não observado $a_i$ é não correlacionado com cada variável explicativa:

$$\mathrm{Cov}(x_{itj}, a_i) = 0, \quad t = 1, 2, \ldots, T; j = 1, 2, \ldots, k. \quad \textbf{(14.8)}$$

Aliás, as hipóteses ideais de efeitos aleatórios 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, deveremos usar a primeira diferença ou os efeitos fixos.

Sob (14.8) e juntamente com as hipóteses dos efeitos aleatórios, como devemos estimar os $\beta_j$? É importante ver que, se acreditarmos que $a_i$ seja não correlacionado com as variáveis explicativas, os $\beta_j$ podem ser consistentemente estimados com o uso de um único corte transversal: não precisamos dos dados em 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: execute o MQO de $y_{it}$ sobre as variáveis explicativas e provavelmente sobre as *dummies* temporais. Isso também produz estimadores consistentes dos $\beta_j$ sob a hipótese de efeitos aleatórios. Porém, isso ignora uma característica fundamental do modelo. Se definirmos o **termo de erro composto** como $v_{it} = a_i + u_{it}$, (14.7) pode ser escrita como

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + v_{it}. \tag{14.9}$$

Como $a_i$ é o erro composto em cada período de tempo, os $v_{it}$ são serialmente correlacionados ao longo do tempo. De fato, sob as hipóteses de efeitos aleatórios,

$$\text{Corr}(v_{it}, v_{is}) = \sigma_a^2/(\sigma_a^2 + \sigma_u^2), \qquad t \neq s,$$

em que $\sigma_a^2 = \text{Var}(a_i)$ e $\sigma_u^2 = \text{Var}(u_{it})$. Essa correlação serial (necessariamente) positiva no termo de 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 autorregressiva. 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. Presumimos 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

$$\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_a^2)]^{1/2}, \tag{14.10}$$

que está entre zero e um. Em seguida, a equação transformada resultará em

$$\begin{aligned} y_{it} - \lambda\bar{y}_i = {} & \beta_0(1 - \lambda) + \beta_1(x_{it1} - \lambda\bar{x}_{i1}) + \ldots \\ & + \beta_k(x_{itk} - \lambda\bar{x}_{ik}) + (v_{it} - \lambda\bar{v}_i), \end{aligned} \tag{14.11}$$

em que a barra superior novamente representa as médias temporais. Essa é uma equação bastante interessante, por envolver **dados quase centrados na média** 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 $\sigma_u^2$, $\sigma_a^2$ 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 no final deste Capítulo.)

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 diferença. Isso é possível porque EA presume 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, presumimos que educação não se correlaciona com $a_i$, que contém aptidão e antecedentes familiares. Em muitas aplicações, a principal razão do uso de dados em painel é possibilitar que o efeito não observado seja correlacionado com as variáveis explicativas.

Na prática, o parâmetro $\lambda$ nunca é conhecido, mas sempre pode ser estimado. Existem maneiras diferentes de fazer isso, que podem se basear no MQO agrupado ou efeitos fixos, por exemplo. Em geral, $\hat{\lambda}$ toma a forma $\hat{\lambda} = 1 - \{1/[1 + T(\hat{\sigma}_a^2/(\hat{\sigma}_u^2)]\}^{1/2}$, em que $\hat{\sigma}_a^2$ é um estimador consistente de $\sigma_a^2$, e $\hat{\sigma}_u^2$ é um estimador consistente de $\sigma_u^2$. Esses estimadores podem estar baseados nos resíduos do MQO agrupado ou dos efeitos fixos. Uma possibilidade é que $\hat{\sigma}_a^2 = [NT(T-1)/2-(k+1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}$, em que os $\hat{v}_{it}$ são os resíduos de estimar (14.9) pelo MQO agrupado. Em seguida, podemos estimar $\sigma_u^2$ usando $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_a^2$, em que $\hat{\sigma}_v^2$ 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 econométricos suportam a estimação de modelos de efeitos aleatórios e automaticamente computam alguma versão de $\hat{\lambda}$. O estimador MQG factível que utiliza $\hat{\lambda}$ em lugar de $\lambda$ é chamado de **estimador de efeitos aleatórios**. Sob as hipóteses dos efeitos aleatórios no Apêndice deste capítulo, o estimador é consistente (não viesado) e distribuído normalmente e assimptoticamente 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 $\sigma_u^2$). É mais comum ser $\sigma_a^2$ grande em relação a $\sigma_u^2$, 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 ideia 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 estimação dos efeitos aleatórios ponderam o efeito não observado em $(1 - \lambda)$. Embora a correlação entre $a_i$ e um ou mais dos $x_{itj}$ cause inconsistência na estimação 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 $\lambda$ estiver próximo de zero, estaremos deixando uma grande fração do efeito não observado no termo de erro e, como consequência, o viés assimptótico do estimador EA será maior.

Nas aplicações de EF e EA, é geralmente instrutivo também calcular as estimativas agrupada de MQO. A comparação desses três conjuntos de estimativas pode nos auxiliar a determinar a natureza dos vieses causados por termos deixado o efeito não observado, $a_i$, inteiramente no termo de erro (como faz os MQO agrupados) ou parcialmente no termo de erro (como faz a transformação EA). Mas devemos nos lembrar que, mesmo se $a_i$ for não correlacionada com todas as variáveis explicativas em todos os períodos de tempo, os erros-padrão dos MQO agrupados e dos testes estatísticos serão, geralmente, inválidos: eles ignoram a frequentemente substancial correlação serial nos erros compostos, $v_{it} = a_i + u_{it}$. Como mencionamos do Capítulo 13 (veja o Exemplo 13.9), é possível calcular os erros-padrão e testes estatísticos que sejam robustos quanto à correlação serial arbitrária (e heteroscedasticidade) na $v_{it}$, e pacotes de programas estatísticos populares muitas vezes permitem esta opção. [Veja, por exemplo, Wooldridge (2002, Capítulo 10).]

**EXEMPLO 14.4**

**(Uma Equação de Salários Usando Dados em 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 fixos. As variáveis de variação temporal são *exper*, *exper*$^2$, *sindicato* e *casado*. Como detalhado na Seção 14.1, a variável *exper* é abandonada na análise EF (mas *exper*$^2$ permanece). Cada regressão também contém um conjunto completo de *dummies* anuais. Os resultados da estimação estão na Tabela 14.2.

**Tabela 14.2**

Três estimadores diferentes de uma equação de salários.

| Variável dependente: log(*salário*) | | | |
|---|---|---|---|
| **Variáveis independentes** | **MQO agrupado** | **Efeitos aleatórios** | **Efeitos fixos** |
| *educ* | 0,091<br>(0,005) | 0,092<br>(0,011) | — |
| *negro* | −0,139<br>(0,024) | −0,139<br>(0,048) | — |
| *hispan* | 0,016<br>(0,021) | 0,022<br>(0,043) | — |
| *exper* | 0,067<br>(0,014) | 0,106<br>(0,015) | — |
| *exper*$^2$ | −0,0024<br>(0,0008) | −0,0047<br>(0,0007) | −0,0052<br>(0,0007) |
| *casado* | 0,108<br>(0,016) | 0,064<br>(0,017) | 0,047<br>(0,018) |
| *sindicato* | 0,182<br>(0,017) | 0,106<br>(0,018) | 0,080<br>(0,019) |

Os coeficientes em *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 estimação pelos efeitos aleatórios. Quando eliminamos totalmente o efeito não observado usando efeitos fixos, o ágio de casado cai para cerca de 4,7%, embora ainda seja estatisticamente significante. A queda do ágio de casado é consistente com a ideia de que os homens mais capazes — como capturado por um efeito não observado mais alto, $a_i$ — são, com maior probabilidade, casados.



**EXEMPLO 14.4 (continuação)**

Portanto, na estimação por MQO agrupado, uma grande parte da magnitude do ágio 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 $\lambda$ para a estimação pelos efeitos aleatórios é $\hat{\lambda} = 0{,}643$, que ajuda a explicar a razão pela qual, nas variáveis de variação temporal, as estimativas por EA ficam mais próximas das estimativas por EF do que das estimativas por MQO agrupado.

**QUESTÃO 14.3**

O ágio 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 Ajustados?

Como os efeitos ajustados permitem correlação arbitrária entre as $a_i$ e as $x_{itj}$, enquanto os efeitos aleatórios não a permitem, os EF são largamente considerados uma ferramenta mais convincente para se estimar efeitos *ceteris paribus*. Mesmo assim, os efeitos aleatórios são aplicados em certas situações. Mais evidentemente, se a principal variável explicativa for constante ao longo do tempo, não poderemos usar os EF para estimarmos seus efeitos na $y$. Por exemplo, na Tabela 14.2, temos de nos valer das estimativas dos EA (ou dos MQO agrupados) do retorno da educação. É claro, somente podemos usar os efeitos aleatórios, pois estamos querendo presumir que o efeito não observado é não correlacionado com as variáveis explicativas. Caracteristicamente, se usarmos efeitos aleatórios, tantos controles constantes no tempo quanto possível serão incluídos entre as variáveis explicativas. (Com uma análise EF não será necessário incluirmos tais controles). Os EA são preferidos aos MQO agrupados, pois os EA geralmente são mais eficientes.

Se nosso interesse estiver numa variável explicativa com variação temporal, haverá algum caso em que será preferível usarmos os EA em vez dos EF? Sim, mas situações nas quais $\mathrm{Cov}(x_{itj}, a_i) = 0$ devem ser consideradas exceção e não a regra. Se a variável política principal for definida experimentalmente — digamos, cada ano as crianças são aleatoriamente designadas a salas de aula de diferentes tamanhos — então os efeitos aleatórios seriam apropriados para estimarmos o efeito do tamanho da sala de aula no desempenho escolar. Infelizmente, na maioria dos casos os regressores são, eles próprios, resultados de processos selecionados e propensos a serem correlacionados com preferências individuais e talentos como capturadas pelas $a_i$.

Ainda é bastante comum vermos pesquisadores aplicarem tanto os efeitos aleatórios como os efeitos ajustados e depois fazerem testes formais das diferenças estatisticamente significantes nos coeficientes das variáveis explicativas com variação temporal. (Assim, na Tabela 14.2, seriam os coeficientes nas *exper*$^2$, *casado*, e *sindicato*). Hausman (1978) foi quem primeiro propôs tal teste e alguns pacotes econométricos calculam rotineiramente o teste de Hausman sob o conjunto total das hipóteses de efeitos aleatórios listadas no Apêndice deste Capítulo. A ideia é que se usem as estimativas de efeitos aleatórios a menos que o teste de Hausman as rejeite (14.8). Na prática, uma falha em rejeitar significa que ou as estimativas EA e as EF são suficientemente próximas que não importa qual será usada, ou a variação amostral é tão grande nas estimativas EF que não se pode concluir se diferenças praticamente significantes são estatisticamente significantes. No último caso, imagina-se se há informação suficiente nos dados para produzir estimativas precisas dos coeficientes. Uma rejeição com o uso do teste de Hausman é considerada como significativa de que a principal hipótese EA (14.8) é falsa e que as estimativas EF são usadas. (Naturalmente, em todas as aplicações de inferência estatística deve-se distinguir entre uma diferença praticamente significante e uma diferença estatisticamente significante.) [Para mais detalhes, veja Wooldridge (2002, Seção 10.7)].

Uma palavra final de alerta: ao ler trabalho empírico você poderá verificar que alguns autores decidem-se entre EF e EA com base em se as $a_i$ são propriamente vistas como parâmetros das estimativas ou como variáveis aleatórias. Essas considerações normalmente são errôneas. Neste capítulo, temos tratado as $a_i$ como variáveis aleatórias no modelo de efeitos não observáveis (14.17), independentemente de como decidimos estimar as $\beta_j$. Como enfatizamos, o item principal que determina se usaremos EF ou EA é se podemos plausivelmente presumir que as $a_i$ são não correlacionadas com todas as $x_{itj}$. No entanto, em algumas aplicações de métodos de dados em painel, não podemos tratar nossa amostra como uma amostra aleatória de uma grande população, especialmente quando a unidade de observação for uma unidade geográfica grande (digamos estados ou municípios). Então, com frequência, faz sentido pensarmos em cada $a_i$ como um intercepto separado para estimar cada unidade de seção transversal. Nesse caso, usamos os efeitos ajustados: lembre-se, usar os EF é, mecanicamente, o mesmo que permitir um intercepto diferente para cada unidade de seção transversal. Felizmente, quer nos envolvamos no debate filosófico da natureza das $a_i$, quer não, os EF são, quase sempre, muito mais convincentes que os EA na análise de política usando dados agregados.

## 14.3 A APLICAÇÃO DE MÉTODOS DE DADOS EM PAINEL A OUTRAS ESTRUTURAS DE DADOS

Os vários métodos de dados em painel podem ser aplicados em certas estruturas de dados que não envolvam tempo. Por exemplo, é comum na demografia usar irmãos (algumas vezes gêmeos) para explicar as características familiares e culturais não observadas. Normalmente queremos permitir que o "efeito familiar", que é comum a todos os irmãos em uma família, seja correlacionado com as variáveis explicativas observadas. Se essas variáveis explicativas variam entre os irmãos em uma família, o diferenciamento entre pares de irmãos — ou, de forma mais geral, usar a transformação interna dentro de uma família — é preferido como um método de estimação. Removendo o efeito não observado, eliminamos o viés potencial causado pela confusão das características do perfil familiar. A implementação dos efeitos fixos em tais estruturas de dados é bastante simples em programas de regressão que suportam estimação EF.

Como exemplo, Geronimus e Korenman (1992) usaram pares de irmãs para estudar os efeitos da gravidez na adolescência sobre as consequências econômicas futuras. Quando o resultado é renda em relação às necessidades — algo que depende do número de filhos —, o modelo é

$$\begin{aligned}\log(rendanec_{fs}) = \beta_0 + \delta_0 irm\tilde{a}2_s + \beta_1 partoad_{fs} \\ + \beta_2 idade_{fs} + outros\ fatores + a_f + u_{fs},\end{aligned} \quad \textbf{(14.12)}$$

em que $f$ indexa as famílias e $s$ indexa uma irmã dentro da família. O intercepto da primeira irmã é $\beta_0$, 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 as consequências econômicas. A solução desse problema é simples: dentro de cada família, aplicamos a diferença de (14.12) entre as irmãs para obter

$$\Delta\log(rendanec) = \delta_0 + \beta_1 \Delta partoad + \beta_2\, \Delta idade + \ldots + \Delta u, \quad \textbf{(14.13)}$$

o que remove o efeito familiar, $a_f$, e a equação resultante poderá ser estimada por MQO. Observe que não há nenhum elemento temporal neste caso: a diferença 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.

**QUESTÃO 14.4**

Ao usar o método da diferença, faz sentido a inclusão de variáveis *dummy* para a etnia da mãe e do pai em (14.12)? Explique.

Utilizando 129 pares de irmãs do *National Longitudinal Survey of Young Women* (Estudo Longitudinal Nacional de Mulheres Jovens) de 1982, Geronimus e Korenman primeiramente estimaram $\beta_1$ 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 Geronimus e Korenman (1992, Tabela 3)]. 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 diferença 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 diferença 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 foram 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 apresenta-se maior que a estimada por MQO agrupado, de 8,4% (que controla sexo, idade e etnia). Ashenfelter e Krueger também estimaram a equação por efeitos aleatórios e obtiveram um retorno da educação de 8,7%. A análise de efeitos aleatórios é, mecanicamente, a mesma que a do caso de dados em painel com dois períodos de tempo.

As amostras usadas por Geronimus e Korenman (1992) e Ashenfelter e Krueger (1994) são exemplos de **amostras pareadas**. De forma geral, os métodos de efeitos fixos e aleatórios podem ser aplicados a uma **amostra por agrupamento**. Esses são conjuntos de dados de corte transversal, mas cada observação pertence a um agrupamento bem definido. Nos exemplos anteriores, cada família é um agupamento. Como outro exemplo, suponha que temos dados sobre a participação em vários planos de previdência, nos quais as empresas oferecem mais de um plano aos seus funcionários. Podemos, então, ver cada empresa como um agrupamento, e é bastante claro que os efeitos não observados das empresas serão fatores importantes na determinação de taxas de participação nos planos de previdência dentro das empresas.

Dados educacionais sobre estudantes escolhidos aleatoriamente de muitas escolas formam uma amostra por agrupamento, na qual cada escola é um agrupamento. Como os resultados em um agrupamento são propensos a serem correlacionados, geralmente é importante considerar um efeito de agrupamento não observado. A estimação por efeitos fixos é preferível quando pensamos que o **efeito de agrupamento** não observado — do qual um exemplo é $a_f$ 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 o mesmo tamanho, de modo que geralmente são requeridos métodos de efeitos fixos para painéis não equilibrados.

Em alguns casos as variáveis explicativas cruciais — frequentemente variáveis políticas — mudam somente ao nível do grupo, não no interior de grupo. Em tais casos o método de efeitos ajustados não é aplicável. Por exemplo, podemos estar interessados nos efeitos da qualidade medida do professor no desempenho estudantil, em que cada agrupamento é uma sala de aula do curso primário. Como todos os alunos dentro de um agrupamento têm o mesmo professor, a eliminação de um "efeito classe" também elimina qualquer medida da qualidade do professor observada. Se tivermos bons controles na equação, pode ser justificável a aplicação de efeitos aleatórios no grupo desequilibrado. Como acontece com dados em painel, o requerimento principal para que o EA produza estimativas convincentes é que as variáveis explicativas sejam não correlacionadas com o efeito de agrupamento não observado. A maioria dos programas econométricos possibilitam a estimação de efeitos aleatórios em agrupamentos desequilibrados sem muito esforço.

MQO agrupados também são comumente aplicados em amostras por agrupamentos quando a eliminação de um efeito de agrupamento via efeitos ajustados é irrealizável ou indesejável. Porém, como acontece com os dados em painel, os erros-padrão dos MQO habituais serão incorretos a menos que não haja efeito de agrupamento, e assim erros-padrão robustos que permitam "correlação de agrupamento" (e heteroscedasticidade) devem ser usados. 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 em painel que descrevemos no Exemplo 13.9 do Capítulo 13. 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 por agrupamento, como também a heteroscedasticidade, no modelo de probabilidade linear.



## RESUMO

Estudamos dois métodos comuns para estimar modelos de dados em painel com efeitos não observados. Comparado com a primeira diferença, 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 diferença, 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 presumir que os motivos pelos quais alguns períodos de tempo estão faltando não são sistematicamente relacionados aos erros idiossincráticos.

O estimador de efeitos aleatórios é adequado 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 de 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 centrados na média. 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 diferença ou dos efeitos fixos (que são todos não viesados, consistentes e assimptoticamente normais).

Finalmente, os métodos de dados em painel estudados nos Capítulos 13 e 14 podem ser usados quando trabalhamos com amostras pareadas ou por agrupamentos. A diferença ou a transformação intragrupo eliminam o efeito de agrupamento. Se o efeito de agrupamento 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.

## PROBLEMAS

**14.1** Suponha que os erros idiossincráticos em (14.4), $\{u_{it}: t = 1, 2, \ldots, T\}$ sejam serialmente não correlacionados com variância constante, $\sigma_u^2$. Mostre que a correlação entre as diferenças adjacentes $\Delta u_{it}$ e $\Delta u_{i,t+1}$ é $-0{,}5$. Portanto, sob as Hipóteses EF ideais, a primeira diferença induz uma correlação serial negativa de valor conhecido.

**14.2** Com uma única variável explicativa, a equação usada para obter o estimador intragrupo é

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_i,$$

em que a barra superior representa a média ao longo do tempo. Podemos presumir 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 $\text{Cov}(x_{it}, a_i) = \sigma_{xa}$ para todo $t$ (e $i$, em razão da 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

$$\text{plim } \tilde{\beta}_1 = \beta_1 + \sigma_{xa}/\text{Var}(\bar{x}_i),$$

em que limite de probabilidade é definido como $N \rightarrow \infty$. [*Sugestão*: Veja as equações (5.5) e (5.6).]

(ii) Suponha também que $x_{it}$, para todo $t = 1,2,\ldots,T$, seja não correlacionado com a variância constante $\sigma_x^2$. Mostre que 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 composto $v_{it} = a_i + u_{it}$, em que $a_i$ é não correlacionado com $u_{it}$ e os $u_{it}$ têm variância constante $\sigma_u^2$ e são serialmente não correlacionados. Defina $e_{it} = v_{it} - \lambda\bar{v}_i$, em que $\lambda$ é dado em (14.10).

(i) Mostre que $E(e_{it}) = 0$.
(ii) Mostre que $\mathrm{Var}(e_{it}) = \sigma_u^2$, $t = 1, \ldots, T$.
(iii) Mostre que para $t \neq s$, $\mathrm{Cov}(e_{it}, e_{is}) = 0$.

**14.4** Para determinar os efeitos do desempenho atlético universitário dos candidatos, você coleta dados das inscrições dos 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 de cronometragem?
(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 frequência 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 (SAT) no curso superior.

(i) Por que você classificaria esse conjunto de dados como uma amostra por agrupamento? 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 frequência 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 presumindo sobre as características não observadas dos alunos que afetam as taxas de desempenho e de frequência? Que papel, com relação a isso, desempenham a nota de ingresso no curso superior (SAT) e a nota média acumulada antes do início do semestre (GPA)?
(iv) Se você julga que a nota de ingresso no curso superior (SAT) e a nota média acumulada antes do início do semestre (GPA) não indicam adequadamente a capacidade dos alunos, como você estimaria o efeito da frequência sobre o desempenho no exame final?

**14.6** Usando a opção "agrupamento" no pacote de econometria Stata®, os erros-padrão totalmente robustos da estimação por MQO agrupado na Tabela 14.2 – isto é, robustos quanto à correlação serial e quanto à heteroscedasticidade nos erros de combinação $\{v_{it} : t = 1, \ldots, T\}$ são obtidos como

$$\mathrm{ep}(\hat{\beta}_{educ}) = 0{,}011,\ \mathrm{ep}(\hat{\beta}_{negro}) = 0{,}051,\ \mathrm{ep}(\hat{\beta}_{hispan}) = 0{,}039,\ \mathrm{ep}(\hat{\beta}_{exper}) = 0{,}020,\ \mathrm{ep}(\hat{\beta}_{exper^2}) = 0{,}0010,\ \mathrm{ep}(\hat{\beta}_{casado}) = 0{,}026,\ \text{e}\ \mathrm{ep}(\hat{\beta}_{sindicato}) = 0{,}027$$



(i) Como esses erros-padrão se comparam, de forma geral, com os não robustos e por quê?
(ii) Como os erros-padrão robustos dos MQO agrupados se comparam com os erros-padrão dos EA? Parece ter importância se a variável explicativa for constante no tempo ou de variação temporal?

## APÊNDICE 14A

### 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 considerada em Wooldridge (2002, Capítulo 10).

**HIPÓTESE EF.1**
Para cada $i$, o modelo é

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

em que os $\beta_j$ são os parâmetros a serem estimados e $a_i$ os efeitos não observados.

**HIPÓTESE EF.2**
Temos uma amostra aleatória na dimensão do corte transversal.

**HIPÓTESE EF.3**
Cada variável explicativa muda ao longo do tempo (para ao menos algum $i$), e não há relações lineares perfeitas entre as variáveis explicativas.

**HIPÓTESE EF.4**
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: $E(u_{it}|\mathbf{X}_i, a_i) = 0$.

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.4. Sob essas mesmas hipóteses, o estimador EF é consistente com um $T$ fixo conforme $N \rightarrow \infty$.

**HIPÓTESE EF.5**
$\mathrm{Var}(u_{it}|\mathbf{X}_i, a_i) = \mathrm{Var}(u_{it}) = \sigma_u^2$ para todo $t = 1, \ldots, T$.

**HIPÓTESE EF.6**
Para todo $t \neq s$, os erros idiossincráticos são não correlacionados (condicionais a todas as variáveis explicativas e $a_i$): $\mathrm{Cov}(u_{it}, u_{is}|\mathbf{X}_i, a_i) = 0$.

Sob as Hipóteses EF.1 a EF.6, o estimador de efeitos fixos dos $\beta_j$ 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 erros idiossincráticos serialmente não correlacionados.

**HIPÓTESE EF.7**
Condicional em $\mathbf{X}_i$ e $a_i$, os $u_{it}$ são independentes e identicamente distribuídos como Normal$(0,\sigma_u^2)$.

A Hipótese EF.7 implica EF.4, EF.5 e EF.6, mas é mais forte, pois presume uma distribuição normal dos erros idiossincráticos. 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 assimptóticas. 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.4, EF.5 e EF.6. (O EF.7 poderia ser adicionado, mas nos trará pouco ganho na prática pois teremos que estimar $\lambda$.) Como estaremos subtraindo somente uma fração das médias temporais, agora podemos permitir variáveis explicativas constantes no tempo. Portanto, o EF.3 é substituído pela

**HIPÓTESE EA.3**
Não existem relacionamentos lineares perfeitos entre as variáveis explicativas.

O custo de se permitir regressores constantes no tempo é que devemos adicionar hipóteses sobre como o efeito não observado, $a_i$, está relacionado com as variáveis explicativas.

**HIPÓTESE EA.4**
Em adição ao EF.4, o valor esperado da $a_i$, dadas todas as variáveis explicativas, é constante: $\mathrm{E}(a_i \mid \mathbf{X}_i) = \beta_0$.

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 presumimos que $a_i$ não é correlacionado com todos os elementos de $\mathbf{x}_{it}$, podemos incluir variáveis explicativas constantes no tempo. (Tecnicamente, a quase centralização na média remove somente uma fração da média temporal, e não a sua totalidade.) Consideramos uma expectativa diferente de zero de $a_i$ na definição da Hipótese EA.4, de forma que o modelo sob as hipóteses dos efeitos aleatórios contenha um intercepto, $\beta_0$, 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.
Também necessitamos impor homoscedasticidade em $a_i$, como segue:



**HIPÓTESE EA.5**
Adicionalmente a EF.5, a variância de $a_i$, dadas todas as variáveis explicativas, é constante: $\mathrm{Var}(a_i|\mathbf{X}_i) = \sigma_a^2$.

Sob as seis hipóteses de efeitos aleatórios (EA.1, EA.2, EA.3, EA.4, EA.5 e EA.6), o estimador EA é consistente e assimptótico e normalmente distribuído na medida em que $N$ se torna maior para $T$ fixo. Na verdade, consistência e normalidade assimptótico são derivadas sob as primeiras quatro hipóteses, mas sem as últimas duas hipóteses os erro-padrão habituais do EA e testes estatísticos não serão válidos. Além disso, sob as seis hipóteses EA, os estimadores de EA são assimptoticamente eficientes. Isso significa que, em amostras grandes, os estimadores de EA terão erros-padrão menores que os correspondentes estimadores de MQO agrupados (quando os erros-padrão robustos apropriados forem usados para os MQO agrupados). Para os coeficientes nas variáveis explicativas com variação temporal (as únicas estimáveis pelo EF), o estimador de EA é mais eficiente que o estimador de EF — frequentemente muito mais eficiente. Mas o propósito do EF não é ser eficiente sob as hipóteses EA; o EF é intencionado a ser robusto quanto à correlação entre as $a_i$ e as $x_{itj}$. Como frequentemente acontece na econometria, existe uma relação de trocas entre robustez e eficiência. Para verificação das afirmações feitas aqui, veja Wooldridge (2002, Capítulo 10)