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: 4 de julho 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 implementear esse estimador na prática (**FGLS**). Explique também como calular a matriz de covariância de $\widehat{\boldsymbol{\beta}}_{FGLS}$.

Modelo de equações lineares

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

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..

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

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

Em notação matricial para um painel:

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

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

   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:

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

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*:

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

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 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 (2.1).
2. $Cov(z, x_2) \neq 0 \implies$ correlação com x_2 após controlar para outras vaariáveis.

Estimação

Indo para o problema de dados de painel, temos:

$$\boldsymbol{y}_i = X_i \boldsymbol{\beta} + \boldsymbol{u}_i \ ; \quad i = 1, \dots, N. \tag{2.2}$$

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 (2.2), 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:

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

Se $L > K$, vamos considerar:

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

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:

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

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

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

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}.

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. Explicite as hipóteses necessárias e indique o estimador apropriado para esse modelo, enfatizando as característica do estimador GLS. Como podemos fazer inferência nesse caso?

Modelo

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

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{3.1}$$

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

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 (3.1) como:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}, \tag{3.2}$$

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

Agora, vamos empilhar os t's e reescrever (3.2) como:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{v}_i, \tag{3.3}$$

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

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

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

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

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

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

2. 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 (3.3) port $\Omega^{-1/2}$ do dois lados, temos:

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

Estimando o modelo acima por POLS:

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

O problema, agora, é estimar Ω. Supondo:

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

Como $\Omega = E(\boldsymbol{v}_i\boldsymbol{v}_i') = E[(c_i\mathbf{1}_T + \boldsymbol{u}_i)(c_i\mathbf{1}_T + \boldsymbol{u}_i)']$, 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 (3.3) 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 (3.5) como:

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

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

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

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. Explicite 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**:

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + c_i + u_{it}, \tag{4.1}$$

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

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

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

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

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

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

Estimação POLS

Aplicando POLS no modelo (4.3)

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

Hipóteses

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

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

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 $\ddot{\mathbf{u}} = (I_N \otimes M^0)\boldsymbol{u}$, 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:

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

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:

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

$Var(\boldsymbol{\beta}^{FE}) =$ 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 $T \times T$.

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

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

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. Explicite 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**:

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

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

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 (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 trasnformação a ser feita é a primeira diferença. Para tanto, seguimos os seguintes passos:

- Reescrevemos (5.1) defasado:

$$y_{it-1} = \boldsymbol{x}_{it-1}\boldsymbol{\beta} + c_i + u_{it-1} \tag{5.2}$$

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

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

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

Reescrevendo (5.3) no formato matricial empilhando T:

$$\Delta \boldsymbol{y}_i = \Delta X_i \boldsymbol{\beta} + \boldsymbol{e}_i \tag{5.4}$$

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 (5.4), assim:

$$\boxed{\boldsymbol{\beta}^{FD} = \left[\sum_{i=1}^{N} \Delta X_i' \Delta X_i\right]^{-1} \left[\sum_{i=1}^{N} \Delta X_i' \Delta \boldsymbol{y}_i\right]} \tag{5.5}$$

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.

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:

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

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

Modelo

No seguinte modelo

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

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

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

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

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

ou seja:

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

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

Para o seguinte modelo:

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

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

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

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

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

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

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

$$\boxed{\Delta y_{it} = \Delta \boldsymbol{z}_{it}\boldsymbol{\gamma} + \rho \Delta y_{it-1} + \Delta u_{it} \,, \qquad t = 3, \ldots, T} \tag{6.1}$$

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

P2SLS

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

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

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

com

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

onde P_Z é a matriz de projeção em Z.

FDIV

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

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

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

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

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

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

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

$e_i = \Delta u_{it}$.

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:

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i. \tag{7.1}$$

Defina y como:

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

temos que:

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

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

$$\boxed{\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to N(0, A_0^{-1} B_0 A_0^{-1})}.$$

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

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

8 ATT, ATE, Propensity Score

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

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 $\boldsymbol{x}$

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

<u>OBS:</u>

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

Métodos Assumindo Ignorabilidade do Tratamento

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

ATE.1': Ignorabilidade da Média.

a) $E(y_0 \,|\, w, \boldsymbol{x}) = E(y_0 \,|\, \boldsymbol{x})$
b) $E(y_1 \,|\, w, \boldsymbol{x}) = E(y_1 \,|\, \boldsymbol{x})$

Vamos definir

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

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

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

ATE.2: *Overlap*
Para todo $\boldsymbol{x}$, $P(w = 1 \,|\, \boldsymbol{x}) \in (0, 1)$, $p(\boldsymbol{x}) = p(w = 1|\boldsymbol{x})$.

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

Para o ATT vamos supor:

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

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

Propensity Score

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

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

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

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

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:

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

$\mu_0(\boldsymbol{x})$

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

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:

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

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

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