

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

**LISTA DE EXERCÍCIOS**

**Autor: Paulo Ferreira Naibert**  
**Professor: Hudson Torrent**

**Porto Alegre**  
**25/06/2020**  
**Revisão: 30 de junho de 2020**

## 1 Panel Data and FGLS

**Questão 1:** Estabeleça o modelo de equações lineares, definindo a notação matricial para dados em painel. Explique quais as hipóteses adequadas para a implementação do estimador **GLS** em um sistema de dados de painel. Explique como implementar esse estimador na prática (**FGLS**). Explique também como calcular a matriz de covariância de  $\hat{\beta}_{FGLS}$ .

### Modelo de equações lineares

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

com  $i = 1, \dots, N$ . Cada  $i$  tem  $G = T$  equações temporais.  $G = T$  para dados em painel, onde  $T$  é o número de equações temporais..

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it} \boldsymbol{\beta} + u_{it} \\ &= x_{i1} \beta_1 + \dots + x_{iK} \beta_K + u_{it} \end{aligned}$$

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

Em notação matricial para um painel:

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iT})$$

Um exemplo de equação com intercepto mais 3 variáveis num intervalo de tempo com  $T = 5$ :

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ y_{i5} \end{bmatrix} = \begin{bmatrix} 1 & x_{1i1} & x_{2i1} & x_{3i1} \\ 1 & x_{1i2} & x_{2i2} & x_{3i2} \\ 1 & x_{1i3} & x_{2i3} & x_{3i3} \\ 1 & x_{1i4} & x_{2i4} & x_{3i4} \\ 1 & x_{1i5} & x_{2i5} & x_{3i5} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix}$$

para  $i = 1, \dots, N$ . Podemos generalizar o modelo acima para  $K$  variáveis e  $T$  períodos de tempo.

### Hipóteses

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

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

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

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

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

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

### Estimação

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

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

Estimando a equação acima por **SOLS**:

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

### FSGLS: SGLS Factível

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

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

$$\hat{\Omega} = N^{-1} \sum_{i=1}^N \mathbf{u}_i \mathbf{u}_i'$$

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

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

Empilhando as  $N$  observações:

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

Reescrevendo a equação acima:

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

### Valor Esperado

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

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

**Variância**

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

Tirando o valor Esperado e supondo que:

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

temos:

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

e temos:

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

## 2 Endogeneity and GMM

**Questão 2:** Explique o problema de endogeneidade. Ressalte quais características um bom instrumento deve possuir. A partir da explicação, motive e estabeleça o estimador **GMM** para dados em painel. Qual a variância assintótica desse estimador? Qual a escolha ótima de  $W$ ? Indique quem é  $W$  a fim de que o estimador de GMM coincida com o estimador de Variáveis Instrumentais.

### Modelo

No seguinte modelo *cross-section*:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i ; \quad i = 1, \dots, N. \quad (2.1)$$

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 (2.1) vamos supor:

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

### Hipóteses

Assim, precisamos encontrar um instrumento  $z_i$  para  $x_2$ , uma vez que queremos estimar  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  de maneira consistente. Para  $z_i$  ser um bom instrumento precisamos que  $z$  tenha:

1.  $Cov(z, \varepsilon) = 0 \implies z$  é exógena em (2.1).
2.  $Cov(z, x_2) \neq 0 \implies$  correlação com  $x_2$  após controlar para outras variáveis.

### Estimação

Indo para o problema de dados de painel, temos:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i ; \quad i = 1, \dots, N. \quad (2.2)$$

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

Se é verdade que há endogeneidade em (2.2), então:

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

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

$$E(\mathbf{Z}_i' \mathbf{u}_i) = 0$$

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

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

Se Usarmos estimadores amostrais:

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

Se  $L > K$ , vamos considerar:

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

onde:

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

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

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

Um estimador mais eficiente pode ser encontrado fazendo:

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

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

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

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

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

**Valor Esperado**

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

**Variância**

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

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

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

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

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

### 3 Random Effects (RE, EA)

**Questão 3:** Usando o problema de variável omitida como motivação (heterogeneidade não observada), explique o modelo de **Efeitos Aleatórios** para dados em painel. Explícite as hipóteses necessárias e indique o estimador apropriado para esse modelo, enfatizando as características do estimador GLS. Como podemos fazer inferência nesse caso?

#### Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (3.1)$$

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

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

Podemos reescrever (3.1) como:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}, \quad (3.2)$$

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 (3.2) como:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i, \quad (3.3)$$

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

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

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

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

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

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

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

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



## Estimação

Premultiplicando (3.3) por  $\Omega^{-1/2}$  de dois lados, temos:

$$\begin{aligned}\Omega^{-1/2}\mathbf{y}_i &= \Omega^{-1/2}X_i\boldsymbol{\beta} + \Omega^{-1/2}\mathbf{v}_i \\ \mathbf{y}_i^* &= X_i^*\boldsymbol{\beta} + \mathbf{v}_i^*,\end{aligned}\tag{3.4}$$

Estimando o modelo acima por POLS:

$$\begin{aligned}\boldsymbol{\beta}^{POLS} &= \left( \sum_{i=1}^N X_i^{*'} X_i^* \right)^{-1} \left( \sum_{i=1}^N X_i^{*'} \mathbf{y}_i^* \right) \\ &= \left( \sum_{i=1}^N X_i' \Omega^{-1} X_i \right)^{-1} \left( \sum_{i=1}^N X_i' \Omega^{-1} \mathbf{y}_i \right) \\ &= (X'(I_N \otimes \Omega^{-1})X)^{-1} (X'(I_N \otimes \Omega^{-1})\mathbf{y}).\end{aligned}\tag{3.5}$$

O problema, agora, é estimar  $\Omega$ . Supondo:

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

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

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

Assim,

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

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

Agora, rodando POLS em (3.3) e guardando os resíduos, temos:

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

e conseguimos estimar  $\sigma_v^2$  e  $\sigma_c^2$  por estimadores amostrais:

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

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

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

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

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

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

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

Com  $\hat{\Omega}$  estimado, reescrevemos (3.5) como:

$$\boldsymbol{\beta}^{RE} = \left[ X'(I_N \otimes \hat{\Omega}^{-1})X \right]^{-1} \left[ X'(I_N \otimes \hat{\Omega}^{-1})\mathbf{y} \right].\tag{3.6}$$

**Valor Esperado**

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

**Variância**

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

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

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

## 4 Fixed Effects (EF, FE)

**Questão 4:** Usando o problema de variável omitida como motivação (heterogeneidade não observada), explique o modelo de **Efeitos Fixos** para dados em painel. Explícite as hipóteses necessárias e indique o estimador apropriado para esse modelo. Como podemos fazer inferência nesse caso? Como podemos fazer inferência robusta nesse caso?

### Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (4.1)$$

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

O modelo contém explicitamente um componente não observado que não varia no tempo  $c_i$ . Abordamos esse componente como parte do erro, não como parâmetro a não observado. No caso da análise de **Efeitos Fixos (EF, FE)**, permitimos que esse componente  $c_i$  seja correlacionado com  $\mathbf{x}_{it}$ . Assim, se decidíssemos estimar o modelo (4.1) por POLS, ignorando  $c_i$ , teríamos problemas de inconsistência devido a **endogeneidade**.

As  $T$  equações do modelo (4.1) podem ser reescritas como:

$$\mathbf{y}_i = X_i\boldsymbol{\beta} + c_1\mathbf{1}_T + \mathbf{u}_i, \quad (4.2)$$

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

### Matriz $M^0$

Definimos a matriz  $M^0$  como:

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

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

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

Podemos transformar o modelo (4.3) ao premultiplicarmos todo o modelo por  $M^0$ .

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

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

$$\ddot{\mathbf{y}}_i = \ddot{X}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i, \quad i = 1, \dots, N. \quad (4.3)$$

### Estimação POLS

Aplicando POLS no modelo (4.3)

$$\boxed{\boldsymbol{\beta}^{FE} = \left[ \sum_{i=1}^N \ddot{X}_i'\ddot{X}_i \right]^{-1} \left[ \sum_{i=1}^N \ddot{X}_i'\ddot{\mathbf{y}}_i \right]} \quad (4.4)$$

## Hipóteses

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

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

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

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

## Valor Esperado

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

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

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

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

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

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

## Variância

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

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

Pão:

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

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

Recheio:

$$E \left[ (\ddot{X}' \ddot{\mathbf{u}}) (\ddot{\mathbf{u}}' \ddot{X}) \right] = E \left[ X'(I_N \otimes M^0)(I_N \otimes M^0) \mathbf{u} \mathbf{u}' (I_N \otimes M^0)(I_N \otimes M^0) X \right]$$

$$= E \left[ X'(I_N \otimes M^0) \mathbf{u} \mathbf{u}' (I_N \otimes M^0) X \right]$$

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

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

## Variância sob Homocedasticidade

Usando **FE.3**, temos

**Recheio':**

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

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

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

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

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

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

## 5 First Difference (FD, PD)

**Questão 5:** Usando o problema de variável omitida como motivação (heterogeneidade não observada), explique o modelo de **Primeira Diferença** para dados em painel. Explícite as hipóteses necessárias e indique o estimador apropriado para esse modelo. Como podemos fazer inferência nesse caso? Como podemos fazer inferência robusta nesse caso?

### Modelo

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

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \quad (5.1)$$

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

O modelo contém explicitamente um componente não observado,  $c_i$ , que não varia no tempo. Tratamos o componente não observado como parte do erro, não como parâmetro a ser estimado. Aqui permitimos que  $c_i$  seja correlacionado com  $\mathbf{x}_{it}$ . Deste modo, **não** podemos ignorar a sua presença e estimar (5.1) por POLS, visto que isso resultaria num estimador inconsistente devido a **endogeneidade**.

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

- Reescrevemos (5.1) defasado:

$$y_{it-1} = \mathbf{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \quad (5.2)$$

- Tiramos a diferença entre (5.2) e (5.1):

$$\begin{aligned} y_{it} - y_{it-1} &= (\mathbf{x}_{it} - \mathbf{x}_{it-1})\boldsymbol{\beta} + c_i - c_i + u_{it} - u_{it-1} \\ \Delta y_{it} &= \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}. \end{aligned} \quad (5.3)$$

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

Reescrevendo (5.3) no formato matricial empilhando  $T$ :

$$\Delta \mathbf{y}_i = \Delta \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i \quad (5.4)$$

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

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

### Estimação POLS

O estimador  $\hat{\boldsymbol{\beta}}^{FD}$  é o POLS da regressão no modelo (5.4), assim:

$$\boxed{\boldsymbol{\beta}^{FD} = \left[ \sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{X}_i \right]^{-1} \left[ \sum_{i=1}^N \Delta \mathbf{X}_i' \Delta \mathbf{y}_i \right]} \quad (5.5)$$

## Hipóteses

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

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

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

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

## Valor Esperado

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

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

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

## Variância

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

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

## Variância sob Homocedasticidade

Usando **FD.3**, temos

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

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

com

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

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

## 6 Exogeneidade Estrita e FDIV

**Questão 6:** Explique a hipótese de exogeneidade estrita dos regressores. Em seguida, argumente mostrando que a hipótese de exogeneidade estrita não se sustenta no seguinte modelo:

$$y_{it} = z_{it}\gamma + \rho y_{it-1} + c_i + u_{it}.$$

Explique detalhadamente como esse modelo pode ser estimado a partir da combinação entre Variáveis Instrumentais e método da Primeira Diferença.

### Modelo

No seguinte modelo

$$y_{it} = x_{it}\beta + u_{it},$$

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

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

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

$$E(u_{it} | x_{i1}, \dots, x_{iT}) = 0, \quad t = 1, \dots, T$$

ou seja:

$$E(y_{it} | x_{i1}, \dots, x_{iT}) = x_{it}\beta, \quad t = 1, \dots, T$$

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

Para o seguinte modelo:

$$y_{it} = z_{it}\gamma + \rho y_{it-1} + c_i + u_{it}, \quad t = 2, \dots, T$$

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

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

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

$$\left. \begin{array}{l} y_{it} = z_{it}\gamma + \rho y_{it-1} + c_i + u_{it} \\ y_{it-1} = z_{it-1}\gamma + \rho y_{it-2} + c_i + u_{it-1} \end{array} \right\} \implies y_{it} = z_{it}\gamma + \rho(z_{it-1}\gamma + \rho y_{it-2} + c_i + u_{it-1}) + c_i + u_{it}.$$

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

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

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

(6.1)

### Estimação

Não podemos estimar o modelo (6.1) por POLS, uma vez que  $Cov(\Delta y_{it-1}, \Delta u_{it}) \neq 0$ . Como saída, podemos estimar por P2SLS, usando instrumentos para  $\Delta y_{it-1}$  (alguns instrumentos para  $\Delta y_{it-1}$  são  $y_{it-2}, y_{it-3}, \dots, y_{i1}$ ).



**P2SLS**

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

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

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

com

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

onde  $P_Z$  é a matriz de projeção em  $Z$ .

**FDIV**

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

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

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

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

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

**Hipóteses**

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

**FDIV.2:**  $\text{Posto}[E(W'_iW_i)] = L$

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

**Estimação FDIV**

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

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

**Valor Esperado**

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

**Variância**

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

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

## 7 Latent Variables, Probit and Logit

**Questão 7:** Usando a motivação de uma **variável latente**, motive a construção do estimador **LOGIT/PROBIT**. Explique o procedimento de estimação de verossimilhança que caracteriza o estimador. Inclua em sua explicação o resultado da distribuição assintótica de  $\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ . Ressalte a forma mais simples da variância assintótica desse estimador, devido ao fato de ser um estimador de máxima verossimilhança.

### Modelo

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

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i. \quad (7.1)$$

Defina  $y$  como:

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

temos que:

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

Além disso, pela definição de  $y_i$ , equação (7.1), temos:

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

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

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

Se  $G(\cdot)$  for uma distribuição:

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

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

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

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

Para estimarmos  $\hat{\boldsymbol{\beta}}$  por máxima verossimilhança, temos de encontrar  $\boldsymbol{\beta} \in B$ , onde  $B$  é o espaço paramétrico, tal que  $\boldsymbol{\beta}$  maximize o valor da distribuição conjunta de  $\mathbf{y}$ , ou seja:

$$\max_{\boldsymbol{\beta} \in B} \prod_{i=1}^N f(y_i | \mathbf{x}_i; \boldsymbol{\beta}).$$

Tirando o logaritmo e dividindo tudo por  $N$  (podemos fazer isso pois são transformações monotônicas e não alteram o lugar onde  $\boldsymbol{\beta}$  ótimo irá parar):

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

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

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

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

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

Definindo o *vector score* da observação  $i$ :

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

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

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

Tendo essas definições, o **Teorema do Valor Médio** (TVM) nos diz que no intervalo  $[a, b]$ , existe um número,  $c$ , tal que:

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

**FAZER DESENHO**

Trocando  $f(\cdot)$  por  $s_i(\cdot)$ ,  $a$  por  $\beta_0$ ,  $b$  por  $\hat{\beta}$  e  $c$  por  $\bar{\beta}$ , temos:

$$H_i(\bar{\beta}) = \frac{s_i(\hat{\beta}) - s_i(\beta_0)}{\hat{\beta} - \beta_0},$$

tirando médias dos dois lados:

$$N^{-1} \sum_{i=1}^N H_i(\bar{\beta}) = \frac{1}{\hat{\beta} - \beta_0} N^{-1} \sum_{i=1}^N [s_i(\hat{\beta}) - s_i(\beta_0)]$$

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

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

Onde

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

Assim, temos que:

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

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

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

## 8 ATT, ATE, Propensity Score

**Questão 8:** Explique como estimar o efeito médio do tratamento ( $\tau_{ATE}$ ) e o efeito médio do tratamento sobre o tratado ( $\tau_{ATT}$ ), considerando a hipótese de Ignorabilidade do Tratamento condicional a um conjunto de covariáveis. Aborde o método *Propensity Score*. Discuta a importância da hipótese *Overlap* para a aplicabilidade desse estimador. Explique resumidamente como o *Propensity Score* pode ser estimado.

### Modelo

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

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

Idealmente, para isolarmos completamente o efeito de  $w = 1$ , gostaríamos de poder calcular:

$$N^{-1} \sum_{i=1}^N (y_{i1} - y_{i0}).$$

Ou seja, o efeito que o tratamento causa sobre um indivíduo com todo o resto permanecendo constante. Em outras palavras, queríamos que houvesse dois mundos paralelos observáveis onde seria possível observar o que acontece com  $y_i$  com e sem tratamento. Infelizmente, para cada indivíduo  $i$ , observamos apenas  $y_{i1}$  ou  $y_{i0}$ , nunca ambos.

Antes de continuarmos, faremos as seguintes definições:

**ATE:**  $E(y_1 - y_0)$

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

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

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

OBS:

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

### Métodos Assumindo Ignorabilidade do Tratamento

**ATE.1:** Ignorabilidade.

$w$  e  $(y_1, y_0)$  são independentes condicionais a  $\mathbf{x}$ .

**ATE.1':** Ignorabilidade da Média.

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

Vamos definir

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

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

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

**ATE.2:** *Overlap*

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

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

Para o *ATT* vamos supor:

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

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

## Propensity Score

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

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

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

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

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

$\mu_1(\mathbf{x})$

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

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

$$\begin{aligned} E(wy | \mathbf{x}) &= \mu_1(\mathbf{x})p(\mathbf{x}) \\ \boxed{\mu_1(\mathbf{x})} &= \frac{E(wy | \mathbf{x})}{p(\mathbf{x})} \end{aligned}$$

$$\mu_0(\mathbf{x})$$

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

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

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

**ATE:**

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

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

**ATT:**

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

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

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

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