

Microeconometrics

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

Autor: Paulo Ferreira Naibert

Porto Alegre
04/07/2020
Revisão: July 4, 2020

1 Conceitos Básicos de Convergência Estatística

Definição 1.1 (Estimador Consistente). *Um estimador $\hat{\theta}$ é **consistente** para um parâmetro θ se*

$$\hat{\theta} \xrightarrow{p} \theta.$$

Definição 1.2 (Convergência em Probabilidade). *Uma sequência de variáveis aleatórias: $\{X_n\}_{n \geq 1}$ converge em probabilidade para uma variável aleatória X se, dado $\varepsilon > 0$,*

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

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

$$X_n \xrightarrow{p} X.$$

Definição 1.3 (Desigualdade de Markov). *Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias com $E|X_n|^K < +\infty$, $K > 0$. Então, dado $\varepsilon > 0$*

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

Definição 1.4.

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

Definição 1.5 (Erro Quadrático Médio).

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

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

Definição 1.6 (LGN – Lei dos Grandes Números). *Seja $\{X_i\}_{i \geq 1}$ uma sequência de variáveis aleatórias iid com $E(X_i) = \mu$. Então,*

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

Definição 1.7 (LGN – Caso Matricial). *Seja $\{\mathbf{x}_i\}_{i=1}^N$, uma sequência iid de vetores aleatórios $K \times 1$ com $E(\mathbf{x}_i \mathbf{x}_i') = Q_{K \times K}$ finita. Então,*

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

Se Q for positiva definida, Q terá inversa.

Multiplicação de Matriz

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

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

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

Definição 1.8.

$$\begin{aligned} P(|X_n - X| > \varepsilon) &\rightarrow 0 \\ X_n - X &\xrightarrow{p} 0 \\ X_n &\xrightarrow{p} X \end{aligned}$$

Definição 1.9 (o_p).

$$\begin{aligned} X_n = o_p(1) &\implies X_n \xrightarrow{p} 0 \\ X_n = o_p(Y_n) &\implies \frac{X_n}{Y_n} = o_p(1) \implies \frac{X_n}{Y_n} \xrightarrow{p} 0 \\ X_n = W_n + o_p(1) &\implies (X_n - W_n) = o_p(1) \implies (X_n - W_n) \xrightarrow{p} 0 \end{aligned}$$

Definição 1.10 (Limitação em Probabilidade: O_p). Dizemos que X_n é **limitado em probabilidade** e denotado por $X_n = O_p(1)$, se existe M maior que zero, tal que para todo ε maior que zero, $P(|X_n| > \varepsilon) < \varepsilon$.

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

Definição 1.11. Dizemos que $X_n = O_p(Y_n)$ se existe M maior que zero, tal que para todo ε maior que zero, $P(|X_n/Y_n| > \varepsilon) < \varepsilon$.

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

Definição 1.12. Se $X_n = O_p(1)$ e $Y_n = o_p(1)$, então

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

Definição 1.13 (Equivalência Assintótica). Seja $\{\mathbf{x}_n\}$ e $\{\mathbf{z}_n\}$ sequências de vetores aleatórios $K \times 1$. Se $\mathbf{z}_n \xrightarrow{d} \mathbf{z}$ e $\mathbf{x}_n - \mathbf{z}_n \xrightarrow{p} \mathbf{0}_K$. Então,

$$\mathbf{x}_n \xrightarrow{d} \mathbf{z}.$$

Definição 1.14 (Convergência em Distribuição). Seja $\{X_n\}_{n \geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória com F_n e F suas respectivas FDAs, então

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

para todo X onde F é contínuo.

Definição 1.15 (Convergência em Distribuição e Limitação em Probabilidade). Se $X_n \xrightarrow{d} X$, X um variável aleatória qualquer; então $X_n = O_p(1)$.

Definição 1.16 (TCL – Teorema Central do Limite). *Seja $\{X_n\}_{n=1}^N$ iid com $E(X_n) = \mu$ e $\text{Var}(X_n) = \sigma^2 < +\infty$. Então, para $S_N = \sum_{n=1}^N X_n$:*

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

Definição 1.17 (TCL – Caso Vetorial). *Seja $\{\mathbf{w}_i\}_{i=1}^n$ uma sequência iid de vetores aleatórios $K \times 1$ com $E(w_{ik}^2) < +\infty$, $k = 1, \dots, K$ e $E(\mathbf{w}_i) = \mathbf{0}_K$. Então,*

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

onde, $B = \text{Var}(\mathbf{w}_i) = E(\mathbf{w}_i \mathbf{w}_i')$.

Definição 1.18. *Seja $\{\mathbf{z}_n\}$ uma sequência de vetores $K \times 1$ aleatórios com*

$$\mathbf{z}_n \xrightarrow{d} N(\mathbf{0}, V).$$

*Então, para qualquer matriz A de dimensão $K \times M$ **não** estocástica,*

$$A' \mathbf{z}_n \xrightarrow{d} N(\mathbf{0}, A'VA).$$

2 Regressão MQO Clássico

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

2.1 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,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u \quad (2.1)$$

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)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. 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 β_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:

$$E(u) = 0, \quad \text{Cov}(x_j, u) = 0, \quad j = 1, \dots, K. \quad (2.2)$$

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 (2.1) and assumption (2.2) is equivalent to defining the **linear projection** of y onto $(1, x_1, \dots, x_K)$ as $\beta_0 + \beta_1 x_1 + \dots + \beta_K x_K$.

Sufficient for assumption (2.2) is the zero conditional mean assumption:

$$E(u|x_1, \dots, x_K) = E(u|\mathbf{x}) = 0. \quad (2.3)$$

Under equation (2.1) and assumption (2.3), we have the **population regression function**

$$E(y|x_1, \dots, x_K) = \beta_0 + \beta_1 x_1 + \dots + \beta_K x_K. \quad (2.4)$$

Endogenous Variable An explanatory variable x_j is said to be **endogenous** in equation (2.1) 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 (2.1). If assumption (2.3) holds, then each explanatory variable is necessarily exogenous.

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

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

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

2.2 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:

$$y = \mathbf{x}\boldsymbol{\beta} + u \quad (2.5)$$

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 aleatórias independentes, identicamente distribuídas, onde \mathbf{x}_i é $1 \times K$ e y_i é escalar. Para cada observação i , temos:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i. \quad (2.6)$$

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

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad (2.7)$$

\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 $N \times 1$;

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

2.2.1 Consistency

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

Assumptions

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

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 (2.3).

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 $\boldsymbol{\beta}$ is **identified**. In the context of models that are linear in the parameters under random sampling, identification of $\boldsymbol{\beta}$ simply

means that β can be written in terms of population moments in observable variables. To see this, we premultiply equation (2.5) 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, β 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 ;

\mathbf{y} é o vetor de dados $N \times 1$ com o i -ésimo elemento de \mathbf{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),

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

Resumo Resumimos os resultados acima com um teorema:

Teorema 2.1 (Consistência do OLS). *Sob as Hipóteses **OLS.1** e **OLS.2**, o estimador de OLS, $\hat{\beta}$ obtido de uma amostra aleatória seguindo o modelo populacional (2.5) é consistente para β .*

Sob as hipóteses do Teorema 2.1, $x\beta$ é uma **projeção linear** de y em x .

2.2.2 Asymptotic Inference Using OLS

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

Hudson OLS

Hipóteses

H1 Estamos usando o modelo correto:

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

H2 X é **não** estocástica;

H3 $\{u_i\}_{i=1}^N$ é *iid* com e para cada $i = 1, \dots, N$:

$$\begin{aligned} E(u_i) &= 0 \\ \text{Var}(u_i) &= E(u_i^2) = \sigma^2 \end{aligned}$$

H2' X é estocástica;

H3'

$$\begin{aligned} E(u_i|X) &= 0 \\ \text{Var}(u_i|X) &= E\left\{[u_i - E(u_i|X)]^2 | X\right\} = E(u_i^2|X) = \sigma^2. \end{aligned}$$

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

Estimação

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y} \tag{2.8}$$

Valor Esperado

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

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

Viés

$$\boxed{B(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}}$$

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

Sob **H2'** e **H3'**:

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

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

Variância Supondo **H2'** e **H3'**:

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

Supondo homocedasticidade e ausência de correlação serial: $E[uu'|X] = \sigma^2 I_N$. Assim,

$$\begin{aligned}
 \text{Var}(\hat{\beta}|X) &= \sigma^2 (X'X)^{-1}X'I_N X (X'X)^{-1} = \sigma^2 (X'X)^{-1}X'X(X'X)^{-1} \\
 \boxed{\text{Var}(\hat{\beta}|X) &= \sigma^2 (X'X)^{-1}}.
 \end{aligned}$$

Ausência de Exogeneidade Estrita

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

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

o erro é correlacionado com o regressor y_{t-1} . Nesse caso, tentaremos obter apenas **consistência** e **variância assintótica** do estimador.

Estimação Lembrando que o estimador de OLS é:

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

onde usamos o modelo

$$y_i = x_i\beta + u_i,$$

e definimos as variáveis

$$X_{N \times K} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}.$$

Assim, representamos $(X'X)^{-1}$ e $(X'y)$ por meio dos seguintes somatórios: temos

$$(X'X)^{-1} = \left(\sum_{i=1}^N x_i' x_i \right)^{-1}, \quad (X'y) = \sum_{i=1}^N x_i' y_i.$$

Valor Esperado e Viés

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

Usando **LGN matricial** (lembrar que as dimensões dos vetores estão invertidas: $1 \times K$ e **não** $K \times 1$), temos:

$$\boxed{N^{-1} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \xrightarrow{p} Q}. \quad (2.9)$$

Supondo $E(\mathbf{x}'_i \mathbf{x}_i) = Q_{K \times K}$, finita e positiva definida, $posto(Q) = K$. Supondo $E(\mathbf{x}'_i \mathbf{u}_i) = 0$, o que corresponde a $Cov(\mathbf{x}_i, \mathbf{u}_i) = 0$, ou seja, o erro \mathbf{u}_i **não** é correlacionado com os regressores da própria equação. Isso é bem menos que exogeneidade estrita.

Então,

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

Logo

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

Então, $(\hat{\beta} - \beta) \xrightarrow{p} 0$ que é equivalente a $\boxed{\hat{\beta} \xrightarrow{p} \beta}$.

Normalidade Assintótica do $\hat{\beta}^{OLS}$

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

Supondo

$$E(x_{ik}^2 u_i^2) < +\infty, \quad k = 1, \dots, K,$$

temos, pelo **TCL**, que

$$\boxed{N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i \xrightarrow{d} N(0, B)} \quad (2.10)$$

onde

$$B = E[\mathbf{x}'_i u'_i u_i \mathbf{x}_i] = E[u_i^2 \mathbf{x}'_i \mathbf{x}_i].$$

E temos que $\boxed{N^{-1/2} \sum_{i=1}^N \mathbf{x}'_i u_i = O_p(1)}.$

Além disso, vamos utilizar a matriz **simétrica** e **não singular** Q da equação (2.9) Assim, temos

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

Podemos inverter Q porque ela tem posto completo (não singular). Pelas propriedades de Q , temos:

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

Então,

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

Usando (2.10) e a definição **XX**

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

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

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

Variância Vamos definir $\boxed{V = Q^{-1} B Q^{-1}}$:

$$\begin{aligned} V &= E[(\mathbf{x}'_i \mathbf{x}_i)]^{-1} E[(\mathbf{x}'_i u'_i u_i \mathbf{x}_i)] E[(\mathbf{x}'_i \mathbf{x}_i)]^{-1} \\ &\boxed{V = E[(\mathbf{x}'_i \mathbf{x}_i)]^{-1} E[(u_i^2 \mathbf{x}'_i \mathbf{x}_i)] E[(\mathbf{x}'_i \mathbf{x}_i)]^{-1}}. \end{aligned}$$

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

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

Estimador Amostral

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

 $\text{Var}(\widehat{\boldsymbol{\beta}})$

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

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

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

A variância **Robusta** é:

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

A variância sob **Homocedasticidade** é:

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

3 System OLS (SOLS)

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

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

3.1 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:

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

onde:

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

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

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

3.2 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 arcabouço de regressão para conseguirmos um estimador de $\boldsymbol{\beta}$ consistente. Essa hipótese permite que alguns elementos de X_i sejam correlacionados com elementos de \mathbf{u}_i . Uma hipótese mais forte seria:

$$E(\mathbf{u}_i | X_i) = \mathbf{0} \quad (3.2)$$

Sob **SOLS.1**, temos:

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

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 $\boldsymbol{\beta}$ precisamos assumir que ele é o único vetor $K \times 1$ que satisfaz $E(X_i' X_i) \boldsymbol{\beta} = E(X_i' \mathbf{y}_i)$. Por isso assumimos **SOLS.2** e sob **SOLS.1** e **SOLS.2**, podemos escrever $\boldsymbol{\beta}$ como:

$$\boxed{\boldsymbol{\beta} = [E(X_i' X_i)]^{-1} E(X_i' \mathbf{y}_i)} \quad (3.3)$$

o que mostra que **SOLS.1** e **SOLS.2** identifica o vetor β . O **princípio da analogia** sugere que estimemos β pelas analogias amostrais de (3.3). Assim, definimos o estimador SOLS de β como:

$$\hat{\beta}^{SOLS} = \left(N^{-1} \sum_{i=1}^N X_i' X_i \right)^{-1} \left(N^{-1} \sum_{i=1}^N X_i' y_i \right). \quad (3.4)$$

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

$$\hat{\beta}^{SOLS} = (X'X)^{-1} (X'y) \quad (3.5)$$

onde

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

$y \equiv (y_1', \dots, y_N')$ é um vetor $NG \times 1$ das observações y_i empilhadas.

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

3.2.1 Consistência

Para provarmos a **consistência** do estimador, usamos a equação (3.4):

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

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:

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

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

3.2.2 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 (3.4) e (3.1), escrevemos:

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

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

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

onde

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

Em particular,

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

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:

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

Resumimos esse resultado com o seguinte Teorema:

Teorema 3.2 (Normalidade Assintótica do SOLS). *Sob Hipóteses **SOLS.1** e **SOLS.2**, temos que a seguinte equação vale:*

$$\boxed{\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, A^{-1} B A^{-1})}. \quad (3.6)$$

3.2.3 Variância Assintótica

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

$$\text{Avar}(\hat{\beta}^{SOLS}) = A^{-1}BA^{-1}/N. \quad (3.7)$$

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:

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

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, $\text{Avar}[\sqrt{N}(\hat{\beta} - \beta)]$ é **consistentemente** estimado por $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$, e $\text{Avar}(\hat{\beta})$ 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 β como $\hat{\beta}$ fosse normalmente distribuído com média β e variância \hat{V} .

4 SOLS para Dados de Painei

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

No caso de dados de painel:

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

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

$$\hat{\beta}^{POLS} = \left(\sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}' \mathbf{x}_{it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T \mathbf{x}_{it}' y_{it} \right). \quad (4.1)$$

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 (4.1) é o mesmo para unidades de *cross section* amostradas em diferentes pontos do tempo. O Teorema 3.1, abaixo, mostra que o estimador POLS é consistente sob as condições de ortogonalidade na hipótese **XX** e uma hipótese de posto completo.

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

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

Onde, Ω 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**:

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

6 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 $\mathbf{y}_i = X_i\beta + \mathbf{u}_i$ via **SOLS** e guardar o resíduo estimado $\hat{\mathbf{u}}_i$.
2. Estimar Ω 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 β^{FSGLS} pela fórmula do β^{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:

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

7 Modelo de Efeitos Não Observados

[Wooldridge \(2010, C.10](#) – Basic Linear Unobserved Effects Panel Data Models)

8 Endogeneity and GMM

Modelo

No seguinte modelo *cross-section*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i ; \quad i = 1, \dots, N. \quad (8.1)$$

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 (8.1) 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 (8.1).
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:

$$\mathbf{y}_i = X_i \boldsymbol{\beta} + \mathbf{u}_i ; \quad i = 1, \dots, N. \quad (8.2)$$

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 (8.2), 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 β 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 β 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 β^{2SLS} .

9 Random Effects (RE, EA)

Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (9.1)$$

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 (9.1) como:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \quad (9.2)$$

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 (9.2) como:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + \mathbf{v}_i, \quad (9.3)$$

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

$$\begin{aligned} \Omega^{-1/2}\mathbf{y}_i &= \Omega^{-1/2}X_i\boldsymbol{\beta} + \Omega^{-1/2}\mathbf{v}_i \\ \mathbf{y}_i^* &= X_i^*\boldsymbol{\beta} + \mathbf{v}_i^*, \end{aligned} \quad (9.4)$$

Estimando o modelo acima por POLS:

$$\begin{aligned}
\beta^{POLS} &= \left(\sum_{i=1}^N X_i^{*'} X_i^* \right)^{-1} \left(\sum_{i=1}^N X_i^{*'} \mathbf{y}_i^* \right) \\
&= \left(\sum_{i=1}^N X_i' \Omega^{-1} X_i \right)^{-1} \left(\sum_{i=1}^N X_i' \Omega^{-1} \mathbf{y}_i \right) \\
&= (X'(I_N \otimes \Omega^{-1})X)^{-1} (X'(I_N \otimes \Omega^{-1})\mathbf{y}).
\end{aligned} \tag{9.5}$$

O problema, agora, é estimar Ω . 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 σ_c^2 .

Agora, rodando POLS em (9.3) e guardando os resíduos, temos:

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

e conseguimos estimar σ_v^2 e σ_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 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 $\hat{\Omega}$, 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 (9.5) como:

$$\beta^{RE} = \left[X'(I_N \otimes \hat{\Omega}^{-1})X \right]^{-1} \left[X'(I_N \otimes \hat{\Omega}^{-1})\mathbf{y} \right]. \tag{9.6}$$

Valor Esperado

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

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

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

10 Fixed Effects (EF, FE)

Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (10.1)$$

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 (10.1) por POLS, ignorando c_i , teríamos problemas de inconsistência devido a **endogeneidade**.

As T equações do modelo (10.1) podem ser reescritas como:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + c_i\mathbf{1}_T + \mathbf{u}_i, \quad (10.2)$$

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 (10.3) ao premultiplicarmos todo o modelo por M^0 .

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

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

$$\ddot{\mathbf{y}}_i = \ddot{X}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i, \quad i = 1, \dots, N. \quad (10.3)$$

Estimação POLS

Aplicando POLS no modelo (10.3)

$$\boxed{\boldsymbol{\beta}^{FE} = \left[\sum_{i=1}^N \ddot{X}_i'\ddot{X}_i \right]^{-1} \left[\sum_{i=1}^N \ddot{X}_i'\ddot{\mathbf{y}}_i \right]} \quad (10.4)$$

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 $T \times T$.

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

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

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

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

11 First Difference (FD, PD)

Modelo

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

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

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 (11.1) 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 (11.1) defasado:

$$y_{it-1} = \mathbf{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \quad (11.2)$$

- Tiramos a diferença entre (11.2) e (11.1):

$$\begin{aligned} y_{it} - y_{it-1} &= (\mathbf{x}_{it} - \mathbf{x}_{it-1})\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{it-1} \\ \Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}. \end{aligned} \quad (11.3)$$

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

Reescrevendo (11.3) no formato matricial empilhando T :

$$\Delta \mathbf{y}_i = \Delta \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \quad (11.4)$$

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 (11.4), assim:

$$\boxed{\boldsymbol{\beta}^{FD} = \left[\sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{y}_i \right]} \quad (11.5)$$

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.

12 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, ρc_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:

$$\begin{aligned} y_{it} - y_{it-1} &= (\mathbf{z}_{it} - \mathbf{z}_{it-1})\boldsymbol{\gamma} + \rho(y_{it-1} - y_{it-2}) + (c_i - c_i) + (u_{it} - u_{it-1}) \\ \boxed{\Delta y_{it} = \Delta \mathbf{z}_{it}\boldsymbol{\gamma} + \rho \Delta y_{it-1} + \Delta u_{it}, \quad t = 3, \dots, T} \end{aligned} \quad (12.1)$$

Estimação

Não podemos estimar o modelo (12.1) 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 instrumentos para Δy_{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

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

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

13 Latent Variables, Probit and Logit

Modelo

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

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i. \quad (13.1)$$

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 (13.1), 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 ε_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 β_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:

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

14 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.

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

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

Vamos definir

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

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

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

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

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

ATE.2: Overlap

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

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

Para o *ATT* vamos supor:

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

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

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

GREENE, WILLIAM H. 2012. *Econometric Analysis*. 7 edn. Boston: Prentice Hall.

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