Convergência do estimador dos mínimos quadrados em modelos lineares

João T. Mexia
FCT/CMA Universidade Nova de Lisboa, Departamento de Matemática
Pedro Corte Real
FCT/CMA Universidade Nova de Lisboa, Departamento de Matemática
Manuel L. Esquível
FCT Universidade Nova de Lisboa, Departamento de Matemática
CMAF Universidade de Lisboa
João Lita da Silva
FCT/CMA Universidade Nova de Lisboa, Departamento de Matemática

Resumo: Condições sobre a matriz de modelo em modelos lineares são impostas para garantir a consistência do estimador dos mínimos quadrados quando a cauda da distribuição dos erros decresce polinomialmente. Alguns dos resultados expostos ainda são válidos quando os erros não são integráveis. Exemplos de matrizes de modelo satisfazendo tais condições são exibidos no final.

Palavras–chave: modelos lineares, consistência, estimador dos mínimos quadrados.

Abstract: Conditions on the design matrix of linear models are given to ensure the consistency of LS estimators when the tail weight of the errors decreases polynomially. Some of the results are valid even in the case where the errors are not integrable. Examples of design matrices satisfying the given conditions are presented.

Keywords: linear models, consistency, least squares estimator.

1 Introdução

Recentemente, a consistência do estimador dos mínimos quadrados (MQ) em modelos lineares (ML) tem sido estudada extensivamente na perspectiva aqui adoptada (ver [2], [3], [4], [9], [10] e [11]). O nosso principal objectivo será estender a consistência do estimador dos mínimos quadrados a modelos onde:

I. os erros são independentes e identicamente distribuídos, tendo cauda de distribuição pesada e um raio espectral exclusivamente associado à matriz de modelo converge para zero com uma velocidade definida *a priori*.

ou alternativamente,

II. os erros satisfazem condições fracas de integrabilidade e apenas se exige a convergência para zero do referido raio espectral associado à matriz de modelo.

Esta abordagem vai, pois, no mesmo sentido da dos autores acima citados. Apesar dos nossos pressupostos serem mais exigentes do que os assumidos em [10], há a clara vantagem de podermos aplicar os nossos resultados a situações em que os erros não admitem valor médio.

O uso prático de modelos lineares de efeitos fixos tais como os estudados aqui é bastante comum. Uma situação típica ocorre em designs de modelos lineares múltiplos. Veja-se [7], onde em cada tratamento de uma base de design temos regressão nalgumas variáveis. Podemos considerar que, para cada tratamento, são feitas observações igualmente espaçadas nas quais as variáveis de controlo tomam como valores potências sucessivas de i para $i = 1, 2, \ldots, n$. Exemplos desta metodologia podem ser vistos ainda em [12] e [14].

Consideremos o modelo linear dado por

$$\mathbf{y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{e}_n$$

em que $\mathbf{e}_n := (e_1, \ldots, e_n)$ sendo os erros $e_1, \ldots, e_n$ variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) ou em alternativa centradas, homocedásticas e não-correlacionadas. A matriz de modelo

$$\mathbf{X}_n := \big[x_{ij}\big]_{\substack{i=1,2,\ldots,n \\ j=1,2,\ldots,\kappa}}$$

é escolhida de acordo com o modelo experimental. O vector parâmetro

$$\boldsymbol{\beta} := (\beta_1, \ldots, \beta_\kappa)$$

é estimado pelas observações $\mathbf{y}_n := (y_1, \ldots, y_n)$ para um conjunto de valores das variáveis controladas.

2 Notações e resultados auxiliares

Vejamos agora alguns conceitos e notações relevantes no que se vai seguir. O *raio espectral* de $\mathbf{A} \in \mathcal{M}_\kappa(\mathbb{R})$ está definido por $\rho(\mathbf{A}) := \sup\big\{\, |\lambda| : \ \lambda \in \mathrm{Spec}(\mathbf{A})\big\}$ onde $\mathrm{Spec}(\mathbf{A})$ é o *espectro* de $\mathbf{A}$. Denotaremos ainda a matriz transposta de $\mathbf{A}$ por $\mathbf{A}^T$ e indicaremos a norma matricial de $\mathbf{A}$ subordinada à norma euclideana por $\|\mathbf{A}\|$ tendo-se então $\|\mathbf{A}\| = \sqrt{\rho(\mathbf{A}^T\mathbf{A})}$. Sinteticamente,

$$\rho_n := \rho\Big((\mathbf{X}_n^T \mathbf{X}_n)^{-1}\Big)$$

e ao longo de todo o texto usaremos as seguintes notações: F_X será a função de distribuição de uma variável aleatória X, $\overline{F}_X := 1 - F_X$, o símbolo $\sim$ será utilizado para mencionar distribuído como, $\asymp$ denotará assimptoticamente equivalente a, $\approx$ significará aproximadamente igual a, $M_n := \max\big(e_1^2, \ldots, e_n^2\big)$, Ω_n indicará o espaço imagem da matriz de modelo $\mathbf{X}_n$ e $\mathbf{P}_{\Omega_n}\mathbf{e}_n$ designará a projecção ortogonal de $\mathbf{e}_n$ sobre Ω_n (notação utilizada também em [6] ou [13]).

Na ausência de multicolineariedade, o estimador $\widetilde{\boldsymbol{\beta}}_n$ dos mínimos quadrados de $\boldsymbol{\beta}$ é dado por:

$$\widetilde{\boldsymbol{\beta}}_n = (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T \, \mathbf{y}_n = \boldsymbol{\beta} + (\mathbf{X}_n^T \mathbf{X}_n)^{-1} \mathbf{X}_n^T \, \mathbf{e}_n \,. \tag{1}$$

Então, são apresentadas de seguida duas estimativas fundamentais em resultados posteriores. Na primeira, destacamos o facto do erro do estimador dos mínimos quadrados ser limitado pelo produto do raio espectral ρ_n por quantidades de crescimento controlado.

Lema 1 *No modelo linear em causa tem-se,*

$$\left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|^2 \leqslant n \, \rho_n \, M_n \quad q.c. \tag{2}$$

Demonstração. Com efeito,

$$\begin{aligned} \left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|^2 &= \left\| \left(\mathbf{X}_n^T \, \mathbf{X}_n\right)^{-1} \mathbf{X}_n^T \, \mathbf{e}_n \right\|^2 \leqslant \\ &\leqslant \rho \left(\mathbf{X}_n \left(\mathbf{X}_n^T \mathbf{X}_n\right)^{-2} \mathbf{X}_n^T \right) \|\mathbf{e}_n\|^2 = \rho \left(\left(\mathbf{X}_n^T \, \mathbf{X}_n\right)^{-1} \right) \|\mathbf{e}_n\|^2 \leqslant n \, \rho_n \, M_n \end{aligned} \tag{3}$$

e a prova está concluída. □

A segunda estimativa mostra claramente o fenómeno da contracção do erro tendo-se, neste caso, que o limite superior do erro do estimador dos mínimos quadrados é dado pelo produto do raio espectral ρ_n por $\mathbf{P}_{\Omega_n} \mathbf{e}_n$.

Lema 2 *No modelo linear em causa tem-se,*

$$\left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|^2 \leqslant \rho_n \, \left\|\mathbf{P}_{\Omega_n} \mathbf{e}_n\right\|^2 \quad q.c. \tag{4}$$

Demonstração. Uma vez que o núcleo de $\mathbf{X}_n^T$ verifica $\mathrm{Ker}(\mathbf{X}_n^T) = \Omega_n^{\perp}$ vem $\mathbf{X}_n^T \mathbf{e}_n = \mathbf{X}_n^T \left(\mathbf{P}_{\Omega_n} \mathbf{e}_n\right)$. Então,

$$\begin{aligned} \left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|^2 &= \left\| \left(\mathbf{X}_n^T \, \mathbf{X}_n\right)^{-1} \mathbf{X}_n^T \, \mathbf{e}_n \right\|^2 = \\ &= \left\| \left(\mathbf{X}_n^T \, \mathbf{X}_n\right)^{-1} \mathbf{X}_n^T \left(\mathbf{P}_{\Omega_n} \mathbf{e}_n\right) \right\|^2 \leqslant \rho \left(\left(\mathbf{X}_n^T \, \mathbf{X}_n\right)^{-1} \right) \left\|\mathbf{P}_{\Omega_n} \mathbf{e}_n\right\|^2 , \end{aligned}$$

o que estabelece a tese. □

3 Convergência do estimador

Começamos por demonstrar que $(\widetilde{\boldsymbol{\beta}}_n)_{n \in \mathbb{N}}$ converge fortemente para $\boldsymbol{\beta}$ se assimptoticamente $\rho_n = O(1/n^s)$ onde o valor de s é suficientemente grande relativamente ao comportamento da cauda de distribuição dos erros. Supomos neste primeiro grupo de resultados que os erros são i.i.d. mas não necessariamente integráveis.

Teorema 1 *Se* $r > 0$ *e*

- *as componentes de* $\mathbf{e}_n$ *são i.i.d. tendo-se* $e_i \sim e$ *para todo o* $i = 1, 2, \ldots, n$
- $\overline{F}_{e^2}(x) = O\big(x^{-1/r}\big)$
- $\exists s > 2r + 1 : \ \rho_n = O\big(1/n^s\big)$,

então $\widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\text{q.c.}} \boldsymbol{\beta}$.

Demonstração. Da estimativa (2) da secção anterior, apenas teremos de mostrar a convergência completa da sucessão $\big\{n\,\rho_n\,M_n,\ n \geqslant 1\big\}$, ou seja,

$$\forall \varepsilon > 0, \quad \sum_{n=1}^{+\infty} \mathbb{P}\big[n\,\rho_n\,M_n > \epsilon\big] < +\infty \,.$$

Mas isto é imediato porque $\displaystyle\sum_{n=1}^{+\infty} n^{\frac{r-s+1}{r}} < +\infty$ e

$$\mathbb{P}[n\,\rho_n\,M_n > \varepsilon] = 1 - \left(1 - \overline{F}_{e^2}\left(\frac{\varepsilon}{n\,\rho_n}\right)\right)^n \leqslant 1 - \left(1 - \frac{K}{\left(\frac{\varepsilon}{n\,\rho_n}\right)^{1/r}}\right)^n =$$
$$= 1 - \left(1 - \frac{K\,n^{\frac{r-s+1}{r}}\,(n^s\rho_n)^{\frac{1}{r}}}{\varepsilon^{1/r}}\right)^n \asymp \frac{K}{\varepsilon^{1/r}}\,n^{\frac{r-s+1}{r}}\,(n^s\rho_n)^{\frac{1}{r}} \asymp n^{\frac{r-s+1}{r}} \,. \tag{5}$$

□

Enfraquecendo a anterior condição sobre a velocidade de convergência para zero do raio espectral ρ_n conseguimos obter ainda a convergência em probabilidade do estimador dos mínimos quadrados.

Corolário 1 *Se* $r > 0$ *e*

- *as componentes de* $\mathbf{e}_n$ *são i.i.d. tendo-se* $e_i \sim e$ *para todo o* $i = 1, 2, \ldots, n$
- $\overline{F}_{e^2}(x) = O\big(x^{-1/r}\big)$
- $\exists s > r + 1 : \ \rho_n = O\big(1/n^s\big)$,

então $\widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \boldsymbol{\beta}$.

Demonstração. Da expressão obtida em (5) tem-se

$$\lim_{n\to+\infty} \mathbb{P}[n\,\rho_n\,M_n > \varepsilon] \leqslant K\,\left(\frac{a}{\varepsilon}\right)^{1/r} \lim_{n\to+\infty} n^{\frac{r-s+1}{r}} = 0,$$

dado que estamos a assumir que $r > 0$ e $s > r+1$. □

Observação 1. Seguem-se algumas observações importantes:

a) A condição sobre o raio espectral em [10] é $(\mathbf{X}_n^T\mathbf{X}_n)^{-1} = O\big(n^{-(2-\alpha)/\alpha}\big)$ onde as componentes do vector dos erros verificam $\mathbb{E}[|e_1|^\alpha] < +\infty$ com $\alpha \geqslant 1$. É de fácil constatação que as hipóteses $\rho_n = O\big(1/n^s\big)$ com $s > 2r+1$ e $\overline{F}_{e^2}(x) = O(x^{-1/r})$ são claramente mais fortes do que as admitidas em [10] mas existe a óbvia vantagem deste resultado poder ser aplicado à situação em que $0 < \alpha < 1$.

b) A condição sobre a função de distribuição do quadrado dos erros pode ser substituída por outra equivalente sobre a função de distribuição dos erros. Visto que $F_{e^2}(x^2) = F_e(x) - F_e(-x)$ tem-se,

$$\overline{F}_{e^2}(x) \leqslant \frac{K}{x^{1/r}} \quad\Longleftrightarrow\quad F_e(x) - F_e(-x) = 1 - \overline{F}_{e^2}(x^2) \geqslant 1 - \frac{K}{x^{2/r}}\ .$$

c) Estes resultados aplicam-se a distribuições tais como Pareto, Student generalizada, Cauchy generalizada (ver [15]) e também a distribuições como a normal e a Fréchet.

A imposição da condição mais restritiva

$$\overline{F}_{e^2}(x) = \frac{K}{x^{1/r}}$$

permite apenas a aplicação à distribuição de Pareto. Para contornar este obstáculo são apresentados, de seguida, outros resultados que possibilitaram alargar a classe de distribuições admitindo de forma natural que $L = L(x)$. Nos próximos resultados faremos hipóteses sobre F_{e^2} que permitiram a convergência em distribuição de M_n.

Teorema 2 *Se $r > 0$ e*

- *as componentes de $\mathbf{e}_n$ são i.i.d. tendo-se $e_i \sim e$ para todo o $i = 1, 2, \ldots, n$*
- *para alguma função de variação lenta $L(x)$,*

$$\overline{F}_{e^2}(x) = x^{-1/r}\,L(x), \quad x \geqslant x_0$$

- $\exists s > 2r+1\colon\ \ \rho_n = O\big(1/n^s\big)$, *então* $\widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\text{q.c.}} \boldsymbol{\beta}$.

Demonstração. Como $\overline{F}_{e^2}$ é de variação regular tem-se

$$c_n^{-1}\, M_n \xrightarrow[n\to+\infty]{\mathrm{d}} \Phi_{1/r} \tag{6}$$

com $c_n = n^r\, L(n)$ para alguma função de variação lenta $L(x)$ e $\Phi_{1/r}(x) = \exp(-x^{-1/r})$ para $x > 0$ a função de distribuição de Fréchet (ver [5]). Visto que $s > 2r+1$, escolhendo $\delta > 0$ tal que $s > 2r+1+\delta$, obtém-se

$$\begin{aligned}\mathbb{P}[n\,\rho_n\, M_n > \varepsilon] = \mathbb{P}\left[c_n^{-1}\, M_n > \frac{\varepsilon}{n\,\rho_n\, c_n}\right] \approx 1 - \Phi_{1/r}\left(\frac{\varepsilon}{n\,\rho_n\, c_n}\right) = \\ = 1 - \exp\left(-\frac{n^{-s/r}\, n^{1/r}\,(n^s\,\rho_n)^{1/r}\, n\,\big(L(n)/n^\delta\big)^{1/r}\, n^{\delta/r}}{\varepsilon^{1/r}}\right) \asymp n^{\frac{\delta+1-s}{r}+1}\,.\end{aligned} \tag{7}$$

Com efeito, $\big(L(n)/n^\delta\big)_{n\in\mathbb{N}}$ converge para zero (ver [1]) e a primeira aproximação em (7) é plenamente justificada uma vez que tomando $a_n = \varepsilon/n\rho_n c_n$ vem

$$\lim_{n\to+\infty}\left|F_{c_n^{-1}M_n}(a_n) - \Phi_{1/r}(a_n)\right| \leqslant \lim_{n\to+\infty}\left(\sup_{x\in\mathbb{R}}\left|F_{c_n^{-1}M_n}(x) - \Phi_{1/r}(x)\right|\right) = 0$$

em virtude de (6) e do facto da distribuição limite $\Phi_{1/r}$ ser contínua. Das hipóteses sobre r, s e δ resulta $\displaystyle\sum_{n=1}^{+\infty} n^{\frac{\delta+1-s}{r}+1} < +\infty$, o que conclui a prova. □

Enfraquecendo novamente a condição sobre a velocidade de convergência para zero do raio espectral ρ_n obtemos a convergência em probabilidade do estimador dos mínimos quadrados.

Corolário 2 *Se* $r > 0$ *e*

- *as componentes de* $\mathbf{e}_n$ *são i.i.d. tendo-se* $e_i \sim e$ *para todo o* $i = 1, 2, \ldots, n$
- *para alguma função de variação lenta* $L(x)$

$$\overline{F}_{e^2}(x) = x^{-1/r}\, L(x), \quad x \geqslant x_0$$

- $\exists s > r+1: \quad \rho_n = O\big(1/n^s\big)$,

então $\widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \boldsymbol{\beta}$.

Demonstração. Sendo b um limite inferior da sucessão $\left(L(n)/n^{\delta}\right)_{n\in\mathbb{N}}$ obtemos da expressão (7) e das condições assumidas para r e s que,

$$\lim_{n\to+\infty} \mathbb{P}[n\,\rho_n\,M_n > \varepsilon] \leqslant \left(\frac{a\,b}{\varepsilon}\right)^{1/r} \lim_{n\to+\infty} n^{\frac{\delta+1-s}{r}+1} = 0,$$

onde $a>0$ e δ é escolhido de modo a que $s > r+1+\delta$. □

Observação 2. Estes resultados podem ser aplicados a todas as distribuições de cauda pesada mencionadas na anterior observação, inclusive, quando F_{e^2} tem distribuição log-gama (ver [5]).

Vejamos agora o que acontece quando $\mathbb{E}[\mathbf{e}_n] = \mathbf{0}$ e se admite que a matriz de covariância dos erros verifica $\mathbb{V}[\mathbf{e}_n] = \sigma^2\,\mathbf{I}_n$ onde $\mathbf{I}_n$ é a matriz identidade de ordem n (significa isto que $e_1,\ldots,e_n$ são variáveis aleatórias centradas, homocedásticas e não-correlacionadas).

Proposição 1 *Se $\mathbb{E}[\mathbf{e}_n] = \mathbf{0}$ e $\mathbb{V}[\mathbf{e}_n] = \sigma^2\,\mathbf{I}_n$ com $\mathbf{I}_n$ a matriz identidade de ordem n então,*

$$\mathbb{E}\left[\left\|\mathbf{P}_{\Omega_n}\mathbf{e}_n\right\|^2\right] = \kappa\,\sigma^2\ . \tag{8}$$

Demonstração. Seja $\mathbf{A}_n$ uma matriz na qual os vectores linha constituem uma base ortonormal de Ω_n. Então $\mathbf{Q}_n := \mathbf{A}_n^T\,\mathbf{A}_n$ é a matriz da projecção ortogonal sobre o espaço Ω_n e como $\mathbf{Q}_n$ é simétrica e idempotente vem,

$$\begin{aligned}\left\|\mathbf{P}_{\Omega_n}\mathbf{e}_n\right\|^2 = \left\|\mathbf{Q}_n\,\mathbf{e}_n\right\|^2 = \mathbf{e}_n^T\,\mathbf{Q}_n^T\,\mathbf{Q}_n\mathbf{e}_n = \mathbf{e}_n^T\,\mathbf{Q}_n^2\,\mathbf{e}_n = \\ = \mathbf{e}_n^T\,\mathbf{Q}_n\,\mathbf{e}_n = \mathbf{e}_n^T\,\mathbf{A}_n^T\,\mathbf{A}_n\,\mathbf{e}_n = \left\|\mathbf{A}_n\,\mathbf{e}_n\right\|^2\ .\end{aligned} \tag{9}$$

Das hipóteses assumidas obtemos ainda $\mathbb{E}\left[\mathbf{A}_n\,\mathbf{e}_n\right] = \mathbf{A}_n\,\mathbb{E}\left[\mathbf{e}_n\right] = \mathbf{0}$ e

$$\mathbb{V}\left[\mathbf{A}_n\,\mathbf{e}_n\right] = \mathbf{A}_n\,\mathbb{V}\left[\mathbf{e}_n\right]\,\mathbf{A}_n^T = \mathbf{A}_n\,\sigma^2\,\mathbf{I}_n\,\mathbf{A}_n^T = \sigma^2\,\mathbf{I}_\kappa$$

o que implica $\mathbb{E}\left[\left\|\mathbf{P}_{\Omega_n}\mathbf{e}_n\right\|^2\right] = \mathbb{E}\left[\left\|\mathbf{A}_n\,\mathbf{e}_n\right\|^2\right] = \kappa\,\sigma^2$. □

Com a ajuda da anterior proposição, alcançamos a convergência em probabilidade e em momento de 2ª ordem do estimador dos mínimos quadrados impondo a mais moderada das hipóteses sobre o raio espectral ρ_n. De referir que o resultado é verdadeiro mesmo quando os erros não são independentes.

Teorema 3 *Se $\mathbb{E}[\mathbf{e}_n] = \mathbf{0}$, $\mathbb{V}[\mathbf{e}_n] = \sigma^2\,\mathbf{I}_n$ e $\lim_{n\to+\infty}\rho_n = 0$ então*

$$\widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \boldsymbol{\beta} \qquad e \qquad \widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\mathcal{L}^2} \boldsymbol{\beta}\ .$$

Demonstração. Em consequência da estimativa (4) e da desigualdade de Markov obtém-se para qualquer $\varepsilon > 0$,

$$\mathbb{P}\left[\left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|^2 > \varepsilon\right] \leqslant \mathbb{P}\left[\left\|\mathbf{P}_{\Omega_n}\mathbf{e}_n\right\|^2 > \frac{\varepsilon}{\rho_n}\right] \leqslant \frac{\rho_n}{\varepsilon}\,\mathbb{E}\left[\left\|\mathbf{P}_{\Omega_n}\mathbf{e}_n\right\|^2\right] = \frac{\kappa\,\rho_n\,\sigma^2}{\varepsilon},$$

o que demonstra a convergência em probabilidade. Para a convergência em momento de 2ª ordem basta observar que da estimativa (4) e da fórmula (8) da anterior proposição se tem

$$\mathbb{E}\left[\left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\|^2\right] \leqslant \rho_n\,\mathbb{E}\left[\left\|\mathbf{P}_{\Omega_n}\mathbf{e}_n\right\|^2\right] = \kappa\,\rho_n\,\sigma^2$$

e o resultado segue da condição imposta sobre ρ_n. □

A consistência forte do estimador $\widetilde{\boldsymbol{\beta}}_n$ está coberta pelo próximo resultado que foi enunciado e demonstrado em [8] constituindo uma extensão da Lei Forte de Kolmogorov para a Regressão Múltipla.

Teorema 4 *Se*

- *as componentes de* $\mathbf{e}_n$ *são i.i.d.* (*com* $e_i \sim e$ *para todo o* $i = 1, 2, \ldots, n$) *e* $\mathbb{E}[\mathbf{e}_n] = \mathbf{0}$
- $\displaystyle\lim_{n\to+\infty} \frac{1}{n}\left(\mathbf{X}_n^T\,\mathbf{X}_n\right) = \mathbf{H}$ *com* $\mathbf{H}$ *definida positiva*
- *as linhas da matriz* $\mathbf{X}_n$ *pertencem a um compacto de* $\mathbb{R}^\kappa$,

então $\widetilde{\boldsymbol{\beta}}_n \xrightarrow[n\to+\infty]{\text{q.c.}} \boldsymbol{\beta}$.

Demonstração. Como existe $n_0 \in \mathbb{N}$ tal que $\rho\left(\left(\frac{1}{n}\,\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}\right) \leqslant 2\,\rho\big(\mathbf{H}^{-1}\big)$, $\forall n \geqslant n_0$ obtém-se da fórmula (1) que

$$\begin{aligned}\left\|\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right\| &= \left\|(\mathbf{X}_n^T\,\mathbf{X}_n)^{-1}\,\mathbf{X}_n^T\,\mathbf{e}_n\right\| \leqslant \\ &\leqslant \rho\left(\left(\frac{1}{n}\,\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}\right)\left\|\frac{1}{n}\mathbf{X}_n^T\,\mathbf{e}_n\right\| \leqslant 2\,\rho\big(\mathbf{H}^{-1}\big)\left\|\frac{1}{n}\,\mathbf{X}_n^T\,\mathbf{e}_n\right\|, \quad \forall n \geqslant n_0.\end{aligned}$$

Resta então mostrar que $\left\|\frac{1}{n}\,\mathbf{X}_n^T\,\mathbf{e}_n\right\|$ converge para zero quase certamente onde

$$\frac{1}{n}\,\mathbf{X}_n^T\,\mathbf{e}_n = \begin{bmatrix} \displaystyle\frac{1}{n}\sum_{i=1}^{n} x_{i1}\,e_i \\ \vdots \\ \displaystyle\frac{1}{n}\sum_{i=1}^{n} x_{i\kappa}\,e_i \end{bmatrix}.$$

Consideremos então dois casos:

1. $\mathbb{E}\big[e^2\big] < +\infty$.

Das hipóteses assumidas vem

$$\begin{aligned} &\mathbb{E}\left[\frac{x_{ij}\, e_i}{i}\right] = 0, \quad \forall i = 1, \ldots, n\,, \quad \forall j = 1, \ldots, \kappa \\ &\exists C > 0\colon\ |x_{ij}| \leqslant C, \quad \forall i = 1, \ldots, n\,, \quad \forall j = 1, \ldots, \kappa, \end{aligned}$$

donde

$$\sum_{i=1}^{+\infty} \mathbb{V}\left[\frac{x_{ij}\, e_i}{i}\right] = \mathbb{E}\big[e^2\big] \sum_{i=1}^{+\infty} \frac{x_{ij}^2}{i^2} \leqslant C^2\, \mathbb{E}\big[e^2\big] \sum_{i=1}^{+\infty} \frac{1}{i^2} < +\infty\,.$$

Utilizando agora um conhecido resultado para somas de variáveis aleatórias independentes com média nula surge que, para cada $j = 1, \ldots, \kappa$,

$$\sum_{i=1}^{+\infty} \frac{x_{ij}\, e_i}{i} \quad \text{converge q.c.}$$

e do lema de Kronecker podemos concluir que cada componente do vector $n^{-1}\, \mathbf{X}_n^T\, \mathbf{e}_n$ converge para zero quase certamente.

2. $\mathbb{E}\big[e^2\big] = +\infty$.

Considerando

$$e_i' := e_i\, I_{\{|e_i| \leqslant i\}} = \begin{cases} e_i & \text{se} \quad |e_i| \leqslant i \\ 0 & \text{se} \quad |e_i| > i \end{cases} \qquad (i = 1, \ldots, n),$$

obtém-se do lema da truncatura de Kolmogorov

$$\begin{aligned} &\textbf{(i)} \quad \lim_{i \to +\infty} \mathbb{E}[e_i'] = \mathbb{E}[e] \\ &\textbf{(ii)} \quad \mathbb{P}\big[\liminf\{e_i' = e_i\}\big] = 1 \\ &\textbf{(iii)} \quad \sum_{i=1}^{+\infty} \frac{\mathbb{V}[e_i']}{i^2} < +\infty, \end{aligned}$$

e raciocinando da forma análoga ao caso **1.** podemos escrever

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} x_{ij}\, e_i' = 0 \quad \text{q.c.}$$

para cada $j = 1, \ldots, \kappa$. Uma vez que $\mathbb{P}\big[\liminf\{e_i' = e_i\}\big] = 1$ concluímos então que cada componente do vector $n^{-1}\,\mathbf{X}_n^T\,\mathbf{e}_n$ converge para zero quase certamente.

Portanto, em ambas as situações

$$\left\| \frac{1}{n}\,\mathbf{X}_n^T\,\mathbf{e}_n \right\| \xrightarrow[n\to+\infty]{\text{q.c.}} 0$$

e a tese está demonstrada. □

4 Exemplos de matrizes de modelo

Uma questão natural ligada aos resultados desenvolvidos anteriormente é a existência de matrizes de modelo com velocidade de convergência para zero arbitrária quando o número de observações cresce infinitamente. Na realidade, é o caso em que o peso nas observações mais recentes cresce polinomialmente.

Teorema 5 *Dado um inteiro positivo p existe uma matriz de modelo $\mathbf{X}_n(p)$ tal que*

$$\lim_{n\to+\infty} n^{2p+1}\cdot\rho\left(\left(\mathbf{X}_n^T(p)\,\mathbf{X}_n(p)\right)^{-1}\right) = (2p+1)\,(2p+2)^2\,.$$

Demonstração. Consideremos a seguinte matriz de modelo

$$\mathbf{X}_n(p) = \begin{bmatrix} 1^p & 1^{p+1} \\ 2^p & 2^{p+1} \\ \vdots & \vdots \\ n^p & n^{p+1} \end{bmatrix}.$$

Então,

$$\mathbf{X}_n^T(p)\,\mathbf{X}_n(p) = \begin{bmatrix} S_{2p}(n) & S_{2p+1}(n) \\ S_{2p+1}(n) & S_{2p+2}(n) \end{bmatrix},$$

onde $S_p(n) := 1^p + 2^p + \ldots + n^p$. Os valores próprios da matriz $\left(\mathbf{X}_n^T(p)\,\mathbf{X}_n(p)\right)^{-1}$ são não-negativos e o seu raio espectral é

$$\rho_n = \frac{S_{2p}(n) + S_{2p+2}(n) + \sqrt{\left(S_{2p}(n) - S_{2p+2}(n)\right)^2 + 4\left(S_{2p+1}(n)\right)^2}}{2\left(S_{2p}(n)\,S_{2p+2}(n) - \left(S_{2p+1}(n)\right)^2\right)}\,.$$

Da fórmula de Faulhaber sabemos que

$$1^p + 2^p + \ldots + n^p = \sum_{j=1}^{p+1} b_{pj}\,n^j,$$

com $b_{pj} = \dfrac{(-1)^{p-j+1} B_{p-j+1}\, p!}{j!\,(p-j+1)!}$ e B_i $(i = 0, 1, \ldots, p)$ os números de Bernoulli (i.e os números definidos pela identidade $\dfrac{x}{e^x - 1} \equiv \sum_{n=0}^{\infty} \dfrac{B_n\, x^n}{n!}$). Deste modo, com $u_{2p+3} = 1/(2p+3)$,

$$S_{2p}(n) + S_{2p+2}(n) = \sum_{j=1}^{2p+3} u_j\, n^j \ ,$$

com $v_{4p+6} = 1/(2p+3)^2$,

$$\Big(S_{2p}(n) - S_{2p+2}(n)\Big)^2 + 4\left(S_{2p+1}(n)\right)^2 = \sum_{j=1}^{4p+6} v_j\, n^j \ ,$$

e, finalmente,

$$2\left(S_{2p}(n)\, S_{2p+2}(n) - \left(S_{2p+1}(n)\right)^2\right) = \sum_{j=1}^{4p+4} z_j\, n^j \ ,$$

com $z_{4p+4} = \dfrac{2}{(2p+1)(2p+2)^2(2p+3)}$. Através destas expressões obtém-se

$$\lim_{n\to+\infty} n^{2p+1} \cdot \rho_n = \frac{\frac{2}{2p+3}}{\frac{2}{(2p+1)\,(2p+2)^2\,(2p+3)}} = (2p+1)\,(2p+2)^2 \ ,$$

o que prova a tese. □

Observação 3. No caso em estudo, podemos concluir que dado um modelo possuindo (eventualmente) erros de elevada potência no índice de cauda é sempre possível encontrar uma matriz de modelo tal que o estimador dos mínimos quadrados seja fortemente consistente.

Referências

[1] Bingham, N. H., Goldie, C. M. e Teugels, J. L. (1987). *Regular Variation.* Cambridge University.

[2] Chen, X. (1994). Some results on consistency of LS estimates. *Chin. Sci. Bull.*, Vol. 39(22), p. 1872-1876.

[3] Chen, X. (1995). Consistency of LS estimates of multiple regression under a lower order moment condition. *Sci. Chin.*, Vol.38(12), p. 1420-1431.

[4] Chen, X. (2001). A note on the consistency of LS estimates in linear models. *Chin. Ann. Math. Ser. B* , Vol. 22(4), p. 471-474.

[5] Embrechts, P., Klüppelberg, C. e Mikosch, T. (1997). *Modelling Extremal Events.* Springer.

[6] Gros, J. (2003). *Linear Regression.* Springer.

[7] Mexia, J. T. (1987). Multitreatment regression designs. *Trabalhos de Investigação da FCT-UNL*, 1.

[8] Mexia, J. T. e Corte Real, P. (2001). Extension of Kolmogorov's strong law to multiple regression. *Revista de Estatística*, N. 24 (2º quadrimestre 2001), p. 277-278.

[9] Mexia, J. T., Corte Real P., Esquível, M. L. e Lita da Silva, J. (2005). Strong convergence of estimators, sub-exponential errors and collapse of radial and related errors (submetido para publicação).

[10] Mingzhong, J. (1998). Some new results of the strong consistency of multiple regression coefficients. *Proceedings of the Second Asian Mathematical Conference 1995* (Tangmanee, S. e Schulz, E, eds.), p. 514-519. World Scientific.

[11] Mingzhong, J. e Chen, X. (1999). Strong consistency of least squares estimate in multiple regression when the error variance is infinite. *Stat. Sin.*, Vol. 9(1), p. 289-296.

[12] Moreira, E. E., Ribeiro, A. B., Mateas, E. P., Mexia, J. T. e Ottosen, L. M. (2005). Regression modeling of electrodialytic removal of Cu, Cr, and As fro CCA treated timber waste. *Wood Science and Technology* (no prelo).

[13] Pestman, W. (1998). *Mathematical Statistics.* Walter de Gruyter, Berlin.

[14] Ribeiro, A. B. e Mexia, J. T. (1997). A dynamical model for the electrokinetic removal of copper from a polluted soil. *Journal of Hazardous Materials*, Vol. 56, p. 257-271.

[15] Reiss, R. D. e Thomas M. (2001). *Statistical Analysis of Extreme Values* (second edition). Birkhäuser.