# PROPOSAL OF A RAO RIDGE TYPE ESTIMATOR

Luzia Aparecida da COSTA[1]
Lucas Monteiro CHAVES[2]
Devanil Jaques de SOUZA[3]

- *ABSTRACT: Based on a geometrical interpretation of Ridge estimators a new Rao Ridge type estimator is proposed. Its advantage is to reach the optimum value for the shrinkage parameter more quickly. The geometry, the predictive capacity, a computational example, an application to real data and comparison with the usual Ridge estimator are developed.*

- *KEYWORDS: Hoerl-Kennard Ridge estimator, geometrical interpretation, optimum Ridge parameter.*

## 1 Introdution

Ridge regression was proposed in a seminal paper by Hoerl and Kennard (1970a). It is an option to circumvent the instability in regression estimators obtained by least squares method in presence of multicolinearity. The Ridge regression coefficients are estimated by a very simple formula $\hat{\boldsymbol{\beta}}(k) = (X'X + kI)^{-1}X'\mathbf{y}$, where k is a shrinkage parameter. The estimator is biased and its mean squared error depends on the shrinkage coefficient and achieve a minimum on a value $k_{opt}$. This value depends on populational parameters and a hard work is required in order to obtain a good estimators. One advantage of the Ridge regression method is the possibility to construct a two dimensional portrait of the behavior of each estimator component, as $k$ increases, by plotting the graphics of

[1]Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - IFET, Campus Formiga, CEP: 35570-000, Formiga, MG, Brasil, E-mail: *luziamatematica@hotmail.com*
[2]Universidade Federal de Lavras - UFLA, Departamento de Ciências Exatas, CEP: 37200-000, Lavras, MG, Brasil, E-mail: *lucas@dex.ufla.br*
[3]Universidade Federal de Lavras - UFLA, Departamento de Estatística, CEP: 37200-000, Lavras, MG, Brasil, E-mail: *devaniljaques@des.ufla.br*

Ridge trace functions $\hat{\beta}_i(k)$, $i$ = 1,2,...,$n$. All these curves goes to zero, but near the optimum value $k_{opt}$ they are relatively stable. Therefore this can be used as a graphical criterion to estimate the optimum value. The Ridge estimators was generalized by Rao (1975), that suggested to use a positive definite matrix G in place of the identity matrix, obtaining the estimator $\hat{\boldsymbol{\beta}}(k) = (X'X + kG)^{-1}X'\mathbf{y}$ (GRUBER, 1998, 2010). This generalization of Ridge estimator showed to be very useful and, in particular, allowed an empirical Bayes approach. In this way, a priori knowledge can be used to choose the matriz $G$. We will call these estimators as Rao Ridge type estimators. The geometry of Rao Ridge type estimators are present in Costa (2014).

In this work, a particular Rao Ridge type estimator, that is, a choice for the matrix $G$, based on geometrical arguments, is proposed. The ideia is to choose a matrix $G$ in such way that the optimum value of the shrinkage factor in achieve more quickly. Some of its properties are explicit and compared with the usual Ridge estimator. As this estimator was derived by geometrical motivation we will rewrite some facts about linear model theory in terms of geometrical concepts.

## 2 The linear model

Consider $\mathbf{Y}$ a $n \times 1$ vector of values obtained from some random phenomena with unknown mean vector $\boldsymbol{\mu} = E[\mathbf{Y}]$. By reasons related to these particular random phenomena, one can conjecture that the unknown mean vector $\boldsymbol{\mu}$ belongs to some known subspace $\mathbf{W}$. That, in essence, characterizes the linear model. The vector $\mathbf{Y}$ stands in the data space and since the dimension of data space is generally higher than the dimension of $\mathbf{W}$, it is plausible to use a lower than $n$ number of variables to describe the $\mathbf{W}$ subspace. Such procedure is called parameterization and can be done in the following way: consider $\mathbf{W}$ to be the image of a linear transformation X, defined in another vector space called parametric space. Then $\mathbf{W}$=Im(X). To avoid technical difficulties, the linear transformation X is considered injective, that is, for each vector $\mathbf{w}$ in $\mathbf{W}$, there is a unique vector $\boldsymbol{\beta}$ in the parameter space such that $\mathbf{w}$=X$\boldsymbol{\beta}$. All of this can be described geometrically by Figure 1.

Model assumptions are clear: $\mathbf{Y}$ is a random vector in the data space, $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, E[$\mathbf{Y}$]=X$\boldsymbol{\beta}$ a vector in $\mathbf{W}$, where $\boldsymbol{\beta}$ is an unknown vector in the parametric space and $\boldsymbol{\varepsilon}$ is the vector of errors. The greatest advantage of representing the linear model by Figure 1 is the description of the estimation process. What is the estimation process? A data vector $\mathbf{y}$ is observed and we have only to choose a vector in $\mathbf{W}$ which we believe to be a good representative of E[$\mathbf{Y}$]. If, eventually, $\mathbf{y}$ happens to belong to the space $\mathbf{W}$, the best choice is $E[\mathbf{Y}] = \mathbf{y}$. But, as the observed vector $\mathbf{y}$ is affected by random errors, almost surely, it will not belong to the subspace $\mathbf{W}$ and a natural procedure to choose the vector in $\mathbf{W}$ closest to $\mathbf{y}$. Such estimation procedure is called *least squares method*. If $P_{\mathbf{W}}$ is the linear orthogonal projection onto $\mathbf{W}$, the chosen vector is $P_{\mathbf{W}}(\mathbf{y})$. The linear transformation X is injective, then the only choice is $P_{\mathbf{W}}(\mathbf{y})$=X$\hat{\boldsymbol{\beta}}$ . The linear transformation $P_W$ can be expressed in matrixial form as $P_W = X(X'X)^{-1}X'$. This equation, in its algebraic form, is

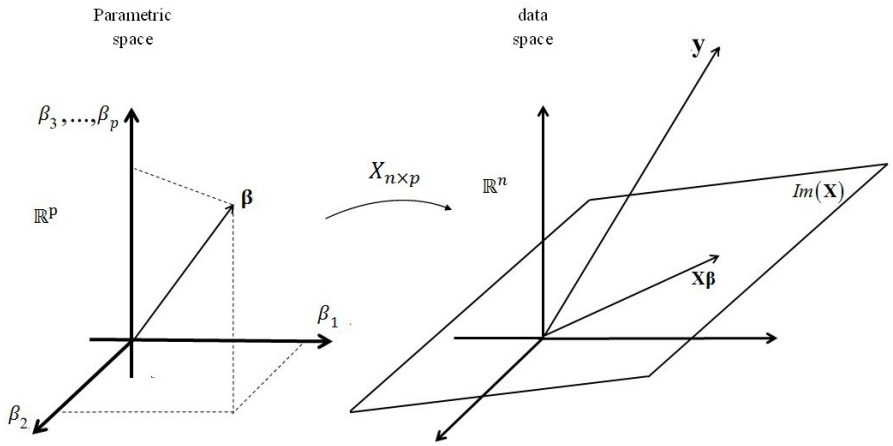


Figure 1 - Geometrical characterization of a linear model.

denominated *normal equation* and $\hat{\boldsymbol{\beta}}$ is the least squares estimator of the parameter vector $\boldsymbol{\beta}$.

## 3 The Rao ridge type estimators

This section is strongly based in Costa (2014). We consider a weighted distance in parametric space defined by a positive definite matrix $G$. The norm is $\|\boldsymbol{\beta}\|^2 = \boldsymbol{\beta}'G\boldsymbol{\beta}$ and distance $d(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2) = (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)'G(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$. The sphere of radius $r$ centered in the origin in this new metric is given by $\boldsymbol{\beta}'G\boldsymbol{\beta} = r^2$ and defines an ellipsoid centered in the origin. As the usual Ridge estimators, we can derive the Rao Ridge type estimators as a solution of a minimization problem involving that weighted distance.

If we want an alternative to the ordinary least squares estimator $\hat{\boldsymbol{\beta}}$, a possibility is to consider a sphere of radius r centered in the orthogonal projection of the observed vector $\mathbf{y}$ in the subspace $\mathbf{W}$. This procedure can be considered as a constrained least squares, or in others words, a penalized least squares method. The pre-image of this sphere in the parametric space is an ellipsoid centered in $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned} \|y - P_W(y)\|^2 &= \|X\boldsymbol{\beta} - P_W(y)\|^2 \\ &= \|X\boldsymbol{\beta} - X\hat{\boldsymbol{\beta}}\|^2 \\ &= (X\boldsymbol{\beta} - X\hat{\boldsymbol{\beta}})'(X\boldsymbol{\beta} - X\hat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ &= r^2. \end{aligned} \tag{1}$$

Now, the idea is to choose, among all $\beta$'s in the ellipsoid, the one with minimum weighted norm. Therefore, the Rao Ridge type estimator in the parametric space is obtained as the solution of the minimization problem:

$$\hat{\beta}\,(r) = \min\left\{\beta' G \beta\right\}, \tag{2}$$

subject to restriction $(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = r^2$.

Using Lagrange multiplier, we can get the explicit $\widehat{\beta_R}\,(k) = \left(X^{'}X + nkG\right)^{-1} X^{'}\mathbf{y}$, where $k$ is a more adequate parameter and is a function of $r$.

This construction can be geometrically described by Figure 2.

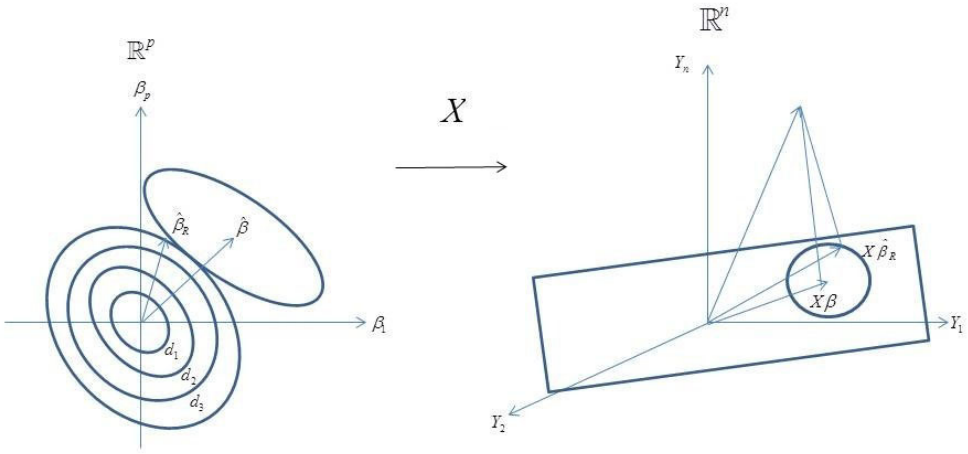


Figure 2 - Geometrical interpretation of Rao Ridge type estimator.

### 3.1 The proposed Rao ridge type estimator

Maybe the simplest example of a Rao Ridge type estimator is obtained by considering $G = X'X$. In this case, the estimator is $\hat{\boldsymbol{\beta}}_R(k) = (X'X + kX'X)^{-1}X'\mathbf{y} = \dfrac{1}{1+k}(X'X)^{-1}X'\mathbf{Y} = \dfrac{1}{1+k}\hat{\boldsymbol{\beta}}$, that is, the estimator is a shrinkage of the ordinary least squares estimator. In this case, there are two parallel ellipsoids and the estimator is obtained by taking the point of tangency of these two ellipsoids, as described in Figure 3. This estimator is known as Mayer-Wilke estimator (MAYER and WILKE, 1973).

How to make a good choice for the matrix G? One the greatest problem in Ridge estimation is to obtain the optimum value of $k$. One of the difficulties is that the mean squared error function can be very flat near the minimum and therefore any numerical procedure is very unstable. This is equivalent to say that the Ridge trace functions goes to zero very slowly. As observed with Mayer-Wilke estimator,

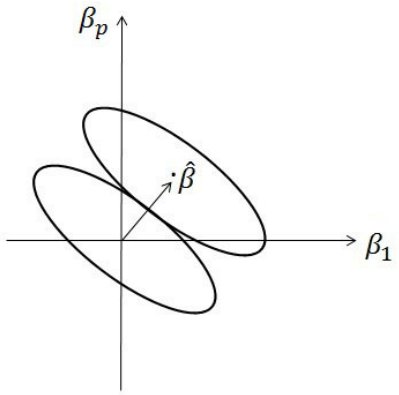


Figure 3 - Paralled ellipsoids and Mayer-Wilke estimator.

for parallel ellipsoid, the velocity goes to zero proportional to the inverse of the shrinkage parameter. The idea is to use some kind of matrix $G$ that increases that velocity. The geometry suggests that the ellipsoid defined by $G$ must be, in some sense, orthogonal to the ellipsoid defined by $X'X$. To make this, we need another ellipsoid with the same principal axes as the ellipsoid defined by $X'X$ but with eigenvalues with inverse values. In this case, major axis of the ellipsoid defined by $X'X$ corresponds the minor axis of the new ellipsoid (Figure 4). We will call these ellipsoids as orthogonal.

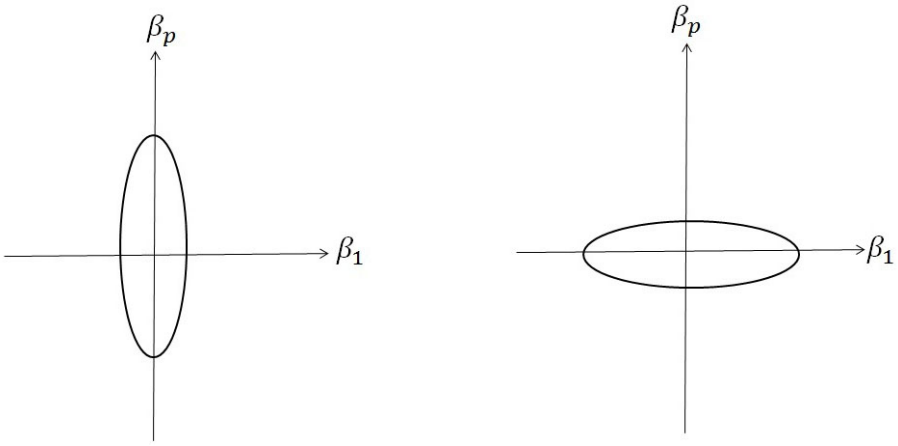


Figure 4 - Orthogonal ellipsoids.

The matrix with eigenvalues equal to the inverse of the eigenvalues of $X'X$ and the same principal axes is the inverse $(X'X)^{-1}$. So we suggested to make $G = (X'X)^{-1}$ and propose the following Rao Ridge type estimator:

$$\hat{\boldsymbol{\beta}}_{prop}(k) = (X'X + nk(X'X)^{-1})^{-1}X'\mathbf{Y}. \tag{3}$$

Geometrically, the proposed estimator is describe in Figure 5.

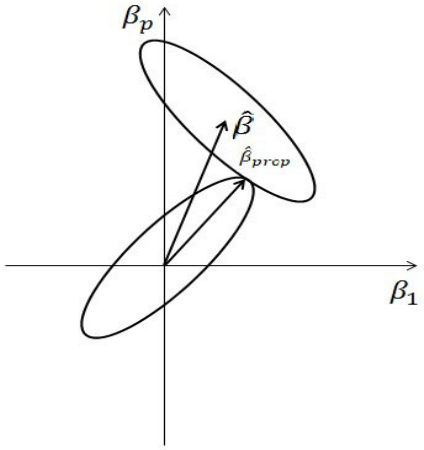


Figure 5 - Geometry of the proposed estimator.

The performance of the proposed estimator will be studied and compared with the usual Ridge estimator.

### 3.2 The predictive performance of the proposed estimator

One of the most important features of an estimator is its predictive capability. One way to try to access this property is to use the Allen's Predictive error of squares (PRESS), that is essentially ordinary cross validation. The process is describe as: let $X_i$ be the matrix $X$ with the ith line omitted and $\mathbf{y}_i = (y_1, \ldots, \widehat{y_i}, \ldots, y_n)'$ the vector of data with ith data value omitted. Then the $\boldsymbol{\beta}^{(i)}(k)$, the Ridge estimate of $\boldsymbol{\beta}$, is obtained as

$$\boldsymbol{\beta}^{(i)}(k) = \left(X_i'X_i + nkI\right)^{-1}X_i'\mathbf{y}_i. \tag{4}$$

The argument is that if $k$ is a good choice for the Ridge parameter, then the ith component $\left[X\boldsymbol{\beta}^{(i)}(k)\right]_i$ should be a good predictor of $y_i$. Therefore, the Allen's PRESS estimate of $k$ is the minimizer of

$$P(k) = \frac{1}{n}\sum_{i=1}^{n}\left(\left[X\boldsymbol{\beta}^{(i)}(k)\right]_i - y_i\right)^2. \tag{5}$$

It is interesting to point out that, for the usual Ridge estimator ($G = I$), the PRESS has a closed form given:

$$P(k) = \frac{1}{n}\|B(k)(I-A)\mathbf{y}\|^2, \tag{6}$$

where $B(k)$ is the diagonal matrix with jjth entry $1/(1-a_{jj}(k))$, $a_{jj}(k)$ being the $jj$th entry of $A(k) = X(X'X+nkI)^{-1}X$ (CHRISTENSEN, 2011).

Golub *et al.* (1979) defined an invariant version of the Allen's PRESS, using the very involving theory of circulant matrices (GELLER *et al.*, 2017) and called it generalized cross-validation method (GCV). The predictive error using GCV is given by:

$$V(k) = \frac{\frac{1}{n}\left\|\left(I-\widetilde{A}(\mathbf{y})\right)\widetilde{\mathbf{y}}\right\|^2}{\left[\frac{1}{n}Tr\left(I-\widetilde{A}(k)\right)\right]^2} \equiv \frac{\frac{1}{n}\sum_{\nu=1}^{n}\left(\frac{nk}{k_{\nu n}+nk}\right)^2 z_\nu^2}{\left[\frac{1}{n}\sum_{\nu=1}^{p}\frac{nk}{k_{\nu n}+nk}+n-p\right]^2}. \tag{7}$$

They proved that this generalization is a simple weighted version of $P(k)$, namely

$$V(k) \equiv \frac{1}{n}\sum_{i=1}^{n}\left(\left[X\boldsymbol{\beta}^{(i)}(k)\right]_i - y_i\right)^2 w_i(k), \tag{8}$$

where

$$w_i(k) = \frac{1-a_{ii}(k)}{1-\frac{1}{n}\mathrm{Tr}A(k)}. \tag{9}$$

For the proposed estimator, the PRESS is obtained in a similar way as follow:

$$\hat{\boldsymbol{\beta}}_{prop}(k) = (X'X+k(X'X)^{-1})^{-1}X'\mathbf{y}. \tag{10}$$

Let $X_i$ and $y_i$ be as before, then

$$\hat{\boldsymbol{\beta}}^{i}_{prop}(k) = (X_i'X_i+k(X_i'X_i)^{-1})^{-1}X_i'\mathbf{y}. \tag{11}$$

The PRESS statistics is defined by

$$P(k) = \frac{1}{n}\sum_{i=1}^{n}\left(\left[X\widehat{\boldsymbol{\beta}}^{i}_{prop}(k)\right]_i - y_i\right)^2. \tag{12}$$

At this point, the authors want to point out that they couldn't find in the literature a closed formula for the PRESS and for GCV, in the case of Rao Ridge type estimator.

In order to compare the proposed estimator and usual Ridge estimator, a simulation with 8 predictors and 12 observations were used. To ensure multicollinearity, we took two pairs of predictors highly correlated. One thousand samples were generated and, for each sample and each of the 51 values of $k$, from 0.00 to 0.12, we computed the mean values of the errors, getting the mean curve of

the prediction errors. For the usual Ridge estimator, the curves were obtained using usual cross validation and generalized cross validation. For the proposed Rao Ridge type estimator, error was computed using only usual cross validation. Results are shown in Figure 6 and in Table 1.

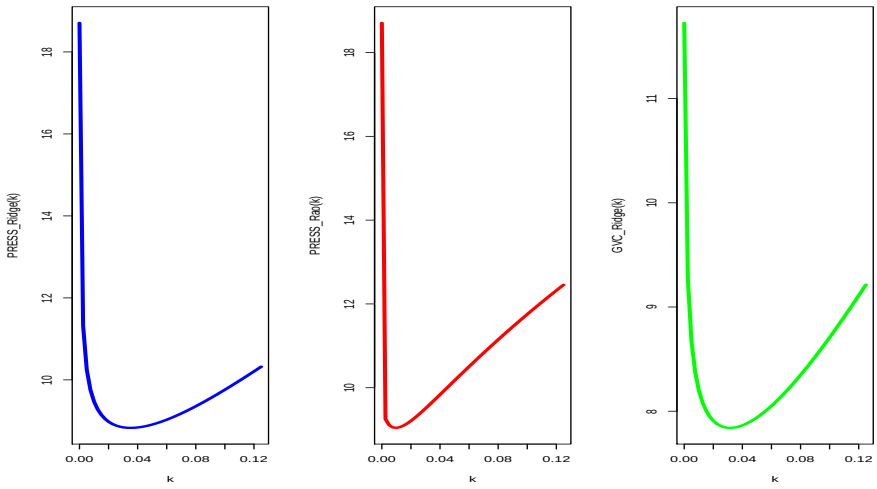


Figure 6 - Mean curves of the predictors errors.

Table 1 - Optimal value of $k$ and mean predictor errors of PRESS Ridge, PRESS Rao and GCV Ridge

| Estimator | $k_{opt}$ | Prediction Error |
|---|---|---|
| PRESS Ridge | 0.04 | 9.16 |
| PRESS Rao | 0.01 | 9.46 |
| GCV Ridge | 0.03 | 8.18 |

The optimum shrinkage parameter is achieve at the minimum of the curves in Figure 6. As expected, the proposed estimator reached the optimal value of $k$ about 3 times faster than the usual Ridge. There is no significant difference among the mean predictor errors.

### 3.3 A computational illustrative example

The example to be describe is as small as possible while still permitting to evaluate the performance of the proposed estimator. This example was analyzed in Marquardt (1970), to compare the usual Ridge estimator with others estimators and it was also studied in the context of Rao Ridge type estimator in Costa (2014).

Consider the linear Ridge regression $Y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with

$$Y = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},\ X = \begin{pmatrix} \frac{3\sqrt{2}}{10} & \frac{4\sqrt{2}}{10} \\ \frac{4\sqrt{2}}{10} & \frac{3\sqrt{2}}{10} \\ \frac{5\sqrt{2}}{10} & \frac{5\sqrt{2}}{10} \end{pmatrix},\ X'X = \begin{pmatrix} 1 & \frac{49}{50} \\ \frac{49}{50} & 1 \end{pmatrix}.$$

The eigenvalues of $X'X$ are 1.98 and 0.02, wich characterize almost multicolinearity.

Supposed the real value of the parameter vector is $\boldsymbol{\beta} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$ and $\sigma^2 = 1$. The response vector obtained by simulation was $\mathbf{y} = \begin{pmatrix} 6.34 \\ 3.94 \\ 5.96 \end{pmatrix}$. With this vector, the Ridge trace are obtained for usual Ridge estimator and for proposed estimator (Figures 7 and 8).

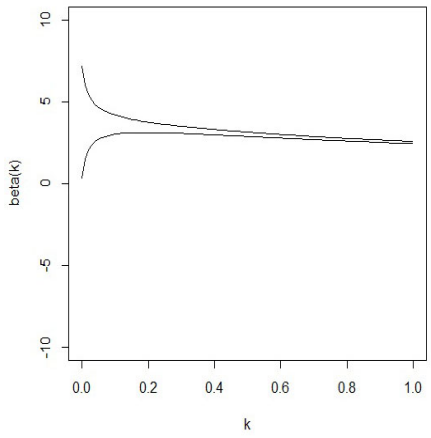


Figure 7 - Ridge trace for usual Ridge estimator.

Using the graphical criteria to obtain the optimum value $k_{opt}$ that occurs when the Ridge trace present certain stability, we get $k_{opt} = 0.30$ for the usual Ridge estimator (FIGURE 7) and $k_{opt} = 0.05$ for the proposed Ridge estimator (FIGURE 8). This example shows, as expected, that the optimum value of the parameter $k$ occurs much more faster for the proposed estimator. With these $k_{opt}$, it can be obtained, from the same Figures 7 and 8, the vectors estimates $\hat{\boldsymbol{\beta}}\,(0.30)$ and $\hat{\boldsymbol{\beta}}_{\text{prop}}\,(0.05)$. For comparison, the least square estimative $\hat{\boldsymbol{\beta}}_{QM}$ is also presented.

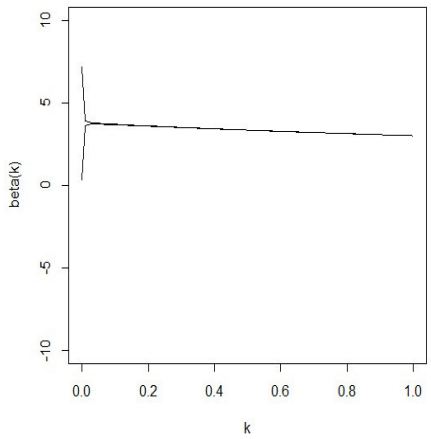


Figure 8 - Ridge trace for the proposed estimator.

By simulation of one thousand response vectors $\mathbf{y}$, the Mean Squared Error (MSE) was presented in the Table 2.

Table 2 - Mean Squared Error (MSE)

| $\hat{\boldsymbol{\beta}}$ | MSE | = | Variance | + | Bias |
|---|---|---|---|---|---|
| $\hat{\boldsymbol{\beta}}_{QM}$ | 50.50 | = | 50.50 | + | 0.00 |
| $\hat{\boldsymbol{\beta}}(0.30)$ | 2.78 | = | 0.59 | + | 2.18 |
| $\hat{\boldsymbol{\beta}}_{prop}(0.05)$ | 2.47 | = | 0.97 | + | 1.51 |

In the Table 2 is possible observe that the MSE for both estimators are closed, as the Prediction Error, seen in the previous section.

### 3.4 The performance of proposed estimator on a set of real data

Hoerl and Kennard (1970b) analyzed a set of real data in Gorman and Toman (1966), as a linear regression on ten factors, using the Ridge estimator. The sample correlation matrix is:

$$X'X = \begin{pmatrix} 1.00 & & & & & & & & & \\ -0.04 & 1.00 & & & & & & & & \\ 0.51 & 0.00 & 1.00 & & & & & & & \\ 0.12 & -0.16 & 0.00 & 1.00 & & & & & & \\ -0.71 & 0.06 & -0.59 & -0.07 & 1.00 & & & & & \\ -0.87 & 0.09 & -0.65 & -0.09 & 0.84 & 1.00 & & & & \\ -0.09 & 0.24 & -0.02 & 0.03 & 0.38 & 0.13 & 1.00 & & & \\ 0.00 & 0.01 & 0.34 & 0.08 & -0.36 & -0.20 & -0.48 & 1.00 & & \\ -0.09 & 0.09 & -0.08 & 0.02 & -0.14 & 0.04 & 0.07 & -0.18 & 1.00 & \\ -0.36 & -0.30 & -0.44 & -0.09 & 0.54 & 0.45 & 0.40 & -0.46 & 0.05 & 1.00 \end{pmatrix}.$$

$$X'Y = (-0.81 \quad -0.10 \quad -0.63 \quad -0.10 \quad 0.56 \quad 0.81 \quad 0.04 \quad 0.06 \quad 0.16 \quad 0.45).$$

The eigenvalues de $X'X$ are

$$(3.69 \quad 1.53 \quad 1.29 \quad 1.05 \quad 0.97 \quad 0.66 \quad 0.35 \quad 0.21 \quad 0.13 \quad 0.07),$$

and again we have almost multicolinearity. The total variance of the model is $33.82\sigma^2$. For the Ridge trace, the covariates were indexed as the numbers in their respective column in the correlation matrix:

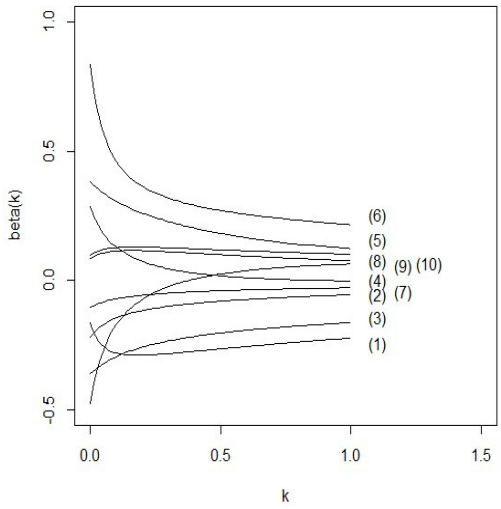


Figure 9 - Ridge trace of the usual ridge estimator.

Analysing the Ridge trace curves for the both estimators, we conclude:

- The least squares estimatives seems to be overestimated because they decrease rapidly in function of $k$. This behavior is more emphasized for factors 5 and 6.
- The factor 5, initially negative, rapidly becomes positive. This is coherent with the fact that factor 5 has sample correlation 0.84 with factor 6 and therefore we expected that they have the same behavior.

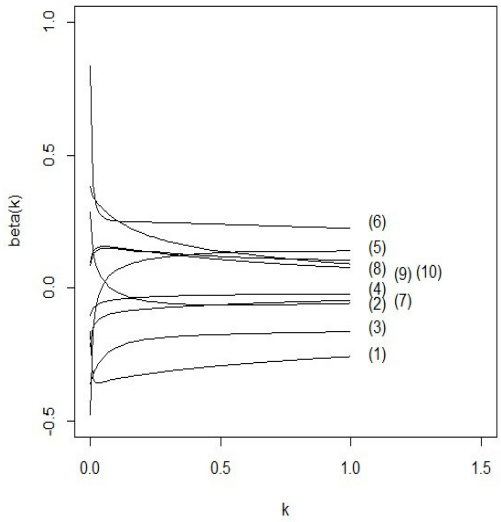


Figure 10 - Ridge trace of proposed estimator.

- Factor 1 is underestimated maybe because the correlation with the others factors.
- Factor 7 is overestimated.
- For the usual ridge estimator the system stabilized for $k$ in the interval $[0.20; 0.30]$. For the proposed estimator the system stabilized in the interval $[0.05; 0.15]$.

Both estimators give the same conclusions, but these conclusions are much ellipse more clear for the proposed estimator. As expected, the Ridge trace stabilized more rapidly for the proposed estimator.

## 4 Conclusion

The proposed estimator that was based in geometrical ideas, seems to have superior properties in relation to the usual Hoerl-Kennard Ridge estimator. One of its promising advantage is to get stability more rapidly and therefore the optimum value for the Ridge parameter $k$ can be estimated more accurately.

## Acknowledgments

We would like to thank reviewers and editors for their suggestions.

COSTA, L. A.; CHAVES, L. M.; SOUZA, D. J. Proposta de um estimador do tipo Rao Ridge. *Rev. Bras. Biom.,* Lavras, v.36, n.3, p.686-699, 2018.

- RESUMO: Com base em uma interpretação geométrica dos estimadores de Ridge, um novo estimador do tipo Rao Ridge é proposto. Sua vantagem é alcançar o valor ótimo para o parâmetro de encolhimento mais rapidamente. A geometria, a capacidade preditiva, um exemplo computacional, uma aplicação a dados reais e uma comparação com o estimador Ridge usual são desenvolvidas.

- PALAVRAS-CHAVE: Estimador de Hoerl-Kennard; interpretação geométrica; valor ótimo do parâmetro Ridge.

## References

CHRISTENSEN, R. *Plane answers to complex questions: The theory of linear models.* 4.ed. New York: Springer, 2011.

COSTA, L. A.; CHAVES, L. M.; SOUZA, D. J. Uma abordagem geométrica do estimador de cumeeira de C.R. Rao. *Revista Brasileira de Biometria*, v.32, n.1, p.28-41, 2014.

GELLER, D.; KRA, I.; POPESCU, S.; SIMANCA, S. *On Circulant matrices.* SD. Available in: `<http://www.math.stonybrook.edu/~sorin/eprints/circulant.pdf>`. Access in: 21 mar. 2017.

GOLUB, G. H.; HEATH, M.; WAHBA, G. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, v.21, n.2, p.215-223, 1979.

GORMAN, J. W.; TOMAN, R. J. Selection of variables for fitting equations to data. *Technometrics*, n.8, p.27-51, 1966.

GRUBER, M. H. J.; *Improving efficiency by shrinkage - The James-Stein and Ridge regression estimators.* Marcel Dekker: New york, 1998. 648p.

GRUBER, M. H. J.; *Regression estimators: a comparative Study.* 2.ed. Baltimore: Johns Hopkins, 2010. 412p.

HOERL, A. E.; KENNARD, R. W. Ridge regression: aplications to nonorthogonal problems. *Technometrics*, v.12, n.1, p. 69-82, 1970a.

HOERL, A. E., KENNARD, R.; Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, v.12, n 1, p.55-67, 1970b.

MARQUARDT, D. W.; Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, *Technometrics*, v.12, n.3, p.591-612, 1970.

MAYER, L. S.; WILKE, T. A. On biased estimation in linear models. *Technometrics*, n.15, p.497-508, 1973.

RAO, C. R.; Simultaneous estimation of parameters in different linear models and applications to biometric problems. *Biometrics*, v.31, n.2, p.545-554, 1975.

Received in 14.07.2017.

Approved after revised in 21.03.2017.