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

Microeconometrics: Lecture Notes

Autor: Paulo Ferreira Naibert
Professor: Hudson Torrent

Porto Alegre
30/06/2020
Revisão: July 23, 2020

1 Regressão MQO Clássico

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

1.1 Modelo de equações lineares

O modelo populacional que estudamos é linear em seus parâmetros,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u \tag{1.1}$$

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.

Notação Vetorial

Wooldridge (2010, Sec. 4.2 – Asymptotic Properties of OLS; p.51)
Por conveniência, escrevemos a equação populacional em forma de vetor:

$$y = \boldsymbol{x}\boldsymbol{\beta} + u \tag{1.2}$$

onde,
$\boldsymbol{x} \equiv (x_1, \ldots, x_K)$ é um vetor $1 \times K$ de regressores;
$\boldsymbol{\beta} \equiv (\beta_1, \ldots, \beta_K)'$ 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, $\{(\boldsymbol{x}_i, y_i); i = 1, 2, \ldots, N\}$ 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:

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta} + u_i. \tag{1.3}$$

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

1.2 Hipóteses

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

OLS.2 **X** é **não** estocástica;

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

$$\mathrm{E}(u_i) = 0$$
$$\mathrm{Var}(u_i) = \mathrm{E}(u_i^2) = \sigma^2$$

OLS.2' $\mathbf{X}$ é estocástica;

OLS.3'

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

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

1.3 Estimação

Usando **OLS.1:**

$$
\begin{aligned}
y_i &= \boldsymbol{x}_i\boldsymbol{\beta} + u_i \\
\boldsymbol{x}_i' y_i &= \boldsymbol{x}_i'\boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{x}_i' u_i \\
\mathrm{E}(\boldsymbol{x}_i' y_i) &= \mathrm{E}(\boldsymbol{x}_i'\boldsymbol{x}_i)\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{x}_i' u_i)
\end{aligned}
$$

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

$$
\begin{aligned}
\mathrm{E}(\boldsymbol{x}_i' y_i) &= \mathrm{E}(\boldsymbol{x}_i'\boldsymbol{x}_i)\boldsymbol{\beta} \\
&\boxed{\boldsymbol{\beta} = [\mathrm{E}(\boldsymbol{x}_i'\boldsymbol{x}_i)]^{-1}\mathrm{E}(\boldsymbol{x}_i' y_i)}.
\end{aligned}
\tag{1.4}
$$

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

$$
\boxed{\widehat{\boldsymbol{\beta}} = \left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i' y_i\right)}. \tag{1.5}
$$

Podemos desenvolver essa equação para:

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

$$
\boxed{\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{x}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\boldsymbol{x}_i'\boldsymbol{u}_i\right)}. \tag{1.6}
$$

1.3.1 Notação Matricial

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

$$
\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{1.7}
$$

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

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

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

$$\widehat{\boldsymbol{\beta}} = (N^{-1}\mathbf{X}'\mathbf{X})^{-1}(N^{-1}\mathbf{X}'\boldsymbol{y}) \implies \boxed{\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{y})} \tag{1.8}$$

1.4 Valor Esperado

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

1.4.1 Viés

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

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

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

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

1.5 Variância

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

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

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

1.5.1 Homocedasticidade

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

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

2 Ausência de Exogeneidade Estrita

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

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

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

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{x}_i' \boldsymbol{x}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \boldsymbol{x}_i' \boldsymbol{u}_i \right).$$

Aqui comeceçaria a seção 16.

2.1 Consistência

Vamos definir a matriz $K \times K$, $\mathbf{A} \equiv \mathrm{E}(\boldsymbol{x}_i' \boldsymbol{x}_i)$. Supondo $\mathbf{A}$, finita e positiva definida, $\text{posto}(\mathbf{A}) = K$. Usando **LGN matricial** (Definição 16.2 na página 47), temos:
[lembrar que as dimensões dos vetores estão invertidas: $1 \times K$ e **não** $K \times 1$]

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

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

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

Logo,

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

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

2.2 Normalidade Assintótica

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

Supondo $\mathrm{E}(x_{ik}^2u_i^2) < +\infty$, $k = 1, \ldots, K$, e definindo $\mathbf{B} = \mathrm{E}[\boldsymbol{x}_i'u_i'u_i\boldsymbol{x}_i] = \mathrm{E}[u_i^2\boldsymbol{x}_i'\boldsymbol{x}_i]$. Temos, pela Definição 16.7 (**TCL**), que

$$
N^{-1/2}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i \xrightarrow{d} N(\mathbf{0},\mathbf{B}) \implies N^{-1/2}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i = \mathrm{O}_p(1) \tag{2.2}
$$

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

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

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

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

Então,

$$
\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \mathrm{o}_p(1)\mathrm{O}_p(1) + \mathbf{A}^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i\right),
$$

Usando (2.2) e o Lema 16.2.

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

Lembrando que $\mathrm{o}_p(1)\mathrm{O}_p(1) = \mathrm{o}_p(1)$, temos:

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

2.3 Variância

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

2.3.1 Homocedasticidade

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

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

2.3.2 Estimador Amostral

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

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

2.3.3 Variância do estimador de OLS

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

A variância **Robusta** é:

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

A variância sob **Homocedasticidade** é:

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

3 System OLS (SOLS)

Wooldridge (2010, C.7 – Estimating Systems of Equations by OLS and GLS, p.143–179)
Wooldridge (2010, Sec.7.3 – System OLS Estimation of a Multivariate Linear System, p.147)

3.1 Modelo Linear

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

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

$\boldsymbol{y}_i$ 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:

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

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

3.2 Hipóteses

Wooldridge (2010, Sec.7.3.1)

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

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

3.3 Estimação

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

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

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

$$\boxed{\boldsymbol{\beta} = \left[\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)\right]^{-1} \mathrm{E}(\mathbf{X}_i'\boldsymbol{y}_i)} \tag{3.2}$$

Usando estimadores amostrais:

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

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

$$\boxed{\widehat{\boldsymbol{\beta}}^{SOLS} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{y})} \tag{3.4}$$

onde

$\mathbf{X} \equiv (\mathbf{X}_1', \dots, \mathbf{X}_N')$ é uma matriz $NG \times K$ dos $\mathbf{X}_i$ empilhados.
$\boldsymbol{y} \equiv (\boldsymbol{y}_1', \dots, \boldsymbol{y}_N')$ é um vetor $NG \times 1$ das observações $\boldsymbol{y}_i$ empilhadas.

3.4 Consistência

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

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

E chegamos em:

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

Por **SOLS.1**:

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

e por **SOLS.2**

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

Resumimos esse resultado pelo seguinte Teorema:

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

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

3.5 Normalidade Assintótica

De (3.5):

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

E chegamos em:

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

Uma vez que $\mathrm{E}(\mathbf{X}_i'\boldsymbol{u}_i) = 0$, sob a hipótese **SOLS.1**, a definição 16.7 (**TCL**) implica que:

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

onde

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

Em particular,

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

Porém,

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

Sendo assim,

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

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

3.6 Variância Assintótica

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

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

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

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

3.6.1 A Matriz Robusta

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

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

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

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

3.6.2 Variância Asstintótica

REVER

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}^{SOLS}) = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}/N. \tag{3.8}$$

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

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

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

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

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

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

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

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

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

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

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

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

4 Dados de Painel (POLS)

Wooldridge (2010, C.7 – Estimating Systems of Equations by OLS and GLS. p.143-179)
Wooldridge (2010, Sec.7.8 – The Linear Panel Data Model, Revisited. p.169)

4.1 Modelo Linear para Dados de Painel

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

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

onde

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

$$\underset{1\times K}{\boldsymbol{x}_{it}} = \begin{bmatrix} x_{1,it} & \dots & x_{K,it} \end{bmatrix} \quad \underset{K\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}$$

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

$\boldsymbol{y}_i$ é um vetor $T \times 1$.
$\boldsymbol{\beta}$ é um vetor $K \times 1$.
$\mathbf{X}_i$ é uma matriz $T \times K$.
$\boldsymbol{u}_{it}$ é um vetor $T \times 1$.

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

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

$\boldsymbol{y}$ é um vetor $NT \times 1$.
$\boldsymbol{\beta}$ é um vetor $K \times 1$.
$\mathbf{X}$ é uma matriz $NT \times K$.
$\boldsymbol{u}$ é um vetor $NT \times 1$.

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

4.2 Estimação

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

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

$$\boxed{\widehat{\boldsymbol{\beta}}^{POLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}'\boldsymbol{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \boldsymbol{x}_{it}'y_{it}\right)}. \tag{4.2}$$

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

4.3 Hipóteses

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

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

5 Alguns Testes

Lembrando a equação do modelo (4.1):
$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + u_{it}$

5.1 Autocorrelação dos Resíduos

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

$$\hat{u}_{it} = y_{it} - \boldsymbol{x}_{it}\widehat{\boldsymbol{\beta}}^{POLS}. \tag{5.1}$$

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

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

Então, testamos

$$H_0 : \delta_1 = 0$$
$$H_1 : \delta_1 \neq 0$$

via teste t (pode ser robusto).

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

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

Então, testamos

$$H_0 : \delta_1 = 0$$
$$H_1 : \delta_1 \neq 0$$

via teste t.

5.2 Heterocedasticidade

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

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

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

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

Então, testamos

$$H_0 : \boldsymbol{\gamma} = \mathbf{0} \quad (\gamma_1 = 0 \text{ e } \gamma_2 = 0)$$
$$H_1 : \boldsymbol{\gamma} \neq \mathbf{0}$$

via teste de Wald.

5.3 Teste de Wald

Se é verdade que

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

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

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

e

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

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

Teste de Wald

$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$
$H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$

A estatística do teste acima é:

$$N\left[\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right]' \left(\mathbf{R}\widehat{\mathbf{V}}\mathbf{R}'\right)^{-1} \left[\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}\right] \overset{a}{\sim} \chi^2_Q$$

Remarks

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

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

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

6 Modelo de Efeitos Não Observados (UEM)

Wooldridge (2010, C.10 – Basic Linear Unobserved Effects Panel Data Models, p.247–291)
Wooldridge (2010, Sec.10.1 – Motivation: The Omitted Variables Problem, p.247)
Wooldridge (2010, Sec.10.2 – Assumptions about the Unobserved Effects and Explanatory Variables, p.251)
Wooldridge (2010, Sec.10.3 – Estimating UEM by POLS, p.256)

6.1 Modelo UEM

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

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

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

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it} \tag{6.2}$$

Ou, em forma de vetor:

$$\boldsymbol{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i \tag{6.3}$$

6.2 Estimação e Consistência

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

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

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

Ou seja, precisamos que:

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

Caso 1: $\mathrm{E}(\boldsymbol{x}'_{it}c_i) = \mathbf{0}$.
POLS é consistente, mas não é eficiente.
Efeitos Aleatórios é consistente e eficiente.
EA é o **FGLS** do modelo.

Caso 2: $\mathrm{E}(\boldsymbol{x}'_{it}c_i) \neq \mathbf{0}$.
Se POLS é **inconsistente**.
Nesse caso, usaremos o Modelo de **Efeitos Fixos** ou **Primeira Diferença**.
EF e PD é o POLS nos modelos transformados.

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

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

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

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

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

7.1 Modelo

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

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

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

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

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

7.2 Transformação *Within*

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

$$\overline{y}_i = \overline{\boldsymbol{x}}_i\boldsymbol{\beta} + \overline{c}_i + \overline{u}_i\,, \qquad i = 1, \dots, N. \tag{7.3}$$

onde:

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

Então, subtraindo (7.3) de (7.1):

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

Finalmente, obtemos:

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

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

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

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

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

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

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

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

A matriz $\mathbf{M}^0$ tem dimensão $T \times T$. Além disso, ela é idempotente ($\mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^0$) e simétrica ($\mathbf{M}^{0\,\prime} = \mathbf{M}^0$).

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

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

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

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

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

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

Após a transformação:

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

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

7.4 Hipóteses

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

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

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

7.5 Estimação POLS

Wooldridge (2010, p.269)
Aplicando POLS no modelo transformado (7.7), temos:

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

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

7.6 Consistência

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

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

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

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

Formato Totalmente Matricial

Empilhando os vetores N vezes, vamos definir:

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

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

$$\underset{N\times N}{\mathbf{I}_N}, \otimes \underset{T\times T}{\mathbf{M}^0}$$

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

$$\begin{aligned} \ddot{\boldsymbol{y}} &= (\mathbf{I}_N \otimes \mathbf{M}^0)\boldsymbol{y}, \\ \ddot{\mathbf{X}} &= (\mathbf{I}_N \otimes \mathbf{M}^0)\mathbf{X} \\ \ddot{\boldsymbol{u}} &= (\mathbf{I}_N \otimes \mathbf{M}^0)\boldsymbol{u}, \end{aligned}$$

E com isso, reescrevermos (7.9) como:

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

Ou ainda, como:

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

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

7.7 Matriz de Covariância Robusta

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

A matriz de covariância assintótica fica:

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

A qual pode ser estimada por

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

Analisando $\mathrm{E}\left(\ddot{\mathbf{X}}_i'\ddot{\boldsymbol{u}}_i\ddot{\boldsymbol{u}}_i'\ddot{\mathbf{X}}_i\right)$:

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

onde usamos as propriedades de simetria e idempotência da matriz $\mathbf{M}^0$ e as definições de $\ddot{\mathbf{X}}_i$ e $\ddot{\boldsymbol{u}}_i$.
Sob **FE.3**, $\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_i'|\ddot{\mathbf{X}}_i) = \sigma_u^2\mathbf{I}_T$, temos:

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

Assim, a matriz de covariância fica:

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

Estimando Elementos da Matriz de Covariância

Queremos estimar σ_u^2 por valores amostrais:

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

$$\sigma_{\ddot{u}}^2 = \mathrm{E}[(u_{it} - \overline{u}_i)^2] = \mathrm{E}(u_{it}^2) + \mathrm{E}(\overline{u}_i^2) - 2\mathrm{E}(u_{it}\overline{u}_i)$$

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

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

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

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

Ajustando os Graus de Liberdade (Cortando Ts e subtraindo K do número de regressores):

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

Teste para Autocorrelação AR(1)

Wooldridge (2010, p.275)

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

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

Vamos testar

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

(ausência de correlação em $\boldsymbol{u}$) na equação:

$$\widehat{\ddot{u}}_{it} = \delta\widehat{\ddot{u}}_{it-1} + e_{it}$$

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

8 Primeira Difereça (First Difference, FD, PD)

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

8.1 Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \qquad i = 1, \dots, N, \text{ e } \quad t = 1, \dots, T. \tag{8.1}$$

$$y_{i,t-1} = \boldsymbol{x}_{i,t-1}\boldsymbol{\beta} + c_i + u_{i,t-1},$$

$$\begin{aligned} y_{it} - y_{i,t-1} &= (\boldsymbol{x}_{it} - \boldsymbol{x}_{i,t-1})\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{i,t-1} \\ \Delta y_{it} &= \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it} \qquad i = 1, \dots, N, \text{ e } \quad t = 2, \dots, T. \end{aligned} \tag{8.2}$$

Definindo $e_{it} = \Delta u_{it}$, reescrvemos (8.2) no formato matricial empilhando T:

$$\Delta \boldsymbol{y}_i = \Delta \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{e}_i \qquad i = 1, \dots, N. \tag{8.3}$$

onde,
$\Delta \boldsymbol{y}_i$ é um vetor $(T-1) \times 1$
$\Delta \mathbf{X}_i$ é uma matriz $(T-1) \times K$
$\Delta \boldsymbol{x}_{it}$ é a $(t-1)$-ésima linha da matriz $\Delta \mathbf{X}_i$.
$\boldsymbol{\beta}$ é um vetor $K \times 1$
$\boldsymbol{e}_i$ é um vetor $(T-1) \times 1$

8.1.1 Matriz D

Vamos definir **D** como a matriz $(T-1) \times T$ como a matriz bidiagonal cuja diagonal inferior é -1 e a diagonal superior é 1. Assim,

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{bmatrix}$$

E podemos escrever $\Delta \boldsymbol{y}_i$ como:

$$\Delta \boldsymbol{y}_i = \mathbf{D} \boldsymbol{y}_i.$$

8.2 Estimação POLS

O estimador $\widehat{\boldsymbol{\beta}}^{FD}$ é o POLS da regressão no modelo (8.3), assim:

$$\boxed{\widehat{\boldsymbol{\beta}}^{FD} = \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \boldsymbol{y}_i \right)} \tag{8.4}$$

8.3 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 $\mathrm{E}(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i)$ (para inverter a matriz). $\text{posto}[\mathrm{E}(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i)] = K$.

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

8.4 Consistência

Usando (8.3) em (8.4):

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

FD.1 é suficiente para $\mathrm{E}(\Delta \mathbf{X}_i' \boldsymbol{e}) = \mathbf{0}$, $i = 1, \ldots, N$. Uma condição necessária para $\widehat{\boldsymbol{\beta}}^{FD} \xrightarrow{p} \boldsymbol{\beta}$ é $\mathrm{E}(\boldsymbol{x}_{it} u_{it}) = \mathrm{E}(\boldsymbol{x}_{it} u_{i,t-1}) = 0$, para $i = 1, \ldots, N$ e $t = 2, \ldots, T$. Note que sob **FD.1**, $\mathrm{E}(\widehat{\boldsymbol{\beta}}^{FD} | \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{it}, c_i) = \boldsymbol{\beta}$. Ou seja, $\widehat{\boldsymbol{\beta}}^{FD}$ é não viciado.

8.5 Variância

$$\mathrm{Cov}(\boldsymbol{\beta}^{FD}) = \mathrm{E}\left(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \mathrm{E}\left(\Delta \mathbf{X}_i' \boldsymbol{e}_i \boldsymbol{e}_i' \Delta \mathbf{X}_i \right) \mathrm{E}\left(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1}$$

Usando estimadores amostrais:

rever

$$\begin{aligned}
\widehat{\mathrm{Cov}}(\boldsymbol{\beta}^{FD}) &= \left(N^{-1} \sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \Delta \mathbf{X}_i' \boldsymbol{e}_i \boldsymbol{e}_i' \Delta \mathbf{X}_i \right) \left(N^{-1} \sum_{i=1}^{N} \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1} \\
&= N (\Delta \mathbf{X}' \Delta \mathbf{X}_i)^{-1} \left(\sum_{i=1}^{N} \Delta \mathbf{X}_i' \boldsymbol{e}_i \boldsymbol{e}_i' \Delta \mathbf{X}_i \right) (\Delta \mathbf{X}' \Delta \mathbf{X})^{-1}
\end{aligned}$$

onde $\Delta \mathbf{X}$ é a matriz $N(T-1) \times K$ das matrizes $\Delta \mathbf{X}_i$ empilhadas.

$$\Delta \mathbf{X} = \begin{bmatrix} \Delta \mathbf{X}_1 \\ \vdots \\ \Delta \mathbf{X}_N \end{bmatrix}$$

8.5.1 Variância sob Homocedasticidade

Usando **FD.3**, temos

$$\begin{aligned}
\mathrm{E}\left(\Delta \mathbf{X}_i' \boldsymbol{e}_i \boldsymbol{e}_i \Delta \mathbf{X}_i \right) &= \mathrm{E}\left[\mathrm{E}\left(\Delta \mathbf{X}_i' \boldsymbol{e}_i \boldsymbol{e}_i \Delta \mathbf{X}_i | \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iT}, c_i \right) \right] \\
&= \mathrm{E}\left(\Delta \mathbf{X}_i' \sigma_e^2 \mathbf{I}_{T-1} \Delta \mathbf{X}_i \right) = \sigma_e^2 \mathrm{E}\left(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)
\end{aligned}$$

Então, sob **FD.3**:

$$\widehat{\mathrm{Cov}}(\boldsymbol{\beta}^{FD}) = \sigma_e^2 \mathrm{E}\left(\Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right)^{-1}$$

onde

$$\sigma_e^2 = \frac{1}{N(T-1) - K} \sum_{i=1}^{N} \sum_{t=2}^{T} \widehat{e}_{it}^2.$$

8.5.2 Teste para autocorrelação AR(1) dos resíduos

A equação do erro e_{it} é

$$e_{it} = u_{it} - u_{i,t-1}$$

rearranjando os termos, encontramos

$$\boxed{u_{it} = u_{i,t-1} + e_{it}}, \qquad t = 2, \dots, T.$$

que é um passeio aleatório. Sob **FD.3**, $e_{it} \sim \mathrm{RB}(0, \sigma_e^2)$ e u_{it} é um passeio aleatório.

$$\hat{e}_{it} = \delta \hat{e}_{it-1} + \varepsilon_{it}, \qquad t = 2, \dots, T.$$

Note que sob **FD.1**,

$$\mathrm{E}\left(e_{it} | \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, c_i\right) = 0$$

$H_0 : \delta = 0$
$H_1 : \delta \neq 0$

Sob **FD.3**,

$$\begin{aligned}\mathrm{E}\left(e_{it}^2\right) &= \mathrm{E}\left[(u_{it} - u_{it-1})^2\right] \\ &= \mathrm{E}\left[u_{it}^2 + u_{it-1}^2 - 2u_{it}u_{it-1}\right] = 2\sigma_u^2\end{aligned}$$

$$\begin{aligned}\mathrm{E}\left(e_{it}e_{i,t-1}\right) &= \mathrm{E}\left[(u_{it} - u_{it-1})(u_{it-1} - u_{it-2})\right] \\ &= \mathrm{E}\left[u_{it}u_{it-1} - u_{it}u_{it-2} - u_{it-1}^2 + u_{it-1}u_{it-2}\right] \\ &= \mathrm{E}(-u_{it-1}^2) = -\mathrm{E}(u_{it-1}^2) = -\sigma_u^2\end{aligned}$$

$$\mathrm{E}\left(e_{it}e_{i,t-2}\right) = \mathrm{E}\left[(u_{it} - u_{it-1})(u_{it-2} - u_{it-3})\right] = 0$$

Assim, temos que, sob **FD.3**, $\mathrm{E}(\boldsymbol{e}_i\boldsymbol{e}_i')$ é uma matriz tridiagonal $XX \times XX$:

$$\mathrm{E}(\boldsymbol{e}_i\boldsymbol{e}_i') = \begin{bmatrix} 2\sigma_u^2 & -\sigma_u^2 & & & 0 \\ -\sigma_u^2 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -\sigma_u^2 \\ 0 & & & -\sigma_u^2 & 2\sigma_u^2 \end{bmatrix}$$

onde

$$\mathrm{Cov}(e_{it}, e_{it-1}) = \frac{-\sigma_u^2}{2\sigma_u^2} = \frac{-1}{2}.$$

Assim, podemos testar $H_0 : \delta = -1/2$ na seguinte equação

$$\hat{e}_{it} = \delta \hat{e}_{it-1} + \varepsilon_{it}.$$

Se **não rejeitar** H_0 temos evidência **favorável** a FE.

8.6 Teste de Exogeneidade Estrita

8.6.1 Teste para o Estimador FD

$\Delta y_{it} = \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{w}_{it}\gamma + e_{it}$

onde $\boldsymbol{w}_{it}$ é um subconjunto de $\boldsymbol{x}_{it}$ (excluindo as dummies de tempo).

Testamos

$H_0 : \boldsymbol{\gamma} = \mathbf{0}$

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

via teste de Wald (Robusto).

8.6.2 Teste para o Estimador FE

$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + \boldsymbol{w}_{it}\gamma + v_{it}$

Testamos

$H_0 : \boldsymbol{\gamma} = \mathbf{0}$

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

Sob H_0, podemos estimar o modelo via FE.

8.7 Alguns Detalhes

Remark. Como no modelo FE, perdemos as variáveis constantes no tempo. Por exemplo, uma variável cujo incremento é igual a 1 a cada período (e.g. experiência profissional), causa multicolinearidade perfeita, pois

$\Delta d_2 + \cdots + \Delta d_T = \Delta exper.$

8.7.1 Avaliação de Políticas

Se $T = 2$:

$y_{i1} = \boldsymbol{x}_{i1}\boldsymbol{\beta} + \delta prog_{i1} + v_{i1}$

$y_{i2} = \boldsymbol{x}_{i2}\boldsymbol{\beta} + \delta prog_{i2} + v_{i2}$

onde $prog_{i1} = 0$ para todo mundo. Vamos considerar

$\Delta y_{i2} = \Delta \boldsymbol{x}_{i2}\boldsymbol{\beta} + \delta prog_{i2} + e_{i2}$

como $prog_{i1} = 0$, então $prog_{i2} = prog_{i2} - prog_{i1} = \Delta prog_{i2}$.

Para $T > 2$:

$\Delta y_{i2} = \Delta \boldsymbol{x}_{i2}\boldsymbol{\beta} + \delta \Delta prog_{i2} + e_{i2}, \qquad t = 2, \ldots, T.$

Exemplo

$\log(scrap_{it}) = \boldsymbol{x}_{it}\boldsymbol{\beta} + \delta_1 grant_t + \delta_2 grant_{t-1} + v_{it}.$

9 System GLS (SGLS)

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

9.1 Modelo Linear

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

onde
$\boldsymbol{y}$ é um vetor $G \times 1$
$\mathbf{X}_i$ é uma matriz $G \times K$
$\boldsymbol{\beta}$ é um vetor $K \times 1$
$\boldsymbol{u}$ é um vetor $G \times 1$

9.2 Hipóteses

SGLS.1: $\mathrm{E}(\mathbf{X}_i \otimes \boldsymbol{u}_i) = \mathbf{0}_{G^2 \times K}$.
Isso implica que se $\mathbf{X}_i$ contém constante,

$$\mathrm{E}(u_{ig}) = 0, \quad g = 1, \dots, G.$$

Nesse caso, implica também que $\mathrm{E}(x_{ikg}u_{ig}) = 0$, $k = 1, \dots, K$, $g = 1, \dots, G$.

SGLS.2: $\boldsymbol{\Omega} = \mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_i')$.
Com $\boldsymbol{\Omega}$ positiva definida (para ter inversa).

SGLS.3: $\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right) = \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)$.
É suficiente para **SGLS.3** supor $\mathrm{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i'|\mathbf{X}_i\right) = \boldsymbol{\Omega}$, pois

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

9.3 Estimação

O estimador SOLS é consistente, nesse caso, mas sua matriz de covariância é dada por:

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}^{SOLS}) = \mathrm{E}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{u}_i\boldsymbol{u}_i'\mathbf{X}_i\right)\mathrm{E}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1},$$

supondo $\mathrm{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i'|\mathbf{X}_i\right) = \boldsymbol{\Omega}$,

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}^{SOLS}) = \mathrm{E}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i\right)\mathrm{E}\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1}.$$

É possível encontrar $\boldsymbol{\Omega}^{1/2}$ e $\boldsymbol{\Omega}^{-1/2}$ tais que:

$$\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2} = \boldsymbol{\Omega} \quad \text{e} \quad \boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}^{-1/2} = \boldsymbol{\Omega}^{-1}$$

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

$$\begin{aligned}
\boldsymbol{\Omega}^{-1/2}\boldsymbol{y}_i &= \boldsymbol{\Omega}^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i \\
\boldsymbol{y}_i^* &= \mathbf{X}_i^*\boldsymbol{\beta} + \boldsymbol{u}_i^* \qquad (9.1)
\end{aligned}$$

$$\mathrm{E}\left(\boldsymbol{u}_i^* \boldsymbol{u}_i^{*\prime}\right) = \mathrm{E}\left(\boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{\Omega}^{-1/2}\right) = \boldsymbol{\Omega}^{-1/2} \underbrace{\mathrm{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i'\right)}_{\boldsymbol{\Omega}=\boldsymbol{\Omega}^{1/2}\boldsymbol{\Omega}^{1/2}} \boldsymbol{\Omega}^{-1/2} = \mathbf{I}_G.$$

$$\mathrm{E}\left(\mathbf{X}_i^{*\prime}\boldsymbol{u}_i^*\right) = \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1/2}\boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i\right) = \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{u}_i\right) = \mathbf{0}_{K\times 1}.$$

Então, o estimador SOLS aplicado ao modelo transformado (9.1) nos fornece um estimador consistente e eficiente.

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

No formato puramente matricial:

$$\boxed{\widehat{\boldsymbol{\beta}}^{SGLS} = \left[\mathbf{X}\left(\mathbf{I}_N\otimes\boldsymbol{\Omega}^{-1}\right)\mathbf{X}\right]^{-1}\left[\mathbf{X}\left(\mathbf{I}_N\otimes\boldsymbol{\Omega}^{-1}\right)\boldsymbol{y}\right]}$$

9.4 Variância

$$\begin{aligned}
\mathrm{Avar}(\widehat{\boldsymbol{\beta}}^{SGLS}) &= \mathrm{E}\left(\mathbf{X}_i^{*\prime}\mathbf{X}_i^*\right)^{-1}\mathrm{E}\left(\mathbf{X}_i^{*\prime}\boldsymbol{u}_i^*\boldsymbol{u}_i^{*\prime}\mathbf{X}_i^*\right)\mathrm{E}\left(\mathbf{X}_i^{*\prime}\mathbf{X}_i^*\right)^{-1} \\
&= \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)^{-1}\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{u}_i\boldsymbol{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)^{-1}.
\end{aligned}$$

Sob **SGLS.3:**

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}^{SGLS}) = \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)^{-1}\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)\mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)^{-1} = \mathrm{E}\left(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right).$$

9.5 Esimtação FSGLS (SGLS Factível)

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

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

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

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

   Note que $\widehat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$.

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

$$\widehat{\boldsymbol{\beta}}^{FGLS} = \left[\sum_{i=1}^N \mathbf{X}_i'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right]^{-1}\left[\sum_{i=1}^N \mathbf{X}_i'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}_i\right] \tag{9.2}$$

Empilhando as N observações, podemos expressar $\widehat{\boldsymbol{\beta}}^{FGLS}$ matricialmente:

$$\widehat{\boldsymbol{\beta}}^{FGLS} = \left[\mathbf{X}'\left(\mathbf{I}_N\otimes\widehat{\boldsymbol{\Omega}}^{-1}\right)\mathbf{X}\right]^{-1}\left[\mathbf{X}'\left(\mathbf{I}_N\otimes\widehat{\boldsymbol{\Omega}}^{-1}\right)\boldsymbol{y}\right] \tag{9.3}$$

9.6 Consistência FGLS

Reescrevendo a equação (9.3):

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

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

Então, $\widehat{\boldsymbol{\beta}}^{FSGLS} \xrightarrow{p} \boldsymbol{\beta}$, se $\widehat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$.

9.7 Variância FGLS

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}^{FGLS}) = \left[\mathbf{X}'\left(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\mathbf{X}\right]^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{u}}_i\widehat{\boldsymbol{u}}_i'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)\left[\mathbf{X}'\left(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\mathbf{X}\right]^{-1}$$

onde $\widehat{\boldsymbol{u}}_i = \boldsymbol{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}^{FGLS}$.

Sob Homocedasticidade (SGLS.3)

$$\widehat{\mathrm{Var}}(\widehat{\boldsymbol{\beta}}^{FGLS}) = \left[\mathbf{X}'\left(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\mathbf{X}\right]^{-1}$$

9.8 Testes

Podemos fazer o teste t e de Wald.
Além disso, podemos comparar o modelo restrito e irrestrito.

$$\text{Est. Teste} = \left[\widetilde{\boldsymbol{u}}'\left(\mathbf{I} \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\widetilde{\boldsymbol{u}} - \widehat{\boldsymbol{u}}'\left(\mathbf{I} \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\widehat{\boldsymbol{u}}\right]$$

Sob H_0, Est. Teste $\sim \chi^2_Q$. Onde,

$\widetilde{\boldsymbol{u}}$ é o vetor de resíduos FGLS do modelo restrito.
$\widehat{\boldsymbol{u}}$ é o vetor de resíduos FGLS do modelo irrestrito.
$\widehat{\boldsymbol{\Omega}}$ é construída com os resíduos SOLS do modelo irrestrito.
Q é o número de restrições.

10 Modelo de Efeitos Aleatórios (EA, Random Effects, RE)

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

10.1 Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \qquad t = 1, \dots, T \quad \text{e} \quad i = 1, \dots, N. \tag{10.1}$$
$$\boldsymbol{y}_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{1} + \boldsymbol{u}_i, \qquad i = 1, \dots, N. \tag{10.2}$$
$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + c_i\mathbf{1} + \boldsymbol{u}. \tag{10.3}$$

10.2 Hipóteses

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

Hipótese (RE.1). Usamos o modelo correto e c_i não é endógeno.

a) $\mathrm{E}(u_{it} \,|\, \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, c_i) = 0, \qquad i = 1, \dots, N.$

b) $\mathrm{E}(c_i \,|\, \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}) = \mathrm{E}(c_i) = 0, \qquad i = 1, \dots, N.$

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

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

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

Hipótese (RE.3).

a) $\mathrm{E}(\boldsymbol{u}_i\boldsymbol{u}_i' \,|\, \boldsymbol{x}_{i1}, \dots, \boldsymbol{x}_{iT}, c_i) = \sigma_u^2\mathbf{I}_T;$

b) $\mathrm{E}(c_{it} \,|\, x_{i1}, \dots, x_{iT}) = \mathrm{E}(c_i) = \sigma_c^2.$

Remark.

RE.1 (a) está presente em FE e FD.
RE.1 (b) está presente em POLS e implica ausência de correlação entre c_i e os regressores do modelo.

10.3 Estimação

Vamos considerar o modelo:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}, \qquad t = 1, \dots, T \quad \text{e} \quad i = 1, \dots, N.$$
$$\boldsymbol{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i, \qquad i = 1, \dots, N.$$
$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}.$$

Sob **RE.3:**

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

Então,

$$\mathrm{E}(\boldsymbol{v}_i\boldsymbol{v}_i') = \begin{bmatrix} \sigma_c^2+\sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2+\sigma_u^2 & \dots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \dots & \sigma_c^2+\sigma_u^2 \end{bmatrix} = \begin{bmatrix} \sigma_c^2+\sigma_u^2 & & \sigma_c^2 \\ & \ddots & \\ \sigma_c^2 & & \sigma_c^2+\sigma_u^2 \end{bmatrix} \equiv \boldsymbol{\Omega}$$

Também poderíamos escrever:

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

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

10.3.1 Estimador RE

$$\widehat{\boldsymbol{\beta}}^{RE} = \left[\mathbf{X}'\left(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\mathbf{X}\right]^{-1}\left[\mathbf{X}'\left(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1}\right)\boldsymbol{y}\right] \tag{10.4}$$

onde

$$\widehat{\boldsymbol{\Omega}} = \begin{bmatrix} \widehat{\sigma}_v^2 & & \widehat{\sigma}_c^2 \\ & \ddots & \\ \widehat{\sigma}_c^2 & & \widehat{\sigma}_v^2 \end{bmatrix}$$

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

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

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

e $\hat{v}_{it}$ é o resíduo de POLS do modelo (10.1)

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

10.4 Inferência

Robusta:

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}^{RE}) = \mathrm{E}\left\{\left[\mathbf{X}'(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1})\mathbf{X}\right]^{-1}\left[\sum_{i=1}^{N}\mathbf{X}_i'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{v}}_i\widehat{\mathbf{v}}_i'\widehat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right]\left[\mathbf{X}'(\mathbf{I}_N \otimes \widehat{\boldsymbol{\Omega}}^{-1})\mathbf{X}\right]^{-1}\right\},$$

Sob Homocedasticidade:

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

10.5 Testes

Teste para presença de componente não observável

$H_0 : \sigma_c^2 = 0$

$H_1 : \sigma_c^2 \neq 0$

Estatística do teste:

$$\text{Est. Teste} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}}{\left[\sum_{i=1}^{N}\left(\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{v}_{it}\hat{v}_{is}\right)^2\right]^{1/2}}$$

Sob H_0, Est. Teste $\sim \mathcal{N}(0,1)$.

Teste de Hausman (FE v. RE)

H_0 : Tanto FE, quanto RE são consistentes
H_1 : Apenas FE é consistente

Estatística do teste:

$$\text{Est. Teste} = \left(\widehat{\boldsymbol{\delta}}_{FE} - \widehat{\boldsymbol{\delta}}_{RE}\right) \left[\widehat{\text{Var}}(\widehat{\boldsymbol{\delta}}_{FE}) - \widehat{\text{Var}}(\widehat{\boldsymbol{\delta}}_{RE})\right]^{-1} \left(\widehat{\boldsymbol{\delta}}_{FE} - \widehat{\boldsymbol{\delta}}_{RE}\right)$$

onde $\widehat{\boldsymbol{\delta}}$ é um vetor $M \times 1$ e M é o número de regressores que variam no tempo.
Sob H_0,

$$\text{Est. Teste} \overset{a}{\sim} \chi^2_M.$$

11 Endogeneity and GMM

11.1 Modelo

No seguinte modelo *cross-section*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \; ; \quad i = 1, \dots, N. \tag{11.1}$$

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 (11.1) vamos supor:

- x_1 é exógena.
- x_2 é endógena.

11.2 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 (11.1).
2. $Cov(z, x_2) \neq 0 \implies$ correlação com x_2 após controlar para outras variáveis.

11.3 Estimação

Indo para o problema de dados de painel, temos:

$$\boldsymbol{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{u}_i \; ; \quad i = 1, \dots, N. \tag{11.2}$$

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

Se é verdade que há endogeneidade em (11.2), então:

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

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).

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

Se Usarmos estimadores amostrais:

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

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

Se $L > K$, vamos considerar:

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

onde:

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

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

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

Um estimador mais eficiente pode ser encontrado fazendo:

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

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

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

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

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

11.4 Valor Esperado

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

11.5 Variância

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

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

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

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

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

12 Exogeneidade Estrita e FDIV

12.1 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:

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

$$\boxed{\Delta y_{it} = \Delta\boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho\Delta y_{it-1} + \Delta u_{it} \;, \qquad t = 3, \ldots, T} \tag{12.1}$$

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

12.2.1 P2SLS

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

- $i = 1, \ldots, N$
- $t = 1, \ldots, T$
- 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.

12.2.2 FDIV

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

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:

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

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

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

12.3 Hipóteses

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

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

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

12.4 Estimação FDIV

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

12.5 Valor Esperado

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

12.6 Variância

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

$e_i = \Delta u_{it}$.

13 Latent Variables, Probit and Logit

13.1 Modelo

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

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i. \tag{13.1}$$

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 (13.1), 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:

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

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

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

14 ATT, ATE, Propensity Score

14.1 Modelo

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

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

Idealmente, para isolarmos completamente o efeito de $w = 1$, gostaríamos de 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).

14.2 ATE e ATT condicional a variáveis dependentes

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

<u>OBS:</u>

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

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

$p(\boldsymbol{x})$ 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': $\mathrm{E}(y_0 \,|\boldsymbol{x}, w) = \mathrm{E}(y_0 \mid \boldsymbol{x})$

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

14.4 Propensity Score

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

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

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

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

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:

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

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

$\mu_0(\boldsymbol{x})$

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

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:

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

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

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

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

15 Álgebra Linear

15.1 Vetores

15.2 Operações com Vetores

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

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

Multiplicação por escalar

Soma

Subtração

Multiplicação de vetores

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

$$\boldsymbol{x}'\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} = \langle \boldsymbol{x} \rangle \boldsymbol{y} = \sum_{i=1}^{N} x_i y_i = x_1 y_1 + \cdots + x_N y_N. \tag{15.1}$$

Podemos utilizar a equação (15.1) para denotarmos a soma dos elementos de um vetor. Para tanto, definimos o vetor $\mathbf{1}_N$ como sendo o vetor cujos elementos são 1 e tem dimensão $N \times 1$. (Greene, 2012, p. 977, A.2.7)

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

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

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

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

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

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

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

Poduto Externo (Outer Product)

$$\mathbf{1}_N\mathbf{1}_N' = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{N\times N} \qquad \mathbf{1}_N\boldsymbol{x}' = \begin{bmatrix} x_1 & \dots & x_N \\ \vdots & \ddots & \vdots \\ x_1 & \dots & x_N \end{bmatrix}_{N\times N} \qquad \boldsymbol{x}\mathbf{1}_N' = \begin{bmatrix} x_1 & \dots & x_1 \\ \vdots & \ddots & \vdots \\ x_N & \dots & x_N \end{bmatrix}_{N\times N}$$

$$\boldsymbol{x}\boldsymbol{x}' = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} x_1 & \dots & x_N \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1x_2 & \dots & x_1x_N \\ x_2x_1 & x_2^2 & \dots & x_2x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_Nx_1 & x_Nx_2 & \dots & x_N^2 \end{bmatrix} \tag{15.2}$$

Distância e Ângulo O coseno do ângulo, α, entre dois vetores é:

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

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

Projeções Ortogonal

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

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

$$\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \frac{\boldsymbol{y}'\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{u}}\boldsymbol{u}$$

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

Por construção, temos:
$\boldsymbol{z} + \widehat{\boldsymbol{y}} = \boldsymbol{y}$.

[FIGURA AQUI]

Outras Notações Projeções Ortogonal Dado que

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

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

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

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

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

Agora, podemos definir o operador rejeição como:

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

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

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

A **rejeição** no eixo de 45 graus:

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

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

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

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

Idempotência: $AA = A$

Simetria: $A' = A$

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

onde podemos denotar

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

$$\mathbf{M}^0\boldsymbol{x} = \boldsymbol{x} - \overline{\boldsymbol{x}}$$

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

15.3 Operações com Matrizes

Multiplicação por escalar

Soma

Subtração

Multiplicação de Matriz

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

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

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

15.3.1 Projeções

$$\mathbf{P_X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$
$$\mathbf{M_X} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$$

Usando o modelo

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

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

$$\boldsymbol{P}_{\mathbf{X}}\boldsymbol{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \mathbf{X}\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{y}}$$
$$\boldsymbol{M}_{\mathbf{X}}\boldsymbol{y} = \mathbf{I}\boldsymbol{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y} = \boldsymbol{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \widehat{\boldsymbol{u}}$$

16 Conceitos Básicos de Convergência Estatística

Definition 16.1 (**Convergência em Probabilidade**)**.** Wooldridge (2010, Def 3.3, p.36)
Uma sequência de variáveis aleatórias: $\{X_N\}_{N\geq 1}$ **converge em probabilidade** para uma variável aleatória X se, dado $\varepsilon > 0$,

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

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

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

Definition 16.2 (**Estimador Consistente**)**.** Wooldridge (2010, Def 3.8, p.40)
Seja $\{\boldsymbol{\theta}_N : N = 1, 2, \dots\}$ uma sequência de estimadores do vetor $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ com dimensão $P \times 1$, onde N indexa o tamanho da amostra. Se

$$\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta} \tag{16.1}$$

Para qualquer valor de $\boldsymbol{\theta}$, então dizemos que $\boldsymbol{\theta}_N$ é um estimador consistente de $\boldsymbol{\theta}$.

Theorem 16.1 (**LGN – Lei dos Grandes Números**)**.** Achar referência
Seja $\{X_i\}_{i\geq 1}$ uma sequência de variáveis aleatórias *iid* com $\mathrm{E}(X_i) = \mu$. Então,

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

Theorem 16.2 (**LGN – Caso Matricial**)**.** Achar referência. O Teorema (16.3), abaixo, é diferente.
Seja $\{\boldsymbol{x}_i\}_{i=1}^{N}$, uma sequência *iid* de vetores aleatórios $K \times 1$ com $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i') = Q_{K\times K}$ finita. Então,

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

Se Q for positiva definida, Q terá inversa.

Theorem 16.3 (**LGNF – WLLN**)**.** Wooldridge (2010, Teo 3.1, p.39)
Seja $\{\boldsymbol{w}_i : i = 1, 2, \dots\}$, uma sequência *iid* de vetores aleatórios $G \times 1$ com $\mathrm{E}(|w_{ig}|) < \infty$ para $g = 1, \dots, G$. Então, a seguência satisfaz a **Lei dos Grandes Números Fraca (WLLN)**:

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

onde $\boldsymbol{\mu}_w \equiv \mathrm{E}(\boldsymbol{w}_i)$.

Definition 16.3 (o_p)**.**

Wooldridge (2010, Def 3.4, p.36)
Wooldridge (2010, Lemma 3.2, p.36)

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

Definition 16.4 (**Limitação em Probabilidade:** O_p)**.** Wooldridge (2010, Def 3.3 (3), p.36)
Dizemos que X_n é **limitado em probabilidade** e denotado por $X_n = \mathrm{O}_p(1)$, se existe M maior que zero, tal que para todo ε maior que zero, $P(|X_n| > 0) < \varepsilon$.

$$X_n = \mathrm{O}_p(1) \implies \exists M > 0\,;\ \forall \varepsilon > 0\,,\ P(|X_n| > 0) < \varepsilon.$$

Definition 16.5. Wooldridge (2010, Def 3.4, p.36)
Dizemos que $X_n = \mathrm{O}_p(Y_n)$ se existe M maior que zero, tal que para todo ε maior que zero, $P(|X_n/Y_n| > 0) < \varepsilon$.

$$X_n = \mathrm{O}_p(Y_n) \implies \exists M > 0\,;\ \forall \varepsilon > 0\,,\ P(|X_n/Y_n| > 0) < \varepsilon.$$

Lemma 16.1. Wooldridge (2010, Lemma 3.2, p.36)
Se $X_n = \mathrm{O}_p(1)$ e $Y_n = \mathrm{o}_p(1)$, então

$$X_n Y_n = \mathrm{O}_p(1)\mathrm{o}_p(1) = \mathrm{o}_p(1).$$

Lemma 16.2 (**Equivalência Assintótica**)**.** Wooldridge (2010, Lemma 3.7, p.39)
Seja $\{\boldsymbol{x}_n\}$ e $\{\boldsymbol{z}_n\}$ sequências de vetores aleatórios $K \times 1$. Se $\boldsymbol{z}_n \xrightarrow{d} \boldsymbol{z}$ e $\boldsymbol{x}_n - \boldsymbol{z}_n \xrightarrow{p} \mathbf{0}_K$. Então,

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

Definition 16.6 (**Convergência em Distribuição**)**.** Wooldridge (2010, Def 3.6, p.38)
Seja $\{X_n\}_{n\geq 1}$ uma sequência de variáveis aleatórias e X uma variável aleatória com F_n e F suas respectivas FDAs, então

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

para todo X onde F é contínuo.

Lemma 16.3 (**Convergência em Distribuição e Limitação em Probabilidade**)**.** Wooldridge (2010, Lemma 3.5, p.39)
Se $X_n \xrightarrow{d} X$, X um variável aleatória qualquer; então $X_n = \mathrm{O}_p(1)$.

Definition 16.7 (**TCL – Teorema Central do Limite**)**.** Achar Referência
Seja $\{X_n\}_{n=1}^N$ *iid* com $\mathrm{E}(X_n) = \mu$ e $\mathrm{Var}(X_n) = \sigma^2 < +\infty$. Então, para $S_N = \sum_{n=1}^N X_n$:

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

Theorem 16.4 (**TCL – Lindeberg-Levy**)**.** Wooldridge (2010, Teo 3.2, p.40)
Seja $\{\boldsymbol{w}_i : i = 1, 2, \dots\}$ uma sequência *iid* de vetores aleatórios $G \times 1$ com $\mathrm{E}(w_{ig}^2) < \infty$ para $g = 1, \dots G$ e $\mathrm{E}(\boldsymbol{w}_i) = \mathbf{0}$. Então, $\{\boldsymbol{w}_i : i = 1, 2, \dots\}$ satisfaz o **Teorema Central do Limite (CLT)**; qual seja:

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

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

Corollary 16.1. Wooldridge (2010, Cor 3.2, p.39)

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

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

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

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

Definition 16.8 (raiz de N assintoticamente normalmente distribuido). Wooldridge (2010, Def 3.9, p.40)

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

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{V}) \tag{16.2}$$

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

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

$$\hat{\boldsymbol{\theta}}_N \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{V}/N). \tag{16.3}$$

sempre que a equação (16.2) for verdade. Por essa razão, $\mathbf{V}/N$ é chamado de **variância assintótica** de $\hat{\boldsymbol{\theta}}_N$, e escrevemos:

$$\text{Avar}(\hat{\boldsymbol{\theta}}_N) = \mathbf{V}/N. \tag{16.4}$$

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

Definition 16.9 (Desigualdade de Markov). Achar referência

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

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

Definition 16.10. Achar referência

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

Definition 16.11 (Erro Quadrático Médio). Achar referência

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

Remark. Achar referência?

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

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, Massachussetts: MIT Press.