Econometria I

Propriedades de Convergência Fraca para Funções de Grandes Amostras: O Caso dos Mínimos Quadrados

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023

Sumário I

Teorema (Teorema de Taylor)

Suponha que $f(x)$ tem r derivadas no ponto a. Então,

$$f(a+\Delta) = f(a) + \Delta f'(a) + \ldots + \frac{\Delta^r}{r!} f^{(r)}(a) + o(\Delta^r) \tag{1}$$

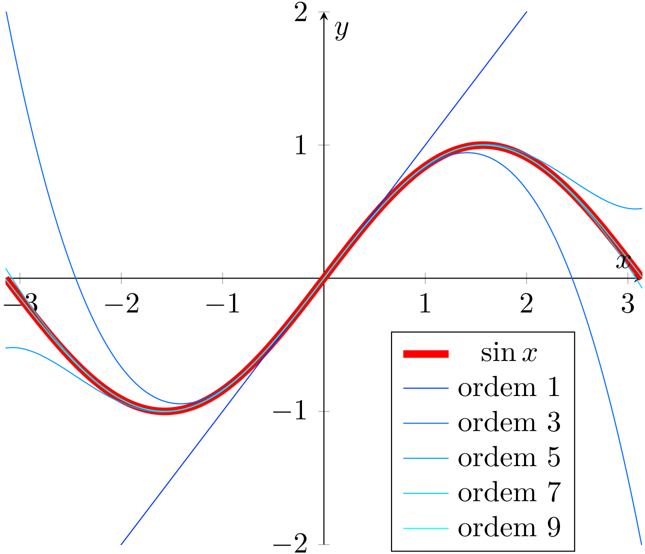
em que o último termo pode ser escrito como

$$\frac{\Delta^r}{r!}\left[f^{(r)}(a) + o(1)\right] \tag{2}$$

Além disso, se a $(r+1)$-ésima derivada de f existe em uma vizinhança de a, então o termo de resto $o(\Delta^r)$ pode ser escrito como

$$R_r = \frac{\Delta^{r+1}}{(r+1)!} f^{(r+1)}(\xi) \tag{3}$$

em que ξ é um ponto entre a e $a+\Delta$.



y
x
2
1
-1
-2
-3
-2
-1
1
2
3
$\sin x$
ordem 1
ordem 3
ordem 5
ordem 7
ordem 9

Teorema (Método Delta)

Se

$$\sqrt{n}\left(T_n - \theta\right) \xrightarrow{L} \mathcal{N}\left(0, \tau^2\right) \tag{4}$$

então,

$$\sqrt{n}\left(f\left(T_n\right) - f(\theta)\right) \xrightarrow{L} \mathcal{N}\left(0, \tau^2\left[f'(\theta)\right]^2\right) \tag{5}$$

se $f'(\theta)$ *existe e é diferente de zero.*

Método Delta

- Este teorema pode parecer surpreendente, uma vez que se X for normalmente distribuído, a distribuição de $f(X)$, por exemplo, $^1/_X$, $\log X$ ou e^X será tipicamente não normal.
- Como $T_n \xrightarrow{P} 0$, temos quase certeza de que quando n é grande, T_n está muito próximo de θ; entretanto, em uma pequena vizinhança, uma função diferenciável é quase linear e uma função linear de uma variável normal é novamente normal.
- O processo de aproximar a diferença $f(T_n) - f(\theta)$ pela função linear $(T_n - \theta)f'(\theta)$ e o resultado em (5) é chamado de método delta.

Método Delta

- O Teorema 2 fornece a base para derivar transformações estabilizadoras de variância, isto é, transformações que levam a uma variância assintótica que é independente do parâmetro.
- Suponha que $X_1, \ldots, X_n$ são variáveis aleatórias iid Poisson com expectativa λ. A variância é λ e segue do TCL que

$$\sqrt{n}\left(\bar{X} - \lambda\right) \xrightarrow{L} \mathcal{N}(0, \lambda) \tag{6}$$

Método Delta

- Para problemas de inferência relativos a λ, muitas vezes é inconveniente que λ ocorra não apenas na expectativa, mas também na variância da distribuição limite.
- É, portanto, de interesse procurar uma função f para a qual $\sqrt{n}\left[f\left(\bar{X}-f(\lambda)\right)\right]$ tende por lei a $\mathcal{N}(0,c^2)$, em que c^2 não depende de λ.
- Do teorema acima $f'(\theta)=\dfrac{c}{\tau(\theta)}$. A transformação resultante f é estabilizadora da variância.

Exemplo

- Suponha que X é uma variável aleatória com média $\mathbb{E}[X] = \mu \neq 0$. Se nós queremos estimar uma função $g(\mu)$, uma aproximação de primeira ordem nos dá $g(X) = g(\mu) + g'(\mu)(X - \mu)$. Se usarmos $g(X)$ como um estimador de $g(\mu)$, então temos

$$\mathbb{E}[g(X)] \approx g(\mu) \tag{7}$$

$$\text{var}(g(X)) \approx [g'(\mu)]^2 \text{var}(X) \tag{8}$$

- Suponha que $g(\mu) = \frac{1}{\mu}$. Estimamos $\frac{1}{\mu}$ a partir de $\frac{1}{X}$, que tem

$$\mathbb{E}\left[\frac{1}{X}\right] \approx \frac{1}{\mu} \tag{9}$$

$$\text{var}\left[\frac{1}{X}\right] \approx \left[\frac{1}{\mu^4}\right] \text{var}(X) \tag{10}$$

Exemplo

- Pelo Teorema 2 , então

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \to \mathcal{N}\left(0, \frac{1}{\mu^4}\text{var}(X_1)\right) \tag{11}$$

em distribuição.

Método Delta

- E o que ocorre quando $f'(\theta) = 0$? Vamos ver o que fazer quando $f'(\theta) = 0$, mas $f''(\theta) \neq 0$. Pela expansão de Taylor, fazemos

$$f(T_n) = f(\theta) + (T_n - \theta)f'(\theta) + \frac{1}{2}(T_n - \theta)^2[f''(\theta) + R_n] \quad (12)$$

e dado que $f'(\theta) = 0$, então

$$k_n\left[f(T_n) - f(\theta)\right] = \frac{k_n}{2}(T_n - \theta)^2 f''(\theta) + o_p\left[k_n((T_n - \theta)^2)\right] \quad (13)$$

- Assim, segue que

$$\frac{n(T_n - \theta)^2}{\tau^2(\theta)} \xrightarrow{L} \chi_1^2 \quad \text{ou} \quad n(T_n - \theta)^2 \to \tau^2(\theta)\chi_1^2 \quad (14)$$

em que χ_1^2 denota a distribuição χ^2 com 1 grau de liberdade.

- Para obter uma distribuição limite não-degenerada é necessário assumir que $k_n = n$ em vez de $\sqrt{n}$. Com isso,

$$n\left[f(T_n) - f(\theta)\right] \xrightarrow{L} \frac{1}{2}\tau^2(\theta)f''(\theta)\chi_1^2 \tag{15}$$

- Segue desse resultado que quando $f'(\theta) = 0$ e $f''(\theta)$, a convergência de $f(T_n)$ para $f(\theta)$ é mais rápida do que quando $f'(\theta) \neq 0$.

- Suponha que o objetivo seja estimar a variância de uma Binomial tal que

$$\sqrt{n}\left(\frac{X}{n}-p\right) \to \mathcal{N}(0,pq) \tag{16}$$

- O estimador de máxima verossimilhança pq é $\delta_n = \dfrac{X}{n}\left(1-\dfrac{X}{n}\right)$. Com $\theta = p$ e $f(\theta) = pq$, segue que

$$\sqrt{n}\left(\delta_n - pq\right) \to \mathcal{N}\left(0, pq(1-2p)^2\right) \tag{17}$$

se $p \neq \dfrac{1}{2}$. Dado que $f'(p) = 1-2p$, segue que em $p = \dfrac{1}{2}$, f' é zero e $f'' = -2$. Assim, fazemos $\dfrac{1}{2}\left(\dfrac{1}{2}\right)^2(-2)$.Logo,

$$n\left(\delta_n - pq\right) \xrightarrow{L} -\frac{1}{4}\chi_1^2 \tag{18}$$

- Seja $\theta = g(\mu)$, em que $\mu = \mathbb{E}[h(y)]$ e $g(\mu)$ é uma função suave. Então, vale o seguinte teorema.

Teorema

Se $y_i \in \mathbb{R}^m$ são iid, $h(u)\colon \mathbb{R}^m \to \mathbb{R}^k$, $\mathbb{E}\|h(y)\| < \infty$, e $g(u)\colon \mathbb{R}^k \to \mathbb{R}^q$ é contínua em μ, então $\widehat{\theta} \xrightarrow{p} \theta$ quando $n \to \infty$.

Teorema

Seja $\mu \in \mathbb{R}^k$ *e* $g(u) : \mathbb{R}^k \to \mathbb{R}^q$*. Se* $\sqrt{n}(\widehat{\mu} - \mu) \to \xi$*, onde* $g(u)$ *é continuamente diferenciável na vizinhança de* μ*, então quando* $n \to \infty$

$$\sqrt{n}\left(g(\widehat{\mu}) - g(\mu)\right) \xrightarrow{d} \boldsymbol{G}'\xi \tag{19}$$

em que $\boldsymbol{G}(u) = \frac{\partial}{\partial u} g(u)'$ *e* $\boldsymbol{G} = \boldsymbol{G}(\mu)$*. Em particular, se* $\xi \sim N(0, \boldsymbol{V})$ *então quando* $n \to \infty$

$$\sqrt{n}\left(g(\widehat{\mu}) - g(\mu)\right) \xrightarrow{d} N(0, \boldsymbol{V}) \tag{20}$$

- Funções diferenciáveis dos estimadores aleatórios assintoticamente normais são assintoticamente normais. Esse teorema estabelece normalidade assintótica.

- É interessante observar as condições para esses resultados:
    1. *Consistência requer que* $\boldsymbol{h(y)}$ *tenha uma média finita, enquanto que normalidade assintótica requer que essa variável tenha variância finita*, em que $\mu = \mathbb{E}[h(y)]$.
    2. *Consistência requer que* $\boldsymbol{g(u)}$ *seja contínua, enquanto que normalidade assintótica requer que* $\boldsymbol{g(u)}$ *seja continuamente diferenciável.* Sendo a última a condição mais forte.

Modelo de Regressão Normal

- O modelo de regressão normal é um modelo de regressão linear sob a restrição que o termo de erro e_i é independente de $\boldsymbol{x}_i$ e tem uma distribuição $N\left(0, \sigma^2\right)$. Formalmente temos

$$e_i | \boldsymbol{x}_i \sim N\left(0, \sigma^2\right) \tag{21}$$

- Este pressuposto implica que

$$y_i | \boldsymbol{x}_i \sim N\left(\boldsymbol{x}_i' \boldsymbol{\beta}, \sigma^2\right) \tag{22}$$

- Podemos derivar a distribuição amostral para o estimador de Mínimos Quadrados, resíduos e para o estimador da variância.

Modelo de Regressão Normal

- Sob o pressuposto de normalidade $e_i|\boldsymbol{x}_i \sim N\left(0, \sigma^2\right)$ combinado com a independência das observações tem a implicação multivariada

$$\boldsymbol{e}|\boldsymbol{X} \sim N\left(\mathbf{0}, \boldsymbol{I}_n\sigma^2\right) \tag{23}$$

- O vetor de erro $\boldsymbol{e}$ é independente de $\boldsymbol{X}$ e é normalmente distribuído. Lembrando que estimador de Mínimos Quadrados satisfaz

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{e} \tag{24}$$

que é uma função linear de $\boldsymbol{e}$.

- Como a função linear de uma normal também é uma normal, isto implica que condicional em $\boldsymbol{X}$

$$\begin{aligned}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X} &\sim (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' N(0, \boldsymbol{I}_n\sigma^2) \\ &\sim N\left(0, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{X} \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \\ &\sim N\left(0, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \qquad (25)\end{aligned}$$

- **Este resultado mostra que sob o pressuposto de erros normais o estimador de Mínimos Quadrados tem uma distribuição normal.**

Teorema

No modelo de regressão normal

$$\widehat{\beta}|\boldsymbol{X} \sim N\left(\boldsymbol{\beta}, \sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \qquad (26)$$

Distribuição do Vetor de Resíduos dos MQO

- Temos que $\widehat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e}$, em que $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Isto mostra que $\widehat{\boldsymbol{e}}$ é linear em $\boldsymbol{e}$. Então, condicional em $\boldsymbol{X}$

$$\widehat{\boldsymbol{e}} = \boldsymbol{M}\boldsymbol{e}|\boldsymbol{X} \sim N\left(0, \sigma^2 \boldsymbol{M}\boldsymbol{M}\right) = N\left(0, \sigma^2 \boldsymbol{M}\right) \tag{27}$$

- É útil encontrar a distribuição conjunta de $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{e}}$. Isto é fácil de ser feito escrevendo os dois como uma função linear do erro $\boldsymbol{e}$. Assim,

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{e}} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{e} \\ \boldsymbol{M}\boldsymbol{e} \end{pmatrix} = \begin{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \\ \boldsymbol{M} \end{pmatrix} \boldsymbol{e} \tag{28}$$

que é uma função linear de $\boldsymbol{e}$.

- O vetor tem uma distribuição normal conjunta com matriz de covariância

$$\boldsymbol{e} \sim N\left(0, \begin{pmatrix} \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} & 0 \\ 0 & \sigma^2\boldsymbol{M} \end{pmatrix}\right) \tag{29}$$

- O bloco fora da diagonal principal é zero dado que $\boldsymbol{X}'\boldsymbol{M} = 0$. Por esse resultado temos que $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{e}}$ são independentes.

Teorema

No modelo de regressão normal,

$$\widehat{\boldsymbol{e}}|\boldsymbol{X} \sim N(0, \sigma^2 \boldsymbol{M}) \tag{30}$$

e é independente de $\widehat{\boldsymbol{\beta}}$.

- Como $\widehat{\boldsymbol{\beta}}$ e $\boldsymbol{e}$ são independentes, isso implica que $\widehat{\boldsymbol{\beta}}$ é independente de qualquer função do vetor de resíduos, incluindo resíduos individuais $\widehat{e}_i$ e os estimadores s^2 e $\widehat{\sigma}^2$.

Distribuição do Estimador da Variância

- Considere o estimador da variância s^2. Esse estimador satisfaz $(n-k)s^2 = \hat{\boldsymbol{e}}'\hat{\boldsymbol{e}} = \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e}$.
- A decomposição espectral de $\boldsymbol{M}$ é dada por $\boldsymbol{M} = \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}'$ onde $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_n$ e $\boldsymbol{\Lambda}$ é uma matriz diagonal com os autovalores de $\boldsymbol{M}$ na diagonal.
- Como $\boldsymbol{M}$ é idempotente com *rank* $(n-k)$ ela tem $(n-k)$ autovalores igualando 1 e k igualando a 0. Assim,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{I}_{n-k} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}_k \end{bmatrix} \tag{31}$$

- Seja $\boldsymbol{u} = \boldsymbol{H}'\boldsymbol{e} \sim N\left(\boldsymbol{0}, \boldsymbol{I}_n\sigma^2\right)$ e a partição $\boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2')'$ onde $\boldsymbol{u}_1 \sim N\left(0, \boldsymbol{I}_{n-k}\sigma^2\right)$.

- Assim,

$$
\begin{aligned}
(n-k)s^2 &= \boldsymbol{e}'\boldsymbol{M}\boldsymbol{e} \\
&= \boldsymbol{e}'\boldsymbol{H}\begin{bmatrix} \boldsymbol{I}_{n-k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{H}'\boldsymbol{e} \\
&= \boldsymbol{u}'\begin{bmatrix} \boldsymbol{I}_{n-k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{u} \\
&= \boldsymbol{u}_1'\boldsymbol{u}_1 \\
&\sim \sigma^2\chi^2_{n-k} \\
\frac{(n-k)s^2}{\sigma^2} &\sim \chi^2_{n-k} \qquad (32)
\end{aligned}
$$

- Com isso temos que no modelo de regressão normal a distribuição exata de s^2 é uma qui-quadrado.
- *Como* $\widehat{\boldsymbol{e}}$ *é independente de* $\widehat{\boldsymbol{\beta}}$ *temos que* s^2 *é independente de* $\widehat{\boldsymbol{\beta}}$ *também.*

Teorema

No modelo de regressão normal,

$$\frac{(n-k)s^2}{\sigma^2} \sim \chi^2_{n-k} \tag{33}$$

e é independente de $\widehat{\boldsymbol{\beta}}$.

Desigualdade para Vetores

- **Decomposição espectral**: se $\boldsymbol{A}$ é $k \times k$ e simétrica então $\boldsymbol{A} = \boldsymbol{H\Lambda H'}$ em que $\boldsymbol{H'H} = \boldsymbol{I}_k$ e $\boldsymbol{A}$ é uma matriz diagonal com os autovalores na diagonal.
- **Desigualdade triangular**: se $\boldsymbol{a} = (a_1, ..., a_m)'$

$$\|\boldsymbol{a}\| \leq \sum_{j=1}^{n} |a_j| \tag{34}$$

- **Desigualdade de Schwarz**: para algum vetor $\boldsymbol{a}$ e $\boldsymbol{b}$ de dimensão $m \times 1$,

$$\|\boldsymbol{a'b}\| \leq \|\boldsymbol{a}\|\|\boldsymbol{b}\| \tag{35}$$

- **Desigualdade triangular**: para algum $m \times 1$ vetor $\boldsymbol{a}$ e $\boldsymbol{b}$,

$$\|\boldsymbol{a}+\boldsymbol{b}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{b}\| \tag{36}$$

Desigualdade para Matrizes

- **Desigualdade de Schwarz**: para alguma matriz $\boldsymbol{A}$ e $\boldsymbol{B}$ com dimensão $m \times k$,

$$\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\| \tag{37}$$

- **Desigualdade triangular**: para alguma matriz $\boldsymbol{A}$ e $\boldsymbol{B}$ com dimensão $m \times k$,

$$\|\boldsymbol{A}+\boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\| \tag{38}$$

Desigualdades em Probabilidade

- **Desigualdade da esperança**: para algum vetor $y \in \mathbb{R}^m$ com $\mathbb{E}\|y\| < \infty$, então

$$\|\mathbb{E}[y]\| \leq \mathbb{E}\|[y]\| \tag{39}$$

- **Desigualdade da esperança condicional**: para algum vetor $r \geq 1$ e para alguma variável aleatória $(y, x) \in \mathbb{R} \times \mathbb{R}^m$ tal que $\mathbb{E}[y]^r < \infty$, então

$$\mathbb{E}|\mathbb{E}(y|x)|^r \leq \mathbb{E}|y|^r < \infty \tag{40}$$

- **Desigualdade de Cauchy-Schwarz**: para algum matriz aleatória X e Y de dimensão $m \times n$,

$$\mathbb{E}\|X'Y\| \leq \left(\mathbb{E}\|X\|^2\right)^{1/2} \left(\mathbb{E}\|Y\|^2\right)^{1/2} \tag{41}$$

- **Desigualdade de Chebyshev**: para algum $x \in \mathbb{R}$ e $\epsilon > 0$

$$\mathbb{P}[|x - \mathbb{E}(x)| > \epsilon] \leq \epsilon^{-2}\text{var}(x) \tag{42}$$

- **Desigualdade de Hölder**: se $p > 1$ e $q > 1$ e $\dfrac{1}{p} + \dfrac{1}{q} = 1$, então para alguma matriz aleatória X e Y com dimensão $m \times n$,

$$\mathbb{E}\|X'Y\| \leq (\mathbb{E}\|X\|^p)^{1/p}(\mathbb{E}\|Y\|^q)^{1/q} \tag{43}$$

Intuição da Convergência

- Derivar a média e a variância do estimador de Mínimos Quadrados no contexto do modelo de regressão linear não nos fornece uma descrição completa da distribuição amostral, nem suficiente para fazermos inferência (testar hipóteses ou intervalos de confiança) sobre os parâmetros desconhecidos.
- Vamos ilustrar essa situação com um exemplo. Considere y_i e x_i seja retirados de uma densidade conjunta.

$$f(x, y) = \frac{1}{2\pi xy} \exp\left(-\frac{1}{2}(\log y - \log x^2)\right) \exp\left(-\frac{1}{2}(\log x)^2\right) \tag{44}$$

- Seja $\widehat{\beta}$ a estimativa do coeficiente de inclinação da regressão de mínimos quadrados de y_i em x_i.

- Usando métodos de simulação, a função densidade de $\widehat{\beta}$ foi calculada e plotada numa Figura para diferentes tamanhos de amostra: $n = 25$, $n = 100$ e $n = 800$.
- Quando o tamanho da amostra aumenta, a densidade se torna mais concentrada ao redor do coeficiente populacional.
- Há uma maneira simples de caracterizar a distribuição amostral de $\widehat{\beta}$?

Figura 1: Densidade Amostral de $\widehat{\beta}$ com $N = 25$

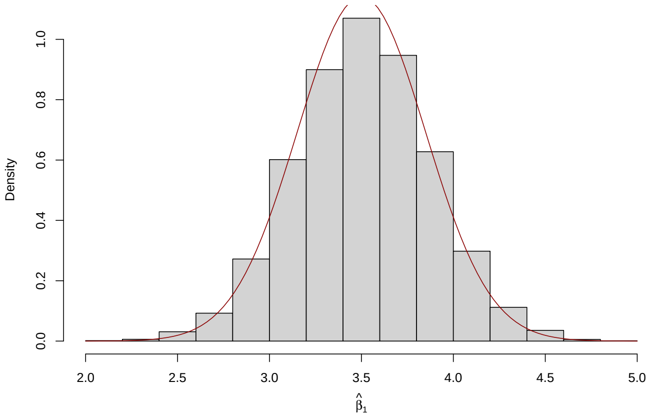


Figura 2: Densidade Amostral de $\widehat{\beta}$ com $N = 100$

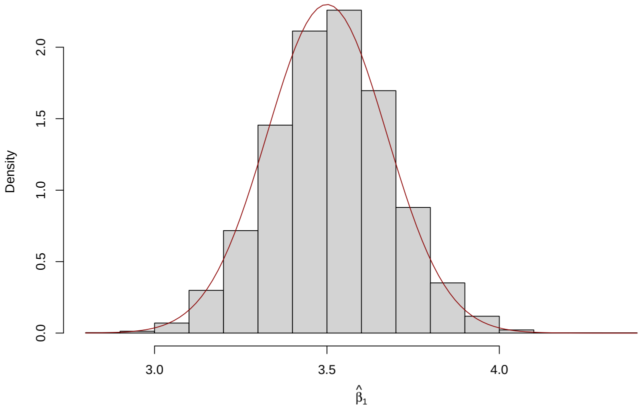
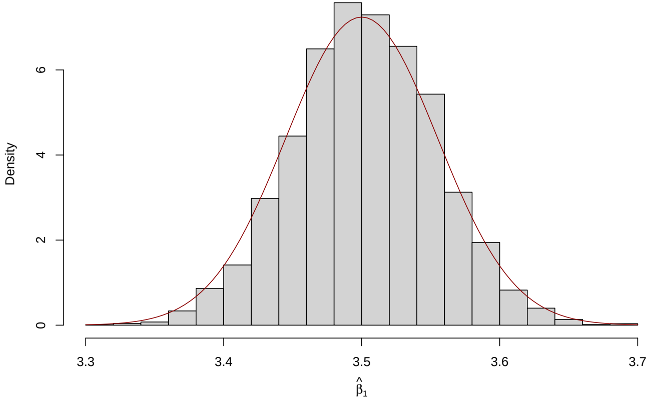


Figura 3: Densidade Amostral de $\widehat{\beta}$ com $N = 1000$



Convergência em Distribuição

- A consistência é importante e pode ser considerada um primeiro passo, mas ela não descreve a distribuição do estimador. Vamos gerar um resultado mais forte, a convergência em distribuição ou a distribuição assintótica.
- Vamos escrever o estimador como uma função dos momentos amostrais. Um dos momentos deve ser escrito como a soma do vetor aleatório com média zero e normalizado de modo que possamos aplicar o **Teorema Central do Limite** (**TCL**).
- Considere a eq. $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1} \widehat{\boldsymbol{Q}}_{\boldsymbol{xe}}$ e vamos pré-multiplicar por $\sqrt{n}$. Com isso temos:

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{x}_i e_i \right) \tag{45}$$

35/78

- A eq. (45) mostra que estimador centrado e normalizado $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ é uma função da média amostral $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'$ e da média amostral normalizada $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{x}_i e_i$.
- O par aleatório $(y_i, \boldsymbol{x}_i)$ é *i.i.d.* e qualquer função de $(y_i, \boldsymbol{x}_i)$ é também *i.i.d.*. Isto também inclui $e_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ e o produto $\boldsymbol{x}_i e_i$. Esse último tem média zero ($\mathbb{E}[\boldsymbol{x}_i e_i] = 0$) e possui matriz de covariância $k \times k$

$$\boldsymbol{\Omega} = \mathbb{E}\left[(\boldsymbol{x}_i e_i)(\boldsymbol{x}_i e_i)\right] = \mathbb{E}\left[\boldsymbol{x}_i\boldsymbol{x}_i' e_i^2\right] \tag{46}$$

que requer que $\boldsymbol{\Omega} < \infty$.

- Pela desigualdade da esperança, temos:

$$\|\boldsymbol{\Omega}\| \leq \mathbb{E}\|\boldsymbol{x}_i \boldsymbol{x}_i' e_i^2\| = \mathbb{E}\|\boldsymbol{x}_i e_i\|^2 \tag{47}$$

ou de forma equivalente $\mathbb{E}\|\boldsymbol{x}_i e_i\|^2 \leq \infty$.

- Usando $\mathbb{E}\|\boldsymbol{x}_i e_i\|^2 = \|\boldsymbol{x}_i\|^2 e_i^2$ e a desigualdade de Cauchy-Schwartz

$$\|\boldsymbol{\Omega}\| \leq \mathbb{E}\|\boldsymbol{x}_i \boldsymbol{x}_i' e_i^2\| = \mathbb{E}\|\boldsymbol{x}_i e_i\|^2 = \mathbb{E}\left(\|\boldsymbol{x}_i\|^2 e_i^2\right) \leqslant \left(\mathbb{E}\|\boldsymbol{x}_i\|^4\right)^{1/2} (\mathbb{E} e_i^4)^{1/2} \tag{48}$$

que é finita se $\boldsymbol{x}_i$ e e_i tem quarto momento finito.

- Como e_i é uma combinação linear de y_i e $\boldsymbol{x}_i$, é suficiente que as observáveis tenha o quarto momento finito. Agora podemos aplicar o **TCL**.

Teorema

Supondo que

1. *as observações* $(y_i, \boldsymbol{x}_i)$, $i = 1, ..., n$ *são independentes e identicamente distribuídos.*
2. $\mathbb{E}y^2 < \infty$.
3. $\mathbb{E}\|\boldsymbol{x}\|^2 < \infty$.
4. $\boldsymbol{Q}_{\boldsymbol{x}\boldsymbol{x}'}$ *é positiva definida.*

temos que

$$\|\boldsymbol{\Omega}\| \leqslant \mathbb{E}\|\boldsymbol{x}_i e_i\|^2 < \infty \tag{49}$$

e

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{x}_i e_i \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}) \tag{50}$$

- Juntando

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i \boldsymbol{x}_i') = \boldsymbol{Q}_{\boldsymbol{xx}} \tag{51}$$

e as equações (45) e (50), temos que:

$$\begin{aligned} \sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) &\xrightarrow{d} \boldsymbol{Q}_{\boldsymbol{xx}}^{-1} N\left(\mathbf{0}, \boldsymbol{\Omega}\right) \\ &= N\left(\mathbf{0}, \boldsymbol{Q}_{\boldsymbol{xx}}^{-1} \boldsymbol{\Omega} \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\right) \end{aligned} \tag{52}$$

quando $n \to \infty$.

- **Nota**: a última igualdade segue da propriedade que uma combinação linear de vetores com distribuição normal também possui distribuição normal.
- Com isso derivamos a aproximação normal assintótica para a distribuição do estimador de mínimos quadrados.

Teorema (Normalidade Assintótica do Estimador de Mínimos Quadrados)

Quando $n \to \infty$,

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\beta}}\right) \tag{53}$$

em que

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{Q}_{\boldsymbol{xx}}^{-1} \boldsymbol{\Omega} \boldsymbol{Q}_{\boldsymbol{xx}}^{-1} \tag{54}$$

com $\boldsymbol{Q}_{\boldsymbol{xx}} = \mathbb{E}\left(\boldsymbol{x}_i \boldsymbol{x}_i'\right)$ *e* $\boldsymbol{\Omega} = \mathbb{E}\left(\boldsymbol{x}_i \boldsymbol{x}_i' e_i^2\right)$.

Variância Assintótica

- A matriz $\boldsymbol{V}_{\boldsymbol{\beta}} = \text{avar}\left(\widehat{\boldsymbol{\beta}}\right)$ é a variância da distribuição assintótica de $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)$.
- A matriz $\boldsymbol{V}_{\boldsymbol{\beta}}$ é chamada de **matriz de covariância assintótica** de $\widehat{\boldsymbol{\beta}}$.
- A expressão $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{\boldsymbol{xx}}^{-1}$ é chamada de forma sanduíche.

- **Importante**: existe uma diferença entre a variância da distribuição assintótica e a variância condicional na amostra finita no modelo CEF.

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \tag{55}$$

- $\boldsymbol{V}_{\boldsymbol{\beta}}$ e $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ são diferentes, mas se aproximam se o n é grande. De fato, quando $n \to \infty$

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}} \tag{56}$$

- Há um caso especial em que $\boldsymbol{\Omega}$ e $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ simplificam. Suponha que

$$\operatorname{cov}\left(\boldsymbol{x}_i\boldsymbol{x}_i', e_i^2\right) = \boldsymbol{0} \tag{57}$$

- A condição dada pela eq. (57) é assegurada num modelo de regressão linear homocedástico, mas é um tanto geral.
- Sob a condição dada pela eq. (57), a fórmula da variância assintótica simplifica quando

$$\begin{aligned} \boldsymbol{\Omega} &= \mathbb{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\mathbb{E}(e_i^2) = \boldsymbol{Q}_{\boldsymbol{xx}}\sigma^2 \\ \boldsymbol{V}_{\boldsymbol{\beta}} &= \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{\boldsymbol{xx}}^{-1} = \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\sigma^2 \equiv \boldsymbol{V}_{\boldsymbol{\beta}}^0 \end{aligned} \tag{58}$$

- **Observação:** na eq. (58) temos que $\boldsymbol{V}_{\boldsymbol{\beta}}^0 = \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\sigma^2$ se $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{V}_{\boldsymbol{\beta}}^0$.
- Chamamos $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}^0$ de **matriz de covariância homocedástica assintótica**.
- *O Teorema da Normalidade Assintótica estabelece que a distribuição amostral do estimador de mínimos quadrados, após reescalar, é aproximadamente normal quando o tamanho da amostra n é suficientemente grande.*

Consistência do Estimador da Variância do E

- Vamos mostrar que $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\widehat{e}_i^2$ e $s^2 = \frac{1}{n-k}\sum_{i=1}^{n}\widehat{e}_i^2$ são estimadores consistentes para σ^2.
- Vamos escrever o resíduos $\widehat{e}_i$ como igual ao erro e_i mais um termo de desvio

$$\begin{aligned}\widehat{e}_i &= y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}} \\ &= e_i + \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}} \\ &= e_i - \boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \qquad (59)\end{aligned}$$

- Assim, elevando ao quadrado ambos os lado temos

$$\widehat{e}_i^2 = e_i^2 - 2e_i\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \qquad (60)$$

- Se tirarmos a média dos resíduos ao quadrado vamos obter a média dos erros ao quadrado mais dois termos que são assintoticamente negligenciáveis:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n} e_i \boldsymbol{x}_i'\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \tag{61}$$

- Usando a LFGN podemos mostrar que

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 \xrightarrow{p} \sigma^2 \tag{62}$$

$$\frac{1}{n}\sum_{i=1}^{n} e_i \boldsymbol{x}_i' \xrightarrow{p} 0 \tag{63}$$

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i \boldsymbol{x}_i') = \boldsymbol{Q}_{\boldsymbol{x}_i \boldsymbol{x}_i} \tag{64}$$

- Pelo Teorema da Consistência dos mínimos quadrados temos que $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$. Com isso, a eq. (61) converge em probabilidade para σ^2 como desejamos.
- Por fim, uma vez que $\dfrac{n}{n-k} \to 1$ quando $n \to \infty$ temos que

$$s^2 = \left(\frac{n}{n-k}\right)\widehat{\sigma}^2 \xrightarrow{p} \sigma^2 \tag{65}$$

- Então, ambos os estimadores são consistentes.

Teorema

$$\widehat{\sigma}^2 \xrightarrow{p} \sigma^2 \quad e \quad s^2 \xrightarrow{p} \sigma^2 \quad quando\ n \to \infty \tag{66}$$

Estimação da Matriz de Covariância Homoce

- Para fazermos inferência assintótica (testes e intervalo de confiança) precisamos de estimativas consistente da sua matriz de covariância.
- O estimador convencional da matriz de covariância é dado por $\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}} = \widehat{\boldsymbol{Q}}^{-1}_{\boldsymbol{xx}} s^2$.
- Como o estimador da matriz de covariância é resultado do produto de duas estimativas dos momentos, o método aqui será mostrar consistência para cada estimador dos momentos, e aplicar o Teorema do Mapeamento Contínuo para o produto.

- Pelo teorema da consistência dos mínimos quadrados, temos: $\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1} \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{xx}}^{-1}$.
- Pelo teorema do estimador da variância do erro consistente, temos que: $s^2 \xrightarrow{p} \sigma^2$.

  Assim, usando o Teorema do Mapeamento Contínuo temos que

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{0} = \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1} s^2 \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{xx}}^{-1} \sigma^2 = \boldsymbol{V}_{\boldsymbol{\beta}}^{0} \tag{67}$$

em que $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{0}$ é consistente para $\boldsymbol{V}_{\boldsymbol{\beta}}^{0}$ como esperado.

Teorema

$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{0} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}^{0}$ *quando* $n \to \infty$

- **Duas observações**
  1. O teorema acima não requer o pressuposto de homocedasticidade!
  2. A $\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\beta}}}$ é consistente para $\boldsymbol{V}^0_{\boldsymbol{\beta}}$ independentemente se a regressão é homocedástica ou heterocedástica.
- Porém, $\boldsymbol{V}^0_{\boldsymbol{\beta}} = \boldsymbol{V}_{\boldsymbol{\beta}} = \text{avar}(\widehat{\boldsymbol{\beta}})$ apenas sob homocedasticidade.

Estimação da Matriz de Covariância Heteroce

- Pelo teorema da normalidade assintótica do estimador de MQO, a variância assintótica de $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)$ é $\boldsymbol{V}_{\boldsymbol{\beta}}=\boldsymbol{Q}_{\boldsymbol{xx}}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{\boldsymbol{xx}}^{-1}$.
- Vamos considerar a estimação dessa matriz de covariância sem impor homocedasticidade. Assim, o estimador dos momentos para $\boldsymbol{\Omega}$ é:

$$\widehat{\boldsymbol{\Omega}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2 \tag{68}$$

levando a um estimador da matriz de covariância dado por

$$\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{HC0}=\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1} \tag{69}$$

- **Observação**
  1. *Este resultado é idêntico ao obtido pelo estimador da matriz de covariância de White* $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC0}$.
- Para verificar a consistência de $\widehat{\Omega}$, vamos substituir $\widehat{e}_i^2$ pelo termo de erro e_i^2 e **mostrar que a diferença entre eles é assintoticamente negligenciável**.

- Façamos:

$$\begin{aligned}\widehat{\boldsymbol{\Omega}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'e_i^2 + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{e}_i^2 - e_i^2) \qquad (70)\end{aligned}$$

- O primeiro termo é uma média das variáveis aleatórias *i.i.d.* $\boldsymbol{x}_i\boldsymbol{x}_i'e_i^2$ e pela LFGN converge em probabilidade para a sua esperança:

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'e_i^2 \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i'e_i^2) = \boldsymbol{\Omega} \qquad (71)$$

- Esse resultado requer que $\boldsymbol{\Omega} < \infty$.

- Para mostrar que $\widehat{\Omega}$ é consistente, resta mostra que

$$\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{e}_i^2 - e_i^2) \xrightarrow{p} 0 \tag{72}$$

- Uma derivação razoável e direta, mas um pouco tediosa, é começar aplicando a desigualdade triangular usando uma norma de matriz:

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{e}_i^2 - e_i^2)\right\| &\leq \frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{e}_i^2 - e_i^2)\right\| \\ &= \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|^2\|\widehat{e}_i^2 - e_i^2\| \end{aligned} \tag{73}$$

- Assim, relembrando da expressão para os resíduos ao quadrado dada por

$$\widehat{e}_i^2 = e_i^2 - 2e_i\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

aplicando a desigualdade triangular e a desigualdade de Cauchy-Schwarz duas vezes, temos:

$$\begin{aligned} \left|\widehat{e}_i^2 - e_i^2\right| &\leq 2\left|e_i\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right| + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= 2|e_i|\left|\boldsymbol{x}_i'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right| + \left|(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{x}_i\right|^2 \\ &\leq 2|e_i|\|\boldsymbol{x}_i\|\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| + \|\boldsymbol{x}_i\|^2\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 \end{aligned} \tag{74}$$

- Combinando a eq. (73) com (74), obtemos

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'(\widehat{e}_i^2-e_i^2)\right\| \le 2\left(\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|^3|e_i|\right)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\| \tag{75}$$

$$+\left(\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i\|^4\right)\left\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right\|^2$$

$$= o_p(1). \tag{76}$$

- A expressão é $o_p(1)$ é uma relação de ordem e indica que o termo $\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\| \xrightarrow{p} 0$ converge em probabilidade para 0 quando $n \to \infty$.
- Ambas as médias nos parenteses são médias de variáveis aleatórias com esperança finita (são $O_p(1)$).

- Usando a desigualdade de Hölder, temos

$$\mathbb{E}\left(\|\boldsymbol{x}_i\|^3|e_i|\right) \leq \left(\mathbb{E}\left(\|\boldsymbol{x}_i\|^3\right)^{4/3}\right)^{3/4}\left(\mathbb{E}e_i^4\right)^{1/4} = \left(\mathbb{E}\|\boldsymbol{x}_i\|^4\right)^{3/4}\left(\mathbb{E}e_i^4\right)^{1/4} < \infty \tag{77}$$

- Com isso, demostramos a eq. (72) como desejado.

Teorema

Sob o pressuposto do quarto momento finito

$$\text{quando } n \to \infty,\ \widehat{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega} \text{ e } \widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC0} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}} \tag{78}$$

- Os estimadores alternativos da matriz de covariância robusta a heterocedasticidade, $(\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{HC2} = n\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC2}$ e $\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}^{HC3} = n\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC3})$, tem a forma da eq. (69) mas com a $\widehat{\Omega}$ substituída por

$$\overline{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{1}^{n}(1-h_{ii})^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2 \tag{79}$$

e

$$\widetilde{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{1}^{n}(1-h_{ii})^{-2}\boldsymbol{x}_i\boldsymbol{x}_i'\widehat{e}_i^2 \tag{80}$$

Teorema

Sob o pressuposto do quarto momento finito, quando $n \to \infty$, $\overline{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$, $\widetilde{\boldsymbol{\Omega}} \xrightarrow{p} \boldsymbol{\Omega}$, $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC1} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}$, $\overline{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC2} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}$ e $\widetilde{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^{HC3} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}$.

- **Duas observações**
  1. os estimadores alternativos da matriz de covariância são consistentes para a matriz covariância assintótica.
  2. podemos usar a notação $\widehat{\boldsymbol{V}}_{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}$ para nos referir a qualquer estimador HC0, HC1, HC2 e HC3 da matriz de covariância consistente a heterocedasticidade já que todos eles tem o mesmo limite assintótico.

Definição

Se $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}$ é um estimador da matriz de covariância de $\widehat{\boldsymbol{\beta}}$, então o erro-padrão é a raiz quadrada dos elementos da diagonal desta matriz.

$$S\left(\widehat{\boldsymbol{\beta}}_j\right) = \sqrt{\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}_j}} = \sqrt{\left[\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}_{jj}}\right]} \quad (81)$$

Geometria do Ajuste dos Mínimos Quadrados

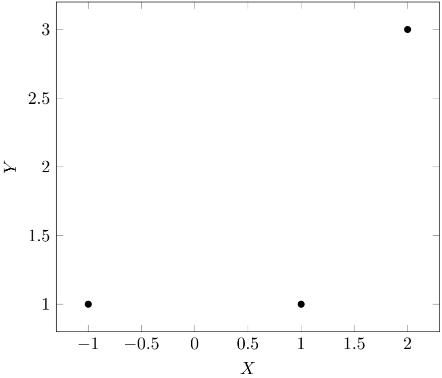
- Suponha que você deseja estimar uma regressão a partir das seguintes observações:

$$X = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}, \quad Y = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \tag{82}$$

- Podemos representar essas três observações em um diagrama de dispersão.
- Por MQO, podemos ajustar uma reta a essas três observações, representada pela equação:

$$Y_i = \beta_1 + \beta_2 X_i \tag{83}$$

em que $i = 1, ., 3$


3
2.5
2
1.5
1
Y
−1
−0.5
0
0.5
1
1.5
2
X

- Substituindo os valores observado, obtemos o sistema linear:

$$1 = \beta_1 + \beta_2(-1) = \beta_1 - \beta_2 \tag{84}$$
$$1 = \beta_1 + \beta_2(1) = \beta_1 + \beta_2 \tag{85}$$
$$3 = \beta_1 + \beta_2(2) = \beta_1 + 2\beta_2 \tag{86}$$

- Esse sistema nas *variáveis* β_1 e β_2 não tem solução. Se o sistema tivesse solução, haveria uma reta que passaria pelos três pontos do diagrama de dispersão, o que é claramente impossível. Essa é a situação típica em econometria.

- Na forma matricial o sistema fica:

$$\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \tag{87}$$

- Na notação que já estamos acostumados

$$\boldsymbol{y} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \ \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \tag{88}$$

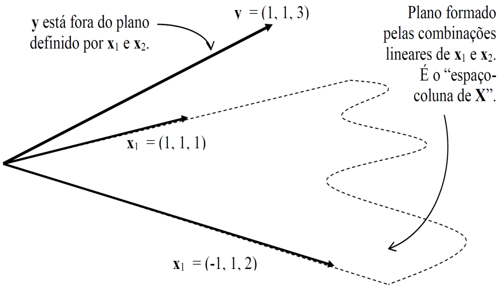
- Em notação vetorial temos:

$$\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \beta_1 + \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix} \beta_2 \tag{89}$$

- Essa segunda forma mostra que resolver o sistema significaria encontrar uma combinação linear dos vetores-coluna da matriz $\boldsymbol{X}$ que fosse igual ao vetor $\boldsymbol{y}$. Os pesos da combinação seriam β_1 e β_2.

- O espaço vetorial gerado pelos vetores linearmente independentes $\boldsymbol{x}_1 = (1, 1, 1)$ e $\boldsymbol{x}_2 = (-1, 1, 2)$ é um plano.
- Se o sistema tivesse solução, $\boldsymbol{y} = (1, 1, 3)$ seria uma dos infinitos vetores nesse plano. **Infelizmente, ele não é e o sistema é insolúvel!**
- Vamos ver graficamente a representação do plano formado pelas combinações lineares de $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$.

Figura 4: Plano formado pelas combinações lineares de $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$



- Como o sistema não tem solução, lançamos mão de um ajuste de mínimos quadrados. As estimativas $\widehat{\beta}_1$ e $\widehat{\beta}_2$ para os parâmetros β_1 e β_2 consistirão em pesos de uma combinação linear de $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$.
- Essa combinação é um vetor muito especial: ele é a ***projeção ortogonal*** de $\boldsymbol{y}$ sobre o espaço-coluna de $\boldsymbol{X}$.
- Justamente por ser ortogonal, a projeção minimiza o comprimento do vetor de erro $\boldsymbol{e}$, definido por:

$$\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{90}$$

em que $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ é a projeção ortogonal de $\boldsymbol{y}$ sobre o plano.

- Esse vetor $\boldsymbol{e}$, por construção, é ortogonal a todo o plano formado por combinações de $\boldsymbol{x}_1$ e $\boldsymbol{x}_2$. Logo:

$$
\begin{aligned}
\boldsymbol{x}_1'(\boldsymbol{y} - \boldsymbol{X}\widehat{\beta}) &= 0 \\
\boldsymbol{x}_2'(\boldsymbol{y} - \boldsymbol{X}\widehat{\beta}) &= 0 \\
\Rightarrow (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) &= \mathbf{0} \\
\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{y} \\
\therefore \quad \widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \qquad (91)
\end{aligned}
$$

- Vamos aplicar a fórmula do estimador de mínimos quadrados para os dados do nosso exemplo.

- As matrizes com os dados são:

$$\boldsymbol{y} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}, \ \boldsymbol{X} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \tag{92}$$

- Vamos calcular $\boldsymbol{X}'\boldsymbol{X}$ e $\boldsymbol{X}'\boldsymbol{y}$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \Longrightarrow \boldsymbol{X}'\boldsymbol{X} = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} \tag{93}$$

$$\boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \Longrightarrow \boldsymbol{X}'\boldsymbol{y} = \begin{pmatrix} 5 \\ 6 \end{pmatrix} \tag{94}$$

- Vamos calcular o determinante de $\boldsymbol{X}'\boldsymbol{X}$:

$$\boldsymbol{X}'\boldsymbol{X} = \begin{vmatrix} 3 & 2 \\ 2 & 6 \end{vmatrix} = 18 - 4 = 14 \tag{95}$$

- A matriz inversa de $\boldsymbol{X}'\boldsymbol{X}$ é:

$$\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} = \begin{pmatrix} 6/14 & -2/14 \\ -2/14 & 3/14 \end{pmatrix} \tag{96}$$

- Calculando o estimador de MQO, temos:

$$\widehat{\boldsymbol{\beta}} = \begin{pmatrix} 6/14 & -2/14 \\ -2/14 & 3/14 \end{pmatrix} \begin{pmatrix} 5 \\ 6 \end{pmatrix} \Longrightarrow \widehat{\boldsymbol{\beta}} = \begin{pmatrix} 9/17 \\ 4/7 \end{pmatrix} \tag{97}$$

- Calculando $\widehat{\boldsymbol{y}}$:

$$\widehat{\boldsymbol{y}} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 9/17 \\ 4/7 \end{pmatrix} \Longrightarrow \widehat{\boldsymbol{y}} = \begin{pmatrix} 5/7 \\ 13/7 \\ 17/7 \end{pmatrix} \tag{98}$$

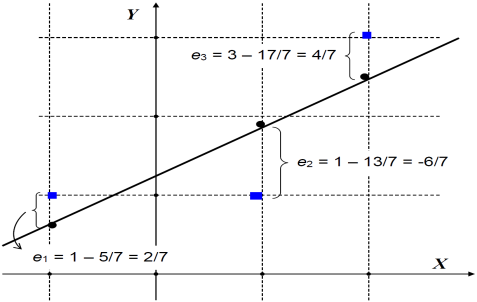
- Calculando os resíduos $\widehat{\boldsymbol{e}} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$:

$$\widehat{\boldsymbol{e}} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 5/7 \\ 13/7 \\ 17/7 \end{pmatrix} \Longrightarrow \widehat{\boldsymbol{e}} = \begin{pmatrix} 2/7 \\ -6/7 \\ 4/7 \end{pmatrix} \tag{99}$$

- A soma dos resíduos ao quadrado, que também é o quadrado do comprimento do vetor $\widehat{\boldsymbol{e}}$, é a menor possível dadas as nossas observações.

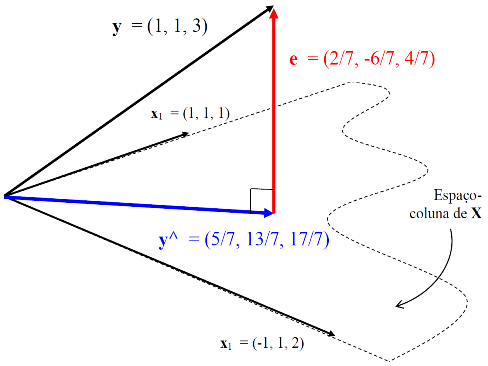
$$\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} = \sum_{i=1}^{3} \widehat{e}_i^2 = \frac{4}{49} + \frac{36}{49} + \frac{16}{49} = \frac{56}{49} \tag{100}$$

Figura 5: Diagrama de Dispersão



$$\mathbf{y} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix} \qquad \hat{\mathbf{y}} = \begin{bmatrix} 5/7 \\ 13/7 \\ 17/7 \end{bmatrix} \qquad \mathbf{e} = \begin{bmatrix} 2/7 \\ -6/7 \\ 4/7 \end{bmatrix}$$

Figura 6: Projeção Ortogonal de $\boldsymbol{y}$ sobre o Espaço-Coluna de $\boldsymbol{X}$



- Vamos relembrar duas matrizes importantes.
- Se substituir a fórmula do estimador de mínimos quadrados na definição do vetor de valores ajustados, $\widehat{\boldsymbol{y}}$, obteremos:

$$\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y} \tag{101}$$

em que

$$\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \tag{102}$$

- É a matriz de projeção de $\boldsymbol{y}$ no espaço coluna de $\boldsymbol{X}$.

- Analogamente, podemos expressar o vetor de resíduos como:

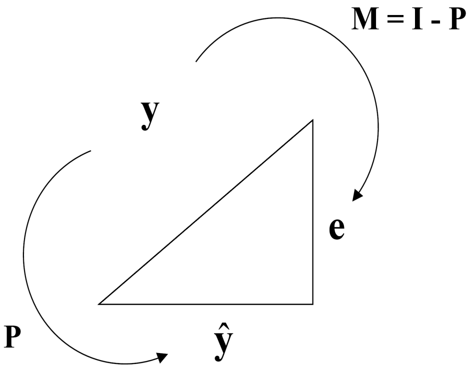
$$\begin{aligned}\widehat{\boldsymbol{e}} &= \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y} \\ &= \boldsymbol{M}\boldsymbol{y} \qquad (103)\end{aligned}$$

em que

$$\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \qquad (104)$$

- É uma matriz que gera o vetor de resíduos dos mínimos quadrados na regressão de $\boldsymbol{y}$ sobre $\boldsymbol{X}$ quando pré-multiplica qualquer vetor $\boldsymbol{y}$.
- Como $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$ temos que: $\boldsymbol{P}\boldsymbol{y} + \boldsymbol{M}\boldsymbol{y} = (\boldsymbol{P} + \boldsymbol{I} - \boldsymbol{P})\boldsymbol{y}$. Isto é, $\boldsymbol{y}$ pode ser decomposto no ajuste da regressão e no resíduo desta. Podemos ver isso graficamente.

Figura 7: Decomposição do $\boldsymbol{y}$ no Ajuste da Regressão e no Resíduo



- **Decomposição da soma dos quadrados**: podemos decompor a variação de $\boldsymbol{y}$ numa parte **explicada** pela regressão e numa parte **não explicada**. Assim, partindo de

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}} \tag{105}$$

- Podemos definir a soma dos quadrados

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{y} &= (\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}})'(\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}) \\ &= \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \\ &= (\boldsymbol{X}\widehat{\boldsymbol{\beta}})'\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \\ &= \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \end{aligned} \tag{106}$$

este resultado é exatamente o Teorema de Pitágoras (em n dimensões), para o triângulo retângulo formado pelos vetores $\boldsymbol{y}$, $\boldsymbol{y}$ ajustado e o resíduo $\widehat{\boldsymbol{e}}$.

- A variação de $\boldsymbol{y}$ ao redor da média é dada por:

$$\begin{aligned}\sum (y_i - \overline{y})^2 &= \sum y_i^2 - 2\overline{y} \sum y_i + n\overline{y}^2 \\ &= \sum y_i^2 - n\overline{y}^2 \\ &= \boldsymbol{y}'\boldsymbol{y} - n\overline{y}^2 \end{aligned} \tag{107}$$

- Temos que

$$\underbrace{\boldsymbol{y}'\boldsymbol{y} - n\overline{y}^2}_{\boldsymbol{S}QT} = \underbrace{(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} - n\overline{y}^2)}_{\boldsymbol{S}QE} + \underbrace{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}_{\boldsymbol{S}QR} \tag{108}$$

- A partir da decomposição dos quadrados são definidos o $\boldsymbol{R}^2$ e o $\boldsymbol{R}^2$ ajustado da regressão.

$$\boldsymbol{R}^2 = \frac{\boldsymbol{S}QE}{\boldsymbol{S}QT} \tag{109}$$

$$\overline{\boldsymbol{R}}^2 = 1 - (1 - \boldsymbol{R}^2)\left({}^{(n-1)}/_{(n-k)}\right) \tag{110}$$

- A aplicação do Teorema de Pitágoras e o $\boldsymbol{R}^2$ ao exemplo geram:

$$\boldsymbol{y}'\boldsymbol{y} = \widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} + \widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} \tag{111}$$

$$(1^2 + 1^2 + 3^2) = \frac{1}{7^2}(5^2 + 13^2 + 17^2) + \frac{1}{7^2}(2^2 + 6^2 + 4^2) \tag{112}$$

$$11 = \frac{483}{49} + \frac{56}{49} \tag{113}$$

e

$$\boldsymbol{R}^2 = \frac{\boldsymbol{S}QE}{\boldsymbol{S}QT} = \frac{483/49 - 5^2/3}{11 - 5^2/3} = \frac{1,52}{2,67} = 57\% \tag{114}$$

e

$$s^2 = \frac{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{n-k} = \frac{56}{49} \tag{115}$$

Econometria I

Propriedades de Convergência Fraca para Funções de Grandes Amostras: O Caso dos Mínimos Quadrados

Victor Oliveira

Núcleo de Economia Internacional e Desenvolvimento Econômico

PPGDE – 2023