Econometria

Parte 5

Prof. Adalto Acir Althaus Junior oe

Sumário

- Panel Data
- Como os dados em painel são úteis
  - ✓ modelo de efeitos fixos
  - ✓ Modelo de efeitos aleatórios
  - ✓ Primeiras diferenças
  - ✓ Modelos defasados (lagged)

Motivação

- Como já observamos, variáveis omitidas representam um obstáculo substancial em nossa capacidade de fazer inferências causais
- E pior ... muitos deles são inerentemente inobserváveis aos pesquisadores e analistas

Motivação

- Considere a estimativa em nível de empresa

$$leverage_{i,j,t} = \beta_0 + \beta_1 profit_{i,j,t-1} + u_{i,j,t}$$

- onde a *leverage* é dívida/ativos para a empresa i, operando na indústria j no ano t, e o *profit* é a receita líquida/ativos da empresa.

- Quais poderiam ser algumas variáveis omitidas não observáveis nessa estimativa?

Motivação

- há tantos…
  - ✓ Talento gerencial e / ou aversão ao risco
  - ✓ Indústria fornecer e / ou exigir choque
  - ✓ Custo do capital
  - ✓ oportunidades de investimento
  - ✓ E assim por diante ...
- Fácil pensar em maneiras pelas quais isso pode estar afetando a alavancagem e estar correlacionado com os lucros.

- Infelizmente, isso é fácil de fazer com outras variáveis dependentes ou independentes…

Motivação

- Usar observações de várias regiões geográficas (por exemplo, estado ou país) abre ainda mais possibilidades…
  - ✓ Pode pensar em algumas variáveis não observadas que podem estar relacionadas com a localização de uma empresa?

  - ✓ Resposta: quaisquer diferenças não observadas no ambiente econômico local, por ex. instituições, proteção dos direitos de propriedade, desenvolvimento financeiro, sentimento do investidor, choques de demanda regional, etc.

Motivação

- Às vezes, podemos controlar essas variáveis não observáveis usando variáveis proxy
  - ✓ Mas, que suposição foi exigida para que uma variável proxy pudesse fornecer estimativas consistentes sobre os outros parâmetros?

  - ✓ Resposta: ela precisa ser uma proxy suficientemente boa para que a variável não observada não possa ser correlacionada com as outras variáveis explicativas depois de controlarmos pela variável proxy... Isso pode ser difícil de encontrar

Panel Data

- Felizmente, os dados em painel podem nos ajudar com um tipo particular de variável não observada ...
  - ✓ Que tipo de variável não observada os dados do painel nos ajudam e por quê?
  - ✓ Resposta = Ajuda-nos com variáveis omitidas invariantes no tempo; agora, vamos ver porque ...
  - ✓ Na verdade, nos ajuda com qualquer variável não observada que não varia dentro de grupos de observações

Panel Data

- Dados em painel = sempre que você tiver várias observações por unidade de observação i (por exemplo, você observa cada empresa ao longo de vários anos)
  - ✓ Vamos assumir N unidades i – N empresas, por exemplo
  - ✓ E, T observações por unidade i [i.e. painel balanceado]
- Ex. # 1 - Você observa 500 empresas em Economatica durante um período de vinte anos [isto é, N = 500, T = 20]
- Ex. # 2 - Você observa 100 CEOs na Thomson durante um período de 10 anos [isto é, N = 100, T = 10]

Panel Data

Variável não observada, invariante no tempo, $\boldsymbol{f}$

- Considere o seguinte modelo:

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta f_i + u_{i,t}$$

- Onde:

$$E(u_{i,t}) = 0$$
$$corr(x_{i,t}, f_i) \neq 0$$
$$corr(f_i, u_{i,t}) = 0$$
$$corr(x_{i,t}, u_{i,s}) = 0 \text{ for all } s, t$$

Isso implica o quê? Resposta: Se não controlar por f, temos OVB, mas se pudéssemos, então não controlaríamos

- **Nota: Esta é uma suposição mais forte do que costumamos fazer; chama-se exogeneidade estrita. Em palavras, essa suposição significa o que?**
- **Resposta: O termo de erro em QUALQUER t não é correlacionado com x em nenhum t**

Panel Data

- Se ignorarmos f, teremos OVB, e estimarmos o modelo…

$$y_{i,t} = \alpha + \beta x_{i,t} + \underbrace{v_{i,t}}_{\delta f_i + u_{i,t}}$$

- x está correlacionado com o erro v (através da sua correlação com a variável não observada, f, que agora faz parte da perturbação, ou seja, do erro)
- Fácil de mostrar $\hat{\beta}^{OLS} = \beta + \boxed{\delta \frac{\sigma_{xf}}{\sigma^2{}_x}}$

Este é o OVB ... coeficiente da regressão da variavel omitida, $\boldsymbol{f}$, em $\boldsymbol{x}$ vezes o verdadeiro coeficiente de $\boldsymbol{f}$

Panel Data

Variável não observada, invariante no tempo, f

- Voltando ao modelo:

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta f_i + u_{i,t}$$

- Onde:

$$E(u_{i,t}) = 0$$

$$corr(x_{i,t}, f_i) \neq 0$$

$$corr(f_i, u_{i,t}) = 0$$

$$corr(x_{i,t}, u_{i,s}) = 0 \text{ for all } s, t$$

Isso implica o quê? Resposta: Se não controlar por f, temos OVB, mas se pudéssemos, então não controlaríamos

- **Como resolver se *f* não é observável?**

Panel Data

- **Podemos resolver transformando os dados**
- Primeiro, observe que se você pegar a média populacional da variável dependente para cada unidade de observação, você terá...

$$\overline{y}_i = \alpha + \beta \overline{x}_i + \delta f_i + \overline{u}_i$$

- Onde:

Assumindo que há $\boldsymbol{T}$ obs. por unidade $\boldsymbol{i}$

$$\overline{y}_i = \frac{1}{T}\sum_t y_{i,t}, \quad \overline{x}_i = \frac{1}{T}\sum_t x_{i,t}, \quad \overline{u}_i = \frac{1}{T}\sum_t u_{i,t}$$

Panel Data

- **Podemos resolver transformando os dados...**
- Agora, se subtrairmos $\overline{y}_i$ de $y_{i,\mathrm{t}}$, temos

$$y_{i,t} - \overline{y}_i = \beta\left(x_{i,t} - \overline{x}_i\right) + \left(u_{i,t} - \overline{u}_i\right)$$

- E olhe! A variável não observada, f_i, desapareceu
- Como é constante, f_i - $\bar{f}$ = 0, porque é invariante no tempo
- Com nossa suposição de exogeneidade estrita antes, é fácil perceber que (x_i, t − x_i) não está correlacionado com a nova perturbação, (u_i, t − u_i), o que significa...

Panel Data

- A estimativa de OLS do modelo transformado produzirá uma estimativa consistente de β
  - ✓ A transformação prévia é chamada de "within transformation" porque reduz da média todas as variáveis dentro de seu grupo
- Neste caso, o "grupo" foi o conjunto de observações ao longo do tempo para cada empresa
- Isso também é chamado de estimador FE – efeitos fixos
  - ✓ Fixed Effects

Panel Data

- Variável não observada, f é muito geral

  ✓ Não somente captura uma variável não observada; captura todas as variáveis não observadas que não variam dentro do grupo

- É por isso que muitas vezes chamamos isso de "heterogeneidade não observada"

Panel Data – Fixed Effects

- Quando você usa o estimador de efeitos fixos (FE) em programas como o Stata, ele faz a “within transformation” para você
  - ✓ Não faça isso sozinho porque ...
  - ✓ Os graus de liberdade (que são usados para obter os erros padrão) às vezes precisam ser ajustados pelo número de unidades transversais, N
  - ✓ E o ajustamento necessário depende de como você agrupa as observações, etc.

Panel Data – LSDV Fixed Effects

- Outra maneira de fazer a estimativa FE é adicionando variáveis indicadoras (dummy)

- Lembre que o coeficiente de f_i, δ, não tem realmente nenhum significado; então você pode apenas redimensionar o não observado para torná-lo igual a 1 ...

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta f_i + u_{i,t}$$

- Agora, para estimar isso, podemos apenas tratar cada f_i como um parâmetro a ser estimado

Panel Data – LSDV Fixed Effects

- crie uma variável dummy para cada unidade i e adicione-a à regressão
- Este é o **least squares dummy variable** model
- Agora, nossa equação de estimativa corresponde exatamente ao modelo subjacente verdadeiro

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta f_i + u_{i,t}$$

- Obtemos estimativas consistentes e SE que são idênticas aqueles que obteríamos com o estimador *"within transformation"*

Panel Data – LSDV Fixed Effects

- **Questão operacional...**
- Como as variáveis dummy serão colineares com a constante, uma delas será descartada na estimativa
  - ✓ Portanto, não tente interpretar o intercepto; é apenas a média y quando todos os x's são iguais a zero para o grupo correspondente à variável dummy perdida

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta f_i + u_{i,t}$$

- No comando xtreg, fe, o intercepto relatado é apenas média de interceptos específicos individuais e não tem interpretação.

Within x LSDV - Fixed Effects

- Pode-se mostrar que LSDV e FE são idênticos, usando resultados de regressão parcial [Como?]
  - ✓ Lembre-se, para controlar alguma variável z, podemos regredir y em x e z, ou podemos apenas parcialmente retirar z de ambos y e x antes de regredir y em x (ou seja, regredir os resíduos da regressão de y em z contra os resíduos da regressão de x em z)

- As variáveis reduzidas da média são os resíduos de uma regressão delas nas dummies!

Within x LSDV - Fixed Effects

- O R^2 reportado será maior com LSDV
  - ✓ Todas as variáveis dummy explicarão grande parte da variação em y, aumentando o R^2
- O R^2 usando *within transformation* como estimador de FE, apenas reporta qual proporção da variação dentro de y é explicada pela variação dentro de x
- **O R^2 usando *within transformation* geralmente é mais interessante para nós**

Within x LSDV - Fixed Effects

- O R^2 usando *within transformation* geralmente é mais interessante, pois descreve o poder explicativo de xxs [após a exclusão parcial do FE]
- Para obter o R^2 *within*, use **xtreg, fe**
- O R^2 ajustado global reportado também é útil
- Para obter o R^2 ajustado global, use o comando areg em vez de xtreg, fe. O "R^2 global" relatado por xtreg não inclui variação explicada por FE, mas o R^2 relatado por areg faz

Panel Data - Fixed Effects

- Existem muitos benefícios do estimador FE
  - ✓ Permite correlação arbitrária entre cada efeito fixo, f_i e cada x dentro do grupo i
  - ✓ isto é, é muito geral e não impõe muita estrutura sobre como os dados subjacentes devem parecer
  - ✓ Interpretação muito intuitiva; coeficiente é identificado usando apenas alterações dentro das cross sections

Panel Data - Fixed Effects

- Também é muito flexível e pode nos ajudar a controlar muitos tipos de heterogeneidades não observadas
  - ✓ Pode adicionar FE para os anos (tempo) se estiver preocupado com a heterogeneidade não observada ao longo do tempo [por exemplo, choques macroeconômicos]
  - ✓ Pode adicionar FE para CEO's se estiver preocupado com a heterogeneidade não observada entre os CEOs [por exemplo, talento, aversão ao risco, ...]
  - ✓ Adicionar FE por setor, se estiver preocupado com a heterogeneidade não observada em todos os setores ao longo do tempo [por exemplo, oportunidades de investimento, choques de demanda]

Panel Data - Fixed Effects

- O Estimador FE é muito geral
  - ✓ Aplica-se a qualquer cenário em que as observações possam ser agrupadas
- Ex. # 1 - As empresas podem ser agrupadas por setor
- Ex. # 2 - As observações dos CEOs (que podem abranger várias empresas) podem ser agrupadas por combinações CEO-empresa
- Agrupamento de unidades i ao longo do tempo é apenas um exemplo (embora o mais comum) – FE por empresa e por tempo

Panel Data - Fixed Effects

- Depois de criar grupos, você pode remover qualquer heterogeneidade não observada a nível de grupo adicionando FE desse grupo

- A consistência requer apenas que haja um grande número de grupos

Panel Data - Fixed Effects

- **Mas o estimador FE também tem seus custos...**
  - ✓ Não é possível identificar variáveis que não variam dentro do grupo
  - ✓ Sujeito a viés de erro de medida potencialmente grande
  - ✓ Pode ser difícil estimar em alguns casos
  - ✓ questões diversas

Panel Data - Fixed Effects

- **Problema 1...**
  - ✓ Se não houver variação dentro do grupo na variável independente, x, de interesse, não será possível desvinculá-lo do grupo de FE
  - ✓ Ela é colinear com o grupo FE; e será descartada por seu computador ou excluída na *"within transformation"*

Panel Data - Fixed Effects

- **Problema 1...**
- Considere seguir a estimativa do nível de CEO

$$\ln(totalpay)_{ijt} = \alpha + \beta_1 \ln(firmsize)_{ijt} + volatility_{ijt}$$

$$+\beta_3 female_i + \delta_t + f_i + \lambda_j + u_{ijt}$$

  - ✓ $Ln\ (totalpay)$ é para CEO i, firma j, ano t
  - ✓ Estimação inclui FE para ano, CEO e empresa
- Que coeficiente não pode ser estimado?
  - ✓ Resposta: β_3! Ser mulher não varia dentro do grupo de observações de cada CEO; ou seja, é colinear com o efeito fixo do CEO

Panel Data - Fixed Effects

- **Problema 1...**
- Tenha cuidado com isso!
  - ✓ Programas como o stata (xtreg) são bons em eliminar a variável female (por exemplo) e não reportar uma estimativa...
- Mas, se você mesmo criar variáveis dummy e inseri-las, o resultado da estimação pode excluir uma delas em vez do indicador female
  - ✓ **Isto é você receberá uma estimativa para β_3, mas não tem significado algum! É apenas um valor do intercepto aleatório que depende inteiramente do FE aleatório excluído pelo Stata**

Panel Data - Fixed Effects

- **Problema 2...**
- Erro de medida da variável independente (e vieses resultantes) pode ser amplificado
- Pense que pode haver dois tipos de variação
  - ✓ Boa variação (significativa)
  - ✓ Variação de ruído porque não medimos perfeitamente a variável de interesse subjacente
- Adicionar FE pode excluir muito da boa variação; e fração da variação remanescente vinda do ruído aumentará [O que isso fará?]

Panel Data - Fixed Effects

- **Problema 2...**
- Resposta: viés de atenuação na variável mal mensurada vai subir!
- Conselhos práticos:
  - ✓ tenha cuidado ao interpretar coeficientes ‘zero’ em regressores potencialmente mal mensurada; pode ser apenas viés de atenuação!
  - ✓ E lembre-se, o sinal do viés em outros coeficientes será geralmente difícil de saber

Panel Data - Fixed Effects

- **Problema 3...**
- Estimar um modelo com vários tipos de FE pode ser computacionalmente difícil
  - ✓ Quando usar mais de um tipo de FE, não é possível remover ambos usando a *"within transformation"*
  - ✓ Geralmente, você só pode excluir um tipo de FE com a *within transformation*; o outro FE deve ser tratado adicionando variáveis dummy ao modelo
- Por exemplo: efeitos fixos de empresas e de ano

Panel Data - Fixed Effects

- **Problema 3...**
- Considere o modelo:

FE por year

FE por firma

$$y_{i,t} = \alpha + \beta x_{i,t} + \boxed{\delta_t} + \boxed{f_i} + u_{i,t}$$

- Para estimar isso no Stata, usaríamos um comando como o seguinte:
  - ✓ xtset firma
  - ✓ xi: xtreg y x i.year, fe

Diz ao Stata que a dimensão do painel é dada pela variável firma

Diz Stata para remover FE para a variável dimensão do painel (ou seja, firma), fazendo a within transformation

Diz ao Stata para criar e adicionar variáveis dummy para a variável year

Panel Data - Fixed Effects

- **Problema 3...**
- Dummies não excluídas pela within transformation serão realmente estimadas
  - ✓ Com FE de ano, isso não é problema porque não há muitos anos de dados
  - ✓ Se tivesse que estimar 1.000 FE de firmas, poderia ser um problema
- Na verdade, é por isso que usamos FE (within) por empresa em vez de FE por ano; Existem mais empresas!

Modelos não lineares - Fixed Effects

- Como não obtemos estimativas consistentes do FE, não podemos estimar modelos não lineares de dados de painéis com FE
  - ✓ Na prática, Logit, Tobit, Probit não deve ser estimado com muitos efeitos fixos
  - ✓ Eles só fornecem estimativas consistentes sob suposições bastante fortes e irrealistas

Panel Data - Random Effects

- **Efeitos aleatórios**
- Bem similar aos efeitos fixos

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + u_{i,t}$$

- Mas uma grande diferença ...
  - ✓ Supõe-se que a heterogeneidade não observada, fi, e os $x's$ observados não sejam correlacionada

Panel Data - Random Effects

- Considere o modelo:

efeito fixo a ser tratado pela técnica de efeitos aleatórios

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + a_i + u_{it}$$

- Sob a hipótese de que o efeito fixo é não correlacionado com cada variável independente:

$$Cov(x_{itj}, a_i) = 0,\ \ t = 1,\ldots,T;\ \ j = 1,\ldots,k$$

- O modelo é chamado de modelo de efeitos aleatórios.

Panel Data - Random Effects

- Neste caso, assumindo também que u_{it} seja independente de x_{itj}, é possível estimar os β_j de forma não viesada por *Pooled* MQO.

- No entanto, este estimador considera que há apenas um termo de erro no modelo ($v_{it} = a_i + u_{it}$) e ignora o importante fato de que este termo tem a seguinte correlação serial:

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

Efeitos aleatórios

- Isto faz com que os desvios-padrão de estimadores de *Pooled* MQO não estejam corretos, comprometendo sua inferência e eficiência.

- Para corrigir este problema pode-se lançar mão de um estimador do tipo MQP (ponderados). Para isto, considere o seguinte termo:

$$\lambda = 1 - \left[\sigma_u^2 / \sigma_u^2 + T\sigma_a^2\right]^{1/2}$$

Panel Data - Random Effects

- Em seguida considere a seguinte transformação no modelo:

$$y_{it} - \lambda \bar{y}_i = \beta_0 (1 - \lambda) + \beta_1 (x_{it1} - \lambda \bar{x}_{it1}) + \dots$$
$$+ \beta_k (x_{itk} - \lambda \bar{x}_{itk}) + (v_{it} - \lambda \bar{v}_i)$$

- Chamada de transformação *quasi time demeaning*.
- A aplicação dos estimadores de *Pooled* MQO neste novo modelo gera os estimadores de MQP que procuramos.
- Este estimador é chamado de estimador de efeitos aleatórios (RE em inglês).

Panel Data - Random Effects

- Os termos contidos em λ são, como se sabe, desconhecidos, pois são as variâncias dos componentes do erro.

- É possível estimar lambda usando estimadores das variâncias dos componentes do erro, a saber:

$$\hat{\sigma}_a^2 = [NT(T-1)/2-(k+1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_a^2$$

$$\hat{\sigma}_v^2 = [NT-(k+1)]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^2$$

Panel Data - Random Effects

- Repare que quando λ se aproxima de zero, o estimador de RE aproxima-se o *Pooled* MQO e quando λ se aproxima de 1, o RE fica próximo do estimador de FE.

- Repare ainda que ambos os estimadores eliminam o viés causado pela presença do efeito fixo, com a diferença de que um o faz de forma mais eficiente.

Panel Data - Random Effects

- Um primeiro ponto a destacar é que as estimativas dos desvios-padrão dos coeficientes são muito menores que as estimativas do modelo de efeitos fixos, mostrando que, de fato, o estimador de efeitos aleatórios é mais eficiente.

- Por outro lado, os coeficientes estimados, apesar de terem o mesmo sinal, são de magnitude muito diferente dos coeficientes do modelo de efeitos fixos, o que pode evidenciar a presença de efeitos fixos correlacionados com as variáveis independentes, os quais não são controlados pelo estimador RE e, portanto, devem torná-lo viesado.

Panel Data - Random Effects

- **Efeitos aleatórios**
- Bem similar aos efeitos fixos, mas....

$$y_{i,t} = \alpha + \beta x_{i,t} + f_i + u_{i,t}$$

- Supõe-se que a heterogeneidade não observada, fi, e os $x's$ observados não sejam correlacionada
  - ✓ O que isso implica na consistência do OLS?
  - ✓ Essa é uma suposição realista, especialmente em economia?

Panel Data - Random Effects

- Resposta # 1 - Essa suposição significa que o OLS lhe daria uma estimativa consistente de β!
- Então, por que se incomodar?
  - ✓ Resposta... potencial ganho de eficiência em relação ao FE
  - ✓ FE não é mais o estimador mais eficiente. Se nossa suposição estiver correta, podemos obter uma estimativa mais eficiente, não eliminando o FE e fazendo os mínimos quadrados generalizados [Nota: não é possível fazer apenas OLS; também será consistente, mas o SE estará errado, já que eles ignoram a correlação serial]

Panel Data - Random Effects

- Resposta # 2 - A suposição de que f e x não são correlacionados é provavelmente irrealista, avalie com cuidado!
  - ✓ A violação desta suposição é toda motivação por trás da razão pela qual fazemos a estimativa do FE!
  - ✓ **Lembre-se de que a correlação entre variáveis não observadas, como talento gerencial, choques de demanda, etc., e x causará viés de variável omitida**

RE x FE

- Em geral, a escolha entre qual estimador usar está associada a que hipótese faz mais sentido:
  - ✓ Se há razões para acreditar que o efeito fixo é não correlacionado com x_{itj}, então o melhor é usar RE.
  - ✓ Mas se acredita-se que o efeito fixo é correlacionado com algum x_{itj}, então recomenda-se o uso de FE.

- Inclusive é possível mostrar que a comparação entre os parâmetros estimados pelos dois métodos **é uma forma de testar se a_i é ou não correlacionado com x_{itj}**, desde que valha a hipótese de que u_t é estritamente exógeno a x_{itj}.

RE x FE

- Vamos testar a existência de efeitos fixos comparando os coeficientes estimados por FE (não viesados na presença destes efeitos) e por RE (viesados quando há efeitos fixos).
- Se os coeficientes forem suficientemente diferentes, então deve haver um viés associado às estimativas de RE e, portanto, deve haver efeitos fixos. Esta é a ideia por trás do **Teste de Hausman**.
- O teste de Hausman, aplicado a este modelo, aponta para um forte viés das estimativas de RE, ou seja, aponta para a existência de efeitos fixos no modelo.
- Com isto, para obter estimativas não viesadas dos parâmetros do modelo recomenda-se o uso do estimador FE, lembrando, porém, que suas estimativas não serão tão precisas quanto as de RE.

RE x FE

- **Teste de Hausman.**

- Execute a estimativa sob FE
  - ✓ xtreg y x1 x2 ...xn, fe
- Em seguida armazene os resultados na variável fe:
  - ✓ estimates store fe
- Execute a estimativa sob RE
  - ✓ xtreg y x1 x2 ...xn, re
- Em seguida armazene os resultados na variável fe:
  - ✓ estimates store re
- Execute o teste de Hausman
  - ✓ Hausman fe re
  - ✓ Se P<0,05 → fe deve ser melhor, pois rejeitamos a hipótese de que a diferença entre as regressões fe e re não é significativa

Panel Data – First Diference Model

- Modelos de primeira diferença é outra maneira de remover heterogeneidades não observadas
  - ✓ Em vez de subtrair a média do grupo da variável de cada variável, você subtrai a observação defasada
- Fácil de ver porque isso também funciona...

Panel Data – First Diference Model

- Note que $y_{i,t} = \alpha + \beta x_{i,t} + f_i + u_{i,t}$

$$y_{i,t-1} = \alpha + \beta x_{i,t-1} + f_i + u_{i,t-1}$$

- Podemos ver que

$$y_{i,t} - y_{i,t-1} = \beta(x_{i,t} - x_{i,t-1}) + (u_{i,t} - u_{i,t-1})$$

- Quando a estimativa de OLS fornecerá uma estimativa consistente de β?
  - ✓ Resposta: Com a mesma exogeneidade estrita da hipótese de FE (isto é, $x_{i,\mathrm{t}}$ e $u_{i,\mathrm{t}}$ não são correlacionados para todos os t e i)

Panel Data – First Diference Model

- tomando a primeira diferença das variáveis do modelo entre os dois períodos chegamos a uma equação do tipo:

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

- Esta equação é então chamada de equação em primeira diferença e, na prática, transforma o painel de dois períodos em uma cross-section simples formada pelas diferenças entre os períodos

Panel Data – First Diference Model

- Sendo assim, podemos aplicar um estimador de MQO (agora sem o termo de erro a_i) sobre estas novas variáveis para estimar β_1.

- O estimador de MQO quando aplicado a esta situação é chamado de estimador de primeira diferença (FD, do inglês).

- Mas, apesar de muito útil para eliminar os efeitos fixos no tempo, como já dito, o estimador FD necessita que Δx e Δu sejam independentes.

Panel Data – First Diference Model

- Isto implica que x_{it} e deve ser independente dos valores passados, presentes e futuros de u_{it}, o que é conhecido como hipótese de exogenidade estrita.

- No caso do *Pooled* MQO basta que x_{it} seja independente de u_{it} em cada período, ou seja, contemporaneamente (exogenidade contemporânea), hipótese bem menos restritiva em muitos casos

Panel Data – First Diference Model

- Esta hipótese de exogenidade estrita é uma extensão da Hip3 dos modelos de *cross-section*. As demais hipóteses também tem seus análogos para os dados em painel:
  - ✓ A variável explicativa deve variar não só entre as observações, mas no tempo (caso contrário delta $x_i = 0$ para todo i), o que é uma extensão da Hip4
  - ✓ Para fins de eficiência e inferência, precisamos garantir que a homocedasticidade valha para todo período t, uma extensão da Hip5.

Panel Data – FE vs FD

- **FE x FD ...**
- Quando existem apenas duas observações por grupo, elas são idênticas uma à outra
- Em outros casos, ambos são consistentes; a diferença é geralmente sobre eficiência
  - ✓ FE é mais eficiente se as perturbações, $u_{i,\mathrm{t}}$, são serialmente não correlacionadas
  - ✓ FD é mais eficiente se a perturbação, $u_{i,\mathrm{t}}$, seguir um passeio aleatório
- Qual delas é verdade?
  - ✓ Isso dificilmente é claro. A verdade é que provavelmente a hipótese mais realista deve ser algo entre elas

Panel Data – FE vs FD

- Se a exogeneidade estrita é violada (isto é, $x_{i,\mathrm{t}}$ é correlacionada com $u_{i,\mathrm{t}}$ para $i \neq t$), FE deve ser melhor
  - ✓ Enquanto acreditarmos que $x_{i,\mathrm{t}}$, e $u_{i,\mathrm{t}}$ não são correlacionados, a inconsistência do FE diminui para 0 na taxa 1 / T; mas, FD não melhora com T maior
  - ✓ Lembre-se: T é o número de observações por grupo (que é o nr., de períodos de tempo nesse caso)
- Mas, se y e x são espuriamente correlacionados, e N é pequeno, T grande, FE pode ser muito ruim

Panel Data – FE vs FD

- Resumindo: não é uma má ideia experimentar os dois...
  - ✓ Se os resultados forem diferentes, você deve tentar entender por quê
  - ✓ Com uma variável omitida ou erro de medição, você receberá respostas diferentes com FD e FE
  - ✓ Na verdade, Griliches e Hausman (1986) mostram que, como a medição causa vieses previsivelmente diferentes em FD e FE, você pode (sob certas circunstâncias) usar as estimativas viesadas para identificar o parâmetro verdadeiro.

Panel Data – Lagged Models

- **Variáveis defasadas...**
- Não podemos estimar facilmente modelos com uma variável dependente defasada e FE não observado

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + f_i + u_{i,t}, \qquad |\rho| < 1$$

- O mesmo que antes, mas agora o modelo verdadeiro contém y defasado como variável independente
  - ✓ Não é possível estimar com o OLS, mesmo que x & f não estejam correlacionados
  - ✓ Não é possível estimar com FE

Panel Data – Lagged Models

- **Variáveis defasadas...**
- Para ver o problema com o OLS, suponha que você calcule o seguinte:

$$y_{i,t} = \alpha + \rho y_{i,t-1} + \beta x_{i,t} + \underbrace{v_{i,t}}_{f_i + u_{i,t}}$$

- Mas: $y_{i,t-1} = \alpha + \rho y_{i,t-2} + \beta x_{i,t-2} + f_i + u_{i,t-2}$

- Assim, $y_{i,t-1}$ e erro composto, $v_{i,t}$ são correlacionados positivamente porque ambos contêm f_i
  - ✓ você terá viés de variável omitida

Panel Data – Lagged Models

- **Variáveis defasadas...**
- Vou pular a matemática, mas é sempre viesado
  - ✓ A ideia básica é que, se você fizer uma *within transformation*, a média defasada de y, que estará nos regressores do modelo agora, será sempre correlacionada negativamente com o erro reduzido da média, u
  - ✓ Nota # 1: Isto é verdade mesmo se não houvesse heterogeneidade não observada, f; FE com valores defasados é sempre uma má ideia
  - ✓ Nota # 2: O mesmo problema se aplica à primeira diferença
- Problema, no entanto, desaparece quando T vai ao infinito

Panel Data – Lagged Models

- **Variáveis defasadas...**
- Judson e Owen (1999) mostram que o viés não é grande para painéis balanceados com T> 30;
  - ✓ Problema é: painéis balanceados não são muito comuns na prática;
  - ✓ Forçar painéis a serem balanceados (isto é, excluir observações que não aparecem em todos os períodos) é muito provável que cause viés de seleção.
- **Basicamente, Não coloque a variável dependente defasada como um dos regressores quando usar dados em painel**

Panel Data

$$y_{i,t} = \alpha + \beta x_{i,t} + \delta f_i + u_{i,t}$$

Panel Data - within transformation

	Ano	Variável Y	Variável X1	Variável X2	...	Variável Xn
Empresa A	2000	100	1	500		30
Empresa A	2001	120	5	300		80
Empresa A	2003	150	3	800		90
Empresa A	20nn	140	2	100		120

média pelo tempo 127,5 2,75 425 80

	Ano	Variável Y	Variável X1	Variável X2	...	Variável Xn
Empresa A	2000	-27,5	-1,75	75		-50
Empresa A	2001	-7,5	2,25	-125		0
Empresa A	2003	22,5	0,25	375		10
Empresa A	20nn	12,5	-0,75	-325		40
Empresa B	2000	...	...	...	...	...
Empresa B	2001	...	...	...	...	...
Empresa B	2003	...	...	...	...	...
Empresa B	20nn	...	...	...	...	...

$$y_{i,t} - \overline{y}_i = \beta\left(x_{i,t} - \overline{x}_i\right) + (u_{i,t} - \overline{u}_i)$$

Tipos de Dados

Cross Sectional

	Ano 2000					
	Variável Y	Variável X1	Variável X2	Variável X3	...	Variável Xn
Empresa A						
Empresa B						
Empresa C						
Empresa D						
Empresa E						
Empresa F						
...	...	...	...	...	...	...
Empresa Z						

Time Series

	Empresa A					
	Variável Y	Variável X1	Variável X2	Variável X3	...	Variável Xn
Ano 2000						
Ano 2001						
Ano 2002						
Ano 2003						
Ano 2004						
Ano 2005						
...	...	...	...	...	...	...
Ano 20nn						

Panel Data

	Ano	Variável Y	Variável X1	Variável X2	...	Variável Xn
Empresa A	2000					
Empresa A	2001					
Empresa A	...					
Empresa A	20nn					
Empresa B	2000					
Empresa B	2001					
Empresa B	...					
Empresa B	20nn					
Empresa C	2000					
Empresa C	2001					
Empresa C	...					
Empresa C	20nn					
...	...	...	...	...	...	...
Empresa Z	2000					
Empresa Z	2001					
Empresa Z	...					
Empresa Z	20nn					

Tipos de Dados - comandos básicos stata

Cross Sectional

	Ano 2000					
	Variável Y	Variável X1	Variável X2	Variável X3	...	Variável Xn
Empresa A						
Empresa B						
Empresa C						
Empresa D						
Empresa E						
Empresa F						
...	...	...	...	...	...	...
Empresa Z						

reg

, vce(rob) ou rob

Time Series

	Empresa A					
	Variável Y	Variável X1	Variável X2	Variável X3	...	Variável Xn
Ano 2000						
Ano 2001						
Ano 2002						
Ano 2003						
Ano 2004						
Ano 2005						
...	...	...	...	...	...	...
Ano 20nn						

Não tratamos neste curso

Panel Data

	Ano	Variável Y	Variável X1	Variável X2	...	Variável Xn
Empresa A	2000					
Empresa A	2001					
Empresa A	...					
Empresa A	20nn					
Empresa B	2000					
Empresa B	2001					
Empresa B	...					
Empresa B	20nn					
Empresa C	2000					
Empresa C	2001					
Empresa C	...					
Empresa C	20nn					
...	...	...	...	...	...	...
Empresa Z	2000					
Empresa Z	2001					
Empresa Z	...					
Empresa Z	20nn					

xtreg

Pooled ols

reg

Inferência pode estar incorreta

, vce(rob) ou rob

, vce(cluster empresa)

, vce(cluster setor)

Interpretação em log

- Quadro resumo:

Modelo	Variável Dependente	Variável Independente	Interpretação de β_1	Interpretação de β_1
Nível-nível	y	x	$\Delta y = \beta_1 \Delta x$	Uma variação de 1 (unidade) na variável X causa uma variação de β_1 (unidades) na variável Y
Nível-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$	Uma variação de 1% na variável X causa uma variação de β_1 (unidades) na variável Y
Log-nível	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$	Uma variação de 1 (unidade) na variável X causa uma variação de β_1 % na variável Y
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \%\Delta x$	Uma variação de 1% na variável X causa uma variação de β_1 % na variável Y

* $p < 0{,}10$
** $p < 0{,}05$
*** $p < 0{,}01$