

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**  
**Microeconometria – 2015/3**

**Microeconometrics: Lecture Notes**

**Autor: Paulo Ferreira Naibert**  
**Professor: Hudson Torrent**

**Porto Alegre**  
**30/06/2020**  
**Revisão: July 15, 2020**

# 1 Regressão MQO Clássico

Wooldridge (2010, C.4 – The Single-Equation Linear Model and OLS Estimation, p.49–76)

## 1.1 Modelo de equações lineares

O modelo populacional que estudamos é linear em seus parâmetros,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u \quad (1.1)$$

onde:

$y, x_1, \dots, x_K$  são escalares aleatórios e observáveis (i.e., conseguimos observá-los em uma amostra aleatória da população);

$u$  é o *random disturbance* não observável, ou erro;

$\beta_0, \beta_1, \dots, \beta_K$  são parâmetros (constantes) que gostaríamos de estimar.

## Notação Vetorial

Wooldridge (2010, Sec. 4.2 – Asymptotic Properties of OLS; p.51)

Por conveniência, escrevemos a equação populacional em forma de vetor:

$$y = \mathbf{x}\boldsymbol{\beta} + u \quad (1.2)$$

onde,

$\mathbf{x} \equiv (x_1, \dots, x_K)$  é um vetor  $1 \times K$  de regressores;

$\boldsymbol{\beta} \equiv (\beta_1, \dots, \beta_K)'$  é um vetor  $K \times 1$ .

Uma vez que a maioria das equações contém um intercepto, assumiremos que  $x_1 \equiv 1$ , visto que essa hipótese deixa a interpretação mais fácil.

## Amostra Aleatória

Assumimos que conseguimos obter uma amostra aleatória de tamanho  $N$  da população para estimarmos  $\boldsymbol{\beta}$ . Dessa forma,  $\{(\mathbf{x}_i, y_i); i = 1, 2, \dots, N\}$  são tratados como variáveis aleatória independentes, identicamente distribuídas, onde  $\mathbf{x}_i$  é  $1 \times K$  e  $y_i$  é escalar. Para cada observação  $i$ , temos:

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i. \quad (1.3)$$

onde  $\mathbf{x}_i$  é um vetor  $1 \times K$  de regressores.

## 1.2 Hipóteses

**OLS.1**  $y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i, \quad i = 1, \dots, N;$

**OLS.2**  $\mathbf{X}$  é não estocástica;

**OLS.3**  $\{u_i\}_{i=1}^N$  é *iid* com e para cada  $i = 1, \dots, N$ :

$$E(u_i) = 0$$

$$\text{Var}(u_i) = E(u_i^2) = \sigma^2$$

**OLS.2'**  $\mathbf{X}$  é estocástica;

**OLS.3'**

$$E(u_i|\mathbf{X}) = 0,$$

$$\text{Var}(u_i|\mathbf{X}) = E\left\{[u_i - E(u_i|\mathbf{X})]^2|\mathbf{X}\right\} = E(u_i^2|\mathbf{X}) = \sigma^2.$$

*Remark.*  $E(u_i|\mathbf{X}) = 0$  implica que  $u_i$  é **não correlacionado** com todos os regressores  $x_k$  para  $k = 1, \dots, K$ . **Exogeneidade estrita.**

### 1.3 Estimação

Usando **OLS.1:**

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i$$

$$\mathbf{x}_i'y_i = \mathbf{x}_i'\mathbf{x}_i\boldsymbol{\beta} + \mathbf{x}_i'u_i$$

$$E(\mathbf{x}_i'y_i) = E(\mathbf{x}_i'\mathbf{x}_i)\boldsymbol{\beta} + E(\mathbf{x}_i'u_i)$$

Usando  $E(\mathbf{x}_i'u_i) = 0$  [Qual seria essa hipótese?]

$$E(\mathbf{x}_i'y_i) = E(\mathbf{x}_i'\mathbf{x}_i)\boldsymbol{\beta}$$

$$\boxed{\boldsymbol{\beta} = [E(\mathbf{x}_i'\mathbf{x}_i)]^{-1}E(\mathbf{x}_i'y_i)} \quad (1.4)$$

Agora, usando o **princípio da analogia** e utilizando **estimadores amostrais**:

$$\boxed{\hat{\boldsymbol{\beta}} = \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'y_i\right)} \quad (1.5)$$

Podemos desenvolver essa equação para:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'(\mathbf{x}_i\boldsymbol{\beta} + \mathbf{u}_i)\right) \\ &= \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i\boldsymbol{\beta}\right) + \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{u}_i\right) \\ \boxed{\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta} + \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{x}_i\right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{x}_i'\mathbf{u}_i\right)} \end{aligned} \quad (1.6)$$

#### 1.3.1 Notação Matricial

Empilhando as  $N$  observações, obtemos a **Notação Matricial**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (1.7)$$

$\mathbf{y}$  é um vetor  $N \times 1$ ;

$\mathbf{X}$  é uma matriz  $N \times K$  de regressores, com  $N$  vetores,  $\mathbf{x}_i$ , de dimensão  $1 \times K$  empilhados;

$\boldsymbol{\beta}$  é um vetor  $K \times 1$ ;

$\mathbf{u}$  é um vetor  $N \times 1$ ;

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}; \quad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{bmatrix}; \quad \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}.$$

As somas de vetores viram simples multiplicações de matrizes e a equação (1.5), vira:

$$\hat{\beta} = (N^{-1}\mathbf{X}'\mathbf{X})^{-1}(N^{-1}\mathbf{X}'\mathbf{y}) \implies \boxed{\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})} \quad (1.8)$$

## 1.4 Valor Esperado

$$\begin{aligned} E(\hat{\beta}) &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u})] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] \\ &= E(\beta) + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] \implies \boxed{E(\hat{\beta}) = \beta + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]} \end{aligned}$$

### 1.4.1 Viés

$$B(\hat{\beta}) = E(\hat{\beta}) - \beta \implies \boxed{B(\hat{\beta}) = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]}$$

*Remark.* Sob **OLS.2'** e **OLS.3'**:

$$E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = E\{E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}]\} = E\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underbrace{E(\mathbf{u}|\mathbf{X})}_{=0}\right\} = 0$$

ou seja,  $B(\hat{\beta}) = 0$ , logo  $\hat{\beta}$  é **não-viciado**. O que também é equivalente a  $E(\hat{\beta}) = \beta$ .

## 1.5 Variância

Supondo **OLS.2'** e **OLS.3'**:

$$\begin{aligned} \text{Var}(\hat{\beta}|\mathbf{X}) &= E\left\{\left[\hat{\beta} - E(\hat{\beta}|\mathbf{X})\right]^2|\mathbf{X}\right\} \\ &= E\left\{\left[\hat{\beta} - E(\hat{\beta}|\mathbf{X})\right]\left[\hat{\beta} - E(\hat{\beta}|\mathbf{X})\right]'\right|\mathbf{X}\right\} \\ &= E\left\{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}][(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}]'\right|\mathbf{X}\right\} \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}] \end{aligned}$$

$$\boxed{\text{Var}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}\mathbf{u}'|\mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}}$$

### 1.5.1 Homocedasticidade

Supondo **homocedasticidade** e ausência de correlação serial:  $E[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \sigma^2\mathbf{I}_N$ . Assim,

$$\text{Var}(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}_N\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \implies \boxed{\text{Var}(\hat{\beta}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}}.$$

## 2 Ausência de Exogeneidade Estrita

Nem sempre poderemos supor **exogeneidade estrita**. Por exemplo, no modelo com variável defasada mostrado abaixo:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{1t} + u_t \\ y_{t-1} &= \beta_0 + \beta_1 y_{t-2} + \beta_2 x_{1t-1} + u_{t-1} \\ y_t &= \beta_0(1 + \beta_1) + \beta_1^2 y_{t-2} + \beta_1 \beta_2 x_{1t-1} + \beta_2 x_{1t} + u_t + \beta_1 u_{t-1}, \end{aligned}$$

o erro é correlacionado com o regressor  $y_{t-1}$ . Nesse caso, tentaremos obter apenas **consistência** e **variância assintótica** do estimador. Para tanto, utilizaremos a equação (1.6):

$$\hat{\beta} = \beta + \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i u_i \right).$$

Aqui começaria a seção 17.

### 2.1 Consistência

Vamos definir a matriz  $K \times K$ ,  $\mathbf{A} \equiv E(\mathbf{x}'_i \mathbf{x}_i)$ . Supondo  $\mathbf{A}$ , finita e positiva definida, posto( $\mathbf{A}$ ) =  $K$ . Usando **LGN matricial** (Definição 17.2 na página 44), temos:

[lembrar que as dimensões dos vetores estão invertidas:  $1 \times K$  e **não**  $K \times 1$ ]

$$N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \xrightarrow{p} \mathbf{A} \implies \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \xrightarrow{p} \mathbf{A}^{-1}. \quad (2.1)$$

Além disso, vamos supor  $E(\mathbf{x}'_i u_i) = 0$ , o que corresponde a  $\text{Cov}(\mathbf{x}_i, u_i) = 0$ , ou seja, o erro  $u_i$  **não** é correlacionado com os regressores da própria equação. Isso é bem menos que exogeneidade estrita. Então,

$$N^{-1} \sum_{i=1}^N \mathbf{x}'_i u_i \xrightarrow{p} E(\mathbf{x}'_i u_i) = \mathbf{0}_K.$$

Logo,

$$\boxed{\hat{\beta} = \beta + \underbrace{\left( \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{x}'_i u_i \right)}_{\xrightarrow{p} 0}}$$

Então,  $(\hat{\beta} - \beta) \xrightarrow{p} 0$  que é equivalente a  $\hat{\beta} \xrightarrow{p} \beta$  e plim  $\hat{\beta} = \beta$ , ou seja,  $\hat{\beta}$  é **consistente** para  $\beta$ .

## 2.2 Normalidade Assintótica

$$\begin{aligned}\hat{\beta} &= \beta + \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i u_i \right) \\ (\hat{\beta} - \beta) &= \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i u_i \right) \\ \sqrt{N}(\hat{\beta} - \beta) &= \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right)\end{aligned}$$

Supondo  $E(x_{ik}^2 u_i^2) < +\infty$ ,  $k = 1, \dots, K$ , e definindo  $\mathbf{B} = E[\mathbf{x}'_i u_i u_i \mathbf{x}_i] = E[u_i^2 \mathbf{x}'_i \mathbf{x}_i]$ . Temos, pela Definição 17.7 (TCL), que

$$N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \xrightarrow{d} N(\mathbf{0}, \mathbf{B}) \implies N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i = O_p(1) \quad (2.2)$$

Além disso, vamos utilizar a matriz **simétrica** e **não singular**  $\mathbf{A}$  da equação (2.1). Assim, temos

$$\begin{aligned}\sqrt{N}(\hat{\beta} - \beta) &= \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right) \\ &= \left[ \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} + \mathbf{A}^{-1} - \mathbf{A}^{-1} \right] \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right) \\ &= \left[ \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} - \mathbf{A}^{-1} \right] \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right) + \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right),\end{aligned}$$

Podemos inverter  $\mathbf{A}$  porque ela tem posto completo (não singular). Pelas propriedades de  $\mathbf{A}$ , temos:

$$N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \xrightarrow{p} \mathbf{A} \implies \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} - \mathbf{A}^{-1} = o_p(1).$$

Então,

$$\sqrt{N}(\hat{\beta} - \beta) = o_p(1)O_p(1) + \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right),$$

Usando (2.2) e o Lema 17.2.

$$\mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \right) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}).$$

Lembrando que  $o_p(1)O_p(1) = o_p(1)$ , temos:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}) \implies \sqrt{N}(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1})$$

## 2.3 Variância

$$\mathbf{V} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$$

$$\mathbf{V} = \mathbb{E}[(\mathbf{x}'_i \mathbf{x}_i)]^{-1} \mathbb{E}[(\mathbf{x}'_i u'_i u_i \mathbf{x}_i)] \mathbb{E}[(\mathbf{x}'_i \mathbf{x}_i)]^{-1}$$

$$\mathbf{V} = \mathbb{E}[(\mathbf{x}'_i \mathbf{x}_i)]^{-1} \mathbb{E}[u_i^2 \mathbf{x}'_i \mathbf{x}_i] \mathbb{E}[(\mathbf{x}'_i \mathbf{x}_i)]^{-1}.$$

### 2.3.1 Homocedasticidade

Sob **Homocedasticidade**, temos  $\mathbf{B} = \mathbb{E}(u_i^2 \mathbf{x}'_i \mathbf{x}_i) = \sigma^2 \mathbb{E}(\mathbf{x}'_i \mathbf{x}_i)$ , logo

$$\boxed{\mathbf{V} = \sigma^2 \mathbb{E}[(\mathbf{x}'_i \mathbf{x}_i)]^{-1}}.$$

### 2.3.2 Estimador Amostral

$$\begin{aligned} \hat{\mathbf{V}} &= \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N u_i^2 \mathbf{x}'_i \mathbf{x}_i \right) \left( N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \\ &= N \left( \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( \sum_{i=1}^N u_i^2 \mathbf{x}'_i \mathbf{x}_i \right) \left( \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \end{aligned}$$

$$\boxed{\hat{\mathbf{V}} = N (\mathbf{X}' \mathbf{X})^{-1} \left( \sum_{i=1}^N u_i^2 \mathbf{x}'_i \mathbf{x}_i \right) (\mathbf{X}' \mathbf{X})^{-1}}.$$

### 2.3.3 Variância do estimador de OLS

$$\text{Var}(\sqrt{N} \hat{\boldsymbol{\beta}}) = \mathbf{V}$$

$$\text{Var}(\hat{\boldsymbol{\beta}}) = N^{-1} \mathbf{V}$$

$$\boxed{\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}' \mathbf{X})^{-1} \left( \sum_{i=1}^N u_i^2 \mathbf{x}'_i \mathbf{x}_i \right) (\mathbf{X}' \mathbf{X})^{-1}}.$$

A variância **Robusta** é:

$$\boxed{\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}' \mathbf{X})^{-1} \left( \sum_{i=1}^N \hat{u}_i^2 \mathbf{x}'_i \mathbf{x}_i \right) (\mathbf{X}' \mathbf{X})^{-1}}.$$

A variância sob **Homocedasticidade** é:

$$\boxed{\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 (\mathbf{X}' \mathbf{X})^{-1}}.$$

### 3 System OLS (SOLS)

Wooldridge (2010, C.7 – Estimating Systems of Equations by OLS and GLS, p.143–179)

Wooldridge (2010, Sec.7.3 – System OLS Estimation of a Multivariate Linear System, p.147)

#### 3.1 Modelo Linear

Assumimos que temos as seguintes observações *cross section iid*:  $\{(\mathbf{X}_i, \mathbf{y}_i) : i = 1, \dots, N\}$ , onde:

$\mathbf{X}_i$  é uma matriz  $G \times K$  e contém as variáveis explicativas que aparecem em qualquer lugar do sistema.

$\mathbf{y}_i$  é um vetor  $G \times 1$ , que contém as variáveis dependentes para todas as equações  $G$  (ou períodos de tempo, no caso de dados de painel).

O modelo linear multivariado para uma observação (draw) aleatória da população pode ser expresso como:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i, \quad i = 1, \dots, N, \quad (3.1)$$

onde:

$\boldsymbol{\beta}$  é um vetor  $K \times 1$  de parâmetros de interesse; e

$\mathbf{u}_i$  é um vetor  $G \times 1$  de não observáveis.

A equação (3.1) explica as  $G$  variáveis  $y_{i1}, \dots, y_{iG}$  em termos de  $\mathbf{X}_i$  e das não observáveis  $\mathbf{u}_i$ . Por causa da hipótese de amostra aleatória podemos escrever tudo em termos de uma observação genérica.

#### 3.2 Hipóteses

Wooldridge (2010, Sec.7.3.1)

**SOLS.1**  $E(\mathbf{X}_i' \mathbf{u}_i) = \mathbf{0}_{K \times 1}$ .

**SOLS.2**  $\mathbf{A} \equiv E(\mathbf{X}_i' \mathbf{X}_i)$  é não singular (tem posto pleno, posto igual a  $K$ ).

#### 3.3 Estimação

Note que, sob **SOLS.1**, temos:

$$E[\mathbf{X}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] = \mathbf{0}$$

$$E(\mathbf{X}_i' \mathbf{X}_i) \boldsymbol{\beta} = E(\mathbf{X}_i' \mathbf{y}_i)$$

$$\boldsymbol{\beta} = [E(\mathbf{X}_i' \mathbf{X}_i)]^{-1} E(\mathbf{X}_i' \mathbf{y}_i) \quad (3.2)$$

Usando estimadores amostrais:

$$\hat{\boldsymbol{\beta}}^{SOLS} = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{y}_i \right). \quad (3.3)$$

Para computar  $\hat{\boldsymbol{\beta}}$  usando linguagem de computação é mais fácil utilizar a notação matricial. Para tanto, cortamos os  $N^{-1}$  e substituímos os somatórios por multiplicações de matrizes.

$$\hat{\boldsymbol{\beta}}^{SOLS} = (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \mathbf{y}) \quad (3.4)$$

onde

$\mathbf{X} \equiv (\mathbf{X}_1', \dots, \mathbf{X}_N')$  é uma matriz  $NG \times K$  dos  $\mathbf{X}_i$  empilhados.

$\mathbf{y} \equiv (\mathbf{y}_1', \dots, \mathbf{y}_N')$  é um vetor  $NG \times 1$  das observações  $\mathbf{y}_i$  empilhadas.



### 3.4 Consistência

Para provarmos a **consistência** do estimador, usamos as equações (3.3) e (3.1):

$$\begin{aligned}\hat{\beta}^{SOLS} &= \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{y}_i \right) \\ &= \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left[ N^{-1} \sum_{i=1}^N \mathbf{X}_i' (\mathbf{X}_i \beta + \mathbf{u}_i) \right] \\ &= \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \beta \right) + \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right).\end{aligned}$$

E chegamos em:

$$\boxed{\hat{\beta}^{SOLS} = \beta + \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right).} \quad (3.5)$$

Por **SOLS.1**:

$$N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \xrightarrow{p} \mathbf{0};$$

e por **SOLS.2**

$$\left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \xrightarrow{p} \mathbf{A}^{-1}.$$

Resumimos esse resultado pelo seguinte Teorema:

**Theorem 3.1** (Consistência do SOLS). Sob Hipóteses **SOLS.1** e **SOLS.2**, temos

$$\boxed{\hat{\beta}^{SOLS} \xrightarrow{p} \beta}.$$

### 3.5 Normalidade Assintótica

De (3.5):

$$\begin{aligned}\hat{\beta} &= \beta + \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right) \\ (\hat{\beta} - \beta) &= \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right).\end{aligned}$$

E chegamos em:

$$\boxed{\sqrt{N}(\hat{\beta} - \beta) = \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right).} \quad (3.6)$$

Uma vez que  $E(\mathbf{X}_i' \mathbf{u}_i) = 0$ , sob a hipótese **SOLS.1**, a definição 17.7 (**TCL**) implica que:

$$N^{-1/2} \sum_{i=1}^N \mathbf{X}_i \mathbf{u}_i \xrightarrow{d} N(\mathbf{0}, \mathbf{B}),$$

onde

$$\mathbf{B} \equiv E(\mathbf{X}_i' \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i) \equiv \text{Var}(\mathbf{X}_i \mathbf{u}_i).$$

Em particular,

$$N^{-1/2} \sum_{i=1}^N \mathbf{X}_i \mathbf{u}_i = O_p(1).$$

Porém,

$$\left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} = (\mathbf{X}' \mathbf{X} / N)^{-1} = \mathbf{A}^{-1} + o_p(1).$$

Sendo assim,

$$\begin{aligned} \sqrt{N}(\hat{\beta} - \beta) &= \left[ \mathbf{A}^{-1} + \left( N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} - \mathbf{A}^{-1} \right] \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right) \\ &= \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right) + [(\mathbf{X}' \mathbf{X} / N)^{-1} - \mathbf{A}^{-1}] \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right) \\ &= \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right) + o_p(1) O_p(1) \\ &= \mathbf{A}^{-1} \left( N^{-1/2} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \right) + o_p(1) \end{aligned}$$

$$\boxed{\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1})} \quad (3.7)$$

### 3.6 Variância Assintótica

**SOLS.3: Homocedasticidade**  $E(\mathbf{X}_i' \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i) = \sigma^2 E(\mathbf{X}_i' \mathbf{X}_i)$ .

De (3.7), vamos definir  $\mathbf{V} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$ . Sob **SOLS.3**,  $\mathbf{V} = \sigma^2 [E(\mathbf{X}_i' \mathbf{X}_i)]^{-1}$ . Estimando:

$$\hat{\sigma}^2 = \frac{1}{NG - K} \sum_{i=1}^N \sum_{g=1}^G \hat{u}_{ig}^2$$

onde  $\hat{u}_{ig} = y_{ig} - \mathbf{x}_{ig} \hat{\beta}^{SOLS}$

#### 3.6.1 A Matriz Robusta

$$\hat{\mathbf{V}} = (\mathbf{X}' \mathbf{X})^{-1} \left( \sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{X}_i \right) (\mathbf{X}' \mathbf{X})^{-1}$$

$$\sum_{i=1}^N \mathbf{X}_i' \hat{\Omega} \mathbf{X}_i \xrightarrow{p} E(\mathbf{X}_i \Omega \mathbf{X}_i)$$

Mas **não** é verdade que  $\hat{\Omega} \xrightarrow{p} \Omega$ .

- Havendo constante, **SOLS.1**  $\implies E(\mathbf{u}_i) = 0$
- Ausência de correlação entre os regressores de uma equação e o erro da própria equação  $\implies$  **SOLS.1**.

### 3.6.2 Variância Assintótica

REVER

$$\text{Avar}(\hat{\beta}^{SOLS}) = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1} / N. \quad (3.8)$$

Assim,  $\text{Avar}(\hat{\beta}^{SOLS})$  tende a zero a uma taxa  $1/N$ , como esperado. Estimação consistente de  $\mathbf{A}$  é:

$$\hat{\mathbf{A}} \equiv \mathbf{X}'\mathbf{X}/N = N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i$$

Um estimador consistente para  $\mathbf{B}$  pode ser achado usando o princípio da analogia.

$$\mathbf{B} = E(\mathbf{X}_i' \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i), \quad N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i \xrightarrow{p} \mathbf{B}.$$

Uma vez que não podemos observar  $\mathbf{u}_i$ , usamos os resíduos da estimação de SOLS:

$$\hat{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{X}_i \hat{\beta} = \mathbf{u}_i - \mathbf{X}_i (\hat{\beta} - \beta).$$

Assim, definimos  $\hat{\mathbf{B}}$  e usando LGN, podemos mostrar que:

$$\hat{\mathbf{B}} \equiv N^{-1} \sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{X}_i \xrightarrow{p} \mathbf{B}.$$

onde supomos que certos momentos envolvendo  $\mathbf{X}_i$  e  $\mathbf{u}_i$  são finitos.

Portanto,  $\text{Avar}[\sqrt{N}(\hat{\beta} - \beta)]$  é **consistentemente** estimado por  $\hat{\mathbf{A}}^{-1} \hat{\mathbf{B}} \hat{\mathbf{A}}^{-1}$ , e  $\text{Avar}(\hat{\beta})$  é estimado como:

$$\hat{\mathbf{V}} \equiv \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{X}_i \right) \left( \sum_{i=1}^N \mathbf{X}_i' \mathbf{X}_i \right)^{-1}.$$

Sob as hipóteses **SOLS.1** e **SOLS.2**, nós fazemos inferência em  $\beta$  como  $\hat{\beta}$  fosse normalmente distribuído com média  $\beta$  e variância  $\hat{\mathbf{V}}$ .

## 4 Dados de Painei (POLS)

Wooldridge (2010, C.7 – Estimating Systems of Equations by OLS and GLS. p.143-179)

Wooldridge (2010, Sec.7.8 – The Linear Panel Data Model, Revisited. p.169)

### 4.1 Modelo Linear para Dados de Painei

No caso de dados de painei, temos a seguinte amostra aleatória:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T. \quad (4.1)$$

onde

$y_{it}$  é um escalar.

$\boldsymbol{\beta}$  é um vetor  $K \times 1$ .

$\mathbf{x}_{it}$  é um vetor  $1 \times K$ .

$u_{it}$  é um escalar.

$$\mathbf{x}_{it} = [x_{1,it} \quad \dots \quad x_{K,it}] \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$$

**Notação Vetorial:**  $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i$ , pra cada  $i = 1, \dots, N$ .

onde

$\mathbf{y}_i$  é um vetor  $T \times 1$ .

$\boldsymbol{\beta}$  é um vetor  $K \times 1$ .

$\mathbf{X}_i$  é uma matriz  $T \times K$ .

$\mathbf{u}_i$  é um vetor  $T \times 1$ .

$$\mathbf{X}_i = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_T \end{bmatrix} = \begin{bmatrix} x_{1,i1} & \dots & x_{K,i1} \\ \vdots & \ddots & \vdots \\ x_{1,iT} & \dots & x_{K,iT} \end{bmatrix} \quad \mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}$$

**Notação Matricial:**  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$

onde

$\mathbf{y}$  é um vetor  $NT \times 1$ .

$\boldsymbol{\beta}$  é um vetor  $K \times 1$ .

$\mathbf{X}$  é uma matriz  $NT \times K$ .

$\mathbf{u}$  é um vetor  $NT \times 1$ .

$$\ddot{\mathbf{X}}_{NT \times K} = \begin{bmatrix} \ddot{\mathbf{X}}_1 \\ \vdots \\ \ddot{\mathbf{X}}_N \end{bmatrix} = \begin{bmatrix} \ddot{x}_{1,11} & \dots & \ddot{x}_{K,11} \\ \vdots & & \vdots \\ \ddot{x}_{1,1T} & \dots & \ddot{x}_{K,1T} \\ \ddot{x}_{1,21} & \dots & \ddot{x}_{K,21} \\ \vdots & & \vdots \\ \ddot{x}_{1,2T} & \dots & \ddot{x}_{K,2T} \\ \vdots & & \vdots \\ \ddot{x}_{1,N1} & \dots & \ddot{x}_{K,N1} \\ \vdots & & \vdots \\ \ddot{x}_{1,NT} & \dots & \ddot{x}_{K,NT} \end{bmatrix} \quad \ddot{\mathbf{y}}_{NT \times 1} = \begin{bmatrix} \ddot{y}_1 \\ \vdots \\ \ddot{y}_N \end{bmatrix} = \begin{bmatrix} \ddot{y}_{11} \\ \vdots \\ \ddot{y}_{1T} \\ \ddot{y}_{21} \\ \vdots \\ \ddot{y}_{2T} \\ \vdots \\ \ddot{y}_{N1} \\ \vdots \\ \ddot{y}_{NT} \end{bmatrix} \quad \ddot{\mathbf{u}}_{NT \times 1} = \begin{bmatrix} \ddot{u}_1 \\ \vdots \\ \ddot{u}_N \end{bmatrix} = \begin{bmatrix} \ddot{u}_{11} \\ \vdots \\ \ddot{u}_{1T} \\ \ddot{u}_{21} \\ \vdots \\ \ddot{u}_{2T} \\ \vdots \\ \ddot{u}_{N1} \\ \vdots \\ \ddot{u}_{NT} \end{bmatrix}$$

## 4.2 Estimação

$$\mathbf{X}'\mathbf{X} = \sum_{i=1}^N \mathbf{X}'_i \mathbf{X}_i = \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}'_{it} \mathbf{x}_{it}; \quad \mathbf{X}'\mathbf{y} = \sum_{i=1}^N \mathbf{X}'_i \mathbf{y}_i = \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}'_{it} y_{it}.$$

Portanto, podemos escrever  $\hat{\beta}$  como:

$$\hat{\beta}^{POLS} = \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}'_{it} \mathbf{x}_{it} \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}'_{it} y_{it} \right). \quad (4.2)$$

Este estimador é chamado **estimador de Mínimos Quadrados Agrupados (POLS)** porque ele corresponde a rodar uma regressão OLS nas observações agrupadas através de  $i$  e  $t$ . O estimador da equação (4.2) é o mesmo para unidades de *cross section* amostradas em diferentes pontos do tempo.

## 4.3 Hipóteses

**POLS.1**  $E(\mathbf{X}'_i \mathbf{u}_i) = E(\mathbf{x}'_{it} u_{it}) = \mathbf{0}_{K \times 1}$ , para cada  $i = 1, \dots, N$  e  $t = 1, \dots, T$ .  
De fato, **POLS.1**  $\implies$  **SOLS.1**.

**Obs:** O modelo (4.1) permite  $y_{i,t-1}$  como regressor, se satisfeita **POLS.1**.

## 5 Alguns Testes

Lembrando a equação do modelo (4.1):

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}$$

### 5.1 Autocorrelação dos Resíduos

Nos dois testes apresentado, primeiro precisamos guardar os resíduos estimado. Para tanto, rodamos a regressão do modelo (4.1) e guardamos os resíduos:

$$\hat{u}_{it} = y_{it} - \mathbf{x}_{it}\hat{\boldsymbol{\beta}}^{POLS}. \quad (5.1)$$

**Com Exogeneidade Estrita** Sob exogeneidade estrita, rodamos a seguinte regressão dos resíduos:

$$\hat{u}_{it} = \delta_0 + \delta\hat{u}_{it-1} + \varepsilon_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T.$$

Então, testamos

$$H_0 : \delta_1 = 0$$

$$H_1 : \delta_1 \neq 0$$

via teste  $t$  (pode ser robusto).

**Sem Exogeneidade Estrita (Apenas exogeneidade contemporânea)** Sem exogeneidade estrita, rodamos a seguinte regressão dos resíduos:

$$\hat{u}_{it} = \mathbf{x}_{it}\boldsymbol{\alpha} + \delta\hat{u}_{it-1} + \varepsilon_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T.$$

Então, testamos

$$H_0 : \delta_1 = 0$$

$$H_1 : \delta_1 \neq 0$$

via teste  $t$ .

### 5.2 Heterocedasticidade

Com os resíduos da equação (5.1), rodamos a seguinte regressão:

$$\hat{u}_{it}^2 = \gamma_0 + \gamma_1\hat{y}'_{it} + \gamma_2\hat{y}_{it}^2 + \varepsilon_{it}$$

onde  $\hat{y}_{it} = \mathbf{x}_{it}\hat{\boldsymbol{\beta}}^{POLS}$ . Definindo  $\mathbf{h}_{it} = (\hat{y}'_{it}, \hat{y}_{it}^2)$ , podemos reescrever a equação acima como

$$\hat{u}_{it}^2 = \gamma_0 + \mathbf{h}_{it}\boldsymbol{\gamma} + \varepsilon_{it}$$

Então, testamos

$$H_0 : \boldsymbol{\gamma} = \mathbf{0} \quad (\gamma_1 = 0 \text{ e } \gamma_2 = 0)$$

$$H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$$

via teste de Wald.

### 5.3 Teste de Wald

Se é verdade que

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \hat{\mathbf{V}}).$$

Seja  $\mathbf{R}$  uma matriz  $Q \times K$  com  $Q \leq K$  e  $\text{posto}(\mathbf{R}) = Q$  (posto pleno), então

$$\sqrt{N}\mathbf{R}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{R}\hat{\mathbf{V}}\mathbf{R}').$$

e

$$\left[ \sqrt{N}\mathbf{R}(\hat{\beta} - \beta) \right]' (\mathbf{R}\hat{\mathbf{V}}\mathbf{R}')^{-1} \left[ \sqrt{N}\mathbf{R}(\hat{\beta} - \beta) \right] \stackrel{a}{\sim} \chi_Q^2$$

O resultado acima vale para  $\hat{\mathbf{V}}$  no lugar de  $\mathbf{V}$ , desde que  $\hat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$ . Ou seja, vale para estimadores **consistentes** de  $\mathbf{V}$ .

#### Teste de Wald

$$H_0 : \mathbf{R}\beta = \mathbf{r}$$

$$H_1 : \mathbf{R}\beta \neq \mathbf{r}$$

A estatística do teste acima é:

$$N \left[ \mathbf{R}\hat{\beta} - \mathbf{r} \right]' \left( \mathbf{R}\hat{\mathbf{V}}\mathbf{R}' \right)^{-1} \left[ \mathbf{R}\hat{\beta} - \mathbf{r} \right] \stackrel{a}{\sim} \chi_Q^2$$

#### Remarks

1.  $\hat{\mathbf{V}}$  pode ser a matriz robusta.
2. Uma aproximação, via distribuição  $F$  é dado por:

$$\frac{\text{Est. Teste}}{Q} \stackrel{a}{\sim} F_{Q, N-K}$$

$$\text{com } \text{Avar}(\hat{\beta}) = \frac{N}{N-K} \hat{\mathbf{V}}.$$

## 6 Modelo de Efeitos Não Observados (UEM)

Wooldridge (2010, C.10 – Basic Linear Unobserved Effects Panel Data Models, p.247–291)

Wooldridge (2010, Sec.10.1 – Motivation: The Omitted Variables Problem, p.247)

Wooldridge (2010, Sec.10.2 – Assumptions about the Unobserved Effects and Explanatory Variables, p.251)

Wooldridge (2010, Sec.10.3 – Estimating UEM by POLS, p.256)

### 6.1 Modelo UEM

O modelo básico de efeitos não observados (UEM) pode ser escrito para uma amostra *cross-section* aleatória  $i$  como:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad t = 1, \dots, T. \quad (6.1)$$

onde  $c_i$  é o efeito não observado (componente não observado, variável latente, heterogeneidade não observada, efeito individual, heterogeneidade individual). Estamos supondo  $c_i$  **não** observável.

Definindo os erros compostos  $v_{it} = c_i + u_{it}$ , temos:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \quad (6.2)$$

Ou, em forma de vetor:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i \quad (6.3)$$

### 6.2 Estimação e Consistência

Wooldridge (2010, Sec.10.3 – Estimating UEM by POLS, p.256).

Se usarmos o estimador POLS na equação (6.1), o estimador será consistente se:

$$E(\mathbf{x}'_{it}v_{it}) = \mathbf{0}, \quad t = 1, \dots, T.$$

Ou seja, precisamos que:

$$E(\mathbf{x}'_{it}u_{it}) = \mathbf{0}, \quad t = 1, \dots, T.$$

$$E(\mathbf{x}'_{it}c_i) = \mathbf{0}, \quad t = 1, \dots, T.$$

**Caso 1:**  $E(\mathbf{x}'_{it}c_i) = \mathbf{0}$ .

**POLS** é consistente, mas não é eficiente.

**Efeitos Aleatórios** é consistente e eficiente.

EA é o **FGLS** do modelo.

**Caso 2:**  $E(\mathbf{x}'_{it}c_i) \neq \mathbf{0}$ .

Se POLS é **inconsistente**.

Nesse caso, usaremos o Modelo de **Efeitos Fixos** ou **Primeira Diferença**.

EF e PD é o POLS nos modelos transformados.

**Obs:** Modelos com variáveis dependentes defasadas em  $\mathbf{x}_{it}$  *devem* violar a hipótese  $E(\mathbf{x}'_{it}u_{it}) = \mathbf{0}$  uma vez que  $y_{i,t-1}$  e  $c_i$  devem ser correlacionados. Considerando  $y_{i,t-1}$  como regressor:

$$\left. \begin{aligned} y_{it} &= \alpha y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \\ y_{it-1} &= \alpha y_{i,t-2} + \mathbf{x}_{i,t-1}\boldsymbol{\beta} + v_{i,t-1} \end{aligned} \right\} \text{Cov}(y_{i,t-1}, v_{it}) \neq 0.$$

Mesmo se  $E(\mathbf{x}'_{it}u_{it}) = \mathbf{0}$  é verdadeiro, os erros compostos serão serialmente correlacionados devido a presença de  $c_i$  em cada período de tempo. Portanto, a inferência do POLS requer um estimador robusto de matriz de covariância e estatísticas robustas de teste.



## 7 Modelo de Efeitos Fixos (EF), (Fixed Effects FE)

Wooldridge (2010, Sec.10.5 – Fixed Effects Methods, p.265)

### 7.1 Modelo

O modelo linear de **efeitos individuais não observados (UEM)**:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (7.1)$$

Estamos supondo  $c_i$  **não** observável. Definindo  $v_{it} = c_i + u_{it}$ .

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (7.2)$$

No modelo FE permitimos  $\text{Cov}(\mathbf{x}_{it}, c_i) \neq 0$ .

### 7.2 Transformação *Within*

Tirando a média do modelo ao longo de  $t = 1, \dots, T$ :

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + \bar{c}_i + \bar{u}_i, \quad i = 1, \dots, N. \quad (7.3)$$

onde:

$$\bar{y}_i = T^{-1} \sum_{t=1}^T y_{it}; \quad \bar{\mathbf{x}}_i = T^{-1} \sum_{t=1}^T \mathbf{x}_{it}; \quad \bar{c}_i = T^{-1} \sum_{t=1}^T c_i = c_i; \quad \bar{u}_i = T^{-1} \sum_{t=1}^T u_{it}.$$

Então, subtraindo (7.3) de (7.1):

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + \underbrace{c_i - \bar{c}_i}_{=0} + u_{it} - \bar{u}_i, \quad i = 1, \dots, N; \quad t = 1, \dots, T.$$

Finalmente, obtemos:

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}, \quad i = 1, \dots, N; \quad t = 1, \dots, T. \quad (7.4)$$

Onde  $\ddot{y}_{it} \equiv y_{it} - \bar{y}_i$ ,  $\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_i$  e  $\ddot{u}_{it} \equiv u_{it} - \bar{u}_i$ . E eliminamos variáveis que não variam ao longo do tempo.

### 7.3 Notação Vetorial e Matriz Centralizadora (*Centering Matrix*) $\mathbf{M}^0$

Utilizando notação vetorial para o modelo linear de **efeitos individuais não observados (UEM)**:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{1} + \mathbf{u}_i, \quad i = 1, \dots, N. \quad (7.5)$$

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_{it}, \quad i = 1, \dots, N. \quad (7.6)$$

Agora, definimos a matriz  $\mathbf{M}^0$  (Wooldridge (2010, p. 268) usa a notação  $\mathbf{Q}_T$  para essa matriz) como:

$$\mathbf{M}^0 = \mathbf{I}_T - \mathbf{1}_T(\mathbf{1}'_T\mathbf{1}_T)^{-1}\mathbf{1}'_T = \mathbf{I}_T - T^{-1}\mathbf{1}_T\mathbf{1}'_T$$

A matriz  $\mathbf{M}^0$  tem dimensão  $T \times T$ . Além disso, ela é idempotente ( $\mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^0$ ) e simétrica ( $\mathbf{M}^{0'} = \mathbf{M}^0$ ).

Podemos transformar o modelo (7.5) ao premultiplicarmos todo o modelo por  $\mathbf{M}^0$ :

$$\begin{aligned}\mathbf{M}^0 \mathbf{y}_i &= (\mathbf{I}_T - T^{-1} \mathbf{1}_T \mathbf{1}_T') \mathbf{y}_i = \mathbf{y}_i - \mathbf{1}_T \bar{y}_i = \ddot{\mathbf{y}}_i \\ \mathbf{M}^0 \mathbf{X}_i &= \mathbf{X}_i - T^{-1} \mathbf{1}_T \mathbf{1}_T' \bar{\mathbf{X}}_i = \mathbf{X}_i - \mathbf{1}_T \bar{\mathbf{x}}_i = \ddot{\mathbf{X}}_i \\ \mathbf{M}^0 \mathbf{u}_i &= (\mathbf{I}_T - T^{-1} \mathbf{1}_T \mathbf{1}_T') \mathbf{u}_i = \mathbf{u}_i - \mathbf{1}_T \bar{u}_i = \ddot{\mathbf{u}}_i \\ \mathbf{M}^0 (c_1 \mathbf{1}_T) &= (\mathbf{I}_T - T^{-1} \mathbf{1}_T \mathbf{1}_T') c_1 \mathbf{1}_T = c_i (\mathbf{I}_T - T^{-1} \mathbf{1}_T \mathbf{1}_T') \mathbf{1}_T = c_i (\mathbf{1}_T - \mathbf{1}_T) = \mathbf{0}_T\end{aligned}$$

onde  $\bar{\mathbf{x}}_i$  é o vetor  $1 \times K$  com a média dos  $K$  regressores.

$$\begin{aligned}\mathbf{M}^0 \mathbf{y}_i &= \mathbf{M}^0 \mathbf{X}_i \boldsymbol{\beta} + \mathbf{M}^0 (c_1 \mathbf{1}_T) + \mathbf{M}^0 \mathbf{u}_i, \quad i = 1, \dots, N. \\ \ddot{\mathbf{y}}_i &= \ddot{\mathbf{X}}_i \boldsymbol{\beta} + \ddot{\mathbf{u}}_i, \quad i = 1, \dots, N.\end{aligned}\tag{7.7}$$

**Exemplo:** (Wooldridge, 2010, p.266) Considere o modelo:

$$y_{it} = \beta_0 + \beta_2 d2_t + \dots + \beta_T dT_t + \mathbf{z}_i \boldsymbol{\delta} + d2_t \mathbf{z}_i \boldsymbol{\delta}_2 + \dots + dT_t \mathbf{z}_i \boldsymbol{\delta}_T + \mathbf{x}_{it} \boldsymbol{\alpha} + v_{it}$$

Após a transformação:

$$\ddot{y}_{it} = \beta_2 \ddot{d2}_t + \dots + \beta_T \ddot{dT}_t + \ddot{d2}_t \mathbf{z}_i \boldsymbol{\delta}_2 + \dots + \ddot{dT}_t \mathbf{z}_i \boldsymbol{\delta}_T + \ddot{\mathbf{x}}_{it} \boldsymbol{\alpha} + \ddot{u}_{it}$$

Então, não podemos estimar o coeficiente da variável sexo do indivíduo, por exemplo. Mas podemos estimar se houve mudança desse efeito ao longo do tempo, em relação a categoria de referência.

## 7.4 Hipóteses

**FE.1:** Exogeneidade Estrita:  $E(u_{it} | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, c_i) = 0$ , para  $t = 1, \dots, T$  e  $i = 1, \dots, N$ .

**FE.2:** Posto pleno de  $E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i)$  (para inverter a matriz).  $\text{posto}[E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i)] = K$ .

**FE.3:** Homoscedasticidade:  $E(\mathbf{u}_i \mathbf{u}_i' | \mathbf{X}_i, c_i) = \sigma_u^2 \mathbf{I}_T$ .

## 7.5 Estimação POLS

Wooldridge (2010, p.269)

Aplicando POLS no modelo transformado (7.7), temos:

$$\boxed{\hat{\boldsymbol{\beta}}^{FE} = \left( \sum_{i=1}^N \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right)^{-1} \left( \sum_{i=1}^N \ddot{\mathbf{X}}_i' \ddot{\mathbf{y}}_i \right) = \left( \sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}' \ddot{y}_{it} \right)}\tag{7.8}$$

Este estimador também é chamado de **estimador within**.

## 7.6 Consistência

Wooldridge (2010, sec.10.5.1 – Consistency of the Fixed Effects Estimator, p.265–269)

Wooldridge (2010, p.269)

Usando a equação (7.7) em (7.8), temos:

$$\hat{\boldsymbol{\beta}}^{FE} = \boldsymbol{\beta} + \left[ N^{-1} \sum_{i=1}^N \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right]^{-1} \left[ N^{-1} \sum_{i=1}^N \ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i \right]\tag{7.9}$$

Nota que para  $E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i) = \mathbf{0}$  é necessário não haver correlação entre todos os erros  $u_{it}$   $t = 1, \dots, T$  e todos os regressores  $\mathbf{x}_{it}'$   $t = 1, \dots, T$ . **FE.1** implica a condição acima. Além disso, sob **FE.1**,  $\hat{\boldsymbol{\beta}}^{FE}$  é não viciado.

## Formato Totalmente Matricial

Empilhando os vetores  $N$  vezes, vamos definir:

$$\ddot{\mathbf{X}}_{NT \times K} = \begin{bmatrix} \ddot{\mathbf{X}}_1 \\ \vdots \\ \ddot{\mathbf{X}}_N \end{bmatrix} = \begin{bmatrix} \ddot{x}_{1,11} & \dots & \ddot{x}_{K,11} \\ \vdots & & \vdots \\ \ddot{x}_{1,1T} & \dots & \ddot{x}_{K,1T} \\ \ddot{x}_{1,21} & \dots & \ddot{x}_{K,21} \\ \vdots & & \vdots \\ \ddot{x}_{1,2T} & \dots & \ddot{x}_{K,2T} \\ \vdots & & \vdots \\ \ddot{x}_{1,N1} & \dots & \ddot{x}_{K,N1} \\ \vdots & & \vdots \\ \ddot{x}_{1,NT} & \dots & \ddot{x}_{K,NT} \end{bmatrix} \quad \ddot{\mathbf{y}}_{NT \times 1} = \begin{bmatrix} \ddot{y}_1 \\ \vdots \\ \ddot{y}_N \end{bmatrix} = \begin{bmatrix} \ddot{y}_{11} \\ \vdots \\ \ddot{y}_{1T} \\ \ddot{y}_{21} \\ \vdots \\ \ddot{y}_{2T} \\ \vdots \\ \ddot{y}_{N1} \\ \vdots \\ \ddot{y}_{NT} \end{bmatrix} \quad \ddot{\mathbf{u}}_{NT \times 1} = \begin{bmatrix} \ddot{u}_1 \\ \vdots \\ \ddot{u}_N \end{bmatrix} = \begin{bmatrix} \ddot{u}_{11} \\ \vdots \\ \ddot{u}_{1T} \\ \ddot{u}_{21} \\ \vdots \\ \ddot{u}_{2T} \\ \vdots \\ \ddot{u}_{N1} \\ \vdots \\ \ddot{u}_{NT} \end{bmatrix}$$

A matriz  $\mathbf{X}$  é  $NT \times K$ , a matriz  $\mathbf{M}^0$  é  $T \times T$  e a matriz  $\mathbf{I}_N$  é  $N \times N$ . A produto **Kronecker** de  $\mathbf{I}_N$  por  $\mathbf{M}^0$ ,

$$\mathbf{I}_{N \times N} \otimes \mathbf{M}_{T \times T}^0$$

é uma matriz  $NT \times NT$ . Dessa forma, podemos definir:

$$\ddot{\mathbf{y}} = (\mathbf{I}_N \otimes \mathbf{M}^0) \mathbf{y},$$

$$\ddot{\mathbf{X}} = (\mathbf{I}_N \otimes \mathbf{M}^0) \mathbf{X}$$

$$\ddot{\mathbf{u}} = (\mathbf{I}_N \otimes \mathbf{M}^0) \mathbf{u},$$

E com isso, reescrevermos (7.9) como:

$$\hat{\beta}^{FE} = \beta + \left( \ddot{\mathbf{X}}' \ddot{\mathbf{X}} \right)^{-1} \ddot{\mathbf{X}}' \ddot{\mathbf{u}} \quad (7.10)$$

Ou ainda, como:

$$\begin{aligned} \hat{\beta}^{FE} &= \beta + [\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{M}^0)(\mathbf{I}_N \otimes \mathbf{M}^0)\mathbf{X}]^{-1} [\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{M}^0)(\mathbf{I}_N \otimes \mathbf{M}^0)\mathbf{u}] \\ \hat{\beta}^{FE} &= \beta + [\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{M}^0)\mathbf{X}]^{-1} [\mathbf{X}'(\mathbf{I}_N \otimes \mathbf{M}^0)\mathbf{u}] \end{aligned} \quad (7.11)$$

onde usamos as propriedades de simetria e idempotência da matriz  $\mathbf{M}^0$ .

## 7.7 Matriz de Covariância Robusta

Wooldridge (2010, sec.10.5.2 – Asymptotic Inference with Fixed Effects, p.269–272)

A matriz de covariância assintótica fica:

$$\mathbf{E} \left( \ddot{\mathbf{X}}' \ddot{\mathbf{X}} \right)^{-1} \mathbf{E} \left( \ddot{\mathbf{X}}' \ddot{\mathbf{u}} \ddot{\mathbf{u}}' \ddot{\mathbf{X}} \right) \mathbf{E} \left( \ddot{\mathbf{X}}' \ddot{\mathbf{X}} \right)^{-1}$$

A qual pode ser estimada por

$$\left( \ddot{\mathbf{X}}' \ddot{\mathbf{X}} \right)^{-1} \left( \sum_{i=1}^N \ddot{\mathbf{x}}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \ddot{\mathbf{x}}_i \right) \left( \ddot{\mathbf{X}}' \ddot{\mathbf{X}} \right)^{-1}$$

Analizando  $E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i \ddot{\mathbf{u}}_i' \ddot{\mathbf{X}}_i)$ :

$$\begin{aligned} E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{u}}_i \ddot{\mathbf{u}}_i' \ddot{\mathbf{X}}_i) &= E[(\mathbf{X}_i' \mathbf{M}^{0'}) (\mathbf{M}^0 \mathbf{u}_i) (\mathbf{u}_i' \mathbf{M}^{0'}) (\mathbf{M}^0 \mathbf{X}_i)] \\ &= E[(\mathbf{X}_i' \mathbf{M}^{0'}) \mathbf{u}_i \mathbf{u}_i' (\mathbf{M}^0 \mathbf{X}_i)] \\ &= E[\ddot{\mathbf{X}}_i' \mathbf{u}_i \mathbf{u}_i' \ddot{\mathbf{X}}_i] \end{aligned}$$

onde usamos as propriedades de simetria e idempotência da matriz  $\mathbf{M}^0$  e as definições de  $\ddot{\mathbf{X}}_i$  e  $\ddot{\mathbf{u}}_i$ .

Sob **FE.3**,  $E(\mathbf{u}_i \mathbf{u}_i' | \ddot{\mathbf{X}}_i) = \sigma_u^2 \mathbf{I}_T$ , temos:

$$E(\ddot{\mathbf{X}}_i' \mathbf{u}_i \mathbf{u}_i' \ddot{\mathbf{X}}_i) = E[E(\ddot{\mathbf{X}}_i' \mathbf{u}_i \mathbf{u}_i' \ddot{\mathbf{X}}_i | \mathbf{X}_i, c_i)] = E[\ddot{\mathbf{X}}_i' E(\mathbf{u}_i \mathbf{u}_i' | \mathbf{X}_i, c_i) \ddot{\mathbf{X}}_i] = \sigma_u^2 \mathbf{I}_T E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i)$$

Assim, a matriz de covariância fica:

$$E(\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1} \sigma_u^2 \mathbf{I}_T E(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i) E(\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1} = \boxed{\sigma_u^2 E(\ddot{\mathbf{X}}' \ddot{\mathbf{X}})^{-1}}$$

### Estimando Elementos da Matriz de Covariância

Queremos estimar  $\sigma_u^2$  por valores amostrais:

$$E(u_{it}^2) = \sigma_u^2 \quad \text{e} \quad E(\ddot{u}_{it}^2) = \sigma_u^2$$

$$\sigma_u^2 = E[(u_{it} - \bar{u}_i)^2] = E(u_{it}^2) + E(\bar{u}_i^2) - 2E(u_{it} \bar{u}_i)$$

utilizando  $\bar{u}_i = T^{-1} \sum_{t=1}^T u_{it}$ :

$$\begin{aligned} E(\ddot{u}_{it}^2) &= E(u_{it}^2) + T^{-1} \sum_{t=1}^T E(u_{it}^2) - 2E(u_{it} T^{-1} \sum_{t=1}^T u_{it}) = E(u_{it}^2) + T^{-1} \sum_{t=1}^T E(u_{it}^2) - 2T^{-1} \sum_{t=1}^T E(u_{it}^2) \\ &= \sigma_u^2 + \sigma_u^2/T - 2\sigma_u^2/T = \sigma_u^2(1 - 1/T) \implies \boxed{\sigma_u^2 = \frac{T}{T-1} \sigma_{\ddot{u}}^2} \end{aligned}$$

Utilizando estimadores amostrais para  $E(\ddot{u}_{it}^2)$ :

$$\hat{\sigma}_u^2 = \frac{T}{T-1} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{\ddot{u}}_{it}^2$$

Ajustando os Graus de Liberdade (Cortando  $T$ s e subtraindo  $K$  do número de regressores):

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1) - K} \sum_{i=1}^N \sum_{t=1}^T \hat{\ddot{u}}_{it}^2. \quad (7.12)$$

### Teste para Autocorrelação AR(1)

Wooldridge (2010, p.275)

$$\begin{aligned} E(\ddot{u}_{it} \ddot{u}_{i,t-1}) &= E[(u_{i,t} - \bar{u}_i)(u_{i,t-1} - \bar{u}_i)] \\ &= E(u_{i,t} u_{i,t-1}) - E(u_{i,t} \bar{u}_i) - E(\bar{u}_i u_{i,t-1}) + E(\bar{u}_i^2) \\ &= 0 - T^{-1} \sigma_u^2 - T^{-1} \sigma_u^2 + T^{-1} \sigma_u^2 \\ E(\ddot{u}_{it} \ddot{u}_{i,t-1}) &= -T^{-1} \sigma_u^2 \end{aligned}$$

$$\text{Corr}(\ddot{u}_{it}, \ddot{u}_{i,t-1}) = \frac{\text{E}(\ddot{u}_{it}\ddot{u}_{i,t-1})}{\text{E}(\ddot{u}_{it}^2)} = \frac{-T^{-1}\sigma_u^2}{\frac{T-1}{T}\sigma_u^2} = \frac{-1}{T-1}$$

Vamos testar

$$H_0 : \delta = \frac{-1}{T-1}$$

(ausência de correlação em ***u***) na equação:

$$\hat{\ddot{u}}_{it} = \delta \hat{\ddot{u}}_{i,t-1} + e_{it}$$

para  $t = 2, \dots, T$ . Fazer teste  $t$  robusto.

## 8 First Difference (FD, PD)

Wooldridge (2010, Sec.10.6 – First Difference Methods, p.279)

### 8.1 Modelo

O modelo linear de **efeitos não observados**:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (8.1)$$

para  $t = 1, \dots, T$  e  $i = 1, \dots, N$ .

O modelo contém explicitamente um componente não observado,  $c_i$ , que não varia no tempo. Tratamos o componente não observado como parte do erro, não como parâmetro a ser estimado. Aqui permitimos que  $c_i$  seja correlacionado com  $\mathbf{x}_{it}$ . Deste modo, **não** podemos ignorar a sua presença e estimar (8.1) por POLS, visto que isso resultaria num estimador inconsistente devido a **endogeneidade**.

Assim, transformamos o modelo para eliminar  $c_i$  e conseguirmos fazer uma estimação consistente de  $\boldsymbol{\beta}$ . A transformação a ser feita é a primeira diferença. Para tanto, seguimos os seguintes passos:

- Reescrevemos (8.1) defasado:

$$y_{it-1} = \mathbf{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \quad (8.2)$$

- Tiramos a diferença entre (8.2) e (8.1):

$$\begin{aligned} y_{it} - y_{it-1} &= (\mathbf{x}_{it} - \mathbf{x}_{it-1})\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{it-1} \\ \Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}. \end{aligned} \quad (8.3)$$

para  $t = 2, \dots, T$  e  $i = 1, \dots, N$ .

Reescrevendo (8.3) no formato matricial empilhando  $T$ :

$$\Delta \mathbf{y}_i = \Delta \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \quad (8.4)$$

com  $\boxed{e_{it} = \Delta u_{it}}$ .

- $\Delta \mathbf{y}_i$  vetor  $(T-1) \times 1$
- $\Delta \mathbf{X}_i$  matriz  $(T-1) \times K$
- $\boldsymbol{\beta}$  vetor  $K \times 1$
- $\mathbf{e}_i$  vetor  $(T-1) \times 1$

### 8.2 Estimação POLS

O estimador  $\hat{\boldsymbol{\beta}}^{FD}$  é o POLS da regressão no modelo (8.4), assim:

$$\boxed{\boldsymbol{\beta}^{FD} = \left[ \sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right]^{-1} \left[ \sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{y}_i \right]} \quad (8.5)$$

### 8.3 Hipóteses

As Hipóteses que usamos para  $\hat{\beta}^{FD}$  são:

**FD.1:** Exogeneidade Estrita:  $E(u_{it} | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}, c_i) = 0$ , para  $t = 1, \dots, T$  e  $i = 1, \dots, N$ .

**FD.2:** Posto completo de  $E(\Delta \mathbf{X}'_i \Delta \mathbf{X}_i)$  (para inverter a matriz).  $\text{posto}[E(\Delta \mathbf{X}'_i \Delta \mathbf{X}_i)] = K$ .

**FD.3:** Homoscedasticidade:  $E(\mathbf{e}_i \mathbf{e}'_i | \mathbf{X}_i, c_i) = \sigma_e^2 \mathbf{I}_{T-1}$ .

### 8.4 Valor Esperado

Usando apenas **FD.1** e **FD.2**:

$$E(\beta^{FD}) = \beta + E \left[ \left( \sum_{i=1}^N \Delta \mathbf{X}'_i \Delta \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \Delta \mathbf{X}'_i \mathbf{e}_i \right) \right]$$

$$E(\beta^{FD}) = \beta + E \left[ (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{e}) \right]$$

### 8.5 Variância

Usando apenas **FD.1** e **FD.2**:

$$\text{Var}(\beta^{FD}) = E \left[ (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \mathbf{e} \mathbf{e}' \Delta \mathbf{X}) (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} \right]$$

### 8.6 Variância sob Homocedasticidade

Usando **FD.3**, temos

$$\text{Var}(\beta^{FD}) = \sigma_e^2 E \left[ (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} (\Delta \mathbf{X}' \Delta \mathbf{X}) (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} \right]$$

$$\text{Var}(\beta^{FD}) = \sigma_e^2 E \left[ (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1} \right]$$

com

$$\sigma_e^2 = [N(T-1) - K]^{-1} \left[ \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2 \right],$$

que é a média de todos  $\hat{e}_{it}^2$  contando  $K$  regressores.

## 9 System GLS (SGLS)

Wooldridge (2010, Sec.7.4 – Consistency and Asymptotic Normality of Generalized Least Squares, p.153)

### 9.1 Modelo Linear

### 9.2 Hipóteses

Para implementarmos o estimador de **GLS** precisamos das seguintes hipótese:

1.  $E(\mathbf{X}_i \otimes \mathbf{u}_i) = 0$ .

Para SGLS ser consistente, precisamos que  $\mathbf{u}_i$  não seja correlacionada com nenhum elemento de  $\mathbf{X}_i$ .

2.  $\Omega$  é positiva definida (para ter inversa).  $E(\mathbf{X}_i' \Omega^{-1} \mathbf{X}_i)$  é **não** singular (para ter inversa).

Onde,  $\Omega$  é a seguinte matriz **simétrica**, positiva-definida:

$$\Omega = E(\mathbf{u}_i \mathbf{u}_i').$$

### 9.3 Estimação

Agora, transformamos o sistema de equações ao realizarmos a pré-multiplicação do sistema por  $\Omega^{-1/2}$ :

$$\Omega^{-1/2} \mathbf{y}_i = \Omega^{-1/2} \mathbf{X}_i \boldsymbol{\beta} + \Omega^{-1/2} \mathbf{u}_i$$

$$\mathbf{y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{u}_i^*$$

Estimando a equação acima por **SOLS**:

$$\begin{aligned} \beta^{SOLS} &= \left( \sum_{i=1} \mathbf{X}_i^{*'} \mathbf{X}_i^* \right)^{-1} \left( \sum_{i=1} \mathbf{X}_i^{*'} \mathbf{y}_i^* \right) \\ &= \left( \sum_{i=1} \mathbf{X}_i' \Omega^{-1/2} \Omega^{-1/2} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1} \mathbf{X}_i' \Omega^{-1/2} \Omega^{-1/2} \mathbf{y}_i \right) \\ &= \left( \sum_{i=1} \mathbf{X}_i' \Omega^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1} \mathbf{X}_i' \Omega^{-1} \mathbf{y}_i \right) \end{aligned}$$



## 10 GLS Factível

Wooldridge (2010, Sec.7.5 – Feasible GLS, p.153)

### 10.1 FSGLS: SGLS Factível

Para obtermos  $\beta^{SGLS}$  precisamos conhecer  $\Omega$ , o que não ocorre na prática. Então, precisamos estimar  $\Omega$  com um estimador consistente. Para tanto usamos um procedimento de dois passos:

1. Estimar  $\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i$  via **SOLS** e guardar o resíduo estimado  $\hat{\mathbf{u}}_i$ .
2. Estimar  $\Omega$  com o seguinte estimador  $\hat{\Omega}$ :

$$\hat{\Omega} = N^{-1} \sum_{i=1}^N \mathbf{u}_i \mathbf{u}_i'$$

Com a estimativa  $\hat{\Omega}$  feita, podemos obter  $\beta^{FSGLS}$  pela fórmula do  $\beta^{SGLS}$ :

$$\beta^{FGLS} = \left[ \sum_i \mathbf{X}_i' \hat{\Omega}^{-1} \mathbf{X}_i \right]^{-1} \left[ \sum_i \mathbf{X}_i' \hat{\Omega}^{-1} \mathbf{y}_i \right]$$

Empilhando as  $N$  observações:

$$\beta^{FGLS} = \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{y} \right]$$

Reescrevendo a equação acima:

$$\begin{aligned} \beta^{FGLS} &= \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) (\mathbf{X}\beta + \mathbf{u}) \right] \\ &= \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left\{ \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X} \right] \beta + \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \right] \right\} \\ &= \beta + \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \right] \end{aligned}$$

### 10.2 Valor Esperado

$$E(\beta^{FGLS}) = \beta + \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \right]$$

Concluimos que, se  $\hat{\Omega} \xrightarrow{p} \Omega$ , então,  $\beta^{FSGLS} \xrightarrow{p} \beta$ ,

### 10.3 Variância

$$\begin{aligned} \text{Var}(\beta^{FGLS}) &= \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \right] \left\{ \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \right] \right\}' \\ &= \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \mathbf{u}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X} \right] \left[ \mathbf{X} (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{X}' \right]^{-1} \end{aligned}$$

Tirando o valor Esperado e supondo que:

$$E(\mathbf{X}_i \Omega^{-1} \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i) = E(\mathbf{X}_i \Omega^{-1})$$

temos:

$$E \left[ \mathbf{X}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1}) \mathbf{u} \mathbf{u}' (\mathbf{I}_N \otimes \hat{\Omega}^{-1})' \mathbf{X} \right] = E(\mathbf{X}' \Omega^{-1} \mathbf{X})$$

e temos:

$$\text{Var}(\beta^{FSGLS}) = [E(\mathbf{X}' \Omega^{-1} \mathbf{X})]^{-1}.$$

## 11 Random Effects (RE, EA)

Wooldridge (2010, Sec.10.4 – Random Effects Methods, p.257)

### 11.1 Modelo

O modelo linear de **efeitos não observados**:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (11.1)$$

onde  $t = 1, \dots, T$  e  $i = 1, \dots, N$ .

O modelo contém explicitamente um componente não observado que não varia no tempo  $c_i$ . Abordamos esse componente como parte do erro, não como parâmetro a ser estimado. Para a análise de **Efeitos Aleatórios, (EA) ou (RE)**, supomos que os regressões  $\mathbf{x}_{it}$  são **não correlacionados** com  $c_i$ , mas fazemos hipóteses mais restritas que o **POLS**; pois assim exploramos a presença de **correlação serial** do erro composto por GLS e garantimos a consistência do estimador de FGLS.

Podemos reescrever (11.1) como:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \quad (11.2)$$

onde  $t = 1, \dots, T$ ,  $i = 1, \dots, N$  e  $\boxed{v_{it} = c_i + u_{it}}$  é o erro composto.

Agora, vamos empilhar os  $t$ 's e reescrever (11.2) como:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i, \quad (11.3)$$

onde  $i = 1, \dots, N$  e  $\boxed{\mathbf{v}_i = c_i\mathbf{1}_T + \mathbf{u}_i}$ .

### 11.2 Hipóteses

As Hipóteses que usamos para  $\hat{\boldsymbol{\beta}}^{RE}$  são:

1. Usamos o modelo correto e  $c_i$  não é endógeno.

- a)  $E(u_{it} | x_{i1}, \dots, x_{iT}, c_i) = 0$ ,  $i = 1, \dots, N$ .
- b)  $E(c_{it} | x_{i1}, \dots, x_{iT}) = E(c_i) = 0$ ,  $i = 1, \dots, N$ .

2. Posto completo de  $E(\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i)$ .

Definindo a matriz  $T \times T$ ,  $\boxed{\Omega \equiv E(\mathbf{v}_i\mathbf{v}_i')}$ , queremos que  $E(\mathbf{X}_i\Omega^{-1}\mathbf{X}_i)$  tenha posto completo (posto =  $K$ ).

A matriz  $\Omega$  é simétrica  $\Omega' = \Omega$  e positiva definida  $\det(\Omega) > 0$ . Assim podemos achar  $\Omega^{1/2}$  e  $\Omega^{-1/2}$  com  $\Omega = \Omega^{1/2}\Omega^{1/2}$  e  $\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$ .

### 11.3 Estimação

Premultiplicando (11.3) por  $\Omega^{-1/2}$  do dois lados, temos:

$$\begin{aligned} \Omega^{-1/2}\mathbf{y}_i &= \Omega^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \Omega^{-1/2}\mathbf{v}_i \\ \mathbf{y}_i^* &= \mathbf{X}_i^*\boldsymbol{\beta} + \mathbf{v}_i^*, \end{aligned} \quad (11.4)$$

Estimando o modelo acima por POLS:

$$\begin{aligned}
\beta^{POLS} &= \left( \sum_{i=1}^N \mathbf{X}_i^{*'} \mathbf{X}_i^* \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i^{*'} \mathbf{y}_i^* \right) \\
&= \left( \sum_{i=1}^N \mathbf{X}_i' \Omega^{-1} \mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^N \mathbf{X}_i' \Omega^{-1} \mathbf{y}_i \right) \\
&= (\mathbf{X}'(\mathbf{I}_N \otimes \Omega^{-1})\mathbf{X})^{-1} (\mathbf{X}'(\mathbf{I}_N \otimes \Omega^{-1})\mathbf{y}). \tag{11.5}
\end{aligned}$$

O problema, agora, é estimar  $\Omega$ . Supondo:

- $E(u_{it}u_{it}) = \sigma_u^2$ ;
- $E(u_{it}u_{is}) = 0$ .

Como  $\Omega = E(\mathbf{v}_i \mathbf{v}_i') = E[(c_i \mathbf{1}_T + \mathbf{u}_i)(c_i \mathbf{1}_T + \mathbf{u}_i)']$ , temos que:

$$\begin{aligned}
E(v_{it}v_{it}) &= E(c_i^2 + 2c_i u_{it} + u_{it}^2) = \sigma_c^2 + \sigma_u^2 \\
E(v_{it}v_{is}) &= E[(c_i + u_{it})(c_i + u_{is})] = E(c_i^2 + c_i u_{is} + u_{it}c_i + u_{it}u_{is}) = \sigma_c^2.
\end{aligned}$$

Assim,

$$\Omega = E(\mathbf{v}_i \mathbf{v}_i') = \sigma_u^2 \mathbf{I}_T + \sigma_c^2 \mathbf{1}_T \mathbf{1}_T'$$

onde  $\sigma_u^2 \mathbf{I}_T$  é uma matriz diagonal, e  $\sigma_c^2 \mathbf{1}_T \mathbf{1}_T'$  é uma matriz com todos os elementos iguais a  $\sigma_c^2$ .

Agora, rodando POLS em (11.3) e guardando os resíduos, temos:

$$\hat{v}_{it}^{POLS} = \hat{y}_{it}^{POLS} - \mathbf{x}_{it} \hat{\boldsymbol{\beta}}^{POLS}$$

e conseguimos estimar  $\sigma_v^2$  e  $\sigma_c^2$  por estimadores amostrais:

- como  $\sigma_v^2 = E(v_{it}^2)$ :

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

- como  $\sigma_c^2 = E(v_{it}v_{is})$ :

$$\hat{\sigma}_c^2 = \left[ N \frac{T(T-1)}{2} - K \right]^{-1} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is}$$

- $N$  indivíduos;
- $T$  elementos da diagonal principal de  $\Omega$
- $\frac{T(T-1)}{2}$  elementos da matriz triangular superior dos elementos fora da diagonal.
- $K$  regressores.

Agora que temos  $\hat{\sigma}_v^2$  e  $\hat{\sigma}_c^2$  podemos achar  $\hat{\sigma}_u^2$  pela equação  $\boxed{\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2}$ . Dessa forma, achamos os  $T^2$  elementos de  $\hat{\Omega}$ , e podemos escrever:

$$\hat{\Omega} = \hat{\sigma}_u^2 \mathbf{I}_T + \hat{\sigma}_c^2 \mathbf{1}_T \mathbf{1}_T'$$

Com  $\hat{\Omega}$  estimado, reescrevemos (11.5) como:

$$\beta^{RE} = [\mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{X}]^{-1} [\mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{y}]. \tag{11.6}$$

### 11.4 Valor Esperado

$$\boxed{E(\boldsymbol{\beta}^{RE}) = \boldsymbol{\beta} + \left[ \mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{X} \right]^{-1} \left[ \mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{v} \right]}.$$

### 11.5 Variância

$$\text{Var}(\boldsymbol{\beta}^{RE}) = E \left\{ \left[ \mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{X} \right]^{-1} \left[ \mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{v}\mathbf{v}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})'\mathbf{X} \right] \left[ \mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{X} \right] \right\},$$

como  $E(\mathbf{v}_i\mathbf{v}_i') = \Omega$ ,

$$\boxed{\text{Var}(\boldsymbol{\beta}^{RE}) = E \left[ \mathbf{X}'(\mathbf{I}_N \otimes \hat{\Omega}^{-1})\mathbf{X} \right]}.$$

## 12 Endogeneity and GMM

### 12.1 Modelo

No seguinte modelo *cross-section*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i ; \quad i = 1, \dots, N. \quad (12.1)$$

A variável explicativa  $x_k$  é dita **endógena** se ela for correlacionada com erro. Se  $x_k$  for não correlacionada com o erro, então  $x_k$  é dita **exógena**.

Endogeneidade surge, normalmente, de três maneiras diferentes:

1. Variável Omitida;
2. Simultaneidade;
3. Erro de Medida.

No modelo (12.1) vamos supor:

- $x_1$  é exógena.
- $x_2$  é endógena.

### 12.2 Hipóteses

Assim, precisamos encontrar um instrumento  $z_i$  para  $x_2$ , uma vez que queremos estimar  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  de maneira consistente. Para  $z_i$  ser um bom instrumento precisamos que  $z$  tenha:

1.  $Cov(z, \varepsilon) = 0 \implies z$  é exógena em (12.1).
2.  $Cov(z, x_2) \neq 0 \implies$  correlação com  $x_2$  após controlar para outras variáveis.

### 12.3 Estimação

Indo para o problema de dados de painel, temos:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i ; \quad i = 1, \dots, N. \quad (12.2)$$

onde  $\mathbf{y}_i$  é um vetor  $T \times 1$ ,  $\mathbf{X}_i$  é uma matriz  $T \times K$ ,  $\boldsymbol{\beta}$  é o vetor de coeficientes  $K \times 1$ ,  $\mathbf{u}_i$  é o vetor de erros  $T \times 1$ .

Se é verdade que há endogeneidade em (12.2), então:

$$E(\mathbf{X}_i' \mathbf{u}_i) \neq 0$$

Definimos  $Z_i$  como uma matriz  $T \times L$  com  $L \geq K$  de variáveis exógenas (incluindo o instrumento). Queremos acabar com a endogeneidade, ou seja:

$$E(Z_i' \mathbf{u}_i) = 0$$

Supondo  $L = K$  (apenas substituímos a variável endógena por um instrumento).

$$\begin{aligned} E[Z_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})] &= 0 \\ E(Z_i' \mathbf{y}_i) - E(Z_i' \mathbf{X}_i) \boldsymbol{\beta} &= 0 \\ E(Z_i' \mathbf{y}_i) &= E(Z_i' \mathbf{X}_i) \boldsymbol{\beta} \end{aligned}$$

$$\boxed{\boldsymbol{\beta} = [E(Z_i' \mathbf{X}_i)]^{-1} [E(Z_i' \mathbf{y}_i)]}$$

Se Usarmos estimadores amostrais:

$$\hat{\beta} = \left[ N^{-1} \sum_{i=1}^N Z_i' \mathbf{X}_i \right]^{-1} \left[ N^{-1} \sum_{i=1}^N Z_i' \mathbf{y}_i \right]$$

$$\boxed{\hat{\beta} = (Z' \mathbf{X})^{-1} (Z' \mathbf{y})}$$

Se  $L > K$ , vamos considerar:

$$\text{Min}_{\beta} E(Z_i \mathbf{u}_i)^2$$

onde:

$$\begin{aligned} E(Z_i \mathbf{u}_i)^2 &= E[(Z_i \mathbf{u}_i)'(Z_i \mathbf{u}_i)] = (Z' \mathbf{y} - Z' \mathbf{X} \beta)'(Z' \mathbf{y} - Z' \mathbf{X} \beta) \\ &= \mathbf{y}' Z Z' \mathbf{y} - \mathbf{y}' Z Z' \mathbf{X} \beta - \beta' \mathbf{X}' Z Z' \mathbf{y} + \beta' \mathbf{X}' Z Z' \mathbf{X} \beta \end{aligned}$$

Derivando em relação em  $\beta$  e igualando a zero:

$$\begin{aligned} -2\mathbf{y}' Z Z' \mathbf{X} + 2\beta' \mathbf{X}' Z Z' \mathbf{X} &= 0 \\ \beta' \mathbf{X}' Z Z' \mathbf{X} &= \mathbf{y}' Z Z' \mathbf{X} \\ \beta' &= (\mathbf{y}' Z Z' \mathbf{X})(\mathbf{X}' Z Z' \mathbf{X})^{-1} \\ \boxed{\beta &= (\mathbf{X}' Z Z' \mathbf{X})^{-1}(\mathbf{X}' Z Z' \mathbf{y})} \end{aligned}$$

Um estimador mais eficiente pode ser encontrado fazendo:

$$\text{Min}_{\beta} E[(Z_i' \mathbf{y} - Z_i' \mathbf{X} \beta)' W (Z_i' \mathbf{y} - Z_i' \mathbf{X} \beta)].$$

Escolhendo  $\widehat{W}$ , a priori, temos:

$$\text{Min}_{\beta} \left\{ \mathbf{y}' Z \widehat{W} Z' \mathbf{y} - \mathbf{y}' Z \widehat{W} Z' \mathbf{X} \beta - \beta' \mathbf{X}' Z \widehat{W} Z' \mathbf{y} + \beta' \mathbf{X}' Z \widehat{W} Z' \mathbf{X} \beta \right\}$$

Derivando em relação em  $\beta$  e igualando a zero:

$$\begin{aligned} -2\mathbf{y}' Z \widehat{W} Z' \mathbf{X} + 2\beta' \mathbf{X}' Z \widehat{W} Z' \mathbf{X} &= 0 \\ \beta' \mathbf{X}' Z \widehat{W} Z' \mathbf{X} &= \mathbf{y}' Z \widehat{W} Z' \mathbf{X} \\ \beta' &= (\mathbf{y}' Z \widehat{W} Z' \mathbf{X})(\mathbf{X}' Z \widehat{W} Z' \mathbf{X})^{-1} \\ \boxed{\beta^{GMM} &= (\mathbf{X}' Z \widehat{W}' Z' \mathbf{X})^{-1}(\mathbf{X}' Z \widehat{W}' Z' \mathbf{y})} \end{aligned}$$

## 12.4 Valor Esperado

$$\boxed{E(\beta^{GMM}) = \beta + E[(\mathbf{X}' Z \widehat{W}' Z' \mathbf{X})^{-1}(\mathbf{X}' Z \widehat{W}' Z' \mathbf{u})]}.$$

## 12.5 Variância

$$\begin{aligned}\text{Var}(\beta^{GMM}) &= E \left\{ \left[ (X'Z\widehat{W}'Z'X)^{-1}(X'Z\widehat{W}'Z'\mathbf{u}) \right] \left[ (X'Z\widehat{W}'Z'X)^{-1}(X'Z\widehat{W}'Z'\mathbf{u}) \right]' \right\} \\ &= E \left\{ (X'Z\widehat{W}'Z'X)^{-1}X'Z\widehat{W}'Z'\mathbf{u}\mathbf{u}'Z\widehat{W}Z'X(X'Z\widehat{W}'Z'X)^{-1} \right\}.\end{aligned}$$

Definindo  $\Delta = E(Z'\mathbf{u}\mathbf{u}'Z)$  com  $\Delta = W^{-1}$ :

$$\begin{aligned}\text{Var}(\beta^{GMM}) &= E \left\{ (X'Z\widehat{W}'Z'X)^{-1}X'Z\widehat{W}'W^{-1}\widehat{W}Z'X(X'Z\widehat{W}'Z'X)^{-1} \right\} \\ &= E \left\{ (X'Z\widehat{W}'Z'X)^{-1}(X'Z\widehat{W}'Z'X)(X'Z\widehat{W}'Z'X)^{-1} \right\}.\end{aligned}$$

$$\boxed{\text{Var}(\beta^{GMM}) = E \left[ (X'Z\widehat{W}'Z'X)^{-1} \right]}.$$

Se tivéssemos definido  $W = (Z'Z)^{-1}$ , teríamos  $\beta^{2SLS}$ .

## 13 Exogeneidade Estrita e FDIIV

### 13.1 Modelo

No seguinte modelo

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it},$$

para  $t = 1, \dots, T$  e  $i = 1, \dots, N$ .

- $y_{it}$  escalar;
- $\mathbf{x}_{it}$  vetor  $1 \times K$ ;
- $\boldsymbol{\beta}$  vetor  $K \times 1$ ;
- $u_{it}$  escalar.

$\{x_{it}\}$  é estritamente **exógeno** se valer:

$$E(u_{it} | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = 0, \quad t = 1, \dots, T$$

ou seja:

$$E(y_{it} | \mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = \mathbf{x}_{it}\boldsymbol{\beta}, \quad t = 1, \dots, T$$

o que é equivalente a hipótese de que utilizamos o modelo linear correto.

Para o seguinte modelo:

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \rho y_{it-1} + c_i + u_{it}, \quad t = 2, \dots, T$$

é **impossível** termos exogeneidade estrita. Isso porque, nesse modelo, de efeitos não observados temos:

$$E(y_{it} | \mathbf{z}_{i1}, \dots, \mathbf{z}_{iT}, y_{it-1}, c_i) \neq 0.$$

Isso ocorre porque,  $y_{it}$  é afetado por  $y_{it-1}$  que contribui para  $y_{it}$  com, pelo menos,  $\rho c_i$ .

$$\left. \begin{aligned} y_{it} &= \mathbf{z}_{it}\boldsymbol{\gamma} + \rho y_{it-1} + c_i + u_{it} \\ y_{it-1} &= \mathbf{z}_{it-1}\boldsymbol{\gamma} + \rho y_{it-2} + c_i + u_{it-1} \end{aligned} \right\} \implies y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \rho(\mathbf{z}_{it-1}\boldsymbol{\gamma} + \rho y_{it-2} + c_i + u_{it-1}) + c_i + u_{it}.$$

Para eliminarmos este efeito, podemos tirar a primeira diferença do modelo:

$$y_{it} - y_{it-1} = (\mathbf{z}_{it} - \mathbf{z}_{it-1})\boldsymbol{\gamma} + \rho(y_{it-1} - y_{it-2}) + (c_i - c_i) + (u_{it} - u_{it-1})$$

$\Delta y_{it} = \Delta \mathbf{z}_{it}\boldsymbol{\gamma} + \rho \Delta y_{it-1} + \Delta u_{it}, \quad t = 3, \dots, T$

(13.1)

### 13.2 Estimação

Não podemos estimar o modelo (13.1) por POLS, uma vez que  $Cov(\Delta y_{it-1}, \Delta u_{it}) \neq 0$ . Como saída, podemos estimar por P2SLS, usando instrumentos para  $\Delta y_{it-1}$  (alguns instrumentos para  $\Delta y_{it-1}$  são  $y_{it-2}, y_{it-3}, \dots, y_{i1}$ ).



### 13.2.1 P2SLS

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it}$$

- $i = 1, \dots, N$
- $t = 1, \dots, T$
- $y_{it}$  escalar;
- $\mathbf{x}_{it}$  vetor  $K \times 1$ ;
- $\boldsymbol{\beta}$  vetor  $K \times 1$ ;
- $u_{it}$  escalar.

$$\boxed{\boldsymbol{\beta}^{P2SLS} = (X'P_ZX)^{-1}(X'P_Z\mathbf{y})}$$

com

$$\boxed{P_Z = Z'(Z'Z)^{-1}Z}$$

onde  $P_Z$  é a matriz de projeção em  $Z$ .

### 13.2.2 FDIV

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T$$

$$\Delta y_{it} = \Delta \mathbf{x}'_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad i = 1, \dots, N, \quad t = 2, \dots, T$$

Vamos supor  $\Delta \mathbf{x}'_{it}$  tem variável endógena ( $y_{it}$ , no caso).  $\mathbf{w}_{it}$  é um vetor  $1 \times L_t$  de instrumentos, onde  $L_t \geq K$ . Se os instrumentos forem diferentes:

$$W_i = \text{diag}(\mathbf{w}'_{i2}, \mathbf{w}'_{i3}, \dots, \mathbf{w}'_{iT})$$

onde  $W_i$  é uma matriz  $(T-1) \times L$

$$L = L_2 + L_3 + \dots + L_T$$

## 13.3 Hipóteses

**FDIV.1:**  $E(\mathbf{w}_{it}\Delta u_{it})$  para  $i = 1, \dots, N, t = 2, \dots, T$ .

**FDIV.2:**  $Posto[E(W'_iW_i)] = L$

**FDIV.3:**  $Posto[E(W'_i\Delta X_i)] = K$

## 13.4 Estimação FDIV

$$\boxed{\boldsymbol{\beta}^{FDIV} = (\Delta X'P_W\Delta X)^{-1}(\Delta X'P_W\Delta \mathbf{y})}$$

$$\boxed{P_W = W(W'W)^{-1}W'}$$

## 13.5 Valor Esperado

$$E(\boldsymbol{\beta}^{FDIV}) = \boldsymbol{\beta} + (\Delta X'P_W\Delta X)^{-1}(\Delta X'P_W\mathbf{e})$$

### 13.6 Variância

$$\begin{aligned}
 \text{Var}(\boldsymbol{\beta}^{FDIV}) &= \text{E} \left\{ [\text{E}(\boldsymbol{\beta}^{FDIV}) - \boldsymbol{\beta}] [\text{E}(\boldsymbol{\beta}^{FDIV}) - \boldsymbol{\beta}]' \right\} \\
 &= \text{E} \left\{ [\Delta X' P_W \Delta X]^{-1} [\Delta X' P_W \mathbf{e}] [\Delta X' P_W \mathbf{e}]' [\Delta X' P_W \Delta X]^{-1} \right\} \\
 &= \text{E} \left[ (\Delta X' P_W \Delta X)^{-1} (\Delta X' P_W \mathbf{e} \mathbf{e}' P_W \Delta X) (\Delta X' P_W \Delta X)^{-1} \right]
 \end{aligned}$$

$$e_i = \Delta u_{it}.$$

## 14 Latent Variables, Probit and Logit

### 14.1 Modelo

Suponha  $y^*$  não observável (**latente**) seguindo o seguinte modelo:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i. \quad (14.1)$$

Defina  $y$  como:

$$y_i = \begin{cases} 1, & y_i^* \geq 0 \\ 0, & y_i^* < 0 \end{cases}$$

temos que:

$$P(y_i = 1 | \mathbf{x}) = p(\mathbf{x})$$

$$P(y_i = 0 | \mathbf{x}) = 1 - p(\mathbf{x}).$$

Além disso, pela definição de  $y_i$ , equação (14.1), temos:

$$\begin{aligned} P(y_i = 1 | \mathbf{x}) &= P(y_i^* \geq 0 | \mathbf{x}) \\ &= P(\mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \geq 0 | \mathbf{x}) \\ &= P(\varepsilon_i \geq -\mathbf{x}_i' \boldsymbol{\beta} | \mathbf{x}). \end{aligned}$$

Agora, supondo que  $\varepsilon_i$  tem FDA,  $G$ , tal que  $G' = g$  é simétrica ao redor de zero:

$$\begin{aligned} P(y_i = 1 | \mathbf{x}) &= 1 - P(\varepsilon_i < -\mathbf{x}_i' \boldsymbol{\beta} | \mathbf{x}) \\ &= 1 - G(-\mathbf{x}_i' \boldsymbol{\beta} | \mathbf{x}) \\ &= G(\mathbf{x}_i' \boldsymbol{\beta}). \end{aligned}$$

Se  $G(\cdot)$  for uma distribuição:

**Normal Padrão:**  $\hat{\boldsymbol{\beta}}$  é o estimador **probit**.

**Logística:**  $\hat{\boldsymbol{\beta}}$  é o estimador **logit**.

Supondo  $\mathbf{y}_i | \mathbf{x} \sim \text{Bernoulli}(p(\mathbf{x}))$ , sua fmp é dada por:

$$f(y_i | \mathbf{x}_i; \boldsymbol{\beta}) = [G(\mathbf{x}_i' \boldsymbol{\beta})]^{y_i} [1 - G(\mathbf{x}_i' \boldsymbol{\beta})]^{1-y_i}, \quad y = 0, 1.$$

Para estimarmos  $\hat{\boldsymbol{\beta}}$  por máxima verossimilhança, temos de encontrar  $\boldsymbol{\beta} \in B$ , onde  $B$  é o espaço paramétrico, tal que  $\boldsymbol{\beta}$  maximize o valor da distribuição conjunta de  $\mathbf{y}$ , ou seja:

$$\text{Max}_{\boldsymbol{\beta} \in B} \prod_{i=1}^N f(y_i | \mathbf{x}_i; \boldsymbol{\beta}).$$

Tirando o logaritmo e dividindo tudo por  $N$  (podemos fazer isso pois são transformações monotônicas e não alteram o lugar onde  $\boldsymbol{\beta}$  ótimo irá parar):

$$\text{Max}_{\boldsymbol{\beta} \in B} \left\{ N^{-1} \sum_{i=1}^N \ln [f(y_i | \mathbf{x}_i; \boldsymbol{\beta})] \right\}.$$

Podemos definir  $\ell_i(\boldsymbol{\beta}) = \ln[f(y_i | \mathbf{x}_i; \boldsymbol{\beta})]$  como sendo a verossimilhança condicional da observação  $i$ :

$$\max_{\beta \in B} \left\{ N^{-1} \sum_{i=1}^N \ell_i(\beta) \right\}.$$

Dessa forma, podemos ver que o problema acima é a analogia amostral de:

$$\max_{\beta \in B} E[\ell_i(\beta)].$$

Definindo o *vector score* da observação  $i$ :

$$s_i(\beta) = [\nabla_{\beta} \ell_i(\beta)]' = \left[ \frac{\partial \ell_i(\beta)}{\partial \beta_1}, \dots, \frac{\partial \ell_i(\beta)}{\partial \beta_K} \right]$$

Definindo a **Matriz Hessiana** da observação  $i$ :

$$H_i(\beta) = \nabla_{\beta} s_i(\beta) = \nabla_{\beta}^2 \ell_i(\beta)$$

Tendo essas definições, o **Teorema do Valor Médio** (TVM) nos diz que no intervalo  $[a, b]$ , existe um número,  $c$ , tal que:

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

### FAZER DESENHO

Trocando  $f(\cdot)$  por  $s_i(\cdot)$ ,  $a$  por  $\beta_0$ ,  $b$  por  $\hat{\beta}$  e  $c$  por  $\bar{\beta}$ , temos:

$$H_i(\bar{\beta}) = \frac{s_i(\hat{\beta}) - s_i(\beta_0)}{\hat{\beta} - \beta_0},$$

tirando médias dos dois lados:

$$N^{-1} \sum_{i=1}^N H_i(\bar{\beta}) = \frac{1}{\hat{\beta} - \beta_0} N^{-1} \sum_{i=1}^N [s_i(\hat{\beta}) - s_i(\beta_0)]$$

Supondo que  $\hat{\beta}$  maximiza  $\ell(\beta | \mathbf{y}, \mathbf{x})$ , temos que:  $N^{-1} \sum_{i=1}^N s_i(\hat{\beta}) = 0$ . E podemos reescrever a equação anterior como:

$$\begin{aligned} \hat{\beta} - \beta_0 &= (-1) \left[ N^{-1} \sum_{i=1}^N H_i(\bar{\beta}) \right]^{-1} N^{-1} \sum_{i=1}^N s_i(\beta_0) \\ \sqrt{N}(\hat{\beta} - \beta_0) &= \left[ -N^{-1} \sum_{i=1}^N H_i(\bar{\beta}) \right]^{-1} \sqrt{N} \cdot N^{-1} \sum_{i=1}^N s_i(\beta_0) \\ \boxed{\sqrt{N}(\hat{\beta} - \beta_0) &= \left[ -N^{-1} \sum_{i=1}^N H_i(\bar{\beta}) \right]^{-1} N^{-1/2} \sum_{i=1}^N s_i(\beta_0).} \end{aligned}$$

Onde

$$\left[ -N^{-1} \sum_{i=1}^N H_i(\hat{\beta}) \right]^{-1} \xrightarrow{p} A_0^{-1}, \quad N^{-1/2} \sum_{i=1}^N s_i(\beta_0) \xrightarrow{d} N(0, B_0).$$

Assim, temos que:

$$\boxed{\sqrt{N}(\hat{\beta} - \beta_0) \rightarrow N(0, A_0^{-1} B_0 A_0^{-1})}.$$

A forma mais simples de achar  $\text{Var}(\hat{\beta})$  é:

$$\boxed{\text{Var}(\hat{\beta}) = -E[H_i(\hat{\beta})]^{-1}}.$$

## 15 ATT, ATE, Propensity Score

### 15.1 Modelo

- $y_1 \rightarrow$  variável de interesse com tratamento
- $y_0 \rightarrow$  variável de interesse sem tratamento

$$w = \begin{cases} 1 & \text{se tratam} \\ 0 & \text{se não tratam} \end{cases}$$

Idealmente, para isolarmos completamente o efeito de  $w = 1$ , gostaríamos de poder calcular:

$$N^{-1} \sum_{i=1}^N (y_{i1} - y_{i0}).$$

Ou seja, o efeito que o tratamento causa sobre um indivíduo com todo o resto permanecendo constante. Em outras palavras, queríamos que houvesse dois mundos paralelos observáveis onde seria possível observar o que acontece com  $y_i$  com e sem tratamento. Infelizmente, para cada indivíduo  $i$ , observamos apenas  $y_{i1}$  ou  $y_{i0}$ , nunca ambos.

Antes de continuarmos, faremos as seguintes definições:

**ATE:**  $E(y_1 - y_0)$

**ATT:**  $E(y_1 - y_0 | w = 1)$  (ATE no tratado).

### 15.2 ATE e ATT condicional a variáveis dependentes

$$ATE(\mathbf{x}) = E(y_1 - y_0 | \mathbf{x})$$

$$ATT(\mathbf{x}) = E(y_1 - y_0 | \mathbf{x}, w = 1)$$

OBS:

$$E(y_1 - y_0) = E[E(y_1 - y_0 | w)]$$

$$E(y_1 - y_0 | w) = E(y_1 - y_0 | w = 0) \cdot P(w = 0) + E(y_1 - y_0 | w = 1) \cdot P(w = 1).$$

### 15.3 Métodos Assumindo Ignorabilidade do Tratamento

**ATE.1:** Ignorabilidade.

$w$  e  $(y_1, y_0)$  são independentes condicionais a  $\mathbf{x}$ .

**ATE.1':** Ignorabilidade da Média.

$$\text{a) } E(y_0 | w, \mathbf{x}) = E(y_0 | \mathbf{x})$$

$$\text{b) } E(y_1 | w, \mathbf{x}) = E(y_1 | \mathbf{x})$$

Vamos definir

$$E(y_0 | \mathbf{x}) = \mu_0(\mathbf{x})$$

$$E(y_1 | \mathbf{x}) = \mu_1(\mathbf{x}).$$

Sob **ATE.1** e **ATE.1'**:

$$ATE(\mathbf{x}) = E(y_1 - y_0 | \mathbf{x}) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$$

$$ATT(\mathbf{x}) = E(y_1 - y_0 | \mathbf{x}, w = 1) = \mu_1(\mathbf{x}) - \mu_0(\mathbf{x})$$

**ATE.2: Overlap**

Para todo  $\mathbf{x}$ ,  $P(w = 1 | \mathbf{x}) \in (0, 1)$ ,  $p(\mathbf{x}) = p(w = 1 | \mathbf{x})$ .

$p(\mathbf{x})$  é o *Propensity Score*, ele representa a probabilidade de  $y_i$  ser tratado dado o valor das covariáveis  $\mathbf{x}$ . Essa hipótese é importante visto que podemos expressar o *ATE* em função de  $p(\mathbf{x})$ .

Para o *ATT* vamos supor:

**ATT.1':**  $E(y_0 | \mathbf{x}, w) = E(y_0 | \mathbf{x})$

**ATT.2: Overlap:** Para todo  $\mathbf{x}$ ,  $P(w = 1 | \mathbf{x}) < 1$ .

**15.4 Propensity Score**

Como foi dito anteriormente, apenas observamos ou  $y_1$  ou  $y_0$  para a mesma pessoa, mas não ambos. Mais precisamente, junto com  $w$ , o resultado observado é:

$$y = wy_1 + (1 - w)y_0$$

como  $w$  é binário,  $w^2 = w$ , assim, temos:

$$\begin{aligned} wy &= w^2y_1 + (w - w^2)y_0 \implies \boxed{wy = wy_1} \\ (1 - w)y &= (w - w^2)y_1 + (w^2 - 2w + 1)y_0 \implies \boxed{(1 - w)y = (1 - w)y_0}. \end{aligned}$$

Fazemos isso para tentar isolar  $\mu_0(\mathbf{x})$  e  $\mu_1(\mathbf{x})$ :

$$\mu_1(\mathbf{x})$$

$$\begin{aligned} E(wy | \mathbf{x}) &= E[E(wy_1 | \mathbf{x}, w) | \mathbf{x}] \\ &= E[w\mu_1(\mathbf{x}) | \mathbf{x}] \\ &= \mu_1(\mathbf{x})E(w | \mathbf{x}). \end{aligned}$$

Como  $w$  é binária:  $E(w | \mathbf{x}) = P(w = 1 | \mathbf{x}) = p(\mathbf{x})$ . Assim:

$$E(wy | \mathbf{x}) = \mu_1(\mathbf{x})p(\mathbf{x})$$

$$\boxed{\mu_1(\mathbf{x}) = \frac{E(wy | \mathbf{x})}{p(\mathbf{x})}}$$

$$\mu_0(\mathbf{x})$$

$$\begin{aligned} E[(1 - w)y | \mathbf{x}] &= E[E((1 - w)y_0 | \mathbf{x}, w) | \mathbf{x}] \\ &= E[(1 - w)\mu_0(\mathbf{x}) | \mathbf{x}] \\ &= \mu_0(\mathbf{x})E(w | \mathbf{x}) \end{aligned}$$

$$E[(1 - w)y | \mathbf{x}] = \mu_0(\mathbf{x})[1 - p(\mathbf{x})] \implies$$

$$\boxed{\mu_0(\mathbf{x}) = \frac{E[(1 - w)y | \mathbf{x}]}{1 - p(\mathbf{x})}}$$

**ATE:**

$$\mu_1(\mathbf{x}) - \mu_0(\mathbf{x}) = E \left[ \frac{[w - p(\mathbf{x})]y}{p(\mathbf{x})[1 - p(\mathbf{x})]} | \mathbf{x} \right]$$

$$\boxed{\widehat{ATE} = N^{-1} \sum_{i=1}^N \frac{[w_i - p(\mathbf{x}_i)]y_i}{p(\mathbf{x}_i)[1 - p(\mathbf{x}_i)]}}$$

**ATT:**

$$E(y_1|\mathbf{x}, w = 1) - E(y_0|\mathbf{x}) = \frac{1}{\hat{P}(w = 1)} E \left[ \frac{[w - \hat{p}(\mathbf{x})]y}{[1 - \hat{p}(\mathbf{x})]} | \mathbf{x} \right]$$

$$\hat{P}(w = 1) = N^{-1} \sum_{i=1}^N w_i$$

$$\widehat{ATT} = \frac{N}{\sum_{i=1}^N w_i} N^{-1} \sum_{i=1}^N \frac{[w_i - \hat{p}(\mathbf{x}_i)]y_i}{[1 - \hat{p}(\mathbf{x}_i)]}$$

$$\boxed{\widehat{ATT} = \frac{1}{\sum_{i=1}^N w_i} \sum_{i=1}^N \frac{[w_i - \hat{p}(\mathbf{x}_i)]y_i}{[1 - \hat{p}(\mathbf{x}_i)]}}$$



## 16 Álgebra Linear

### 16.1 Vetores

### 16.2 Operações com Vetores

Vamos definir os vetores  $\mathbf{x}$  e  $\mathbf{y}$  com dimensão  $1 \times N$ :

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

**Multiplicação por escalar**

**Soma**

**Subtração**

**Multiplicação de vetores**

**Produto Interno (Produto Escalar, Dot Product, Inner Product)**

$$\mathbf{x}'\mathbf{y} = \mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x} \rangle \mathbf{y} = \sum_{i=1}^N x_i y_i = x_1 y_1 + \cdots + x_N y_N. \quad (16.1)$$

Podemos utilizar a equação (16.1) para denotarmos a soma dos elementos de um vetor. Para tanto, definimos o vetor  $\mathbf{1}_N$  como sendo o vetor cujos elementos são 1 e tem dimensão  $N \times 1$ . (Greene, 2012, p. 977, A.2.7)

$$x_1 + \cdots + x_N = \sum_{i=1}^N x_i = \mathbf{x}'\mathbf{1}_N = \mathbf{1}_N' \mathbf{x} = (\mathbf{x}'\mathbf{1}_N)'$$

Com a definição do vetor  $\mathbf{1}_N$ , também podemos escrever:  $\mathbf{1}_N' \mathbf{1}_N = N$ .

Usando a definição de **média aritmética**, também podemos representá-la da seguinte forma:

$$\frac{x_1 + \cdots + x_N}{N} = \bar{x} = N^{-1} \sum_{i=1}^N x_i = N^{-1} \mathbf{x}'\mathbf{1}_N$$

Usando a definição de **média ponderada**, também podemos representá-la da seguinte forma:

$$w_1 x_1 + \cdots + w_N x_N = \sum_{i=1}^N w_i x_i = \mathbf{w}' \mathbf{x}$$

onde  $\mathbf{w}'\mathbf{1} = 1$

**Produto Externo (Outer Product)**

$$\begin{aligned}
\mathbf{1}_N \mathbf{1}'_N &= \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{N \times N} & \mathbf{1}_N \mathbf{x}' &= \begin{bmatrix} x_1 & \dots & x_N \\ \vdots & \ddots & \vdots \\ x_1 & \dots & x_N \end{bmatrix}_{N \times N} & \mathbf{x} \mathbf{1}'_N &= \begin{bmatrix} x_1 & \dots & x_1 \\ \vdots & \ddots & \vdots \\ x_N & \dots & x_N \end{bmatrix}_{N \times N} \\
\mathbf{x} \mathbf{x}' &= \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_N \\ x_2 x_1 & x_2^2 & \dots & x_2 x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_N x_1 & x_N x_2 & \dots & x_N^2 \end{bmatrix} \quad (16.2)
\end{aligned}$$

**Distância e Ângulo** O coseno do ângulo,  $\alpha$ , entre dois vetores é:

$$\cos(\alpha) = \frac{\mathbf{u}'\mathbf{v}}{\|\mathbf{u}\| \cdot \|\mathbf{v}\|}$$

Dois vetores são **ortogonais** se  $\mathbf{u}'\mathbf{v} = 0$ .

**Projeções Ortogonal**

$$\text{proj}_{\mathbf{u}}(\mathbf{y}) = \hat{\mathbf{y}} = \frac{\mathbf{y}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}\mathbf{u}$$

é a **projeção ortogonal** de  $\mathbf{y}$  em  $\mathbf{u}$ .

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \frac{\mathbf{y}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}\mathbf{u}$$

é a **componente de  $\mathbf{y}$  ortogonal a  $\mathbf{u}$  (Rejeição)**.

Por construção, temos:

$$\mathbf{z} + \hat{\mathbf{y}} = \mathbf{y}.$$

[FIGURA AQUI]

**Outras Notações Projeções Ortogonal** Dado que

$$\text{proj}_{\mathbf{u}}(\mathbf{y}) = \frac{\mathbf{y}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}\mathbf{u}$$

é a **projeção ortogonal** de  $\mathbf{y}$  em  $\mathbf{u}$ . Usando  $\mathbf{y}'\mathbf{u} = \mathbf{u}'\mathbf{y}$ , temos

$$\text{proj}_{\mathbf{u}}(\mathbf{y}) = \frac{\mathbf{u}\mathbf{u}'}{\mathbf{u}'\mathbf{u}}\mathbf{y}$$

Tirando o  $\mathbf{y}$  da equação, obtemos o **operador de projeção (Matriz de projeção em  $\mathbf{u}$ ?)**:

$$\frac{\mathbf{u}\mathbf{u}'}{\mathbf{u}'\mathbf{u}} = \mathbf{u}(\mathbf{u}'\mathbf{u})^{-1}\mathbf{u}' = (\mathbf{u}'\mathbf{u})^{-1}\mathbf{u}\mathbf{u}' = (\|\mathbf{u}\|^2)^{-1}\mathbf{u}\mathbf{u}'$$

Agora, podemos definir o **operador rejeição** como:

$$\mathbf{I}_N - \frac{\mathbf{u}\mathbf{u}'}{\mathbf{u}'\mathbf{u}}$$

Usando o caso especial onde  $\mathbf{u}$  é um vetor de uns ( $\mathbf{u} = \mathbf{1}_N$ ), temos a **projeção** no eixo de 45 graus:

$$(\mathbf{1}'_N \mathbf{1}_N)^{-1} \mathbf{1}_N \mathbf{1}'_N = N^{-1} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{N \times N} = \begin{bmatrix} 1/N & \dots & 1/N \\ \vdots & \ddots & \vdots \\ 1/N & \dots & 1/N \end{bmatrix}_{N \times N}$$

A **rejeição** no eixo de 45 graus:

$$\mathbf{I}_N - \frac{\mathbf{1}_N \mathbf{1}'_N}{\mathbf{1}'_N \mathbf{1}_N} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} - \begin{bmatrix} 1/N & 1/N & \dots & 1/N \\ 1/N & 1/N & \dots & 1/N \\ \vdots & \vdots & \ddots & \vdots \\ 1/N & 1/N & \dots & 1/N \end{bmatrix} = \begin{bmatrix} (N-1)/N & 1/N & \dots & 1/N \\ 1/N & (N-1)/N & \dots & 1/N \\ \vdots & \vdots & \ddots & \vdots \\ 1/N & 1/N & \dots & (N-1)/N \end{bmatrix}$$

**Centering Matrix** (Greene, 2012, p. 978, A.28)

$$\mathbf{M}^0 = \mathbf{I}_N - \mathbf{1}_N (\mathbf{1}'_N \mathbf{1}_N)^{-1} \mathbf{1}'_N = \mathbf{I}_N - N^{-1} \mathbf{1}_N \mathbf{1}'_N$$

A Matriz  $\mathbf{M}^0$  é **idempotente** e **simétrica**.

**Idempotência:**  $AA = A$

**Simetria:**  $A' = A$

$$\mathbf{M}^0 \mathbf{x} = (\mathbf{I}_N - N^{-1} \mathbf{1}_N \mathbf{1}'_N) \mathbf{x} = \mathbf{x} - N^{-1} \mathbf{1}_N (\mathbf{1}'_N \mathbf{x}) = \mathbf{x} - \mathbf{1}_N \bar{x}$$

onde podemos denotar

$$\bar{x} = \mathbf{1}_N \bar{x} = \begin{bmatrix} \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix}$$

$$\mathbf{M}^0 \mathbf{x} = \mathbf{x} - \bar{x}$$

$$\mathbf{M}^0 \mathbf{1} = (\mathbf{I}_N - N^{-1} \mathbf{1}_N \mathbf{1}'_N) \mathbf{1}_N = \mathbf{1}_N - N^{-1} \mathbf{1}_N (\mathbf{1}'_N \mathbf{1}_N) = \mathbf{1}_N - \mathbf{1}_N = \mathbf{0}_N$$

### 16.3 Operações com Matrizes

**Multiplicação por escalar**

**Soma**

**Subtração**

**Multiplicação de Matriz**

$$A_{2 \times 2} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B_{2 \times 3} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}$$

$$[AB]_{2 \times 3} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} \end{bmatrix} + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} \begin{bmatrix} b_{21} & b_{22} & b_{23} \end{bmatrix} \implies AB = \sum_{i=1}^2 a_i b_i$$

onde  $a_i$  é a  $i$ -ésima **coluna** da matriz  $A$ .  $b_i$  é a  $i$ -ésima **linha** da matriz  $B$ .

### 16.3.1 Projeções

$$\mathbf{P}_X = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

$$\mathbf{M}_X = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

Usando o modelo

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

E definindo  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ , temos

$$\mathbf{P}_X\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} = \hat{\mathbf{y}}$$

$$\mathbf{M}_X\mathbf{y} = \mathbf{I}\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{y} - \hat{\mathbf{y}} = \hat{\mathbf{u}}$$

## 17 Conceitos Básicos de Convergência Estatística

**Definition 17.1 (Convergência em Probabilidade).** Wooldridge (2010, Def 3.3, p.36 )

Uma sequência de variáveis aleatórias:  $\{X_N\}_{N \geq 1}$  **converge em probabilidade** para uma variável aleatória  $X$  se, dado  $\varepsilon > 0$ ,

$$P(|X_N - X| > \varepsilon) \rightarrow 0,$$

quando  $N \rightarrow +\infty$ . E denotamos

$$\text{plim } X_N = X, \quad \text{ou} \quad X_N \xrightarrow{p} X, \quad \text{ou} \quad X_N - X \xrightarrow{p} 0.$$

**Definition 17.2 (Estimador Consistente).** Wooldridge (2010, Def 3.8, p.40 )

Seja  $\{\theta_N : N = 1, 2, \dots\}$  uma sequência de estimadores do vetor  $\theta \in \Theta$  com dimensão  $P \times 1$ , onde  $N$  indexa o tamanho da amostra. Se

$$\hat{\theta} \xrightarrow{p} \theta \tag{17.1}$$

Para qualquer valor de  $\theta$ , então dizemos que  $\theta_N$  é um estimador consistente de  $\theta$ .

**Theorem 17.1 (LGN – Lei dos Grandes Números).** Achar referência

Seja  $\{X_i\}_{i \geq 1}$  uma sequência de variáveis aleatórias *iid* com  $E(X_i) = \mu$ . Então,

$$N^{-1} \sum_{i=1}^N X_i \xrightarrow{p} \mu.$$

**Theorem 17.2 (LGN – Caso Matricial).** Achar referência. O Teorema (17.3), abaixo, é diferente.

Seja  $\{x_i\}_{i=1}^N$ , uma sequência *iid* de vetores aleatórios  $K \times 1$  com  $E(x_i x_i') = Q_{K \times K}$  finita. Então,

$$N^{-1} \sum_{i=1}^N x_i x_i' \xrightarrow{p} Q.$$

Se  $Q$  for positiva definida,  $Q$  terá inversa.

**Theorem 17.3 (LGNF – WLLN).** Wooldridge (2010, Teo 3.1, p.39 )

Seja  $\{w_i : i = 1, 2, \dots\}$ , uma sequência *iid* de vetores aleatórios  $G \times 1$  com  $E(|w_{ig}|) < \infty$  para  $g = 1, \dots, G$ . Então, a sequência satisfaz a **Lei dos Grandes Números Fraca (WLLN)**:

$$N^{-1} \sum_{i=1}^N w_i \xrightarrow{p} \mu_w,$$

onde  $\mu_w \equiv E(w_i)$ .

**Definition 17.3** ( $o_p$ ).

Wooldridge (2010, Def 3.4, p.36 )

Wooldridge (2010, Lemma 3.2, p.36 )

$$X_n = o_p(1) \implies X_n \xrightarrow{p} 0$$

$$X_n = o_p(Y_n) \implies \frac{X_n}{Y_n} = o_p(1) \implies \frac{X_n}{Y_n} \xrightarrow{p} 0$$

$$X_n = W_n + o_p(1) \implies (X_n - W_n) = o_p(1) \implies (X_n - W_n) \xrightarrow{p} 0$$

**Definition 17.4 (Limitação em Probabilidade:  $O_p$ ).** Wooldridge (2010, Def 3.3 (3), p.36 )

Dizemos que  $X_n$  é **limitado em probabilidade** e denotado por  $X_n = O_p(1)$ , se existe  $M$  maior que zero, tal que para todo  $\varepsilon$  maior que zero,  $P(|X_n| > 0) < \varepsilon$ .

$$X_n = O_p(1) \implies \exists M > 0; \forall \varepsilon > 0, P(|X_n| > 0) < \varepsilon.$$

**Definition 17.5.** Wooldridge (2010, Def 3.4, p.36)

Dizemos que  $X_n = O_p(Y_n)$  se existe  $M$  maior que zero, tal que para todo  $\varepsilon$  maior que zero,  $P(|X_n/Y_n| > 0) < \varepsilon$ .

$$X_n = O_p(Y_n) \implies \exists M > 0; \forall \varepsilon > 0, P(|X_n/Y_n| > 0) < \varepsilon.$$

**Lemma 17.1.** Wooldridge (2010, Lemma 3.2, p.36)

Se  $X_n = O_p(1)$  e  $Y_n = o_p(1)$ , então

$$X_n Y_n = O_p(1) o_p(1) = o_p(1).$$

**Lemma 17.2 (Equivalência Assintótica).** Wooldridge (2010, Lemma 3.7, p.39)

Seja  $\{x_n\}$  e  $\{z_n\}$  sequências de vetores aleatórios  $K \times 1$ . Se  $z_n \xrightarrow{d} z$  e  $x_n - z_n \xrightarrow{p} \mathbf{0}_K$ . Então,

$$x_n \xrightarrow{d} z.$$

**Definition 17.6 (Convergência em Distribuição).** Wooldridge (2010, Def 3.6, p.38)

Seja  $\{X_n\}_{n \geq 1}$  uma sequência de variáveis aleatórias e  $X$  uma variável aleatória com  $F_n$  e  $F$  suas respectivas FDAs, então

$$X_n \xrightarrow{d} X, \text{ se } F_n(X) \rightarrow F(X)$$

para todo  $X$  onde  $F$  é contínuo.

**Lemma 17.3 (Convergência em Distribuição e Limitação em Probabilidade).** Wooldridge (2010, Lemma 3.5, p.39)

Se  $X_n \xrightarrow{d} X$ ,  $X$  um variável aleatória qualquer; então  $X_n = O_p(1)$ .

**Definition 17.7 (TCL – Teorema Central do Limite).** Achar Referência

Seja  $\{X_n\}_{n=1}^N$  iid com  $E(X_n) = \mu$  e  $\text{Var}(X_n) = \sigma^2 < +\infty$ . Então, para  $S_N = \sum_{n=1}^N X_n$ :

$$\frac{S_N - N\mu}{\sqrt{N}\sigma} = \frac{N(\bar{X} - \mu)}{\sqrt{N}\sigma} = \frac{N(\bar{X} - \mu)}{\sqrt{N}\sigma} = \boxed{\frac{\sqrt{N}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0, 1)}.$$

**Theorem 17.4 (TCL – Lindeberg-Levy).** Wooldridge (2010, Teo 3.2, p.40)

Seja  $\{w_i : i = 1, 2, \dots\}$  uma sequência iid de vetores aleatórios  $G \times 1$  com  $E(w_{ig}^2) < \infty$  para  $g = 1, \dots, G$  e  $E(w_i) = \mathbf{0}$ . Então,  $\{w_i : i = 1, 2, \dots\}$  satisfaz o **Teorema Central do Limite (CLT)**; qual seja:

$$N^{-1/2} \sum_{i=1}^N w_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{B}),$$

onde,  $\mathbf{B} = \text{Var}(w_i) = E(w_i w_i')$  é necessariamente positiva semidefinida. Para nossos propósitos,  $\mathbf{B}$  será sempre positiva definida.

**Corollary 17.1.** Wooldridge (2010, Cor 3.2, p.39)

Seja  $\{\mathbf{z}_N\}$  uma sequência de vetores  $K \times 1$  aleatórios tal que  $\mathbf{z}_N \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V})$ , então:

1. Para qualquer matriz  $\mathbf{A}$  de dimensão  $K \times M$  **não** estocástica, temos:

$$\mathbf{A}'\mathbf{z}_N \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A}).$$

$$2. \mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \xrightarrow{d} \chi_K^2 \text{ (ou } \mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \overset{a}{\sim} \chi_K^2).$$

**Definition 17.8** (raiz de  $N$  assintoticamente normalmente distribuido). Wooldridge (2010, Def 3.9, p.40)

Seja  $\{\hat{\boldsymbol{\theta}}_N : N = 1, 2, \dots\}$  uma sequência de estimadores do vetor  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  com dimensão  $P \times 1$ . Suponha que

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}) \quad (17.2)$$

onde  $\mathbf{V}$  é uma matriz  $P \times P$  positiva semidefinida. Então dizemos que  $\hat{\boldsymbol{\theta}}_N$  é  $\sqrt{N}$ -asintoticamente normalmente distribuido e  $\mathbf{V}$  é a **variância assintótica** de  $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ , denotada  $\text{Avar}\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) = \mathbf{V}$

*Remark.* Apesar de  $\mathbf{V}/N = \text{Var}(\hat{\boldsymbol{\theta}}_N)$  ser verdade apenas em casos especiais, e raramente  $\hat{\boldsymbol{\theta}}_N$  ter uma distribuição exatamente normal, tratamos  $\hat{\boldsymbol{\theta}}_N$  como se

$$\hat{\boldsymbol{\theta}}_N \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{V}/N). \quad (17.3)$$

sempre que a equação (17.2) for verdade. Por essa razão,  $\mathbf{V}/N$  é chamado de **variância assintótica** de  $\hat{\boldsymbol{\theta}}_N$ , e escrevemos:

$$\text{Avar}(\hat{\boldsymbol{\theta}}_N) = \mathbf{V}/N. \quad (17.4)$$

Abaixo, temos as definições necessárias para mostrar como verificar se um estimador é consistente por EQM. Quero saber se mantemos essa parte. Se sim, precisamos de referências.

**Definition 17.9** (Desigualdade de Markov). Achar referência

Seja  $\{X_n\}_{n \geq 1}$  uma sequência de variáveis aleatórias com  $E|X_n|^K < +\infty$ ,  $K > 0$ . Então, dado  $\varepsilon > 0$

$$P(|X_n| > \varepsilon) \leq \frac{E|X_n|^K}{\varepsilon^K}$$

**Definition 17.10.** Achar referência

$$0 \leq P(|\hat{\theta} - \theta| > \varepsilon) \leq \frac{E|X_n|^2}{\varepsilon^2}$$

**Definition 17.11** (Erro Quadrático Médio). Achar referência

$$EQM(\hat{\theta}) = E \left[ \left( \hat{\theta} - \theta \right)^2 \right] = \left[ \text{Bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) \right]$$

*Remark.* Achar referência?

Então, se  $\text{Bias}(\hat{\theta}) \rightarrow 0$  e  $\text{Var}(\hat{\theta}) \rightarrow 0$ , temos que  $EQM(\hat{\theta}) \rightarrow 0$ . Pelo **Teorema do Sanduíche**,  $P(|\hat{\theta} - \theta| > \varepsilon) \rightarrow 0$ ; logo,  $\hat{\theta} \xrightarrow{p} \theta$ .

---

[Bryant \(1984\)](#) [Herr \(1980\)](#) [Sydsaeter \*et al.\* \(2005\)](#) [Ruud \(2000\)](#) [Davidson & MacKinnon \(2004\)](#) [Faraway \(2015\)](#) [Boyd & Vandenberghe \(2018\)](#)



Table 1: Vector Space Ideas Expressed in Different Ways

Vector Space Idea	Geometric Expression		Statistical Expression	Probabilistic Expression
	Coordinate Free Form	Analytic Form		
Vector $z$	Directed Line Segment from 0 to $z$	$z = (z_1, \dots, z_n)$	Data $z_1, \dots, z_n$	Random Variable $z$
Vector $z - y$	Directed Line Segment from $y$ to $z$	$z - y = (z_1 - y_1, \dots, z_n - y_n)$	Differences $z_1 - y_1, \dots, z_n - y_n$	Random variable $z - y$
Inner Product of $x$ and $z$ , $\langle x, z \rangle$	Measure of Nonorthogonality	$x_1 z_1 + \dots + x_n z_n$	Sums of cross products	$E(xz)$
$ z ^2 = \langle z, z \rangle$	Squared distance, from 0 to $z$	$z_1^2 + \dots + z_n^2$	Sum of Squares of $z$	$E(z^2)$
$ z - y ^2$	Squared distance, from $y$ to $z$	$(z_1 - y_1)^2 + \dots + (z_n - y_n)^2$	Sum of Squared differences	$E(z - y)^2$
$\langle x, z \rangle = 0$	Orthogonality: $x, z$ perpendicular			$E(xz) = 0$
All vectors of the form $ax + by + cz + \dots$	Plane $L(x, y, z, \dots)$ through 0 determined by vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}, \dots$		All models of the form $ax + by + cz + \dots$	All square-integrable functions of $x, y, z, \dots$
A vector $P$ in $L(x, y, \dots)$ such that $\langle x - P, Q \rangle = 0$ for every $Q$ in $L$	Projection $P_L(z)$ of $z$ onto $L(x, y, \dots)$		The best-fitting model of the form $ax + by + c + \dots$	$E(z x, y, \dots)$
“45-degree” vector $J$	$J = (1, 1, \dots, 1)$		Constant $(1, 1, \dots, 1)$	Constant 1
$ x - P_J(x) ^2$	$\sum (x_i - \bar{x})^2$		$(n - 1)x$ sample variance of $x$	Variance of $x$
$\frac{\langle x - P_J(x), z - P_J(z) \rangle}{ x - P_J(x)  \cdot  z - P_J(z) }$	cosine of angle between $x - P_J(x)$ and $z - P_J(z)$		sample correlation of $x$ and $z$	Correlation of $x$ and $z$
$\langle x - P_J(x), z - P_J(z) \rangle = 0$	$x - P_J(x)$ and $z - P_J(z)$ are perpendicular		$x$ and $z$ are uncorrelated	$x$ and $z$ are uncorrelated

Source: [Bryant \(1984, Table 1, p.43\)](#)

Figure 1: Vector Addition

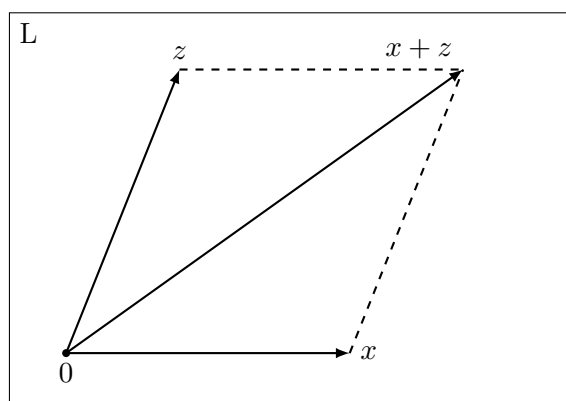


Figure 2: Vector Subtraction

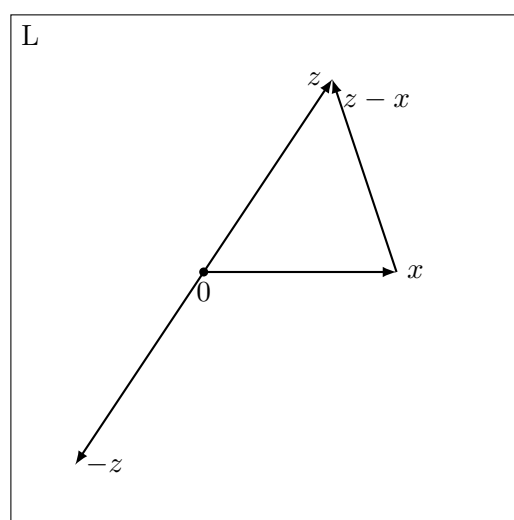


Figure 3: Scalar Multiplication and Subspaces

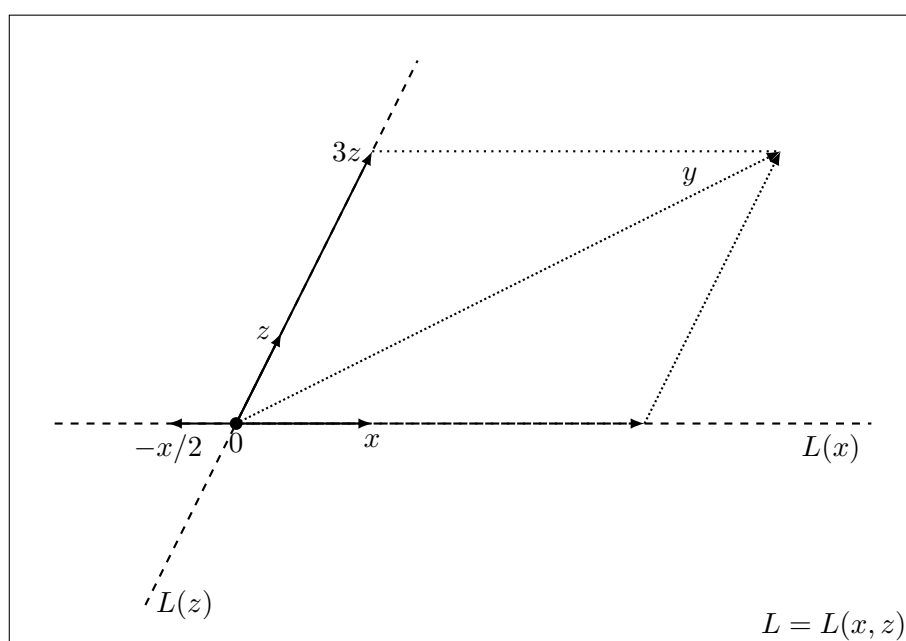


Figure 4: The Law of Cosines 1

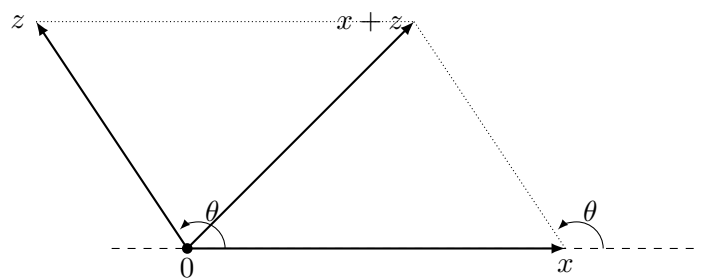


Figure 5: The Law of Cosines 2

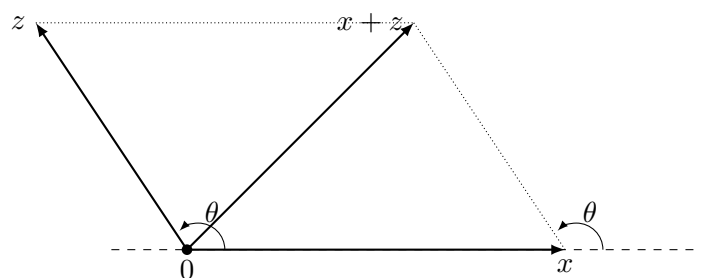
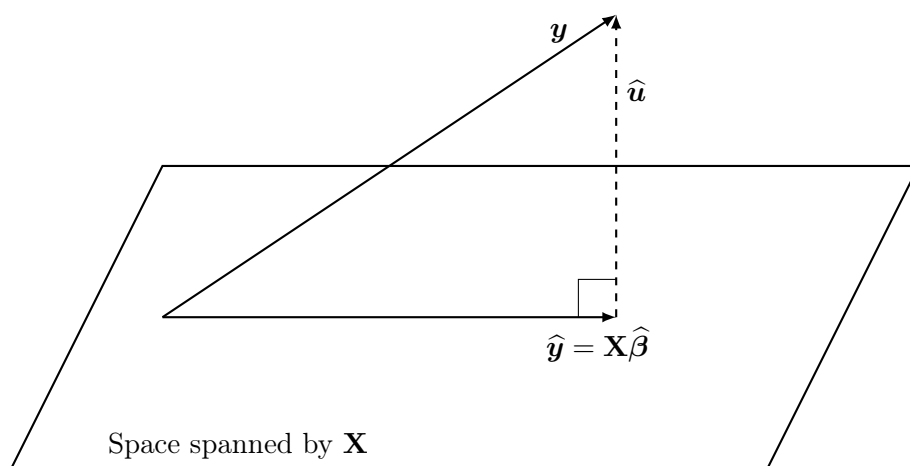
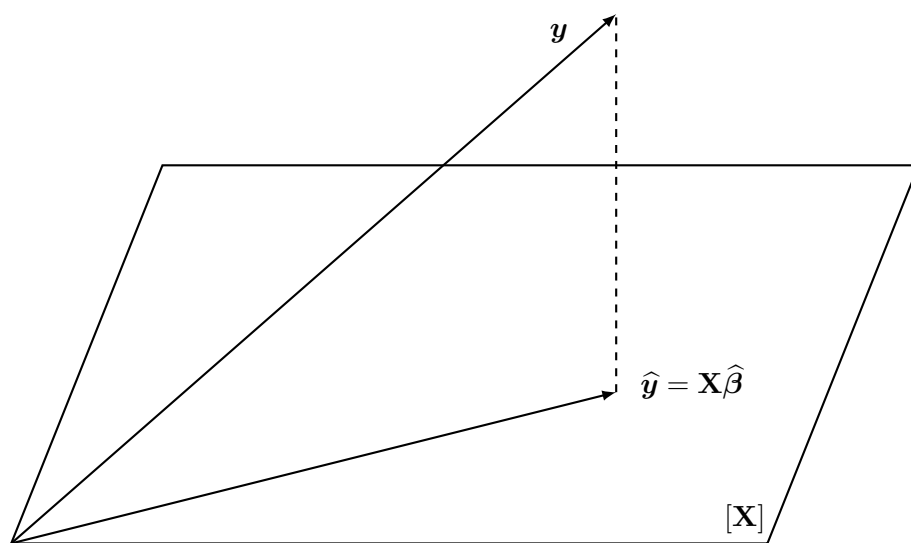


Figure 6: Projection



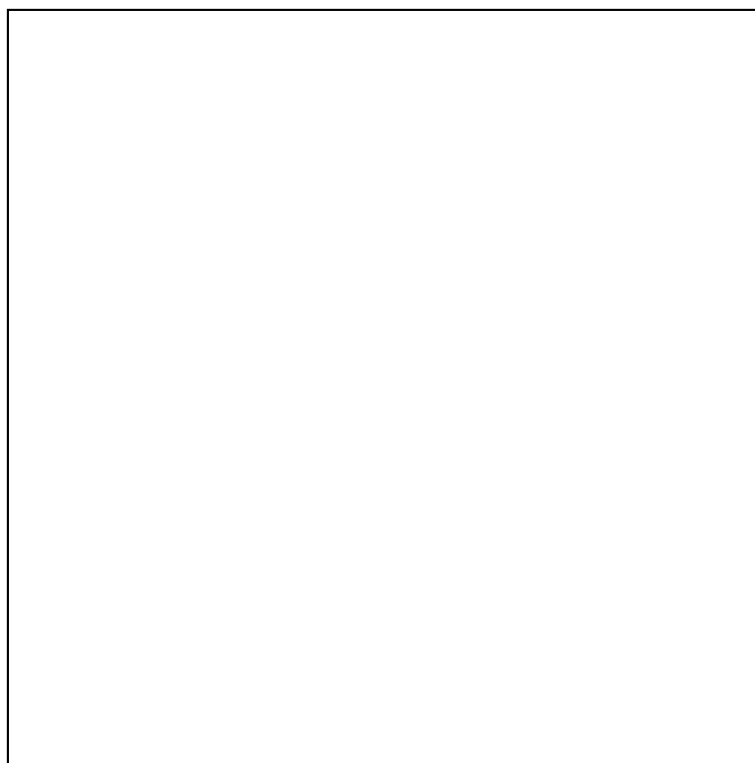
Faraway (2015, p.15)

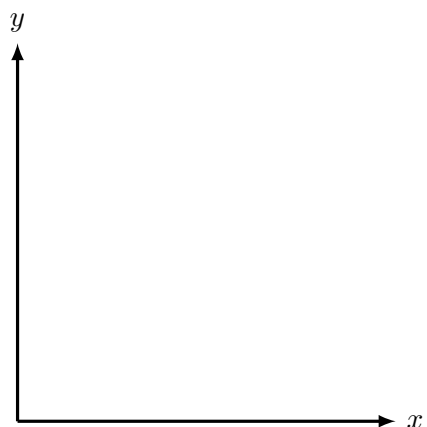
Figure 7: Least Squares Estimate



Herr (1980, p.44)

Figure 8: Square





## References

- BOYD, STEPHEN, & VANDENBERGHE, LIEVEN. 2018. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. 1st edn. Cambridge: Cambridge University Press.
- BRYANT, PETER. 1984. Geometry, Statistics, Probability: Variations on a Common Theme. *The American Statistician*, **38**(1), 38–48.
- DAVIDSON, RUSSEL, & MACKINNON, JAMES G. 2004. *Econometric Theory and Methods*. 1 edn. New York: Oxford University Press.
- FARAWAY, JULIAN J. 2015. *Linear Models with R*. 2nd edn. Boca Raton: CRC Press.
- GREENE, WILLIAM H. 2012. *Econometric Analysis*. 7 edn. Boston: Prentice Hall.
- HERR, DAVID G. 1980. On the History of the Use of Geometry in the General Linear Model. *The American Statistician*, **34**(1), 43–47.
- RUUD, PAUL A. 2000. *An Introduction to Classical Econometric Theory*. 1st edn. New York: Oxford University Press.
- SYDSAETER, KNUT, STROM, ARNE, & BERCK, PETER. 2005. *Economists' Mathematical Manual*. 4th edn. Heidelberg: Springer.
- WOOLDRIDGE, JEFFREY M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2 edn. Boston, Massachusetts: MIT Press.