

# Microeconometrics

Microeconometrics: Lecture Notes [Wooldridge](#) (2010)

Autor: Paulo Ferreira Naibert

Porto Alegre  
04/07/2020  
Revisão: July 5, 2020

**Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2ED. Boston, Massachusetts: MIT Press.**

I Introduction and Background

1. Introduction
2. Conditional Expectations and Related Concepts
3. Basic Asymptotic Theory

II Linear Models

4. The Single-Equation Linear Model and OLS Estimation
5. Instrumental Variables Estimation of Single-Equation Linear Models
6. Additional Single-Equation Topics
7. Estimating Systems of Equations by OLS and GLS
8. System Estimation by Instrumental Variables
9. Simultaneous Equations Models
10. Basic Linear Unobserved Effects Panel Data Models
11. More Topics in Linear Unobserved Effects Models

III General Approaches to Nonlinear Estimation

12. M-Estimation
13. Maximum Likelihood Methods
14. Generalized Method of Moments and Minimum Distance Estimation

IV Nonlinear Models and Related Topics

15. Discrete Response Models
16. Corner Solution Outcomes and Censored Regression Models
17. Sample Selection, Attrition, and Stratified Sampling
18. Estimating Average Treatment Effects
19. Count Data and Related Models
20. Duration Analysis

# 1 Introduction

[Wooldridge \(2010, C.1, pp. 3–9\)](#)

- 1.1. Causal Relationships and Ceteris Paribus Analysis
- 1.2. The Stochastic Setting and Asymptotic Analysis
  - 1.2.1. Data Structures
  - 1.2.2. Asymptotic Analysis
- 1.3. The Stochastic Setting and Asymptotic Analysis
- 1.4. Some Examples
- 1.5. Why Not Fixed Explanatory Analysis

## 2 Conditional Expectations and Related Concepts

[Wooldridge \(2010, C.2, pp. 13–34\)](#)

- 2.1. The Role of Conditional Expectations in Econometrics
- 2.2. Features of Conditional Expectations
  - 2.2.1. Definition and Examples
  - 2.2.2. Partial Effects, Elasticities, and Semielasticities
  - 2.2.3. The Error Form of Models of Conditional Expectations
  - 2.2.4. Some Properties of Conditional Expectations
  - 2.2.5. Average Partial Effects
- 2.3. Linear Projections
  - A Appendices
    - A.1 Properties of Conditional Expectations
    - A.2 Properties of Conditional Variances
    - A.3 Properties of Linear Projections

### 2.1 The Role of Conditional Expectations in Econometrics

### 2.2 Features of Conditional Expectations

#### 2.2.1 Definition and Examples

#### 2.2.2 Partial Effects, Elasticities, and Semielasticities

#### 2.2.3 The Error Form of Models of Conditional Expectations

#### 2.2.4 Some Properties of Conditional Expectations

#### 2.2.5 Average Partial Effects

### 2.3 Linear Projections

### 2.4 Appendices

#### 2.4.1 Properties of Conditional Expectations

CE.1

CE.2

CE.3

CE.4

CE.5

CE.6

CE.7

CE.8

### **2.4.2   Properties of Conditional Variances**

**CV.1**

**CV.2**

**CV.3**

**CV.4**

### **2.4.3   Properties of Linear Projections**

**LP.1**

**LP.2**

**LP.3**

**LP.4**

**LP.5**

**LP.6**

**LP.7**

### 3 Basic Asymptotic Theory

[Wooldridge \(2010, C.3, p.35–45\)](#)

- 3.1. Convergence of Deterministic Sequences
- 3.2. Convergence in Probability and Bounded in Probability
- 3.3. Convergence in Distribution
- 3.4. Limit Theorems for Random Samples
- 3.5. Limiting Behavior of Estimators and Test Statistics
  - 3.5.1. Asymptotic Properties of Estimators
  - 3.5.2. Asymptotic Properties of Test Statistics

[White \(1984\)](#).

#### 3.1 Convergence of Deterministic Sequences

**Definition 3.1.**

1. A sequence of nonrandom numbers  $\{a_N; N = 1, 2, \dots\}$  **converges** to  $a$  (has limit  $a$ ) if for all  $\varepsilon > 0$  there exists  $N_\varepsilon$  such that if  $N > N_\varepsilon$ , then  $|a_N - a| < \varepsilon$ . We write  $a_N \rightarrow a$  as  $N \rightarrow +\infty$ .
2. A sequence  $\{a_N; N = 1, 2, \dots\}$  is **bounded** if and only if there is some  $b < \infty$  such that  $|a_N| < b$  for all  $N = 1, 2, \dots$ . Otherwise, we say that  $\{a_N\}$  is **unbounded**.

**Definition 3.2.**

1. A sequence  $\{a_N\}$  is  $O(N^\lambda)$  (*at most of order  $N^\lambda$* ) if  $N^{-\lambda}a_N$  is bounded. When  $\lambda = 0$ ,  $\{a_N\}$  is bounded, and we also write  $a_N = O(1)$  (*big oh one*).
2. A sequence  $\{a_N\}$  is  $o(N^\lambda)$  if  $N^{-\lambda}a_N \rightarrow 0$ . When  $\lambda = 0$ ,  $a_N$  converges to zero, and we also write  $a_N = o(1)$  (*little oh one*).

#### 3.2 Convergence in Probability and Bounded in Probability

**Definition 3.3.**

1. A sequence of random variables  $\{x_N; N = 1, 2, \dots\}$  **converges in probability** to the constant  $a$  if for all  $\varepsilon > 0$ ,

$$P[|x_N - a| > \varepsilon] \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

We write  $x_N \xrightarrow{p} a$  and say that  $a$  is the **probability limit** (plim) of  $x_N$ ;  $\text{plim } x_N = a$ .

2. In the special case where  $a = 0$  ( $\text{plim } x_N = 0$ ), we also say that  $\{x_N\}$  is  $o_p(1)$  (*little oh p one*). We also write  $x_N = o_p(1)$  or  $x_N \xrightarrow{p} 0$  or
3. A sequence of random variables  $\{x_N\}$  is **bounded in probability** if and only if for every  $\varepsilon > 0$ , there exists a  $b_\varepsilon < \infty$  and an integer  $N_\varepsilon$  such that

$$P[|x_N| \geq b_\varepsilon] < \varepsilon \quad \text{for all } N \geq N_\varepsilon.$$

We write  $x_N = O_p(1)$  ( $x_N$  is *big oh p one*).

**Lemma 3.1.** If  $x_N \xrightarrow{p} a$ , then  $x_N = O_p(1)$ . This lemma also holds for vectors and matrices.

**Definition 3.4.**

1. A random sequence  $\{x_N : N = 1, 2, \dots\}$  is  $o_p(a_N)$ , where  $\{a_N\}$  is a nonrandom, positive sequence if  $x_N/a_N = o_p(1)$ . We write  $x_N = o_p(a_N)$ .
2. A random sequence  $\{x_N : N = 1, 2, \dots\}$  is  $O_p(a_N)$ , where  $\{a_N\}$  is a nonrandom, positive sequence if  $x_N/a_N = O_p(1)$ . We write  $x_N = O_p(a_N)$ .

*Remark.* We could have started by defining a sequence  $\{x_N\}$  to be  $o_p(N^\delta)$  for  $\delta \in \mathbb{R}$  if  $N^{-\delta}x_N \xrightarrow{p} 0$ , in which case we obtain the definition of  $o_p(1)$  when  $\delta = 0$ . This is where the one in  $o_p(1)$  comes from. A similar remark holds for  $O_p(1)$ .

**Lemma 3.2.** If  $w_N = o_p(1)$ ,  $x_N = o_p(1)$ ,  $y_N = O_p(1)$ , and  $z_N = O_p(1)$ , then

1.  $w_N + x_N = o_p(1)$
2.  $y_N + z_N = O_p(1)$
3.  $y_N z_N = O_p(1)$
4.  $x_N z_N = o_p(1)$

*Remark.*

1.  $o_p(1) + o_p(1) = o_p(1)$
2.  $O_p(1) + O_p(1) = O_p(1)$
3.  $O_p(1)O_p(1) = O_p(1)$
4.  $o_p(1)O_p(1) = o_p(1)$

Because a  $o_p(1)$  sequence is also  $O_p(1)$ , we also have:

5.  $o_p(1) + O_p(1) = O_p(1)$
6.  $o_p(1)o_p(1) = o_p(1)$

*Remark.* All of the previous definitions apply element by element to sequences of random vectors or matrices.

**Lemma 3.3.** Let  $\{\mathbf{Z}_N : N = 1, 2, \dots\}$  be a sequence of  $J \times K$  matrices such that  $\mathbf{Z}_N = o_p(1)$ , and let  $\{\mathbf{x}_N\}$  be a sequence of  $K \times 1$  random vectors such that  $\mathbf{x}_N = O_p(1)$ . Then  $\mathbf{Z}_N \mathbf{x}_N = o_p(1)$ .

**Lemma 3.4** (Slutsky's Theorem). Let  $\mathbf{g} : \mathbb{R}^K \rightarrow \mathbb{R}^J$  be a function continuous at some point  $\mathbf{c} \in \mathbb{R}^K$ . Let  $\{\mathbf{x}_N : N = 1, 2, \dots\}$  be a sequence of  $K \times 1$  vectors such that  $\mathbf{x}_N \xrightarrow{p} \mathbf{c}$ . Then  $\mathbf{g}(\mathbf{x}_N) \xrightarrow{p} \mathbf{g}(\mathbf{c})$  as  $N \rightarrow \infty$ . In other words:

$$\text{plim } \mathbf{g}(\mathbf{x}_N) = \mathbf{g}(\text{plim } \mathbf{x}_N) \quad (3.1)$$

if  $\mathbf{g}(\cdot)$  is continuous at  $\text{plim } \mathbf{x}_N$ .

**Definition 3.5.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A sequence of events:  $\{\Omega_N : N = 1, 2, \dots\} \subset \mathcal{F}$  is said to occur with **probability approaching one (wpa1)** if and only if  $P(\Omega_N) \rightarrow 1$  as  $N \rightarrow \infty$ .

**Corollary 3.1.** Let  $\{\mathbf{Z}_N : N = 1, 2, \dots\}$  be a sequence of random  $K \times K$  matrices, and let  $\mathbf{A}$  be a nonrandom, invertible  $K \times K$  matrix. If  $\mathbf{Z}_N \xrightarrow{p} \mathbf{A}$  then:

1.  $\mathbf{Z}_N^{-1}$  exists with wpa1;
2.  $\mathbf{Z}_N^{-1} \xrightarrow{p} \mathbf{A}^{-1}$  or  $\text{plim } \mathbf{Z}^{-1} = \mathbf{A}^{-1}$  (in an appropriate sense).

### 3.3 Convergence in Distribution

**Definition 3.6.** A sequence of random variables  $\{x_N : N = 1, 2, \dots\}$  **converges in distribution** to the continuous random variable  $x$  if and only if

$$F_N(\xi) \rightarrow F(\xi) \quad \text{as } N \rightarrow \infty \text{ for all } \xi \in \mathbb{R}$$

where  $F_N$  is the cumulative distribution function (cdf) of  $x_N$  and  $F$  is the (continuous) cdf of  $x$ . We write  $x_N \xrightarrow{d} x$ .

*Remark.* When  $x \sim N(\mu, \sigma^2)$  we write  $x_N \xrightarrow{d} N(\mu, \sigma^2)$  or  $x_N \overset{a}{\sim} N(\mu, \sigma^2)$  ( $x_N$  is **asymptotically normal**).

**Definition 3.7.** A sequence of  $K \times 1$  random vectors  $\{\mathbf{x}_N : N = 1, 2, \dots\}$  converges in distribution to the continuous random vector  $\mathbf{x}$  if and only if for any  $K \times 1$  nonrandom vector  $\mathbf{c}$  such that  $\mathbf{c}'\mathbf{c} = 1$ ,  $\mathbf{c}'\mathbf{x}_N \xrightarrow{d} \mathbf{c}'\mathbf{x}$ , and we write  $\mathbf{x}_N \xrightarrow{d} \mathbf{x}$ .

*Remark.* When  $\mathbf{x} \sim N(\mathbf{m}, \mathbf{V})$  the requirement for the definition is that **XXX**

**Lemma 3.5.** If  $\mathbf{x}_N \xrightarrow{d} \mathbf{x}$ , where  $\mathbf{x}$  is any  $K \times 1$  random vector, then  $\mathbf{x}_N = O_p(1)$ .

**Lemma 3.6** (Continuous mapping theorem). Let  $\{\mathbf{x}_N\}$  be a sequence of  $K \times 1$  random vectors such that  $\mathbf{x}_N \xrightarrow{d} \mathbf{x}$ . If  $\mathbf{g} : \mathbb{R}^K \rightarrow \mathbb{R}^J$  is a continuous function, then  $\mathbf{g}(\mathbf{x}_N) \xrightarrow{d} \mathbf{g}(\mathbf{x})$ .

**Corollary 3.2.** If  $\{\mathbf{z}_N\}$  is a sequence of  $K \times 1$  random vectors such that  $\mathbf{z}_N \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$ , then

1. For any  $K \times M$  nonrandom matrix  $\mathbf{A}$ ,  $\mathbf{A}'\mathbf{z}_N \xrightarrow{d} N(\mathbf{0}, \mathbf{A}'\mathbf{V}\mathbf{A})$ .
2.  $\mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \xrightarrow{d} \chi_K^2$  (or  $\mathbf{z}_N'\mathbf{V}^{-1}\mathbf{z}_N \overset{a}{\sim} \chi_K^2$ )

**Lemma 3.7** (Asymptotic Equivalence Lemma). Let  $\{\mathbf{x}_N\}$  and  $\{\mathbf{z}_N\}$  be sequences of  $K \times 1$  random vectors. If  $\mathbf{z}_N \xrightarrow{d} \mathbf{z}$  and  $\mathbf{x}_N - \mathbf{z}_N \xrightarrow{p} \mathbf{0}$ , then  $\mathbf{x}_N \xrightarrow{d} \mathbf{z}$ .

### 3.4 Limit Theorems for Random Samples

**Theorem 3.1 (Weak Law of Large Numbers (WLLN)).** Let  $\{\mathbf{w}_i : i = 1, 2, \dots\}$  be a sequence of independent, identically distributed (iid)  $G \times 1$  random vectors such that  $E(|w_{ig}|) < \infty$ ,  $g = 1, \dots, G$ . Then the sequence satisfies the **weak law of large numbers (WLLN)**:

$$N^{-1} \sum_{i=1}^N \mathbf{w}_i \xrightarrow{p} \boldsymbol{\mu}_w.$$

where  $\boldsymbol{\mu}_w \equiv E(\mathbf{w}_i)$

**Theorem 3.2 (Central Limit Theorem (CLT), Lindberg-Levy).** Let  $\{\mathbf{w}_i : i = 1, 2, \dots\}$  be a sequence of independent, identically distributed (iid)  $G \times 1$  random vectors such that  $E(|w_{ig}|) < \infty$ ,  $g = 1, \dots, G$ , and  $E(\mathbf{w}_i) = \mathbf{0}$ . Then  $\{\mathbf{w}_i : i = 1, 2, \dots\}$  satisfies the **central limit theorem (CLT)**; that is:

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

where  $\mathbf{B} = \text{Var}(\mathbf{w}_i) = E(\mathbf{w}_i\mathbf{w}_i')$  is necessarily positive semidefinite. For our purposes,  $\mathbf{B}$  is almost always positive definite.



### 3.5 Limiting Behavior of Estimators and Test Statistics

#### 3.5.1 Asymptotic Properties of Estimators

**Definition 3.8.** Let  $\{\hat{\theta}_N : N = 1, 2, \dots\}$  be a sequence of estimators of the  $P \times 1$  vector  $\theta \in \Theta$ , where  $N$  indexes the sample size. If

$$\hat{\theta}_N \xrightarrow{p} \theta \quad (3.2)$$

for any value of  $\theta$ , then we say  $\hat{\theta}_N$  is a **consistent estimator** of  $\theta$ .

**Definition 3.9.** Let  $\{\hat{\theta}_N : N = 1, 2, \dots\}$  be a sequence of estimators of the  $P \times 1$  vector  $\theta \in \Theta$ . Suppose that

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(\mathbf{0}, \mathbf{V}) \quad (3.3)$$

where  $\mathbf{V}$  is a  $P \times P$  positive semidefinite matrix. Then we say that  $\hat{\theta}$  is  **$\sqrt{N}$ -asymptotically normally distributed** and  $\mathbf{V}$  is the **asymptotic variance** of  $\sqrt{N}(\hat{\theta}_N - \theta)$  denoted  $\text{Avar}[\sqrt{N}(\hat{\theta}_N - \theta) = \mathbf{V}]$ .

**Definition 3.10.** If  $\sqrt{N}(\hat{\theta}_N - \theta) \overset{a}{\sim} N(\mathbf{0}, \mathbf{V})$  where  $\mathbf{V}$  is positive definite with  $j$ -th diagonal  $v_{jj}$ , and  $\hat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$ , then the **asymptotic standard error** of  $\hat{\theta}_{Nj}$ , denoted  $\text{se}(\hat{\theta}_{Nj})$ , is  $(\hat{v}_{Njj}/N)^{1/2}$ .

**Definition 3.11.** Let  $\hat{\theta}_N$  and  $\tilde{\theta}_N$  be estimators of  $\theta$  each satisfying statement (3.3), with asymptotic variances  $\mathbf{V} = \text{Avar}\sqrt{N}(\hat{\theta}_N - \theta)$  and  $\mathbf{D} = \text{Avar}\sqrt{N}(\tilde{\theta}_N - \theta)$  (these generally depend on the value of  $\theta$ , but we suppress that consideration here).

1.  $\hat{\theta}_N$  is **asymptotically efficient relative to  $\tilde{\theta}_N$**  if  $\mathbf{D} - \mathbf{V}$  is positive semidefinite for all  $\theta$ .
2.  $\hat{\theta}_N$  and  $\tilde{\theta}_N$  are  **$\sqrt{N}$ -equivalent** if  $\sqrt{N}(\hat{\theta}_N - \tilde{\theta}_N) = o_p(1)$

**Definition 3.12.** Partition  $\hat{\theta}_N$  satisfying statement (3.3) into vectors  $\hat{\theta}_{N1}$  and  $\hat{\theta}_{N2}$ . Then  $\hat{\theta}_{N1}$  and  $\hat{\theta}_{N2}$  are **asymptotically independent** if

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 \end{bmatrix}$$

where  $\mathbf{V}_1$  is the asymptotic variance of  $\sqrt{N}(\hat{\theta}_{N1} - \theta_1)$  and similarly for  $\mathbf{V}_2$ . In other words, the asymptotic variance of  $\sqrt{N}(\hat{\theta}_N - \theta)$  is block diagonal.

#### 3.5.2 Asymptotic Properties of Test Statistics

**Definition 3.13.**

1. The **asymptotic size** of a testing procedure is defined as the limiting probability of rejecting  $H_0$  when it is true. Mathematically, we can write this as  $\lim_{N \rightarrow \infty} P_N(\text{reject } H_0 | H_0)$ , where the  $N$  subscript indexes the sample size.
2. A test is said to be **consistent** against alternative  $H_1$  if the null hypothesis is rejected with probability approaching one when  $H_1$  is true:  $\lim_{N \rightarrow \infty} P_N(\text{reject } H_0 | H_1) = 1$ ,

**Lemma 3.8.** Suppose that statement (3.3) holds, where  $\mathbf{V}$  is positive definite. Then for any nonstochastic matrix  $Q \times P$  matrix  $\mathbf{R}$ ,  $Q \leq P$ , with  $\text{rank}(\mathbf{R}) = Q$ ,

$$\sqrt{N}\mathbf{R}(\hat{\theta}_N - \theta) \overset{a}{\sim} N(\mathbf{0}, \mathbf{R}\mathbf{V}\mathbf{R}')$$

and

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

In addition, if  $\text{plim } \hat{\mathbf{V}}_N = \mathbf{V}$ , then

$$\left[ \sqrt{N}\mathbf{R}(\hat{\theta}_N - \theta) \right]' \left[ \mathbf{R}\hat{\mathbf{V}}_N\mathbf{R}' \right]^{-1} \left[ \sqrt{N}\mathbf{R}(\hat{\theta}_N - \theta) \right] = (\hat{\theta}_N - \theta)' \mathbf{R}' \left[ \mathbf{R}(\hat{\mathbf{V}}_N/N)\mathbf{R}' \right]^{-1} \mathbf{R}(\hat{\theta}_N - \theta) \overset{a}{\sim} \chi_Q^2$$

*Remark.* For testing the null hypothesis  $H_0 : \mathbf{R}\boldsymbol{\theta} = \mathbf{r}$ , where  $\mathbf{r}$  is a  $Q \times 1$  nonrandom vector, define the **Wald Statistic** for testing  $H_0$  against  $H_1 : \mathbf{R}\boldsymbol{\theta} \neq \mathbf{r}$  as

$$W_N \equiv (\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r})' \left[ \mathbf{R}(\hat{\mathbf{V}}_N/N)\mathbf{R}' \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\theta}}_N - \mathbf{r}) \quad (3.4)$$

Under  $H_0$ ,  $W_N \stackrel{a}{\sim} \chi_Q^2$ . If we abuse the asymptotics and treat  $\hat{\boldsymbol{\theta}}_N$  as being distributed as  $N(\boldsymbol{\theta}, \hat{\mathbf{V}}_N/N)$ , we get equation (3.4) exactly.

**Lemma 3.9.** Suppose that statement (3.3) holds, where  $\mathbf{V}$  is positive definite. Let  $\mathbf{c} : \boldsymbol{\Theta} \rightarrow \mathbb{R}^Q$  be a continuously differentiable function on the parameter space  $\boldsymbol{\Theta} \subset \mathbb{R}^P$ , where  $Q \leq P$ , and assume that  $\boldsymbol{\theta}$  is in the interior of the parameter space. Define  $\mathbf{C}(\boldsymbol{\theta}) \equiv \nabla_{\boldsymbol{\theta}} \mathbf{c}(\boldsymbol{\theta})$  as the  $Q \times P$  Jacobian of  $\mathbf{c}$ . Then

$$\sqrt{N}[\mathbf{c}(\hat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta})] \stackrel{a}{\sim} N[\mathbf{0}, \mathbf{C}(\boldsymbol{\theta})\mathbf{V}\mathbf{C}(\boldsymbol{\theta})'] \quad (3.5)$$

and

$$\left\{ \sqrt{N}[\mathbf{c}(\hat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta})] \right\}' [\mathbf{C}(\boldsymbol{\theta})\mathbf{V}\mathbf{C}(\boldsymbol{\theta})']^{-1} \left\{ \sqrt{N}[\mathbf{c}(\hat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta})] \right\} \stackrel{a}{\sim} \chi_Q^2$$

Define  $\hat{\mathbf{C}} = \mathbf{C}(\hat{\boldsymbol{\theta}}_N)$ . The  $\text{plim } \hat{\mathbf{C}}_N = \mathbf{C}(\boldsymbol{\theta})$ . If  $\text{plim } \hat{\mathbf{V}}_N = \mathbf{V}$ , then

$$\left\{ \sqrt{N}[\mathbf{c}(\hat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta})] \right\}' \left[ \hat{\mathbf{C}}_N \hat{\mathbf{V}}_N \hat{\mathbf{C}}_N' \right]^{-1} \left\{ \sqrt{N}[\mathbf{c}(\hat{\boldsymbol{\theta}}_N) - \mathbf{c}(\boldsymbol{\theta})] \right\} \stackrel{a}{\sim} \chi_Q^2 \quad (3.6)$$

## 4 The Single-Equation Linear Model and OLS Estimation

Wooldridge (2010, C.4 p.49–76)

- 4.1 Overview of the Single-Equation Linear Model
- 4.2 Asymptotic Properties of OLS
  - 4.2.1. Consistency
  - 4.2.2. Asymptotic Inference Using OLS
  - 4.2.3. Heteroskedasticity-Robust Inference
  - 4.2.4. Lagrange Multiplier (Score) Tests
- 4.3 OLS Solutions to the Omitted Variables Problem
  - 4.31. OLS Ignoring the Omitted Variables
  - 4.32. The Proxy Variable–OLS Solution
  - 4.33. Models with Interactions in Unobservables
- 4.4 Properties of OLS under Measurement Error
  - 4.41. Measurement Error in the Dependent Variable
  - 4.42. Measurement Error in an Explanatory Variable

### 4.1 Modelo de equações lineares

Wooldridge (2010, Sec. 4.1 – Overview of the Single-Equation Linear Model; p.49)

O modelo populacional que estudamos é linear em seus parâmetros,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u \quad (4.1)$$

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.

**Structural Model** Goldberger (1972)<sup>1</sup> defines a **structural model** as one representing a causal relationship, as opposed to a relationship that simply captures statistical associations. A structural equation can be obtained from an economic model, or it can be obtained through informal reasoning. Sometimes the structural model is directly estimable. Other times we must combine auxiliary assumptions about other variables with algebraic manipulations to arrive at an **estimable model**. In addition, we will often have reasons to estimate **nonstructural equations**, sometimes as a precursor to estimating a structural equation.

**Zero Mean Error** The key condition for OLS to consistently estimate the  $\beta_j$  (assuming we have available a random sample from the population) is that the error (in the population) has mean zero and is uncorrelated with each of the regressors:

$$E(u) = 0, \quad \text{Cov}(x_j, u) = 0, \quad j = 1, \dots, K. \quad (4.2)$$

The zero-mean assumption is for free when an intercept is included. It is the zero covariance of  $u$  with each  $x_j$  that is important. From XX we know that equation (4.1) and assumption (4.2) is equivalent to defining the **linear projection** of  $y$  onto  $(1, x_1, \dots, x_K)$  as  $\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K$ .

<sup>1</sup>Goldberger (1972), “Structural Equation Methods in the Social Sciences,” *Econometrica* 40, 979–1001.

Sufficient for assumption (4.2) is the zero conditional mean assumption:

$$E(u|x_1, \dots, x_K) = E(u|\mathbf{x}) = 0. \quad (4.3)$$

Under equation (4.1) and assumption (4.3), we have the **population regression function**

$$E(y|x_1, \dots, x_K) = \beta_0 + \beta_1 x_1 + \dots \beta_K x_K. \quad (4.4)$$

**Endogenous Variable** An explanatory variable  $x_j$  is said to be **endogenous** in equation (4.1) if it is correlated with  $u$ . The usage of the word **endogenous**, in econometrics, is used broadly to describe any situation where an explanatory variable is correlated with the disturbance. If  $x_j$  is uncorrelated with  $u$ , then  $x_j$  is said to be **exogenous** in equation (4.1). If assumption (4.3) holds, then each explanatory variable is necessarily exogenous.

In applied econometrics, endogeneity usually arises in one of three ways:

1. **Ommited Variables**;
2. **Measurement Error**;
3. **Simultaneity**.

## 4.2 Asymptotic Properties of OLS

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

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

$$y = \mathbf{x}\boldsymbol{\beta} + u \quad (4.5)$$

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:

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i. \quad (4.6)$$

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

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (4.7)$$

$\mathbf{y}$  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  $N \times 1$ ;

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

### 4.2.1 Consistency

Wooldridge (2010, Sec. 4.2.1 – Consistency; p.52-4)

#### Assumptions

**OLS.1 population orthogonality condition:**  $E(\mathbf{x}'u) = 0$ ;

**OLS.2** *posto* $[E(\mathbf{x}'\mathbf{x})] = K$ .

Because  $\mathbf{x}$  contains a constant, **OLS.1** is equivalent to saying that  $u$  has zero mean and is uncorrelated with each regressor. Sufficient for **OLS.1** is the zero conditional mean assumption (4.3).

Since  $E(\mathbf{x}'\mathbf{x})$  is symmetric  $K \times K$  matrix, **OLS.2** is equivalent to assuming that  $E(\mathbf{x}'\mathbf{x})$  is positive definite.

Under assumptions **OLS.1** and **OLS.2** the parameter  $\beta$  is **identified**. In the context of models that are linear in the parameters under random sampling, identification of  $\beta$  simply means that  $\beta$  can be written in terms of population moments in observable variables. To see this, we premultiply equation (4.5) by  $\mathbf{x}'$  and take expectations:

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

Because  $(\mathbf{x}, y)$  is observed,  $\beta$  is identified.

**Analogy Principle** The **analogy principle** for choosing an estimator says to turn to the population problem into its **sample counterparts** (Goldberger, 1968; Manski, 1988). In the current application this step leads to the **method of moment**: replace the population moments  $E(\mathbf{x}'\mathbf{x})$  and  $E(\mathbf{x}'y)$  with the corresponding **sample averages**. Doing so leads to the OLS estimator:

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

Isso pode ser escrito na forma matricial:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1} X'y \\ \hat{\beta} &= \beta + (X'X)^{-1} X'u, \end{aligned}$$

onde,

$X$  é a matriz de dados  $N \times K$  dos regressores com linha  $i$  igual a  $\mathbf{x}_i$ ;

$y$  é o vetor de dados  $N \times 1$  com o  $i$ -ésimo elemento de  $y$  sendo representado por  $y_i$ .

Under **OLS.2**  $X'X$  is nonsingular with probability approaching one and

$$\text{plim} \left[ \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \right] = \mathbf{A}^{-1},$$

where,  $\boxed{\mathbf{A} \equiv E(\mathbf{x}'\mathbf{x})}$  (see [Corollary 3.1](#)).

Further, under **OLS.1**

$$\text{plim} \left[ \left( N^{-1} \sum_{i=1}^N \mathbf{x}_i' u_i \right)^{-1} \right] = E(\mathbf{x}'u) = \mathbf{0}.$$

Therefore, by [Slutsky's Theorem](#) ([Lemma 3.4](#)),

$$\text{plim} \hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{A}^{-1} \cdot \mathbf{0} = \boldsymbol{\beta}$$

**Resumo** Resumimos os resultados acima com um teorema:

[Consistência do OLS] Sob as Hipóteses **OLS.1** e **OLS.2**, o estimador de OLS,  $\hat{\boldsymbol{\beta}}$  obtido de uma amostra aleatória seguindo o modelo populacional [\(4.5\)](#) é consistente para  $\boldsymbol{\beta}$ .

Sob as hipóteses do Teorema [4.2.1](#),  $\mathbf{x}\boldsymbol{\beta}$  é uma **projecção linear** de  $y$  em  $\mathbf{x}$ .

#### 4.2.2 Asymptotic Inference Using OLS

[Wooldridge \(2010, Sec. 4.2.2 – Asymptotic INference Using OLS; p.54-5\)](#)

## 5 Estimating Systems of Equations by OLS and GLS

[Wooldridge \(2010, C.7, p.143–179\)](#)

- 7.1. Introduction
- 7.2. Some Examples
- 7.3. System OLS Estimation of a Multivariate Linear System
  - 7.3.1. Preliminaries
  - 7.3.2. Asymptotic Properties of System OLS
  - 7.3.3. Testing Multiple Hypotheses
- 7.4. Consistency and Asymptotic Normality of Generalized Least Squares
  - 7.4.1. Consistency
  - 7.4.2. Asymptotic Normality
- 7.5. Feasible GLS
  - 7.5.1. Asymptotic Properties
  - 7.5.2. Asymptotic Variance of FGLS under a Standard Assumption
- 7.6. Testing Using FGLS
- 7.7. Seemingly Unrelated Regressions, Revisited
  - 7.7.1. Comparison between OLS and FGLS for SUR Systems
  - 7.7.2. Systems with Cross Equation Restrictions
  - 7.7.3. Singular Variance Matrices in SUR Systems
- 7.8. The Linear Panel Data Model, Revisited
  - 7.8.1. Assumptions for Pooled OLS
  - 7.8.2. Dynamic Completeness
  - 7.8.3. A Note on Time Series Persistence
  - 7.8.4. Robust Asymptotic Variance Matrix
  - 7.8.5. Testing for Serial Correlation and Heteroskedasticity after Pooled OLS
  - 7.8.6. Feasible GLS Estimation under Strict Exogeneity

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

### 5.1 Preliminares

[Wooldridge \(2010, Sec.7.3.1\)](#)

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

$X_i$  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:

$$\mathbf{y}_i = X_i \boldsymbol{\beta} + \mathbf{u}_i, \quad i = 1, \dots, N, \quad (5.1)$$

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

## 5.2 Propriedades Assintóticas do SOLS

Wooldridge (2010, Sec.7.3.1)

**SOLS.1**  $E(X_i' \mathbf{u}_i) = 0$ .

**SOLS.2**  $A \equiv E(X_i' X_i)$  é não singular (tem posto pleno, posto igual a  $K$ ).

A hipótese **SOLS.1** é a mais fraca que podemos impor num aracabouço de regressão para conseguirmos um estimador de  $\beta$  consistente. Essa hipótese permite que alguns elementos de  $X_i$  sejam correlacionados com elementos de  $\mathbf{u}_i$ . Uma hipótese mais forte seria:

$$E(\mathbf{u}_i | X_i) = \mathbf{0} \quad (5.2)$$

Sob **SOLS.1**, temos:

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

Para cada  $i$ ,  $X_i \mathbf{y}_i$  é um vetor aleatório  $K \times 1$  e  $X_i' X_i$  é uma matriz  $K \times K$  aleatória simétrica, positiva semidefinida. Então,  $E(X_i' X_i)$  é sempre uma matriz  $K \times K$  não aleatória simétrica, positiva semidefinida. Para conseguirmos estimar  $\beta$  precisamos assumir que ele é o único vetor  $K \times 1$  que satisfaz  $E(X_i' X_i) \beta = E(X_i' \mathbf{y}_i)$ . Por isso assumimos **SOLS.2** e sob **SOLS.1** e **SOLS.2**, podemos escrever  $\beta$  como:

$$\beta = [E(X_i' X_i)]^{-1} E(X_i' \mathbf{y}_i) \quad (5.3)$$

o que mostra que **SOLS.1** e **SOLS.2** identifica o vetor  $\beta$ . O princípio da analogia sugere que estimemos  $\beta$  pelas analogias amostrais de (5.3). Assim, definimos o estimador SOLS de  $\beta$  como:

$$\hat{\beta}^{SOLS} = \left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N X_i' \mathbf{y}_i \right). \quad (5.4)$$

Para computar  $\hat{\beta}$  usando linguagem de computação é mais fácil utilizar a notação matricial

$$\hat{\beta}^{SOLS} = (X' X)^{-1} (X' \mathbf{y}) \quad (5.5)$$

onde

$X \equiv (X_1', \dots, X_N')$  é uma matriz  $NG \times K$  dos  $X_i$  empilhados.

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

**SOLS para SUR** Estimação SOLS para um modelo SUR é equivalente a OLS equação a equação.



### 5.2.1 Consistência

Para provarmos a **consistência** do estimador, usamos a equação (5.4):

$$\begin{aligned}
 \hat{\beta}^{SOLS} &= \left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N X_i' y_i \right) \\
 &= \left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left[ N^{-1} \sum_{i=1}^N X_i' (X_i \beta + u_i) \right] \\
 &= \left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N X_i' X_i \beta \right) + \left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N X_i' u_i \right) \\
 &= \beta + \left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left( N^{-1} \sum_{i=1}^N X_i' u_i \right)
 \end{aligned}$$

Por **SOLS.1**,  $N^{-1} \sum_{i=1}^N X_i' u_i \xrightarrow{p} \mathbf{0}$ ; e por **SOLS.2**  $\left( N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \xrightarrow{p} A^{-1}$ .

Resumimos esse resultado pelo seguinte Teorema:

[Consistência do SOLS] Sob Hipóteses **SOLS.1** e **SOLS.2**, temos

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

### 5.2.2 Normalidade Assintótica

Para fazermos **Inferência**, precisamos achar a variância assintótica do estimador de OLS sob, essencialmente, as mesmas duas hipóteses. Tecnicamente, a seguinte derivação exige os elementos de  $X_i' u_i u_i' X_i$  tenham *finite expected absolute value*. De (5.4) e (5.1), escrevemos:

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

Uma vez que  $E(X_i' u_i) = 0$ , sob a hipótese **SOLS.1**, o CLT implica que:

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

onde

$$B \equiv E(X_i' u_i u_i' X_i) \equiv \text{Var}(X_i' u_i).$$

Em particular,

$$N^{-1/2} \sum_{i=1}^N X_i' u_i = O_p(1).$$

Porém,

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

Sendo Assim,

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

Portanto, com apenas *single-equation OLS* and *2SLS*, obtemos a representação assintótica para  $\sqrt{N}(\hat{\beta} - \beta)$  que é uma combinação linear não aleatória de somas parciais que satisfazem o CLT. Usando o **lema de equivalência assintótica**, temos:

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, A^{-1}BA^{-1})$$

Resumimos esse resultado com o seguinte Teorema:

[Normalidade Assintótica do SOLS] Sob Hipóteses **SOLS.1** e **SOLS.2**, temos que a seguinte equação vale:

$$\boxed{\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, A^{-1}BA^{-1})}. \quad (5.6)$$

### 5.2.3 Variância Assintótica

A variância assintótica de  $\hat{\beta}^{SOLS}$  é:

$$\text{Avar}(\hat{\beta}^{SOLS}) = A^{-1}BA^{-1}/N. \quad (5.7)$$

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

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

Um estimador consistente para  $B$  pode ser achado usando o princípio da analogia.

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

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

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

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

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

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

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

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

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

## 6 SOLS para Dados de Painei

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

No caso de dados de painei:

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

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

$$\hat{\beta}^{POLS} = \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}' \mathbf{x}_{it} \right)^{-1} \left( \sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}' y_{it} \right). \quad (6.1)$$

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 (6.1) é o mesmo para unidades de *cross section* amostradas em diferentes pontos do tempo. O Teorema 5.2.1, abaixo, mostraa que o estimador POLS é consistente sob as condições de ortogonalidade na hipótese **XX** e uma hipótese de posto completo.

$$\begin{aligned} &E(\mathbf{x}_{it}) \\ X_i' \mathbf{u}_i &= \sum_{t=1}^T \mathbf{x}_{it}' u_{it} \end{aligned}$$

## 7 System GLS (SGLS)

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

### Hipóteses

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

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

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

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

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

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

### Estimação

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

$$\begin{aligned}\Omega^{-1/2} \mathbf{y}_i &= \Omega^{-1/2} X_i \boldsymbol{\beta} + \Omega^{-1/2} \mathbf{u}_i \\ \mathbf{y}_i^* &= X_i^* \boldsymbol{\beta} + \mathbf{u}_i^*\end{aligned}$$

Estimando a equação acima por **SOLS**:

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

## 8 GLS Factível

[Wooldridge \(2010, Sec.7.5 – Feasible GLS, p.153\)](#)

### 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  $\mathbf{y}_i = X_i\beta + \mathbf{u}_i$  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  $\beta^{SGLS}$ :

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

Empilhando as  $N$  observações:

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

Reescrevendo a equação acima:

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

### Valor Esperado

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

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

### Variância

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

Tirando o valor Esperado e supondo que:

$$E(X_i \Omega^{-1} u_i u_i' X_i) = E(X_i \Omega^{-1})$$

temos:

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

e temos:

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

## 9 Basic Linear Unobserved Effects Panel Data Models

[Wooldridge \(2010, C.10, pp.247–291 \)](#)

10.1. Motivation: The Omitted Variables Problem

10.2. Assumptions about the Unobserved Effects and Explanatory Variables

10.2.1. Random of Fixed Effects?

10.2.2. Strict Exogeneity Assumptions on the Explanatory Variables

10.2.3. Some Examples of Unobserved Effects Panel Data Models

10.3. Estimating Unobserved Effects Models by Pooled OLS

10.4. **Random Effects Methods**

10.4.1. Estimation and Inference under the Basic Random Effects Assumptions

10.4.2. Robust Variance Matrix Estimator

10.4.3. A General FGLS Analysis

10.4.4. Testing for the Presence of an Unobserved Effect

10.5. **Fixed Effects Methods**

10.5.1. Consistency of the Fixed Effects Estimator

10.5.2. Asymptotic Inference with Fixed Effects

10.5.3. The Dummy Variable Regression

10.5.4. Serial Correlation and the Robust Variance Matrix Estimator

10.5.5. Fixed Effects GLS

10.5.6. Using Fixed Effects Estimation for Policy Analysis

10.6. **First Differencing Methods**

10.6.1. Inference

10.6.2. Robust Variance Matrix

10.6.3. Testing for Serial Correlation

10.6.4. Policy Analysis Using First Differencing

10.7. Comparison of Estimators

10.7.1. Fixed Effects versus First Differencing

10.7.2. The Relationship between the Random Effects and Fixed Effects Estimators

10.7.3. The Hausman Test Comparing the RE and the FE Estimators

## 10 Endogeneity and GMM

### Modelo

No seguinte modelo *cross-section*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i ; \quad i = 1, \dots, N. \quad (10.1)$$

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 (10.1) vamos supor:

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

### 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 (10.1).
2.  $Cov(z, x_2) \neq 0 \implies$  correlação com  $x_2$  após controlar para outras variáveis.

### Estimação

Indo para o problema de dados de painel, temos:

$$\mathbf{y}_i = X_i \boldsymbol{\beta} + \mathbf{u}_i ; \quad i = 1, \dots, N. \quad (10.2)$$

onde  $\mathbf{y}_i$  é um vetor  $T \times 1$ ,  $X_i$  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 (10.2), então:

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

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).

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

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



Se Usarmos estimadores amostrais:

$$\hat{\beta} = \left[ N^{-1} \sum_{i=1}^N Z_i' X_i \right]^{-1} \left[ N^{-1} \sum_{i=1}^N Z_i' y_i \right]$$

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

Se  $L > K$ , vamos considerar:

$$\min_{\beta} E(Z_i u_i)^2$$

onde:

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

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

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

Um estimador mais eficiente pode ser encontrado fazendo:

$$\min_{\beta} E[(Z_i' y - Z_i' X \beta)' W (Z_i' y - Z_i' X \beta)].$$

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

$$\min_{\beta} \left\{ y'Z\widehat{W}Z'y - y'Z\widehat{W}Z'X\beta - \beta'X'Z\widehat{W}Z'y + \beta'X'Z\widehat{W}Z'X\beta \right\}$$

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

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

**Valor Esperado**

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

**Variância**

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

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

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

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

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

## 11 Random Effects (RE, EA)

### 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:

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

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

### Hipóteses de $\hat{\boldsymbol{\beta}}^{RE}$

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  $E(X_i'\Omega^{-1}X_i)$ .

Definindo a matriz  $T \times T$ ,  $\boxed{\Omega \equiv E(\mathbf{v}_i\mathbf{v}_i')}$ , queremos que  $E(X_i\Omega^{-1}X_i)$  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  $\Omega^{-1} = \Omega^{-1/2}\Omega^{-1/2}$ .

### Estimação

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

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

Estimando o modelo acima por POLS:

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

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,

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

onde  $\sigma_u^2 I_T$  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:

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

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

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

**Valor Esperado**

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

**Variância**

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

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

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

## 12 Fixed Effects (EF, FE)

### Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (12.1)$$

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 não observado. No caso da análise de **Efeitos Fixos (EF, FE)**, permitimos que esse componente  $c_i$  seja correlacionado com  $\mathbf{x}_{it}$ . Assim, se decidíssemos estimar o modelo (12.1) por POLS, ignorando  $c_i$ , teríamos problemas de inconsistência devido a **endogeneidade**.

As  $T$  equações do modelo (12.1) podem ser reescritas como:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + c_i\mathbf{1}_T + \mathbf{u}_i, \quad (12.2)$$

com  $\mathbf{v}_i = c_i\mathbf{1}_T + \mathbf{u}_i$  sendo os erros compostos.

### Matriz $M^0$

Definimos a matriz  $M^0$  como:

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

A matriz  $M^0$  é idempotente e simétrica.

$$M^0\mathbf{x} = \mathbf{x} - \bar{\mathbf{x}}\mathbf{1}_T = \ddot{\mathbf{x}}.$$

Podemos transformar o modelo (12.3) ao premultiplicarmos todo o modelo por  $M^0$ .

$$M^0\mathbf{y}_i = M^0X_i\boldsymbol{\beta} + M^0(c_i\mathbf{1}_T) + M^0\mathbf{u}_i, \quad i = 1, \dots, N.$$

$$M^0(c_i\mathbf{1}_T) = (I_T - T^{-1}\mathbf{1}_T\mathbf{1}_T')c_i\mathbf{1}_T = c_i\mathbf{1}_T - T^{-1}c_i\mathbf{1}_T\mathbf{1}_T'\mathbf{1}_T = c_i\mathbf{1}_T - c_i\mathbf{1}_T \implies \boxed{M^0(c_i\mathbf{1}_T) = 0}$$

$$\ddot{\mathbf{y}}_i = \ddot{X}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i, \quad i = 1, \dots, N. \quad (12.3)$$

### Estimação POLS

Aplicando POLS no modelo (12.3)

$$\boxed{\boldsymbol{\beta}^{FE} = \left[ \sum_{i=1}^N \ddot{X}_i'\ddot{X}_i \right]^{-1} \left[ \sum_{i=1}^N \ddot{X}_i'\ddot{\mathbf{y}}_i \right]} \quad (12.4)$$

### Hipóteses

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

**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 completo de  $E(X_i'\Omega^{-1}X_i)$  (para inverter a matriz).  $\text{posto}[E(X_i'\Omega^{-1}X_i)] = K$ .

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

## Valor Esperado

Usando **FE.1** e **FE.2**, apenas.

$$E(\beta^{FE}) = \beta + E \left[ \left( \sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left( \sum_{i=1}^N \ddot{X}_i' \ddot{u}_i \right) \right]$$

$$E(\beta^{FE}) = \beta + E \left[ (\ddot{X}' \ddot{X})^{-1} (\ddot{X}' \ddot{u}) \right]$$

Sabendo que  $\ddot{X} = (I_N \otimes M^0)X$  e  $\ddot{u} = (I_N \otimes M^0)u$ , definimos:

$$E(\beta^{FE}) = \beta + E \left\{ [X'(I_N \otimes M^0)(I_N \otimes M^0)X]^{-1} [X'(I_N \otimes M^0)(I_N \otimes M^0)u] \right\}$$

$$E(\beta^{FE}) = \beta + E \left\{ [X'(I_N \otimes M^0)X]^{-1} [X'(I_N \otimes M^0)u] \right\}$$

## Variância

Usamos a variância do estimador para inferência. Usando **FE.1** e **FE.2**, apenas:

$$\text{Var}(\beta^{FE}) = E \left[ (\ddot{X}' \ddot{X})^{-1} (\ddot{X}' \ddot{u})(\ddot{u}' \ddot{X})(\ddot{X}' \ddot{X})^{-1} \right]$$

**Pão:**

$$\begin{aligned} E \left[ (\ddot{X}' \ddot{X})^{-1} \right] &= E \left\{ [X'(I_N \otimes M^0)(I_N \otimes M^0)X]^{-1} \right\} \\ &= E \left\{ [X'(I_N \otimes M^0)X]^{-1} \right\} \end{aligned}$$

**Recheio:**

$$\begin{aligned} E \left[ (\ddot{X}' \ddot{u})(\ddot{u}' \ddot{X}) \right] &= E \left[ X'(I_N \otimes M^0)(I_N \otimes M^0)uu'(I_N \otimes M^0)(I_N \otimes M^0)X \right] \\ &= E \left[ X'(I_N \otimes M^0)uu'(I_N \otimes M^0)X \right] \end{aligned}$$

$$\text{Var}(\beta^{FE}) = \text{Pão Recheio Pão}$$

$$\text{Var}(\beta^{FE}) = E \left\{ [X'(I_N \otimes M^0)X]^{-1} \right\} E \left[ X'(I_N \otimes M^0)uu'(I_N \otimes M^0)X \right] E \left\{ [X'(I_N \otimes M^0)X]^{-1} \right\}$$

## Variância sob Homocedasticidade

Usando **FE.3**, temos

**Recheio':**

$$E \left[ X'(I_N \otimes M^0) \right] \sigma_u^2 I_{NT} E \left[ (I_N \otimes M^0)X \right] = \sigma_u^2 E \left[ X'(I_N \otimes M^0)X \right]$$

$(I_N \otimes M^0)$  é uma matrix de dimensão  $NT \times NT$ , visto que  $I_N$  é  $N \times N$  e  $M^0$  é  $T \times T$ .

$$\text{Var}(\beta^{FE}) = \text{Pão Recheio' Pão}$$

$$= E \left\{ [X'(I_N \otimes M^0)X]^{-1} \right\} \sigma_u^2 E \left[ X'(I_N \otimes M^0)X \right] E \left\{ [X'(I_N \otimes M^0)X]^{-1} \right\}$$

$$= E \left\{ [X'(I_N \otimes M^0)X]^{-1} \right\} \sigma_u^2 I_{NT}$$

$$\text{Var}(\beta^{FE}) = \sigma_u^2 \cdot E \left[ X'(I_N \otimes M^0)X \right]$$

## 13 First Difference (FD, PD)

### Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (13.1)$$

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 (13.1) 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 (13.1) defasado:

$$y_{it-1} = \mathbf{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \quad (13.2)$$

- Tiramos a diferença entre (13.2) e (13.1):

$$\begin{aligned} y_{it} - y_{it-1} &= (\mathbf{x}_{it} - \mathbf{x}_{it-1})\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{it-1} \\ \Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}. \end{aligned} \quad (13.3)$$

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

Reescrevendo (13.3) no formato matricial empilhando  $T$ :

$$\Delta \mathbf{y}_i = \Delta \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \quad (13.4)$$

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  $(T-1) \times 1$

### Estimação POLS

O estimador  $\hat{\boldsymbol{\beta}}^{FD}$  é o POLS da regressão no modelo (13.4), assim:

$$\boxed{\boldsymbol{\beta}^{FD} = \left[ \sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right]^{-1} \left[ \sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{y}_i \right]} \quad (13.5)$$

### Hipóteses

As Hipóteses que usamos para  $\hat{\boldsymbol{\beta}}^{FD}$  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). *posto* $[E(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i)] = K$ .

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



**Valor Esperado**

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

$$E(\beta^{FD}) = \beta + E \left[ \left( \sum_{i=1}^N \Delta X_i' \Delta X_i \right)^{-1} \left( \sum_{i=1}^N \Delta X_i' e_i \right) \right]$$

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

**Variância**

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

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

**Variância sob Homocedasticidade**

Usando **FD.3**, temos

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

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

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.

## 14 Exogeneidade Estrita e FDIV

### 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:

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

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

(14.1)

### Estimação

Não podemos estimar o modelo (14.1) 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  $y_{it-2}, y_{it-3}, \dots, y_{i1}$ ).

**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  $Z$ .

**FDIV**

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

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  $(T-1) \times L$

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

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

**Estimação FDIV**

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

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

**Valor Esperado**

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

**Variância**

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

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

## 15 Latent Variables, Probit and Logit

### Modelo

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

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i. \quad (15.1)$$

Defina  $y$  como:

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

temos que:

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

Além disso, pela definição de  $y_i$ , equação (15.1), 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  $i$ :

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

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

$$\text{Max}_{\beta \in B} E[\ell_i(\beta)].$$

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:

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

Onde

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

Assim, temos que:

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

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

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

## 16 ATT, ATE, Propensity Score

### 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).

**ATE e ATT condicional a variáveis  $\mathbf{x}$**

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

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

OBS:

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

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

### 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.

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

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  $\mathbf{x}$ ,  $P(w = 1 | \mathbf{x}) < 1$ .

**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:**

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

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

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

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

**TODO**

## References

WOOLDRIDGE, JEFFREY M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2 edn. Boston, Massachusetts: MIT Press.