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

NOTAS DE AULA

Autor: Paulo Ferreira Naibert
Professor: Hudson Torrent

Porto Alegre
30/06/2020
Revisão: 2 de julho de 2020

1 Regressão MQO Clássico

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

Modelo de equações lineares

O modelo populacional que estudamos é linear em seus parâmetros,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_K x_K + u$$

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.

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

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

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:

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

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

Empilhando as N observações, obtemos a **notação matricial**:

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

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

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

Hipóteses

H1 Estamos usando o modelo correto:

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

H2 X é **não** estocástica;

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

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

H2' X é estocástica;

H3'

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

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

Estimação

$$\widehat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y} \tag{1.1}$$

Valor Esperado

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

Viés

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

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

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

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

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

Variância Supondo **H2'** e **H3'**:

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

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

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

Ausência de Exogeneidade Estrita

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

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

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

Estimação Lembrando que o estimador de OLS é:

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

onde usamos o modelo

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

e definimos as variáveis

$$X_{N\times K} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_N \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad \boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix}.$$

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

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

Valor Esperado e Viés

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

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

$$\boxed{N^{-1}\sum_{i=1}^{N} \boldsymbol{x}_i'\boldsymbol{x}_i \xrightarrow{p} Q}. \tag{1.2}$$

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

Então,

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

Logo

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

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

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

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

Supondo

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

temos, pelo **TCL**, que

$$\boxed{N^{-1/2}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i \xrightarrow{d} N(0,B)} \tag{1.3}$$

onde

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

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

Além disso, vamos utilizar a matriz **simétrica** e **não singular** Q da equação (1.2) Assim, temos

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

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

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

Então,

$$\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = 0_p(1)O_p(1) + Q^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\boldsymbol{x}_i'u_i\right),$$

Usando (1.3) e a definição XX

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

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

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

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

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

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

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

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

Estimador Amostral

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

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

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

A variância **Robusta** é:

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

A variância sob **Homocedasticidade** é:

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

Conceitos Básicos de Convergência Estatística

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

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

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

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

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

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

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

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

Definição.

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

Definição (Erro Quadrático Médio).

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

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

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

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

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

Multiplicação de Matriz

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

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

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

Definição.

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

Definição (o_p)**.**

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

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

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

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

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

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

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

Definição (Equivalência Assintótica)**.** *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,*

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

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

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

para todo X onde F é contínuo.

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

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

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

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

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

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

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

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

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

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

2 Wooldridge

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

$$y = \boldsymbol{x\beta} + u \tag{2.1}$$

OLS.1 Population Orthogonality Condition: $E(\boldsymbol{x}'u) = 0$

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

$$\boldsymbol{\beta} = \left[E(\boldsymbol{x}'\boldsymbol{x})\right]^{-1} E(\boldsymbol{x}'y)$$

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

Isso pode ser escrito na forma matricial $(X'X)^{-1}X'\boldsymbol{y}$, onde X é a matriz de dados $N \times K$ dos regressores com linha i igual a $\boldsymbol{x}_i$, e $\boldsymbol{y}$ é o vetor de dados $N \times 1$ com o i=ésimo elemento de $\boldsymbol{y}$ sendo representado por y_i.

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

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

3 System OLS (SOLS)

Wooldridge (2010, C.7 – The Single-Equation Linear Model and OLS Estimation)
Wooldridge (2010, C.10 – The Single-Equation Linear Model and OLS Estimation)

Hipóteses

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

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

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

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

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

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

Estimação

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

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

Estimando a equação acima por **SOLS**:

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

FSGLS: SGLS Factível

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

1. Estimar $\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{u}_i$ via **SOLS** e guardar o resíduo estimado $\widehat{\boldsymbol{u}}_i$.

2. Estimar Ω com o seguinte estimador $\widehat{\Omega}$:

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

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

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

Empilhando as N observações:

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

Reescrevendo a equação acima:

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

Valor Esperado

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

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

Variância

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

Tirando o valor Esperado e supondo que:

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

temos:

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

e temos:

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

4 Endogeneity and GMM

Modelo

No seguinte modelo *cross-section*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \; ; \quad i = 1, \ldots, N. \tag{4.1}$$

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 (4.1) vamos supor:

- x_1 é exógena.
- x_2 é endógena.

Hipóteses

Assim, precisamos encontrar um instrumento z_i para x_2, uma vez que queremos estimar β_0, β_1 e β_2 de maneira consistente. Para z_i ser um bom instrumento precisamos que z tenha:

1. $Cov(z, \varepsilon) = 0 \implies z$ é exógena em (4.1).
2. $Cov(z, x_2) \neq 0 \implies$ correlação com x_2 após controlar para outras vaariáveis.

Estimação

Indo para o problema de dados de painel, temos:

$$\boldsymbol{y}_i = X_i \boldsymbol{\beta} + \boldsymbol{u}_i \; ; \quad i = 1, \ldots, N. \tag{4.2}$$

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

Se é verdade que há endogeneidade em (4.2), então:

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

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

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

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

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

Se Usarmos estimadores amostrais:

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

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

Se $L > K$, vamos considerar:

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

onde:

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

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

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

Um estimador mais eficiente pode ser encontrado fazendo:

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

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

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

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

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

Valor Esperado

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

Variância

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

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

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

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

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

5 Random Effects (RE, EA)

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{5.1}$$

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

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

Podemos reescrever (5.1) como:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}, \tag{5.2}$$

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

Agora, vamos empilhar os t's e reescrever (5.2) como:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{v}_i, \tag{5.3}$$

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

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

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

1. Usamos o modelo correto e c_i não é endógeno.
   [label =)]$E(u_{it} \,|\, x_{i1}, \dots, x_{iT}, c_i) = 0$, $i = 1, \dots, N$. $E(c_{it} \,|\, x_{i1}, \dots, x_{iT}) = E(c_i) = 0$, $i = 1, \dots, N$.

(b2) Posto completo de $E(X_i'\Omega^{-1}X_i)$.

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

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

Estimação

Premultiplicando (5.3) port $\Omega^{-1/2}$ do dois lados, temos:

$$\begin{aligned} \Omega^{-1/2}\boldsymbol{y}_i &= \Omega^{-1/2}X_i\boldsymbol{\beta} + \Omega^{-1/2}\boldsymbol{v}_i \\ \boldsymbol{y}_i^* &= X_i^*\boldsymbol{\beta} + \boldsymbol{v}_i^*, \end{aligned} \tag{5.4}$$

Estimando o modelo acima por POLS:

$$
\begin{aligned}
\boldsymbol{\beta}^{POLS} &= \left(\sum_{i=1}^{N} X_i^{*\prime} X_i^{*}\right)^{-1} \left(\sum_{i=1}^{N} X_i^{*\prime} \boldsymbol{y}_i^{*}\right) \\
&= \left(\sum_{i=1}^{N} X_i^{\prime} \Omega^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X_i^{\prime} \Omega^{-1} \boldsymbol{y}_i\right) \\
&= \left(X'(I_N \otimes \Omega^{-1})X\right)^{-1} \left(X'(I_N \otimes \Omega^{-1})\boldsymbol{y}\right). \qquad (5.5)
\end{aligned}
$$

O problema, agora, é estimar Ω. Supondo:

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

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

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

Assim,

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

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

Agora, rodando POLS em (5.3) e guardando os resíduos, temos:

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

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

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

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

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

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

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

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

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

Com $\widehat{\Omega}$ estimado, reescrevemos (5.5) como:

$$
\boldsymbol{\beta}^{RE} = \left[X'(I_N \otimes \widehat{\Omega}^{-1})X\right]^{-1} \left[X'(I_N \otimes \widehat{\Omega}^{-1})\boldsymbol{y}\right]. \qquad (5.6)
$$

Valor Esperado

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

Variância

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

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

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

6 Fixed Effects (EF, FE)

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{6.1}$$

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

O modelo contém explicitamente um componente não observado que não varia no tempo c_i. Abordamos esse componente como parte do erro, não como parâmetro a não observado. No caso da análise de **Efeitos Fixos (EF, FE)**, permitimos que esse componente c_i seja correlacionado com $\boldsymbol{x}_{it}$. Assim, se decidíssemos estimar o modelo (6.1) por POLS, ignorando c_i, teríamos problemas de inconsistência devido a **endogeneidade**.

As T equações do modelo (6.1) podem ser reescritas como:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + c_1\mathbf{1}_T + \boldsymbol{u}_i, \tag{6.2}$$

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

Matriz M^0

Definimos a matriz M^0 como:

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

A matriz M^0 é idempotente e simétrica.

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

Podemos transformar o modelo (6.3) ao premultiplicarmos todo o modelo por M^0.

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

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

$$\ddot{\boldsymbol{y}}_i = \ddot{X}_i\boldsymbol{\beta} + \ddot{\boldsymbol{u}}_i, \quad i = 1, \dots, N. \tag{6.3}$$

Estimação POLS

Aplicando POLS no modelo (6.3)

$$\boxed{\boldsymbol{\beta}^{FE} = \left[\sum_{i=1}^{N} \ddot{X}_i'\ddot{X}_i\right]^{-1}\left[\sum_{i=1}^{N} \ddot{X}_i'\ddot{\boldsymbol{y}}_i\right]} \tag{6.4}$$

Hipóteses

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

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

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

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

Valor Esperado

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

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

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

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

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

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

Variância

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

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

Pão:

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

Recheio:

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

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

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

Variância sob Homocedasticidade

Usando **FE.3**, temos

Recheio':

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

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

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

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

7 First Difference (FD, PD)

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{7.1}$$

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

O modelo contém explicitamente um componente não observado, c_i, que não varia no tempo. Tratamos o componente não observado como parte do erro, não como parâmetro a ser estimado. Aqui permitimos que c_i seja correlacionado com $\boldsymbol{x}_{it}$. Deste modo, **não** podemos ignorar a sua presença e estimar (7.1) por POLS, visto que isso resultaria num estimador inconsistente devido a **endogeneidade**.

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

- Reescrevemos (7.1) defasado:

$$y_{it-1} = \boldsymbol{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \tag{7.2}$$

- Tiramos a diferença entre (7.2) e (7.1):

$$\begin{aligned} y_{it} - y_{it-1} &= (\boldsymbol{x}_{it} - \boldsymbol{x}_{it-1})\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{it-1} \\ \Delta y_{it} &= \Delta \boldsymbol{x}_{it}\boldsymbol{\beta} + \Delta u_{it}. \end{aligned} \tag{7.3}$$

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

Reescrevendo (7.3) no formato matricial empilhando T:

$$\Delta \boldsymbol{y}_i = \Delta X_i \boldsymbol{\beta} + \boldsymbol{e}_i \tag{7.4}$$

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

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

Estimação POLS

O estimador $\widehat{\boldsymbol{\beta}}^{FD}$ é o POLS da regressão no modelo (7.4), assim:

$$\boxed{\boldsymbol{\beta}^{FD} = \left[\sum_{i=1}^{N} \Delta X_i' \Delta X_i\right]^{-1} \left[\sum_{i=1}^{N} \Delta X_i' \Delta \boldsymbol{y}_i\right]} \tag{7.5}$$

Hipóteses

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

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

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

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

Valor Esperado

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

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

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

Variância

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

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

Variância sob Homocedasticidade

Usando **FD.3**, temos

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

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

com

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

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

8 Exogeneidade Estrita e FDIV

Modelo

No seguinte modelo

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

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

- 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:

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

ou seja:

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

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

Para o seguinte modelo:

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

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

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

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

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

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

$$\begin{aligned} y_{it} - y_{it-1} &= (\boldsymbol{z}_{it} - \boldsymbol{z}_{it-1})\boldsymbol{\gamma} + \rho(y_{it-1} - y_{it-2}) + (c_i - c_i) + (u_{it} - u_{it-1}) \\ &\boxed{\Delta y_{it} = \Delta \boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho \Delta y_{it-1} + \Delta u_{it} \ , \qquad t = 3, \dots, T} \end{aligned} \tag{8.1}$$

Estimação

Não podemos estimar o modelo (8.1) 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 $y_{it-2}, y_{it-3}, \dots, y_{i1}$).

P2SLS

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

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

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

com

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

onde P_Z é a matriz de projeção em Z.

FDIV

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

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

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

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

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

Hipóteses

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

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

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

Estimação FDIV

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

Valor Esperado

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

Variância

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

$e_i = \Delta u_{it}$.

9 Latent Variables, Probit and Logit

Modelo

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

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i. \tag{9.1}$$

Defina y como:

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

temos que:

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

Além disso, pela definição de y_i, equação (9.1), 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:

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

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

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

10 ATT, ATE, Propensity Score

Modelo

- $y_1 \to$ variável de interesse com tratamento
- $y_0 \to$ 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: $E(y_1 - y_0)$

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

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

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

<u>OBS:</u>

$$E(y_1 - y_0) = E\left[E(y_1 - y_0 \,|w)\right]$$
$$E(y_1 - y_0 \,|w) = E(y_1 - y_0 \,|w = 0) \cdot P(w = 0) + E(y_1 - y_0 \,|w = 1) \cdot P(w = 1).$$

Métodos Assumindo Ignorabilidade do Tratamento

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

ATE.1': Ignorabilidade da Média.
[label =)]$E(y_0 \,|\, w, \boldsymbol{x}) = E(y_0 \,|\, \boldsymbol{x})$ $E(y_1 \,|\, w, \boldsymbol{x}) = E(y_1 \,|\, \boldsymbol{x})$

Vamos definir

$$E(y_0 \,|\, \boldsymbol{x}) = \mu_0(\boldsymbol{x})$$
$$E(y_1 \,|\, \boldsymbol{x}) = \mu_1(\boldsymbol{x}).$$

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

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

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

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

Propensity Score

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

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

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

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

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

$\mu_1(\boldsymbol{x})$

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

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

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

$\mu_0(\boldsymbol{x})$

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

ATE:

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

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

ATT:

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

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

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

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

Appêndice

Sums of Values

(Greene, 2012, p. 977, A.2.7)

$$\mathbf{1}_N'\mathbf{1}_N = N \quad ; \qquad \mathbf{1}_N\mathbf{1}_N' = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{N\times N}$$

Defining $\boldsymbol{x}$ with dimension $1 \times N$:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

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

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

$$E(\boldsymbol{x}) = \overline{\boldsymbol{x}} = N^{-1}\sum_{i=1}^{N} x_i = N^{-1}\boldsymbol{x}'\mathbf{1}_N$$

Important Idempotent Matrices

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

Centering Matrix

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

A Matriz M^0 é **idempotente** e **simétrica**.

Idempotência: $AA = A$

Simetria: $A' = A$

$$M^0\boldsymbol{x} = (I_N - N^{-1}\mathbf{1}_N\mathbf{1}_N')\boldsymbol{x} = \boldsymbol{x} - N^{-1}\mathbf{1}_N(\mathbf{1}_N'\boldsymbol{x}) = \mathbf{1}_N\overline{\boldsymbol{x}} = \begin{bmatrix} \overline{\boldsymbol{x}} \\ \vdots \\ \overline{\boldsymbol{x}} \end{bmatrix}$$

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

Referências

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.