

ECONOMETRIA I

O ESTIMADOR DE MÍNIMOS QUADRADOS ORDINÁRIOS

Luis A. F. Alvarez

26 de março de 2025

AMBIENTE

- Pesquisador observa n pares (Y_i, X_i) , $i = 1 \dots, n$, para os quais supõe um modelo linear da forma:

$$Y_i = X_i' \beta + \epsilon_i \dots i = 1, \dots, n, \quad (1)$$

onde $\beta \in \mathbb{R}^k$ é um parâmetro desconhecido, e ϵ_i , $i = 1, \dots, n$ são variáveis aleatórias não observadas.

- Recorde-se, da última aula, que cabe ao pesquisador postular o modelo e a interpretação dos coeficientes.
- No caso mais comum, $(Y_i, X_i) \stackrel{d}{=} (Y, X)$ para $i = 1, \dots, n$, onde a distribuição de (Y, X) representa a distribuição das variáveis numa população de interesse.
 - Por exemplo, podemos ter que $\{(Y_i, X_i)\}_{i=1}^n$ é uma amostra aleatória de uma população com distribuição $\mathbb{P}_{Y,X}$, para a qual postulamos um modelo linear.
 - Mas também podemos ter que as observações entre pares apresentem dependência entre si, embora com leis $\mathbb{P}_{(Y_i, X_i)}$ comuns a todo i .
- De modo mais geral, no entanto, pode ser que os (Y_i, X_i) não possuam a mesma distribuição conjunta, mas haja uma relação comum e estável ao longo de i .

NOTAÇÃO MATRICIAL

- No que segue, definimos as seguintes matrizes aleatórias:

$$\mathbf{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X'_1 \\ X'_2 \\ \vdots \\ X'_n \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

- Com base na notação acima introduzida, podemos reescrever (1) em notação matricial como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

ESTIMADOR DE MÍNIMOS QUADRADOS ORDINÁRIOS

- O estimador de mínimos quadrados ordinários de β , denotado por \hat{b} , consiste em estimar β minimizando a distância, na norma Euclidiana, entre \mathbf{y} e uma combinação linear das colunas de \mathbf{X} , i.e.

$$\hat{b} \in \operatorname{argmin}_{b \in \mathbb{R}^k} \|\mathbf{y} - \mathbf{X}b\|_2^2 = \operatorname{argmin}_{b \in \mathbb{R}^k} \frac{1}{n} \sum_{i=1}^n (Y_i - X_i' b)^2,$$

- Em outras palavras, encontramos o coeficiente b que maximiza a contribuição dos X_i à explicação de Y_i , tal qual medida pela média da distância ao quadrado entre os Y_i e $X_i' b$.
- Estimador é obtido como solução a uma otimização de uma função convexa, sem restrições \implies condição de primeira ordem caracteriza o conjunto de soluções.
- Condições de primeira ordem podem ser escritas como:

$$\mathbf{0}_{k \times 1} = \sum_{i=1}^n X_i (Y_i - X_i' b) = \mathbf{X}'(\mathbf{y} - \mathbf{X}b)$$

CONDIÇÃO DE POSTO E UNICIDADE DO MÍNIMO

Sob a condição

HIPÓTESE (H1-POSTO)

A matriz \mathbf{X} apresenta posto k .

Temos que $\mathbf{X}'\mathbf{X}$ é invertível (por quê), de modo que existe uma única solução ao problema de otimização, dada por:

$$\hat{b} = (\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{y}).$$

- Condição de posto requer que nenhuma das colunas seja escrita como combinação linear das demais.
 - Se a primeira entrada dos X_i corresponde a um intercepto (i.e. $X_{i,1} = 1$ para todo $i = 1, \dots, n$), nenhuma das colunas pode ser escrita como função afim das demais
- Observe que, como $\text{rank}(\mathbf{X}) \leq \min\{n, k\}$, condição implica que $n \geq k$.

UM CASO SIMPLES

- Considere, para fixar as ideias, o caso em que $X_i = \begin{bmatrix} 1 & T_i \end{bmatrix}'$.
- Nesse caso, a matriz $\mathbf{X}'\mathbf{X}$ é dada por:

$$\begin{bmatrix} n & \sum_{i=1}^n T_i \\ \sum_{i=1}^n T_i & \sum_{i=1}^n T_i^2 \end{bmatrix} \quad (3)$$

de modo que a condição de posto é equivalente a

$$\widehat{V(T)} = \frac{1}{n} \sum_{i=1}^n T_i^2 - \left(\frac{1}{n} \sum_{i=1}^n T_i \right)^2 > 0.$$

- Se condição de posto é satisfeita, estimador de MQO é dado por:

$$\hat{b}_2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(T_i - \bar{T})}{\sum_{i=1}^n (T_i - \bar{T})^2} = \frac{\widehat{\text{cov}(T, Y)}}{\widehat{V(T)}}$$

MQO: Propriedades Algébricas

MATRIZ DE PROJEÇÃO

- Definimos a matriz de projeção de \mathbf{X} como:

$$P = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

- Observe que a matriz de projeção é tal que $X\hat{b} = P\mathbf{y}$.
- $P\mathbf{z}$ é a projeção (em termos de minimização da distância Euclidiana) de $\mathbf{z} \in \mathbb{R}^n$ no espaço gerado pelas colunas de \mathbf{X} .
- Matriz de projeção tem as seguintes propriedades:
 - Simétrica.
 - Idempotente ($P^2 = P$).
 - Os autovalores de P são ou 0 ou 1.
 - $\text{trace}(P) = k = \text{rank}(P)$.

MATRIZ RESIDUALIZADORA

- A matriz residualizadora (*residual-maker*) de X é dada por:

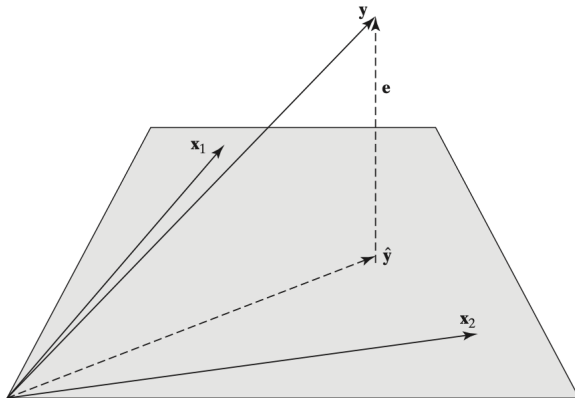
$$M = (I - P)$$

- Matriz residualizadora tem as seguintes propriedades:
 - Simétrica.
 - Idempotente ($M^2 = M$).
 - Os autovalores de M são ou 0 ou 1.
 - $\text{trace}(M) = n - k = \text{rank}(M)$.
 - $\mathbf{X}'M = 0$, $M\mathbf{X} = 0$, $PM = MP = 0$.
- Matriz residualizadora devolve o erro de projeção $\mathbf{z} - P\mathbf{z}$.
 - Como $P\mathbf{z}$ é o minimizador da distância Euclidiana no espaço gerado pelas colunas de \mathbf{X} , temos que:

$$(\mathbf{Mz}) \cdot (\mathbf{XI}) = \mathbf{z}'\mathbf{MXI} = 0 \quad \forall \mathbf{z} \in \mathbb{R}^n, \mathbf{I} \in \mathbb{R}^k.$$

VISUALIZAÇÃO GRÁFICA

FIGURE 3.2 Projection of \mathbf{y} into the Column Space of \mathbf{X} .



FÓRMULA DA INVERSA PARTICIONADA

- Suponha que particionemos a matriz $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}$ onde \mathbf{X}_1 e \mathbf{X}_2 são matrizes de dimensão $n \times k_1$ e $n \times k_2$.
- Nesse caso, a fórmula da inversa particionada nos indica que:

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{X}_1'\mathbf{X}_1)^{-1} + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\mathbf{F}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} & -(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\mathbf{F} \\ -\mathbf{F}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1} & \mathbf{F} \end{bmatrix}$$

onde $\mathbf{F} = (\mathbf{X}_2'\mathbf{X}_2 - \mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2)^{-1} = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}$ onde \mathbf{M}_1 é a residualizadora de \mathbf{X}_1 .

- Inversa de $\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2$ existe pois matriz \mathbf{X} tem posto cheio.

TEOREMA DE FRISCH-WAUGH-LOVELL

-

- Da propriedade acima, segue o importante resultado abaixo:

TEOREMA (FRISCH-WAUGH-LOVELL)

$$\hat{b}_2 = (\mathbf{X}_2' M_1 \mathbf{X}_2)^{-1} (\mathbf{X}_2' M_1 \mathbf{y}) = ((M_1 \mathbf{X}_2)' (M_1 \mathbf{X}_2))^{-1} ((M_1 \mathbf{X}_2)' (M_1 \mathbf{y}))$$

- Resultado acima nos mostra que estimadores de MQO associados a \mathbf{X}_2 em uma regressão que inclui \mathbf{X}_1 e \mathbf{X}_2 são idênticos a:
 - Regredir \mathbf{y} em \mathbf{X}_1 , e guardar os resíduos \mathbf{e}_y .
 - Para cada $j = 1, \dots, k_2$, regredir a j -ésima coluna de \mathbf{X}_2 em \mathbf{X}_1 , e guardar os resíduos \mathbf{e}_j .
 - Regredir \mathbf{e}_y em $\mathbf{e}_1, \dots, \mathbf{e}_{k_2}$ e recuperar os coeficientes.

MQO: Propriedades Estatísticas em Amostras Finitas

REGRESSORES FIXOS

- Nesta seção, analisaremos as propriedades de risco do estimador de MQO.
- Essas propriedades serão analisadas com respeito à distribuição de \mathbf{y} condicionalmente a \mathbf{X} .
 - Em outras palavras, estamos pensando nas propriedades do estimador sob amostras repetidas (realizações alternativas da incerteza) em que o valor dos regressores é o mesmo da amostra observada.
 - Ou seja, estamos efetivamente tratando os regressores como fixos sob amostras repetidas.
 - **Exemplo:** se Y_i é a taxa de inflação no período i , e X_i a taxa de desemprego no período $i - 1$, analisaremos as propriedades dos estimadores sob realizações alternativas do cenário econômico em que a taxa de desemprego é igual à observada nos períodos $i = 1, \dots, n$.

NÃO VIÉS

- Sob a restrição de que:

HIPÓTESE (H2-EXOGENEIDADE)

$$\mathbb{E}[\mathbf{e}|\mathbf{X}] = 0$$

- O estimador de MQO é não viciado para β , isto é, qualquer que seja o valor de $\beta \in \mathbb{R}^k$, temos que:

$$\mathbb{E}[\hat{\mathbf{b}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbb{E}[\mathbf{y}|\mathbf{X}] = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbb{E}[\mathbf{e}|\mathbf{X}] = \beta$$

- Observe que a condição de exogeneidade implica que:

$$\mathbb{E}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\beta$$

INTERPRETAÇÃO DA CONDIÇÃO DE EXOGENEIDADE

- Sob amostragem aleatória de uma população, i.e. $(Y_i, X_i) \stackrel{iid}{\sim} (X, Y)$, condição implica $\mathbb{E}[Y_i|\mathbf{X}] = \mathbb{E}[Y_i|X_i] = X_i'\beta \implies \mathbb{E}[Y|X] = X'\beta$.
 - A esperança condicional das variáveis (Y, X) que representam a distribuição das características na população de interesse é linear em X , e coincide com o modelo linear de interesse.
 - Se o modelo postulado é **preditivo**, isso significa que o parâmetro-alvo do melhor preditor linear consiste também no melhor preditor dentro das funções não lineares de X (ou, de modo equivalente, a melhor aproximação linear a $\mathbb{E}[Y|X]$ é exata).
 - Se modelo postulado é **causal**, isso significa que as causas não observadas ϵ apresentam o mesmo valor médio nas diferentes subpopulações definidas pelos valores de X . Trata-se de condição mais forte que (i.e. que implica) a condição de identificação $\text{cov}(X, \epsilon) = 0$.
- Quando as observações (X_i, Y_i) apresentam dependência entre si, esta condição impõe restrições adicionais.
 - Por exemplo, se as observações estão ordenadas no tempo e o modelo postulado é causal, $\mathbb{E}[\epsilon_i|\mathbf{X}] = 0$ implica que causas não observadas do fenômeno no período i não exibem associação sistemática com as causas em i e em nenhum outro período.
 - Nesse caso, condição limita retroalimentação entre causas observadas e não observadas no tempo.

VIÉS DE VARIÁVEL OMITIDA

- Considere um modelo da forma:

$$\hat{b} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon,$$

em que $\mathbb{E}[\epsilon|\mathbf{X}_1, \mathbf{X}_2] = 0$.

- Por exemplo, num modelo causal, é suficiente observar as causas \mathbf{X}_1 e \mathbf{X}_2 para que as causas não observadas restantes estejam balanceadas nas subpopulações definidas pelos valores de $(\mathbf{X}_1, \mathbf{X}_2)$.
- Suponha que você **não observe ou desconsidere** \mathbf{X}_2 , e considere o estimador de MQO \tilde{b}_1 de \mathbf{y} em \mathbf{X}_1 .
 - Sob quais condições esse estimador é não viciado para β_1 ?
- Um cálculo simples nos mostra que:

$$\begin{aligned}\tilde{b}_1 &= (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} = \beta_1 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\epsilon \\ \tilde{b}_1 &= \beta_1 + \hat{\gamma}\beta_2 + (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\epsilon,\end{aligned}$$

onde $\hat{\gamma}$ é uma matrix $k_1 \times k_2$ em que cada coluna representa os coeficientes do estimador de MQO da j -ésima coluna de \mathbf{X}_2 em \mathbf{X}_1 .

VIÉS DE VARIÁVEL OMITIDA (CONT.)

- Da propriedade da torre, temos que

$\mathbb{E}[\epsilon|\mathbf{X}_1] = \mathbb{E}[\mathbb{E}[\epsilon|\mathbf{X}_1, \mathbf{X}_2]|\mathbf{X}_1] = 0$. Portanto, temos que:

$$\mathbb{E}[\tilde{b}_1|\mathbf{X}_1] = \beta_1 + \mathbb{E}[\hat{\gamma}|\mathbf{X}_1]\beta_2$$

- Se uma das duas condições abaixo for satisfeita, estimador será não viciado para β_1 :

1. $\beta_2 = 0$.

- Se o modelo postulado é causal, essa hipótese significa que as variáveis \mathbf{X}_2 não possuem efeito causal sobre \mathbf{y} .
- Se o modelo postulado é preditivo, essa hipótese significa que as variáveis \mathbf{X}_2 não possuem informação preditiva sobre \mathbf{y} , uma vez que usamos \mathbf{X}_1 na predição.
 - Nesse caso, estimador de MQO do melhor preditor linear que inclui \mathbf{X}_2 produzirá, em média, o mesmo resultado que o estimador que inclui \mathbf{X}_1 e \mathbf{X}_2 .

2. $\mathbb{E}[\hat{\gamma}|\mathbf{X}_1] = 0$.

- Essa hipótese é satisfeita se as variáveis em \mathbf{X}_1 não possuem capacidade preditiva sobre nenhuma das variáveis em \mathbf{X}_2 .

ESTIMADOR LINEAR

- Para obter propriedades de otimalidade para o estimador de MQO, necessitamos introduzir definições adicionais.
- Um estimador do parâmetro β é dito **linear** (em \mathbf{y}) se:

$$\phi(\mathbf{X}, \mathbf{y}) = A(\mathbf{X})\mathbf{y},$$

onde $A : \mathbb{R}^{n \times k} \mapsto \mathbb{R}^{k \times n}$.

- Note que o estimador de MQO é um estimador linear.

HOMOCEDASTICIDADE

- O estimador de MQO terá propriedades de otimalidade se, além de $H1-H2$ restringirmos que:

HIPÓTESE (H3-HOMOCEDASTICIDADE)

Existe $\sigma^2 > 0$ tal que $\mathbb{V}[\epsilon|\mathbf{X}] = \sigma^2 \mathbb{I}_{n \times n}$.

- A hipótese de homocedasticidade requer que condicionalmente a \mathbf{X} os termos de erro $\epsilon_i, \epsilon_j, j \neq i$, sejam não correlacionados e exibam variância idêntica e não dependente do valor de \mathbf{X} .
 - Sob amostragem aleatória de uma população (X, Y) em que vale para o modelo postulado que $\mathbb{E}[\epsilon|\mathbf{X}] = 0$ (H2), temos que, para $i \neq j$,
 $\text{cov}(\epsilon_i, \epsilon_j|\mathbf{X}) = \mathbb{E}[\epsilon_i \epsilon_j|\mathbf{X}] = \mathbb{E}[\epsilon_i \epsilon_j|X_i, X_j] = \mathbb{E}[\mathbb{E}[\epsilon_i|\epsilon_j, X_i, X_j]\epsilon_j|X_i, X_j] = \mathbb{E}[\mathbb{E}[\epsilon_i|X_i]\epsilon_j|X_i, X_j] = 0$. Portanto, o requerimento de não covariância condicional não impõe restrições adicionais a H2.
 - Por outro lado, nesse caso, a hipótese impõe que $\mathbb{V}[Y|\mathbf{X}] = \mathbb{V}[\epsilon|\mathbf{X}] = \sigma^2$, i.e. que a dispersão de Y nas subpopulações definidas pelos diferentes valores de X sejam as mesmas.

TEOREMA DE GAUSS-MARKOV

- Defina o seguinte conjunto de distribuições condicionais de \mathbf{y} dado \mathbf{X} :

$$\mathcal{F}(\mathbf{X}) = \{P_{\mathbf{y}|\mathbf{X}} : \exists \delta \in \mathbb{R}^k, \mathbb{E}_{P_{\mathbf{y}|\mathbf{X}}}[\mathbf{y}|\mathbf{X}] = \mathbf{X}\delta, \exists \sigma^2 > 0, \mathbb{V}_{P_{\mathbf{y}|\mathbf{X}}}[\mathbf{y}|\mathbf{X}] = \sigma^2 \mathbb{I}_{n \times n}\}$$

- Essa é a classe de distribuições condicionais de \mathbf{y} dado \mathbf{X} em que a esperança condicional é linear em \mathbf{X} e a hipótese de homocedasticidade é satisfeita.
 - Denote, para uma distribuição condicional $F \in \mathcal{F}(\mathbf{X})$, $\delta(F)$ o parâmetro que entra na esperança condicional.
- Com base nas definições acima, temos o seguinte resultado.

TEOREMA (GAUSS-MARKOV)

*Sob H1, o estimador de MQO é o estimador **linear** não viciado para $\delta(F)$ de variância uniformemente mínima na classe $\mathcal{F}(\sigma^2, \mathbf{X})$, i.e. para qualquer outro estimador **linear** ψ tal que $\mathbb{E}_F[\psi|\mathbf{X}] = \delta(F)$ para todo $F \in \mathcal{F}(\mathbf{X})$, temos que, para todo $F \in \mathcal{F}(\mathbf{X})$:*

$$\mathbb{V}_F(\psi|\mathbf{X}) - \mathbb{V}_F(\hat{b}|\mathbf{X}) \quad \text{é positiva semidefinida}$$

ETAPAS DA DEMONSTRAÇÃO

1. Mostrar que $\{\delta(F) : F \in \mathcal{F}(\mathbf{X})\} = \mathbb{R}^k$.
 - Basta considerar, para $\delta \in \mathbb{R}^k$, $\mathbf{y} = \mathbf{X}\delta + \mathbf{e}$, com $\mathbf{e} \sim \mathcal{N}(0, \mathbb{I}_{n \times n})$ e \mathbf{e} independente de \mathbf{X} .
2. Com base no resultado acima, mostrar que qualquer estimador linear $A(\mathbf{X})\mathbf{y}$ não viciado em $\mathcal{F}(\mathbf{X})$ deve satisfazer

$$A(\mathbf{X})\mathbf{X} = \mathbb{I}_{k \times k}$$

3. Mostrar que, para todo $F \in \mathcal{F}(\mathbf{X})$, $\text{cov}(A(\mathbf{X})\mathbf{y} - \hat{b}, \hat{b} | \mathbf{X}) = 0_{k \times k}$
4. Concluir que, para todo $F \in \mathcal{F}(\mathbf{X})$:

$$\mathbb{V}_F(A(\mathbf{X})\mathbf{y} - \hat{b} | \mathbf{X}) = \mathbb{V}_F(A(\mathbf{X})\mathbf{y} | \mathbf{X}) - \mathbb{V}_F(\hat{b} | \mathbf{X})$$