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

Definição

- A distribuição amostral de $\widehat{\beta}$ é uma função da distribuição conjunta de (y_i, x_i) e do tamanho da amostra n. Na prática, esta função é extremamente complicada e assim não é fácil analiticamente calcular a distribuição exata de $\widehat{\beta}$. Por isso, tipicamente partimos para os métodos de aproximação.
- *O método mais conhecido é a teoria assintótica que aproxima uma distribuição amostral tomando o limite da distribuição amostral finita quando o tamanho da amostra n tende para infinito.*
- As ferramentas básicas da teoria assintótica são a Lei Fraca dos Grandes Números (LFGN), o Teorema Central do Limite (TCL) e o Teorema do Mapeamento Contínuo (TMP). Com essas ferramentas podemos aproximar a distribuição amostral da maior parte dos estimadores econométricos.

- **Análise assintótica** é um método de aproximação obtido tomando o limite adequado.
- Existe mais de um método pata tomar o limite, porém o **método mais comum em estatística e econometria é aproximar distribuições amostrais tomando o limite quando o tamanho da amostra tende para mais infinito**, isto é, $n \to +\infty$.
- O primeiro passo na análise assintótica é conhecer o conceito de limite de uma sequência.

Definição (Limite (informal))

Se f é alguma função, então

$$\lim_{x \to a} f(x) = L, \tag{1}$$

é lido "o limite de $f(x)$ quando x se aproxima de a é L". Isso significa que se você escolher valores de x próximos, mas não iguais a a, então $f(x)$ estará próximo do valor L; além disso, $f(x)$ se aproxima cada vez mais de L quando x se aproxima mais e mais de a.

A seguinte notação alternativa é usada às vezes:

$$f(x) \to L \text{ quando } x \to a. \tag{2}$$

Se $f(x) = x + 3$, então $\lim_{x \to 4} f(x) = 7$ é verdade, porque se você substituir números de x perto de 4 em $f(x) = x + 3$, o resultado será próximo de 7.

Definição (Limite (formal))

Dizemos que L é o limite de $f(x)$ quando $x \to a$ se:

1. $f(x)$ não precisa ser definido em $x = a$, mas deve ser definido para todos os outros valores de x em algum intervalo que contenha a.
2. para todo $\varepsilon > 0$ pode-se encontrar um $\delta > 0$ tal que para todo x no domínio de f tem-se que

$$|x - a| < \delta \text{ implica } |f(x) - L| < \varepsilon. \tag{3}$$

Graficamente, a noção de limite pode ser vista como segue:

Figura 1: Noção Gráfica de Limites

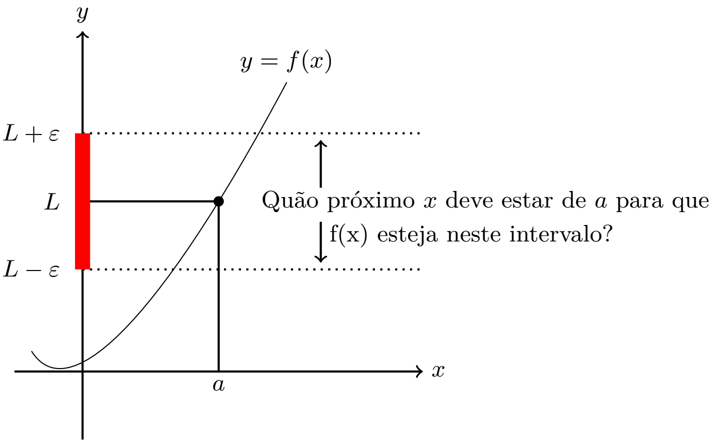
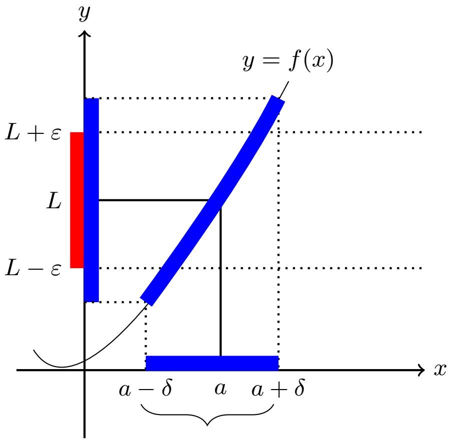
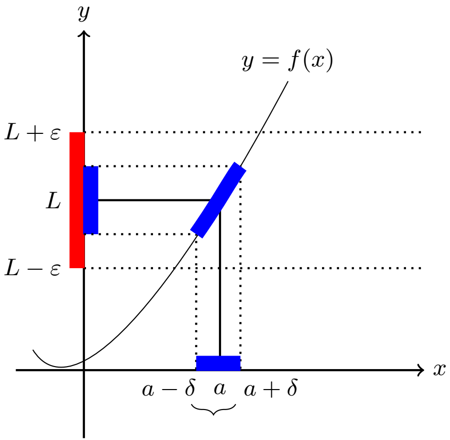


Figura 2: Noção Gráfica de Limites



Para algum x neste intervalo $f(x)$
não está entre $L - \varepsilon$ e $L + \varepsilon$.
Portanto, δ é muito grande para o
ε dado.

Figura 3: Noção Gráfica de Limites



Se escolhermos x neste intervalo $f(x)$ estará entre $L - \varepsilon$ e $L + \varepsilon$.

Portanto, δ é pequeno o suficiente para o ε dado.

- Por que os valores absolutos? A quantidade $|x-a|$ é a distância entre os pontos x e a na reta e pode-se medir quão próximo x é de a calculando-se $|x-a|$. A desigualdade $|x-a| < \delta$ diz que a distância entre x e a é menor que δ ou que x e a estão próximos que δ é muito pequeno.
- O que são ε e δ? A quantidade ε é quão próximo você gostaria que $f(x)$ estivesse de seu limite L; a quantidade δ é o quão perto você tem que escolher x para conseguir isso. Para provar que $\lim_{x \to a} f(x) = L$, você deve assumir que alguém lhe deu um valor desconhecido $\varepsilon > 0$ e, em seguida, encontrar um $\delta > 0$ pelo qual (3) é verdadeiro. O valor de δ encontrado dependerá de ε. Por isso, muitas vezes é escrito como $\delta(\varepsilon)$.

- Dizer que o número real a é o limite da sequência (x_n) significa afirmar que, para valores muito grandes de n, os termos x_n tornam-se e se mantêm tão próximos de a quanto se deseje.
- Especificamente, estipulando-se um erro por meio de um número real $\varepsilon > 0$, existe um índice n_0 tal que todos os termos x_n da sequência que têm índice n maior do que n_0 são valores aproximados de a com erro inferior a ε.
- O índice n_0 deve depender de ε, sendo de se esperar que, para valores cada vez menores de ε, necessita-se tomar n_0 cada vez maior. Com isso, temos a seguinte definição.

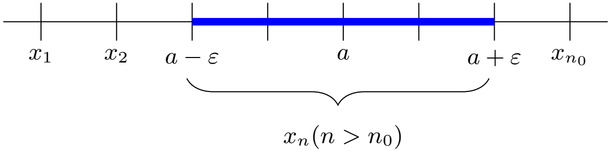
Definição (Limite de uma sequência)

Diz-se que o número real a é limite da sequência (x_n) de números reais, e escreve-se $a = \lim x_n$, ou $a = \lim\limits_{n} x_n$, ou $a = \lim\limits_{n\to\infty} x_n$, quando para cada número real $\varepsilon > 0$, dado arbitrariamente, for possível obter um inteiro $n_0 \in \mathbb{N}$ tal que $|x_n - a| < \varepsilon$, sempre que $n > n_0$. Isto é,

$$\lim x_n = a. \equiv .\forall \varepsilon > 0 \ \exists \ n_0 \in \mathbb{N}; n > n_0 \Rightarrow |x_n - a| < \varepsilon. \tag{4}$$

- Assim, a definição acima significa que, para todo número real $\varepsilon > 0$, existe um número natural n_0 tal que $n > n_0$ implica $|x_n - a| < \varepsilon$. Observamos que se $\lim x_n = a$ então qualquer intervalo $(a-\varepsilon, a+\varepsilon)$, de centro a e raio $\varepsilon > 0$, contém todos os termos x_n da sequência, com exceção no máximo de um número finito de índices n. Com efeito, dado o intervalo $(a-\varepsilon, a+\varepsilon)$, como $\lim x_n = a$, obtemos $n_0 \in \mathbb{N}$ tal que $n > n_0 \Rightarrow |x_n - a| < \varepsilon$. Ou seja, $n > n_0 \Rightarrow x_n \in (a-\varepsilon, a+\varepsilon)$.
- Graficamente, temos a seguinte situação.

Figura 4: Noção Geométrica de Limite



Convergência em Probabilidade

- Considere a média amostral $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ para uma dada amostra aleatória com n observações. Quando n aumenta, a distribuição de $\overline{y}$ muda.
- Em que sentido podemos descrever o "limite" de $\bar{y}$? Em que sentido ele converge?
- **Como $\bar{y}$ é uma variável aleatória, não podemos aplicar o conceito determinístico de uma sequência de números. Vamos requerer a definição de convergência que é apropriada para variáveis aleatórias.**
- A definição mais comum utilizada é chamada de convergência em probabilidade, também conhecida como convergência fraca.

Definição (Convergência em Probabilidade)

Uma variável aleatória $z_n \in \mathbb{R}$ converge em probabilidade para z quando $n \to \infty$, denotado por $z_n \xrightarrow{p} z$, ou alternativamente quando $p \lim_{n\to\infty} z_n = z$, se para todo $\delta > 0$,

$$\lim_{n\to\infty} \Pr(|z_n - z| \leq \delta) = 1 \tag{5}$$

podemos chamar z o **limite em probabilidade** ou (**plim**) de z.

- Essa definição formaliza o conceito de uma sequência de variáveis aleatórias que se concentram em um ponto.
- Para qualquer $\delta > 0$, o evento $\{|z_n - z| \leq \delta\}$ ocorre quando z_n está dentro de δ a partir do ponto z. $\Pr(|z_n - z| \leq \delta)$ é a probabilidade desse evento.

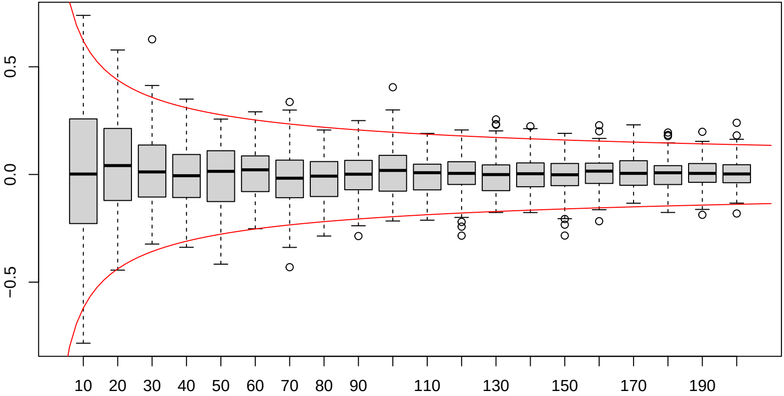
- A eq. (5) afirma que essa probabilidade se aproxima de 1 à medida que o tamanho da amostra n aumenta. A definição requer que isso seja válido para qualquer δ. Portanto, para qualquer intervalo pequeno sobre z, a distribuição de z_n concentra-se nesse intervalo para n grande.
- Note que a definição diz respeito a distribuição das variáveis aleatória z_n, não de suas realizações.

Exemplo

- Se S_n é o número de sucessos em n tentativas (binomial) com probabilidade de sucesso p, seja $Y_n = \dfrac{S_n}{n}$ e $z = p$. Então,

$$\mathbb{E}\left[\frac{S_n}{n} - p\right]^2 = \text{Var}\left[\frac{S_n}{n}\right] = \frac{npq}{n^2} \to 0 \tag{6}$$

quando $n \to \infty$. Disso segue que $\dfrac{S_n}{n} \xrightarrow{p} p$.


0.5
0.0
−0.5
10
20
30
40
50
60
70
80
90
110
130
150
170
190

LFGN

- **A LFGN é o primeiro passo, mas ela não dá uma aproximação para a distribuição de um estimador**. Uma aproximação assintótica pode ser obtida usando o conceito de convergência em distribuição.

Definição

Seja $\boldsymbol{z}_n$ um vetor aleatório com distribuição $F_n(\boldsymbol{u}) = \Pr(\boldsymbol{z}_n < \boldsymbol{u})$. Dizemos que $\boldsymbol{z}_n$ converge em distribuição para $\boldsymbol{z}$ quando $n \to \infty$, denotado por $\boldsymbol{z}_n \xrightarrow{d} \boldsymbol{z}$, se para todo $\boldsymbol{u}$ em qual $F(\boldsymbol{u}) = Pr(\boldsymbol{z} \leq \boldsymbol{u})$ é contínua, $F_n(\boldsymbol{u}) \to F(\boldsymbol{u})$ quando $n \to \infty$.

- Quando $\boldsymbol{z}_n \xrightarrow{d} \boldsymbol{z}$, é comum se referir a $\boldsymbol{z}$ como a **distribuição assintótica** ou **distribuição limite** de $\boldsymbol{z}$.

Exemplo

- Seja X_i variáveis aleatórias iid com média μ e seja $A = \dfrac{X_1 + \ldots + X_n}{n}$.
- Quando $n \to \infty$, a média amostral é igual a média populacional μ de cada variável,

$$\bar{A} = \frac{X_1 + \ldots + X_n}{n} = \frac{n}{n}\mu = \mu \tag{7}$$

- Além disso,

$$\begin{aligned}
\mathrm{var}(A) &= \mathrm{var}\left(\frac{X_1 + \ldots + X_n}{n}\right) \\
&= \mathrm{var}\left(\frac{X_1}{n}\right) + \ldots + \mathrm{var}\left(\frac{X_n}{n}\right) \\
&= \frac{\sigma^2}{n^2} + \ldots + \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}
\end{aligned} \tag{8}$$

- Por Chebyshev,

$$P\left(|X-\mu| \geq \delta\right) \leq \frac{\mathrm{var}(X)}{\delta^2} = \frac{\sigma^2}{n\delta^2} \tag{9}$$

- Quando $n \to \infty$, temos que

$$\lim_{n\to\infty} P\left(|X-\mu| \geq \delta\right) \to 0 \tag{10}$$

Definição

Se y_i é *i.i.d.* e $\mathbb{E}|y| < \infty$, então quando $n \to \infty$,

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \xrightarrow{p} \mathbb{E}(y) \tag{11}$$

- A LFGN mostra que o estimador $\overline{y}$ converge em probabilidade para a verdadeira média populacional μ. Em geral, um estimador que converge em probabilidade para o valor populacional é chamado de **consistente**.

Definição

Um estimador $\widehat{\theta}$ de um parâmetro θ é consistente se $\widehat{\theta} \xrightarrow{p} \theta$.

- Consistência é um boa propriedade para um estimador. Significa que para qualquer distribuição de dados, existe um tamanho de amostra n suficientemente grande tal que o estimador $\widehat{\theta}$ será arbitrariamente próximo do verdadeiro valor θ com alta probabilidade.

Teorema

Se y_i são i.i.d e $\mathbb{E}|y| < \infty$, então $\widehat{\mu} = \overline{y}$ é consistente para a média populacional μ.

Teorema Central do Limite

- A forma típica de estabelecer a convergência em distribuição é através do Teorema Central do Limite (TCL), que estabelece que uma média amostral padronizada converge em distribuição para um vetor aleatório normal.

Teorema

Se $y_i \in \mathbb{R}^k$ *é i.i.d. e* $\mathbb{E}\|\boldsymbol{y}\|^2 < \infty$, *então quando* $n \to \infty$

$$\sqrt{n}(\overline{\boldsymbol{y}} - \boldsymbol{\mu}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}) \tag{12}$$

em que $\boldsymbol{\mu} = \mathbb{E}\boldsymbol{y}$ *e* $\boldsymbol{V} = \mathbb{E}\left((\boldsymbol{y} - \boldsymbol{\mu})(\boldsymbol{y} - \boldsymbol{\mu})'\right)$.

- **O TCL mostra que a distribuição da média amostral é aproximadamente normal em grandes amostras**. Para alguma aplicação deve ser útil observar que o teorema não impõe qualquer restrição sobre $\boldsymbol{V}$ além do que os elementos são finitos.
- O teorema central do limite (TCL) não é um teorema único, mas abrange uma variedade de resultados relacionados com a soma de um grande número de variáveis aleatórias que, adequadamente normalizadas, têm uma distribuição limite normal.

- Uma coisa interessante sobre o TCL é que não importa qual seja a distribuição dos X_i's. Os X_i podem ser variáveis aleatórias discretas, contínuas ou mistas.
- Para ter uma ideia do TCL, vejamos alguns exemplos. Vamos supor que os X_i sejam Bernoulli(p).
- Então $\mathbb{E}[X_i] = p$ e $\text{var}(X_i) = p(1-p)$. Além disso, $Y_n = X_1 + X_2 + \ldots + X_n$ tem distribuição Binomial(n, p).
- Portanto,

$$Z_n = \frac{Y_n - np}{\sqrt{np(1-p)}} \tag{13}$$

Exemplo

- Suponha que $X_1, X_2, \ldots, X_n$ seguem uma Bernoulli com parâmetro p.
- Assim,

$$Z_n = \frac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}} \tag{14}$$

- Escolhemos $p = \dfrac{1}{2}$.

Figura 5: Binonial $(p = 0,5)$

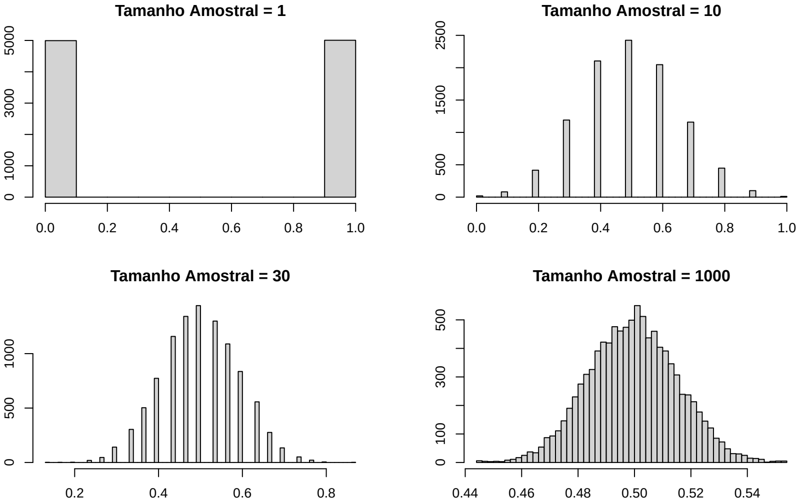
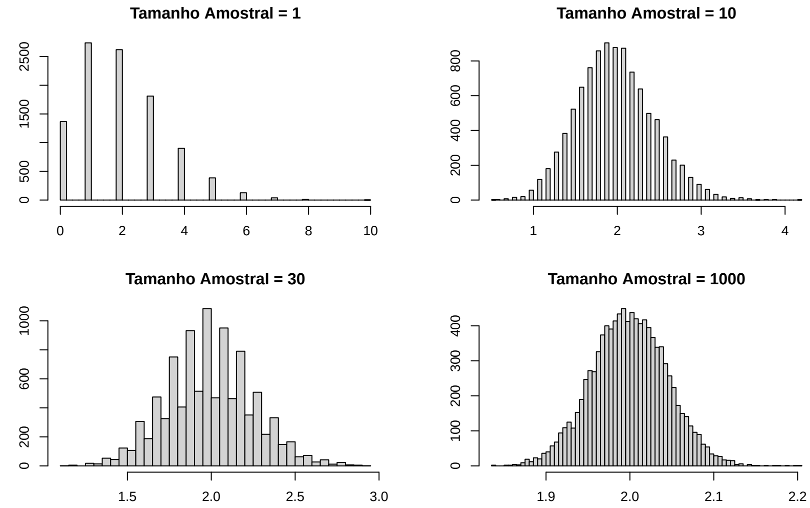


Figura 6: Poisson ($\lambda = 2$)



Exemplo

- Um caixa de banco atende os clientes que estão na fila, um por um. Suponha que o tempo de atendimento X_i para o cliente i tenha média $\mathbb{E}[X_i] = 2$ (minutos) e $\text{Var}(X_i) = 1$. Assumimos que os tempos de atendimento para diferentes clientes do banco são independentes. Seja Y o tempo total que o caixa do banco gasta atendendo 50 clientes. Encontre $P(90 < Y < 110)$.

$$\begin{aligned}
P(90 < Y \leq 110) &= P\left(\frac{90 - n\mu}{\sqrt{n}\sigma} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{90 - 100}{\sqrt{50}} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \frac{110 - 100}{\sqrt{50}}\right) \\
&= P\left(-\sqrt{2} < \frac{Y - n\mu}{\sqrt{n}\sigma} < \sqrt{2}\right) \\
&\approx \Phi(\sqrt{2}) - \Phi(-\sqrt{2}) = 0.8427 \qquad (15)
\end{aligned}$$

Exemplo

- Seja X_i uma variável aleatória indicadora do i-ésimo erro no pacote estatístico. Ou seja, $X_i = 1$ se há um erro e $X_i = 0$ caso contrário. Então X_i é i.i.d. e $X_i \sim \text{Bernoulli}(p)$. Se Y é o número total de erros no pacote, com $n = 1000$, temos:

$$Y = X_1 + \ldots + X_n \tag{16}$$

- Se $X_i \sim \text{Bernoulli}(0, 1)$, então

$$\mathbb{E}[X_i] = 0.1 \tag{17}$$

$$\text{var}(X_i) = \sigma^2 = p(1 - p) = 0.09 \tag{18}$$

- Assim, qual o valor de $P(Y > 120)$?

$$
\begin{aligned}
P(Y > 120) &= P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - n\mu}{\sqrt{n}\sigma}\right) \\
&= P\left(\frac{Y - n\mu}{\sqrt{n}\sigma} > \frac{120 - 100}{\sqrt{90}}\right) \\
&\approx 1 - \Phi\left(\frac{20}{\sqrt{90}}\right) \\
&= 0.0175 \qquad (19)
\end{aligned}
$$

Exemplo

- Há 100 homens em um avião. Seja X_i o peso do i-ésimo homem no avião. Suponha que X_i é i.i.d., e $\mathbb{E}X_i = \mu = 70$ e $\sigma(X_i) = \sigma = 30$. Encontre a probabilidade de que o peso total dos homens no avião exceda 8000 kg.

- Se W é o peso total, então $W = X_1 + \ldots + X_n$, em que $n = 100$. Então,

$$\begin{aligned} \mathbb{E}[W] &= n\mu \\ &= (100)(70) \\ &= 7000, \qquad (20) \\ \text{var}(W) &= 100\text{var}(X_i) \\ &= (100)(30)^2 \\ &= 90000 \qquad (21) \end{aligned}$$

- Assim,

$$
\begin{aligned}
P(W > 18000) &= P\left(\frac{W - 7000}{300} > \frac{8000 - 7000}{300}\right) \\
&= P\left(\frac{W - 7000}{300} > \frac{10}{3}\right) \\
&= 1 - \Phi\left(\frac{10}{3}\right) \\
&\approx 4.3 \times 10^{-4}
\end{aligned} \tag{22}
$$

Teorema (Teorema do Mapeamento Contínuo)

Seja $\boldsymbol{z}_n \in \mathbb{R}^k$ *e* $g(u)$*:* $\mathbb{R}^k \to \mathbb{R}^q$*. Se* $\boldsymbol{z}_n \xrightarrow{p} c$ *quando* $n \to \infty$ *e* $g(u)$ *é contínua em* c *então* $g(\boldsymbol{z}_n) \xrightarrow{p} g(c)$ *quando* $n \to \infty$*.*

Consistência da Estimação de MQO

- A teoria assintótica da estimação de mínimos quadrados se aplica igualmente ao modelo de projeção e ao modelo CEF linear. Vamos considerar aqui os resultados para o modelo de projeção mais amplo. Recorde que o modelo é dado por

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + e_i \tag{23}$$

para $i = 1, \ldots, n$ em que a projeção linear de $\boldsymbol{\beta}$ é dada por

$$\boldsymbol{\beta} = \left(\mathbb{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right)^{-1} \boldsymbol{E}(\boldsymbol{x}_i y_i) \tag{24}$$

- Vamos precisar de alguns pressupostos estabelecidos anteriormente.

- Muitos dos resultados a seguir se mantém sobre o pressuposto de **amostra aleatória** e o pressuposto de **segundo momento finitos**.

Pressuposto

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.*
5. $\mathbb{E}y^4 < \infty$ *e* $\mathbb{E}\|\boldsymbol{x}_i\|^4 < \infty$.

- Para mostrar que o estimador de mínimos quadrados $\widehat{\boldsymbol{\beta}}$ é consistente para o coeficiente de projeção $\boldsymbol{\beta}$, vamos usar a Lei Fraca dos Grandes Números (**LFGN**) e o Teorema do Mapeamento Contínuo (**TMP**).
- A derivação é baseada em três etapas:
  1. *O **estimador de MQO** pode ser escrito como uma função contínua de um conjunto de momentos amostrais.*
  2. *A **LFGN** mostra que os momentos amostrais convergem probabilidade para os momentos populacionais.*
  3. *O **TMP** estabelece que as funções contínuas preservam a convergência em probabilidade.*

- Observe que o estimador de MQO é uma função dos momentos amostrais $\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i$ e $\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i y_i$, isto é,

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i y_i\right) \\
&= \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} \qquad (25)
\end{aligned}
$$

- Pela aplicação da LFGN estes momentos converge em probabilidade para momentos populacionais.
- Como $(y_i, \boldsymbol{x}_i)$ são mutuamente **i.i.d.**, qualquer função de $(y_i, \boldsymbol{x}_i)$ é **i.i.d.**, incluindo $\boldsymbol{x}\boldsymbol{x}_i'$ e $\boldsymbol{x}y_i$. Vale lembrar que essas variáveis tem esperança finitas.
- Assim,

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xx}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i\boldsymbol{x}_i') = \boldsymbol{Q}_{\boldsymbol{xx}} \tag{26}$$

e

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{y}_i \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i\boldsymbol{y}_i) = \boldsymbol{Q}_{\boldsymbol{xy}} \tag{27}$$

- O TMC permite combinar estas equações para mostrar que o $\widehat{\boldsymbol{\beta}}$ converge em probabilidade para $\boldsymbol{\beta}$. Especificamente, quando $n \to \infty$,

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1} \widehat{\boldsymbol{Q}}_{\boldsymbol{xy}} \\ &\xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{xx}}^{-1} \boldsymbol{Q}_{\boldsymbol{xy}} \\ &= \boldsymbol{\beta} \end{aligned} \tag{28}$$

- Dessa forma, temos que $\widehat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$ quando $n \to \infty$.

- Esse mesmo resultado ser obtido de uma forma alternativa. Considere que

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{xe}} \tag{29}$$

em que

$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xe}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{e}_i \tag{30}$$

- Pela LFGN e $\mathbb{E}(\boldsymbol{x}e_i) = 0$, temos:

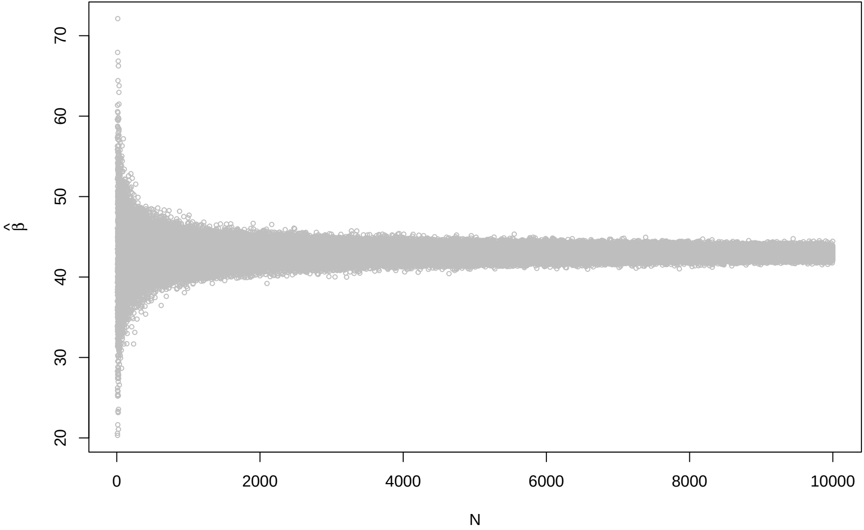
$$\widehat{\boldsymbol{Q}}_{\boldsymbol{xe}} \xrightarrow{p} \boldsymbol{Q}_{\boldsymbol{xe}} \tag{31}$$

- Como $\widehat{\boldsymbol{Q}}_{\boldsymbol{xe}} \xrightarrow{p} \mathbb{E}(\boldsymbol{x}_i e_i) = 0$, temos que

$$\begin{aligned} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1}\widehat{\boldsymbol{Q}}_{\boldsymbol{xe}} \\ &\xrightarrow{p} \widehat{\boldsymbol{Q}}_{\boldsymbol{xx}}^{-1}\boldsymbol{0} = \boldsymbol{0} \end{aligned} \tag{32}$$

- Vamos ilustrar esse resultado da consistência da estimação de Mínimos Quadrados com o gráfico a seguir considerando a estimação da equação de salários, onde iremos plotar o coeficiente associado a escolaridade obtido para diferentes tamanhos da amostra.

Figura 7: $\widehat{\beta}_1$ como Função do Tamanho da Amostra



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