Microeconometrics

Microeconometrics: Lecture Notes Wooldridge (2010)

Autor: Paulo Ferreira Naibert

Porto Alegre
04/07/2020
Revisão: July 6, 2020

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data.* 2ED. Boston, Massachussetts: 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. Simulteneous 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 Molds
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 Let $a_1(\boldsymbol{x}), \ldots, a_G(\boldsymbol{x})$ and $b(\boldsymbol{x})$ be scalar functions of $\boldsymbol{x}$, and let $y_1, \ldots, y_G$ be random scalars. Then

$$\mathrm{E}\left(\sum_{j=1}^{G} a_j(\boldsymbol{x})y_j + b(\boldsymbol{x})|\boldsymbol{x}\right) = \sum_{j=1}^{G} a_j(\boldsymbol{x})\mathrm{E}(y_j|\boldsymbol{x}) + b(\boldsymbol{x})$$

provided that $\mathrm{E}(|y_j|) < \infty$ $\mathrm{E}(|a_j(\boldsymbol{x})y_j|) < \infty$, and $\mathrm{E}(|b(\boldsymbol{x})|) < \infty$. This is the sense in which the conditional expectations is a **linear operator**.

CE.2 $\mathrm{E}(y) = \mathrm{E}[\mathrm{E}(y|\boldsymbol{x})] \equiv \mathrm{E}[\mu(\boldsymbol{x})]$.

CE.3

1. $\mathrm{E}(y|\boldsymbol{x}) = \mathrm{E}[\mathrm{E}(y|)|\boldsymbol{x}]$, where $\boldsymbol{x}$ and are vectors with $\boldsymbol{x} = ()$ for some nonstochastic function $(\cdot)$. (This is the genral version of the **law of iterated expectations**.)
2. A special case of part 1, $\mathrm{E}(y|\boldsymbol{x}) = \mathrm{E}[\mathrm{E}(y|\boldsymbol{x},)|\boldsymbol{x}]$ for vectors $\boldsymbol{x}$ and .

CE.4 If $(\boldsymbol{x}) \in \mathbb{R}^J$ is a function of $\boldsymbol{x}$ such that $\mathrm{E}(y|\boldsymbol{x}) = g[(\boldsymbol{x})]$ for some scalar function $g(\cdot)$, then $\mathrm{E}[y|(\boldsymbol{x})] = E(y|\boldsymbol{x})$.

CE.5 If the vector $(\boldsymbol{u},)$ is independent of the vector $\boldsymbol{x}$, then $\mathrm{E}(\boldsymbol{u}|\boldsymbol{x},) = \mathrm{E}(\boldsymbol{u}|)$.

CE.6 If $u \equiv= y - \mathrm{E}(y|\boldsymbol{x})$, then $\mathrm{E}[(\boldsymbol{x})|u] = \mathbf{0}$ for any function $(\boldsymbol{x})$, provided that $\mathrm{E}[|g_j(\boldsymbol{x})u|] < \infty$, $j = 1, \ldots, J$, and $\mathrm{E}(|u|) < \infty$. In particular, $\mathrm{E}(u) = 0$ and $\mathrm{Cov}(x_j, u) = 0$, $j = 1, \ldots, K$.

CE.7 (Conditional Jensen's Inequality): If $c : \mathbb{R} \to \mathbb{R}$ is a convex function defined on $\mathbb{R}$ and $E(|y|) < \infty$, then

$c[E(y|\boldsymbol{x})] \leq E[c(y)|\boldsymbol{x}]$.

CE.8 If $E(y^2) < \infty$ and $\mu(\boldsymbol{x}) \equiv E(y|\boldsymbol{x})$, then μ is a solution to (argmin)

$\min_{m \in \mathcal{M}} E[(y - m(\boldsymbol{x}))^2]$

where $\mathcal{M}$ is the set of functions $m : \mathbb{R}^K \to \mathbb{R}$ such that $E[m(\boldsymbol{x})] < \infty$. In other words, $\mu(\boldsymbol{x})$ is the best mean square predictor of y based on information contained in $\boldsymbol{x}$.

2.4.2 Properties of Conditional Variances

The **conditional variance** of y given $\boldsymbol{x}$ is defined as

$\mathrm{Var}(y|\boldsymbol{x}) \equiv \sigma^2(\boldsymbol{x}) \equiv \mathrm{E}\left[\{y - E(y|\boldsymbol{x})\}^2|\boldsymbol{x}\right] = \mathrm{E}(y^2|\boldsymbol{x}) - [E(y|\boldsymbol{x})]^2$

CV.1 $\mathrm{Var}[a(\boldsymbol{x})y + b(\boldsymbol{x})|\boldsymbol{x}] = [a(\boldsymbol{x})]^2\mathrm{Var}(y|\boldsymbol{x})$.

CV.2 $\mathrm{Var}(y) = \mathrm{E}[\mathrm{Var}(y|\boldsymbol{x})] + \mathrm{Var}[\mathrm{E}(y|\boldsymbol{x})] = E[\sigma^2(\boldsymbol{x})] + \mathrm{Var}[\mu(\boldsymbol{x})]$.

CV.3 $\mathrm{Var}(y|\boldsymbol{x}) = \mathrm{E}[\mathrm{Var}(y|\boldsymbol{x},)|\boldsymbol{x}] + \mathrm{Var}[E(y|\boldsymbol{x},)|\boldsymbol{x}]$

CV.4 $\mathrm{E}[\mathrm{Var}(y|\boldsymbol{x})] \geq \mathrm{E}[\mathrm{Var}(y|\boldsymbol{x},)]$

Remark. For any function $m(\cdot)$ define the **mean squared error** as

$\mathrm{MSE}(y; m) \equiv \mathrm{E}[(y - m(\boldsymbol{x}))^2]$

Then **CV.4** can be loosely stated as

$\mathrm{MSE}[y; \mathrm{E}(y|\boldsymbol{x})] \geq \mathrm{MSE}[y; \mathrm{E}(y|\boldsymbol{x},)]$.

In other words, in the population one never does worse for predicting y when additional variables are conditioned on. In particular, if $\mathrm{Var}(y|\boldsymbol{x})$ and $\mathrm{Var}(y|\boldsymbol{x},)$ are both constant, then $\mathrm{Var}(y|\boldsymbol{x}) \geq \mathrm{Var}(y|\boldsymbol{x},)$.

2.4.3 Properties of Linear Projections

In what follow, y is a scalar, $\boldsymbol{x}$ is a $1 \times K$ vector, and is a $1 \times J$ vector. We allow the first element of $\boldsymbol{x}$ to be unity (regression with constant), although the following properties hold in either case. All of the variable are assumed to have finite second moments, and the apropriate variance matrices are assumed to be nonsingular.

LP.1 If $\mathrm{E}(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}$, then $\mathrm{L}(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}$. More generally, if

$$\mathrm{E}(y|\boldsymbol{x}) = \beta_1 g_1(\boldsymbol{x}) + \cdots + \beta_M g_M(\boldsymbol{x})$$

then

$$\mathrm{L}(y|w_1, \ldots, w_M) = \beta_1 w_1 + \cdots + \beta_M w_M$$

where $w_j \equiv g_j(\boldsymbol{x})$, $j = 1, \ldots, M$. This property tells us that, if $\mathrm{E}(y|\boldsymbol{x})$ is known to be linear in some functions $g_j(\boldsymbol{x})$, then this linear function also represents a linear projection.

LP.2 Define $u \equiv y - \mathrm{L}(y|\boldsymbol{x}) = y - \boldsymbol{x}\boldsymbol{\beta}$. Then $E(\boldsymbol{x}'u) = \mathbf{0}$.

LP.3 Suppose y_j, $j = 1, \ldots, G$ are each random scalars, and $a_1, \ldots, a_G$ are constants. Then

$$\mathrm{L}\left(\sum_{j=1}^{G} a_j y_j | \boldsymbol{x}\right) = \sum_{j=1}^{G} a_j \mathrm{L}(y_j|\boldsymbol{x})$$

Thus, the linear projection is a linear operator.

LP.4 (Law of Iterated Projections): $\mathrm{L}(y|\boldsymbol{x}) = \mathrm{L}[\mathrm{L}(y|\boldsymbol{x},)|\boldsymbol{x}]$. More precisely, let

$$\mathrm{L}(y|\boldsymbol{x}, \mathbf{z}) \equiv \boldsymbol{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\gamma} \quad \text{and} \quad \mathrm{L}(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\delta} \tag{2.1}$$

For each element of $\mathbf{z}$, write $\mathrm{L}(z_j|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\pi}_j$, $j = 1, \ldots, J$, where $\boldsymbol{\pi}_j$ is $K \times 1$. Then $\mathrm{L}(\mathbf{z}|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\Pi}$ where $\boldsymbol{\Pi}$ is the $K \times J$ matrix $\boldsymbol{\Pi} \equiv (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_J)$. Property **LP.4** implies that

$$\begin{aligned}\mathrm{L}(y|\boldsymbol{x}) = \mathrm{L}(\boldsymbol{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\gamma}|\boldsymbol{x}) &= \mathrm{L}(\boldsymbol{x}|\boldsymbol{x})\boldsymbol{\beta} + \mathrm{L}(\mathbf{z}|\boldsymbol{x})\boldsymbol{\gamma} \\ &= \boldsymbol{x}\boldsymbol{\beta} + (\boldsymbol{x}\boldsymbol{\Pi})\boldsymbol{\gamma} = \boldsymbol{x}(\boldsymbol{\beta} + \boldsymbol{\Pi})\boldsymbol{\gamma}.\end{aligned} \tag{2.2}$$

Thus, we have shown that $\boldsymbol{\delta} = \boldsymbol{\beta} + \boldsymbol{\Pi}\boldsymbol{\gamma}$. This is, in fact, the population analogue of the**omitted variable bias** formula from the standar regression theory.

LP.5 $\mathrm{L}(y|\boldsymbol{x}) = \mathrm{L}[\mathrm{E}(y|\boldsymbol{x},)|\boldsymbol{x}]$

Proof. ■

LP.6 $\boldsymbol{\beta}$ is a solution to

$$\min_{\in \mathbb{R}^K} \mathrm{E}[(y - \boldsymbol{x}\mathbf{b})^2] \tag{2.3}$$

Proof. For any $\mathbf{b}$, write $y - \boldsymbol{x}\mathbf{b} = (y - \boldsymbol{x}\boldsymbol{\beta}) + (\boldsymbol{x}\boldsymbol{\beta} - \boldsymbol{x}\mathbf{b})$. Then

$$\begin{aligned}(y - \boldsymbol{x}\mathbf{b})^2 &= (y - \boldsymbol{x}\boldsymbol{\beta})^2 + (\boldsymbol{x}\boldsymbol{\beta} - \boldsymbol{x}\mathbf{b})^2 + 2(y - \boldsymbol{x}\boldsymbol{\beta})(\boldsymbol{x}\boldsymbol{\beta} - \boldsymbol{x}\mathbf{b}) \\ &= (y - \boldsymbol{x}\boldsymbol{\beta})^2 + (\boldsymbol{\beta} - \mathbf{b})'\boldsymbol{x}'\boldsymbol{x}(\boldsymbol{\beta} - \mathbf{b}) + 2(\boldsymbol{\beta} - \mathbf{b})'\boldsymbol{x}'(y - \boldsymbol{x}\boldsymbol{\beta})\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{E}[(y - \boldsymbol{x}\mathbf{b})^2] &= \mathrm{E}[(y - \boldsymbol{x}\boldsymbol{\beta})^2] + (\boldsymbol{\beta} - \mathbf{b})'\mathrm{E}(\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{\beta} - \mathbf{b}) + 2(\boldsymbol{\beta} - \mathbf{b})'\mathrm{E}[\boldsymbol{x}'(y - \boldsymbol{x}\boldsymbol{\beta})] \\
&= \mathrm{E}[(y - \boldsymbol{x}\boldsymbol{\beta})^2] + (\boldsymbol{\beta} - \mathbf{b})'\mathrm{E}(\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{\beta} - \mathbf{b}) \qquad (2.4)
\end{aligned}$$

Because $\mathrm{E}[\boldsymbol{x}'(y-\boldsymbol{x}\boldsymbol{\beta})] = \mathbf{0}$ by **LP.2**. When $\mathbf{b} = \boldsymbol{\beta}$, on the RHS of equation (2.4) is minimized. Further, if $\mathrm{E}(\boldsymbol{x}'\boldsymbol{x})$ is Positive Definite, $(\boldsymbol{\beta} - \mathbf{b})'E(\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{\beta} - \mathbf{b}) > 0$ if $\mathbf{b} \neq \boldsymbol{\beta}$; so in this case $\boldsymbol{\beta}$ is the unique minimizer. ■

Remark. Property **LP.6** states that the linear projection is the minimum mean square *linear* predictor. It is not necessarily the minimum square predictor: if $\mathrm{E}(y|\boldsymbol{x}) = \mu(\boldsymbol{x})$ is not linear in $\boldsymbol{x}$, then

$$\mathrm{E}[(y - \mu(\boldsymbol{x}))^2] < \mathrm{E}[(y - \boldsymbol{x}\boldsymbol{\beta})^2] \qquad (2.5)$$

LP.7 This is a partioned projection formula, which is useful in a variety of circumstances. Write

$$\mathrm{L}(y|\boldsymbol{x}, \mathbf{z}) = \boldsymbol{x}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\gamma} \qquad (2.6)$$

Define the $1 \times K$ vector of population residuals from the projection of $\boldsymbol{x}$ on $\mathbf{z}$ as $\mathbf{r} \equiv \boldsymbol{x} - \mathrm{L}(\boldsymbol{x}|\mathbf{z})$. Further, define the population residual from the projection of y on $\mathbf{z}$ as $v \equiv y - L(y|\mathbf{z})$. Then the following are true:

$$\mathrm{L}(v|\mathbf{r}) = \mathbf{r}\boldsymbol{\beta} \qquad (2.7)$$

and

$$\mathrm{L}(y|\mathbf{r}) = \mathbf{r}\boldsymbol{\beta} \qquad (2.8)$$

Remark. The point is that the $\boldsymbol{\beta}$ in equation (2.7) and (2.8) is the *same* as that appearing in equation (2.6). Another way of stating this result is

$$\boldsymbol{\beta} = [\mathrm{E}(\mathbf{r}'\mathbf{r})]^{-1}\mathrm{E}(\mathbf{r}'v) = [\mathrm{E}(\mathbf{r}'\mathbf{r})]^{-1}E(\mathbf{r}'y). \qquad (2.9)$$

Proof. ■

3 Basic Aymptotic 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 $\{a_N; N = 1, 2, \ldots\}$ **converges** to a (has limit a) if for all $\varepsilon > 0$ there exists N_ε such that if $N > N_\varepsilon$, then $|a_N - a| < \varepsilon$. We write $a_N \to a$ as $N \to +\infty$

2. A sequence $\{a_N; N = 1, 2, \ldots\}$ is **bounded** if and only if there is some $b < \infty$ such that $|a_N| < b$ for all $N = 1, 2, \ldots$. Otherwise, we say that $\{a_N\}$ is **unbounded**.

Definition 3.2.

1. A sequence $\{a_N\}$ is $\mathrm{O}(N^\lambda)$ (*at most of order* N^λ) if $N^{-\lambda}a_N$ is bounded. When $\lambda = 0$, $\{a_N\}$ is bounded, and we also write $a_N = \mathrm{O}(1)$ *(big oh one)*.

2. A sequence $\{a_N\}$ is $\mathrm{o}(N^\lambda)$ if $N^{-\lambda}a_N \to 0$. When $= 0$, a_N converges to zero, and we also write $a_N = \mathrm{o}(1)$ *(little oh one)*.

3.2 Convergence in Probability and Bounded in Probability

Definition 3.3.

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

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

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

2. In the special case where $a = 0$ ($\operatorname{plim} x_N = 0$), we also say that $\{x_N\}$ is $\mathrm{o}_p(1)$ (*little oh p one*). We also write $x_N = \mathrm{o}_p(1)$ 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_ε such that

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

   We write $x_N = \mathrm{O}_p(1)$ (x_N is *big oh p one*).

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

Definition 3.4.

1. A random sequence $\{x_N : N = 1, 2, \ldots\}$ is $\mathrm{o}_p(a_N)$, where $\{a_N\}$ is a nonrandom, positive sequence if $x_N/a_N = \mathrm{o}_p(1)$. We write $x_N = \mathrm{o}_p(a_n)$.

2. A random sequence $\{x_N : N = 1, 2, \ldots\}$ is $\mathrm{O}_p(a_N)$, where $\{a_N\}$ is a nonrandom, positive sequence if $x_N/a_N = \mathrm{O}_p(1)$. We write $x_N = \mathrm{O}_p(a_n)$.

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

Lemma 3.2. If $w_N = \mathrm{o}_p(1)$, $x_N = \mathrm{o}_p(1)$, $y_N = \mathrm{O}_p(1)$, and $z_N = \mathrm{O}_p(1)$, then

1. $w_N + x_N = \mathrm{o}_p(1)$
2. $y_N + z_N = \mathrm{O}_p(1)$
3. $y_N z_N = \mathrm{O}_p(1)$
4. $x_N z_N = \mathrm{o}_p(1)$

Remark.

1. $\mathrm{o}_p(1) + \mathrm{o}_p(1) = \mathrm{o}_p(1)$
2. $\mathrm{O}_p(1) + \mathrm{O}_p(1) = \mathrm{O}_p(1)$
3. $\mathrm{O}_p(1)\mathrm{O}_p(1) = \mathrm{O}_p(1)$
4. $\mathrm{o}_p(1)\mathrm{O}_p(1) = \mathrm{o}_p(1)$

Because a $\mathrm{o}_p(1)$ sequence is also $\mathrm{O}_p(1)$, we also have:

5. $\mathrm{o}_p(1) + \mathrm{O}_p(1) = \mathrm{O}_p(1)$
6. $\mathrm{o}_p(1)\mathrm{o}_p(1) = \mathrm{o}_p(1)$

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

Lemma 3.3. Let $\{\mathbf{Z}_N : N = 1, 2, \ldots\}$ be a sequence of $J \times K$ matrices such that $\mathbf{Z}_N = \mathrm{o}_p(1)$, and let $\{\boldsymbol{x}\}_N$ be a sequence of $K \times 1$ random vectors such that $\boldsymbol{x}_N = \mathrm{O}_p(1)$. Then $\mathbf{Z}_N\boldsymbol{x}_N = \mathrm{o}_p(1)$.

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

$$\operatorname{plim} \boldsymbol{g}(\boldsymbol{x}_N) = \mathbf{g}(\operatorname{plim} \boldsymbol{x}_N) \tag{3.1}$$

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

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

Corollary 3.1. Let $\{\mathbf{Z}_N : N = 1, 2, \ldots\}$ 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 $\operatorname{plim} \mathbf{Z}^{-1} = \mathbf{A}^{-1}$ (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

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

where F_N is the cumulative distribution function (cdf) of x_N and F is the (countinuous) 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 $\{\boldsymbol{x}_N : N = 1, 2, \dots\}$ converges in distribution to the continuous randoom vector $\boldsymbol{x}$ if and only if for any $K \times 1$ nonrandom vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{c} = 1$, $\mathbf{c}'\boldsymbol{x}_N \xrightarrow{d} \mathbf{c}\boldsymbol{x}$, and we write $\boldsymbol{x}_N \xrightarrow{d} \boldsymbol{x}$.

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

Lemma 3.5. If $\boldsymbol{x}_N \xrightarrow{d} \boldsymbol{x}$, where $\boldsymbol{x}$ is any $K \times 1$ random vector, then $\boldsymbol{x}_N = \mathrm{O}_p(1)$.

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

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

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

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

3.4 Limit Theorems for Random Samples

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

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

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

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

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

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

3.5 Limiting Behavior of Estimators and Test Statistics

3.5.1 Asymptotic Properties of Estimators

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

$$\hat{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta} \tag{3.2}$$

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

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

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

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

Definition 3.10. If $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \overset{a}{\sim} N(\mathbf{0}, \mathbf{V})$ where $\mathbf{V}$ is positive definite with j-th diagonal v_{jj}, and $\widehat{\mathbf{V}} \xrightarrow{p} \mathbf{V}$, 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 $\hat{\boldsymbol{\theta}}_N$ and $\tilde{\boldsymbol{\theta}}_N$ be estimators of $\boldsymbol{\theta}$ each satisfying statement (3.3), with asymptotic variances $\mathbf{V} = \text{Avar}\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ and $\mathbf{D} = \text{Avar}\sqrt{N}(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ (these generally depend on the value of $\boldsymbol{\theta}$, but we suppress that consideration here).

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

Definition 3.12. Partition $\hat{\boldsymbol{\theta}}_N$ satisfying statemente (3.3) into vectors $\hat{\boldsymbol{\theta}}_{N1}$ and $\hat{\boldsymbol{\theta}}_{N2}$. Then $\hat{\boldsymbol{\theta}}_{N1}$ and $\hat{\boldsymbol{\theta}}_{N2}$. 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 $\sqrt{N}(\hat{\boldsymbol{\theta}}_{N1} - \boldsymbol{\theta}_1)$ and similarly for $\mathbf{V}_2$. In other words, the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ is block diagonal.

3.5.2 Asymptotic Properties of Test Statistics

Definition 3.13.

1. The **asymptotic size** of a testing procedure is defnied as the limiting probability of rejecting H_0 when it is true. Mathematically, we can write this as $\lim_{N\to\infty} \text{P}_N(\text{reject } H_0|H_0)$, 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: $\lim_{N\to\infty} \text{P}_N(\text{reject } H_0|H_1) = 1$,

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

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

and

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

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

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

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

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

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

Lemma 3.9. Suppose that statement (3.3) holds, where $\mathbf{V}$ is positive definite. Let $: \boldsymbol{\Theta} \to \mathbb{R}^Q$ 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 $\mathbf{C}(\boldsymbol{\theta}) \equiv \nabla_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ as the $Q \times P$ Jacobian off . Then

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

and

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

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

$$\left\{\sqrt{N}[(\hat{\boldsymbol{\theta}}_N) - (\boldsymbol{\theta})]\right\}' \left[\widehat{\mathbf{C}}_N\widehat{\mathbf{V}}_N\widehat{\mathbf{C}}'_N\right]^{-1} \left\{\sqrt{N}[(\hat{\boldsymbol{\theta}}_N) - (\boldsymbol{\theta})]\right\} \overset{a}{\sim} \chi^2_Q \tag{3.6}$$

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 Porperties of OLS

4.2.1. Consitency
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 Porxy 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,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u \tag{4.1}$$

onde

- $y, x_1, \ldots, x_K$ 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;
- $\beta_0, \beta_1, \ldots, \beta_K$ são parâmetros (constantes) que gostaríamos de estimar.

Structural Model Goldberger (1972) [1] 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. Someteimes the structural model is directly estimable. Other times times we must combine auxiliary assumptions about other variables with algebraic manipulations to arrive at an **estimable model**. In addtion, 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 consitently estimate the β_j (assuming we have availabe a random sample from the population) is that the error (in the population) has mean zero and is uncorrelated with each of the regressors:

$$\mathrm{E}(u) = 0\,, \quad \mathrm{Cov}(x_j, u) = 0\,, \quad j = 1, \ldots, K. \tag{4.2}$$

The zero-mean assumptions 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 $(1, x_1, \ldots, x_K)$ as $\beta_0 + \beta_1 x_1 + \ldots \beta_K x_K$.

[1] Goldberger (1972), "Structural Equation Methods in the Social Sciences," *Econometrica* 40, 979–1001.

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

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

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

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

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:

$$y = \boldsymbol{x\beta} + u \tag{4.5}$$

onde,
$\boldsymbol{x} \equiv (x_1,\dots,x_K)$ 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, $\{(\boldsymbol{x}_i, y_i);\ i = 1, 2, \dots, N\}$ são tratados como variáveis aleatória independentes, identicamente distribuídas, onde $\boldsymbol{x}_i$ é $1 \times K$ e y_i é escalar. Para cada observação i, temos:

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + u_i. \tag{4.6}$$

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

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

$$\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{u} \tag{4.7}$$

$\boldsymbol{y}$ é um vetor $N \times 1$;
X é uma matriz $N \times K$ de regressores, com N vetores, $\boldsymbol{x}_i$, de dimensão $1 \times K$ empilhados;
$\boldsymbol{\beta}$ é um vetor $K \times 1$;
$\boldsymbol{u}$ é um vetor $N \times 1$;

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

4.2.1 Consistency

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

Assumptions

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

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

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

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

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

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

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

Analogy Principle The **analogy principle** for choosing and 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 $\mathrm{E}(\boldsymbol{x}'\boldsymbol{x})$ and $\mathrm{E}(\boldsymbol{x}'y)$ with the corresponding **sample averages**. Doing so leads to the OLS estimator:

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

Isso pode ser escrito na forma matricial:

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

onde,

X é a matriz de dados $N \times K$ dos regressores com linha i igual a $\boldsymbol{x}_i$;
é o vetor de dados $N \times 1$ com o i-ésimo elemento de $\boldsymbol{y}$ sendo representado por y_i.

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

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

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

Further, under **OLS.1**

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

Therefore, by Slutsky's Theorem (Lemma 3.4),

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

Resumo Resumimos os resultados acima com um teorema:

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

Sob as hipóteses do Teorema 4.2.1, $\boldsymbol{x}\boldsymbol{\beta}$ é uma **projeção linear** de y em $\boldsymbol{x}$.

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 Porperties of System OLS
7.3.3. Testing Multiple Hypotheses
7.4. Consistency and Asymptotic Normality of Generalized Least Squares
7.4.1. Consitency
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*: $\{(X_i, \boldsymbol{y}_i) : i = 1, \ldots, N\}$, onde:

X_i é uma matriz $G \times K$ e contém as variáveis explicativas que aparecem em qualquer lugar do sistema.

$\boldsymbol{y}_i$ 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:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{u}_i\,, \quad i = 1, \ldots, N, \tag{5.1}$$

onde:

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

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

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

5.2 Propriedades Assintóticas do SOLS

Wooldridge (2010, Sec.7.3.1)

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

SOLS.2 $A \equiv \mathrm{E}(X_i'X_i)$ 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 $\boldsymbol{\beta}$ consistente. Essa hipótese permite que alguns elementos de X_i sejam correlacionados com elementos de $\boldsymbol{u}_i$. Uma hipótese mais forte seria:

$$\mathrm{E}(\boldsymbol{u}_i|X_i) = \mathbf{0} \tag{5.2}$$

Sob **SOLS.1**, temos:

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

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

$$\boxed{\boldsymbol{\beta} = \left[\mathrm{E}(X_i'X_i)\right]^{-1}\mathrm{E}(X_i'\boldsymbol{y}_i)} \tag{5.3}$$

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

$$\boxed{\widehat{\boldsymbol{\beta}}^{SOLS} = \left(N^{-1}\sum_{i=1}^{N} X_i'X_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N} X_i'\boldsymbol{y}_i\right)}. \tag{5.4}$$

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

$$\boxed{\widehat{\boldsymbol{\beta}}^{SOLS} = (X'X)^{-1}(X'\boldsymbol{y})} \tag{5.5}$$

onde
$X \equiv (X_1', \ldots, X_N')$ é uma matriz $NG \times K$ dos X_i empilhados.
$\boldsymbol{y} \equiv (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_N')$ é um vetor $NG \times 1$ das observações $\boldsymbol{y}_i$ 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):

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

Por **SOLS.1**, $N^{-1}\sum_{i=1}^{N} X_i'\boldsymbol{u}_i \xrightarrow{p} \mathbf{0}$; 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

$$\boxed{\widehat{\boldsymbol{\beta}}^{SOLS} \xrightarrow{p} \boldsymbol{\beta}}.$$

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 $X_i'\boldsymbol{u}_i\boldsymbol{u}_i'X_i$ tenham *finite expected absolute value.* De (5.4) e (5.1), escrevemos:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \boldsymbol{\beta}+\left(N^{-1}\sum_{i=1}^{N} X_i' X_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N} X_i'\boldsymbol{u}_i\right)\\
(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= \left(N^{-1}\sum_{i=1}^{N} X_i' X_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N} X_i'\boldsymbol{u}_i\right)
\end{aligned}
$$

$$\boxed{\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \left(N^{-1}\sum_{i=1}^{N} X_i' X_i\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N} X_i'\boldsymbol{u}_i\right)}.$$

Uma vez que $\mathrm{E}(X_i'\boldsymbol{u}_i)=0$, sob a hipótese **SOLS.1**, o CLT implica que:

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

onde

$$B \equiv \mathrm{E}(X_i'\boldsymbol{u}_i\boldsymbol{u}_i'X_i) \equiv \mathrm{Var}(X_i\boldsymbol{u}_i).$$

Em particular,

$$N^{-1/2}\sum_{i=1}^{N} X_i\boldsymbol{u}_i = \mathrm{O}_p(1).$$

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,

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

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

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

Resumimos esse resultado com o seguinte Teorema:

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

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

5.2.3 Variância Assintótica

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

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

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

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

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

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

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

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

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

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

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

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

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

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

6 SOLS para Dados de Painel

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

No caso de dados de painel:

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

Portanto, podemos escrever $\widehat{\boldsymbol{\beta}}$ como:

$$\boxed{\widehat{\boldsymbol{\beta}}^{POLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}' \boldsymbol{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}' y_{it}\right)}. \tag{6.1}$$

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.

E$(\boldsymbol{x}_{it})$

$X_i' \boldsymbol{u}_i = \sum_{t=1}^{T} \boldsymbol{x}_{it}' u_{it}$

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. $\mathrm{E}(X_i \otimes \boldsymbol{u}_i) = 0$.

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

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

   Onde, Ω é a seguinte matriz **simétrica**, positiva-definida:

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

Estimação

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

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

Estimando a equação acima por **SOLS**:

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

8 GLS Factível

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

FSGLS: SGLS Factível

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

1. Estimar $\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{u}_i$ via **SOLS** e guardar o resíduo estimado $\widehat{\boldsymbol{u}}_i$.
2. Estimar Ω com o seguinte estimador $\widehat{\Omega}$:

$$\widehat{\Omega} = N^{-1}\sum_{i=1}^{N}\boldsymbol{u}_i\boldsymbol{u}_i'$$

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

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

Empilhando as N observações:

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

Reescrevendo a equação acima:

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

Valor Esperado

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

Concluímos que, se $\widehat{\Omega}\xrightarrow{p}\Omega$, então, $\beta^{FSGLS}\xrightarrow{p}\beta$,

Variância

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

Tirando o valor Esperado e supondo que:

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

temos:

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

e temos:

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

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*:

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

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 β_0, β_1 e β_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 vaariáveis.

Estimação

Indo para o problema de dados de painel, temos:

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

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

Se é verdade que há endogeneidade em (10.2), então:

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

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:

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

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

$$\begin{aligned}
\mathrm{E}[Z_i'(\boldsymbol{y}_i - X_i \boldsymbol{\beta})] &= 0 \\
\mathrm{E}(Z_i' \boldsymbol{y}_i) - \mathrm{E}(Z_i' X_i)\boldsymbol{\beta} &= 0 \\
\mathrm{E}(Z_i' \boldsymbol{y}_i) &= \mathrm{E}(Z_i' X_i)\boldsymbol{\beta} \\
&\boxed{\boldsymbol{\beta} = \left[\mathrm{E}(Z_i' X_i)\right]^{-1} \left[\mathrm{E}(Z_i' \boldsymbol{y}_i)\right]}
\end{aligned}$$

Se Usarmos estimadores amostrais:

$$\hat{\boldsymbol{\beta}} = \left[N^{-1}\sum_{i=1}^{N} Z_i'X_i\right]^{-1}\left[N^{-1}\sum_{i=1}^{N} Z_i'\boldsymbol{y}_i\right]$$
$$\boxed{\hat{\boldsymbol{\beta}} = (Z'X)^{-1}(Z'\boldsymbol{y})}$$

Se $L > K$, vamos considerar:

$$\underset{\boldsymbol{\beta}}{\text{Min}}\ \text{E}(Z_i\boldsymbol{u}_i)^2$$

onde:

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

Derivando em relação em $\boldsymbol{\beta}$ e igualando a zero:

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

Um estimador mais eficiente pode ser encontrado fazendo:

$$\underset{\boldsymbol{\beta}}{\text{Min}}\ \text{E}[(Z_i'\boldsymbol{y} - Z'X\boldsymbol{\beta})'W(Z_i'\boldsymbol{y} - Z'X\boldsymbol{\beta})].$$

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

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

Derivando em relação em $\boldsymbol{\beta}$ e igualando a zero:

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

Valor Esperado

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

Variância

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

Definindo $\Delta = \text{E}(Z'\boldsymbol{u}\boldsymbol{u}'Z)$ 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 β^{2SLS}.

11 Random Effects (RE, EA)

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{11.1}$$

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 $\boldsymbol{x}_{it}$ 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 consitência do estimador de FGLS.

Podemos reescrever (11.1) como:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}, \tag{11.2}$$

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:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{v}_i, \tag{11.3}$$

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

Hipóteses de $\widehat{\boldsymbol{\beta}}^{RE}$

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

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

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

2. Posto completo de $\mathrm{E}(X_i'\Omega^{-1}X_i)$.

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

A matriz Ω 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) port $\Omega^{-1/2}$ do dois lados, temos:

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

Estimando o modelo acima por POLS:

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

O problema, agora, é estimar Ω. Supondo:

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

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

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

Assim,

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

onde $\sigma_u^2 I_T$ é uma matriz diagonal, e $\sigma_c^2\mathbf{1}_T\mathbf{1}_T^{\prime}$ é uma matriz com todos os elementos iguais a σ_c^2.
Agora, rodando POLS em (11.3) e guardando os resíduos, temos:

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

e conseguimos estimar σ_v^2 e σ_c^2 por estimadores amostrais:

- como $\sigma_v^2 = \mathrm{E}(v_{it}^2)$:
$$\hat{\sigma}_v^2 = (NT - K)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{v}_{it}^2$$
- como $\sigma_c^2 = \mathrm{E}(v_{it}v_{is})$:
$$\hat{\sigma}_c^2 = \left[N\frac{T(T-1)}{2} - K\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T} \hat{v}_{it}\hat{v}_{is}$$
- N indivíduos;
- T elementos da diagonal principal de Ω
- $\frac{T(T-1)}{2}$ 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 $\widehat{\Omega}$, e podemos escrever:

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

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

$$\boldsymbol{\beta}^{RE} = \left[X^{\prime}(I_N \otimes \widehat{\Omega}^{-1})X\right]^{-1} \left[X^{\prime}(I_N \otimes \widehat{\Omega}^{-1})\boldsymbol{y}\right]. \qquad (11.6)$$

Valor Esperado

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

Variância

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

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

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

12 Fixed Effects (EF, FE)

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{12.1}$$

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 $\boldsymbol{x}_{it}$. 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:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + c_1\mathbf{1}_T + \boldsymbol{u}_i, \tag{12.2}$$

com $\boldsymbol{v}_i = c_i\mathbf{1}_T + \boldsymbol{u}_i$ 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.

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

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

$$M^0\boldsymbol{y}_i = M^0X_i\boldsymbol{\beta} + M^0(c_1\mathbf{1}_T) + M^0\boldsymbol{u}_i, \quad i = 1, \dots, N.$$

$$M^0(c_1\mathbf{1}_T) = (I_T - T^{-1}\mathbf{1}_T\mathbf{1}_T')c_i\mathbf{1}_T = c_i\mathbf{1}_T - T^{-1}c_i\mathbf{1}_T\mathbf{1}_T'\mathbf{1}_T = c_i\mathbf{1}_T - c_i\mathbf{1}_T \implies \boxed{M^0(c_1\mathbf{1}_T) = 0}$$

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

Estimação POLS

Aplicando POLS no modelo (12.3)

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

Hipóteses

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

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

FE.2: Posto completo de $\mathrm{E}(X_i'\Omega^{-1}X_i)$ (para inverter a matriz). $posto[\mathrm{E}(X_i'\Omega^{-1}X_i)] = K$.

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

Valor Esperado

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

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

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

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

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

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

Variância

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

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

Pão:

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

Recheio:

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

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

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

Variância sob Homocedasticidade

Usando **FE.3**, temos

Recheio':

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

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

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{\beta}^{FE}) &= \text{Pão Recheio' Pão} \\
&= \mathrm{E}\left\{\left[X'(I_N \otimes M^0)X\right]^{-1}\right\}\sigma_u^2\mathrm{E}\left[X'(I_N \otimes M^0)X\right]\mathrm{E}\left\{\left[X'(I_N \otimes M^0)X\right]^{-1}\right\} \\
&= \mathrm{E}\left\{\left[X'(I_N \otimes M^0)X\right]^{-1}\right\}\sigma_u^2 I_{NT}
\end{aligned}$$

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

13 First Difference (FD, PD)

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{13.1}$$

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 $\boldsymbol{x}_{it}$. 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 trasnformação a ser feita é a primeira diferença. Para tanto, seguimos os seguintes passos:

- Reescrevemos (13.1) defasado:

$$y_{it-1} = \boldsymbol{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \tag{13.2}$$

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

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

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

Reescrevendo (13.3) no formato matricial empilhando T:

$$\Delta \boldsymbol{y}_i = \Delta X_i \boldsymbol{\beta} + \boldsymbol{e}_i \tag{13.4}$$

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

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

Estimação POLS

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

$$\boxed{\boldsymbol{\beta}^{FD} = \left[\sum_{i=1}^{N} \Delta X_i' \Delta X_i\right]^{-1} \left[\sum_{i=1}^{N} \Delta X_i' \Delta \boldsymbol{y}_i\right]} \tag{13.5}$$

Hipóteses

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

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

FD.2: Posto completo de $\mathrm{E}(\Delta X_i' \Delta X_i)$ (para inverter a matriz). $posto[\mathrm{E}(\Delta X_i' \Delta X_i)] = K$.

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

Valor Esperado

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

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

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

Variância

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

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

Variância sob Homocedasticidade

Usando **FD.3**, temos

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

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

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

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

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

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

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

$$\mathrm{E}(u_{it} \,|\, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}) = 0 \,, \qquad t = 1, \ldots, T$$

ou seja:

$$\mathrm{E}(y_{it} \,|\, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}) = \boldsymbol{x}_{it}\boldsymbol{\beta} \,, \qquad t = 1, \ldots, T$$

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

Para o seguinte modelo:

$$y_{it} = \boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho y_{it-1} + c_i + u_{it}. \,, \qquad t = 2, \ldots, T$$

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

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

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

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

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

$$\begin{aligned} y_{it} - y_{it-1} &= (\boldsymbol{z}_{it} - \boldsymbol{z}_{it-1})\boldsymbol{\gamma} + \rho(y_{it-1} - y_{it-2}) + (c_i - c_i) + (u_{it} - u_{it-1}) \\ &\boxed{\Delta y_{it} = \Delta \boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho \Delta y_{it-1} + \Delta u_{it} \,, \qquad t = 3, \ldots, T} \end{aligned} \tag{14.1}$$

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 Δy_{it-1} (alguns intrumentos para Δy_{it-1} são $y_{it-2}, y_{it-3}, \ldots, y_{i1}$).

P2SLS

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

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

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

com

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

onde P_Z é a matriz de projeção em Z.

FDIV

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

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

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

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

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

Hipóteses

FDIV.1: $\mathrm{E}(\boldsymbol{w}_{it}\Delta u'_{it})$ para $i = 1, \dots, N$, $t = 2, \dots, T$.

FDIV.2: $Posto\,[\mathrm{E}(W'_i W_i)] = L$

FDIV.3: $Posto\,[\mathrm{E}(W'_i \Delta X_i)] = K$

Estimação FDIV

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

Valor Esperado

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

Variância

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

$e_i = \Delta u_{it}$.

15 Latent Variables, Probit and Logit

Modelo

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

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i. \tag{15.1}$$

Defina y como:

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

temos que:

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

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

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

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

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

Se $G(\cdot)$ for uma distribuição:

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

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

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

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

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 $\boldsymbol{y}$, ou seja:

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

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):

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

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

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

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

$$\underset{\boldsymbol{\beta}\in B}{\text{Max}}\ \mathrm{E}\left[\ell_i(\boldsymbol{\beta})\right].$$

Definindo o *vector score* da observação i:

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

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

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

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 $\boldsymbol{\beta}_0$, b por $\widehat{\boldsymbol{\beta}}$ e c por $\bar{\boldsymbol{\beta}}$, temos:

$$H_i(\bar{\boldsymbol{\beta}}) = \frac{s_i(\widehat{\boldsymbol{\beta}}) - s_i(\boldsymbol{\beta}_0)}{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0},$$

tirando médias dos dois lados:

$$N^{-1}\sum_{i=1}^{N} H_i(\bar{\boldsymbol{\beta}}) = \frac{1}{\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0} N^{-1}\sum_{i=1}^{N}\left[s_i(\widehat{\boldsymbol{\beta}}) - s_i(\boldsymbol{\beta}_0)\right]$$

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

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = (-1)\left[N^{-1}\sum_{i=1}^{N} H_i(\bar{\boldsymbol{\beta}})\right]^{-1} N^{-1}\sum_{i=1}^{N} s_i(\boldsymbol{\beta}_0)$$

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left[-N^{-1}\sum_{i=1}^{N} H_i(\bar{\boldsymbol{\beta}})\right]^{-1} \sqrt{N}\cdot N^{-1}\sum_{i=1}^{N} s_i(\boldsymbol{\beta}_0)$$

$$\boxed{\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left[-N^{-1}\sum_{i=1}^{N} H_i(\bar{\boldsymbol{\beta}})\right]^{-1} N^{-1/2}\sum_{i=1}^{N} s_i(\boldsymbol{\beta}_0)}.$$

Onde

$$\left[-N^{-1}\sum_{i=1}^{N} H_i(\bar{\boldsymbol{\beta}})\right]^{-1} \xrightarrow{p} A_0^{-1}, \qquad N^{-1/2}\sum_{i=1}^{N} s_i(\boldsymbol{\beta}_0) \xrightarrow{d} N(0, B_0).$$

Assim, temos que:

$$\boxed{\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(0, A_0^{-1}B_0A_0^{-1})}.$$

A forma mais simples de achar $\mathrm{Var}(\widehat{\boldsymbol{\beta}})$ é:

$$\boxed{\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = -\mathrm{E}[H_i(\widehat{\boldsymbol{\beta}})]^{-1}}.$$

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 pode 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 ccada indivíduo i, observamos apenas y_{i1} ou y_{i0}, nunca ambos.

Antes de continuarmos, faremos as seguintes definições:

ATE: $\mathrm{E}(y_1 - y_0)$

ATT: $\mathrm{E}(y_1 - y_0 \,|\, w = 1)$ (ATE no tratado).

ATE e ATT condicional a variáveis $\boldsymbol{x}$

$$\begin{aligned} ATE(\boldsymbol{x}) &= \mathrm{E}(y_1 - y_0 \,|\boldsymbol{x}) \\ ATT(\boldsymbol{x}) &= \mathrm{E}(y_1 - y_0 \,|\boldsymbol{x}, w = 1) \end{aligned}$$

<u>OBS:</u>

$$\begin{aligned} \mathrm{E}(y_1 - y_0) &= \mathrm{E}\left[\mathrm{E}(y_1 - y_0 \,|w)\right] \\ \mathrm{E}(y_1 - y_0 \,|w) &= \mathrm{E}(y_1 - y_0 \,|w = 0) \cdot P(w = 0) + \mathrm{E}(y_1 - y_0 \,|w = 1) \cdot P(w = 1). \end{aligned}$$

Métodos Assumindo Ignorabilidade do Tratamento

ATE.1: Ignorabilidade.
w e (y_1, y_0) são independentes condicionais a $\boldsymbol{x}$.

ATE.1': Ignorabilidade da Média.
a) $\mathrm{E}(y_0 \,|\, w, \boldsymbol{x}) = \mathrm{E}(y_0 \,|\, \boldsymbol{x})$
b) $\mathrm{E}(y_1 \,|\, w, \boldsymbol{x}) = \mathrm{E}(y_1 \,|\, \boldsymbol{x})$

Vamos definir

$$\begin{aligned} \mathrm{E}(y_0 \,|\, \boldsymbol{x}) &= \mu_0(\boldsymbol{x}) \\ \mathrm{E}(y_1 \,|\, \boldsymbol{x}) &= \mu_1(\boldsymbol{x}). \end{aligned}$$

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

$$\begin{aligned} ATE(\boldsymbol{x}) &= \mathrm{E}(y_1 - y_0 \,|\boldsymbol{x}) = \mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x}) \\ ATT(\boldsymbol{x}) &= \mathrm{E}(y_1 - y_0 \,|\boldsymbol{x}, w = 1) = \mu_1(\boldsymbol{x}) - \mu_0(\boldsymbol{x}) \end{aligned}$$

ATE.2: *Overlap*

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

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

Para o ATT vamos supor:

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

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

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:

$$wy = w^2 y_1 + (w - w^2)y_0 \implies \boxed{wy = wy_1}$$

$$(1 - w)y = (w - w^2)y_1 + (w^2 - 2w + 1)y_0 \implies \boxed{(1 - w)y = (1 - w)y_0}.$$

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

$\mu_1(\boldsymbol{x})$

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

Como w é binaria: $\mathrm{E}(w|\boldsymbol{x}) = P(w = 1|\boldsymbol{x}) = p(\boldsymbol{x})$. Assim:

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

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

$\mu_0(\boldsymbol{x})$

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

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

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

ATE:

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

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

ATT:

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

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

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

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

TODO

References

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