Econometria I

Regressão de Mínimos Quadrados

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024

Sumário I

Modelo de Regressão Linear

- Vamos mostrar que o estimador de MQO é não viesado no modelo de regressão linear.
- Sabemos que

$$\mathbb{E}(y_i|\boldsymbol{X}) = \mathbb{E}(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{1}$$

- **Nota**: a *primeira igualdade* estabelece que a esperança condicional de y_i dado $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ depende apenas de $\boldsymbol{x}_i$ dado que as observações são independentes entre i. A *segunda igualdade* é o pressuposto de uma média condicional linear.

- Assim,

$$\mathbb{E}\left(\boldsymbol{y}|\boldsymbol{X}\right) = \begin{pmatrix} \vdots \\ \mathbb{E}\left(y_i|\boldsymbol{X}\right) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \boldsymbol{x}_i'\boldsymbol{\beta} \\ \vdots \end{pmatrix} = \boldsymbol{X}\boldsymbol{\beta} \tag{2}$$

- De forma similar

$$\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}\right) = \begin{pmatrix} \vdots \\ \mathbb{E}\left(e_i|\boldsymbol{X}\right) \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ \mathbb{E}\left(e_i|\boldsymbol{x}_i\right) \\ \vdots \end{pmatrix} = \mathbf{0} \tag{3}$$

Estimador de MQO

- Usando a definição de que $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y})$, o teorema do condicionamento e as propriedades de matriz inversa, temos:

$$\begin{aligned}
\mathbb{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) &= \left(\mathbb{E}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{y})|\boldsymbol{X}\right) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbb{E}(\boldsymbol{y}|\boldsymbol{X}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\
&= \boldsymbol{\beta} \qquad (4)
\end{aligned}$$

Estimador de MQO

- Outra forma de obter o mesmo resultado é inserindo $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e}$ em $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{y}\right)$:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{e})\right) \\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{e}\right) \\ &= \boldsymbol{\beta} + \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e} \qquad (5)\end{aligned}$$

- Essa decomposição para $\widehat{\boldsymbol{\beta}}$ tem o verdadeiro parâmetro $\boldsymbol{\beta}$ e o componente estocástico $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e}$.

Viés

- Podemos calcular que

$$\begin{aligned}\mathbb{E}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X}\right) &= \mathbb{E}\left(\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e}|\boldsymbol{X}\right)\\ &= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\mathbb{E}\left(\boldsymbol{e}|\boldsymbol{X}\right)\\ &= \mathbf{0} \qquad (6)\end{aligned}$$

- Independente do método temos que $\mathbb{E}\left(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}\right) = \boldsymbol{\beta}$. Usando a lei das expectativas iteradas temos:

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right) = \mathbb{E}\left(\mathbb{E}\left(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}\right)\right) = \boldsymbol{\beta} \qquad (7)$$

Teorema (Esperança do Estimador de Mínimos Quadrados)

Em um modelo de regressão linear

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}\right) = \boldsymbol{\beta} \tag{8}$$

e

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta} \tag{9}$$

- Eq. (9) diz que o estimador $\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right)$ é não visado para $\boldsymbol{\beta}$, indicando que a distribuição de $\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right)$ está centrada em $\boldsymbol{\beta}$.
- Eq. (8) diz que o estimador condicional é não viesado. **Esse é um resultado mais forte**! Ele indica que $\mathbb{E}\left(\widehat{\boldsymbol{\beta}}\right)$ é não viesado para qualquer realização da matriz de regressores $\boldsymbol{X}$.

Variância do Estimador de MQO

- Vamos calcular a variância condicional do estimador de MQO.
- Para qualquer vetor aleatório $\boldsymbol{Z}$ definimos uma matriz de covariância $r \times r$

$$\begin{aligned} \operatorname{var}(\boldsymbol{Z}) &= \mathbb{E}(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])' \\ &= \mathbb{E}(\boldsymbol{Z}\boldsymbol{Z}') - (\mathbb{E}\boldsymbol{Z})(\mathbb{E}\boldsymbol{Z})' \end{aligned} \tag{10}$$

  e para qualquer par $(\boldsymbol{Z}, \boldsymbol{X})$ definimos a matriz de variância condicional

$$\operatorname{var}(\boldsymbol{Z}|\boldsymbol{X}) = \mathbb{E}\left[(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}|\boldsymbol{X}])\,(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}|\boldsymbol{X}])'\right] \tag{11}$$

- Definimos $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} \overset{\text{def}}{=} \operatorname{var}[\widehat{\boldsymbol{\beta}}|\boldsymbol{X}]$ como a matriz de covariância condicional dos estimadores do coeficientes de regressão.

- A matriz de covariância condicional $n \times 1$ do erro de regressão $\boldsymbol{e}$ é a matriz $n \times n$

$$\operatorname{var}[\boldsymbol{e}|\boldsymbol{X}] = \mathbb{E}\left[\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}\right] \stackrel{\text{def}}{=} \boldsymbol{D} \tag{12}$$

- O *i-ésimo* elemento da diagonal de $\boldsymbol{D}$ é

$$\mathbb{E}\left(e_i^2|\boldsymbol{X}\right) = \mathbb{E}\left(e_i^2|\boldsymbol{x}_i\right) = \sigma_i^2 \tag{13}$$

enquanto o *j-ésimo* elemento fora da diagonal principal de $\boldsymbol{D}$ é

$$\mathbb{E}(e_i e_j|\boldsymbol{X}) = \mathbb{E}(e_i|\boldsymbol{x}_i)\mathbb{E}(e_j|\boldsymbol{x}_j) = 0. \tag{14}$$

em que a primeira igualdade usa a independência das observações e a segunda sai do fato de que $\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) = 0$.

- Então $\boldsymbol{D}$ é uma matriz diagonal com o *i-ésimo* elemento da diagonal σ_i^2:

$$\boldsymbol{D} = \operatorname{diag}\left(\sigma_1^2, ..., \sigma_n^2\right) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \tag{15}$$

- No caso especial do modelo de regressão linear homocedástico, temos:

$$\mathbb{E}\left(e_i^2|\boldsymbol{x}_i\right) = \mathbb{E}\left(e_i^2\right) = \sigma^2 \tag{16}$$

- Simplificando $\boldsymbol{D} = \boldsymbol{I}_n\sigma^2$.
- No modelo de regressão linear homocedástico, $\boldsymbol{D} = \boldsymbol{I}_n\sigma^2$, então $\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X} = \boldsymbol{X}'\boldsymbol{X}\sigma^2$, e a matriz de covariância é simplificada para $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \boldsymbol{X}'\boldsymbol{X}\sigma^2$.

Teorema (Estimador da Variância dos Mínimos Quadrados)

Em um modelo de regressão linear com amostra i.i.d.

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \mathrm{var}[\widehat{\boldsymbol{\beta}}|\boldsymbol{X}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{17}$$

em que $\boldsymbol{D}$ é definida pela eq. (15)*. Se o termo de erro é homocedástico, então a eq.* (17) *é simplificada para*

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{18}$$

Teorema de Gauss-Markov

- Considere a classe de estimadores de $\boldsymbol{\beta}$ que são funções lineares do vetor $\boldsymbol{y}$ e que pode ser escrito como

$$\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{y} \tag{19}$$

em que $\boldsymbol{A}$ é uma matriz $n \times k$ função de $\boldsymbol{X}$.

- O estimador de MQO é um caso especial fazendo $\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$.
- **Questão**: qual é a melhor escolha de $\boldsymbol{A}$?
- *Pelo Teorema de Gauss-Markov, o estimador de mínimos quadrados é a melhor escolha entre os estimadores lineares não viesados porque possui a menor variância entre todos os estimadores lineares não viesados.*

- Como $\mathbb{E}[\boldsymbol{y}|\boldsymbol{X}] = \boldsymbol{X}\boldsymbol{\beta}$, então para qualquer estimador linear $\widetilde{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{y}$, temos

$$\mathbb{E}\left[\widetilde{\beta}|\boldsymbol{X}\right] = \boldsymbol{A}'\mathbb{E}[\boldsymbol{y}|\boldsymbol{X}] = \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} \tag{20}$$

- Assim, $\widetilde{\boldsymbol{\beta}}$ é não viesado se e apenas se $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$ e

$$\text{var}\left[\widetilde{\beta}|\boldsymbol{X}\right] = \text{var}\left[\boldsymbol{A}'\boldsymbol{y}|\boldsymbol{X}\right] = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{A}\sigma^2 \tag{21}$$

a última igualdade vem do pressuposto de homocedasticidade $\boldsymbol{D} = \boldsymbol{I}_n\sigma^2$.

- O **melhor** estimador linear não viesado é obtido encontrando a matriz $\boldsymbol{A}_0$ que satisfaz a condição $\boldsymbol{A}_0'\boldsymbol{X} = \boldsymbol{I}_k$ tal que $\boldsymbol{A}_0'\boldsymbol{A}_0$ é minimizado no sentido positivo definido, o que significa que para qualquer outra matriz $\boldsymbol{A}$ que satisfaz $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$ então $\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{A}_0'\boldsymbol{A}_0$ é positiva semi-definida.

Teorema (Gauss-Markov)

Em um modelo de regressão linear homocedástico com amostra i.i.d., se $\widetilde{\beta}$ é um estimador linear não viesado de β, então

$$\operatorname{var}\left[\widetilde{\boldsymbol{\beta}}|\boldsymbol{x}\right] \geqslant \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{22}$$

***Intuição**: o teorema Gauss-Markov fornece um limite inferior na matriz de covariância de estimadores lineares não viesados sob a suposição de homocedasticidade. Por esse teorema, nenhum estimador linear não viesado pode ter uma matriz de variância menor (no sentido positivo definido) do que* $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Ausência de Viés

- Seja $\boldsymbol{A}$ ser uma função $n \times k$ de $\boldsymbol{X}$ tal que $\boldsymbol{A}'\boldsymbol{X} = \boldsymbol{I}_k$.
- O estimador $\boldsymbol{A}'\boldsymbol{y}$ é não viesado para $\boldsymbol{\beta}$ e tem variância $\boldsymbol{A}'\boldsymbol{A}\sigma^2$.
- Como o estimador de mínimos quadrados é não viesado e tem variância $(\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma^2$, é suficiente para mostrar que a diferença nas duas matrizes de variância é positiva semi-definida, ou:

$$\boldsymbol{A}'\boldsymbol{A} \geq \left(X'X\right)^{-1} \tag{23}$$

- Seja $\boldsymbol{C} = \boldsymbol{A} - \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$.
- Note que $\boldsymbol{X}'\boldsymbol{C} = 0$.

- Podemos calcular que:

$$
\begin{aligned}
\boldsymbol{A}'\boldsymbol{A} - (\boldsymbol{X}'\boldsymbol{X})^{-1} &= \left(\boldsymbol{C} + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)' \left(\boldsymbol{C} + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \\
&\quad - (\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \boldsymbol{C}'\boldsymbol{C} + \boldsymbol{C}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C} \\
&\quad + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} - (\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \boldsymbol{C}'\boldsymbol{C} \geq 0. \qquad (24)
\end{aligned}
$$

- A desigualdade final estabelece que a matriz $\boldsymbol{C}'\boldsymbol{C}$ é positiva semi-definida que é uma propriedade das formas quadráticas. Com isso demostramos a eq. (23).

Resíduos

- Vamos olhar as propriedades dos resíduos, $\widehat{e}_i = y_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}$, e dos erros de previsão $\widetilde{e}_i = y_i - \boldsymbol{x}_i\widehat{\boldsymbol{\beta}}_{(-i)}$. Podemos escrever os resíduos como:

$$\widehat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e} \tag{25}$$

em que $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ é a matriz de projeção ortogonal.

- Usando as propriedades da esperança condicional

$$\begin{aligned} \mathbb{E}(\widehat{\boldsymbol{e}}|\boldsymbol{X}) &= \mathbb{E}(\boldsymbol{M}\boldsymbol{e}|\boldsymbol{X}) \\ &= \boldsymbol{M}\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) = \boldsymbol{0} \end{aligned} \tag{26}$$

e

$$\begin{aligned} \operatorname{var}(\widehat{\boldsymbol{e}}|\boldsymbol{X}) &= \operatorname{var}(\boldsymbol{M}\boldsymbol{e}|\boldsymbol{X}) \\ &= \boldsymbol{M}\operatorname{var}(\boldsymbol{e}|\boldsymbol{X})\boldsymbol{M} = \boldsymbol{M}\boldsymbol{D}\boldsymbol{M} \end{aligned} \tag{27}$$

- Podemos simplificar essa expressão sob o pressuposto de homocedasticidade condicional: $\mathbb{E}(e^2|\boldsymbol{x}_i) = \sigma^2$. Neste caso, a eq. (27) é simplificada para

$$\text{var}(\widehat{\boldsymbol{e}}|\boldsymbol{X}) = \boldsymbol{M}\sigma^2$$

- Para uma única observação i, podemos encontrar a variância de $\widehat{e}_i$ considerando o *i-ésimo* elemento da diagonal de $\text{var}(\widehat{\boldsymbol{e}}|\boldsymbol{X}) = \boldsymbol{M}\sigma^2$. Assim temos:

$$\text{var}(\widehat{e}_i|\boldsymbol{X}) = \mathbb{E}(\widehat{e}_i^2|\boldsymbol{X}) = (1 - h_{ii})\sigma^2 \tag{28}$$

em que $h_{ii} = \boldsymbol{X_i'}\left(\boldsymbol{X'X}\right)^{\boldsymbol{-1}}\boldsymbol{X_i}$. Como esta variância é uma função de h_{ii} e, portanto, $\boldsymbol{x}_i$, os resíduos $\widehat{e}_i$ são heterocedásticos mesmo se os erros e_i são homocedásticos. Observe que eq. (28) implica $\widehat{e}_i^2$ é um estimador viesado de σ^2.

Erros de Previsão

- Vamos escrever os erros de previsão $\widetilde{e}_i = (1-h_{ii})^{-1}\widehat{e}_i$ em notação vetorial como $\widetilde{\boldsymbol{e}} = \boldsymbol{M}^*\widehat{\boldsymbol{e}}$ em que $\boldsymbol{M}^*$ é a matriz diagonal com *i-ésimo* elemento $(1-h_{ii})^{-1}$. Então $\widetilde{\boldsymbol{e}} = \boldsymbol{M}^*\boldsymbol{M}\boldsymbol{e}$ pode ser calculada como

$$\mathbb{E}(\widetilde{\boldsymbol{e}}|\boldsymbol{X}) = \boldsymbol{M}^*\boldsymbol{M}\mathbb{E}(\boldsymbol{e}|\boldsymbol{X}) = \boldsymbol{0} \tag{29}$$

e

$$\begin{aligned} \operatorname{var}(\widetilde{\boldsymbol{e}}|\boldsymbol{X}) &= \boldsymbol{M}^*\boldsymbol{M}\operatorname{var}(\boldsymbol{e}|\boldsymbol{X})\boldsymbol{M}\boldsymbol{M}^* \\ &= \boldsymbol{M}^*\boldsymbol{M}\boldsymbol{D}\boldsymbol{M}\boldsymbol{M}^* \end{aligned} \tag{30}$$

- Que simplifica sob homocedasticidade para

$$\begin{aligned} \operatorname{var}(\widetilde{\boldsymbol{e}}|\boldsymbol{X}) &= \boldsymbol{M}^*\boldsymbol{M}\boldsymbol{M}\boldsymbol{M}^*\sigma^2 \\ &= \boldsymbol{M}^*\boldsymbol{M}\boldsymbol{M}^*\sigma^2 \end{aligned} \tag{31}$$

- A variância do *i-ésimo* erro de previsão é então dada por:

$$
\begin{aligned}
\operatorname{var}(\widetilde{\boldsymbol{e}}|\boldsymbol{X}) &= \mathbb{E}(\widetilde{e}_i^2|\boldsymbol{X}) \\
&= (1-h_{ii})^{-1}(1-h_{ii})(1-h_{ii})^{-1}\sigma^2 \\
&= (1-h_{ii})^{-1}\sigma^2 \qquad (32)
\end{aligned}
$$

Erros Padronizados

- Um resíduo com variância condicional constante pode ser obtido reescalando. Os **resíduos padronizados** são:

$$\overline{e}_i = (1 - h_{ii})^{-1/2} \widehat{e}_i^2 \tag{33}$$

- Em notação vetorial

$$\overline{\boldsymbol{e}} = (\overline{e}_1, \ldots, \overline{e}_n)' = (\boldsymbol{M}^*)^{1/2} \boldsymbol{M}\boldsymbol{e} \tag{34}$$

- Pelos cálculos anteriores, sob homocedasticidade temos

$$\mathrm{var}(\overline{\boldsymbol{e}}|\boldsymbol{X}) = (\boldsymbol{M}^*)^{1/2}\, \boldsymbol{M}\, (\boldsymbol{M}^*)^{1/2}\, \sigma^2 \tag{35}$$

e

$$\mathrm{var}(\overline{e}_i|\boldsymbol{X}) = \mathbb{E}\left(\overline{e}_i^2|\boldsymbol{X}\right) = \sigma^2 \tag{36}$$

- Estes resíduos padronizados possuem o mesmo viés e variância como os erros originais quando os últimos são homocedásticos.

Estimação da Variância do Erro

- A variância do erro $\sigma^2 = \mathbb{E}\left(e_i^2\right)$ mensura a variação na parte "*não explicada*" da regressão.
- O seu estimador do método dos momentos é a média amostral dos resíduos ao quadrado:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{e}_i^2 \tag{37}$$

- No modelo de regressão linear podemos calcular a média de $\widehat{\sigma}^2$. Usando as propriedades das matrizes de projeção e o operador do traço, observe que:

$$\begin{aligned}\widehat{\sigma}^2 &= \frac{1}{n}\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} = \frac{1}{n}\boldsymbol{e}'\boldsymbol{M}\boldsymbol{M}\boldsymbol{e} = \frac{1}{n}\left(\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}\right)\\ &= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}\right) = \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{e}'\boldsymbol{e}\right)\end{aligned} \tag{38}$$

- Então,

$$\begin{aligned}\mathbb{E}\left(\widehat{\sigma}^2|\boldsymbol{X}\right) &= \frac{1}{n}\operatorname{tr}\left(\mathbb{E}(\boldsymbol{M}\boldsymbol{e}\boldsymbol{e}')|\boldsymbol{X}\right) \\ &= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\mathbb{E}(\boldsymbol{e}\boldsymbol{e}')|\boldsymbol{X}\right) \\ &= \frac{1}{n}\operatorname{tr}\left(\boldsymbol{M}\boldsymbol{D}\right) \qquad (39)\end{aligned}$$

- Inserindo o pressuposto de homocedasticidade $\mathbb{E}\left(e_i^2|\boldsymbol{x}_i\right) = \sigma^2$, então $\boldsymbol{D} = \boldsymbol{I}_n\sigma^2$.
- Assim, a eq. (39) é simplificada para

$$\begin{aligned}
\mathbb{E}(\widehat{\sigma}^2|\boldsymbol{X}) &= \frac{1}{n}\,\mathrm{tr}(\boldsymbol{M}\sigma^2) \\
&= \frac{\sigma^2}{n}\,\mathrm{tr}(\boldsymbol{M}) \\
&= \frac{\sigma^2}{n}\,\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') \\
&= \frac{\sigma^2}{n}\,\mathrm{tr}(\boldsymbol{I}_n - \boldsymbol{I}_k) \\
&= \sigma^2\left(\frac{n-k}{n}\right) \qquad (40)
\end{aligned}$$

- Esse resultado mostra que a σ^2 é viesada para zero. A ordem do viés depende de $\frac{k}{n}$, a razão do número de coeficientes estimados para o tamanho da amostra. Como o viés toma a forma escalar, um método clássico para obter um estimador não viesado é reescalando o estimador. Defina

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n} e_i^2 \tag{41}$$

- Pelas contas anteriores temos

$$\mathbb{E}\left(s^2|\boldsymbol{X}\right) = \widehat{\sigma}^2 \tag{42}$$

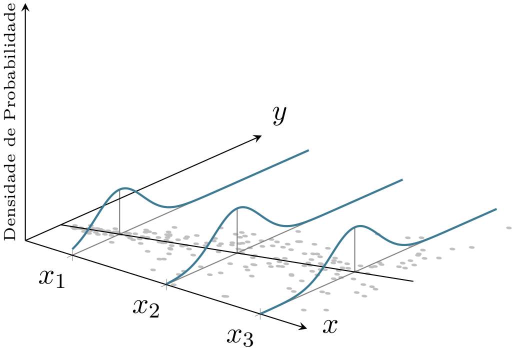
- Assim,

$$\mathbb{E}\left(s^2\right) = \sigma^2 \tag{43}$$

- O estimador s^2 é não viesado para σ^2.

Estimação da Matriz de Covariância sob Homocedasticidade

Figura 1: Caso Homocedástico



- Para fazer inferência necessitamos estimar a matriz de covariância $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ do estimador de mínimos quadrados. Vamos considerar o modelo de regressão linear.
- Sob homocedasticidade, a matriz de covariância toma uma forma relativamente simples

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2 \tag{44}$$

que é conhecida para o desconhecido escalar σ^2. O estimador mais comum utilizado para σ^2 é s^2, levando ao estimador clássico da matriz de covariância

$$\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}s^2 \tag{45}$$

- Como s^2 é condicionalmente não viesado para σ^2, é fácil calcular que $\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}}$ é condicionalmente não viesada para $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ sob o pressuposto de homocedasticidade:

$$
\begin{aligned}
\mathbb{E}\left(\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}}\middle|\boldsymbol{X}\right) &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\mathbb{E}\left(s^2\middle|\boldsymbol{X}\right) \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\sigma^2 \\
&= \boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} \qquad (46)
\end{aligned}
$$

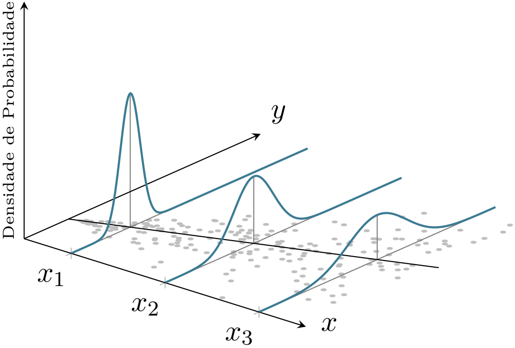
- Este estimador era o estimador de matriz de covariância dominante em econometria aplicada nos últimos tempos, e ainda é *default* na maioria dos *softwares* de regressão.

- Se o estimador dado pela eq. (45) é utilizado, mas o erro da regressão é **heterocedástico**, é possível que $\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}}$ seja totalmente viesado para a matriz de covariância correta:

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{47}$$

Estimação da Matriz de Covariância sob Heterocedasticidade

Figura 2: Caso Heterocedástico



- Para qualquer matriz $n \times r$ $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{X})$,

$$\text{var}(\boldsymbol{A}'\boldsymbol{y}|\boldsymbol{X}) = \text{var}(\boldsymbol{A}'\boldsymbol{e}|\boldsymbol{X}) = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} \tag{48}$$

- Em particular, podemos escrever $\widehat{\boldsymbol{\beta}} = \boldsymbol{A}'\boldsymbol{y}$ onde $\boldsymbol{A} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ e assim

$$\text{var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{A}'\boldsymbol{D}\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{49}$$

- É útil notar que

$$(\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}) = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\sigma_i^2 \tag{50}$$

é uma versão ponderada de $\boldsymbol{X}'\boldsymbol{X}$.

- Se a hipótese de variância constante, homocedasticidade, não se verificar, o estimador da matriz de variância pode ser altamente viesado.
- A forma geral da matriz de covariância é dada por:

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{51}$$

- Como pode ser observado, essa matriz depende de $\boldsymbol{D}$ que é uma matriz desconhecida e que pode ser escrita como:

$$\boldsymbol{D} = \operatorname{diag}\left\{\sigma_1^2, \ldots, \sigma_n^2\right\} = \mathbb{E}\left(\boldsymbol{e}\boldsymbol{e}'|\boldsymbol{X}\right) = \mathbb{E}(\boldsymbol{D}_0|\boldsymbol{X}) \tag{52}$$

em que $\boldsymbol{D}_0 = \operatorname{diag}\left(e_1^2, \ldots, e_n^2\right)$. Assim, $\boldsymbol{D}_0$ é um estimador não viesado para $\boldsymbol{D}$.

Matriz de Variância Ótima

- Se os erros ao quadrado e_i^2 fossem observados, poderíamos construir um estimador não viesado dado por

$$\begin{aligned}\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{ideal} &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}_0\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\ &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' e_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \qquad (53)\end{aligned}$$

- Usando o operador de esperança condicional, temos:

$$
\begin{aligned}
\mathbb{E}\left(\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{ideal}|\boldsymbol{X}\right) &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}_0\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\mathbb{E}\left(e_i^2\right)\Bigg|\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\sigma_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} \qquad (54)
\end{aligned}
$$

- Mostrando que $\mathbb{E}\left(\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{ideal}|\boldsymbol{X}\right)$ é não viesado para $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$.
- Como e_i^2 não são observados, $\mathbb{E}\left(\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{ideal}|\boldsymbol{X}\right)$ **não é um estimador factível**.
- **Qual é a alternativa?** Podemos usar os resíduos ao quadrado, $\widehat{e}_i^2$, os erros de previsão, $\widetilde{e}_i^2$ ou os resíduos padronizados, $\overline{e}_i^2$, para construir um estimador factível.

- Seja

$$\widehat{\boldsymbol{D}} = \operatorname{diag}\left(\widehat{e}_i^2, \ldots, \widehat{e}_n^2\right), \tag{55}$$

$$\widetilde{\boldsymbol{D}} = \operatorname{diag}\left(\widetilde{e}_i^2, \ldots, \widetilde{e}_n^2\right), \tag{56}$$

$$\overline{\boldsymbol{D}} = \operatorname{diag}\left(\overline{e}_i^2, \ldots, \overline{e}_n^2\right) \tag{57}$$

- Obtemos os estimadores substituindo essas matrizes na eq. (51) para $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$. Assim,

$$\begin{aligned} \widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC0} &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\frac{1}{n}\boldsymbol{X}'\widehat{\boldsymbol{D}}\boldsymbol{X}\right) \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\ &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2\right) \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \end{aligned} \tag{58}$$

em que o rótulo HC0 se refere a ser estimador *baseline* da matriz de covariância consistente a heterocedasticidade (White, 1980, Econometrica).

- Como $\widehat{e}_i^2$ é viesado para zero, para estimar a σ^2 o estimador não viesado s^2 pode ajustar o estimador de momentos $\widehat{\sigma}^2$ pelos graus de liberdade $\left(\frac{n}{n-k}\right)$.
- O estimador HC1 é recomendado em relação ao HC0.

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC1} = \left(\frac{n}{n-k}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{59}$$

- Uma alternativa é ajustar pela matriz de projeção P (MacKinnon and White, 1985):

$$\begin{aligned}
\overline{\boldsymbol{V}}^{HC2}_{\widehat{\boldsymbol{\beta}}} &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\overline{\boldsymbol{D}}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\overline{e}_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(1-h_{ii})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}
\end{aligned} \tag{60}$$

- Por fim, temos uma variação do HC2 (Davidson and MacKinnon, 1993):

$$
\begin{aligned}
\widetilde{\boldsymbol{V}}_{\widehat{\beta}}^{HC3} &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\widetilde{\boldsymbol{D}}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\widetilde{e}_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \\
&= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(1-h_{ii})^{-2}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}
\end{aligned}
\tag{61}
$$

- Os estimadores HC0, HC1, HC2 e HC3 são chamados de estimadores da matriz de variância robusto ou robusto à heterocedasticidade.
- **HC0** é o estimador da matriz de variância de White; **HC1** é o estimador da matriz de variância robusto ajustado pelos graus de liberdade.

- **Resumindo**: temos cinco estimadores da matriz de variância, incluindo o estimador homocedástico $\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}}$ e os quatro estimadores HC.
- **Qual deveríamos utilizar?**
  1. A escolha do estimador $\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}}$ é pobre na medida em que ele é válido apenas sobre a improvável restrição de homocedasticidade.
  2. Dos estimadores robustos *HC*, o HC1 é o mais usado é *default* em programas como o Stata. Mas os estimadores HC2 e HC3 são preferíveis.
- O estimador HC2 é não viesado sob homocedasticidade e o estimador HC3 é conservador para qualquer $\boldsymbol{X}$. Na maioria das aplicações HC1, HC2 e HC3 serão similares.

Erro Padrão

- O estimador da variância tal como $n^{-1}\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}$ é um estimador da variância da distribuição de $\widehat{\boldsymbol{\beta}}$. Uma medida facilmente interpretável do risco é sua raiz quadrada – desvio-padrão.

Definição

Um **erro padrão** $s(\widehat{\boldsymbol{\beta}})$ para um estimador β com valor real é uma estimativa do desvio-padrão da distribuição de $\widehat{\boldsymbol{\beta}}$. Isto é

$$s\left(\widehat{\boldsymbol{\beta}_j}\right) = \sqrt{n^{-1}\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}_j}} = n^{-1/2}\sqrt{\left[\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}_{jj}}}\right]} \tag{62}$$

Mínimos Quadrados Generalizados

- Considere o modelo de regressão linear:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \tag{63}$$

- Considere uma situação onde as observações dos **erros são heterocedásticos**. Suponha que

$$\mathbb{E}[\boldsymbol{e}|\boldsymbol{X}] = \mathbf{0} \tag{64}$$

$$\operatorname{var}[\boldsymbol{e}|\boldsymbol{X}] = \Omega \tag{65}$$

- A matriz de variância Ω possui dimensão $n \times n$ e é possivelmente uma função de $\boldsymbol{X}$. A estrutura da amostra é *i.i.d.* em que $\Omega = \boldsymbol{D}$, permitindo matrizes de covariâncias não diagonais também. A matriz Ω é simétrica e positiva semi-definida.

- Considerando esses pressupostos, podemos calcular a média e a variância do estimador de mínimos quadrados:

$$\mathbb{E}\left[\widehat{\beta}|\boldsymbol{X}\right] = \beta \tag{66}$$

$$\operatorname{var}\left[\widehat{\beta}|\boldsymbol{X}\right] = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}(\boldsymbol{X}'\Omega\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{67}$$

- Considere que a matriz Ω seja conhecida. Suponha que ela tenha a forma $\Omega = c^2\Sigma$ em que $c^2 > 0$ é um numero real e Σ é conhecida e tem dimensão $n \times n$.
- Pré-multiplique o modelo linear $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ por $\Sigma^{-1/2}$. Isto irá gerar a equação

$$\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \widetilde{\boldsymbol{e}} \tag{68}$$

em que $\widetilde{\boldsymbol{y}} = \Sigma^{-1/2}\boldsymbol{y}$, $\widetilde{\boldsymbol{X}} = \Sigma^{-1/2}\boldsymbol{X}$ e $\widetilde{\boldsymbol{e}} = \Sigma^{-1/2}\boldsymbol{e}$.

- O estimador de mínimos quadrados de β nesta equação é dado por:

$$\begin{aligned}
\widetilde{\boldsymbol{\beta}}_{MQG} &= \left(\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}}\right)^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{y}} \\
&= \left(\left(\Sigma^{-1/2}\boldsymbol{X}\right)'\left(\Sigma^{-1/2}\boldsymbol{X}\right)^{-1}\right)\boldsymbol{X}'\boldsymbol{y} \\
&= \left(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X}\right)\boldsymbol{X}'\Sigma^{-1}\boldsymbol{y} \qquad (69)
\end{aligned}$$

- Este é o estimador de **Mínimos Quadrados Generalizados (MQG)** de β. Podemos calcular:

$$\mathbb{E}\left[\widetilde{\boldsymbol{\beta}}_{MQG}|\boldsymbol{X}\right] = \boldsymbol{\beta} \qquad (70)$$

$$\text{var}\left[\widetilde{\boldsymbol{\beta}}_{MQG}|\boldsymbol{X}\right] = \left(\boldsymbol{X}'\Omega^{-1}\boldsymbol{X}\right)^{-1} \qquad (71)$$

- Isto mostra que o estimador de MQG é não viesado e possui uma matriz de variância que é igual ao limite inferior do Teorema de Gauss-Markov Generalizado.

- Isso mostra que o limite inferior é nítido quando Σ é conhecido. **MQG, portanto, é eficiente na classe dos estimadores lineares não viesados**.

Teorema (Gauss-Markov Generalizado)

No modelo de regressão linear e $\Sigma > 0$, se $\widetilde{\boldsymbol{\beta}}$ é um estimador linear não viesado de $\boldsymbol{\beta}$, então

$$\operatorname{var}\left[\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}\right] \geq \left(\boldsymbol{X}'\Omega^{-1}\boldsymbol{X}\right)^{-1} \tag{72}$$

- **Intuição**: o teorema fornece um limite inferior na matriz de variância dos estimadores lineares não viesados.
- O limite é diferente da matriz de variância do estimador de mínimos quadrados exceto quando $\Sigma = \boldsymbol{I}_n\sigma^2$.

- No modelo de regressão linear com observações independentes e matriz variância condicional conhecida, tal que $\Omega = \Sigma = \boldsymbol{D} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)$, o estimador de MQG toma a forma:

$$\begin{aligned}\widetilde{\beta}_{MQG} &= \left(\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{y} \\ &= \left(\sum_{i=1}^{n}\sigma_i^{-2}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\sum_{i=1}^{n}\sigma_i^{-2}\boldsymbol{x}_i\boldsymbol{y}_i\right) \qquad (73)\end{aligned}$$

- O pressuposto $\Omega > 0$ neste caso se reduz para $\sigma_i^2 > 0$ para $i = 1, \ldots, n$.
- **Nota**: na prática, a matriz de variância Ω é não conhecida e, assim, o estimador factível como descrito aqui não é factível.

Medidas de Ajuste

- A medida mais comum reportada do ajuste da regressão é o R^2 definido como

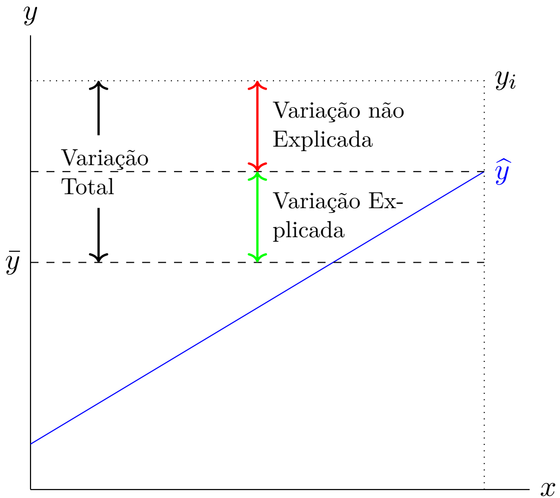
$$\boldsymbol{R}^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{e}_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = 1 - \frac{\widehat{\sigma}^2}{\widehat{\sigma}_y^2} \tag{74}$$

em que $\widehat{\sigma}_y^2 = n^{-1} \sum_{i=1}^{n} (y_i - \overline{y})^2$. O R^2 pode ser visto como um estimador do parâmetro populacional

$$\rho^2 = \frac{\text{var}(\boldsymbol{x}_i' \boldsymbol{\beta})}{\text{var}(y_i)} = 1 - \frac{\sigma^2}{\sigma_y^2} \tag{75}$$

- Contudo, $\widehat{\sigma}^2$ e $\widehat{\sigma}_y^2$ são estimadores viesados.

Figura 3: Geometria do R^2



- Theil (1961) propôs substituir estes pela versão não viesada s^2 e $\tilde{\sigma}_y^2 = (n-1)^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$ gerando o R^2 ajustado $\left(\overline{R}^2\right)$:

$$\begin{aligned} \overline{\boldsymbol{R}}^2 &= 1 - \frac{s^2}{\tilde{\sigma}_y^2} \\ &= 1 - \frac{(n-1)^{-1}\sum_{i=1}^{n}\widehat{e}_i^2}{(n-k)^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^2} \end{aligned} \tag{76}$$

Intuição

- **Caso 1:** Considere a matriz a seguir

$$\boldsymbol{X} = \begin{bmatrix} 2 & 1 & 1 \\ 5 & 2 & 3 \\ -3 & 1 & -4 \end{bmatrix} \text{ e } (\boldsymbol{X}'\boldsymbol{X}) = \begin{bmatrix} 2 & 5 & -3 \\ 1 & 2 & 1 \\ 1 & 3 & -4 \end{bmatrix}' \times \begin{bmatrix} 2 & 1 & 1 \\ 5 & 2 & 3 \\ -3 & 1 & -4 \end{bmatrix}$$

- Na matriz $\boldsymbol{X}$ a coluna 1 é uma combinação linear das colunas 2 e 3. Isso faz com que a matriz $\boldsymbol{X}'\boldsymbol{X}$ seja singular. Por isso, $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ e $\widehat{\boldsymbol{\beta}}$ não são definidos.

- **Caso 2:** Considere a matriz

$$\boldsymbol{X} = \begin{bmatrix} 2 & 1 & 1 \\ 5 & 2 & 3 \\ -3 & 1 & -6 \end{bmatrix} \text{ e } (\boldsymbol{X}'\boldsymbol{X}) = \begin{bmatrix} 2 & 5 & -3 \\ 1 & 2 & 1 \\ 1 & 3 & -6 \end{bmatrix}' \times \begin{bmatrix} 2 & 1 & 1 \\ 5 & 2 & 3 \\ -3 & 1 & -6 \end{bmatrix}$$

- Na matriz $\boldsymbol{X}$ a coluna 1 é quase uma combinação linear das colunas 2 e 3. Isso faz com que a matriz $\boldsymbol{X}'\boldsymbol{X}$ seja quase singular. Mesmo assim, $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ e $\widehat{\boldsymbol{\beta}}$ são definidos.

Caso 1 Pode ser definido como **estrita multicolinearidade** ou **perfeita multicolinearidade**.

- Ocorre quando as colunas de $\boldsymbol{X}$são linearmente dependentes, isto é, existe algum $\boldsymbol{\alpha} \neq 0$ tal que $\boldsymbol{X\alpha} = \mathbf{0}$.
- Esse caso **ocorre quando um conjunto de regressores são incluídos que são identicamente relacionados**.
- **Por exemplo:** se $\boldsymbol{X}$ inclui ambos o log de dois preços e o log dos preços relativos ($\log(p_1), \log(p_2)$ e $\log(p_1/p_2)$), então $\boldsymbol{X'X}$ necessariamente será singular.
- Não será possível calcular a $(\boldsymbol{X'X})^{-1}$ e obter o $\widehat{\boldsymbol{\beta}}$.

Caso 2 Pode ser definido como **quase multicolinearidade** ou **multicolinearidade**.

- Esta situação ocorre quando os regressores são altamente correlacionados.
- Uma implicação da multicolinearidade é que as estimativas dos coeficientes individuais serão imprecisas. Não chega a ser um problema para análise econométrica se os erros-padrão reportados são precisos (eficientes).
- Porém, erros-padrão robustos pode ser sensíveis a pequenas alterações nos dados sob multicolinearidade. Isto leva a uma situação indesejável em que as estimativas são imprecisas, embora os erros-padrão sejam enganosamente pequenos.
- Não será possível calcular $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ e obter o $\widehat{\boldsymbol{\beta}}$.

- Vamos olhar esse último caso mais detalhadamente. Considere um modelo de regressão linear com apenas dois regressores:

$$y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + e_i \tag{77}$$

e

$$\frac{1}{n}\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{78}$$

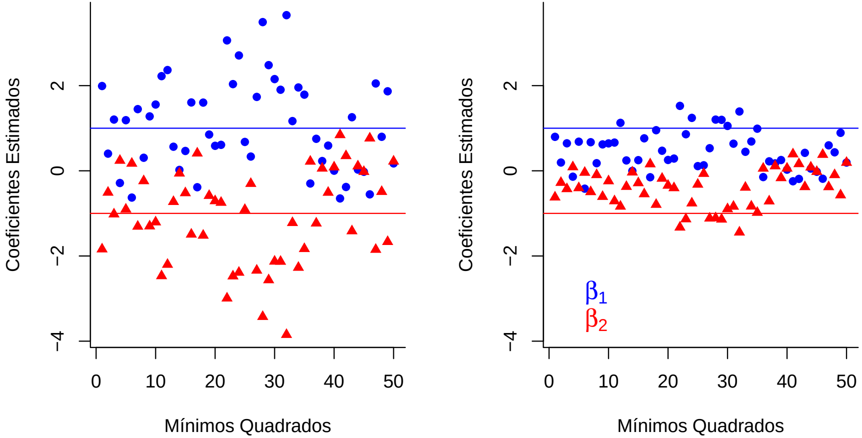
- Neste caso temos:

$$\operatorname{var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \frac{\sigma^2}{n}\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \frac{\sigma^2}{n(1-\rho^2)}\begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \tag{79}$$

- A correlação ρ indexa a colinearidade. **Quando ρ aproxima de 1, a matriz se torna singular**.

- E a precisão das estimativas dos coeficientes? Para ver o efeito da multicolinearidade na precisão das estimativas, podemos observar a variância de uma estimativa dos coeficientes.
- Quando ρ se aproxima de 1, o termo $\sigma^2[n(1-\rho^2)]^{-1}$ se aproxima do infinito. **Por isso, quanto mais "colinear" são os regressores, pior é a precisão das estimativas dos coeficientes individuais**.
- O que está acontecendo é que, quando os regressores são altamente dependentes, é estatisticamente difícil separar o impacto de β_1 de β_2. Como consequência, a precisão das estimativas individuais são reduzidas.

Figura 4: Exemplo Geométrico da Multicolinearidade



Perspectiva Alternativa

- Vamos utilizar o modelo linear $\boldsymbol{y} = \boldsymbol{X}\beta + e$, onde e é o erro aleatório, $\boldsymbol{X}$ é o regressor e y é o regressando. Às vezes $\boldsymbol{X}$ vai ser um único regressando, às vezes X vai ser uma matriz de variáveis.
- Teremos n observações e p variáveis – então X é a matriz $n \times p$.
- No curso de econometria I da graduação, aprendemos que o estimador de mínimos quadrados para o caso univariado é:

$$\widehat{\beta}_{MQO} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{80}$$

sendo $\bar{x}$ a média de x e $\bar{y}$ a média de y.

- A versão da pós do estimador de mínimos quadrados é:

$$\widehat{\beta}_{MQO} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{81}$$

- Isso é exatamente a mesma coisa que a versão acima, mas como agora tem mais de uma variável nós usamos matrizes.
- $\boldsymbol{X}'\boldsymbol{X}$ é a variância de $\boldsymbol{X}$ e $\boldsymbol{X}'\boldsymbol{y}$ a covariância entre $\boldsymbol{X}$ e $\boldsymbol{y}$.
- Para essa fórmula funcionar, $\boldsymbol{X}'\boldsymbol{X}$ tem que ser inversível, e portanto tem que ter *rank* completo. Nós usualmente pensamos isso como X não podendo ter duas variáveis que são funções afim uma da outra.

- Existem várias maneiras de introduzir o estimador de mínimos quadrados. A primeira advém da condição de momentos $\mathbb{E}[\boldsymbol{e}|\boldsymbol{X}] = 0$, ou mais fraco, a covariância entre $\boldsymbol{X}$ e $\boldsymbol{e}$ é zero.
- Outra maneira é pensar no problema de minimização:

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i'\beta)^2 \tag{82}$$

- Que pode ser escrito como

$$\|\boldsymbol{y} - \boldsymbol{X}\beta\|_2^2 = \sum_{i=1}^{n} (y_i - x_i'\beta)^2 \tag{83}$$

- Ainda existe outra maneira de motivar mínimos quadrados: suponha que y e x são dois vetores no $\mathbb{R}^n$.
- Vamos usar $\langle x, y\rangle$ como o produto interno $\left(\langle x, y\rangle = \sum_i x_i y_i\right)$.
- Como nós podemos transformar y de maneira que $\langle y - cx, x\rangle = 0$?
- Noutras palavras, nós queremos deixar os dois vetores ortogonais.

- No $\mathbb{R}^2$, isso significa formar um ângulo reto:

$$\langle y - cx, x\rangle = 0 \therefore \sum_i (y_i - cx_i)x_i = \sum_i y_i x_i - c\sum_i x_i^2 = 0$$
$$\sum_i x_i y_i = c\sum_i x_i^2$$
$$c = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \qquad (84)$$

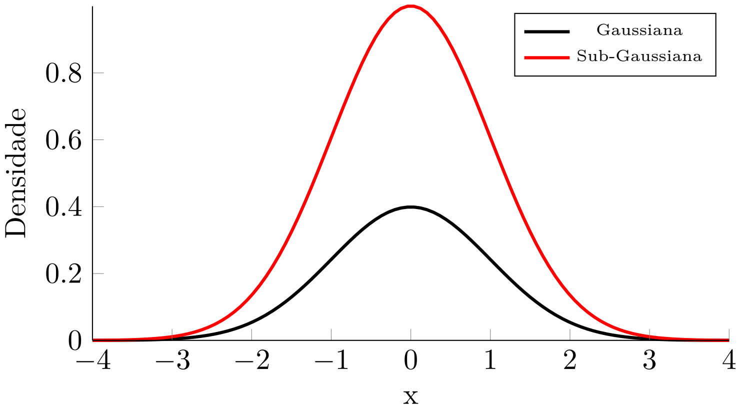
- Isso é exatamente a primeira expressão que escrevemos.

Prova da Consistência

1. Para variáveis subgaussianas vale a seguinte desigualdade

$$P(|X| > t) \leq e^{-\frac{t^2}{2\sigma^2}} \tag{85}$$

- Cotar a probabilidade do máximo delas ser maior que um valor t é importante por alguns motivos, entre eles:
    - Nós frequentemente trabalhamos com estimadores que minimizam ou maximizam alguma função: mínimos quadrados, máxima verossimilhança. É natural que estes estimadores dependam do máximo de uma variável aleatória.
    - Se você está trabalhando com algum processo aleatório, muitas vezes o máximo pode ser mortal: qual é o máximo que um ativo pode perder se a distribuição dos retornos é subgaussiana, por exemplo?


Gaussiana
Sub-Gaussiana
Densidade
0
0.2
0.4
0.6
0.8
−4
−3
−2
−1
0
1
2
3
4
x

2. Seja a definição de norma euclidiana[1]:

$$\|X\|_2^2 = X'X = \sum_i x_i^2 \tag{86}$$

Nós vamos usar o formato de multiplicação de matriz $(X'X)$ para facilitar as contas.

3. Se uma matriz é de rank completo, então nenhum autovalor é zero.

4. Seja $\lambda_{\min}$ o menor autovalor da matriz $\dfrac{X'X}{n}$. Então:

$$\lambda_{\min} \leq \frac{\frac{1}{n}\|Xv\|_2^2}{\|v\|_2^2} \tag{87}$$

para qualquer v.

[1] A norma euclidiana associa um número real a cada matriz X, e podemos interpretar $\|X\|_2^2$ como o valor singular máximo de X $\left(\|X\|_2 = \left(\lambda_{\max}\left(X'X\right)\right)^{1/2}\right)$.

5. Seja $x \in \mathbb{R}^n$, então a norma p é representada por $\|x\|_p$ e

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p} \tag{88}$$

- Veja que nós podemos pensar a norma do vetor x como a distância entre o vetor x e a origem. A norma euclidiana é o caso $p = 2$.
- Em particular, nós definimos a "norma sup" como o máximo do módulo do vetor e representamos por $\|x\|_\infty$:

$$\|X\|_\infty = \max_{i=1,\ldots,n} |x_i| \tag{89}$$

6. Nós vamos usar a desigualdade de Hölder, que diz que se p e q são tais que $\frac{1}{p} + \frac{1}{q} = 1$, então:

$$|\langle x, y\rangle| \leq \|x\|_p \|y\|_q \tag{90}$$

- Lembre que $\langle x, y\rangle$ representa o produto interno.
- A desigualdade de Hölder vale para $p = 1$ e $q = \infty$:

$$|\langle x, y\rangle| \leq \|x\|_1 \|y\|_\infty \tag{91}$$

- Um caso particular de Hölder é a desigualdade de Cauchy-Schwartz:

$$|\langle x, y\rangle| \leq \|x\|_2 \|y\|_2 \tag{92}$$

- Uma aplicação que vai ser útil de Cauchy-Schwartz é a seguinte desigualdade de normas

$$\|x\|_1 \leq \sqrt{n}\|x\|_2 \tag{93}$$

- Veja que $\|x\|_1 = \sum_i |x_i|$. Nós podemos representar o módulo como a multiplicação de x_i pela função sinal de x_i:

$$\text{sinal}(x_i) = \begin{cases} 1, \text{ se } x_i \geq 0 \\ -1, \text{ se } x_i < 0 \end{cases} \tag{94}$$

- Se x_i é positivo, então sinal$(x_i) = 1$; se x_i é negativo, sinal$(x_i) = -x_i$.

- Logo:

$$\|x\|_1 = \langle \text{sinal}(x), x\rangle \leq \|\text{sinal}(x)\|_2 \|x\|_2 \tag{95}$$

em que sinal(x) é só um vetor $(\text{sinal}(x_1), \ldots, \text{sinal}(x_n))$.

- Agora use a definição de $\|.\|_2$:

$$\|\text{sinal}(x)\|_2 = \left(\sum_{i=1}^{n} (\text{sinal}(x_i))^2\right)^{1/2} \tag{96}$$

- Como o sinal é sempre 1 ou -1, então o quadrado é sempre 1 e nós temos:

$$\|\text{sinal}(x)\|_2 = \left(\sum_{i=1}^{n} (\text{sinal}(x_i))^2\right)^{1/2} = \left(\sum_{i=1}^{n} 1\right)^{1/2} = (n)^{1/2} \tag{97}$$

- Sejam y e X os dados. Vamos supor que o ruído e é subgaussiano com parâmetro σ. Basicamente, estamos exigindo que a cauda da distribuição caia tão rápido quanto a gaussiana. Distribuições que atendem a esse requisito são chamadas de subgaussianas.

- Daqui por diante, nós representamos o estimador de Mínimos Quadrados por $\widehat{\beta}$.

- Como o estimador de mínimos quadrados resolve um problema de minimização de $\sum_i (y_i - X\beta)^2$ (que representamos como $\|Y - X\beta\|_2^2$), então para qualquer outro vetor β:

$$\frac{1}{n}\|Y - X\beta\|_2^2 \leq \frac{1}{n}\|Y - X\widehat{\beta}\|_2^2 \tag{98}$$

- Como isso é verdade para qualquer outro vetor β, isso também é verdade para o vetor β_0, de parâmetros verdadeiros:

$$\frac{1}{n}\|Y - X\widehat{\beta}\|_2^2 \leq \frac{1}{n}\|Y - X\beta_0\|_2^2 \tag{99}$$

- Nós sabemos que $Y = X\beta_0 + e$. Vamos substituir isso no resultado acima:

$$\begin{aligned} & \frac{1}{n}\|X\beta_0 + e - X\widehat{\beta}\|_2^2 \leq \frac{1}{n}\|X\beta_0 + e - X\beta_0\|_2^2 \\ \therefore \quad & \frac{1}{n}\|X(\beta_0 - \widehat{\beta}) + e\|_2^2 \leq \frac{1}{n}\|e\|_2^2 \end{aligned} \tag{100}$$

- Usando o ponto 2 acima no termo $\|X(\beta_0 - \widehat{\beta}) + e\|_2^2$, temos:

$$
\begin{aligned}
&\|X(\beta_0 - \widehat{\beta}) + e\|_2^2 = (X(\beta_0 - \widehat{\beta}) + e)'(X(\beta_0 - \widehat{\beta}) + e) \\
&= ((\beta_0 - \widehat{\beta})'X' + e')(X(\beta_0 - \widehat{\beta}) + e) \\
&= (\beta_0 - \widehat{\beta})'X'X(\beta_0 - \widehat{\beta}) + (\beta_0 - \widehat{\beta})'X'e + e'X(\beta_0 - \widehat{\beta}) + e'e \\
&= \|X(\beta_0 - \widehat{\beta})\|_2^2 + 2e'X(\beta_0 - \widehat{\beta}) + \|e\|_2^2 \qquad (101)
\end{aligned}
$$

- Sabendo que $e'X(\beta_0 - \widehat{\beta}) = (\beta_0 - \widehat{\beta})'X'e$, encontramos:

$$
\frac{1}{n}\|X(\beta_0 - \widehat{\beta})\|_2^2 + \frac{2}{n}e'X(\beta_0 - \widehat{\beta}) + \frac{1}{n}\|e\|_2^2 \leq \frac{1}{n}\|e\|_2^2 \qquad (102)
$$

- Podemos cancelar $\frac{1}{n}\|e\|_2^2$:

$$\begin{aligned} \frac{1}{n}\|X(\beta_0 - \widehat{\beta})\|_2^2 + \frac{2}{n}e'X(\beta_0 - \widehat{\beta}) &\leq 0 \\ \frac{1}{n}\|X(\beta_0 - \widehat{\beta})\|_2^2 &\leq \frac{2}{n}e'X(\widehat{\beta} - \beta_0) \end{aligned} \tag{103}$$

- Agora, $e'X(\widehat{\beta} - \beta_0)$ é um escalar e nós podemos ver isso como o produto interno de e e $X(\widehat{\beta} - \beta_0)$. Nós podemos usar Hölder:

$$\frac{1}{n}\left\langle e'X, (\widehat{\beta} - \beta_0)\right\rangle \leq \frac{1}{n}\left|\left\langle e'X, (\widehat{\beta} - \beta_0)\right\rangle\right| \leq \|e'X\|_\infty \|\widehat{\beta} - \beta_0\|_1 \tag{104}$$

- Vamos sair pela tangente aqui para tratar de $\|e'X\|_\infty$.
- Veja que isso é a norma de uma variável aleatória (já que e é aleatório). Pelo ponto 1 acima e fazendo X_i representar a i-ésima coluna de X:

$$\mathbb{P}\left(\max_{i=1,\ldots,p} \frac{|e'x_i|}{n} > t\right) \leq p \exp\left(-\frac{t^2}{2\sigma^2/n^2}\right) \tag{105}$$

- Para $t = \frac{\sigma}{n}\sqrt{2(\log(p) + \delta)}$, com $\delta > 0$, temos:

$$\begin{aligned} P\left(\max_{i=1,\ldots,p} \frac{|e'x_i|}{n} > t\right) &\leq p \exp\left(-\frac{\frac{2\sigma^2}{n^2}\left(\log(p) + \delta\right)}{\frac{2\sigma^2}{n^2}}\right) \\ &= p \exp\left(-\log(p) - \delta\right) \\ &= \exp(-\delta) \end{aligned} \tag{106}$$

- Com alta probabilidade, $\frac{\sigma}{n}\sqrt{2(\log(p)+\delta)}$ é maior que o máximo. Vamos substituir esse valor na nossa cota:

$$\begin{aligned} \left\langle e'X, (\widehat{\beta}-\beta_0)\right\rangle \leq \left|\left\langle e'X, (\widehat{\beta}-\beta_0)\right\rangle\right| \leq \|e'X\|_\infty \|\widehat{\beta}-\beta_0\|_1 \\ \left\langle e'X, (\widehat{\beta}-\beta_0)\right\rangle \leq \frac{\sigma}{n}\sqrt{2(\log(p)+\delta)}\|\widehat{\beta}-\beta_0\|_1 \end{aligned} \tag{107}$$

- Argumentamos que $\|x\|_1 \leq \sqrt{n}\|x\|_2$ (ponto 6). Vamos usar isso agora com $\|\widehat{\beta}-\beta_0\|_1$.

$$\begin{aligned} \left\langle e'X, (\widehat{\beta}-\beta_0)\right\rangle &\leq \frac{\sigma}{n}\sqrt{2(\log(p)+\delta)}\|\widehat{\beta}-\beta_0\|_1 \\ &\leq \frac{\sigma}{n}\sqrt{2(\log(p)+\delta)}\sqrt{n}\|\widehat{\beta}-\beta_0\|_2 \end{aligned} \tag{108}$$

- Vamos jogar isso de volta na equação (103):

$$\frac{1}{n}\|X(\widehat{\beta}-\beta_0)\|_2^2 \leq 2\sigma\sqrt{\frac{2(\log(p)+\delta)}{n}}\|\widehat{\beta}-\beta_0\|_2 \tag{109}$$

- Multiplique e divida o lado direito por $\|\widehat{\beta}-\beta_0\|_2$:

$$\frac{1}{n}\|X(\widehat{\beta}-\beta_0)\|_2^2 \leq 2\sigma\sqrt{\frac{2(\log(p)+\delta)}{n}}\frac{\|\widehat{\beta}-\beta_0\|_2^2}{\|\widehat{\beta}-\beta_0\|_2} \tag{110}$$

- Reorganizando:

$$\frac{1}{n}\frac{\|X(\widehat{\beta}-\beta_0)\|_2^2}{\|\widehat{\beta}-\beta_0\|_2^2} \leq 2\sigma\sqrt{\frac{2(\log(p)+\delta)}{n}}\frac{1}{\|\widehat{\beta}-\beta_0\|_2} \tag{111}$$

- Uma boa hora para usar o nosso resultado 4, sobre o autovalor da matriz:

$$\lambda_{\min} \leq \frac{1}{n}\frac{\|X(\widehat{\beta}-\beta_0)\|_2^2}{\|\widehat{\beta}-\beta_0\|_2^2} \leq 2\sigma\sqrt{\frac{2(\log(p)+\delta)}{n}}\frac{1}{\|\widehat{\beta}-\beta_0\|_2} \tag{112}$$

- Reorganizando a expressão acima:

$$\|\widehat{\beta} - \beta_0\|_2 \leq \frac{2\sigma}{\lambda_{\min}} \sqrt{\frac{2(\log(p) + \delta)}{n}} \tag{113}$$

Isso é bem legal, porque nos diz várias coisas:

- Conforme n cresce, a diferença entre a estimativa de MQO e o vetor verdadeiro cai para zero. Isso é consistência.
- Quanto mais variáveis nós temos, pior a nossa vida em termos de consistência. Mas veja que o termo em cima piora com a raiz quadrada do log de p. Isso é extremamente benevolente.
- Veja que se X for uma matriz de variáveis descorrelacionadas, então o menor autovalor de $X'X$ é a menor variância das variáveis do lado direito da equação.
- Veja que na verdade podemos fazer todas as contas sem a hipótese de ortogonalidade entre X e e. Foi necessário cotar a covariância entre x e e.

Econometria I

Regressão de Mínimos Quadrados

Victor Oliveira
Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2024