

Análise de Dados em Paineis

Prof. Alexandre Gori Maia

Universidade Estadual de Campinas

Disciplina: Econometria Aplicada II

Universidade Nacional Agraria La Molina

Ementa

Dados empilhados

Efeitos Fixos

Efeitos Aleatórios

Bibliografia

Baltagi, B. Econometric analysis of panel data. Third Edition. John Wiley & Sons. 2005, Chapters 1-4.

Wooldridge, J. M. 2001. Econometric analysis of cross section and panel data. Cap. 10.

Dados em Painel - Definição

- Exemplos de amostras:

Corte Transversal (*Cross-Section*)

$$Y_i$$

$$i = 1, 2, \dots, n$$

Unidades distintas observadas em um mesmo período do tempo

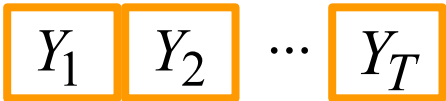


Séries Temporais (*Time Series*)

$$Y_t$$

$$t = 1, 2, \dots, T$$

Mesma unidade observada em diferentes períodos do tempo



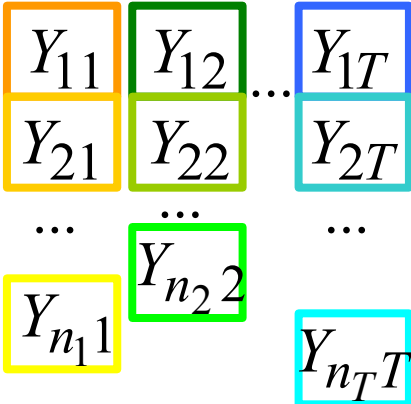
Dados Empilhados (*Pooled*)

$$Y_{it}$$

$$i = 1, 2, \dots, n_T$$

$$t = 1, 2, \dots, T$$

Unidades não necessariamente idênticas em diferentes períodos do tempo



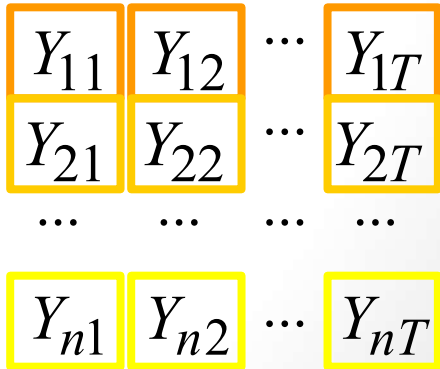
Dados em Painel (*Panel Data*)

$$Y_{it}$$

$$i = 1, 2, \dots, n$$

$$t = 1, 2, \dots, T$$

Mesmas unidades em diferentes períodos do tempo



Dados em Painel - Definição

- Exemplos de amostras de dados em painel:

Painel Balanceado (balanced)

Y_{it}

O mesmo número de períodos para cada unidade de corte transversal.

Y_{11}	Y_{12}	...	Y_{1T}
Y_{21}	Y_{22}	...	Y_{2T}
...	...	...	...
Y_{n1}	Y_{n2}	...	Y_{nT}

Painel Desbalanceado (Unbalanced)

Y_{it}

Uma unidade de corte transversal não é necessariamente observada em todos os períodos do tempo.

Y_{11}	Y_{12}	...	
Y_{21}		...	Y_{21}
...	...	...	...
	Y_{n2}	...	Y_{n3}

Painel Rotativo (Rotating Panel)

Y_{it}

Há rotação entre as unidades, de tal forma que cada uma seja observada em apenas um determinado número de períodos.

Y_{11}			
Y_{21}	Y_{22}		
	Y_{32}	...	
			Y_{n-1T}
			Y_{nT}

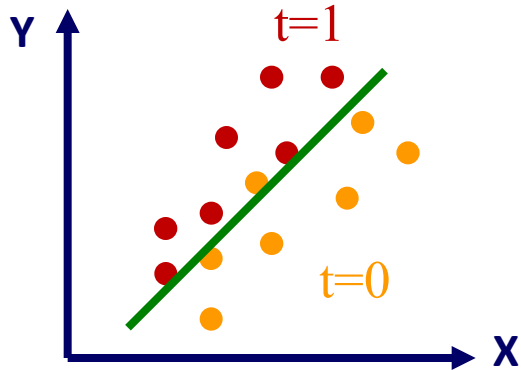
Painel Dividido (Split Panel)

Y_{it}

Combina características de dados em painel com dados em corte transversal, ou seja, uma parte da amostra é observada continuamente e outra é formada por unidades distintas em cada período do tempo.

Y_{11}	Y_{12}	...	Y_{1T}
Y_{21}	Y_{22}	...	Y_{2T}
...	...	...	...
$Y_{n_1 1}$	$Y_{n_2 2}$	...	$Y_{n_T T}$

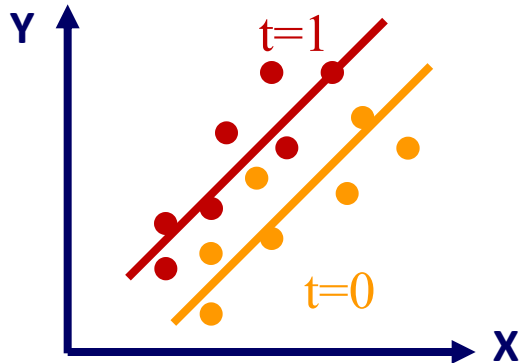
Dados Empilhados – Tipos de Modelos



Intercepto e coeficientes angulares constantes

$$Y = \alpha + \beta X + e$$

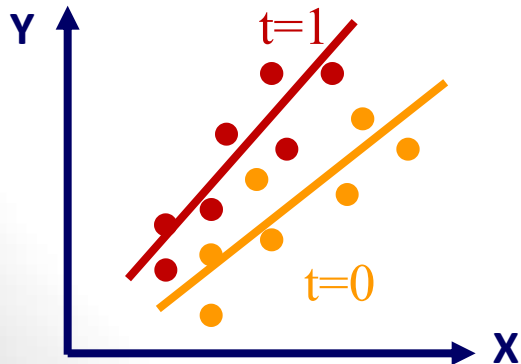
Pressupõe que as relações de causa e efeito sejam as mesmas em qualquer período do tempo.



Intercepto variável e coeficientes angulares constantes

$$Y = \alpha + \beta X + \delta t + e$$

Pressupõe variação de Y no tempo mas relações marginais constantes entre X e Y.



Intercepto e coeficientes angulares variáveis

$$Y = \alpha + \beta X + \delta t + \theta(t \times X) + e$$

Pressupõe quebra estrutural da relação entre X e Y no tempo.

Dados Empilhados - Definição

- A análise de dados empilhados apresenta uma série de vantagens em relação aos dados de corte transversal. Por exemplo, o maior tamanho da amostra e a possibilidade de identificar mudanças estruturais na relação entre a variável dependente e as variáveis independentes;
- Se pressupormos que as relações são as mesmas no tempo, teremos:

$$Y = \beta_0 + \sum_{j=1}^k \beta_j X_j + e_i$$

- Se pressupormos que os valores de Y variam no tempo ($t=0$ ou 1 , por exemplo) para um mesmo valor de X (interceptos variáveis), teremos:

$$Y = \beta_0 + \sum_{j=1}^k \beta_j X_j + \delta t + e_i$$

- Se pressupormos mudanças estruturais da relação entre Y e X no tempo, teremos:

$$Y = \beta_0 + \sum_{j=1}^k \beta_j X_j + \delta t + \sum_{j=1}^k \theta_j X_j \times t + e_i$$

Exercícios

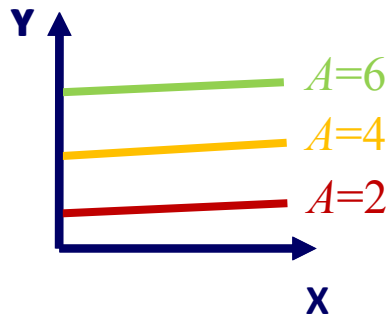
- 1) O arquivo *Data_AgricultureClimate.xls* contém dados da produção agrícola municipal e variáveis climáticas ([GORI MAIA, A., MIYAMOTO, B. C, GARCIA, J. R. Climate change and agriculture: Do environmental preservation and ecosystem services matter? Ecological Economics, v. 152 \(October 2018\), 2018](#)):
 - a) Elabore um modelo de dados empilhados para a relação entre o (log) valor da produção agrícola, (log) área, temperatura e precipitação;
 - b) Controle pelas variações temporais da produção;

Viés de Omissão - Conceito

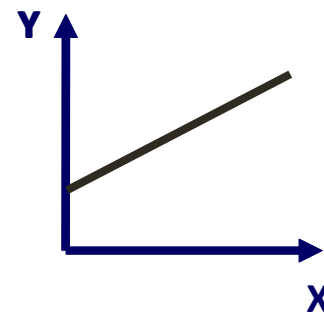
- Suponha que a produção Y dependa do crédito (X) e da área da propriedade A ;
- Entretanto, se não observarmos os dados da área A da propriedade, a relação da produção Y com o crédito (X) pode ser tendenciosa;

$A=2$	$A=2$	$A=4$	$A=4$	$A=6$	$A=6$
$Y=2000$	$Y=2200$	$Y=4000$	$Y=4000$	$Y=6200$	$Y=6000$
$X=2$	$X=4$	$X=6$	$X=8$	$X=10$	$X=12$

$$Y = \alpha + \beta_1 X + \beta_2 A + e$$



$$Y = \alpha + \beta X + e$$



Efeitos não Observados - Conceito

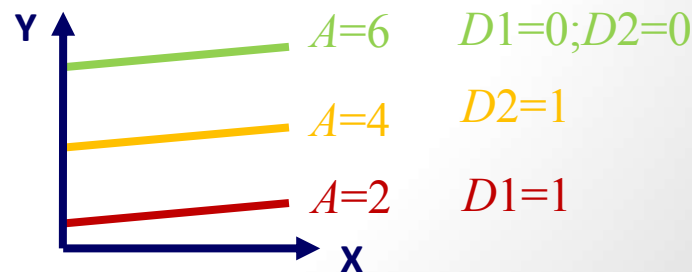
- Suponha agora que cada propriedade ($i=1,2,3$) seja observada em dois períodos distintos ($t=0,1$);
- Se supormos que a área seja diferente entre as propriedades, mas constante no tempo, podemos controlar seu efeito com binárias para discriminar cada propriedade ($D1=1$ para $i=1$, $D2=1$ para $i=2$, propriedade 3 como referência);
- Em outras palavras, mesmo que A não seja observado, podemos controlá-lo por um efeito c , chamado *efeito não observado*.

$D1=1; D2=0$		$D1=0; D2=1$		$D1=0; D2=0$	
$A=2$	$A=2$	$A=4$	$A=4$	$A=6$	$A=6$
$i=1$	$i=1$	$i=2$	$i=2$	$i=3$	$i=3$
$t=0$	$t=1$	$t=0$	$t=1$	$t=0$	$t=1$
$Y=2000$	$Y=2200$	$Y=4000$	$Y=4000$	$Y=6200$	$Y=6000$
$X=2$	$X=4$	$X=6$	$X=8$	$X=10$	$X=12$

$$Y_{it} = \alpha + \beta_1 X_{it} + \beta_2 A_{it} + e_{it}$$

$$Y_{it} = \alpha + \beta_1 X_{it} + c_1 D1_i + c_2 D2_i + e_{it}$$

$$Y_{it} = \alpha + \beta_1 X_{it} + c_i + e_{it}$$



Efeito não Observado

- Vamos assumir que a relação entre y e $\mathbf{x} \equiv (X_1, X_2, \dots, X_k)$ seja dada por:

$$E(y | \mathbf{x}, c) = \mathbf{x}\boldsymbol{\beta} + c$$

Onde c é uma variável aleatória não observada, também chamada de **efeito não observado**. Uma pressuposição em análise de dados em painel é que o efeito de c seja constante no tempo. Assim, teríamos a seguinte especificação:

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + c + e_t$$

Onde $E(e_t | \mathbf{x}_t, c) = 0$

- Caso c não seja correlacionado às variáveis independentes – $\text{Cov}(X_j, c) = 0$ – então c será apenas mais um dos fatores não observáveis a comporem o erro do modelo de regressão. Nesse sentido, podemos aplicar MQO de maneira análoga à amostra de dados empilhados. Entretanto, caso $\text{Cov}(X_j, c) \neq 0$, as estimativas de dados empilhados serão viesadas e inconsistentes.

Efeitos Fixos *One-Way* - Binárias

- Seja o modelo com efeitos não observados dado por:

$$Y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + e_{it}$$

- c_i é chamado de **efeito individual** ou **heterogeneidade individual**. O termo e_{it} é denominado **erro idiossincrático**, já que varia aleatoriamente para todos os indivíduos e períodos.
- Uma solução simples para controlar a heterogeneidade individual é elaborar um **modelo de efeitos fixos com variáveis binárias**. Este pressupõe que c_i represente parâmetros da população a serem estimados (por isso o nome efeito fixo):

$$Y_{it} = \alpha + \sum_{j=1}^k \beta_j X_{jit} + c_2 I_{2i} + \dots + c_n I_{ni} + e_{it}$$

Onde $I_{ji}=1$ se $j=i$, $I_{ji}=0$ se $j \neq i$. Os estimadores de c_j são denominados **estimadores de variáveis binárias**. O intercepto α refere-se ao efeito para o indivíduo utilizado como referência. Na especificação proposta, para $i=1$. Assim, cada coeficiente c_j representará a diferença em relação à unidade de referência ($i=1$).

Efeitos Fixos *One-Way - Within*

- Uma limitação do modelo de efeitos fixos com variáveis binárias é que consome muitos graus de liberdade, já que incorpora muitas variáveis no modelo. As estimativas tendem a ser insignificantes caso a amostra não seja grande o suficiente.
- Alternativamente, pode-se realizar uma transformação algébrica para obter os coeficientes pelo **modelo *within* de efeitos fixos**.

Seja o modelo com heterogeneidade individual:

$$Y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + e_{it}$$

A igualdade será também válida para os valores médios de cada indivíduo:

$$\bar{Y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{e}_i \quad \text{como } c_i \text{ é constante no tempo, sua média será o próprio valor } c_i.$$

Subtraindo as duas equações teremos:

$$\underbrace{(Y_{it} - \bar{Y}_i)}_{\tilde{Y}_{ij}} = \underbrace{(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)}_{\tilde{\mathbf{x}}_{ij}}\boldsymbol{\beta} + (c_i - c_i) + \underbrace{(e_{it} - \bar{e}_i)}_{\tilde{e}_{ij}} \quad \Rightarrow \quad \tilde{Y}_{it} = \tilde{\mathbf{x}}_{it}\boldsymbol{\beta} + \tilde{e}_{it}$$

Efeitos Fixos *Two-Way* – Binárias

- O modelo de efeitos fixos *two-way* considera que o intercepto possa variar entre as unidades de corte transversal (indivíduos) e entre os períodos. Em outras palavras:

$$Y_{it} = \alpha_t + \sum_{j=1}^k \beta_j X_{jit} + c_i + e_{it}$$

- O modelo de efeitos fixos *two-way* pode ser representado por variáveis binárias:

$$Y_{it} = \alpha + \sum_{j=1}^k \beta_j X_{jit} + c_i + ct_2 P_{2_t} + \dots + ct_T P_{T_t} + e_{it}$$

Onde $P_{ji}=1$ se $j=t$, $P_{ji}=0$ se $j \neq t$. O intercepto α refere-se ao efeito para o indivíduo e o período utilizado como referência. Na especificação proposta, para $i=1$ e $t=1$.

Dados em Painel - Vantagens

Algumas das vantagens de trabalhar com dados em painel:

- 1) Discriminação de diferenças individuais e temporais:** Permite identificar efeitos que não seriam detectados isoladamente com dados em corte transversal ou séries temporais: por exemplo, verificar simultaneamente como características da empresa (corte transversal) e mudanças tecnológicas (dados temporais) afetam produtividade de indústrias;
- 2) Maior número de graus de liberdade:** o tamanho da amostra é o produto do número de observações individuais (amostra de dados transversais) pelo tamanho da série temporal;
- 3) Diminui dificuldades inerentes às variáveis omitidas:** por exemplo, a variação na renda média pode ser devida a características específicas não observáveis. Em uma regressão convencional esse efeito estaria diluído na média; em dados em painel pode ser captado exclusivamente pelos coeficientes individuais;

Exercícios

- 1) O arquivo *Data_AgricultureClimate.xls* contém dados da produção agrícola municipal e variáveis climáticas ([GORI MAIA, A., MIYAMOTO, B. C, GARCIA, J. R. Climate change and agriculture: Do environmental preservation and ecosystem services matter? Ecological Economics, v. 152 \(October 2018\), 2018](#)):
 - a) Elabore um modelo para dados em painel com efeitos fixo (one-way) para a relação entre o (log) valor da produção agrícola, (log) área, temperatura e precipitação;
 - b) Elabore um modelo com efeitos fixos two-way;

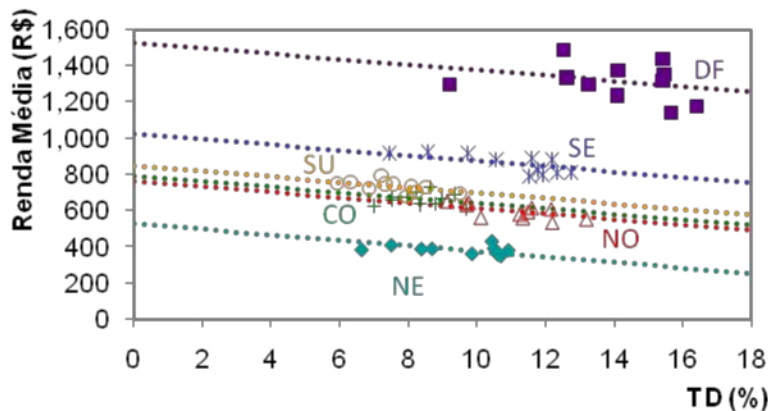
Efeitos Fixos - Resumo

- Seja o modelo de efeitos fixos para a relação entre as rendas e TDs regionais:

$$Y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + e_{it} \quad \Rightarrow \quad Y_{it} = \alpha + \sum_{j=1}^k \beta_j X_{jit} + c_2 I_{2i} + \dots + c_n I_{ni} + e_{it}$$

ou

$$(Y_{it} - \bar{Y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (c_i - c_i) + (e_{it} - \bar{e}_i)$$

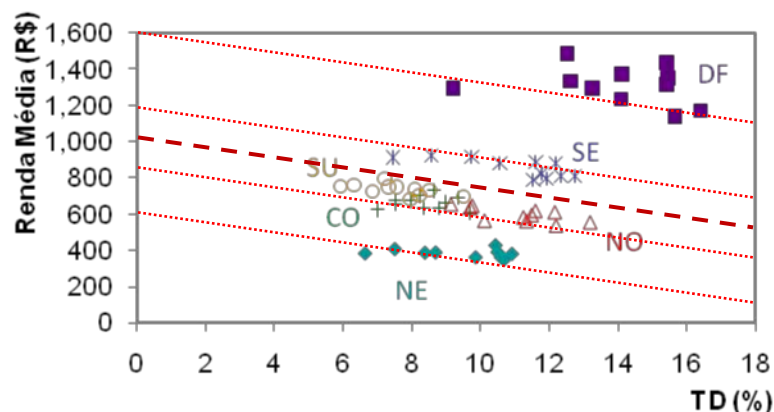


O tratamento da heterogeneidade individual (c_i) como parâmetros do modelo permite identificar variações no intercepto de cada indivíduo i . Tanto o modelo com variáveis binárias quanto o modelo *within* controlam possíveis vieses de omissão de variáveis, que ocorreriam caso a heterogeneidade individual esteja associada aos regressores X .

- Desvantagem:** embora de simples implementação, a inclusão de muitas variáveis binárias ou parâmetros $\bar{Y}_i$ no modelo tende a consumir muitos graus de liberdade e comprometer a significância do efeito isolado das variáveis;

Efeitos Aleatórios - Conceitos

- Outra maneira de analisarmos o problema seria considerar que o intercepto varie aleatoriamente entre os indivíduos:



No caso, o modelo que queremos ajustar é:

$$\text{Renda}_{it} = \alpha + \beta TD_{it} + e_{it}$$

Considerando que o intercepto possa variar aleatoriamente entre as regiões, teríamos:

$$\text{Renda}_{it} = \alpha + c_i + \beta TD_{it} + e_{it}$$

Ou simplesmente:

$$\text{Renda}_{it} = \alpha + \beta TD_{it} + w_{it} \quad \text{Onde: } w_{it} = c_i + e_{it}$$

- O **modelo de correção de erros**, ou **efeitos aleatórios**, considera que o intercepto seja uma variável aleatória e não uma constante. Em outras palavras, as variações regionais seriam identificadas por oscilações aleatórias (c_i) em torno de um valor médio constante (α).
- Para estimar um modelo de efeitos aleatórios pode-se utilizar MQG ou MV.

Efeitos Aleatórios - Definição

O modelo de efeitos aleatórios pressupõe que os efeitos individuais estejam aleatoriamente distribuídos em torno de uma média β_0 constante. Em outras palavras:

$$Y_{it} = \alpha + \sum_{j=1}^k \beta_j X_{jit} + c_i + e_{it}$$

Ou simplesmente:

$$Y_{it} = \alpha + \sum_{j=1}^k \beta_j X_{jit} + w_{it} \text{ onde } w_{it} = c_i + e_{it}$$

O modelo de efeitos aleatórios também é chamado de modelo de correção de erros, justamente por considerar que o erro composto w_{it} possa, na verdade, ser desagregado em dois componentes: i) variação entre indivíduos; ii) variação geral entre observações.

Efeitos Fixos ou Aleatórios?

1) Graus de liberdade: O modelo de efeitos fixos é operacionalmente mais simples mas pode consumir muitos graus de liberdade quando há muitas unidades de corte transversal e/ou temporal. Dessa forma, o modelo de efeitos aleatórios seria especialmente atraente nas situações em que haja uma caracterização substancialmente grande das variações de corte transversal e séries temporais. Seus estimadores seriam mais eficientes (menores variabilidades).

Efeitos Fixos ou Aleatórios?

2) Variabilidade do regressor: Em situações onde os valores de X de uma unidade de corte transversal variam pouco entre os períodos, os modelos de efeitos fixos produzirão estimativas pouco precisas. Em outras palavras, a baixa variabilidade dos fatores explanatórios gerará elevadas variâncias. Nessas situações, o modelo de efeitos aleatórios será mais indicado. Na ausência de variabilidade para um regressor entre períodos de tempo, este será excluído do modelo.

Efeitos Fixos ou Aleatórios?

3) Correlação entre erros e regressor: os erros de um modelo não podem ter relação com as variáveis independentes. Assim, em um modelo de efeitos aleatórios, devemos ter:

$$Cov(X_{it}, c_i) = 0$$

Essa pressuposição nem sempre é válida em modelo de efeitos aleatórios pois implicariam, por exemplo, que o fato de uma região apresentar uma renda superior à das demais (devido a fatores não presentes no modelo) não teria relação com a sua taxa de desemprego (variável independente). Mas sabemos que normalmente regiões mais desenvolvidas apresentam as maiores rendas e também as maiores taxas de desemprego, por exemplo, devido à atração de um expressivo contingente de migrantes. A mesma pressuposição não é necessária no modelo de efeitos fixos pois não são os resíduos que determinam as variações das rendas entre as regiões

Teste de Especificação - Hausman

A idéia do teste de Hausman é comparar as estimativas de efeitos aleatórios com as de efeitos fixos. Diferenças significativas entre elas sugerem a inconsistência dos estimadores de efeitos aleatórios. Em outras palavras, sejam os estimadores dos modelos de efeitos fixos (β_{EF}) e de efeitos aleatórios (β_{EA}):

$$y = X\beta_{EF} + w \quad \text{e} \quad y = X\beta_{EA} + w$$

Na ausência de correlação entre os efeitos (individuais ou temporais) e os regressores X , ambos os estimadores β_{EF} e β_{EA} serão consistentes e β_{EA} será relativamente mais eficiente. Como condição para a hipótese nula, o teste de Hausman assume a consistência dos estimadores de efeitos aleatórios. Em outras palavras:

$$H_0 : \beta_{EA} \text{ são consistentes} \quad \{\text{ausência de correlação entre efeitos e regressores}\}$$

A estatística utilizada no teste, também chamada de estatística m será dada por:

$$m = (\beta_{EF} - \beta_{EA})' (S_{\beta_{EF}} - S_{\beta_{EA}})^{-1} (\beta_{EF} - \beta_{EA})$$

A qual terá uma distribuição χ^2 com k graus de liberdade, onde k é o número de fatores explanatórios de cada modelo.

Rejeitar a hipótese nula significa afirmar que há correlação entre os efeitos e os regressores e, conseqüentemente, os estimadores do modelo de efeitos aleatórios não serão consistentes. Em outras palavras, deve-se trabalhar com efeitos fixos.

Exercícios

- 1) O arquivo *Data_AgricultureClimate.xls* contém dados da produção agrícola municipal e variáveis climáticas ([GORI MAIA, A., MIYAMOTO, B. C, GARCIA, J. R. Climate change and agriculture: Do environmental preservation and ecosystem services matter? Ecological Economics, v. 152 \(October 2018\), 2018](#)):
 - a) Elabore um modelo para dados em painel com efeitos aleatórios para a relação entre o (log) valor da produção agrícola, (log) área, temperatura e precipitação;
 - b) Compare as estimativas obtidas com efeitos fixos e aleatórios;