1. Publicação nº *INPE-2234-PRE/022*	2. Versão	3. Data	5. Distribuição ☐ Interna ☒ Externa ☐ Restrita
4. Origem *MEDEP*	Programa *DES*		

6. Palavras chaves - selecionadas pelo(s) autor(es)

ESTIMADOR "RIDGE"
MODELOS DE REGRESSÃO LINEAR
MULTICOLINEARIDADE

7. C.D.U.: *519.233.5*

8. Título *INPE-2234-PRE/022*	10. Páginas: *18*
ESTUDO DE PROBLEMAS DE REGRESSÃO MAL-CONDICIONADOS OU COM PRESENÇA DE MULTICOLINEARIDADE	11. Última página: *17*
	12. Revisada por *Carlos A. Cappelletti*
9. Autoria *Luiz Antonio Nogueira Lorena* *José Roberto Reis*	13. Autorizada por *Nelson de Jesus Parada* *Diretor*
Assinatura responsável	

14. Resumo/Notas

O trabalho apresenta um estudo de modelos de regressão linear mal-condicionados ou com presença de multicolinearidade. O desenvolvimento é dirigido para ressaltar os graves efeitos que são provocados pelas instabilidades nos coeficientes estimados, no caso em que o sistema de equações normais apresenta-se mal-condicionado. O número de condição da matriz normal (apresentada na forma de matriz de correlações) é relacionado ao problema da multicolinearidade. O estimador "ridge" é apresentado como alternativo ao de mínimos quadrados, para casos onde há forte multicolinearidade, evidenciando as relações entre eles em um exemplo no final do trabalho.

15. Observações *Este trabalho foi submetido para apresentação no III Encontro Brasileiro de Econometria, a se realizar em Olinda, Pernambuco, de 08 a 11 de dezembro de 1981.*

ESTUDO DE PROBLEMAS DE REGRESSÃO MAL-CONDICIONADOS OU COM PRESENÇA DE MULTICOLINEARIDADE

Luiz Antonio Nogueira Lorena
José Roberto Reis
Instituto de Pesquisas Espaciais
Conselho Nacional de Desenvolvimento Científico e Tecnológico
Caixa Postal 515 - 12200 - São José dos Campos, SP

RESUMO

O trabalho apresenta um estudo de modelos de regressão linear mal-condicionados ou com presença de multicolinearidade. O desenvolvimento é dirigido para ressaltar os graves efeitos que são provocados pelas instabilidades nos coeficientes estimados, no caso em que o sistema de equações normais apresenta-se mal-condicionado. O número de condição da matriz normal (apresentada na forma de matriz de correlações) é relacionado ao problema da multicolinearidade. O estimador "ridge" é apresentado como alternativo ao de mínimos quadrados, para casos onde há forte multicolinearidade, evidenciando as relações entre eles em um exemplo no final do trabalho.

ABSTRACT

This paper presents a study of ill-conditioned linear regression models or with multicolinearity. The objective is to point out the bad effects resulting from perturbations in the estimated coefficients, when the system of normal equations is ill-conditioned. The condition number of the normal matrix (presented in the correlation matrix form) is related to the multicolinearity problem. The ridge estimator is presented as alternative to the least squares one, when there exists strong multicolinearity. The relationship between these two estimators is showed in an example at the end of the paper.

1. INTRODUÇÃO

O presente trabalho tem o objetivo de descrever o método de estimação "ridge", sugerido por Hoerl e Kennard [1], para atenuar os efeitos de forte multicolinearidade em modelos de regressão linear múltipla.

O trabalho está voltado unicamente para o aspecto de estimação, sendo evidenciado os efeitos da multicolinearidade na solução de sistemas de equações normais mal-condicionados.

Na Seção 2 é apresentada a normalização do modelo linear geral de regressão múltipla. Na Seção 3 é enfatizada a relação existente entre mal-condicionamento e multicolinearidade. O estimador "ridge" é apresentado na Seção 4, sendo dado um exemplo de sua aplicação na Seção 5.

2. NORMALIZAÇÃO DO MODELO DE REGRESSÃO MÚLTIPLA

Seja o modelo de regressão linear

$$\underset{\sim}{Y} = X \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} , \tag{2.1}$$

onde $X(n \times p)$ é a matriz de observações das variáveis independentes do modelo, e admite-se que $r(X) = p < n$ (X tem característica p),

$\underset{\sim}{Y}(n \times 1)$ é o vetor de observações da variável dependente do modelo,

$\underset{\sim}{\beta}(p \times 1)$ é o vetor de parâmetros, e

$\underset{\sim}{\varepsilon}(n \times 1)$ é o vetor de erros, tal que $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $E(\underset{\sim}{\varepsilon}\, \underset{\sim}{\varepsilon}^t) = \sigma^2 I_n$, onde I_n é a matriz identidade de ordem n, e t denota transposição.

Usando-se o estimador não-viciado de variância mínima, ou o estimador de máxima-verossimilhança quando o vetor aleatório $\underset{\sim}{\varepsilon}$ é normal, tem-se

$$\hat{\underset{\sim}{\beta}} = (X^t X)^{-1} X^t \underset{\sim}{Y} ,$$

como estimador de $\underset{\sim}{\beta}$, que proporciona a mínima soma de resíduos ao quadrado:

$$S(\hat{\underset{\sim}{\beta}}) = (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}})^t \, (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}) .$$

Seja

$$X = [\underset{\sim}{x}_1 \vdots \underset{\sim}{x}_2 \vdots \cdots \vdots \underset{\sim}{x}_p] ,$$

onde $\underset{\sim}{x}_i = (x_{1i}, x_{2i}, \ldots, x_{ni})^t$, $i = 1, \ldots, p$, e defina-se o produto interno padrão por

$$\langle \underset{\sim}{x}_i, \underset{\sim}{x}_j \rangle = \sum_{k=1}^{n} x_{ki}\, x_{kj} = \underset{\sim}{x}_j^t\, \underset{\sim}{x}_i \text{ , e}$$

e a norma euclidiana do vetor $\underset{\sim}{x}_i$ por

$$\|\underset{\sim}{x}_i\| = \sqrt{\langle \underset{\sim}{x}_i, \underset{\sim}{x}_i \rangle} = \sqrt{\sum_{k=1}^{n} x_{ki}^2} \; .$$

Assim,

$$\tilde{X} = \left[\frac{\underset{\sim}{x}_1}{\|\underset{\sim}{x}_1\|} \;\vdots\; \frac{\underset{\sim}{x}_2}{\|\underset{\sim}{x}_2\|} \;\vdots\; \cdots \;\vdots\; \frac{\underset{\sim}{x}_p}{\|\underset{\sim}{x}_p\|} \right]$$

é a matriz X "normalizada", e

$$\tilde{X}^t\tilde{X} = \begin{bmatrix} \dfrac{\langle \underset{\sim}{x}_1, \underset{\sim}{x}_1 \rangle}{\|\underset{\sim}{x}_1\|^2} & \dfrac{\langle \underset{\sim}{x}_2, \underset{\sim}{x}_1 \rangle}{\|\underset{\sim}{x}_2\| \cdot \|\underset{\sim}{x}_1\|} & \cdots & \dfrac{\langle \underset{\sim}{x}_p, \underset{\sim}{x}_1 \rangle}{\|\underset{\sim}{x}_p\| \cdot \|\underset{\sim}{x}_1\|} \\ \dfrac{\langle \underset{\sim}{x}_1, \underset{\sim}{x}_2 \rangle}{\|\underset{\sim}{x}_1\| \cdot \|\underset{\sim}{x}_2\|} & \dfrac{\langle \underset{\sim}{x}_2, \underset{\sim}{x}_2 \rangle}{\|\underset{\sim}{x}_2\|^2} & \cdots & \dfrac{\langle \underset{\sim}{x}_p, \underset{\sim}{x}_2 \rangle}{\|\underset{\sim}{x}_p\| \cdot \|\underset{\sim}{x}_2\|} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\langle \underset{\sim}{x}_1, \underset{\sim}{x}_p \rangle}{\|\underset{\sim}{x}_1\| \cdot \|\underset{\sim}{x}_p\|} & \dfrac{\langle \underset{\sim}{x}_2, \underset{\sim}{x}_p \rangle}{\|\underset{\sim}{x}_2\| \cdot \|\underset{\sim}{x}_p\|} & \cdots & \dfrac{\langle \underset{\sim}{x}_p, \underset{\sim}{x}_p \rangle}{\|\underset{\sim}{x}_p\|^2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & r_{21} & \cdots & r_{p1} \\ r_{12} & 1 & \cdots & r_{p2} \\ \cdots & \cdots & \cdots & \cdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix} \qquad (2.2)$$

onde

$$r_{ij} = \frac{<\underset{\sim}{x}_i, \underset{\sim}{x}_j>}{\|\underset{\sim}{x}_i\| \cdot \|\underset{\sim}{x}_j\|} = \frac{\sum_{k=1}^{n} x_{ki}\, x_{kj}}{\sqrt{\sum_{k=1}^{n} x_{ki}^2} \cdot \sqrt{\sum_{k=1}^{n} x_{kj}^2}} \,.$$

O modelo $\underset{\sim}{Y} = X\, \underset{\sim}{\beta}$, ou

$$Y_i = \beta_1\, x_{i1} + \beta_2\, x_{i2} + \ldots + \beta_p\, x_{ip} \tag{2.3}$$

transforma-se (depois de normalizar $\underset{\sim}{Y}$) em:

$$\frac{Y_i}{\|\underset{\sim}{Y}\|} = \beta_1'\, \frac{x_{i1}}{\|\underset{\sim}{x}_1\|} + \beta_2'\, \frac{x_{i2}}{\|\underset{\sim}{x}_2\|} + \ldots + \beta_p'\, \frac{x_{ip}}{\|\underset{\sim}{x}_p\|}$$

ou $$Y_i = \frac{\|\underset{\sim}{Y}\|}{\|\underset{\sim}{x}_1\|}\, \beta_1'\, x_{i1} + \frac{\|\underset{\sim}{Y}\|}{\|\underset{\sim}{x}_2\|}\, \beta_2'\, x_{i2} + \ldots + \frac{\|\underset{\sim}{Y}\|}{\|\underset{\sim}{x}_p\|}\, \beta_p'\, x_{ip} \,. \tag{2.4}$$

Comparando-se (2.3) e (2.4), observa-se que:

$$\beta_i = \frac{\|\underset{\sim}{Y}\|}{\|\underset{\sim}{x}_i\|}\, \beta_i' \,, \qquad\qquad i = 1,\ldots,p \tag{2.5}$$

Desta forma, pode-se usar o modelo (2.4), obtendo-se $\hat{\beta}_i'$ por mínimos quadrados, e os valores de $\hat{\beta}_i$ são obtidos por (2.5).

O cálculo de $\hat{\beta}_i'$ para o modelo (2.4) implica a resolução das "equações normais":

$$(\tilde{X}^t\tilde{X})\ \underset{\sim}{\hat{\beta}}' = \tilde{X}^t\underset{\sim}{\tilde{Y}} \,, \tag{2.6}$$

onde $\underset{\sim}{\tilde{Y}} = \frac{\underset{\sim}{Y}}{\|\underset{\sim}{Y}\|}$. Se em (2.2), $r_{ij} = 0$, $\forall i,j$ $(i \neq j)$, os vetores $\underset{\sim}{x}_i$ e $\underset{\sim}{x}_j$ são ortogonais e a matriz $\tilde{X}^t\tilde{X}$ é igual a I_p (matriz identidade). Neste caso, $\underset{\sim}{\hat{\beta}}'$ pode ser calculado em (2.6) de modo imediato por:

$$\underset{\sim}{\hat{\beta}}' = \tilde{X}^t\, \underset{\sim}{\tilde{Y}} \,.$$

Quando $\tilde{X}^t\tilde{X}$ se "distancia" de I_p, X não é ortogonal, $r_{ij} \neq 0$ para algum $i \neq j$ e o Sistema (2.6) pode apresentar-se instável e sujeito a erros computacionais no cálculo de $\hat{\underset{\sim}{\beta}}'$. Observando-se que r_{ij} é o coeficiente de correlação simples entre $\underset{\sim}{x}_i$ e $\underset{\sim}{x}_j$, neste caso as variáveis x_i e x_j estarão correlacionadas. Caso $1 > r_{ij} >> 0$, a multicolinearidade estará presente.

3. NÚMERO DE CONDIÇÃO DE UMA MATRIZ

Seja o sistema de "equações normais" (2.6) na forma

$$A\,\underset{\sim}{x} = \underset{\sim}{b}\,, \tag{3.1}$$

fazendo-se $A = \tilde{X}^t\tilde{X}$, $\underset{\sim}{x} = \underset{\sim}{\beta}'$, e $\underset{\sim}{b} = \tilde{X}^t\underset{\sim}{\tilde{Y}}$.

Para motivar e caracterizar a definição do número de condição de uma matriz, considere-se a sensibilidade de $\underset{\sim}{x}$ em (3.1) a pequenas perturbações nos elementos de A e $\underset{\sim}{b}$. Com respeito somente a pequenas variações, $\delta\underset{\sim}{b}$ em $\underset{\sim}{b}$, a nova solução $\underset{\sim}{x} + \delta\underset{\sim}{x}$ de (3.1) será definida por

$$A(\underset{\sim}{x} + \delta\underset{\sim}{x}) = \underset{\sim}{b} + \delta\underset{\sim}{b}\,. \tag{3.2}$$

Subtraindo-se (3.1) de (3.2) resulta $A\,\delta\underset{\sim}{x} = \delta\underset{\sim}{b}$, ou

$$\delta\underset{\sim}{x} = A^{-1}\,\delta\underset{\sim}{b}\,. \tag{3.3}$$

Define-se norma da matriz A, e indica-se por $\|A\|$, como uma função de valor real dos elementos da matriz A, satisfazendo as seguintes propriedades:

a) $\|A\| > 0$, com $\|A\| = 0$ somente se $A = 0$;

b) $\|c\,A\| = |c|\cdot\|A\|$ para qualquer escalar c;

c) $\|A+B\| \leq \|A\| + \|B\|$ (desigualdade do triângulo); e

d) $\|A\cdot B\| \leq \|A\|\cdot\|B\|$, [5] .

Então, de (3.3) vem:

$$\|\delta\,\underset{\sim}{x}\| = \|A^{-1}\,\delta\underset{\sim}{b}\| \leq \|A^{-1}\|\cdot\|\delta\underset{\sim}{b}\|\,, \tag{3.4}$$

onde as normas de matriz e de vetor são compatíveis [8]. De (3.1)

$$\|\underset{\sim}{b}\| \leq \|A\| \cdot \|\underset{\sim}{x}\| ,$$

e segue de (3.4) que o erro relativo, $\dfrac{\|\delta\underset{\sim}{x}\|}{\|\underset{\sim}{x}\|}$, na solução de (3.1) satisfaz

$$\frac{\|\delta\underset{\sim}{x}\|}{\|\underset{\sim}{x}\|} \leq \frac{\|A^{-1}\| \cdot \|\delta\underset{\sim}{x}\|}{\|A\|^{-1} \cdot \|\underset{\sim}{b}\|} = \|A\| \cdot \|A^{-1}\| \frac{\|\delta\underset{\sim}{b}\|}{\|\underset{\sim}{b}\|} \tag{3.5}$$

Como $\|A\| \cdot \|A^{-1}\|$ é um indicador da sensibilidade da solução de (3.1) a perturbações, define-se $K(A) = \|A\| \cdot \|A^{-1}\|$ como número de condição da matriz A.

Define-se a norma espectral de uma matriz genérica B, e indica-se por $\|B\|_e$, como:

$$\|B\|_e = \sqrt{\xi_{máx}} ,$$

onde $\xi_{máx}$ é o maior autovalor de B^tB. Como a matriz $A = \tilde{X}^t\tilde{X}$ é simétrica e definida positiva (veja 2.2), seus autovalores são todos reais e positivos e então $\|A\|_e = \lambda_{máx}$, onde $\lambda_{máx}$ é o maior autovalor de $A^tA = A^2$, e é portanto o maior autovalor de A [5].

Define-se o número de condição espectral de uma matriz genérica B, denotado por $K_e(B)$, como:

$$K_e(B) = \|B\|_e \cdot \|B^{-1}\|_e .$$

Para a matriz A,

$$K_e(A) = \|A\|_e \cdot \|A^{-1}\|_e = \frac{\lambda_{máx}}{\lambda_{mín}} , \tag{3.6}$$

onde $\lambda_{mín}$ é o menor autovalor de A, e portanto $1/\lambda_{mín}$ é o maior autovalor de A^{-1}.

Portanto, a Relação (3.6) fornece um número que indica a sensibilidade a perturbações da solução das equações normais. Observa-se que

$$K_e(A) = \|A\|_e \cdot \|A^{-1}\|_e \geq \|A\,A^{-1}\|_e = 1 ,$$

com a igualdade valendo somente no caso de A ser ortogonal. Esta última afirmação estabelece a relação existente entre um problema de regressão com variáveis independentes correlacionadas (isto é, com presença de multicolinearidade) e o número de condição espectral para as equações normais, ou seja, conforme $K_e(A) >> 1$ existe multicolinearidade e instabilidades na obtenção de $\hat{\underset{\sim}{\beta}}'$.

4. TRATAMENTOS DE MODELOS DE REGRESSÃO LINEAR MAL-CONDICIONADOS

Seja o modelo de regressão linear (2.1) e o estimador de mínimos quadrados de $\underset{\sim}{\beta}$,

$$\hat{\underset{\sim}{\beta}} = (X^t X)^{-1} X^t \underset{\sim}{Y} .$$

Define-se o quadrado da distância entre $\underset{\sim}{\beta}$ e $\hat{\underset{\sim}{\beta}}$ por:

$$D^2 = ||\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}||^2 = <\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}, \hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}> = (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta})^T (\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) . \tag{4.1}$$

Então [6]:

$$E(D^2) = \sigma^2 \cdot \text{Traço}(X^t X)^{-1} , \tag{4.2}$$

ou $$E(\hat{\underset{\sim}{\beta}}^t \hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}^t \underset{\sim}{\beta} + \sigma^2 \cdot \text{Traço}(X^t X)^{-1} , \tag{4.3}$$

e, quando $\underset{\sim}{\varepsilon}$ é normalmente distribuído, tem-se:

$$\text{Var}(D^2) = 2\sigma^4 \cdot \text{Traço}(X^t X)^{-2} . \tag{4.4}$$

Como $X^t X$ é simétrica e definida positiva (pois $r(X) = p$), possui p autovalores reais positivos. Sejam:

$$\lambda_{máx} = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p = \lambda_{mín} > 0 ,$$

os p autovalores de $X^t X$, então, de (4.2), o valor médio do quadrado da distância entre $\hat{\underset{\sim}{\beta}}$ e $\underset{\sim}{\beta}$ é dado por:

$$E(D^2) = \sigma^2 \sum_{i=1}^{p} \left(\frac{1}{\lambda_i} \right) , \tag{4.5}$$

e de (4.4) sua variância, quando o erro é normal, será[1]:

$$\text{Var}(D^2) = 2\sigma^4 \sum_{i=1}^{p} \left(\frac{1}{\lambda_i}\right)^2 . \qquad (4.6)$$

Portanto, limites inferiores para a média e variância são, respectivamente:

$$\sigma^2/\lambda_{\text{mín}} \quad \text{e} \quad 2\,\sigma^4/\lambda_{\text{mín}} .$$

Usando-se $\tilde{X}^t\tilde{X}$, verifica-se que quando $0 < \lambda_{\text{mín}} << 1$, $K_e(\tilde{A}) >> 1$ existirá multicolinearidade. Como consequência, verifica-se que a distância entre $\hat{\beta}$ e β tende a ser grande e, de acordo com (4.3),

$$E(\hat{\beta}^t\hat{\beta}) = E(\|\hat{\beta}\|^2) = \|\beta\|^2 + \sigma^2/\lambda_{\text{mín}} , \qquad (4.7)$$

e $\|\hat{\beta}\|$ estará "inflacionada".

Verifica-se, então, que o mal-condicionamento implica um aumento em $\|\hat{\beta}\|$. Quando o valor de r_{ij}, $i \neq j$, é próximo a 1 (um) para as variáveis x_i e x_j, o reflexo em $\hat{\beta}$ pode ser sério, com grandes instabilidades nos coeficientes do Modelo (2.1).

Para contornar o problema da "inflação" e instabilidades associadas ao estimador $\hat{\beta}$, Hoerl e Kennard [1] sugeriram usar outro estimador $\hat{\beta}^*$, dado por:

$$\hat{\beta}^* = [X^tX + KI]^{-1} X^tY \; ; \quad K \geq 0 . \qquad (4.8)$$

Tal estimador recebe o nome de estimador "ridge".

A seguir são apresentadas algumas propriedades do estimador "ridge":

---

[1] *Está sendo admitido sem demonstração que* X^tX *é diagonalizável, e que se* $\Lambda = S^{-1} X^tX\, S$ *é a matriz diagonal semelhante a* X^tX *(onde* S *é a matriz modal), então*

$$\text{Traço } \Lambda = \text{Traço}(X^tX) = \sum_{i=1}^{p} \lambda_i .$$

a) Seja $K \geq 0$ arbitrário, então $\hat{\underset{\sim}{\beta}}^*$ minimiza a soma dos quadrados dos resíduos

$$S(\hat{\underset{\sim}{\beta}}^*) = (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}^*)^t \, (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}}^*)$$

para a esfera centrada na origem e cujo raio ao quadrado é igual a $\hat{\underset{\sim}{\beta}}^{*t}\hat{\underset{\sim}{\beta}}^* = \|\hat{\underset{\sim}{\beta}}^*\|^2$. Além disso, $S(\hat{\underset{\sim}{\beta}}^*)$ é uma função monótona crescente em K e, portanto, a solução "ridge" requer algum acréscimo da soma dos quadrados dos resíduos em relação à soma dos quadrados dos resíduos para o estimador de mínimos quadrados.

b) $\|\hat{\underset{\sim}{\beta}}^*\|$ é uma função monótona decrescente contínua em K, tal que quando $K \to \infty$, $\|\hat{\underset{\sim}{\beta}}^*\| \to 0$, ou seja, o problema da "inflação" estará atenuado para $\hat{\underset{\sim}{\beta}}^*$;

c) O estimador $\hat{\underset{\sim}{\beta}}^*$ é uma transformação linear de $\hat{\underset{\sim}{\beta}}$, e que depende unicamente de X e K. Seja

$$\hat{\underset{\sim}{\beta}}^* = (X^tX + KI)^{-1} \, X^t\underset{\sim}{Y} \; ,$$

mas $X^t\underset{\sim}{Y} = (X^tX)\hat{\underset{\sim}{\beta}}$, então $\hat{\underset{\sim}{\beta}}^* = (X^tX + KI)^{-1} \, (X^tX)\hat{\underset{\sim}{\beta}}$, ou $\hat{\underset{\sim}{\beta}}^* = Z_K \, \hat{\underset{\sim}{\beta}}$. Segue-se, imediatamente, que $\hat{\underset{\sim}{\beta}}^*$ é um estimador viciado de $\underset{\sim}{\beta}$, pois $E(\hat{\underset{\sim}{\beta}}^*) = Z_K \, \underset{\sim}{\beta}$;

d) Se $\underset{\sim}{\beta}^t\underset{\sim}{\beta} = \|\underset{\sim}{\beta}\|^2$ é limitada, então existe um K tal que $E(D^2)$ para $\hat{\underset{\sim}{\beta}}^*$ é menor que $E(D^2)$ para $\hat{\underset{\sim}{\beta}}$;

e) O estimador "ridge" é equivalente a um estimador de mínimos quadrados, quando os dados são suplementados por um conjunto fictício de dados, tomados de um experimento ortogonal H_K, onde os valores correspondentes para $\underset{\sim}{Y}$ serão nulos.

Seja aumentar X por H_K, então as equações normais ficam:

$$(X^t \vdots H_K^t) \begin{bmatrix} X \\ \cdots \\ H_K \end{bmatrix} \hat{\underset{\sim}{\beta}}^* = (X^t \vdots H_K^t) \begin{bmatrix} \underset{\sim}{Y} \\ \cdots \\ \underset{\sim}{0} \end{bmatrix}$$

ou $(X^tX + H_K^t H_K) \, \hat{\underset{\sim}{\beta}}^* = X^t\underset{\sim}{Y}$. Como H_K é ortogonal, $H_K^t H_K$ é um múltiplo escalar de I_p, e para qualquer valor de K, a matriz H_K pode ser tal que $H_K^t H_K = K \cdot I_p$, por exemplo $H_K = K^{1/2} I_p$. Hoerl e Kennard [1] apresentaram uma representação bidimensional, em K e $\hat{\underset{\sim}{\beta}}^*(K)$, denominada traço "ridge" em que, conside rando-se K variando no intervalo $[0,1]$, um certo número (geralmente quinze, para K = 0; 0,02; 0,04; 0,06; 0,08; 0,1; 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8; 0,9; 1) de estimações de $\hat{\beta}^*(K)$ são realizadas. A cada valor de $K > 0$, $K \to 1$, e de acordo com os valores de r_{ij}, $i \neq j$, as instabilidades no valor dos coeficientes são diminuídas e os autores sugerem uma seleção visual para o melhor valor de K. Evidentemente, para K = 0, tem-se $\hat{\beta}^*(0) = \hat{\beta}$, o estimador por mínimos quadrados.

Na Seção 5, apresenta-se um exemplo de traço "ridge", aplicado a uma função demanda por vestuário, em que se usa o critério visual para a seleção do "melhor" estimador "ridge".

Usa-se também, no exemplo, o critério de comparar cada estimativa com a anterior, ou seja

$$|\hat{\beta}_i^*(K_{atual}) - \hat{\beta}_i^*(K_{anterior})| \leq \Delta, \tag{4.9}$$

onde Δ é um valor pequeno prefixado.

5. EXEMPLO

O exemplo apresentado é discutido por Koutsoyiannis [3], págs 240 a 242, para mostrar os efeitos da multicolinearidade nos coeficientes estimados em modelos de regressão múltipla.

Os dados (apresentados na Tabela 5.1), relativos aos anos de 1959 a 1968, incluem gastos com vestuário (G.V.), renda disponível (R.D.), ativo líquido (A.L.), índice de preços para vestuário (I.P.V.) e índice geral de preços (I.G.P.).

O modelo estimado por mínimos quadrados é:

$$(G.V.) = -13.53 + 0{,}097(R.D.) + 0.015(A.L.) - 0{,}199(I.P.V.) + \\ + 0.34(I.G.P.) \ , \tag{5.1}$$

e tem-se ainda que:

$$R^2 = 0{,}998, \quad d = 3{,}4$$

$$\sum \hat{Y}^2 = 28{,}15, \quad \sum e^2 = 0{,}33,$$

$$F^* = \frac{\sum \hat{Y}^2/(K-1)}{\sum e^2/(n-K)} = \frac{28,15/4}{0,33/5} = 15,6, \quad [2] .$$

Como $F_{0,05}$, com $v_1 = K-1 = 5$ e $v_2 = n-K = 5$ graus de liberdade, é 5,19, rejeitou-se a hipótese nula, indicando que existe um relacionamento significativo ($p < 0,05$) entre os gastos com vestuário e as variáveis explanatórias.

TABELA 5.1

DADOS PARA ESTIMAÇÃO DA FUNÇÃO DEMANDA POR VESTUÁRIO

ANO	GASTOS COM VESTUÁRIO (£m) (G.V.)	RENDA DISPONÍVEL (£m) (R.D.)	ATIVO LÍQUIDO (£m) (A.L.)	ÍNDICE DE PREÇOS PARA VESTUÁRIO (1963 = 100) (I.P.V.)	ÍNDICE GERAL DE PREÇOS (1963=100) (I.G.P.)
1959	8,4	82,9	17,1	92	94
1960	9,6	88,0	21,3	93	96
1961	10,4	99,9	25,1	96	97
1962	11,4	105,3	29,0	94	97
1963	12,2	117,7	34,0	100	100
1964	14,2	131,0	40,0	101	101
1965	15,8	148,2	44,0	105	104
1966	17,9	161,8	49,0	112	109
1967	19,3	174,2	51,0	112	111
1968	20,8	184,7	53,0	112	111

Entretanto, as variáveis explanatórias são seriamente multicolineares, como pode ser visto na matriz de correlações abaixo:

	R.D.	A.L.	I.P.V.	I.G.P.
R.D.	1.	0.988315	0.980356	0.987666
A.L.	0.988315	1.	0.969962	0.969477
I.P.V.	0.980356	0.969962	1.	0.991796
I.G.P.	0.987666	0.969477	0.991796	1.

e, desta forma, o autor usa o "critério R^2", descrito por Neter e Wasserman [4], pág. 376, e também considera a significância dos coeficientes estima dos através de seus desvios padrões, concluindo pela escolha do modelo G.V. = f(R.D., I.P.V., I.G.P.) como o melhor (veja a Tabela 5.2), excluindo, portanto, a variável A.L. do Modelo (5.1) por não apresentar coeficiente significativo.

O professor Koutsoyiannis usa considerações a respeito do acréscimo em R^2, quando são incluídas variáveis independentes até compor o Modelo (5.1), e verifica, para cada relação, a significância dos coeficientes de regressão.

TABELA 5.2

RESULTADOS DAS REGRESSÕES EFETUADAS POR KOUTSOYIANNIS [3]

	$\hat{\beta}_0$	$\hat{\beta}_1$ (R.D.)	$\hat{\beta}_2$ (A.L.)	$\hat{\beta}_3$ (I.P.V.)	$\hat{\beta}_4$ (I.G.P.)	R^2
G.V.=f(R.D.)	-1,24 (0,37)	0,118 (0,002)	-	-	-	0,995
G.V.=f(R.D.,I.P.V.)	1,40 (4,92)	0,126 (0,01)	-	-0,036 (0,07)	-	0,996
G.V.=f(R.D.,I.P.V., A.L.)	0,94 (5,17)	0,138 (0,02)	-0,037 (0,05)	-0,034 (0,06)	-	0,996
G.V.=f(R.D.,I.P.V., I.G.P.)	-12,76 (6,52)	0,104 (0,01)	-	-0,188 (0,07)	0,319 (0,12)	0,997
G.V.=f(R.D.,I.P.V., A.L.,I.G.P.)	-13,53 (7,5)	0,097 (0,03)	0,015 (0,05)	-0,199 (0,09)	0,34 (0,15)	0,998

Nota: Os números entre parênteses são os desvios padrões dos coeficientes estimados.

Com a finalidade de mostrar as instabilidades estimativas dos coeficientes do Modelo (5.1), usaram-se, neste trabalho, os dados da Tabela 5.1 como entrada em um programa de computador para determinação dos estimadores "ridge" dos parâmetros (coeficientes) da função demanda por vestuário. Os resultados estão resumidos abaixo:

a) Para K = 0 em (4.8) (estimador de mínimos quadrados) ou (5.1), tem-se

$$\left.\begin{array}{l}\lambda_{mín} = 0,0036 \\ \lambda_{máx} = 3,9438\end{array}\right\} \Rightarrow K_e(\tilde{A}) = \frac{\lambda_{máx}}{\lambda_{mín}} = 1084,3228 >>> 1 ,$$

ou seja, o problema é extremamente mal-condicionado (existe multicolinearidade) e, portanto, a solução será extremamente instável;

b) A Figura 5.1 mostra o traço "ridge" para o exemplo. Conforme K → 1 as instabilidades diminuem, sendo que visualmente, para K = 0,16 as instabilidades estão praticamente eliminadas. Neste caso, pode-se usar o seguinte modelo estimado

$$(G.V.) = -20,2186 + 0,0340(R.D.) + 0,0802(A.L.) + 0,0967(I.P.V.) + \\ + 0,1674(I.G.P.) \qquad (5.2)$$

como alternativo ao Modelo (5.1). Observa-se pela Tabela 5.1 que as variáveis I.P.V. e I.G.P. apresentam dados muito próximos, altamente correlacionados, evidenciando que numericamente deveriam ser uma única variável. A mudança de sinal do coeficiente da variável I.P.V. é justificada pela relação quase direta entre ela e a variável dependente (G.V.) (veja a Tabela 5.1).
A variável R.D. é a mais estável conforme mostra a Figura 5.1, embora ligeiramente superestimada para K = 0 (solução de mínimos quadrados), enquanto a variável A.L. apresenta-se subestimada neste caso.

c) Considerando-se o critério (4.9) para Δ = 0,01, seleciona-se como estimador "ridge":

$$(G.V.) = -18,5820 + 0,0466(R.D.) + 0,0689(A.L.) - 0,0107(I.P.V.) + \\ + 0,2200(I.G.P.). \qquad (5.3)$$

Neste caso, o coeficiente de I.P.V. é muito próximo a zero, sendo esta variável uma séria candidata a ser excluída do modelo. A forma correta de se fazer tal exclusão, seria considerar em cada predição de G.V. o valor médio de I.P.V., que é equivalente a desconsiderar o efeito desta variável na predição de G.V..

A Figura 5.2 mostra o traço "ridge" para o caso em que foi eliminada a variável A.L., conforme sugere o professor Koutsoyiannis. Nota-se que as instabilidades continuam, evidenciando que o problema ainda é mal-condicionado.

d) Os números de condição para os Modelos (5.2) e (5.3) foram 25,07 e 62,91, respectivamente, evidenciando a melhor estabilidade destas soluções.

6. CONCLUSÕES

Devido ao estimador "ridge" ser um estimador viciado, suas propriedades estatísticas são desconhecidas e não se deve pensar em critérios tais como R^2, para comparar o estimador "ridge" e o de mínimos quadrados. Hoerl e Kennard [1] mostraram que existe $K > 0$ inversamente proporcional aos coeficientes estimados de regressão (para $K = 0$), tal que o Erro Médio Quadrático para o estimador "ridge" é menor que para o estimador por mínimos quadrados. A determinação de tal valor de K tem preocupado pesquidadores (veja o trabalho de Wichern e Churchill [7] e suas referências bibliográficas). Contudo, no aspecto de condicionamento das equações normais, a solução "ridge" é no mínimo mais estável que a de mínimos quadrados.

Deve-se observar que para o exemplo na Seção 5, os resultados obtidos pela análise "ridge" refletem os dados apresentados na Tabela 5.1, não sendo considerados, neste trabalho, os aspectos de teoria econômica referente ao sinal dos coeficientes de regressão, possivelmente explicáveis à luz dos dados disponíveis.

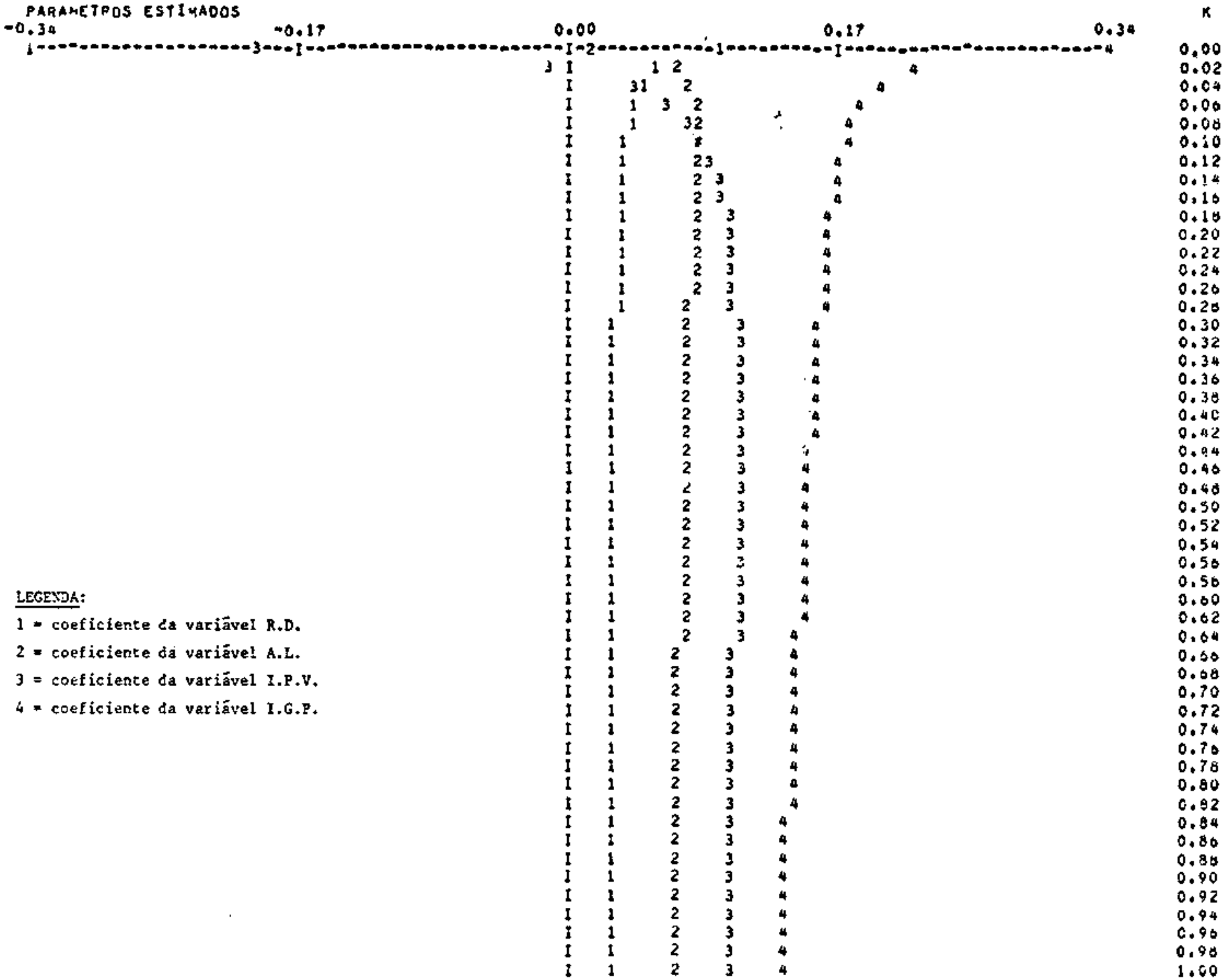

Fig. 5.1 - Traço "ridge" do modelo completo.

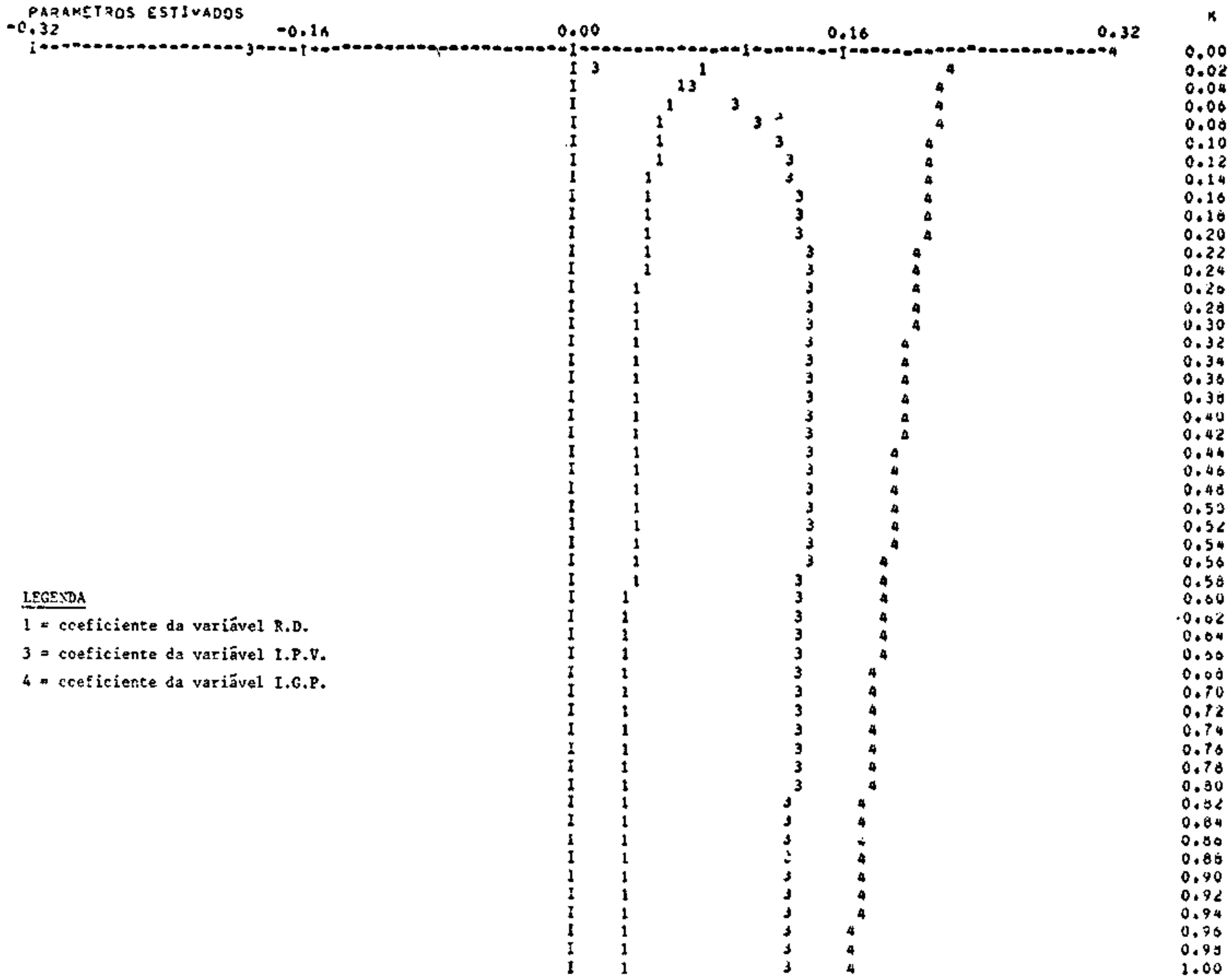

Fig. 5.2 - Traço "ridge" do modelo em três variáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] HOERL, A.E.; KENNARD, R.W. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1): 55-67, Feb. 1970.

[2] JOHNSTON, J. *Econometrics methods*. New York, NY, McGraw-Hill, 1963.

[3] KOUTSOYIANNIS, A. *Theory of econometrics*. 2 ed. London, UK, The Macmillan, 1981.

[4] NETER, J.; WASSERMAN, W. *Applied linear statistical models*. Homewood, IL, Richard D. Irwin, 1974.

[5] SCHWARZ, H.R.; RUTISHAUSER, H.; STIEFEL, E. *Numerical analysis of symmetric matrices*. Englewood Cliffs, NJ, Prentice-Hall, 1973.

[6] SEARLE, S.R. *Linear models*. New York, NY, John Wiley, 1971.

[7] WICHERN, D.W.; CHURCHILL, G.A. A comparison of ridge estimators. *Technometrics*, 20(3): 301-311, Aug. 1978.

[8] WILKINSON, J.H. *The algebraic engenvalue problem*. Oxford, UK, Clarendon, 1965.