

2m11.2577.3

Université de Montréal

QUELQUES ESTIMATEURS BIAISÉS DANS LE  
MODÈLE DE RÉGRESSION LINÉAIRE

par

Elmiloud Bensaïd

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de Maître ès sciences (M.SC.)  
en statistique

Juillet 1997

© ELMILOUD BENSALD, MCMXCVII



QA  
3  
U54  
1998  
V.003

Université de Montréal

QUELQUES ESTIMATEURS BIAISÉS DANS LE  
MODELE DE RÉGRESSION LINÉAIRE

par

Elmoud Bensaïd

ÉCOLE NATIONALE DE STATISTIQUE  
CITÉ DE LA PLÉIADE

Je soussigné, soussignée, le 18 1998, déclare avoir écrit et soumis  
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)  
en statistique

LE 18 1998



Document communiqué par le service de la bibliothèque

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

**QUELQUES ESTIMATEURS BIAISÉS DANS LE  
MODÈLE DE RÉGRESSION LINÉAIRE**

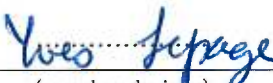
présenté par

**Elmiloud Bensaid**

a été évalué par un jury composé des personnes suivantes :

  
\_\_\_\_\_  
(président-rapporteur)

*Martin Bilodeau*  
\_\_\_\_\_  
(directeur de recherche)

  
\_\_\_\_\_  
(membre du jury)

Mémoire accepté le :

26-11-1997.

## SOMMAIRE

---

L'objectif du présent mémoire est l'application de quelques estimateurs biaisés dans un modèle de régression linéaire.

Des simulations ont été faites sur les quatre estimateurs suivants: l'estimateur ridge, l'estimateur de James et Stein à partie positive [Baranchik (1970)], Strawderman (1978) et Srivastava et Bilodeau (1989). Le but de ces simulations est d'évaluer à travers une fonction de perte fixée le risque de chacun de ces estimateurs.

Nous avons alors comparé ces différents risques afin d'établir lequel de ces estimateurs présente la meilleure performance.

# TABLE DES MATIÈRES

---

<b>SOMMAIRE</b>	<b>iii</b>
<b>LISTE DES FIGURES</b>	<b>vi</b>
<b>LISTE DES TABLEAUX</b>	<b>viii</b>
<b>REMERCIEMENTS</b>	<b>1</b>
<b>INTRODUCTION</b>	<b>2</b>
<b>CHAPITRE 1. Méthodes d'estimation des paramètres dans un modèle linéaire</b>	<b>3</b>
1.1. INTRODUCTION	3
1.2. ESTIMATEURS LINÉAIRES	4
1.2.1. Estimateur ridge	4
1.2.2. Estimateur inverse généralisé	8
1.2.3. Estimateur Bayésien empirique	9
1.2.4. Estimateur de James-Stein	10
1.2.5. Autres estimateurs Minimax	11
1.2.6. Estimateur linéaire d'erreur quadratique moyenne minimale	13
1.2.7. Représentation spectrale de certains estimateurs	14
<b>CHAPITRE 2. Les méthodes de sélection du paramètre ridge</b>	<b>18</b>
2.1. INTRODUCTION	18
2.2. MÉTHODES DE SÉLECTION	18
2.2.1. Méthode de validation croisée	18
2.2.2. Méthode de validation croisée généralisée	21

2.2.3. Le critère de Mallows	23
2.2.4. Méthode de McDonald et Galarneau	25
2.2.5. La méthode du bootstrap	27
<b>CHAPITRE 3. Comparaison entre certains estimateurs</b>	<b>28</b>
3.1. INTRODUCTION	28
3.2. SIMULATIONS	28
3.2.1. Plan des simulations	28
3.2.2. Le choix du paramètre ridge	29
3.2.3. Construction des variables explicatives	30
3.2.4. Résultats	30
3.2.5. Conclusion	31
<b>CONCLUSION</b>	<b>44</b>
<b>APPENDICES</b>	<b>45</b>
<b>RÉFÉRENCES</b>	<b>54</b>

## LISTE DES FIGURES

---

1.1	La variance totale de l'estimateur ridge	5
1.2	Le carré du biais de l'estimateur ridge	5
1.3	Le carré de la norme de l'estimateur ridge	6
1.4	L'erreur quadratique moyenne	6
1.5	Les valeurs propres de la matrice $A$ associées à chaque estimateur de la forme $\tilde{\beta} = P'APX'Y$ .	17
2.6	La fonction $CVC(k)$	20
2.7	La fonction $GCV(k)$	22
3.8	La distribution expérimentale de $\hat{\sigma}^2$ , pour $\sigma^2 = 1; 0, 1; 25; 100$ , $\beta = (0, 0, 0, 0)'$ et $S(\beta) = 0$	35
3.9	La distribution expérimentale de la première composante de chaque estimateur, pour $\sigma^2 = 1$ , $\beta = (0, 0, 0, 0)'$ , $S(\beta) = 0$ et $Q = I_4$	36
3.10	La distribution expérimentale de la deuxième composante de chaque estimateur, pour $\sigma^2 = 1$ , $\beta = (0, 0, 0, 0)'$ , $S(\beta) = 0$ et $Q = I_4$	37
3.11	La distribution expérimentale de la troisième composante de chaque estimateur, pour $\sigma^2 = 1$ , $\beta = (0, 0, 0, 0)'$ , $S(\beta) = 0$ et $Q = I_4$	38
3.12	La distribution expérimentale de la quatrième composante de chaque estimateur, pour $\sigma^2 = 1$ , $\beta = (0, 0, 0, 0)'$ , $S(\beta) = 0$ et $Q = I_4$	39
3.13	La distribution expérimentale de la première composante de chaque estimateur, pour $\sigma^2 = 1$ , $\beta = (0, 0, 0, 0)'$ , $S(\beta) = 0$ et $Q = X'X$	40
3.14	La distribution expérimentale de la première composante de chaque estimateur, pour $\sigma^2 = 1$ , $\beta = (0, 0, 0, 0)'$ , $S(\beta) = 0$ et $Q = (X'X)^2$	41

- 3.15 La distribution expérimentale de la deuxième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (1, 078; 0, 764; -1, 257; 0, 234)'$ ,  $S(\beta) = 100$  et  $Q = I_4$  42
- 3.16 La distribution expérimentale de la deuxième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (1, 078; 0, 764; -1, 257; 0, 234)'$ ,  $S(\beta) = 100$  et  $Q = X'X$  43



## LISTE DES TABLEAUX

---

3.1	Les risques estimés des estimateurs, pour $\sigma^2 = 0, 1$ , $S(\beta) = 0, 1$ , $Q = I_4$ et $\beta = \beta_1, \beta_2, \beta_3, \beta_4$	32
3.2	Les risques estimés des estimateurs, pour $\sigma^2 = 0, 1$ , $S(\beta) = 0, 1$ , $Q =$ $X'X$ et $\beta = \beta_1, \beta_2, \beta_3, \beta_4$	32
3.3	Les risques estimés des estimateurs, pour $\sigma^2 = 0, 1$ , $S(\beta) = 0, 1$ , $Q =$ $(X'X)^2$ et $\beta = \beta_1, \beta_2, \beta_3, \beta_4$	32
3.4	Les risques estimés des estimateurs, pour $\sigma^2 = 1$ , $S(\beta) = 0; 0, 1; 1; 10; 100$ , $Q = I_4$ et $\beta = \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$	33
3.5	Les risques estimés des estimateurs, pour $\sigma^2 = 1$ , $S(\beta) = 0; 0, 1; 1; 10; 100$ , $Q = X'X$ et $\beta = \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$	33
3.6	Les risques estimés des estimateurs, pour $\sigma^2 = 1$ , $S(\beta) = 0; 0, 1; 1; 10; 100$ , $Q = (X'X)^2$ et $\beta = \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$	33
3.7	Les risques estimés des estimateurs, pour $\sigma^2 = 50$ , $S(\beta) = 100$ , $Q = I_4$ et $\beta = \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$	34
3.8	Les risques estimés des estimateurs, pour $\sigma^2 = 50$ , $S(\beta) = 100$ , $Q =$ $X'X$ et $\beta = \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$	34
3.9	Les risques estimés des estimateurs, pour $\sigma^2 = 50$ , $S(\beta) = 100$ , $Q =$ $(X'X)^2$ et $\beta = \beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}$	34

## REMERCIEMENTS

---

Je tiens à témoigner toute ma gratitude à mon directeur de recherche Martin Bilodeau, dont l'appui et la grande disponibilité ont permis à ce mémoire de voir le jour.

Je tiens à remercier également le "FCAR" pour la bourse qu'il m'a accordée durant ma recherche.

Enfin, je désire également remercier toute ma famille et mes amis qui m'ont soutenu tout au long de mes études.

## INTRODUCTION

---

La régression linéaire multiple est une méthode statistique très utilisée dans divers domaines, comme la technologie, les sciences sociales, l'économie, la finance, etc...

Dans le modèle linéaire usuel, on estime souvent le vecteur des paramètres de régression par la méthode des moindres carrés, qui en présence de multicolinéarité, possède une variance tendant vers l'infini. Hoerl et Kennard (1970a, b) pour pallier au problème de multicolinéarité ont proposé l'estimateur ridge. Suivant une autre direction, plusieurs résultats sur des estimateurs de type minimax ont été publiés en commençant par la recherche séminale de James et Stein (1956), ainsi que beaucoup d'autres qui suivirent, comme celles de Baranchik (1970), Strawderman (1978) et Srivastava et Bilodeau (1989).

Dans le premier chapitre, nous faisons une revue de la littérature de ces estimateurs. Nous expliquons certaines de leurs propriétés, et soulignons des similitudes entre ces différents estimateurs. L'estimateur de Stein à partie positive ainsi que celui de ridge réduisent, tous deux, en norme l'estimateur des moindres carrés. Ces estimateurs ont tous en commun qu'ils dilatent les valeurs propres de la matrice  $X'X$ .

Dans le deuxième chapitre nous étudions certaines méthodes permettant de sélectionner le paramètre ridge, comme le critère de la validation croisée, la validation croisée généralisée, le critère de Mallows, la méthode du bootstrap et la méthode de McDonald et Galarneau.

Dans le troisième chapitre nous présentons une simulation numérique qui sert à comparer l'estimateur ridge avec certains estimateurs minimax, comme l'estimateur de James et Stein à partie positive [Baranchik (1970)], de Strawderman (1978) et de Srivastava et Bilodeau (1989).

# CHAPITRE 1

---

## Méthodes d'estimation des paramètres dans un modèle linéaire

### 1.1. INTRODUCTION

Le modèle standard de régression linéaire multiple est

$$Y = X\beta + \varepsilon$$

où  $Y$ :  $n \times 1$  est le vecteur des observations dépendantes,  $X$ :  $n \times p$  est une matrice connue des variables explicatives avec  $p \leq n$ ,  $\beta$ :  $p \times 1$  est le vecteur des coefficients inconnus,  $\varepsilon$ :  $n \times 1$  est le vecteur des erreurs tel que  $\varepsilon \sim N(0, \sigma^2 I_p)$ ,  $I_p$  est la matrice unité d'ordre  $p$  et  $\sigma^2$  est une constante inconnue.

En situation idéale, c'est-à-dire lorsque les données ne présentent pas le problème de multicolinéarité, expliqué ultérieurement, l'estimateur des moindres carrés produira une bonne estimation. Toutefois, en d'autres circonstances, certains estimateurs bien que biaisés, pourront constituer une meilleure alternative par rapport à l'estimateur des moindres carrés. L'estimateur des moindres carrés,  $\hat{\beta} = (X'X)^{-1}X'Y$  est sans biais et sa matrice des covariances est  $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$ . Il est aussi l'estimateur du maximum de vraisemblance, car sous la normalité, le modèle s'écrit aussi comme:  $Y \sim N(X\beta, \sigma^2 I_p)$ . La fonction d'erreur quadratique moyenne d'un estimateur quelconque  $\tilde{\beta}$  est  $EQM(\tilde{\beta}) = E\|\tilde{\beta} - \beta\|^2$ , où  $\|\cdot\|$  est la norme euclidienne.

En particulier, pour l'estimateur des moindres carrés, on a  $EQM(\hat{\beta}) = \sigma^2 \sum_{i=1}^p 1/\lambda_i$ , où  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$  désignent les valeurs propres de la matrice  $(X'X)$ . Lorsqu'une valeur propre  $\lambda_i$  est très proche de zéro, on dit qu'il y a un problème de multicollinéarité. Dans ce cas  $\hat{\beta}$  n'est pas stable, ce qui se traduit par une grande variance, puisque  $EQM(\hat{\beta}) = \sigma^2 \sum_{i=1}^p 1/\lambda_i$  est grand quand  $\lambda_p$  approche zéro. Certains auteurs, ont alors proposé des estimateurs linéaires biaisés, comme l'estimateur ridge qui sera étudié dans ce mémoire.

## 1.2. ESTIMATEURS LINÉAIRES

### 1.2.1. Estimateur ridge.

L'estimateur ridge est  $\beta^* \equiv \beta^*(k) = (X'X + kI_p)^{-1}X'Y$ , où  $k \geq 0$  est un scalaire à déterminer. Cet estimateur a été introduit par Hoerl et Kennard (1970a), comme solution au problème des matrices de corrélation  $(X'X)$ , qui présentent la multicollinéarité. Il est une transformation linéaire de  $\hat{\beta}$ , puisqu'on peut réécrire  $\beta^* = (X'X + kI_p)^{-1}(X'X)\hat{\beta}$  et sa matrice des covariances est

$$Var(\beta^*) = \sigma^2(X'X + kI_p)^{-1}(X'X)(X'X + kI_p)^{-1}.$$

Comme cet estimateur est biaisé, nous allons nous intéresser à son erreur quadratique moyenne

$$EQM(\beta^*) = V_T(\beta^*) + ||Biais(\beta^*)||^2$$

où

$$\begin{aligned} V_T(\beta^*) &= tr[Var(\beta^*)] \\ &= \sigma^2 \sum_{i=1}^p \lambda_i / (\lambda_i + k)^2 \end{aligned}$$

est la variance totale, où  $tr(\cdot)$  est la trace d'une matrice donnée. D'autre part  $||Biais(\beta^*)||^2 = \sum_{i=1}^p \alpha_i^2 / \left(1 + \frac{\lambda_i}{k}\right)^2$ , où  $\alpha = P\beta$  et  $P$  est la matrice des vecteurs

propres de la matrice  $(X'X)$  dans la décomposition  $(X'X) = P' \text{diag}(\lambda_i) P$ ,  $\text{diag}(\lambda_i)$  est une matrice diagonale pour éléments diagonaux  $\lambda_1, \dots, \lambda_p$ . La fonction  $V_T(\beta^*)$  est décroissante en  $k$  alors que  $\| \text{Biais}(\beta^*) \|^2$  est une fonction croissante. Pour  $\sigma^2 = 1$ ,  $\lambda_1 = 1,647$ ,  $\lambda_2 = 17,832$ ,  $\lambda_3 = 30,807$ ,  $\lambda_4 = 1689,882$ ,  $\beta = (-0,247; -0,209; 0,042; -0,009)'$  et  $\alpha = (-0,1652; 0,123; -0,060; -0,246)'$ , nous avons tracé les fonctions  $V_T(\beta^*)$  et  $\| \text{Biais}(\beta^*) \|^2$  comme fonction de  $k$  (voir la figure 1.1 et 1.2).

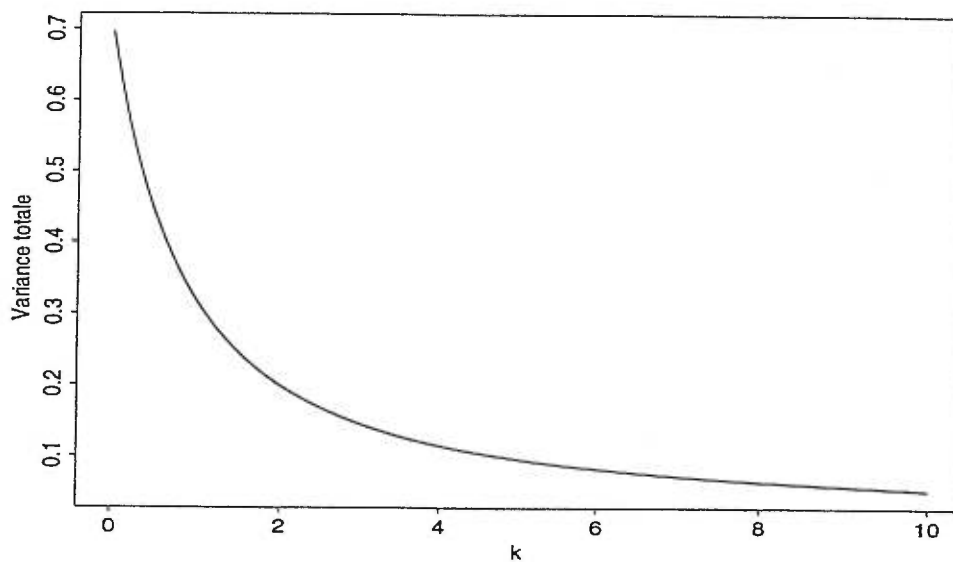


FIGURE 1.1. La variance totale de l'estimateur ridge

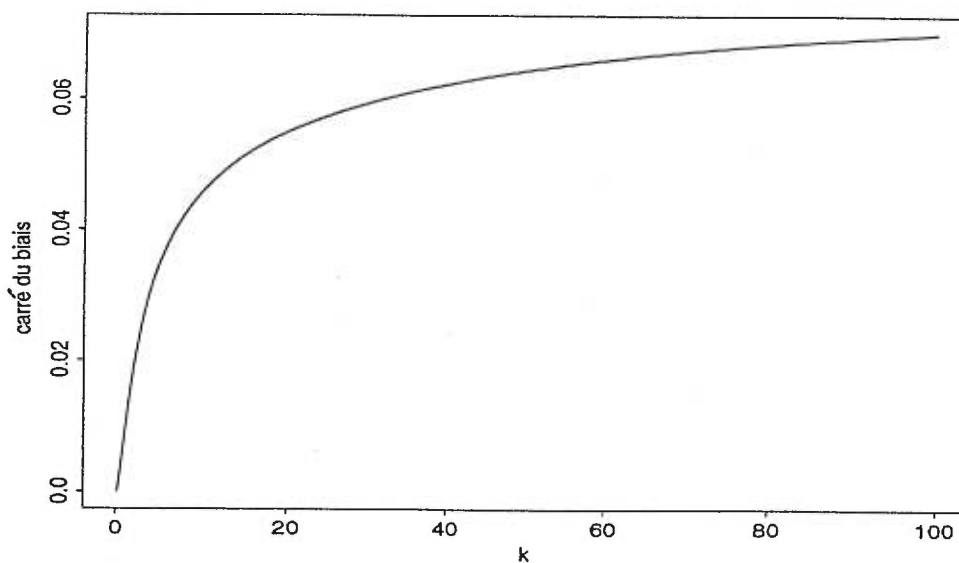


FIGURE 1.2. Le carré du biais de l'estimateur ridge

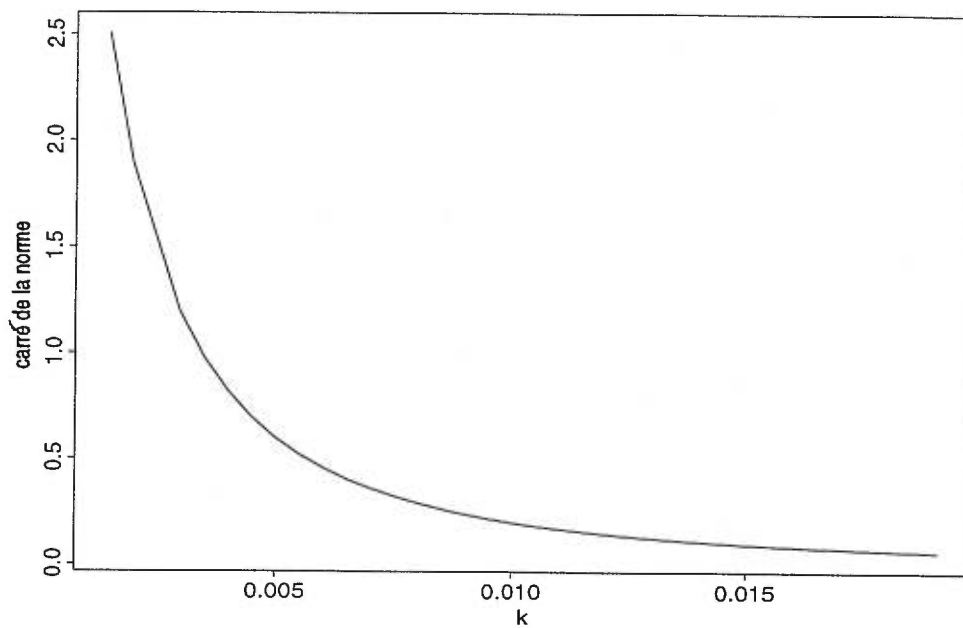


FIGURE 1.3. Le carré de la norme de l'estimateur ridge

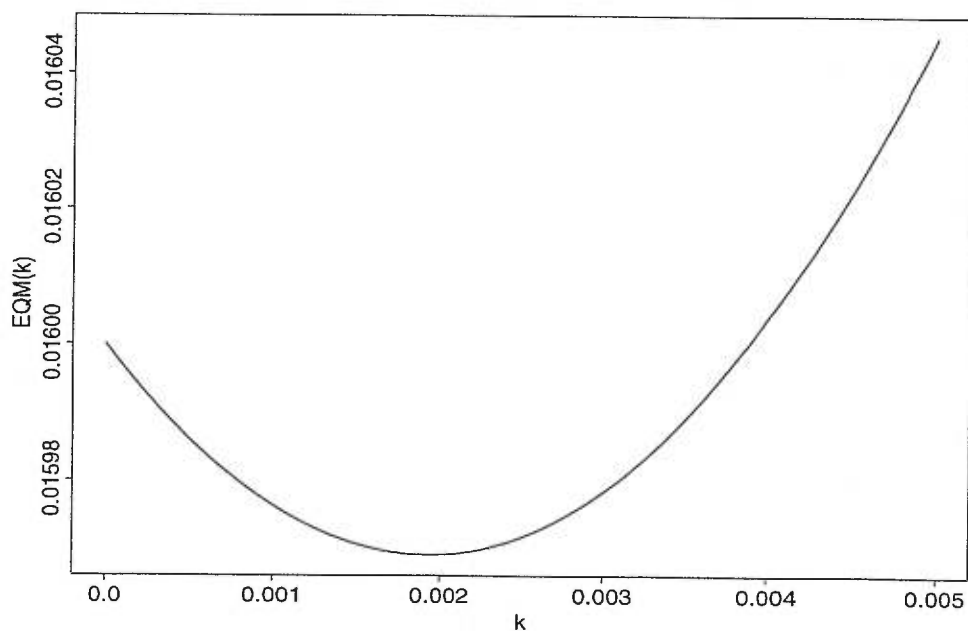


FIGURE 1.4. L'erreur quadratique moyenne

De plus on a

$$\lim_{k \rightarrow \infty} \|\text{Biais}(\beta^*)\|^2 = \beta' \beta$$

et

$$\lim_{k \rightarrow \infty} V_T(\beta^*) = 0$$

de telle façon que

$$\lim_{k \rightarrow \infty} EQM(\beta^*) = \beta' \beta.$$

Dans le cas où  $X'X = I_p$

$$EQM(\beta^*) = (1 + k)^{-2}(p\sigma^2 + k^2\beta' \beta). \quad (1.2.1)$$

Par exemple, lorsque  $p = 4$ ,  $\sigma^2 = 0,004$  et  $\beta' \beta = 8,256$ , on trouve  $EQM(\beta^*) = (1 + k)^{-2}(0.016 + 8.256k^2)$ . Le graphe obtenu à la figure 1.3 montre bien l'allure de cette fonction. Dans ce cas, la fonction d'erreur quadratique moyenne atteint son minimum pour  $k \simeq 0,002$ . On note qu'il y a un compromis à faire entre la variance totale et le biais. Lorsque le biais diminue la variance totale augmente, et vice versa. En sus de ces propriétés, on sait que la norme de  $\beta^*$ , qui est une fonction décroissante de  $k$ , est plus petite que celle de  $\hat{\beta}$ ,  $\|\beta^*\| \leq \|\hat{\beta}\|$ , avec

$$\lim_{k \rightarrow \infty} \|\beta^*\| = 0.$$

En utilisant la matrice  $X$  et le vecteur  $Y$  de l'appendice  $B$  et  $C$ , nous avons tracé le carré de la norme, comme fonction de  $k$  (voir la figure 1.4). De plus, on a  $\|\beta^*\|^2 < \frac{\lambda_1}{(\lambda_1 + k)} \|\hat{\beta}\|^2$  car  $\lambda_1/(\lambda_1 + k)$  est la plus grande valeur propre de la matrice  $(X'X + kI_p)^{-1}(X'X)$ . Toutes ces propriétés sont facilement vérifiables, elles se trouvent dans Hoerl et Kennard (1970a). Généralement on cherche à augmenter la précision de  $\hat{\beta}$  en introduisant un biais. Hoerl et Kennard (1970a) montre que pour chaque  $X$  et  $\beta$  fixes, il existe un  $k \geq 0$  tel que  $EQM(\beta^*) < V_T(\hat{\beta})$ . Cependant, Thistel (1978) a aussi démontré que pour chaque  $X$  et  $k > 0$  fixes, il existe une régression (i.e une valeur de  $\beta$ ), telle que  $V_T(\hat{\beta}) < EQM(\beta^*)$ . Donc, si l'objectif est de réduire l'erreur quadratique moyenne, il faudra estimer  $k$  à l'aide des données. Plusieurs de ces méthodes seront présentées au chapitre 2. Hemmerle (1975) a utilisé l'estimateur ridge généralisé par Hoerl et Kennard (1970a), en posant  $X^* = XP'$  et  $\alpha = P\beta$  où  $P$  est la matrice des vecteurs propres de  $X'X$ , réécrit le modèle sous la forme  $Y = X^*\alpha + \varepsilon$ , pour lequel, l'estimateur ridge généralisé est défini par  $\hat{\alpha}^* = (X^{*'}X^* + K)^{-1}X^{*'}Y$ , où



$K = \text{diag}(k_i)$  avec  $k_i \geq 0$  et  $1 \leq i \leq p$ . On note que  $X^{*'}X^* = \text{diag}(\lambda_i)$ . L'erreur quadratique moyenne devient alors,

$$EQM(\hat{\alpha}^*) = \sum_{i=1}^p \frac{\sigma^2 \lambda_i + \alpha_i^2 k_i^2}{(\lambda_i + k_i)^2}. \quad (1.2.2)$$

En minimisant cette quantité, on obtient alors, la valeur optimale  $k_i = \sigma^2/\alpha_i^2$ , où les  $\alpha_i$  et  $\sigma^2$  restent à estimer. Certains auteurs ont utilisé la matrice d'erreur quadratique moyenne définie par

$$\begin{aligned} M(k) &= E [(\beta^* - \beta)(\beta^* - \beta)'] \\ &= \sigma^2(X'X + kI_p)^{-1}X'X(X'X + kI_p)^{-1} + \\ &\quad k^2(X'X + kI_p)^{-1}\beta\beta'(X'X + kI_p)^{-1}. \end{aligned}$$

Hoerl et Kennard (1970a) et Theobald (1974) montrent que si  $0 < k < k_0$  avec  $k_0 = 2\sigma^2/\beta'\beta$ , alors  $M(0)-M(k)$  est définie positive. Cette amélioration se produit pour chaque composante, c'est-à-dire que  $EQM(\beta_i^*) < EQM(\hat{\beta}_i)$ . On note que  $EQM(\hat{\beta}) = \text{tr}(M(0))$ ,  $EQM(\beta^*) = \text{tr}(M(k))$ , et sous la même condition  $EQM(\beta^*) < EQM(\hat{\beta})$ . Finalement, une dernière interprétation de l'estimateur ridge est celle d'un estimateur des moindres carrés pénalisés. En effet, la valeur de  $\beta$  qui minimise,

$$\sum_{i=1}^n (Y_i - X_i'\beta)^2 + \lambda\beta'\beta,$$

où les  $X_i'$  sont les vecteurs lignes de la matrice  $X$ , est  $\beta = (X'X + \lambda I_p)^{-1}X'Y$ . L'estimateur ridge peut donc être considéré, comme celui qui minimise la somme des carrés résiduels, en imposant toutefois une pénalité sur la norme de l'estimateur.

### 1.2.2. Estimateur inverse généralisé.

Supposons que les  $r$  premières valeurs propres de  $(X'X)$  sont non nulles, et que les autres  $(p - r)$  valeurs propres sont nulles. Posons alors,

$$(X'X)_r = \sum_{i=1}^r \frac{1}{\lambda_i} P_i P_i'$$

où  $P$  est la même matrice de la section précédente et  $P_1, \dots, P_p$  désignent les vecteurs colonnes de la matrice  $P$ . L'estimateur inverse généralisé [Marquardt (1970)] est défini alors de la façon suivante:  $\beta^+ = (X'X)_r X'Y$ . Cet estimateur minimise la somme des carrés résiduels dans l'espace engendré par les vecteurs  $P_1, \dots, P_p$ . On peut alors écrire

$\beta^+ = (X'X)_r(X'X)\hat{\beta}$ . C'est donc un estimateur qui a la même forme que celui de ridge (i.e  $\beta^+ = A\hat{\beta}$ , où  $A$  est une matrice symétrique). Le carré de sa norme  $\|\beta^+\|^2$  est une fonction en escalier croissante avec  $r$ . Son espérance mathématique est  $E(\beta^+) = (X'X)_r(X'X)\beta$  et sa matrice des covariances est  $Var(\beta^+) = \sigma^2 Z_r(X'X)^{-1}Z_r'$  avec  $Z_r = (X'X)_r(X'X)$ . Son erreur quadratique moyenne est  $EQM(\beta^+) = \sigma^2 \sum_{i=1}^r 1/\lambda_i + \beta'(Z_r - I_p)'(Z_r - I_p)\beta$ . Le terme de variance,  $\sigma^2 \sum_{i=1}^r 1/\lambda_i$  croît avec  $r$  et le carré du biais,  $\beta'(Z_r - I_p)'(Z_r - I_p)\beta$  décroît avec  $r$ . Marquardt (1970) a aussi proposé l'estimateur combiné ridge et inverse généralisé, comme suit

$$\tilde{\beta} = [(X'X)_r + kI]^{-1} X'Y.$$

Il est conçu de façon à éliminer non seulement l'effet des  $(p - r)$  valeurs propres nulles, mais également celui des valeurs propres proches de zéro. Le principe consiste à chercher, les valeurs de  $k$  qui donnent l'erreur quadratique moyenne minimale pour  $r$  fixe.

### 1.2.3. Estimateur Bayésien empirique.

Nous supposons ici que

$$Y = \gamma 1_n + X\beta + \varepsilon,$$

où  $X$ :  $n \times p$  exclut l'ordonnée à l'origine,  $1_n = (1, \dots, 1)$  est un vecteur  $n \times 1$ ,  $\gamma$  est un scalaire, et que les variables explicatives sont centrées réduites pour que  $X'X$  soit une matrice de corrélation. Efron et Morris (1973) supposent que  $\beta$  suit la distribution a priori  $\beta \sim N(0, \sigma_s^2 I_p)$  avec  $\sigma_s^2$  scalaire inconnu. La distribution a posteriori de  $\beta$  est alors  $N\left((X'X + \frac{\sigma^2}{\sigma_s^2})^{-1}X'Y, C\right)$  pour une certaine matrice  $C$  donnée. L'espérance mathématique de la distribution a posteriori de  $\beta$  est donc

$$E(\beta | Y) = \left(X'X + \frac{\sigma^2}{\sigma_s^2} I_p\right)^{-1} X'Y$$

qu'on adopte, comme estimateur de Bayes pour  $\beta$ . On note la similarité avec l'estimateur ridge. La forme de l'estimateur bayésien dans le modèle  $Y = \gamma 1_n + X^*\alpha + \varepsilon$ , où  $X^* = XP'$  et  $PX'XP' = \text{diag}(\lambda_i)$ , est  $\hat{\alpha}_{B,i} = \frac{\lambda_i}{\lambda_i + \sigma^2/\sigma_s^2} \hat{\alpha}_i$ , où on a supposé de façon

équivalente la distribution a priori  $\alpha \sim N(0, \sigma_s^2 I_p)$ . Le rapport  $\sigma_s^2/\sigma^2$  est ensuite estimé par l'estimateur

$$\frac{1}{p\hat{\sigma}^2} \sum_{i=1}^r \lambda_i \hat{\alpha}_i^2 - 1.$$

On peut démontrer, par construction, que  $E(\hat{\alpha}_i|\alpha_i) = \alpha_i$ ,  $Var(\hat{\alpha}_i|\alpha_i) = \sigma^2/\lambda_i$ , où  $\alpha_i \sim N(0, \sigma_s^2)$ . D'où  $E(\lambda_i \hat{\alpha}_i^2|\alpha_i) = \sigma^2 + \lambda_i \alpha_i^2$  et  $E(\lambda_i \hat{\alpha}_i^2) = \sigma^2 + \lambda_i \sigma_s^2$ . Enfin,

$$\begin{aligned} E\left(\sum_{i=1}^r \frac{\lambda_i \alpha_i^2}{p\sigma^2}\right) - 1 &= \sum_{i=1}^r \frac{\lambda_i \sigma_s^2}{p\sigma^2} \\ &= \sigma_s^2/\sigma^2 \end{aligned} \quad (1.2.3)$$

où  $tr(X'X) = \sum_{i=1}^p \lambda_i = p$  puisque  $X'X$  est une matrice de corrélation. On peut consulter Lawless et Wang (1976), pour plus de détails concernant ces résultats.

#### 1.2.4. Estimateur de James-Stein.

Dans un modèle linéaire, le principe de minimaxité consiste à trouver l'estimateur  $\tilde{\beta}$ , qui minimise le maximum de la fonction de risque

$$R(\tilde{\beta}; A) = E \left[ (\tilde{\beta} - \beta)' A (\tilde{\beta} - \beta) \right],$$

où  $A$  est une matrice fixe. Formellement, l'estimateur  $\hat{\beta}_n$  est dit minimax si

$$R(\hat{\beta}_n; A) = \inf_{\tilde{\beta}} \sup_{\beta} R(\tilde{\beta}; A).$$

Soient  $Z: p \times 1$  de loi  $N(\mu, \sigma^2 I_p)$  et  $S \sim \sigma^2 \chi_n^2$  des variables indépendantes, où  $\chi_n^2$  désigne la loi du khi-deux a  $n$  degrés de liberté et  $\mu$  est un vecteur  $p \times 1$ . en considérant l'estimateur de  $\mu$  par rapport à la fonction de risque  $R(\hat{\mu}, I_p) = E [(\hat{\mu} - \mu)'(\hat{\mu} - \mu)]$ . James et Stein (1961) ont proposé

$$\hat{\mu} = \mu_{J,S} = \left( 1 - \frac{cS}{\|Z\|^2} \right) Z.$$

Cet estimateur est minimax si  $0 \leq c \leq 2(p-2)/(n+2)$ . Dans le modèle linéaire, il suffit de remplacer  $Z$  par  $(X'X)^{1/2} \hat{\beta}$  et  $S$  par  $\sum_{i=1}^n \hat{\epsilon}_i^2 = \|Y - X\hat{\beta}\|^2 \sim \sigma^2 \chi_{n-p}^2$  pour conclure que

$$\beta_{J,S} = \left( 1 - \frac{c\|Y - X\hat{\beta}\|^2}{\hat{\beta}' X' X \hat{\beta}} \right) \hat{\beta}$$

est un estimateur minimax par rapport à la fonction de risque  $R(\tilde{\beta}; X'X)$  si  $0 \leq c \leq 2(p-2)/(n-p+2)$ . L'estimateur de James et Stein est aussi de la forme  $\beta_{J,S} = A\hat{\beta}$ ,

où  $A$  est une matrice symétrique (dans ce cas  $A = \left(1 - \frac{c\|Y-X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta}}\right) I_p$ ) comme celui de ridge. Ils cherchent tous les deux à réduire la norme de l'estimateur usuel  $\hat{\beta}$ , pour obtenir une meilleure estimation. Dans le cas où le facteur de réduction en norme est constant, considérons  $(1-w)\hat{\beta}$  avec  $0 \leq w \leq 1$ . Il est facile de vérifier que l'erreur quadratique moyenne est

$$\begin{aligned} EQM\left((1-w)\hat{\beta}\right) &= w^2\beta'\beta + (1-w)^2EQM(\hat{\beta}) \\ &= w^2\beta'\beta + (1-w)^2\sigma^2tr(X'X)^{-1}. \end{aligned}$$

Le minimum est atteint pour  $w = \sigma^2tr(X'X)^{-1} / [\sigma^2tr(X'X)^{-1} + \beta'\beta]$ . On peut toujours remplacer  $\sigma^2$  et  $\beta$  par les estimateurs usuels, mais alors on ne sait pas si l'estimateur est minimax contrairement à l'estimateur de James et Stein. Dans le chapitre 3, l'estimateur ridge sera comparé à des estimateurs de type James et Stein par des simulations numériques.

### 1.2.5. Autres estimateurs Minimax.

Baranchik (1970) a utilisé la fonction d'erreur quadratique moyenne  $R(\tilde{\mu}; I_p) = \|\tilde{\mu} - \mu\|^2/\sigma^2$ , pour estimer la moyenne  $\mu$  d'une loi multinormale,  $N(\mu, \sigma^2 I_p)$ , où  $\sigma^2$  est paramètre inconnu. Il a proposé l'estimateur

$$\varphi(Z, S) = \left(1 - \frac{r(F)}{F}\right)Z,$$

où  $Z$ :  $p \times 1$  de loi  $N(\mu, \sigma^2 I_p)$ ,  $S \sim \sigma^2 \chi_n^2$  des variables indépendantes et  $F = \|Z\|^2/S$ . Cet estimateur est minimax si  $p \geq 3$  et  $r(\cdot)$  est une fonction monotone non décroissante satisfaisant  $0 \leq r(\cdot) \leq 2(p-2)/(n+2)$ . Dans le cas où  $r(\cdot) = c$  (constante) on obtient l'estimateur de type James et Stein pour la moyenne  $\mu_{J,S} = \left(1 - \frac{cS}{\|Z\|^2}\right)Z$ ; et si  $r(F) = c/(1+cF^{-1})$ , on a  $\varphi(F, S) = [\|Z\|^2/(\|Z\|^2 + cS)]Z$ , avec  $0 \leq c \leq (p-2)/(n+2)$ , qui définit l'estimateur de Alam et Thompson (1964). Si on choisit  $r(F) = \min(F, c)$ , alors  $1 - r(F)/F = \max(0, 1 - c/F)$ , et on trouve  $\varphi(Z, S) = \max(0, 1 - c/F)Z$  qui définit l'estimateur à partie positive de type James et Stein, ( $\min(a, b)$  et  $\max(a, b)$  désignent respectivement le plus petit et le plus grand nombre entre  $a$  et  $b$ ). Strawderman (1973) a généralisé l'estimateur de Baranchik;

son estimateur est

$$\varphi(Z, S) = [1 - \frac{r(F, S)}{F}]Z$$

où  $r(., S)$  est croissante pour  $S$  fixe et  $r(F, .)$  est décroissante pour  $F$  fixe et  $0 \leq r(., .) \leq 2(p-2)/(n+2)$ . On peut transposer facilement ces résultats à la régression.

Strawderman (1978) a montré que l'estimateur suivant

$$\tilde{\beta} = \left[ I_p + \frac{a \|Y - X\hat{\beta}\|^2 AX'X}{\hat{\beta}'X'X\hat{\beta} + g \|Y - X\hat{\beta}\|^2 + h} \right]^{-1} \hat{\beta}$$

est minimax par rapport à la fonction de risque  $R(\tilde{\beta}; A^{-1})$  lorsque  $h \geq 0$ ,  $g \geq 2p/(n-p+2)$ ,  $0 \leq a \leq 2(p-2)/(n-p+2) \cdot [1/\max_{1 \leq i \leq p} \gamma_i]$ , avec les  $\gamma_i$  sont les valeurs propres de la matrice  $AX'X$  et  $\max_{1 \leq i \leq p} \gamma_i$  désigne la plus grande valeurs des  $\gamma_i$  et  $A$  est une matrice donnée. Si  $A^{-1} = (X'X)^2$ , on trouve

$$\tilde{\beta} = \left[ I_p + \frac{a \|Y - X\hat{\beta}\|^2 (X'X)^{-1}}{\hat{\beta}'X'X\hat{\beta} + g \|Y - X\hat{\beta}\|^2 + h} \right]^{-1} \hat{\beta}$$

qui est de la forme d'un estimateur ridge général. Lorsque les erreurs du modèle de régression suivent une distribution elliptique, et la fonction de risque est  $R(\tilde{\beta}; Q)$ , où  $Q$  est une matrice définie positive, Srivastava et Bilodeau (1989) ont démontré que l'estimateur suivant

$$\hat{\beta}_{SB} = \left[ I_p - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'XQ^{-1}X'X\hat{\beta}} Q^{-1}X'X \right] \hat{\beta}$$

est minimax si  $0 \leq c \leq 2(p-2)/(n-p+2)$ . En particulier, si on prend  $Q = I_p$  on trouve

$$\hat{\beta}_{SB} = \left[ I_p - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'(X'X)^2\hat{\beta}} X'X \right] \hat{\beta}$$

et pour  $Q = (X'X)^2$

$$\hat{\beta}_{SB} = \left[ I_p - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'\hat{\beta}} (X'X)^{-1} \right] \hat{\beta}.$$

Une autre notion du caractère minimax est utilisée par Bibby et Toutenburg (1977, pp. 150-151), avec la même fonction de risque  $R(\tilde{\beta}; A^{-1})$ . Un estimateur  $\hat{\beta}_n$  est dit minimax si

$$R(\hat{\beta}_n; A^{-1}) = \inf_{\tilde{\beta}} \sup_{\beta \in B} R(\tilde{\beta}; A^{-1})$$

où  $B$  est une boule de rayon  $h$  ( $B = \{\beta : \beta'T\beta \leq h\}$ ), avec  $T$  matrice connue définie positive.

Considérons l'estimateur linéaire de la forme  $\tilde{\beta} = C'Y$ , où  $C$  est une matrice quelconque. Cet estimateur est minimax à condition que la matrice  $C$  minimise

$$\sup_{\beta \in B} R(\tilde{\beta}; A) = \sigma^2 \text{tr}(AC'C) + h\lambda_1(\tilde{A})$$

où  $\tilde{A} = T^{-1/2}(C'X - I_p)'A(C'X - I_p)T^{-1/2}$ ,  $\lambda_1(\tilde{A})$  est la plus grande valeur propre de la matrice  $\tilde{A}$ . Läuter (1975) a examiné ce problème, dans le cas où  $\text{rang}(A)=1$ , et a démontré que l'estimateur de la forme  $C'Y$  est minimax par rapport  $R(\tilde{\beta}; A^{-1})$  si  $C = (h^{-1}\sigma^2T + X'X)^{-1}X'$ . Notons cependant que le cas  $\text{rang}(A) = 1$  donne une fonction de risque qui n'est pas intéressante d'un point de vue statistique, puisqu'on estime une seule combinaison linéaire.

### 1.2.6. Estimateur linéaire d'erreur quadratique moyenne minimale.

Theil (1971) a démontré que l'estimateur linéaire  $\beta_T$  qui est défini par

$$\beta_T = \beta\beta'X'(X\beta\beta'X' + \sigma^2I_n)^{-1}Y$$

est un estimateur d'erreur quadratique moyenne matricielle minimale. En effet, posons

$$b = \beta_T + AY$$

où  $A: p \times n$  est une matrice fixe. La matrice d'erreur quadratique moyenne de  $b$  devient

$$M(b) = \left\{1 - \beta'X'(X\beta\beta'X' + \sigma^2I_n)^{-1}X\beta\right\} \beta\beta' + A(X\beta\beta'X' + \sigma^2I_n)A'.$$

On note que  $M(b)$  est minimale pour  $A = 0$  car  $X\beta\beta'X' + \sigma^2I_n$  est définie positive, d'où  $\beta_T$  est l'estimateur linéaire avec une erreur quadratique moyenne minimale.

Farebrother (1975) a démontré que  $\beta_T$  s'écrit aussi:

$$\beta_T = \frac{\beta'X'Y\beta}{\sigma^2 + \beta'X'X\beta}.$$

Si on remplace  $\beta$  par  $\hat{\beta}$  et  $\sigma^2$  par  $S^2$ , on obtient un estimateur de la forme de celui de Baranchik

$$\begin{aligned} \hat{\beta}_B &= \frac{\hat{\beta}'X'Y\hat{\beta}}{S^2 + \hat{\beta}'X'X\hat{\beta}} \\ &= \left(1 - \frac{1}{1 + \hat{\beta}'X'X\hat{\beta}/S^2}\right) \hat{\beta}. \end{aligned}$$

D'autre part Farebrother (1975) a montré que

$$\begin{aligned}
 \beta_T &= \beta \left\{ \sigma^2 + \beta' X' X \beta \right\}^{-1} \beta' X' Y \\
 &= \beta \left\{ \beta' \left( \frac{\sigma^2}{\beta' \beta} I_p + X' X \right) \beta \right\}^{-1} \beta' X' Y \\
 &= \underbrace{\beta (\beta' V \beta)^{-1} \beta' V}_A \underbrace{(X' X + \frac{\sigma^2}{\beta' \beta} I_p)^{-1} X' Y}_B
 \end{aligned}$$

où  $V = X' X + \frac{\sigma^2}{\beta' \beta} I_p$ . L'estimateur  $\beta_T$  est ainsi composé de deux parties. La partie  $B$  forme justement l'estimateur ridge avec  $k = \sigma^2 / \beta' \beta$  et la partie  $A$  forme la régression de l'estimateur ridge sur  $\beta$  avec  $V^{-1}$  pour matrice de covariance.

### 1.2.7. Représentation spectrale de certains estimateurs.

Un objectif de ce mémoire est de trouver quelle méthode ou quel estimateur corrige bien le problème de multicolinéarité de la matrice  $(X' X)$ . Tous les estimateurs étudiés dans ce mémoire augmentent les valeurs propres, ou d'une manière équivalente réduisent en norme l'estimateur des moindres carrés. Dans cette section, on décrit certains estimateurs suivant la décomposition spectrale de  $X' X = P' \text{diag}(\lambda_i) P$ . Considérons d'abord l'estimateur ridge,

$$\begin{aligned}
 \beta^*(k) &= (X' X + k I_p)^{-1} X' Y \\
 &= P' \text{diag}(\lambda_i + k)^{-1} P X' Y.
 \end{aligned}$$

Dans ce cas, on remplace  $\lambda_i$  par  $f_1(\lambda_i) = \lambda_i + k \geq \lambda_i$ .

En deuxième lieu, on considère certains estimateurs minimax par rapport à la fonction de risque

$$R(\tilde{\beta}; Q) = E \left[ (\tilde{\beta} - \beta)' Q (\tilde{\beta} - \beta) \right].$$

Si  $Q = I_p$ , cela revient à utiliser la fonction d'erreur quadratique moyenne ordinaire. Parmi les estimateurs minimax par rapport à  $R(\tilde{\beta}; I_p)$  on retrouve celui de Srivastava et Bilodeau (1989),

$$\begin{aligned}
 \hat{\beta}_{SB} &= \left[ I_p - c \frac{\|Y - X \hat{\beta}\|^2}{\hat{\beta}' (X' X)^2 \hat{\beta}} X' X \right] \hat{\beta} \\
 &= P' \text{diag} \left( \frac{1}{\lambda_i} - c \frac{\|Y - X \hat{\beta}\|^2}{\hat{\beta}' (X' X)^2 \hat{\beta}} \right) P X' Y.
 \end{aligned}$$

Cela revient ici à remplacer chaque  $\lambda_i$  par  $f_2(\lambda_i) = \left( \frac{1}{\lambda_i} - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'(X'X)^2\hat{\beta}} \right)^{-1}$ . On remarque que  $f_2(\lambda_i) > \lambda_i$  pour tout  $1 \leq i \leq p$ .

Il y a aussi celui de Strawderman (1978),

$$\begin{aligned} \tilde{\beta} &= \left[ I_p + \frac{a\|Y - X\hat{\beta}\|^2(X'X)}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right]^{-1} \hat{\beta} \\ &= P' \text{diag} \frac{1}{\lambda_i} \left( 1 + \lambda_i \frac{a\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right)^{-1} PX'Y. \end{aligned}$$

Dans ce cas, on remplace  $\lambda_i$  par

$$f_3(\lambda_i) = \lambda_i \left( 1 + \lambda_i \frac{a\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right)$$

ainsi, les valeurs propres sont également augmentées. Si  $Q = X'X$ , on considère les estimateurs minimax. D'abord celui de Srivastava et Bilodeau (1989) est

$$\begin{aligned} \hat{\beta}_{SB} &= \left[ 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'(X'X)^2\hat{\beta}} \right] \hat{\beta} \\ &= P' \text{diag} \frac{1}{\lambda_i} \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'(X'X)^2\hat{\beta}} \right) PX'Y, \end{aligned}$$

où l'on remplace  $\lambda_i$

$$f_4(\lambda_i) = \lambda_i \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'(X'X)^2\hat{\beta}} \right)^{-1},$$

puis l'estimateur de Strawderman (1978)

$$\begin{aligned} \tilde{\beta} &= \left[ 1 + \frac{a\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right]^{-1} \hat{\beta} \\ &= P' \text{diag} \frac{1}{\lambda_i} \left( 1 + \frac{a\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right)^{-1} PX'Y \end{aligned}$$

où  $\lambda_i$  sera remplacée par

$$f_5(\lambda_i) = \lambda_i \left( 1 + \frac{a\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right).$$

Enfin, il y a l'estimateur de Stein à partie positive

$$\begin{aligned} \hat{\beta}_S &= \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta}} \right)_+ \hat{\beta} \\ &= P' \text{diag} \frac{1}{\lambda_i} \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta}} \right)_+ PX'Y, \end{aligned}$$



et dans ce cas on remplace  $\lambda_i$  par

$$f_6(\lambda_i) = \lambda_i \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'X'X\hat{\beta}} \right)_+^{-1},$$

où  $(C)_+ = \max(0, C)$ . Si  $Q = (X'X)^2$ , les estimateurs qui restent minimax par rapport à  $R(\tilde{\beta}; Q)$  sont celui de Srivastava et Bilodeau (1989)

$$\begin{aligned} \hat{\beta}_{SB} &= \left[ I_p - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'\hat{\beta}} (X'X)^{-1} \right] \hat{\beta} \\ &= P' \text{diag} \frac{1}{\lambda_i} \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\hat{\beta}'\hat{\beta}\lambda} \right) PX'Y, \end{aligned}$$

où l'on remplace  $\lambda_i$  par

$$f_7(\lambda_i) = \lambda_i \left( 1 - c \frac{\|Y - X\hat{\beta}\|^2}{\lambda_i \hat{\beta}'\hat{\beta}} \right)^{-1},$$

puis, celui de Strawderman (1978)

$$\begin{aligned} \tilde{\beta} &= \left[ I_p + \frac{a\|Y - X\hat{\beta}\|^2(X'X)^{-1}}{\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h} \right]^{-1} \hat{\beta} \\ &= P' \text{diag} \frac{1}{\lambda_i} \left( 1 + \frac{a\|Y - X\hat{\beta}\|^2}{\lambda_i (\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h)} \right)^{-1} PX'Y, \end{aligned}$$

où  $\lambda_i$  devient

$$f_8(\lambda_i) = \lambda_i \left( 1 + \frac{a\|Y - X\hat{\beta}\|^2}{\lambda_i (\hat{\beta}'X'X\hat{\beta} + g\|Y - X\hat{\beta}\|^2 + h)} \right).$$

Dans le cas où,  $h=0$ ,  $k=0,25$ ,  $c=0,032$ ,  $g=0,213$ ,  $\hat{\beta}'\hat{\beta} = 0,256$ ,  $\hat{\beta}'X'X\hat{\beta} = 2,815$ ,  $\|Y - X\hat{\beta}\|^2 = 48.794$  et  $Q = X'X$  (la matrice  $X$  est donnée à l'appendice B), nous avons calculé,  $f_1(\lambda_i) = 0,25 + \lambda_i$ ,  $f_4(\lambda_i) = 1,554\lambda_i$ ,  $f_5(\lambda_i) = 1,118\lambda_i$  et  $f_6(\lambda_i) = 2,235\lambda_i$ , où  $\lambda_i$  est un nombre réel positif. Ces fonctions correspondent respectivement aux estimateurs  $\beta^*(0,25) = P' \text{diag}(1/(\lambda_i+0,25))PX'Y$ ,  $\hat{\beta}_{SB} = P' \text{diag}(1/1,554\lambda_i)PX'Y$ ,  $\tilde{\beta} = P' \text{diag}(1/1,118\lambda_i)PX'Y$  et  $\hat{\beta}_S = P' \text{diag}(1/2,235\lambda_i)X'Y$ . Pour l'estimateur de Stein à partie positive,  $\hat{\beta}_S$  nous avons considéré  $Q = X'X$ , qui correspond au seul cas où cet estimateur est minimax. Par la suite, nous avons tracé les fonctions ci-dessus avec la fonction  $f_0(\lambda_i) = \lambda_i$ , correspondante à l'estimateur des moindres carrés;  $\hat{\beta} = P' \text{diag}(1/\lambda_i)PX'Y$  (voir figure 1.5). On peut remarquer d'après cette figure que les fonctions,  $f_1$ ,  $f_4$ ,  $f_5$  et  $f_6$  sont supérieures à la fonction  $f_0$ . Par conséquent, les éléments

des matrices,  $diag(1/f_1(\lambda_i))$ ,  $diag(1/f_4(\lambda_i))$ ,  $diag(1/f_5(\lambda_i))$  et  $diag(1/f_6(\lambda_i))$  sont inférieurs à  $1/\lambda_i$ , où  $i = 1, \dots, p$ . On peut déduire facilement que, la norme des estimateurs,  $\beta^*(0, 25)$ ,  $\hat{\beta}_{SB}$ ,  $\tilde{\beta}$  et  $\hat{\beta}_S$  dont les expressions sont données ci-dessus est inférieure à celle des moindres carrés. Dans cet exemple numérique, on remarque que ces estimateurs réduisent la norme de l'estimateur des moindres carrés.

(1): $f_0(\lambda_i)$ , (2): $f_5(\lambda_i)$ , (3): $f_1(\lambda_i)$ , (4): $f_4(\lambda_i)$ , (5): $f_6(\lambda_i)$

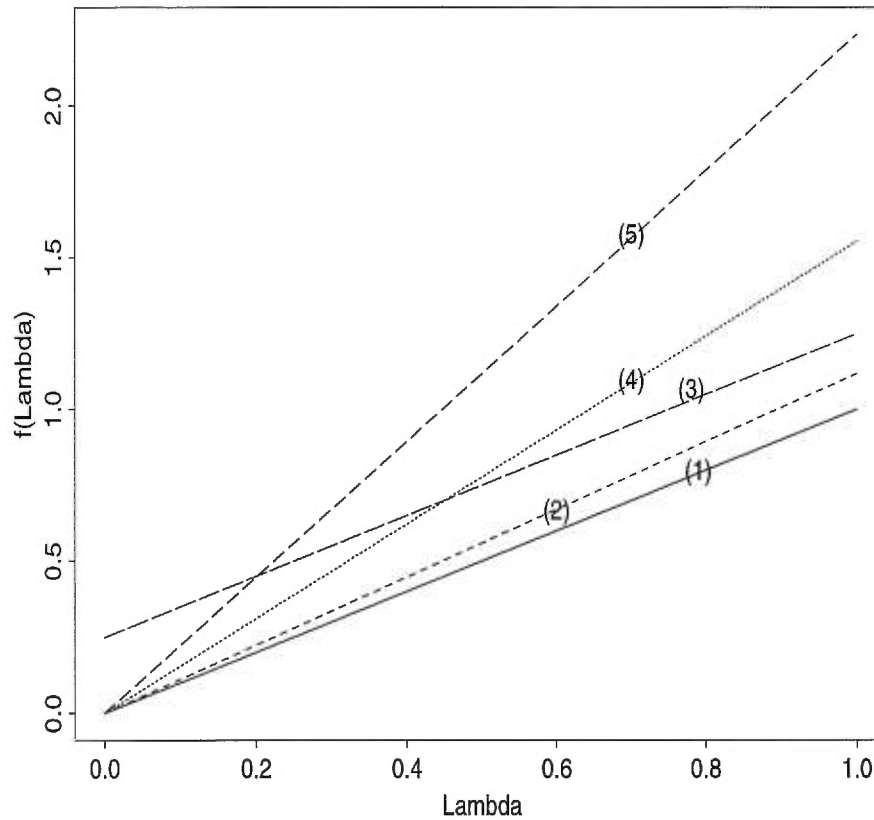


FIGURE 1.5. Les valeurs propres de la matrice  $A$  associées à chaque estimateur de la forme  $\tilde{\beta} = P'APX'Y$ .

## CHAPITRE 2

---

### Les méthodes de sélection du paramètre ridge

#### 2.1. INTRODUCTION

Lorsque  $X'X = I_p$ , on peut trouver à l'aide de (1.2.1) la valeur de  $k$  qui rend  $EQM(\beta^*)$  minimale,  $k = p\sigma^2/\beta'\beta$ . Dans le cas de l'estimateur ridge général  $\hat{\alpha}^* = (X^*X^* + K)^{-1}X^{*'}Y$ , où  $K = \text{diag}(k_i)$  avec  $k_i \leq 0$  et  $1 \leq i \leq p$ . La fonction d'erreur quadratique moyenne donnée par (1.2.2) est minimale lorsque  $k_i = \sigma^2/\alpha_i^2$ . Dans le cas général, une forme explicite est difficile à trouver et de toute façon elle dépend toujours de paramètres inconnus. Plusieurs méthodes faisant appel à des simulations ou à des algorithmes ont été proposées afin de trouver une solution numérique. Depuis 1970, il y a eu beaucoup de méthodes proposées pour choisir le paramètre ridge. Dans ce chapitre nous allons présenter les méthodes les plus utilisées.

#### 2.2. MÉTHODES DE SÉLECTION

##### 2.2.1. Méthode de validation croisée.

Considérons l'estimateur de  $X\beta$ , à l'aide des observations  $Y$ , par rapport à la fonction de perte

$$L(k) = \frac{1}{n} \|X\beta^*(k) - X\beta\|^2$$

et soit  $R(k) = EL(k)$  la fonction de risque. Une mesure de précision étroitement liée à  $R(k)$  est le risque de prévision. Supposons que l'on dispose de  $n$  nouvelles observations  $Y^*$  du même modèle,  $Y^* = X\beta + \varepsilon^*$ , indépendantes de  $Y$ . On peut évaluer la performance de  $X\beta^*$  pour prévoir les nouvelles observations  $Y^*$  par le risque de prévision

$$P(k) = E \frac{1}{n} \|Y^* - X\beta^*\|^2.$$

Par l'indépendance de  $Y^*$  et  $Y$  on obtient immédiatement

$$P(k) = \sigma^2 + R(k).$$

Le problème revient donc à minimiser  $P(k)$  ou à trouver l'estimateur  $\hat{R}(k)$ . Nous verrons maintenant comment obtenir  $\hat{P}(k)$ . Soit

$$EQM(k) = \frac{1}{n} \|Y - X\beta^*\|^2$$

l'erreur quadratique moyenne. Un estimateur sans biais de  $P(k)$  [Eubank (1988), p. 25] lorsque  $\sigma^2$  est connue est

$$\hat{P}(k) = EQM(k) + \frac{2}{n} \sigma^2 \text{tr} H$$

où  $H = X(X'X + kI_p)^{-1}X'$ . Allen (1971, 1974) a proposé la méthode "PRESS" ("prediction sum of squares") communément appelée validation croisée qui peut servir à sélectionner le paramètre ridge  $k$  sans avoir à connaître ou à estimer  $\sigma^2$ . L'idée est inspirée du risque de prévision  $P(k)$ , mais plutôt que d'utiliser de nouvelles observations dont on ne dispose pas on fait appel à des sous-échantillons. La méthode consiste à enlever une observation, puis à estimer le modèle à partir des  $n - 1$  autres observations. Par la suite, on fait une prévision pour l'observation manquante, en utilisant chaque valeur de  $k$  choisie sur un treillis. On répète cette procédure en enlevant à tour de rôle chacune des observations. La valeur de  $k$  optimale est celle pour laquelle l'erreur quadratique moyenne de prévision est minimale. Le critère de validation croisée consiste à minimiser la fonction

$$CVC(k) = \frac{1}{n} \sum_{j=1}^n [Y_j - X'_j \beta^*(j)]^2$$

où  $\beta^*(j)$  est l'estimateur ridge calculé sans l'observation  $Y_j$  et  $X'_j$  est la  $j$ -ième ligne de la matrice  $X$ . Il faut noter que  $CVC(k)$  est un estimateur biaisé de  $P(k)$  car on utilise  $n - 1$  plutôt que  $n$  observations pour faire la prévision. L'évaluation directe du

*CVC* nous oblige à ajuster  $n$  modèles pour chaque valeur de  $k$  et est numériquement très laborieuse. L'identité du *CVC* donne l'expression équivalente [Eubank (1988), p.30]

$$CVC(k) = \frac{1}{n} \sum_{i=1}^n [e_i(k)]^2 / [1 - h_{ii}(k)]^2$$

où  $e_1(k), \dots, e_n(k)$  désignent les résidus de la régression ridge et  $h_{ii}(k)$  sont les éléments diagonaux de la matrice  $H$ . A l'aide de cette identité l'évaluation du *CVC* pour une valeur de  $k$  donnée nécessite l'ajustement d'un seul modèle et est numériquement beaucoup plus avantageuse. En utilisant la même matrice  $X$  et le vecteur  $Y$  de l'appendice B et C, on a obtenu le graphe de la figure 2.1 suivant, qui montre bien un minimum dans cet exemple numérique. On note également une asymptote

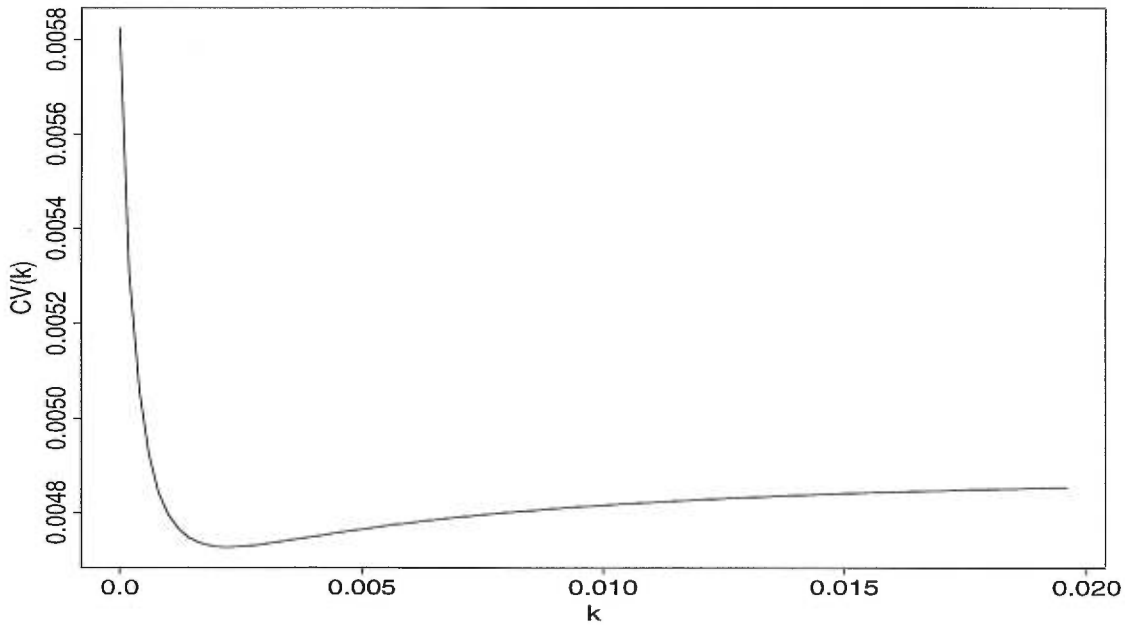


FIGURE 2.6. La fonction  $CVC(k)$

horizontale car

$$\lim_{k \rightarrow \infty} CVC(k) = \frac{1}{n} \sum_{i=1}^n y_i^2.$$

Mentionnons finalement que, dans le modèle linéaire, on peut écrire

$$CVC(k) = \frac{1}{n} Y'(I_n - H) \text{diag} [(1 - h_{ii}(k))^{-2}] (I_n - H) Y.$$

Cette fonction n'est pas invariante par transformation orthogonale ( $Y \mapsto \Gamma Y$ ,  $X \mapsto \Gamma X$  où  $\Gamma$  est une matrice orthogonale). Une version invariante appelée *GCV* fait l'objet de la prochaine section.

### 2.2.2. Méthode de validation croisée généralisée.

Golub, Heath et Wahba (1979) ont utilisé la méthode de validation croisée généralisée pour choisir le paramètre ridge. Cette méthode est une version invariante de la validation croisée. Au lieu de minimiser  $EQM(k)$ , on minimise la fonction

$$GCV(k) = EQM(k) / \left[ n^{-1} \text{tr}(I_n - H) \right]^2.$$

La meilleure motivation pour le critère *GCV* se retrouve dans Craven et Wahba (1979). On peut démontrer dans le cas de la régression ridge [Eubank (1988), p. 93]

$$|E(GCV(k)) - P(k)| / R(k) \leq 3p/n + O\left((p/n)^2\right).$$

Donc le biais relatif du *GCV* comme estimateur du risque de prévision  $P(k)$  est négligeable pour les grandes valeurs de  $n$ . Dans le contexte bayésien sous les conditions  $Y \sim N(X\beta, \sigma^2 I_n)$  et  $\beta$  suit la distribution a priori  $\beta \sim N(0, \sigma_s^2 I_p)$ , Golub, Heath et Wahba (1979) ont montré que  $E_\beta \{E[GCV(k)]\}$  et  $E_\beta [R(k)]$  sont toutes deux minimales pour  $k = \sigma^2 / \sigma_s^2$ . En se rappelant que  $P(k) = R(k) + \sigma^2$ , même si  $GCV(k)$  est généralement biaisé pour  $P(k)$ , il est quand même surprenant qu'après avoir calculé la moyenne sur la distribution a priori, les deux fonctions ont leur minimum au même point. Dans une décomposition singulière, posons  $X = UDV'$  et  $Z = U'Y$  où  $U : n \times n$  et  $V : p \times p$  sont orthogonales et  $D : n \times p$  est de la forme

$$\begin{pmatrix} \text{diag}(\lambda_i^{1/2}) \\ 0 \end{pmatrix}.$$

On a alors,

$$\begin{aligned} GCV(k) &= \frac{1/n \sum_{i=1}^n [nk / (\lambda_i + nk)]^2 Z_i^2}{[1/n (\sum_{i=1}^p nk / (\lambda_i + nk) + n - p)]^2} \\ &= \frac{\left(1/n \sum_{i=1}^p [nk / (\lambda_i + nk)]^2 Z_i^2 + \frac{1}{n} \sum_{i=p+1}^n Z_i^2\right)}{[1/n (\sum_{i=1}^p nk / (\lambda_i + nk) + n - p)]^2} \end{aligned}$$

où les  $\lambda_i$  sont les valeurs propres de la matrice  $X'X$  et  $\lambda_i = 0$  pour tout  $p < i \leq n$ . On peut montrer facilement cette décomposition et elle se trouve aussi dans Golub, Heath et Wahba (1979). La fonction  $GCV(k)$  est une fonction continue et bornée avec

$$GCV(0) = \frac{n}{(n-p)^2} \sum_{i=p+1}^n Z_i^2$$

et

$$\lim_{k \rightarrow \infty} GCV(k) = \frac{1}{n} \sum_{i=1}^n Z_i^2.$$

En utilisant  $X$  et  $Y$  de l'annexe A, on obtient le graphe 2.2, qui montre l'existence d'un minimum de la fonction  $GCV(k)$  dans cet exemple. Dans Eubank (1988), il y

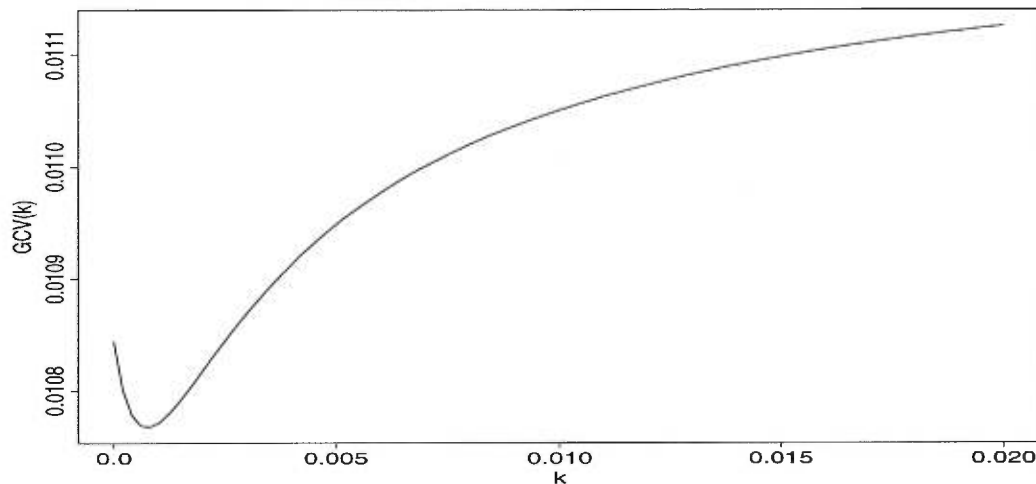


FIGURE 2.7. La fonction  $GCV(k)$

a d'autres critères qui sont à peu près équivalents lorsque  $tr(H)/n$  est petit, ce qui est le cas dans la régression ridge si  $n$  est grand. On peut aussi citer d'autres critères de sélection  $k$  [Eubank (1988)]:

$$\begin{aligned} AIC(k) &= n \log EQM(k) + 2P_k, \\ FPE(k) &= \frac{(n + P_k)}{(n - P_k)} EQM(k), \\ S(k) &= (n + 2P_k) EQM(k), \\ U(k) &= \frac{n}{n - P_k} \left( 1 + \frac{P_k}{n - 1 - P_k} \right) EQM(k) \end{aligned}$$

avec  $P_k = \text{tr}(H)$ . Quand  $\text{tr}(H)/n$  est petit

$$\begin{aligned} \exp\{-P_k/n\} &\simeq 1 - \frac{P_k}{n} \\ &= \frac{1}{n} \text{tr}(I_n - H), \\ \frac{n + P_k}{n - P_k} &= \frac{1 + P_k/n}{1 - P_k/n} \simeq \frac{1}{(1 - P_k/n)^2}, \\ n + 2P_k &= n(1 + 2P_k/n) \simeq n/(1 - P_k/n)^2, \\ \frac{n}{n - P_k} \left(1 + \frac{P_k}{n - 1 - P_k}\right) &\simeq \frac{1}{(1 - P_k/n)^2}. \end{aligned}$$

Donc on observe les approximation suivantes

$$\begin{aligned} \exp\{AIC(k)/n\} &= EQM(k) \exp\{2P_k/n\} \\ &= \frac{EQM(k)}{(\exp\{-P_k/n\})^2} \\ &\simeq GCV(k), \\ FPE(k) &\simeq \frac{EQM(k)}{(1 - P_k/n)^2} = GCV(k), \\ S(k) &\simeq \frac{nEQM(k)}{(1 - P_k/n)^2} = nGCV(k), \\ U(k) &\simeq \frac{EQM(k)}{(1 - P_k/n)^2} = GCV(k). \end{aligned}$$

### 2.2.3. Le critère de Mallows.

Comme à la section 1.2.3 soit  $Y = \gamma 1_n + X\beta + \varepsilon$ , où  $X$ :  $n \times p$  exclut l'ordonnée à l'origine et  $X'X$  est une matrice de corrélation. Mallows a considéré la fonction de perte suivante

$$J_L = \frac{1}{\sigma^2} \left[ \|X\beta^* - X\beta\|^2 + n(\bar{Y} - \gamma)^2 \right]$$

et la fonction de risque

$$EJ_L = 1 + \text{tr}(X'XLL') + \frac{1}{\sigma^2} \beta'(LX - I_p)'X'X(LX - I_p)\beta$$

où  $L = (X'X + kI_p)^{-1}X'$  avec  $L1_n = 0$  où  $1'_n = (1, 1, \dots, 1)$ . Dans Mallows (1973, p. 672) un estimateur  $C_L$  de  $EJ_L$  est

$$C_L = \frac{1}{\hat{\sigma}^2} RSS_L - n + 2 + 2\text{tr}(XL)$$

qui définit la statistique de Mallows où  $RSS_L = \|Y - \bar{Y}1_n - X\beta^*\|^2$ . Le critère consiste à minimiser la fonction  $C_L$  en fonction de  $k$ .



Considérons la décomposition  $X'X = P'DP$ , où  $P$  est une matrice orthogonale,  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$  et  $Z = PX'Y$ , nous avons

$$RSS_L(k) - RSS_L(0) = \sum_{i=1}^p k^2 Z_i^2 / [\lambda_i(k + \lambda_i)^2].$$

De plus on a  $C_L(0) = 2p - n + 2 + \frac{1}{\hat{\sigma}^2} RSS_L(0)$  et

$$\lim_{k \rightarrow \infty} C_L = 2 - n + \frac{1}{\hat{\sigma}^2} \left( \sum_{i=1}^p Z_i^2 / \lambda_i + RSS_L(0) \right).$$

Hoerl et Kennard (1970b), partant des données de 10 variables analysées par Gorman et Toman (1966), ont trouvé que pour toute valeur de  $k$  appartenant à l'intervalle  $(0, 2; 0, 3)$ , les estimateurs des coefficients seront plus proches de  $\beta$ .

Comme cas particulier d'une telle situation, si on considère le cas où les variables explicatives sont orthogonales, c'est-à-dire  $X'X = I_p$ , alors la fonction  $EJ_L$  devient  $\sum_{i=1}^p E(\hat{\beta}_i - \beta_i)^2$  qui est la fonction d'erreur quadratique moyenne utilisée par Hoerl et Kennard (1970a). Dans ce même cas  $P = I_p$  et  $X'X = D = I_p$  et donc  $Z = \hat{\beta} = X'Y$  qui est l'estimateur des moindres carrés. Mallows indique aussi que le paramètre ridge vérifie la relation

$$(1 + k)/k = \sum_{i=1}^p \beta_i^2 / (p\hat{\sigma}^2)$$

et l'estimateur ajusté proposé est donc

$$\left[ 1 - k\hat{\sigma}^2 / \hat{\beta}'\hat{\beta} \right] \hat{\beta}$$

qui est la forme suggérée par Stein (1956). Dans le deuxième cas où les régresseurs sont tels que:

$$X'X = I_p + \rho(1_n 1_n' - I_p)$$

c'est-à-dire le cas d'équicorrélation, où  $\rho$  est le coefficient de corrélation. Lorsque ce dernier est proche de l'unité,  $C_L$  est minimisé pour  $k$  proche de  $(1 - \rho)g$ , où  $g$  vérifie la relation suivante

$$\frac{1 + g}{g} = \frac{1 - \rho}{(p - 1)\hat{\sigma}^2} \sum_{i=1}^p (\hat{\beta}_i - \hat{\bar{\beta}})^2$$

où

$$\hat{\bar{\beta}} = \sum_{i=1}^p \hat{\beta}_i / p.$$

Dans ce cas l'estimateur ajusté de ridge est donné approximativement par

$$\beta^* = \frac{1_n' \hat{\beta}}{n} 1_n + \left( 1 - \frac{g}{1 + g} \right) \left[ \hat{\beta} - \frac{1_n' \hat{\beta}}{n} 1_n \right],$$

qui est un estimateur de réduction en norme de  $\hat{\beta}$  vers  $\hat{\beta}1_n$ . On peut montrer facilement et cela se trouve aussi dans Golub, Heath et Wahba (1979, p. 216) que la minimisation de  $C_L$  est équivalente à minimiser

$$\frac{n\hat{R}(k)}{\hat{\sigma}^2}$$

où

$$\hat{R}(k) = \frac{1}{n} \|(I_n - H)Y\|^2 - 2\frac{\hat{\sigma}^2}{n} \text{tr}(I_n - H) + \hat{\sigma}^2$$

est l'estimateur sans biais de la fonction de risque  $R(k) = E\frac{1}{n}\|X\beta^* - X\beta\|^2$  avec

$$\hat{\sigma}^2 = \frac{1}{n-p} \|(I_n - X(X'X)^{-1}X')Y\|^2$$

et

$$H = X(X'X + kI_p)^{-1}X'.$$

On note que l'estimateur ridge est un cas particulier pour le critère de Mallows car ce dernier traite les estimateurs linéaires en général.

#### 2.2.4. Méthode de McDonald et Galarneau.

Nous nous sommes basés sur la fonction d'erreur quadratique moyenne pour trouver le meilleur estimateur ridge, mais cette fonction dépend de  $\sigma^2$  et de  $\beta$ . L'estimateur de  $\beta$  dépend des données et du degré de multicolinéarité. Pour une valeur de  $k$ , fixe, on peut facilement trouver une valeur de  $\beta$  pour laquelle  $EQM(\beta^*(k)) \geq EQM(\hat{\beta})$ . Alors, aucune valeur constante de  $k$  ne peut assurer que l'estimateur ridge est meilleur que l'estimateur des moindres carrés. McDonald et Galarneau (1975) ont proposé un critère de sélection de  $k$ , basé sur la minimisation de la somme des carrés résiduels sous une contrainte portant sur la norme de l'estimateur. Pour une telle valeur de  $k$ , la norme de l'estimateur ridge est plus petite que celle de l'estimateur des moindres carrés. Ils ont fait des simulations basés sur la norme de l'estimateur. En posant,  $G = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 \sum_{i=1}^p 1/\lambda_i$ , où  $\hat{\beta}$  et  $\hat{\sigma}^2$  sont les estimateurs des moindres carrés, on a  $E(G) = \beta'\beta$ . Il en découle que  $G$  est un estimateur sans biais du carré de la norme de  $\beta$ . Ils se sont basés sur trois règles:  $R_0$ ,  $R_1$  et  $R_2$ . La première, considérée comme "optimale", nécessite la connaissance de  $\beta$ , consiste à déterminer une valeur de  $k$  qui minimise  $\sum_{i=1}^p [\beta_j - \beta_j^*(k)]^2$ , pour une valeur de  $\beta$  donnée.

La deuxième consiste à choisir un  $k$  tel que

$$\beta^*(k)' \beta^*(k) = \begin{cases} G & \text{si } G > 0 \\ \hat{\beta}' \hat{\beta} & \text{sinon.} \end{cases} \quad (2.2.1)$$

La troisième consiste à choisir un  $k$  tel que

$$\beta^*(k)' \beta^*(k) = \begin{cases} G & \text{si } G > 0 \\ 0 & \text{sinon.} \end{cases} \quad (2.2.2)$$

Si  $G > 0$  alors  $R_1$  et  $R_2$  définissent le même estimateur ridge. Cependant, si  $G \leq 0$  la règle  $R_1$  résulte en l'estimateur des moindres carrés ( $k = 0$ ) alors que la règle  $R_2$  estime  $\beta$  par le vecteur nul ( $k = \infty$ ). McDonald et Galarneau ont généré des données de taille  $n = 100$  et  $p = 3$ . Les variables explicatives sont calculées par l'équation  $x_{ij} = (1 - \alpha^2)^{1/2} z_{ij} + \alpha z_{i4}$  avec  $1 \leq i \leq n$ ,  $1 \leq j \leq p$ ,  $0 \leq \alpha \leq 1$  et  $z_{i1}, z_{i2}, z_{i3}, z_{i4}$  des variables d'une loi normale centrée réduite. Ces variables sont ensuite standardisées de sorte que  $X'X$  soit une matrice de corrélation.

Pour le choix du paramètre  $\beta$ , ces auteurs se sont basés sur Newhouse et Oman (1971) qui ont choisi un  $\beta$  de norme égale à un. Dans cette situation Newhouse et Oman (1971) ont montré que la fonction d'erreur quadratique moyenne  $EQM$  est minimale si  $\beta$  est le vecteur propre associé à la plus grande valeur propre de la matrice  $X'X$ . Le vecteur propre correspondant est noté  $\beta_M$ . De même  $EQM$  est maximal si  $\beta$  est le vecteur propre associé à la plus petite valeur propre. On désigne ce vecteur par  $\beta_m$ . Les variables dépendantes sont construites par le modèle

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_i$$

où  $\varepsilon_i \sim N(0, \sigma^2)$  et  $i = 1, \dots, n$ .

L'erreur quadratique moyenne de l'estimateur ridge a été calculée pour 201 valeurs de  $k$ , en respectant la règle  $R_1$  et  $R_2$ . Par la suite, ils ont défini  $M$  comme le rapport de  $EQM(\beta^*)$  à  $EQM(\hat{\beta})$  pour voir la performance relative de l'estimateur ridge par rapport à l'estimateur des moindres carrés. Ils ont étudié l'évolution de ce rapport en fonction de  $\sigma^2$  et  $\alpha$ , en fixant l'un des deux et en variant l'autre et ce pour les deux valeurs de  $\beta$ ,  $\beta_M$  et  $\beta_m$ . Pour toutes les valeurs de  $\alpha$  et  $\sigma^2$  il ressort dans le cas par McDonald et Galarneau (1975) que la règle  $R_0$  domine les règles  $R_1$  et  $R_2$ . En effet, le gain en terme d'erreur quadratique moyenne  $EQM$  en utilisant  $R_0$  est de l'ordre

de 0,1% jusqu'à 87,5%. Lorsqu'on applique les règles  $R_1$  et  $R_2$ , l'estimateur ridge est toujours meilleur que l'estimateur des moindres carrés sauf quand pour des petites valeurs de  $\sigma^2$  lorsque  $\beta$  est fixé à  $\beta = \beta_m$ .

### 2.2.5. La méthode du bootstrap.

Dans cette section, nous allons présenter la méthode de bootstrap pour sélectionner le paramètre ridge. Cette méthode se trouve dans Delaney et Chatterjee (1986). L'idée est de minimiser l'erreur de prévision en utilisant des échantillons 'bootstrap'. Le vecteur des résidus de la régression,  $\hat{\epsilon} = Y - X\hat{\beta}$ , est calculé afin d'obtenir une approximation de la distribution  $F$  du vecteur  $\epsilon$ . Soit  $\hat{\epsilon} = \hat{\epsilon} - \bar{\hat{\epsilon}}$ , le vecteur des résidus centré, avec  $\bar{\hat{\epsilon}} = (1/n) \sum_{i=1}^n \hat{\epsilon}_i$ . La distribution  $F$  de  $\epsilon$  est approximée par  $\hat{F}_n(x) = (1/n) \sum_{i=1}^n I(\hat{\epsilon}_i \leq x)$ , la distribution expérimentale des résidus centrés. Par la suite, on génère des vecteurs  $\epsilon^*$  de la distribution expérimentale  $\hat{F}_n(x) = (1/n) \sum_{i=1}^n I(\hat{\epsilon}_i \leq x)$ . On obtient le modèle bootstrap  $Y^* = X\hat{\beta} + \epsilon^*$ , où  $\hat{\beta}$  est l'estimateur des moindres carrés. En effet,  $\epsilon^*$  est un vecteur aléatoire provenant d'un tirage simple avec remise parmi les résidus centrés  $\hat{\epsilon}$ . À partir de ce modèle bootstrap, nous calculons un estimé bootstrap de  $\hat{\beta}$ , le paramètre du modèle bootstrap, donné par

$$\hat{\beta}^*(k_g) = (X'X + k_g I_p)^{-1} X'Y^*,$$

avec  $k_g \geq 0$  et  $g = 1, \dots, G$  ( $G$  est un entier naturel fixe) formant un treillis. On génère  $B$  échantillons bootstrap ( $B$  grand), et pour chaque échantillon  $j$  ( $j = 1, \dots, B$ ) on calcule le vecteur de prévision  $\hat{Y}_{L_j}(k_g) = X_{fut} \hat{\beta}^*(k_g)$  pour  $L_j$  observations non choisies. La matrice  $X_{fut}$  dénote les lignes de  $X$  correspondantes aux observations non choisies. On définit l'erreur quadratique moyenne de prévision par

$$EQM_j(k_g) = (\hat{Y}_{L_j}(k_g) - Y_{L_j})' (\hat{Y}_{L_j}(k_g) - Y_{L_j}) / L_j.$$

On répète cette procédure  $B$  fois puis on calcule la moyenne pondérée sur tous les échantillons. L'erreur quadratique moyenne pondérée est définie par

$$EQMP(k_g) = \sum_{j=1}^B (EQM_j(k_g)) (L_j) / \sum_{j=1}^B L_j.$$

L'estimateur bootstrap de  $k$  est donné par  $k_{boot} = \inf_{k_g} EQMP(k_g)$  (le minimum de  $EQMP(k_g)$  sur tous les  $k_g$ ).

## CHAPITRE 3

---

### Comparaison entre certains estimateurs

#### 3.1. INTRODUCTION

Dans les chapitres précédents, nous avons présenté des estimateurs de  $\beta$  et des méthodes pour choisir le paramètre ridge  $k$ . Dans ce chapitre, nous allons exposer le plan des simulations et une méthode numérique pour le choix du paramètre ridge. Nous allons étudier les résultats de plusieurs simulations afin de comparer les estimateurs de Stein à partie positive [Baranchik (1970)], Strawderman (1978), et celui de Srivastava et Bilodeau (1989) avec l'estimateur des moindres carrés. Nous allons étudier aussi l'estimateur de  $\sigma^2$ .

#### 3.2. SIMULATIONS

##### 3.2.1. Plan des simulations.

On estime le risque d'un estimateur  $\tilde{\beta}$ ,  $R(\tilde{\beta}; Q) = E[(\tilde{\beta} - \beta)'Q(\tilde{\beta} - \beta)]$ , par  $\hat{R}(\tilde{\beta}; Q) = \frac{1}{m} \sum_{i=1}^m (\tilde{\beta}_i - \beta)'Q(\tilde{\beta}_i - \beta)$  par rapport à la fonction de perte  $L(\tilde{\beta}; Q) = (\tilde{\beta} - \beta)'Q(\tilde{\beta} - \beta)$ . La matrice  $Q$  prend les valeurs  $I_p, X'X, (X'X)^2$  et  $\tilde{\beta}_i = f(X, Y_i)$ , où  $Y_i = X\beta + \varepsilon_i$ ,  $\varepsilon_i \sim N(0, \sigma^2 I_{50})$ ,  $i = 1, \dots, m$ ,  $f$  est une fonction donnée suivant l'estimateur utilisé et  $m$  est un entier naturel fixe désignant le nombre de simulations. Par

la suite, on compare  $\hat{R}(\tilde{\beta}; Q)$  avec le risque de l'estimateur des moindres carrés,  $R(\hat{\beta}; Q) = \sigma^2 \text{tr}[Q(X'X)^{-1}]$ . Nous allons calculer le risque estimé  $\hat{R}(\tilde{\beta}; Q)$  pour chaque estimateur, et cela pour différentes valeurs de  $\beta$ ,  $Q$ ,  $\sigma^2$  et de  $S(\beta)$ , le rapport du signal au bruit blanc défini un peu plus loin. À cet effet, nous avons écrit un programme pour calculer ce risque pour différentes valeurs de  $\beta$ ,  $S(\beta)$  et de  $\sigma^2$  (voir appendice D). Pour des valeurs données de  $\beta$ ,  $Q$ ,  $S(\beta)$  et  $\sigma^2$  ce programme génère  $m$  fois le vecteur des erreurs,  $\varepsilon$ , d'une loi normale  $N(0, \sigma^2 I_{50})$ . Pour chacune des  $m$  générations, il calcule le vecteur  $Y = X\beta + \varepsilon$  puis la perte  $L(\tilde{\beta}, Q)$  des estimateurs de Strawderman, Srivastava et Bilodeau et de Stein à partie positive. On a répété le même calcul pour différentes valeurs de  $Q$ ,  $\beta$ ,  $S(\beta)$  et  $\sigma^2$ . Les valeurs de  $\beta$  sont construites en fixant la valeur du rapport du signal au bruit blanc défini par

$$S(\beta) = \frac{\beta' X' X \beta}{\sigma^2}$$

et la valeur de  $\sigma^2$ .

Le risque de l'estimateur ridge est estimé en utilisant le même programme. La difficulté additionnelle est que nous devons sélectionner le paramètre ridge,  $k$ , par la méthode de validation croisée généralisée et ce pour chacune des  $m$  générations. Cette méthode fait l'objet de la prochaine section.

### 3.2.2. Le choix du paramètre ridge.

Le paramètre ridge est sélectionné par la méthode de la validation croisée généralisée. Cette méthode a été étudiée à la section 2.2.2. Nous avons écrit un programme qui minimise globalement la fonction  $GCV(k)$  (voir appendice E). Plus précisément, ce dernier évalue la fonction  $GCV(k)$  sur un treillis de plusieurs valeurs de  $k \geq 0$ . Ensuite, il prend  $k = k_0$  tel que  $GCV(k_0) \leq GCV(k)$  comme valeur initiale de l'algorithme de Newton Raphson. Ce dernier algorithme donne la valeur précise de  $k_0$  qui minimise la fonction  $GCV(k)$ . Cette minimisation a été faite pour chacune des  $m$  générations. Par la suite, nous avons calculé  $L(\beta^*(k); Q)$  et  $\hat{R}(\beta^*(k); Q)$ . Ce calcul est répété pour différentes valeurs de  $Q$ ,  $\sigma^2$ ,  $S(\beta)$  et  $\beta$ .

### 3.2.3. Construction des variables explicatives.

La matrice des variables explicatives que nous avons utilisée dans ce mémoire a été construite comme suit. Les trois premiers vecteurs colonnes,  $X_1$ ,  $X_2$  et  $X_3$  ont été générés d'une loi normale  $N(0, I_{50})$  et le quatrième vecteur a été calculé en fonction des ces vecteurs,  $X_4 = X_1 + 3,5X_2 + 7X_3 + 1,51_{50}$ , où  $1_{50} = (1, \dots, 1)$  est un vecteur  $50 \times 1$ . Cette matrice est désignée par  $X$  (voir appendice B). Nous avons choisi cette matrice de telle sorte que le rapport entre les valeurs propres maximale et minimale de la matrice  $(X'X)$  soit d'environ à 1000. Ceci est un bon exemple de multicollinéarité. En effet, les valeurs propres de la matrice  $(X'X)$  sont  $\lambda_1 \simeq 1,647$ ,  $\lambda_2 \simeq 17,832$ ,  $\lambda_3 \simeq 30,807$  et  $\lambda_4 \simeq 1689,882$ . Alors le rapport des valeurs propre est

$$\frac{\lambda_{\max}}{\lambda_{\min}} \simeq 1026,036.$$

### 3.2.4. Résultats.

Toutes les simulations seront effectuées pour  $m = 1000$ . Les vecteurs  $\beta_1, \dots, \beta_{13}$  sont présentés à l'appendice C. En premier lieu, les simulations ont été faites pour  $\sigma^2 = 0, 1$ ,  $S(\beta) = 0, 1$ ,  $\beta = \beta_1, \beta_2, \beta_3, \beta_4$  et pour différentes valeurs de  $Q$ . Nous avons calculé aussi bien le risque estimé de chaque estimateur que la variance échantillonnale du risque de l'estimateur par

$$\hat{V}ar(L(\tilde{\beta}; Q)) = \frac{1}{m-1} \left[ \sum_{i=1}^m L_i^2 - \frac{(\sum_{i=1}^m L_i)^2}{m} \right],$$

où  $L_i = (\tilde{\beta}_i - \beta)'Q(\tilde{\beta}_i - \beta)$ . Les tableaux 3.1, 3.2 et 3.3 représentent ces résultats. Dans chaque tableau, la valeur entre parenthèses représente l'écart type échantillonnale du risque de l'estimateur, et l'autre valeur représente le risque estimé de l'estimateur. Le risque minimax pour les différentes valeurs de  $Q$  est  $R(\hat{\beta}; I_4; \sigma^2) = 0,0696$ ,  $R(\hat{\beta}; X'X; \sigma^2) = 0,4$ ,  $R(\hat{\beta}; (X'X)^2; \sigma^2) = 174,017$  pour  $\sigma^2 = 0, 1$ . Les risques estimés vont être comparés avec ces risques minimax. En deuxième lieu, nous avons fixé  $\sigma^2 = 1$ ,  $S(\beta) = 0; 0, 1; 1; 10; 100$ , et  $\beta = \beta_5, \dots, \beta_9$ . Les risques minimax sont:  $R(\hat{\beta}; I_4; \sigma^2) = 0,696$ ,  $R(\hat{\beta}; X'X; \sigma^2) = 4$ ,  $R(\hat{\beta}; (X'X)^2; \sigma^2) = 1740,17$ . Les résultats obtenus sont présentés dans les tableaux 3.4, 3.5 et 3.6. En dernier lieu, nous avons fixé  $\sigma^2 = 50$ ,  $S(\beta) = 100$  et  $\beta = \beta_{10}, \dots, \beta_{13}$ . Pour les différentes valeurs de  $Q$ , nous avons calculé les risques estimés des estimateurs et le risque minimax. On trouve  $R(\hat{\beta}; I_4; \sigma^2) = 34,8$ ,

$R(\hat{\beta}; X'X; \sigma^2) = 200$ ,  $R(\hat{\beta}; (X'X)^2; \sigma^2) = 87008,48$ . Nous avons présenté les résultats dans les tableaux 3.7, 3.8 et 3.9. Pour étudier la performance de ces estimateurs, nous avons présenté les distributions expérimentales des composantes des estimateurs obtenus pour les  $m$  générations, en fixant certaines valeurs de  $\beta$ ,  $\sigma^2$ ,  $S(\beta)$  et  $Q$  (voir les figures 3.9 à 3.16).

Le paramètre  $\sigma^2$  a été estimé par  $\hat{\sigma}^2 = \frac{\|Y - X\hat{\beta}\|^2}{n-p}$ . Nous avons présenté aussi, la distribution expérimentale de cet estimateur pour certaines valeurs de  $\sigma^2$ ,  $S(\beta)$ ,  $\beta$  et cela pour  $m$  générations (voir la figure 3.8).

### 3.2.5. Conclusion.

Nous avons estimé le risque des estimateurs par rapport à la fonction de perte  $L(\tilde{\beta}; Q) = (\tilde{\beta} - \beta)'Q(\tilde{\beta} - \beta)$ . Les tableaux 3.1 à 3.9 montrent que le risque de tous les estimateurs est inférieur au risque minimax, sauf pour l'estimateur de Stein à partie positive dans le cas où  $Q = (X'X)^2$ ,  $\beta = \beta_8$ ,  $\sigma^2 = 1$  et  $S(\beta) = 10$ . Cet effet s'explique par le fait que cet estimateur est minimax uniquement dans le cas où  $Q = X'X$ . Nous remarquons que dans tous les cas étudiés, le risque estimé de l'estimateur ridge est inférieur à celui des autres estimateurs.

Les figures 3.9 à 3.16 montrent les distributions expérimentales des composantes de chaque estimateur pour certaines valeurs de  $\sigma^2$ ,  $\beta$ , et  $S(\beta)$ . Les distributions expérimentales des composantes de chaque estimateur que nous avons présentées nous montre que l'étendue de la distribution expérimentale de chaque composante de l'estimateur ridge est plus petite que l'étendue des autres distributions correspondantes.

On remarque aussi que ces distributions sont centrées par rapport à la valeur réelle. Nous pouvons alors conclure que de tous les estimateurs étudiés dans ce travail, l'estimateur ridge performe le mieux en présence de forte multicolinéarité.

Les distributions expérimentales de  $\hat{\sigma}^2$  présentées à la figure 3.8 sont également centrées par rapport aux valeurs réelles.



L'estimateur	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
Ridge	0,012 (0,003)	0,018 (0,005)	0,01 (0,002)	0,015 (0,006)
Bilodeau	0,026 (0,002)	0,027 (0,002)	0,025 (0,001)	0,028 (0,002)
Strawderman	0,066 (0,003)	0,070 (0,003)	0,068 (0,002)	0,070 (0,003)
Stein à partie positive	0,064 (0,002)	0,068 (0,003)	0,065 (0,002)	0,067 (0,003)

TABLEAU 3.1. Les risques estimés des estimateurs, pour  $\sigma^2 = 0,1$ ,  $S(\beta) = 0,1$ ,  $Q = I_4$  et  $\beta = \beta_1, \beta_2, \beta_3, \beta_4$

L'estimateur	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
Ridge	0,034 (0,008)	0,022 (0,006)	0,033 (0,007)	0,018 (0,005)
Bilodeau	0,171 (0,007)	0,154 (0,007)	0,159 (0,006)	0,157 (0,007)
Strawderman	0,313 (0,008)	0,298 (0,007)	0,301 (0,007)	0,298 (0,007)
Stein à partie positive	0,213 (0,009)	0,206 (0,011)	0,213 (0,011)	0,212 (0,009)

TABLEAU 3.2. Les risques estimés des estimateurs, pour  $\sigma^2 = 0,1$ ,  $S(\beta) = 0,1$ ,  $Q = X'X$  et  $\beta = \beta_1, \beta_2, \beta_3, \beta_4$

L'estimateur	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
Ridge	7,193 (1,358)	10,018 (2,479)	13,351 (2,352)	24,630 (8,076)
Bilodeau	80,115 (4,734)	76,898 (4,857)	70,674 (4,422)	73,840 (4,5412)
Strawderman	178,990 (7,864)	177,054 (7,773)	167,154 (7,354)	169,490 (7,348)
Stein à partie positive	181,843 (7,733)	174,250 (7,700)	164,578 (7,255)	167,135 (7,271)

TABLEAU 3.3. Les risques estimés des estimateurs, pour  $\sigma^2 = 0,1$ ,  $S(\beta) = 0,1$ ,  $Q = (X'X)^2$  et  $\beta = \beta_1, \beta_2, \beta_3, \beta_4$

L'estimateur	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$
Ridge	0,049 (0,006)	0,032 (0,003)	0,023 (0,002)	0,029 (0,003)	0,033 (0,004)
Bilodeau	0.252 (0.016)	0.273 (0.018)	0.292 (0.016)	0.514 (0.022)	0.693 (0.028)
Strawderman	0.665 (0.026)	0.703 (0.028)	0.676 (0.025)	0.701 (0.027)	0.718 (0.030)
Stein à partie positive	0.650 (0.026)	0.6778 (0.027)	0.660 (0.024)	0.702 (0.027)	0.718 (0.030)

TABLEAU 3.4. Les risques estimés des estimateurs, pour  $\sigma^2 = 1$ ,  $S(\beta) = 0; 0, 1; 1; 10; 100$ ,  $Q = I_4$  et  $\beta = \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$

L'estimateur	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$
Ridge	2,783 (0,534)	2,43 (0,30)	0,450 (0,160)	1,48 (0,11)	2,14 (0,28)
Bilodeau	1,840 (0,800)	3,256 (1,142)	1,404 (0,256)	3,106 (0,475)	2,625 (0,398)
Strawderman	3,295 (0,855)	4,615 (1,194)	2,136 (0,493)	3,150 (0,548)	2,622 (0,393)
Stein à partie positive	3,149 (1,159)	3,924 (1,132)	2,130 (0,337)	3,106 (0,475)	2,625 (0,398)

TABLEAU 3.5. Les risques estimés des estimateurs, pour  $\sigma^2 = 1$ ,  $S(\beta) = 0; 0, 1; 1; 10; 100$ ,  $Q = X'X$  et  $\beta = \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$

L'estimateur	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$	$\beta_9$
Ridge	172,397 (21,148)	141,211 (18,274)	223,037 (26,970)	199,737 (32,554)	166,774 (22,434)
Bilodeau	758,670 (456,837)	271,863 (85,513)	1452,430 (567,696)	2052,039 (995,485)	2067,308 (716,656)
Strawderman	1639,105 (635,230)	1313,351 (258,592)	2392,039 (987,544)	2554,196 (1011,044)	1793,470 (639,071)
Stein à partie positive	1613,388 (627,849)	1289,790 (254,368)	2357,698 (982,585)	4106,739 (1142,341)	1797,764 (639,647)

TABLEAU 3.6. Les risques estimés des estimateurs, pour  $\sigma^2 = 1$ ,  $S(\beta) = 0; 0, 1; 1; 10; 100$ ,  $Q = (X'X)^2$  et  $\beta = \beta_5, \beta_6, \beta_7, \beta_8, \beta_9$

L'estimateur	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_{13}$
Ridge	0,465 (0,029)	0,552 (0,032)	0,585 (0,033)	0,627 (0,034)
Bilodeau	32,059 (1,254)	33,868 (1,345)	32,564 (1,211)	33,843 (1,327)
Strawderman	33,299 (1,303)	35,163 (1,393)	33,843 (1,257)	35,084 (1,375)
Stein à partie positive	33,299 (1,303)	35,163 (1,393)	33,843 (1,257)	35,084 (1,375)

TABLEAU 3.7. Les risques estimés des estimateurs, pour  $\sigma^2 = 50$ ,  $S(\beta) = 100$ ,  $Q = I_4$  et  $\beta = \beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{13}$

L'estimateur	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_{13}$
Ridge	3,578 (0,126)	3,7296 (0,155)	4,057 (0,148)	3,386 (0,119)
Bilodeau	155,425 (34,570)	125,923 (18,844)	244,929 (64,817)	229,748 (60,698)
Strawderman	154,535 (33,975)	125,916 (19,135)	244,325 (65,087)	229,570 (60,891)
Stein à partie positive	155,425 (34,570)	125,924 (18,844)	244,929 (64,817)	229,748 (60,698)

TABLEAU 3.8. Les risques estimés des estimateurs, pour  $\sigma^2 = 50$ ,  $S(\beta) = 100$ ,  $Q = X'X$  et  $\beta = \beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{13}$

L'estimateur	$\beta_{10}$	$\beta_{11}$	$\beta_{12}$	$\beta_{13}$
Ridge	196,975 (14,873)	204,040 (12,908)	176,914 (13,265)	182,604 (10,646)
Bilodeau	9094,496 (3131,976)	4747,225 (1265,679)	5918,092 (1089,796)	2304,438 (710,240)
Strawderman	9039,288 (3173,225)	4732,578 (1264,181)	6104,345 (1081,196)	2250,574 (697,758)
Stein à partie positive	9481,075 (3111,192)	4440,628 (1181,685)	5621,408 (1023,144)	2655,08 (869,468)

TABLEAU 3.9. Les risques estimés des estimateurs, pour  $\sigma^2 = 50$ ,  $S(\beta) = 100$ ,  $Q = (X'X)^2$  et  $\beta = \beta_{10}$ ,  $\beta_{11}$ ,  $\beta_{12}$ ,  $\beta_{13}$

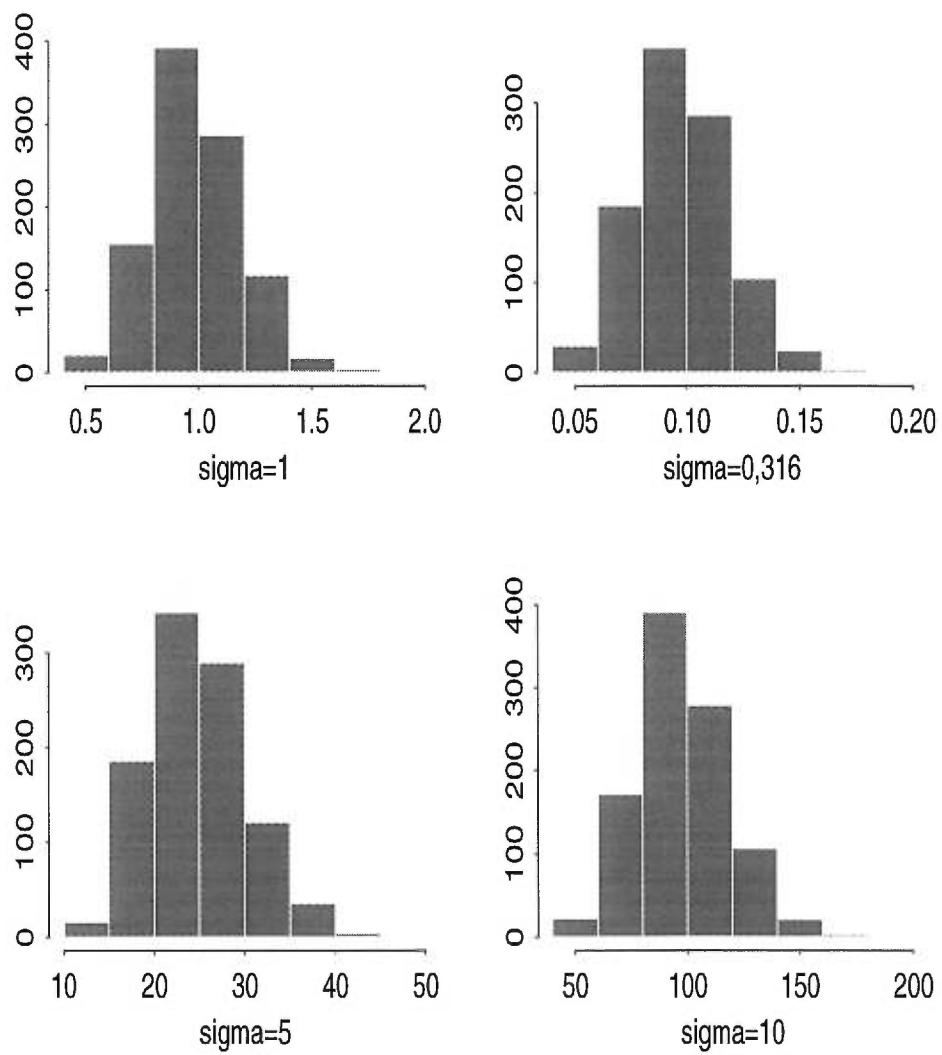


FIGURE 3.8. La distribution expérimentale de  $\hat{\sigma}^2$ , pour  $\sigma^2 = 1; 0,1; 25; 100$ ,  $\beta = (0, 0, 0, 0)'$  et  $S(\beta) = 0$

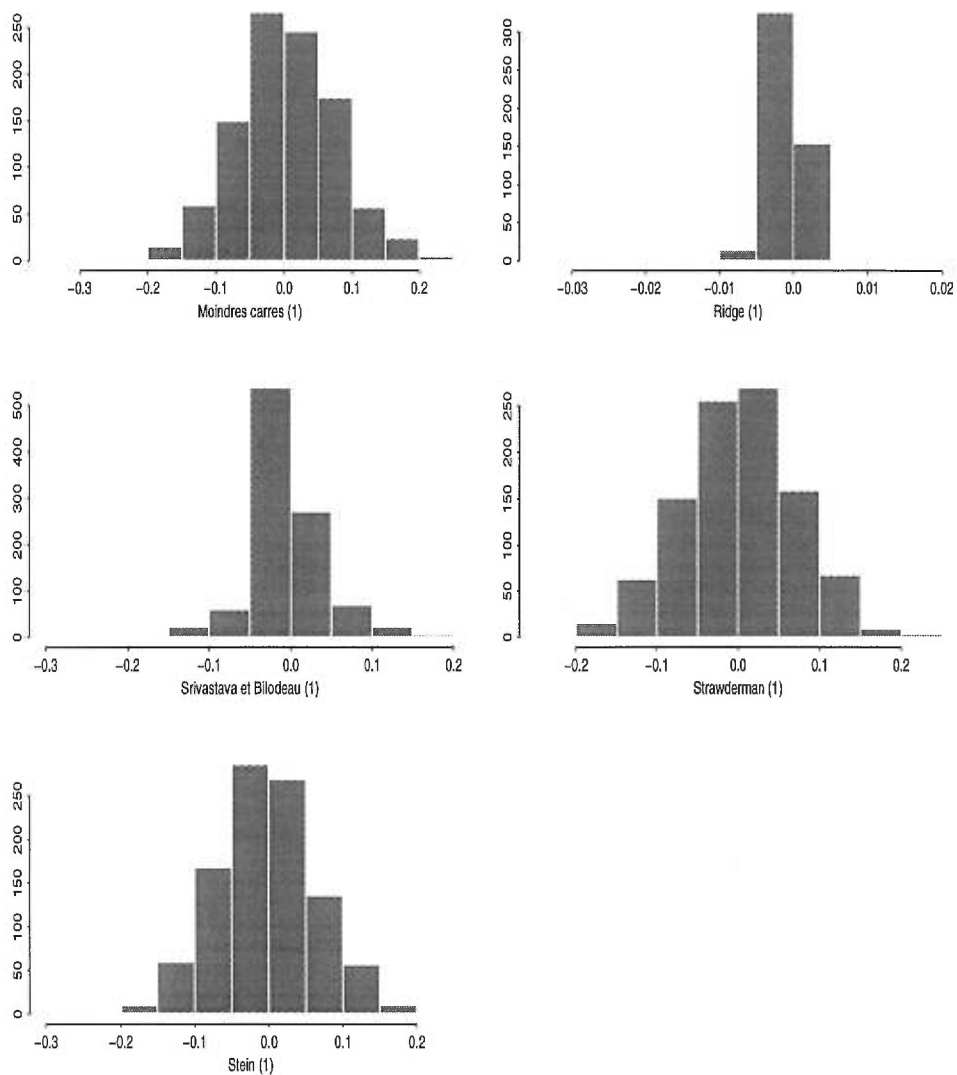


FIGURE 3.9. La distribution expérimentale de la première composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (0, 0, 0, 0)'$ ,  $S(\beta) = 0$  et  $Q = I_4$

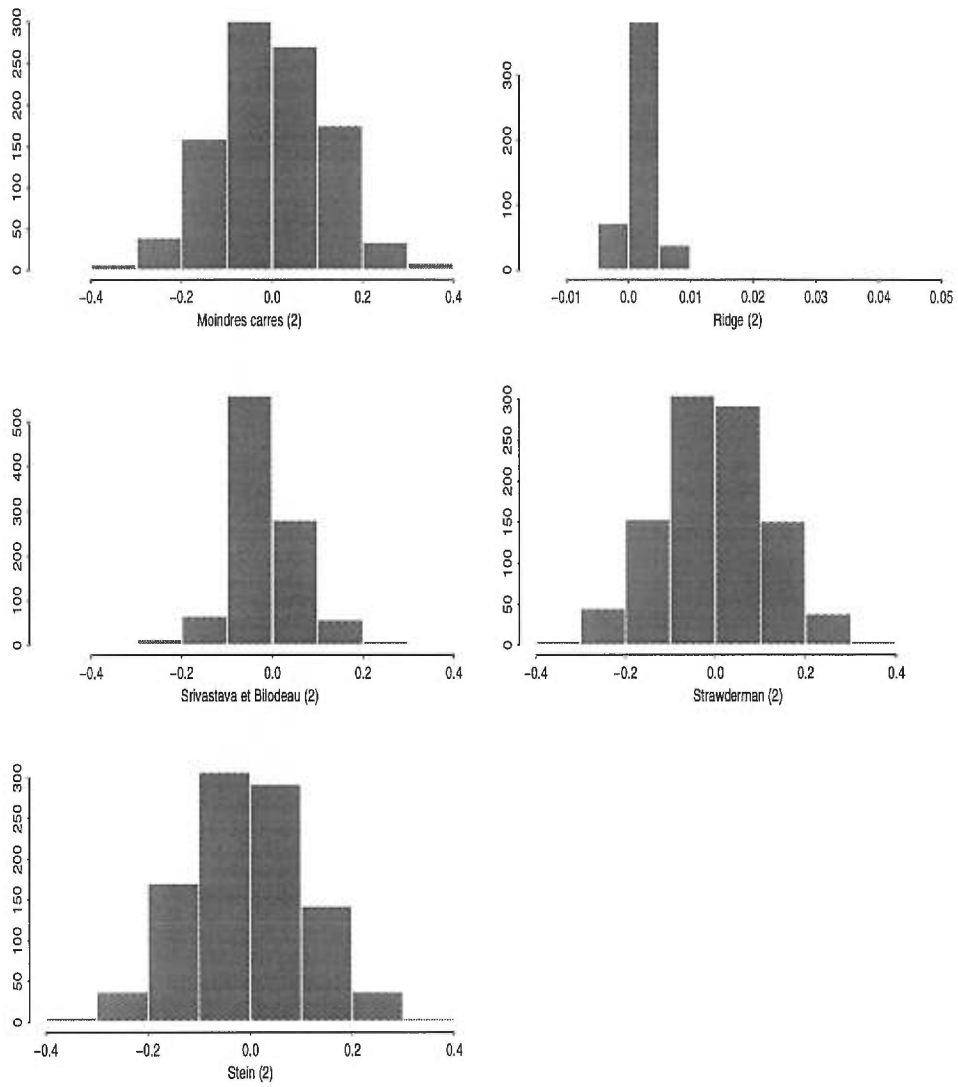


FIGURE 3.10. La distribution expérimentale de la deuxième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (0, 0, 0, 0)'$ ,  $S(\beta) = 0$  et  $Q = I_4$

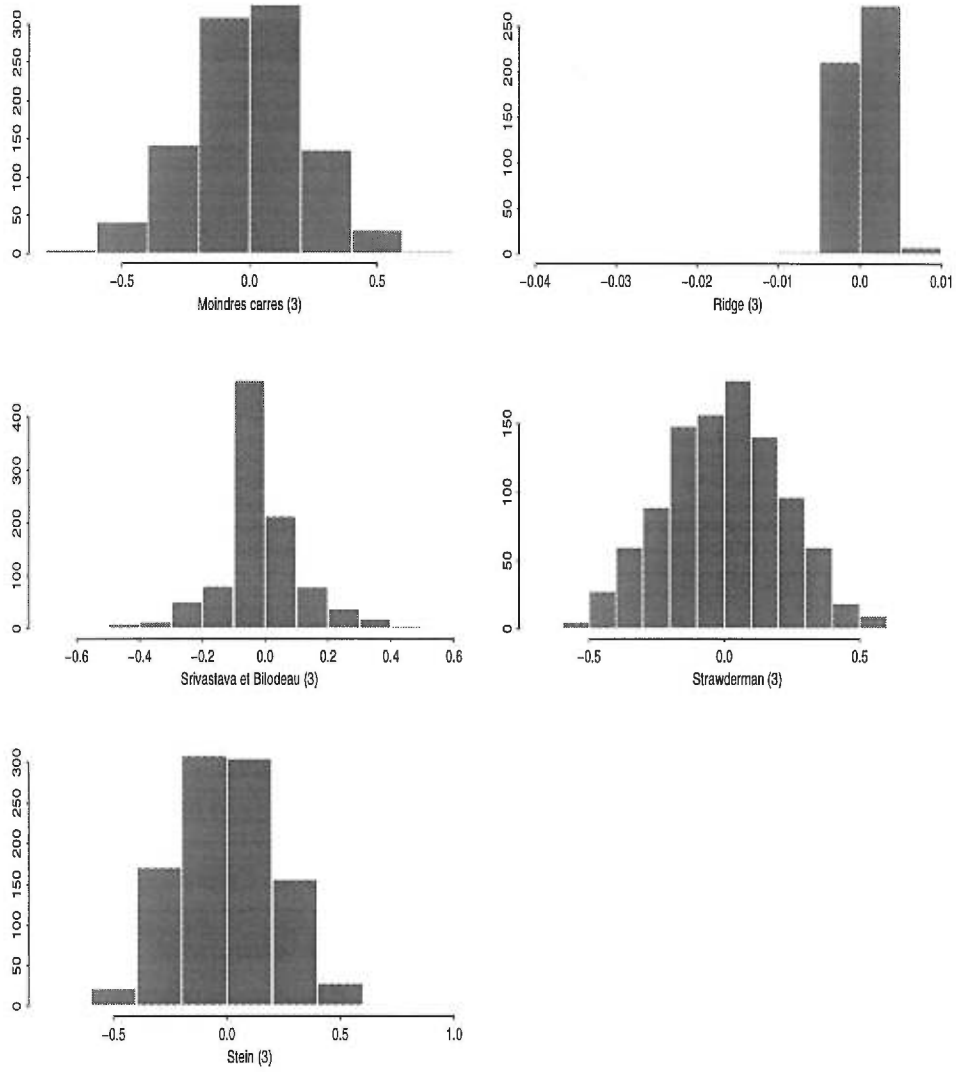


FIGURE 3.11. La distribution expérimentale de la troisième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (0, 0, 0, 0)'$ ,  $S(\beta) = 0$  et  $Q = I_4$

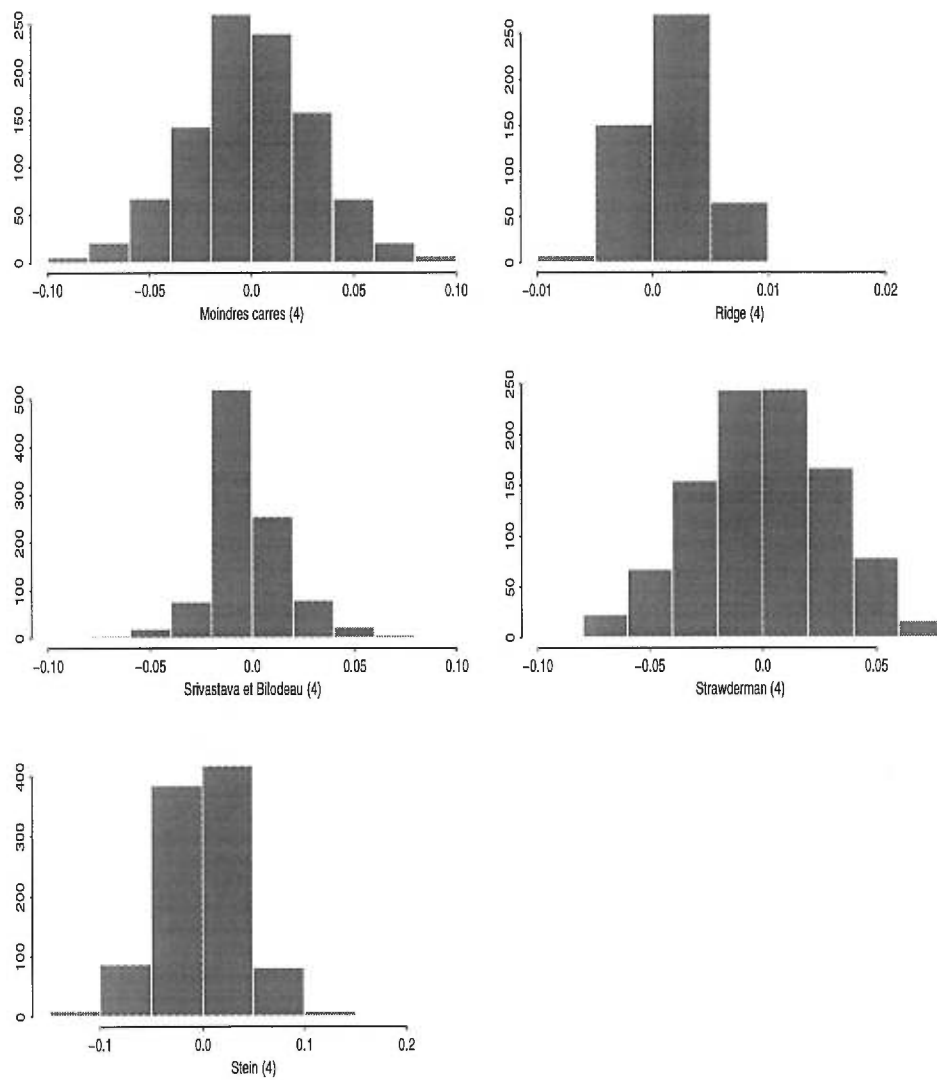


FIGURE 3.12. La distribution expérimentale de la quatrième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (0, 0, 0, 0)'$ ,  $S(\beta) = 0$  et  $Q = I_4$



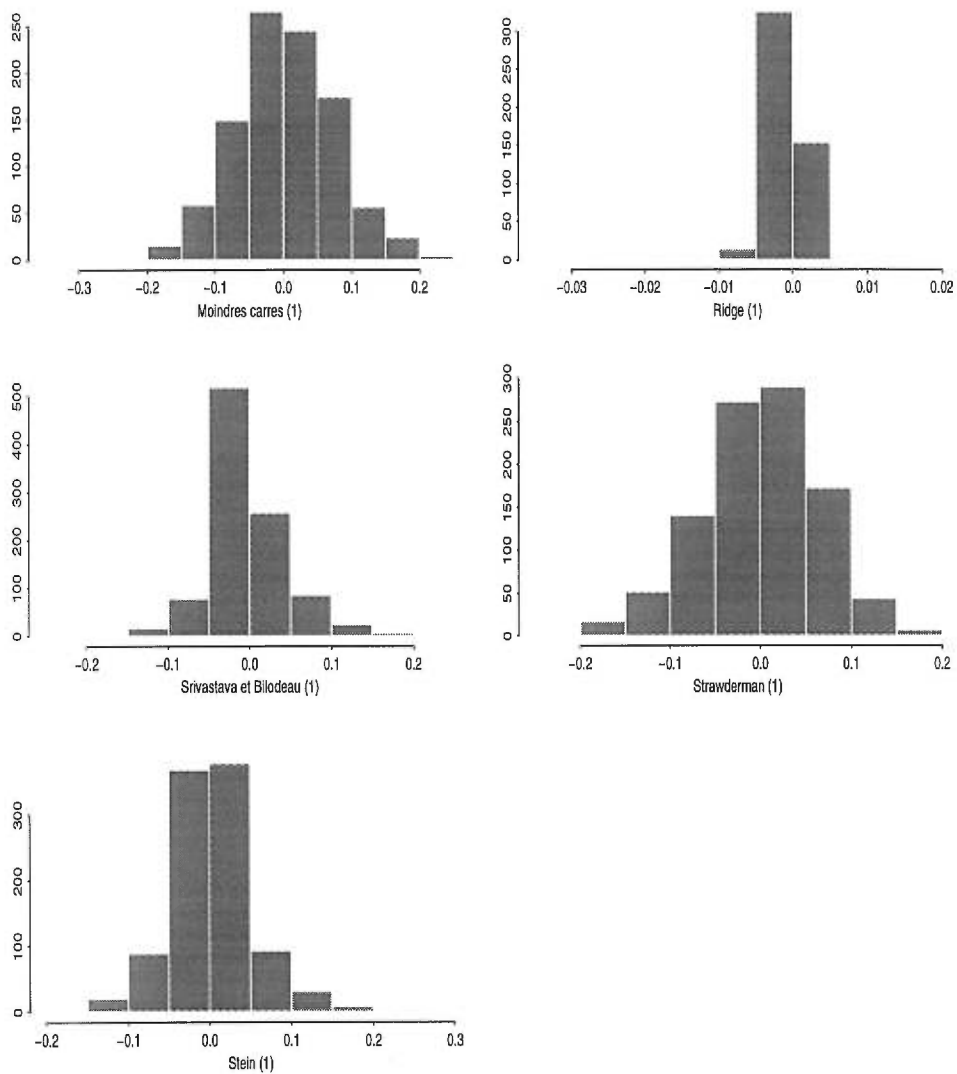


FIGURE 3.13. La distribution expérimentale de la première composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (0, 0, 0, 0)'$ ,  $S(\beta) = 0$  et  $Q = X'X$

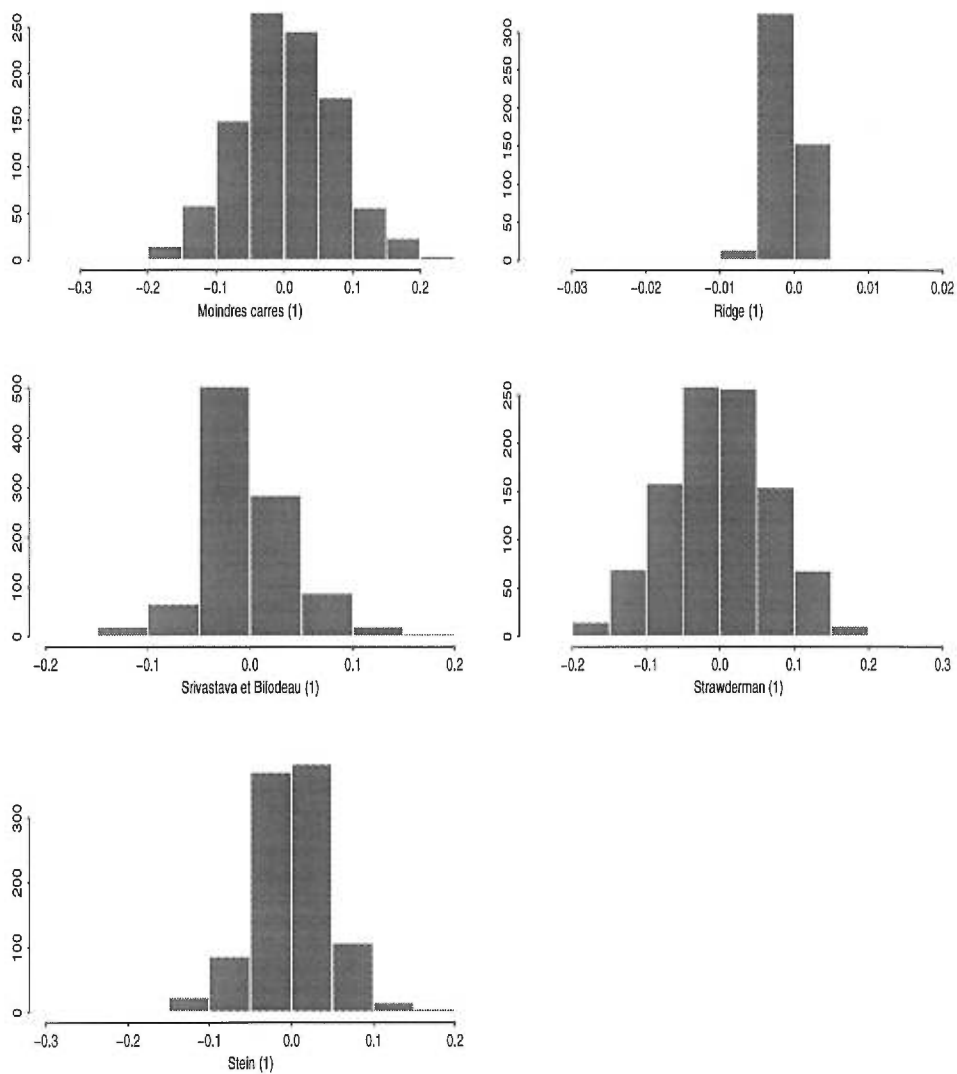


FIGURE 3.14. La distribution expérimentale de la première composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (0, 0, 0, 0)'$ ,  $S(\beta) = 0$  et  $Q = (X'X)^2$

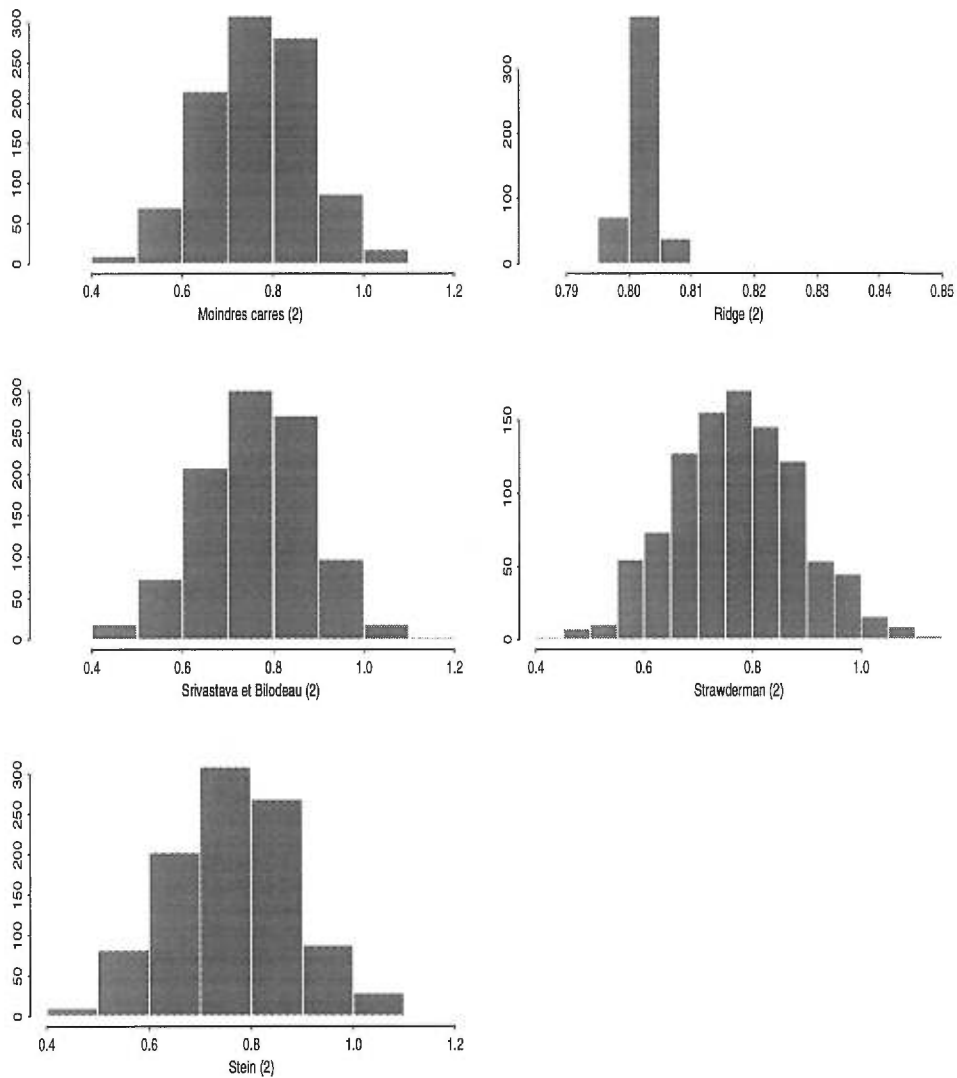


FIGURE 3.15. La distribution expérimentale de la deuxième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (1, 078; 0, 764; -1, 257; 0, 234)'$ ,  $S(\beta) = 100$  et  $Q = I_4$

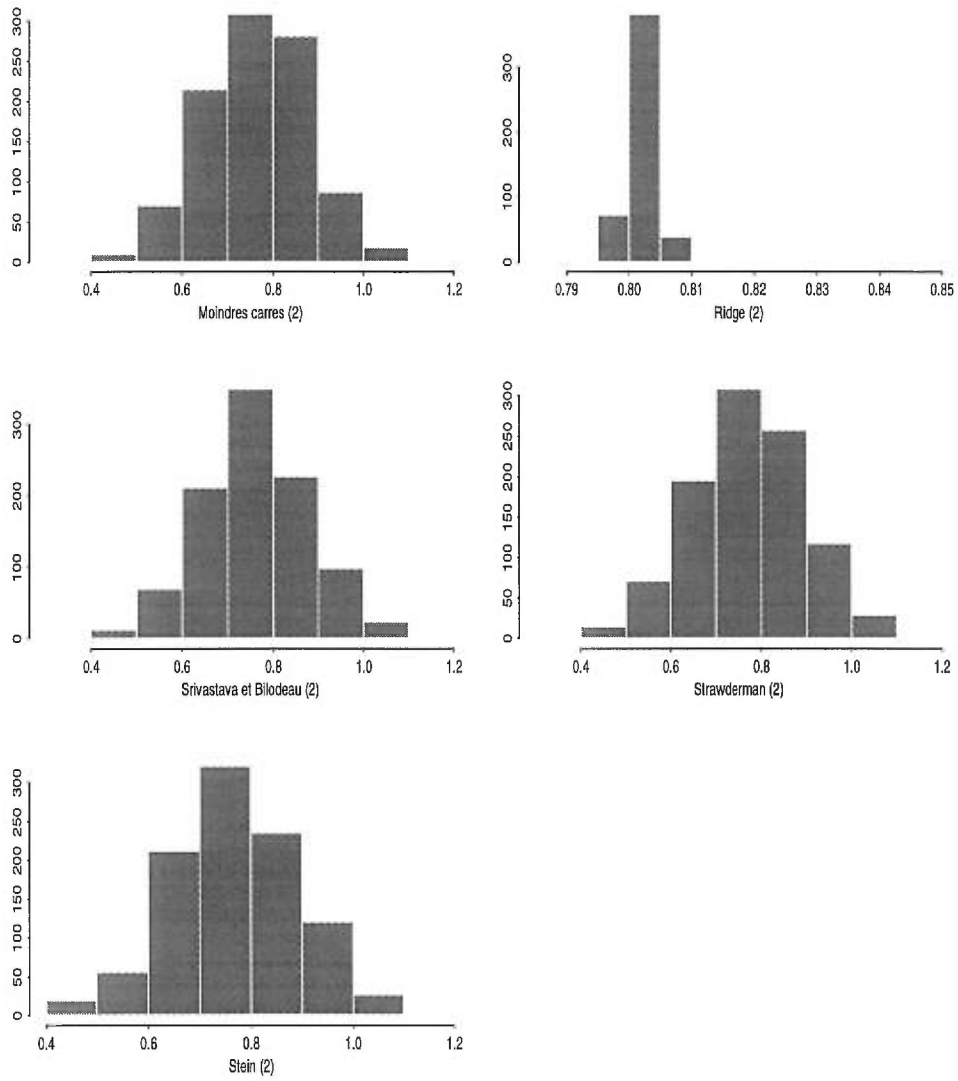


FIGURE 3.16. La distribution expérimentale de la deuxième composante de chaque estimateur, pour  $\sigma^2 = 1$ ,  $\beta = (1, 078; 0, 764; -1, 257; 0, 234)'$ ,  $S(\beta) = 100$  et  $Q = X'X$

## CONCLUSION

---

Nous avons présenté certains estimateurs et leur comportement avec des données qui présentent la multicolinéarité. Plus précisément nous avons concentré notre étude sur l'estimateur ridge. Nous avons écrit un programme pour un estimateur ridge adaptatif qui minimise globalement la fonction  $GCV$ . Par la suite nous avons mentionné d'autres critères de sélection, qui asymptotiquement sont équivalents au  $GCV$ . En plus, nous avons étudié le comportement de certains estimateurs minimax, comme l'estimateur de James et Stein à partie positive [Baranchik (1970)], Strawderman (1978) et Srivastava et Bilodeau (1989). Le risque de ces estimateurs est inconnu par rapport à une fonction de perte donnée. Alors nous avons fait des simulations pour l'estimer et le comparer avec le risque des moindres carrés qui est connu. Les résultats des simulations ont montré que les estimateurs minimax ci-haut sont bons par rapport à la fonction d'erreur quadratique moyenne et par rapport à deux autres fonctions de perte. Cependant, l'estimateur ridge corrige le mieux le problème de multicolinéarité.

## APPENDICES

---

### APPENDICE A

La matrice  $X'X$  :

$$\begin{pmatrix} 23,133 & -4,934 & 3,859 & 32,330 \\ -4,934 & 26,814 & 0,387 & 91,452 \\ 3,859 & 0,387 & 23,525 & 170,867 \\ 32,330 & 91,452 & 170,867 & 1666,696 \end{pmatrix}$$

La matrice  $(X'X)^{-1}$  :

$$\begin{pmatrix} 0,055 & 0,038 & 0,053 & -0,008 \\ 0,038 & 0,147 & 0,216 & -0,031 \\ 0,053 & 0,216 & 0,485 & -0,063 \\ -0,008 & -0,031 & -0,063 & 0,009 \end{pmatrix}$$

La matrice  $(X'X)^2$  :

$$\begin{pmatrix} 1619,632 & 2711,703 & 5702,305 & 54840,54 \\ 2711,703 & 9106,983 & 15626,599 & 154781,56 \\ 5702,305 & 15626,599 & 29764,158 & 288963,99 \\ 54840,538 & 154781,562 & 288963,988 & 2816481,78 \end{pmatrix}$$

APPENDICE B  
La matrice des variables explicatives X:

-0,1120	-0,932	-0,218	-3,400
-0,3120	-0,977	-1,981	-16,098
0,4440	-0,718	0,190	0,761
0,6780	-1,322	0,605	1,786
0,8410	-0,351	-1,231	-7,504
1,0800	0,429	0,127	4,970
-0,4380	0,189	0,644	6,231
-0,0210	0,797	-0,894	-1,989
-0,2430	0,632	1,307	12,618
0,1700	0,066	-0,030	1,691
0,4340	0,081	-0,450	-0,932
1,2580	-0,201	-0,220	0,514
0,3200	-0,107	-0,082	0,871
0,7670	-1,049	0,420	1,535
-0,0530	0,218	0,195	3,575
0,0010	0,065	-0,040	1,448
0,9010	0,772	0,337	7,462
-0,6900	-0,040	-0,354	-1,808
1,1651	-0,118	0,447	5,381
-0,1500	0,215	0,894	8,360
-0,3970	-0,177	0,189	1,80
-0,4940	1,336	0,378	8,328
1,2130	-0,190	0,295	4,113
0,5310	0,343	-0,338	0,865
-0,6180	0,357	-0,814	-3,566
-0,3470	-1,419	0,111	-3,036
1,1200	-0,077	-0,050	2,0005
-1,5150	-0,361	-0,683	-6,059
-0,0960	0,230	-0,017	2,090
-0,6580	0,270	-0,398	-0,999
-0,3760	-0,345	0,474	3,234
-0,2490	1,006	-0,897	-1,507
0,3000	1,112	1,045	13,007
-0,1290	-0,030	-0,324	-1,002
-1,4690	0,924	0,088	3,881
-0,5630	0,737	0,643	8,017
0,2670	-0,289	-0,150	-0,29
-0,2450	-0,936	0,691	2,816
-0,5480	-1,383	0,518	-0,262
-0,3980	0,734	-0,506	0,129
0,1310	0,103	0,462	5,225
-1,0010	1,100	-0,933	-2,182
-0,9880	0,172	-0,370	-1,476
1,0190	-1,392	0,833	3,478
-0,0290	1,195	1,078	13,199
0,8390	0,121	0,950	9,412
-0,3180	-1,271	-0,053	-3,637
-0,656	-0,821	-0,003	-2,050
-0,875	0,983	0,586	8,167
0,145	0,205	-1,818	-10,363

## APPENDICE C

Le vecteur des variables dépendantes  $Y$  :

$Y = (1,528; -0,007; -0,960; -0,211; 0,005; -0,290; 0,751; -1,305; -0,075; -0,928; 0,262; -1,699; -0,296; 0,025; 0,674; -1,096; 1,443; 0,218; -0,776; -0,121; 0,660; -1,461; -0,417; -0,834; 0,563; 1,223; -0,971; 0,223; -0,140; -0,879; 0,890; 1,058; -0,837; 0,965; -1,623; -0,493; 1,262; -0,063; -0,866; 0,833; 1,535; 0,556; 0,938; -0,669; -1,357; 1,859; -1,327; 3,079; -0,313; -0,682)'$ . Ce vecteur a été généré d'une loi  $N(0, I_{50})$ .

Les vecteurs de  $\beta$  qui ont été construits pour  $\sigma^2 = 0,1$  et  $S(\beta) = 0,1$  :

$$\beta_1 = (0,004; 0,003; 0,006; 0,001)'$$

$$\beta_2 = (0,008; -0,002; 0,004; -0,003)'$$

$$\beta_3 = (-0,008; 0,004; 0,004; 0,002)'$$

$$\beta_4 = (0,002; 0,004; 0,002; -0,003)'$$

Les vecteurs de  $\beta$  qui ont été construits pour  $\sigma^2 = 1$  et  $S(\beta) = 0; 0,1; 1; 10; 100$  :

$$\beta_5 = (0; 0; 0; 0)'$$

$$\beta_6 = (-0,018; 0,015; 0,020; -0,010)'$$

$$\beta_7 = (0,025; 0,012; 0,016; 0,021)'$$

$$\beta_8 = (0,023; 0,082; -0,059; 0,078)'$$

$$\beta_9 = (1,078; 0,764; -1,257; 0,234)'$$

Les vecteurs de  $\beta$  qui ont été construits pour  $\sigma^2 = 50$ , et  $S(\beta) = 100$  :

$$\beta_{10} = (0,449; 1,609; 1,148; 1,513)'$$

$$\beta_{11} = (-1,821; 0,756; 1,695; 1,535)'$$

$$\beta_{12} = (0,690; -0,951; -0,912; -1,595)'$$

$$\beta_{13} = (0,306; 8,526; 0,530; 0,927)'$$



## APPENDICE D

Le programme de "SPLUS" qui calcule le risque estimé de l'estimateur de Bilodeau et Srivastava, de Strawderman et celui de Stein à partie positive:

```

function(x, m,ba)
{
  x <- as.matrix(x)
  L <- matrix(0:0, ncol = m, nrow = 3)
  L1 <- matrix(0:0, ncol = m, nrow = 4)
  L2 <- matrix(0:0, ncol = m, nrow = 4)
  L3 <- matrix(0:0, ncol = m, nrow = 4)
  L4 <- matrix(0:0, ncol = m, nrow = 4)
  v <- matrix(0:0, ncol = m, nrow = 1)
  n <- nrow(x)
  p <- ncol(x)
  I <- diag(p)
  w <- (p - 2)/(n - p + 2)
  g <- (2 * p)/(n - p + 2)
  qrx <- qr(x)
  txx <- t(x) %*% x
  eig <- eigen(txx)
  eigva <- eig$values
  eigve <- eig$vectors
  Q <- array(0, dim = c(p, p, 3))
  Q[, , 1] <- diag(p)
  Q[, , 2] <- txx
  Q[, , 3] <- txx %*% txx
  for(j in (1:3)) {
    Eig <- eigen(solve(Q[, , j]) %*% Q[, , 2],symmetric=T,only.values = T )values
    print("la valeur de Eig est")
    print(Eig)
    a <- (p - 2)/((n - p + 2)* max(Eig))
    print("la matrice Q est")
  }
}

```

```

print(Q[, , j])

for(b in (1:ncol(ba))) {

be <- ba[, b]

print("le vecteur beta est:")

print(be)

for(i in (1:m)) {

e <- rnorm(50, 0, sigma)

y <- x %*% be + e

bet <- qr.coef(qrx, y)

sig <- sum(qr.resid(qrx, y) ^2)

E2 <- (I - c((w * sig)/(t(bet) %*% Q[, , 2] %*%

solve(Q[, , j]) %*% Q[, , 2] %*% bet))*

solve(Q[, , j]) %*% txx) %*% bet

E3 <- solve(I + c((a* sig)/(t(bet) %*% Q[, , 2]

%*% bet + g * sig))*solve(Q[, , j])*Q[, , 2]) %*% bet

E4 <- max(0, 1 - (w* sig)/(t(bet) %*% Q[, , 2] %*% bet))*bet

Est <- cbind(E2, E3, E4)

Est <- as.matrix(Est)

for(k in (1:ncol(Est))) {

E <- Est[, k]

L[k, i] <- t(be - E) %*%E

L1[,i]<-bet

L2[,i]<-E2

L3[,i]<-E3

L4[,i]<-E4

v[,i]<-sig/(n-p)

} }

print(L1)

print(L2)

print(L3)

print(L4)

```

```
print(v)

d1 <- apply(L, 1, sum)
d2 <- apply(L2, 1, sum)

r <- d1/m

print("le vecteur risque est")

print(r)

va <- ((1/(m - 1)) * (d2 - (d1)2/m))

ec <- sqrt(va)

print("le vecteur des ecarts est")

print(ec)

}} }
```

## APPENDICE E

Le programme de "SPLUS" qui calcule le risque estimé de l'estimateur ridge:

```

function(x,n1, b1, m, beta)
{
  x <- as.matrix(x)
  L <- matrix(0:0, ncol = n1, nrow = 1)
  L5<-matrix(0:0, ncol = n1, nrow = 1)
  n <- nrow(x)
  p <- ncol(x)
  m1 <- 100
  m2 <- 1000000
  I <- diag(p)
  qrx <- qr(x)
  txx <- t(x) %*% x
  eig <- eigen(txx)
  eigva <- eig$values
  eigve <- eig$vectors
  Q <- array(0, dim = c(p, p, 3))
  Q[, , 1] <- diag(p)
  Q[, , 2] <- txx
  Q[, , 3] <- txx %*% txx
  for(j in (1:3)) {
    print("la matrice Q est")
    print(Q[, , j])
    for(b in (1:ncol(beta1))) {
      be <- beta1[, b]
      print("le vecteur beta est:")
      print(be)
      for(i in (1:n1)) {
        e <- rnorm(50, 0, sigma)
        y <- x %*% be + e
      }
    }
  }
}

```

```

bet <- qr.coef(qrx, y)

for(k in (0:m)) {

P0 <- (k * b1)/m

z <- t(svd(x, nrow(x)) $u) %*% y

s1 <- (1/n) * sum(z^2)

a11 <- ((n * P0)/(eigva + n * P0))^2 * z^2

a1 <- (1/n) * sum(a11[1:p])

a22 <- z^2

a2 <- (1/n) * sum(a22[5:n])

a3 <- (1/n) * (sum((n * P0)/(eigva[1:p] + n * P0)[1:p]) + n - p)

gcv <- (a1 + a2)/a3^2

v <- gcv

if(v < m2) {

m1 <- P0

m2 <- v

} }

print("la valeur de m1 qui donne un gcv petit est:")

print(m1)

E <- solve((txx + m1 * I)) %*% txx %*% bet

L[i] <- t(be - E) %*% Q[, , j] %*% (be - E)

L5[i] <-L[i]

} d1 <- apply(L, 1, sum)

d2 <- apply(L^2, 1, sum)

r <- d1/m

print(L5)

print("le risque est")

print(r)

v <- ((1/(n1 - 1)) * (d2 - (d1)^2/n1))

ec <- sqrt(v)

print("l'ecart est")

```

```
print(ec)
}}}
```

## RÉFÉRENCES

---

- [1] ALAM, A., ET THOMPSON, J.R. (1964). Estimation of the mean of a multivariate normal distribution. *Indiana Univ. Technical Report*.
- [2] ALLEN, D. M. (1971). Mean square error of prediction as a criterion for selecting variables. *Technometrics*. **13**, 469-475.
- [3] ALLEN, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*. **16**, 125-127.
- [4] BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41**, 642-645.
- [5] BIBBY, J., ET TOUTENBURG, H. (1977). *Prediction and improved estimation in linear models*. Wiley, Berlin.
- [6] CRAVEN, P., ET WAHBA, G. (1979). Smoothing noisy data with spline function: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31**, 377-403.
- [7] DELANEY, N. J., ET CHATTERJEE, S. (1986). Use of the bootstrap and cross-validation in ridge regression. *Journal of Business and Economic Statistics*. **4**, 255-262.
- [8] EFRON, B., ET MORRIS, C. (1973). Stein's estimation rule and its competitors an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68**, 117-130.
- [9] EUBANK, R. L. (1988). *Spline smoothing and nonparametric regression*. Marcel Dekker, New York.
- [10] FAREBROTHER, R. W. (1975). The minimum mean square error linear estimator and ridge regression. *Technometrics*. **17**, 127-128.

- [11] GOLUB, G. H., HEATH, M., ET WAHBA, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*. **21**, 215-223.
- [12] GORMAN, J.W., ET TOMAN, R.J. (1966). Selection of variables for fitting equations to data. *Technometrics*. **8**, 27-51.
- [13] HEMMERLE, W. J. (1975). An explicit solution for generalized ridge regression. *Technometrics*. **17**, 309-314.
- [14] HOERL, A. E., ET KENNARD, R.W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. **12**, 55-67.
- [15] HOERL, A. E., ET KENNARD, R.W. (1970b). Ridge regression: Applications to nonorthogonal problems. *Technometrics*. **12**, 69-82.
- [16] HOERL, A. E., ET KENNARD, R.W. (1975). Ridge regression: Some simulations. *Communications in Statistics*. **4**, 105-123.
- [17] LÄUTER, H. (1975). A minimax linear estimator for linear parameters under restrictions in form of inequalities. *Math. Operationsforsch. Statist.* **6**, 689-695.
- [18] LAWLESS, J. F., ET WANG, P. (1976). A simulation study of ridge and other regression estimators. *Commun. Statis. Theor. Meth. A<sub>5</sub>*, 307-323.
- [19] LI, T. F. (1982). Generalized James-Stein estimators. *Commun. Statis. Theor. Meth.* **11**, 2249-2257.
- [20] MACDONALD, G. C., ET GALARNEAU, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*. **70**, 407-416.
- [21] MALLOWS, C. L. (1973). Some comments on  $C_p$ . *Technometrics*. **15**, 661-674.
- [22] MARQUARDT, D. W. (1970). Generalized, ridge regression, biased linear estimation, and non linear estimation. *Technometrics*. **12**, 359-609.
- [23] NEWHOUSE, J. P., ET OMAN, S. D. (1971). An evaluation of ridge estimators. *Rand Report*. No. R-716-PR, 1-28.
- [24] SRIVASTAVA, M. S., ET BILODEAU, M. (1989). Stein estimation under elliptical distribution. *Journal of Multivariate Analysis*. **28**, 247-258.



- [25] STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the third Berkeley Symposium*. **1**, 197-206.
- [26] STRAWDERMAN, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *The Annals of Mathematical Statistics*. **42**, 385-388.
- [27] STRAWDERMAN, W. E. (1973). Proper Bayes minimax estimators of the multivariate normal mean vector for the case of common unknown variances. *The Annals of Statistics*. **1**, 1189-1194.
- [28] STRAWDERMAN, W. E. (1978). Minimax adaptive generalized ridge regression estimators. *Journal of the American Statistical Association*. **73**, 623-627.
- [29] THEIL, H. (1971). *Principles of econometrics*. North Holland Publishing company, Amsterdam.
- [30] THEOBALD, C. M. (1974). Generalization of mean square error applied to ridge regression. *J. Roy. Statist. Soc. B-36*, 103-106.
- [31] THISTED, R. A. (1978). On generalized ridge regression. *University of Chicago Department of Statistics. Technical Report*.