ANNALES D'ÉCONOMIE ET DE STATISTIQUE. - N° 10 — 1988

Performances d'estimateurs à rétrécisseur en situation de multicolinéarité

Christian ROBERT *

RÉSUMÉ. – La présence de quasi-colinéarité au sein des variables explicatives d'un modèle linéaire engendre des effets indésirables sur l'estimateur des moindres carrés, en particulier l'augmentation notable des variances des estimateurs des composantes et la détérioration de la fiabilité de cet estimateur. Après avoir défini précisément la notion de multicolinéarité, nous montrons pourquoi un estimateur de risque uniformément inférieur à celui de l'estimateur des moindres carrés ne peut apporter de solution aux problèmes de multicolinéarité. Nous proposons ensuite un estimateur généralisant les estimateurs en composantes principales qui se comporte de manière relativement satisfaisante en situation de multicolinéarité.

Shrinkage Estimators and Multicollinearity Problems

ABSTRACT. – Near collinearities among explicative variables in a regression model have unwanted effects on the least squares estimator. They inflate the variances of least squares regression coefficient estimates and introduce a lack of fiability for this estimator. In this paper, we define precisely the notion of multicollinearity, then we show why a minimax shrinkage estimator cannot bring any significant improvement over the LSE in the estimation of the components responsible for multicollinearity. In the last part, we propose a generalization of the principal components estimators which performs rather well in a bounded neighborhood of 0.

* C. ROBERT : Université de Rouen (UA CNRS 759, Calcul des Probabilités et Statistique). L'auteur est actuellement *Visiting Professor* à Purdue University (West Lafayette, IN 47907, USA). Cette Recherche a été financée par l'Institut National de la Statistique et des Études Économiques.

1 Introduction

Lorsqu'on considère le modèle classique en régression linéaire, les situations de *multicolinéarité* apparaissent quand deux ou plusieurs des variables explicatives sont presque colinéaires. Les effets sur l'estimateur des moindres carrés sont à la fois numériques et statistiques. Tout d'abord, la matrice intervenant dans l'estimation des coefficients du modèle ayant une ou plusieurs valeurs propres proches de zéro, le calcul de son inverse entraîne d'importantes erreurs numériques. D'autre part, cette matrice inverse est également la matrice de variance-covariance de l'estimateur des moindres carrés; certaines de ces composantes auront donc nécessairement des variances élevées. Enfin, si l'on considère le risque de l'estimateur des moindres carrés relativement à un coût quadratique scalaire, il sera significativement aggravé en situation de multicolinéarité, ce fait étant bien sûr relié au précédent. Pour une vue plus numérique de la multicolinéarité, on pourra consulter STEWARD [1987].

Une première alternative à l'estimateur des moindres carrés apportant une solution partielle à ce problème est la classe des *estimateurs en composantes principales*. Tout d'abord, on considère l'estimateur des moindres carrés dans une base où ses composantes sont indépendantes, ce changement de base étant dit *passage en composantes principales*; cette technique a donné lieu à de nombreux développements en statistique; cf, par exemple, FLURY [1988]. Relativement à cette base, les estimateurs en composantes principales remplacent alors par 0 les composantes dites *responsables de la multicolinéarité* puisque celles-ci ne sont plus fiables (voir JOHNSTON, REIMER et ROTHROCK [1973] et HOCKING [1976]).

Une seconde solution a été apportée par les *estimateurs ridges*, introduits par HOERL et KENNARD [1970]. Elle consiste à ajouter à la matrice à inverser un multiple de la matrice identité afin de rendre plus stable la matrice inverse. Le facteur intervenant dans cet estimateur est déterminé en fonction des observations de manière relativement imprécise (procédé de la «*ridge trace*»); les auteurs ont donc par la suite considéré ce facteur comme dépendant explicitement des observations. CASELLA [1980] a ainsi généralisé le résultat de STRAWDERMAN [1978] pour établir des conditions très générales de domination de l'estimateur des moindres carrés, pour un risque quadratique arbitraire. Mais cet auteur a ensuite démontré que, pour les estimateurs ridges, les deux objectifs :

– amélioration uniforme du risque de l'estimateur des moindres carrés;

– réduction de la contribution au risque des composantes touchées dans certaines régions,

n'étaient pas compatibles (CASELLA [1985]). En situation de multicolinéarité, le gain obtenu est illusoire, face à la grave détérioration du risque.

A contrario, les *estimateurs à rétrécisseur* ont constitué le premier exemple d'estimateurs améliorant uniformément l'estimateur des moindres carrés (JAMES et STEIN [1961]).

JUDGE et BOCK [1978] considèrent que les critères de stabilisation et de domination uniforme ne sont pas compatibles et déconseillent l'usage d'estimateurs à rétrécisseur en situation de multicolinéarité. Par une étude quantitative, HILL et ZIEMER [1982] aboutissent à la même conclusion. Dans un second article (HILL et ZIEMER [1984]), ils montrent qu'en fait la possibilité d'amélioration significative dépend de la « forme » de la multicolinéarité, comme l'avaient suggéré FOMBY et HILL [1979] et considèrent que la « *multicolinéarité sévère* » (i. e. lorsqu'une seule composante est responsable de la multicolinéarité) est un obstacle absolu pour la domination de l'estimateur des moindres carrés. Nous montrerons au paragraphe 4 qu'il est possible d'obtenir un estimateur à rétrécisseur améliorant uniformément cet estimateur, quelle que soit la situation de multicolinéarité. Cependant, le gain restant borné, on ne peut compenser la dégradation due à une situation de multicolinéarité sévère.

Cette incompatibilité est conforme au résultat de STEIN [1956], établissant l'admissibilité de l'estimateur des moindres carrés lorsqu'on estime un ou deux coefficients. En effet, pour un coût quadratique arbitraire, la contribution de la composante touchée par la multicolinéarité au risque devient prépondérante. Diminuer significativement le risque de l'estimateur des moindres carrés revient donc à améliorer l'estimation de cette seule composante. Le problème étant essentiellement unidimensionnel, il est alors impossible de garantir une uniforme domination de l'estimateur des moindres carrés. Il faut choisir entre les deux objectifs définis ci-dessus.

Le second critère étant en pratique plus important, nous proposons au paragraphe 5, un estimateur, généralisant les estimateurs en composantes principales, dont le but est d'améliorer l'estimation des composantes touchées par la multicolinéarité. Il est « intuitif » dans le sens où il rétrécit d'autant plus vers 0 que la variance de la composante principale est grande. Nous comparons les performances de cet estimateur à celles de l'estimateur des moindres carrés et des estimateurs en composantes principales. Les simulations effectuées indiquent que cet estimateur est une solution possible au problème de la multicolinéarité sévère pour des valeurs modérées des paramètres.

Nous donnons au paragraphe 2. 2, un exemple introductif avant d'exposer au paragraphe 3, une définition de la multicolinéarité qui permet d'intégrer son *aspect dynamique*, essentiel à l'analyse de ce problème et souvent mal formalisé dans la littérature sur le sujet.

2 Modèle étudié

2.1. Cas général

Les problèmes de multicolinéarité se présentent souvent dans le cadre du modèle linéaire; on observe un vecteur aléatoire n-dimensionnel $y = X\beta + \sigma z$, où β, inconnu, appartient à $\mathbb{R}^k (k < n)$, X est une matrice (n, k), de rang k, connue, $\sigma (>0)$ est inconnu et z est un vecteur aléatoire n-dimensionnel, centré, de matrice de variance V connue (on la supposera ici définie positive).

Les risques des estimateurs de β sont calculés relativement à des coûts quadratiques scalaires normalisés, c'est-à-dire tels que le coût de l'estimation de β par $\hat{\beta}$ soit pris égal à $\frac{1}{\sigma^2}(\hat{\beta}-\beta)'Q(\hat{\beta}-\beta)$, où Q est une matrice symétrique définie positive, dite *matrice de coût.*

Dans la suite de l'étude, nous supposerons V fixé; par contre, nous ferons varier X ainsi que, dans certains cas, Q. Par conséquent, nous pouvons, au prix d'un éventuel changement de base de $\mathbb{R}^n$, supposer V égale à l'identité (i. e. les composantes de z normales centrées réduites et indépendantes). Nous nous placerons dans ce cas par la suite.

Il est bien connu que l'estimateur des moindres carrés de β est $\hat{\beta}^0 = (X'X)^{-1}X'y$, de matrice de variance-covariance $\sigma^2(X'X)^{-1}$. Son risque est constant, égal à $\mathrm{tr}((X'X)^{-1}Q)$.

C'est essentiellement à l'impact des « variations » de $(X'X)$ sur le risque de divers estimateurs (estimateur des moindres carrés, estimateurs en composantes principales, estimateurs à rétrécisseur matriciel) que nous allons nous intéresser; plus précisément, nous proposerons (en 5.3) un estimateur relativement robuste face à des variations « arbitraires » du spectre de $X'X$.

Avant de présenter un exemple introductif du problème de la multicolinéarité, rappelons que $(X'X)$ a une double interprétation :

— algébrique, puisque l'élément générique, t_{ij}, de $(X'X)$ est le produit scalaire dans $\mathbb{R}^n$ des vecteurs colonnes e_i et e_j de X;

— probabiliste, puisque l'élément générique, u_{ij}, de $(X'X)^{-1}$ est, à σ^2 près, la covariance des composantes $\hat{\beta}_i^0$ et $\hat{\beta}_j^0$ de $\hat{\beta}^0$.

2.2. Exemple

Soit, pour $k=3$ et $n=4$, la matrice

$$X_\theta=\begin{bmatrix}\cos\theta & \cos\theta & 0\\ \sin\theta & -\sin\theta & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}$$

Q étant prise de terme générique q_{ij}.

Ainsi les $e_i (1 \leqslant i \leqslant 3)$, vecteurs colonnes unitaires de X_θ, sont tels que e_1 et e_2 font entre eux l'angle 2θ, où $0<\theta\leqslant\pi/4$, e_3 étant orthogonal à e_1 et e_2. On a

$$X'_\theta X_\theta=\begin{bmatrix}1 & \cos 2\theta & 0\\ \cos 2\theta & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

et le spectre de $(X'_\theta X_\theta)$ est $\{1+\cos 2\theta, 1-\cos 2\theta, 1\}$. Les sous-espaces propres sont engendrés par les vecteurs unitaires $\frac{f_1+f_2}{\sqrt{2}}, \frac{f_1-f_2}{\sqrt{2}}$, f_3 où (f_1, f_2, f_3) est la base canonique de $\mathbb{R}^3$. Remarquons que, dans cet exemple, la base propre est indépendante de θ. Il vient

$$(X'_\theta X_\theta)^{-1}=\begin{bmatrix}\dfrac{1}{\sin^2 2\theta} & -\dfrac{\cos 2\theta}{\sin^2 2\theta} & 0\\ -\dfrac{\cos 2\theta}{\sin^2 2\theta} & \dfrac{1}{\sin^2 2\theta} & 0\\ 0 & 0 & 1\end{bmatrix}$$

Les composantes $\hat\beta_1^0$, $\hat\beta_2^0$, $\hat\beta_3^0$ de l'estimateur des moindres carrés ont donc pour variances respectives $\frac{\sigma^2}{\sin^2 2\theta}, \frac{\sigma^2}{\sin^2 2\theta}, \sigma^2$; $\hat\beta_3^0$ est non corrélée avec $\hat\beta_1^0$ et avec $\hat\beta_2^0$ mais la corrélation entre $\hat\beta_1^0$ et $\hat\beta_2^0$ vaut $-\cos 2\theta$.

Le risque de $\hat\beta^0$ s'écrit $\frac{1}{\sin^2 2\theta}(q_{11}+q_{22}-2q_{12}\cos 2\theta)+q_{33}$. On voit donc que, si θ tend vers 0, les deux vecteurs e_1 et e_2 tendent à se confondre, d'où le terme de *quasi-colinéarité* ou *multicolinéarité* appliqué à ce phénomène. Plusieurs conséquences en découlent :

- la plus petite valeur propre de $X'_\theta X_\theta$, 1-cos 2θ, tend vers 0, donc les composantes $\hat\beta_1^0$ et $\hat\beta_2^0$ de l'estimateur des moindres carrés deviennent instables numériquement; les variances de $\hat\beta_1^0$ et de $\hat\beta_2^0$ tendent vers l'infini et la corrélation entre $\hat\beta_1^0$ et $\hat\beta_2^0$ tend vers -1;

– quelle que soit la forme quadratique de coût Q, le risque de $\hat\beta^0$ tend vers l'infini (car, Q étant définie positive, $q_{11}+q_{22}-2q_{12}\neq 0$).

La nature de cette aggravation du risque est bien mise en évidence si l'on utilise la diagonalisation de $X'_\theta X_\theta$ dans la base propre

$$g_1 = \frac{f_1 - f_2}{\sqrt{2}}, \qquad g_2 = \frac{f_1 + f_2}{\sqrt{2}}, \qquad g_3 = f_3;$$

relativement à cette base, $(X'_\theta X_\theta)^{-1}$ se transforme en la matrice diagonale de termes diagonaux $(1-\cos 2\theta)^{-1}$, $(1+\cos 2\theta)^{-1}$, 1, et Q en une matrice Q′, non nécessairement diagonale, dont les éléments diagonaux sont

$$q'_{11} = \frac{q_{11}+q_{22}-2q_{12}}{2}, \qquad q'_{22} = \frac{q_{11}+q_{22}+2q_{12}}{2} \qquad \text{et} \qquad q'_{33} = q_{33}.$$

Soient $(\eta_1, \eta_2, \beta_3)$ et $(\hat\eta_1^0, \hat\eta_2^0, \hat\beta_3^0)$ les composantes respectives de β et de $\hat\beta^0$ dans cette base.

On retrouve donc le risque de $\hat\beta^0$ sous la forme

$$\frac{q'_{11}}{1-\cos 2\theta} + \frac{q'_{22}}{1+\cos 2\theta} + q'_{33},$$

somme de trois termes dont *seul* le premier tend vers l'infini. On pourra dire dans ce cas que la première composante est *responsable de la multicolinéarité* (ou est *touchée par la multicolinéarité*). Si l'on note R_1 le risque associé au coût quadratique $q'_{11}(\hat\eta_1^0 - \eta_1)^2$, le risque de $\hat\beta^0$ s'écrit $R_1 + o(R_1)$, où $o(R_1)$ tend vers 0 quand θ tend vers 0; on est donc ramené à un problème unidimensionnel: l'estimation de la composante touchée par la multicolinéarité. Pour cette situation, STEIN [1956] a démontré qu'il n'existait pas d'estimateur uniformément meilleur que l'estimateur des moindres carrés quelle que soit la valeur de β.

3 Situations de multicolinéarité

3.1. Définition

L'exemple ci-dessus s'intègre de manière évidente dans la situation suivante, qui nous apparaît contenir les différents contextes dans lesquels les problèmes de multicolinéarité sont évoqués dans la littérature statistique.

Soit ξ la moyenne du vecteur y; on suppose qu'elle appartient à un sous-espace de dimension k de $\mathbb{R}^n$, Θ. Si X, matrice (n, k) est une « base » de ce sous-espace (au sens où les vecteurs colonnes de X forment une base de Θ), il existe $\beta \in \mathbb{R}^k$ tel que $\xi = X\beta$. En pratique, la base X est donnée et l'on s'intéresse plus à l'estimation de β qu'à celle de ξ. Afin de montrer la sensibilité de divers estimateurs de β au choix de X, nous allons considérer une *famille de matrices* (n, k), $(X_t)_{t\in T}$, de rang k, telle que les matrices $X_t' X_t$ admettent toutes une même base propre $(g_1, \ldots, g_k)$, orthonormale relativement à la forme quadratique usuelle sur $\mathbb{R}^k$. Cette situation est purement artificielle puisque, en pratique, l'expérimentateur ne dispose que d'une matrice X, observée; cependant cette formalisation permet de mettre en évidence les modifications induites par des variations au sein du spectre de $X'X$. En particulier, il faut bien voir que l'indice t ne correspond généralement à rien en pratique (ce n'est pas un temps ou un nombre d'observations).

Pour l'exemple 2.2, on peut prendre $t=\theta\in T=\left]0, \frac{\pi}{4}\right]$. Mais un autre choix, comme $t=\cos\theta$ et $T=\left[\frac{1}{\sqrt{2}}, 1\right[$, est aussi possible. Nous supposerons toujours σ, V et ξ indépendants de t. Le modèle considéré s'écrit

$$y=X_t\,\beta^{(t)}+\sigma^2 \tag{1}$$

et l'on s'intéresse à l'estimation de $\beta^{(t)}\in\mathbb{R}^k$ (notons que $X_t\beta^{(t)}=\xi$ pour tout t). On note $\hat\beta^{0\,(t)}$ l'estimateur des moindres carrés de $\beta^{(t)}$.

Soit $(\lambda_1^{(t)}, \ldots, \lambda_k^{(t)})$ le spectre de $X_t' X_t$. Nous allons considérer l'effet de variations des $\lambda_i^{(t)}$ (en fonction de t) sur les risques de divers estimateurs. Tout d'abord, soit $(q'_{11}, \ldots, q'_{kk})$ la suite des éléments diagonaux de la matrice Q′ transformée de Q dans la base propre; Q′ n'est pas nécessairement diagonale et les q'_{ii} sont strictement positifs; le risque de $\hat\beta^{0\,(t)}$ s'écrit alors

$$\sum_{i=1}^{k}\frac{q'_{ii}}{\lambda_i^{(t)}} \tag{2}$$

Ce risque est non borné sur T si et seulement si

$$\inf_{t\in T}\ \inf_i \lambda_i^{(t)}=0.$$

Multiplier tous les termes de la matrice X par un même facteur s revient à multiplier les valeurs propres de $(X'X)$ par s^2, donc à diviser le risque de $\hat\beta^0$ par s^2. Ce type de variation sur $X'X$ étant sans intérêt, on l'éliminera en supposant que la famille $(X_t)_{t\in T}$ est, en un certain sens, « normalisée supérieurement », c'est-à-dire que la famille des valeurs propres $\lambda_i^{(t)}$ ($t\in T$; $1\leqslant i\leqslant k$) est bornée supérieurement (on peut par exemple supposer, comme dans CASELLA [1985], que $\max_{1\leqslant i\leqslant k}\lambda_i^{(t)}=1$, $\forall t\in T$). Ne considérant que des variations relatives des valeurs propres de $X_t' X_t$, il est donc naturel de dire que la famille $(X_t)_{t\in T}$ *présente une situation de multicolinéarité* si

$$\inf_{t\in T}\frac{\inf_i \lambda_i^{(t)}}{\sup_i \lambda_i^{(t)}}=0. \tag{3}$$

Pour reprendre une caractérisation introduite dans BROWN [1975], (3) équivaut à: *quel que soit le cône C de* $(\mathbb{R}_+)^k\cup(0)$, *il existe* $t\in T$ *tel que* $(\lambda_1^{(t)}, \ldots, \lambda_k^{(t)})$ *n'appartient pas à* C.

C'est donc en la considération des risques des différents estimateurs en fonction des coefficients $\inf_{t \in T} \frac{\lambda_i^{(t)}}{\sup_j \lambda_j(t)}$, et surtout de $\inf_{t \in T} \frac{\inf_i \lambda_i(t)}{\sup_i \lambda_i(t)}$, que consiste la majeure partie des études (théoriques et numériques) consacrées à la multicolinéarité. Le rapport $\kappa_t = \frac{\sup_i \lambda_i^{(t)}}{\inf_i \lambda_i^{(t)}}$ est d'ailleurs fréquemment appelé *degré de multicolinéarité*; il est utilisé en analyse numérique (sous la qualification de *condition number*) comme indicateur numérique de la dégradation vis-à-vis de la situation orthonormale (où ce degré vaut 1). Du point de vue statistique, il est parfois critiqué comme trop imprécis : STEWARD [1987] propose de lui substituer un *ensemble d'indices* pour établir un diagnostic plus fiable. Cependant THISTED [1987] note que le « condition number » est un indicateur de précision relative pour l'estimation des combinaisons linéaires des paramètres. De plus, son importance dans les problèmes de minimaxité apparaît nettement dans CASELLA [1985] (voir aussi Section 4).

Dans l'exemple 2.2, $\kappa_\theta = (\operatorname{cotg} \theta)^2$, qui tend vers $+\infty$ quand θ tend vers 0. Notons que, au risque d'une certaine ambiguïté, on peut substituer t à (κ_t) comme indicateur de la multicolinéarité.

3.2. Composantes principales

Les composantes de $\beta^{(t)}$ dans la base propre $(g_1, \ldots, g_k)$ sont dites « *composantes principales* » de $\beta^{(t)}$. Nous les noterons $\eta_i^{(t)}$ $(1 \leqslant i \leqslant k)$. Les composantes de $\hat{\eta}^{0\,(t)}$, estimateur des moindres carrés de $\eta^{(t)}$, sont indépendantes, de variances respectives $\frac{\sigma^2}{\lambda_i^{(t)}}(1 \leqslant i \leqslant k)$. Ainsi, quel que soit le coût, le risque apparaît comme la somme des risques des estimateurs des $\eta_i^{(t)}$, pour les coûts $\frac{q'_{ii}}{\sigma^2}(\eta_i^{(t)} - \hat{\eta}_i^{(t)})^2$.

Comme nous l'avons déjà remarqué dans l'exemple introductif, l'intérêt de la décomposition en composantes principales réside donc en la mise en évidence des *composantes principales responsables de la multicolinéarité*, i. e. de l'ensemble des i tels que

$$\inf_{t \in T} \frac{\lambda_i^{(t)}}{\sup_j \lambda_j^{(t)}} = 0.$$

Cet ensemble peut éventuellement se réduire à un seul indice, comme dans l'exemple 2.2; dans ce cas, nous dirons que nous sommes en *situation de multicolinéarité* sévère car il est alors impossible d'améliorer uniformément l'estimateur des moindres carrés (voir §4 et CASELLA [1985]). Si un plus grand nombre de composantes principales interviennent, on peut alors satisfaire aux deux objectifs exposés dans l'introduction; de plus, cette situation est caractéristique d'un problème mal posé et doit remettre en cause le modèle choisi. C'est pourquoi nous nous attacherons plus particulièrement à l'étude des situations de multicolinéarité sévère.

Si le passage dans la base des composantes principales ne résout pas le problème de la multicolinéarité, il peut fournir un indicateur pour le choix d'estimateurs remédiant (partiellement) à l'aggravation du risque en situation de multicolinéarité. Les *estimateurs en composantes principales* s'appuient sur cette décomposition (voir § 5. 1) : pour une matrice X donnée, ils consistent à remplacer l'estimation par les moindres carrés des composantes principales contribuant le plus à la multicolinéarité [i. e. correspondant aux plus petites valeurs propres de $(X'X)$] par 0, considérant que l'estimation par les moindres carrés n'est plus fiable, puis à effectuer le changement de base permettant d'obtenir les estimations des composantes initiales $\beta_i (1 \leqslant i \leqslant k)$. Le nombre de composantes principales estimées par 0 est dit *ordre de l'estimateur.* Nous verrons en quoi les estimateurs que nous proposons généralisent ces estimateurs en composantes principales (voir § 5).

Remarque: L'hypothèse faite sur la base propre commune à toutes les matrices $X_t' X_t$ peut paraître restrictive. Insistons à nouveau sur le fait que cette base est donnée, en pratique, par la matrice X et que seules les amplitudes des valeurs propres de $X'X$ sont importantes pour la multicolinéarité. L'aspect « dynamique » de la notion de multicolinéarité peut donc être introduit en associant à X une famille $(X_t)_{t \in T}$ dont X est l'un des éléments, et telle que les matrices $X_t' X_t$ admettent une base propre identique à celle de $X'X$.

3.3. Le cas particulier de la prédiction

Comme on peut le remarquer dans la formule (2), la multicolinéarité est indépendante du choix de la matrice Q, cette dernière ne déterminant que la « vitesse de convergence » du risque vers l'infini. Cependant, lorsque le coût dépend de $X_t' X_t$, il n'en est plus de même. HILL et ZIEMER [1982] envisagent des matrices de la forme $(X_t' X_t)^m (m \geqslant 1)$ qui donnent, pour l'estimateur des moindres carrés, un risque égal à $\sum_{i=1}^{k} (\lambda_i^{(t)})^{m-1}$. Dans ce cas, le phénomène de multicolinéarité disparaît; on peut proposer des estimateurs améliorant significativement l'estimateur des moindres carrés pour toute valeur du spectre $\{\lambda_1^{(t)}, \ldots, \lambda_k^{(t)}\}$.

Ce résultat peut sembler artificiel mais il présente une application importante lorsque l'on s'intéresse à la *prédiction* (voir THISTED et MORRIS [1980]) car la matrice de coût usuelle pour ce problème est $(X'X)$ (ce qui revient à estimer ξ), modèle dans lequel le risque de l'estimateur des moindres carrés, $\hat{\xi}^0$, est k, indépendant de la multicolinéarité.

Nous allons donc, dans les paragraphes suivants, étudier des estimateurs en fonction du spectre de $X_t' X_t$, pour un coût quadratique constant.

4 Estimateurs à rétrécisseur matriciel

4.1. Définition

Nous nous intéressons ici aux *estimateurs à rétrécisseur* de β de la forme

(4) $$\varphi(y) = \hat{\beta}^0(y) - h(\hat{\beta}^{0\prime} \mathrm{B} \hat{\beta}^0, s^2(y)) . \mathrm{C} \hat{\beta}^0(y)$$

où

- C est une matrice (k, k);
- B est une matrice (k, k) symétrique définie positive;
- s^2 est l'estimateur usuel de la variance σ^2,

$$s^2 = \frac{1}{n-k} \| y - \mathrm{X} \hat{\beta}^0 \|^2;$$

- h est une application de $(\mathbb{R}_+)^2$ dans $\mathbb{R}_+$.

Sur la forme et l'origine de ces estimateurs, on pourra consulter Judge et Bock [1978] ou Cellier, Fourdrinier et Robert [1986]. En particulier, ces auteurs établissent des conditions, fonctions de Q, sur C, B et h qui assurent que le risque de φ est uniformément inférieur à celui de $\hat{\beta}^0$. Ces conditions font intervenir des notions de *contrôle* sur les variations de h.

A la suite de Cellier *et al.*, nous dirons qu'une fonction f de $\mathbb{R}_+$ est *i-contrôlée* s'il existe $\lambda \in \mathbb{R}_+$ tel que $x \to x^\lambda f(x)$ est une application croissante (« *i* » pour « increasing »); on dira alors que f est *i-contrôlée à hauteur* λ. De même, nous dirons que f est *d-contrôlée à hauteur* μ si $x \to x^\mu f(x)$ est une application décroissante (« *d* » pour « decreasing »).

Si A est une matrice (k, k), on note respectivement $\mathrm{tr}(\mathrm{A})$ et $\mathrm{pgvp}(\mathrm{A})$ la trace et la plus grande valeur propre de A.

Dans tous les cas, une condition nécessaire et suffisante pour qu'il existe φ améliorant uniformément l'estimateur des moindres carrés est que

(5) $$\mathrm{tr}((\mathrm{X}'\mathrm{X})^{-1} \mathrm{QC}) - 2\, \mathrm{pgvp}((\mathrm{X}'\mathrm{X})^{-1} \mathrm{QC}) \geq 0$$

(voir Brown [1975]). Cette condition implique évidemment des contraintes sur le choix de C, tout particulièrement lorsqu'on travaille en situation de multicolinéarité, comme nous allons le voir ci-dessous.

4.2. Calcul du risque et condition suffisante de domination

En restreignant légèrement la classe (4), on obtient des conditions suffisantes de domination de l'estimateur des moindres carrés.

Ainsi, s'il existe une base orthonormale dans laquelle (X′X), Q, C et B diagonalisent simultanément et si les valeurs propres de C sont toutes positives ou nulles, on déduit de CELLIER *et al.* [1986] la proposition suivante :

> PROPOSITION 1 : Une condition suffisante pour que l'estimateur φ défini en (4.1) domine uniformément l'estimateur des moindres carrés est qu'il existe deux nombres réels positifs μ_1, μ_2 tels que
>
> (i) pour tout $u>0$, $h(.,u)$ soit *i-contrôlée* à hauteur μ_1;
>
> (ii) pour tout $v>0$, $h(v,.)$ soit *d-contrôlée à hauteur* μ_2;
>
> (iii) $\forall (v,u)\in(\mathbb{R}_+)^2$,
>
> $$\frac{v}{u}h(v,u)\leqslant 2\frac{\operatorname{tr}((X'X)^{-1}QC)-2\mu_1\operatorname{pgvp}((X'X)^{-1}QC)}{\operatorname{pgvp}(C'QCB^{-1})}\times\frac{n-k}{n-k-2\mu_2}$$

Cette hypothèse de diagonalisation simultanée revient en fait à travailler dans la base en composantes principales.

D'autre part, si la fonction de rétrécissement h est différentiable, on peut, sans que soient nécessaires les hypothèses de diagonalisation simultanée, fournir un estimateur sans biais du gain de φ, i.e. de la différence entre le risque de l'estimateur des moindres carrés et celui de φ, $G_\varphi(\xi,\sigma)$. On a en effet d'après FRAISSE, ROBERT et ROY [1987, Proposition 6]

> PROPOSITION 2 : Si $\mathbb{E}_{\xi,\sigma}[\hat\beta^{0'}B\hat\beta^0 h^2(\hat\beta^{0'}B\hat\beta^0,s^2(y))]$ est fini,
>
> $$\begin{aligned}G_\varphi(\xi,\sigma)=\mathbb{E}_{\xi,\sigma}[&2h(\hat\beta^{0'}B\hat\beta^0,s^2(y))\operatorname{tr}((X'X)^{-1}QC)\\&+4h_1'(\hat\beta^{0'}B\hat\beta^0,s^2(y))\hat\beta^{0'}B(X'X)^{-1}QC\hat\beta^0\\&-\hat\beta^{0'}C'QC\hat\beta^0\frac{h(\hat\beta^{0'}B\hat\beta^0,s^2(y))}{(n-k)s^2(y))}\\&\times\{h(\hat\beta^{0'}B\hat\beta^0,s^2(y))(n-k+2)+h_2'(\hat\beta^{0'}B\hat\beta^0,s^2(y))s^2(y)\}]\end{aligned}$$

Ce gain dépend, de manière indirecte, des valeurs propres $\lambda_i(1\leq i\leq k)$ de X′X. Il est possible de trouver C et B (dépendant de X′X) tels que ce gain soit uniformément positif quel que soit le spectre de X′X, i.e. d'obtenir un estimateur minimax en situation de multicolinéarité. Nous allons voir dans les paragraphes suivants en quoi ce résultat est artificiel.

4.3. Une condition nécessaire de domination uniforme

La condition (5) entraîne des restrictions assez importantes sur le choix de C en situation de multicolinéarité. Étant donnée la famille $(X_t)_{t \in T}$, on considère les estimateurs

$$(6) \qquad \varphi_t(y) = \beta^{0\,(t)}(y) - h_t(\hat\beta^{0\,(t)\prime} B_t \hat\beta^{0\,(t)}, s^2(y)) \,.\, C_t \hat\beta^{0\,(t)}(y).$$

Supposons Q diagonalisable dans la base $(g_1, \ldots, g_k)$ (c'est en particulier le cas si $Q = I_k$); on obtient alors le résultat suivant :

Proposition 3 : En situation de multicolinéarité sévère, si C_t est indépendante de t, φ_t ne peut dominer uniformément l'estimateur des moindres carrés quel que soit t.

Démonstration: Notons D la transformée de QC_t dans la base $(g_1, \ldots, g_k)$, supposée indépendante de t par hypothèse. Soit d_{ij}, $1 \leqslant i,\ j \leqslant k$, l'élément générique de D. On a alors, pour tout $1 \leqslant i,\ j \leqslant k$,

$$\begin{aligned} \operatorname{pgvp}((X_t' X_t)^{-1} QC_t) &\geqslant \|(X_t' X_t)^{-1} QC_t g_i\| \\ &= \left\| \sum_{j=1}^{k} \frac{d_{ji}}{\lambda_j^{(t)}} g_j \right\| \\ &= \left(\sum_{j=1}^{k} \left(\frac{d_{ji}}{\lambda_j^{(t)}} \right)^2 \right)^{1/2} \\ &\geqslant \frac{|d_{ji}|}{\lambda_j^{(t)}} \end{aligned}$$

Or la proposition (5) s'écrit

$$\sum_{i=1}^{k} \frac{d_{ii}}{\lambda_i^{(t)}} - 2 \operatorname{pgvp}((X_t' X_t)^{-1} QC_t) \geqslant 0$$

On a donc, pour tout $1 \leqslant j \leqslant k$,

$$(7) \qquad \sum_{i=1}^{k} \frac{d_{ii}}{\lambda_i^{(t)}} - 2 \frac{|d_{ji}|}{\lambda_j^{(t)}} \geqslant 0.$$

Par hypothèse, il existe i_0 tel que $\lim\limits_{\varkappa_t \to \infty} \dfrac{\lambda_{i_0}^{(t)}}{\lambda_j^{(t)}} = 0$ pour $j \neq i_0$. Donc, si $d_{i_0 i_0} \neq 0$, le premier terme de (7) tend vers $-\infty$; s'il existe j_0 tel que $d_{i_0 j_0} \neq 0$, le second terme tend vers $-\infty$ avec $-\dfrac{|d_{i_0 j_0}|}{\lambda_i}$. Si $d_{i_0 j} = 0$ pour $1 \leqslant j \leqslant k$, il se peut que (7) soit positive pour tout t. Mais, dans ce cas, φ_t ne modifie pas la composante touchée par la multicolinéarité; par conséquent, le gain tend vers 0 quand κ_t tend vers l'infini. □

On déduit de la démonstration précédente que, pour que φ_t domine l'estimateur des moindres carrés pour tout t, il faut que QC_t soit

« proportionnel » à $(X_t' X_t)$ asymptotiquement en κ_t, i.e. qu'il existe une matrice constante A telle que $QC_t = (X_t' X_t) A$. Nous allons étudier dans le paragraphe suivant le cas particulier $A = I_k$.

4.4. Un estimateur minimax en situation de multicolinéarité

Considérant $C_t = Q^{-1}(X_t' X_t)$ pour un choix convenable de B, on obtient des matrices qui diagonalisent simultanément. Si $h_t(v, u) = \alpha_t \dfrac{u}{v}$, fonction de rétrécissement introduite par James et Stein, on peut appliquer les résultats rappelés en paragraphe 4.2.

PROPOSITION 4 : L'estimateur

(8)
$$\tilde{\varphi}_t(y) = \hat{\beta}^{0\,(t)} - \alpha_t \frac{s^2(y)}{\hat{\beta}^{0\,(t)'} B_t \hat{\beta}^{0\,(t)}} Q^{-1} X_t' X_t \hat{\beta}^{0\,(t)}$$

domine l'estimateur des moindres carrés pour le risque associé à Q si

$$\alpha_t \leqslant 2 \frac{(k-2)(n-k)}{(n-k+2)\,\mathrm{pgvp}(X_t' X_t Q^{-1} X_t' X_t B_t^{-1})}.$$

Le gain de $\tilde{\varphi}_t$ est

(9)
$$G^t(\xi, \sigma) = \alpha_t \mathbb{E}_{\xi, \sigma}\left[\frac{1}{\hat{\beta}^{0\,(t)'} B_t \hat{\beta}^{0\,(t)}} \times \left\{ -\alpha_t \frac{n-k+2}{n-k} \frac{\hat{\beta}^{0\,(t)'} X_t' X_t Q X_t' X_t \hat{\beta}^{0\,(t)}}{\hat{\beta}^{0\,(t)'} B_t \hat{\beta}^{0\,(t)}} + 2(k-2) \right\}\right].$$

La proposition ci-dessus entraîne des contraintes sur la matrice B_t. Puisqu'elle est diagonalisable dans la base $(g_1, \ldots, g_k)$, *si* B_t *est indépendante de* t, il existe $b_1, \ldots, b_k$ tels que $\hat{\beta}^{0\,(t)'} B_t \hat{\beta}^{0\,(t)} = \sum_{i=1}^{k} b_i \hat{\eta}_i^{0\,(t)2}$; cette expression tend vers $+\infty$ avec κ_t. Par conséquent, $G^t(\xi, \sigma)$ étant majoré par $\alpha_t \mathbb{E}_{\xi, \sigma}\left[\dfrac{2(k-2)}{\hat{\beta}^{0\,(t)'} B \hat{\beta}^{0\,(t)}}\right]$, le gain de $\hat{\varphi}_t$ tend vers 0 quand κ_t tend vers $+\infty$.

Une alternative relativement naturelle est de considérer $B_t = X_t' X_t$. Dans ce cas, $G^t(\xi, \sigma)$ est égal à

$$\alpha_t \mathbb{E}_{\xi, \sigma}\left[\frac{1}{\xi_t^{0'} \xi_t^0} \left\{ -\alpha_t \frac{n-k+2}{n-k} \frac{\xi_t^{0'} X_t Q^{-1} X_t \xi_t^0}{\xi_t^{0'} \xi_t^0} + 2(k-2) \right\}\right]$$

où $\hat{\xi}_t^0 = X_t \hat{\beta}^{0(t)}$. Par conséquent, si

$$\alpha_t = \frac{(k-2)(n-k)}{(n-k+2)\,\mathrm{pgvp}(X_t' X_t Q^{-1})},$$

$G^t(\xi, \sigma)$ est minoré par

$$\frac{(n-k)(k-2)}{(n-k+2)\,\mathrm{pgvp}(X_t' X_t Q^{-1})} \mathbb{E}_{\xi,\sigma}[(\hat{\xi}_t^{0\prime} \hat{\xi}_t^0)^{-1}],$$

qui, bien que dépendant de t, demeure strictement positif quel que soit t. Mais il ne tend pas vers l'infini avec κ_t. Par conséquent, en situation de multicolinéarité sévère, le gain relatif tend vers 0 quand κ_t tend vers $+\infty$.

Dans le cas particulier où $n=8$, $k=4$, nous avons effectué des simulations pour évaluer la décroissance du gain. Les valeurs propres de $(X_t' X_t)$ sont $\lambda_1 = (0,5)^{t-1}$, $\lambda_2 = 1,5$, $\lambda_3 = 2$ et $\lambda_4 = 10$, pour $t = 1, 2, \ldots, 15$; σ est pris égal à 1. Le risque de

$$\tilde{\varphi}_t(y) = \hat{\beta}^{0(t)} - \frac{2}{15} \frac{s^2(y)}{\hat{\xi}_t^{0\prime} \hat{\xi}_t^0} X_t' X_t \hat{\beta}^{0(t)} \tag{10}$$

est donné dans le tableau 1 pour le coût quadratique usuel et plusieurs valeurs du paramètre β

$$\beta^{(1)} = (0,1;\ -0,3;\ -0,25;\ 0,15),$$
$$\beta^{(2)} = (1,36;\ -2,3;\ -1,22;\ 0,97),$$
$$\beta^{(3)} = (4,22;\ 10,3;\ -4,52;\ 1,39).$$

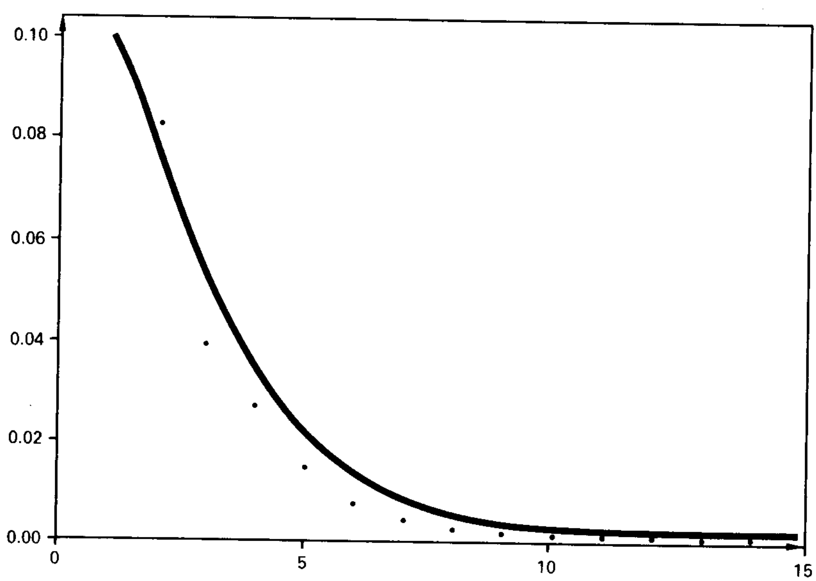
Comme on peut le remarquer dans le tableau 1 ou sur le graphe 1, même si le gain de l'estimateur $\tilde{\varphi}_t$ augmente avec la multicolinéarité, le gain relatif tend à s'annuler. Ce mauvais comportement des estimateurs minimax est conforme aux considérations évoquées dans l'introduction : en situation de multicolinéarité sévère, le problème devient essentiellement unidimensionnel et le risque d'un estimateur est asymptotiquement (i.e. lorsque κ_t tend vers $+\infty$) égal à celui de l'estimateur de la composante touchée par la multicolinéarité. Il n'est donc pas étonnant que les estimateurs minimax de cette composante tendent tous vers l'estimateur des moindres carrés correspondant : c'est l'unique estimateur possible.

A l'opposé, nous allons introduire dans le paragraphe suivant une classe d'estimateurs améliorant l'estimation des composantes touchées par la multicolinéarité.

TABLEAU 1

Risque de l'estimateur des moindres carrés, $\hat{\beta}^0$, et risques et gains de l'estimateur (10) pour diverses valeurs du paramètre β, en fonction de t

		0		$\beta^{(1)}$		$\beta^{(2)}$		$\beta^{(3)}$	
t	$\hat{\beta}^0$	R	G	R	G	R	G	R	G
1....	2,40	2,18	0,22	2,22	0,19	2,38	0,02	2,40	0
2....	3,12	2,88	0,24	2,91	0,21	3,10	0,02	3,12	0
3....	5,80	5,58	0,22	5,60	0,20	5,78	0,02	5,60	0
4....	9,32	9,07	0,25	9,09	0,23	9,30	0,02	9,31	0,01
5....	16,56	16,31	0,25	16,34	0,22	16,52	0,04	16,54	0,02
6....	34,79	34,52	0,27	34,55	0,84	34,73	0,06	34,74	0,05
7....	62,26	61,95	0,31	61,98	0,28	62,16	0,10	62,18	0,08
8....	144,12	143,72	0,40	143,74	0,38	143,93	0,19	143,94	0,18
9....	254,00	253,45	0,55	253,48	0,52	253,68	0,22	253,68	0,22
10....	511,68	510,82	0,86	510,85	0,83	511,03	0,65	511,08	0,60
11....	1 076,70	1 075,15	1,55	1 075,18	1,53	1 075,34	1,36	1 075,37	1,33
12....	1 966,47	1 963,83	2,64	1 943,85	2,62	1 964,04	2,43	1 964,06	2,41
13....	4 358,50	4 352,92	5,58	4 352,95	5,53	4 353,10	5,40	4 353,15	5,40
14....	8 405,07	8 394,55	10,52	8 394,58	10,49	8 394,71	10,36	8 994,77	10,30
15....	15 613,50	15 994,10	21,40	15 595,10	21,40	15 594,20	21,30	15 594,30	21,00



GRAPHIQUE 1

Gains relatifs de (10) par rapport à l'estimateur des moindres carrés en fonction de t pour $\beta = 0$.

5 Composantes principales

5.1. Estimateurs en composantes principales

Dans ce paragraphe, nous considérons le risque associé au coût quadratique usuel (la généralisation à une matrice de coût diagonalisable dans la base propre de $X_t' X_t$ est relativement aisée mais alourdit considérablement les écritures).

Considérant l'estimateur en composantes principales d'ordre p $(1 \leqslant p \leqslant k)$, introduit en paragraphe 3.2, on voit qu'il peut s'écrire sous la forme (4) en prenant h égal à 1, C_t égal à $P' \begin{bmatrix} c_1^t & & \\ & \vdots & \\ & & c_k^t \end{bmatrix} P$, où P est la matrice de passage dans la base propre $(P' X_t' X_t P = \mathrm{diag}(\lambda_i))$ et où c_i^t vaut 1 s'il correspond à l'indice de l'une des p plus petites valeurs propres de $X_t' X_t$, 0 sinon.

Si $J_p^t = \{ i ; 1 \leqslant i \leqslant k, c_i^t = 1 \}$, on a :

Proposition 5 : Le gain de l'estimateur en composantes principales d'ordre p est

$$G_p^t(\xi, \sigma) = \sum_{J_p^t} \left\{ \frac{1}{\lambda_i^{(t)}} - \frac{(\eta_i^{(t)})^2}{\sigma^2} \right\} \tag{11}$$

Démonstration: La différence des risques s'écrit

$$G_p^t(\xi, \sigma) = \sum_{i=1}^{k} \frac{1}{\lambda_i^{(t)}} - \frac{1}{\sigma^2} \{ \sum_{J_p^t} (\eta_i^{(t)})^2 + \sum_{(J_p^t)^c} \mathbb{E}_{\xi, \sigma} [(\hat{\eta}_i^{\alpha\,(t)} - \eta_i^{(t)})^2] \}. \quad \square$$

On voit donc bien que l'estimateur en composantes principales ne peut dominer uniformément l'estimateur des moindres carrés, même à t fixé, puisque, pour tout p, il existe des valeurs des $\eta_i^{(t)}$ $((i \in J_p^t)$ telles que (11) soit négatif. Par contre, la différence des risques croît avec la multicolinéarité, approximativement à la même vitesse que le risque de l'estimateur des moindres carrés $\left(\sum_{i=1}^{k} (\lambda_i^{(t)})^{-1} \right)$, d'où l'intérêt des estimateurs en composantes principales.

Si, en pratique, le choix de l'ordre p ne pose pas de problèmes (une fois identifiée la présence de multicolinéarité), la forme de l'estimateur en composantes principales paraît quelque peu brutale. Ainsi, dans l'exemple 2.2, l'ordre est manifestement 1; la première composante principale de β_θ

sera donc estimée par 0. L'estimateur introduit en paragraphe 5.3 généralise de manière naturelle cet estimateur en proposant un ordre dépendant de l'observation y. Au contraire de l'estimateur en composantes principales, il présente également l'avantage d'avoir un risque uniformément borné (en ξ).

5.2. Un estimateur « intuitif »

Dans son article de 1985, Casella remarque à plusieurs reprises que les estimateurs ridges qu'il propose ont un comportement *anti-intuitif*: ils rétrécissent d'autant moins des composantes que celles-ci ont une grande variance, donc sont plus instables. Il en va de même des estimateurs minimax étudiés en paragraphe 4, puisque les matrices de rétrécissement sont proportionnels à $X_t' X_t$.

Afin de stabiliser les composantes avec une grande variance, qui sont en fait les composantes touchées par la multicolinéarité, un choix naturel est de prendre C_t proportionnel à $(X_t' X_t)^{-1}$. On en définit donc

$$(12) \qquad \varphi_t'(y) = \hat\beta^{0\,(t)} - \alpha_t \frac{s^2(y)}{\hat\beta^{0\,(t)'} X_t' X_t \hat\beta^{0\,(t)}} (X_t' X_t)^{-1} \hat\beta^{0\,(t)},$$

en conservant la fonction de rétrécissement de James et Stein.

D'autre part, on définit *l'indice de multicolinéarité* de X_t comme

$$(13) \qquad \gamma_t = \sum_{i=1}^{k} (\lambda_i^{(t)})^{-2} - 2 \max((\lambda_i^{(t)})^{-2}.$$

Il correspond à la notion d'« effective dimension » (ED) introduite par THISTED [1982] pour séparer les situations de multicolinéarité; si $ED \leqslant 2$ (i. e. si $\gamma_t \leqslant 0$) il est impossible de proposer des estimateurs (12) dominant uniformément l'estimateur des moindres carrés. On dira que l'on est en *situation de multicolinéarité partielle* si cette domination uniforme peut être atteinte, i. e. si $\gamma_t > 0$. L'énoncé suivant découle alors de la proposition 1.

PROPOSITION 6 : En situation de multicolinéarité partielle, l'estimateur φ_t' domine uniformément l'estimateur des moindres carrés si

$$\alpha_t \leqslant 2\gamma_t \frac{n-k}{n-k+2} \min(\lambda_i^{(t)})^3$$

Cette notion est plus forte que celle introduite par CASELLA [1985]

$$\sum_{i=1}^{k} \left(\frac{\max \lambda_i^{(t)}}{\lambda_i^{(t)}}\right)^2 - \left(\frac{\max \lambda_j^{(t)}}{\min \lambda_j^{(t)}}\right)^2 > 1 + \frac{\max \lambda_j^{(t)}}{\min \lambda_j^{(t)}}.$$

Dans l'exemple 2.2, l'indice de multicolinéarité vaut

$$\gamma(X_\theta)=\frac{1}{4\cos^4\theta}+\frac{1}{4\sin^4\theta}+1-\frac{1}{2\sin^4\theta}=1-4\frac{\cos 2\theta}{\sin^2 2\theta}$$

Nous serons donc en situation de multicolinéarité partielle pour $\theta>38°$, ce qui est assez restrictif.

5.3. Généralisation des estimateurs en composantes principales

Comme l'ont montré JUDGE et BOCK [1978, p. 239], on améliore toujours, du point de vue du risque quadratique, un estimateur du type (12) en le remplaçant par sa version tronquée grâce à la «*positive part*», c'est-à-dire en remplaçant par 0 les estimations des composantes qui ne sont pas du signe de l'estimateur des moindres carrés de ces composantes. Nous allons appliquer ce résultat à l'estimateur φ_t' en travaillant dans la base en composantes principales.

On considère donc le nouvel estimateur

$$(14)\qquad \varphi_t^+(y)=P'\left[I_k-\alpha_t\frac{s^2(y)}{\hat\beta^{0\,(t)'}X_t'X_t\hat\beta^{0\,(t)}}\begin{bmatrix}(\lambda_1^{(t)})^{-1} & & \\ & \ddots & \\ & & (\lambda_k^{(t)})^{-1}\end{bmatrix}\right]^+\hat\eta^{0\,(t)}.$$

Par conséquent, pour $1\leqslant i\leqslant k$, $\eta_i^{(t)}$ sera estimé par

$$(15)\qquad \text{soit}\begin{cases}\hat\eta_i^{0\,(t)}-\alpha_t\dfrac{s^2(y)}{\sum\limits_{j=1}^{k}\lambda_j^{(t)}(\hat\eta_j^{0\,(t)})^2}\dfrac{\hat\eta_i^{0\,(t)}}{\lambda_i^{(t)}} & \\ \qquad\qquad \text{si}\quad \alpha_t s^2(y)\leqslant\lambda_i^{(t)}\sum\limits_{j=1}^{k}\lambda_j^{(t)}(\hat\eta_j^{0\,(t)})^2, & \\ 0\qquad \text{sinon.} & \end{cases}$$

La décomposition (15) de l'estimateur φ_t^+ montre l'intérêt de la troncature en 0 et du choix de la matrice C_t: les premières composantes de $\eta(t)$ estimées par 0 seront celles correspondant aux plus petites valeurs propres. Ainsi, cet estimateur affine les estimateurs en composantes principales car il donne une technique de sélection de l'ordre qui dépend non seulement des valeurs propres mais aussi des observations.

Si $\gamma_t>0$, pour un choix convenable de α_t (donné par la proposition 6), φ_t^+ domine l'estimateur des moindres carrés. Si, au contraire, $\gamma_t\leqslant 0$, il découle de la condition (5) que φ^+ n'est pas minimax. Cependant en état de multicolinéarité sévère, lorsque $\min\lambda_i^{(t)}$ est suffisamment petit, φ_t^+ estime dans la plupart des cas la composante responsable de la multicolinéarité par 0. Par conséquent, dans un voisinage de 0 (pour la composante estimée par 0), le gain obtenu sur les autres composantes permet d'obtenir un risque inférieur à celui des moindres carrés (voir simulation ci-dessous).

Quelle que soit la situation de multicolonéarité, φ_t^+ *ne peut dominer uniformément les estimateurs en composantes principales.* En effet, considérons l'estimateur d'ordre 1 et supposons, pour simplifier, que la composante touchée par la multicolinéarité corresponde à l'indice 1. Si $\eta_1^{(t)}=0$, il découle de la proposition 5 que le risque de cet estimateur est $\sum_{i=2}^{k} (\lambda_i^{(t)})^{-1}$, i.e. le risque de l'estimateur des moindres carrés pour les *autres* composantes. Supposons d'autre part que $\eta_2^{(t)}, \ldots, \eta_k^{(t)}$ soient relativement grands. Même si φ_t^+ domine l'estimateur des moindres carrés, le risque de φ_t^+ sera approximativement égal au risque de l'estimateur des moindres carrés, soit $\sum_{i=1}^{k} (\lambda_i^{(t)})^{-1}$, donc supérieur au risque de l'estimateur en composantes principales.

On a cependant le résultat asymptotique suivant (en supposant $n-k>2$).

PROPOSITION 7 : $\mathrm{I}(\lambda_i^{(t)})=\mathbb{E}_{\xi,\sigma}[(\hat{\eta}_1^{(t)+}-\eta_1^{(t)})^2]-(\eta_1^{(t)})^2$ tend vers 0 avec $\lambda_i^{(t)}$.

Démonstration: Nous allons effectuer la démonstration dans le cas où $\eta_1^{(t)}=0$. Le cas général se déduit de manière directe. Il vient

$$
\begin{aligned}
\mathrm{I}(\lambda_i^{(t)}) &= \mathbb{E}_{\xi,\sigma}((\hat{\eta}_1^{(t)})^+)^2 \\
&\leqslant \mathbb{E}_{\xi,\sigma}\left[(\hat{\eta}_1^{0(t)})^2\, \mathbb{1}_{[\alpha/\lambda_i^{(t)},\,+\infty[}\left(\frac{\sum_{i=1}^{k}\lambda_i^{(t)}(\hat{\eta}_i^{0\,(t)})^2}{s^2(y)}\right)\right] \\
&\leqslant \left\{\mathbb{E}_{\xi,\sigma}((\hat{\eta}_1^{0\,(t)})^4)\,\mathbb{E}_{\xi,\sigma}\left[\mathbb{1}_{[\alpha/\lambda_i^{(t)},\,+\infty[}\left(\frac{\sum_{i=1}^{k}\lambda_i^{(t)}(\hat{\eta}_i^{0\,(t)})^2}{s^2(y)}\right)\right]\right\}^{1/2} \\
&\leqslant \sqrt{3}\,\frac{\sigma^2}{\lambda_i^{(t)}}\left(\mathrm{P}\left(\mathrm{F}\geqq\frac{\alpha}{k}\,\frac{1}{\lambda_i^{(t)}}\right)\right)^{1/2},
\end{aligned}
$$

où F suit la loi de Fisher décentrée $\mathscr{F}\left(k, n-k;\ \sum_{i=1}^{n}\lambda_i^{(t)}(\eta_i^{(t)})^2\right)$. L'inégalité de Tchebychev entraîne

$$
\mathrm{P}\left(\mathrm{F}\geqslant\frac{\alpha}{k}\,\frac{1}{\lambda_i^{(t)}}\right)<\frac{\mathbb{E}(\mathrm{F}^{5/2})}{\alpha^{5/2}}\,k^{5/2}\,(\lambda_i^{(t)})^{5/2}
$$

donc $\mathrm{I}(\lambda_i^{(t)})$ est bornée supérieurement par un terme proportionnel à $(\lambda_i^{(t)})^{1/4}$. □

On déduit de ce résultat que, pour toute valeur du paramètre $\eta^{(t)}$, il existe une situation de multicolinéarité suffisamment sévère pour que φ_t^+ domine l'estimateur en composantes principales correspondant. Afin d'obtenir la domination pour les composantes non touchées par la multicolinéarité, il suffit de prendre, si la composante touchée par la multicolinéarité correspond à l'indice 1,

$$(16) \qquad \alpha \leqslant 2 \sum_{i=2}^{k} \frac{(\lambda_i^{(t)})^{-2} - 2(\min\limits_{i \geqslant 2} \lambda_i^{(t)})^{-2}}{(\min\limits_{i \geqslant 2} \lambda_i^{(t)})^{-3}} \; \frac{n-k+1}{n-k+3}.$$

Si $k=3$, comme dans l'exemple 2.2, la domination uniforme est bien sûr impossible.

Sous les conditions exposées à la fin du paragraphe 4.4, les risques de

$$(17) \qquad \varphi_t^+(y) = \mathrm{P}' \left[\mathrm{I}_4 - (0{,}7) \frac{s^2(y)}{\hat{\beta}^{0\,(t)\prime} \mathrm{X}_t' \mathrm{X}_t \hat{\beta}^{0\,(t)}} \begin{bmatrix} (\lambda_1^{(t)})^{-1} & & & \\ & (\lambda_2^{(t)})^{-1} & & \\ & & (\lambda_3^{(t)})^{-1} & \\ & & & (\lambda_4^{(t)})^{-1} \end{bmatrix} \right]^+ \hat{\eta}^{0\,(t)}(y)$$

pour diverses valeurs de t et de β ont été évalués par simulation (voir tableau 2); la condition (16) n'étant pas remplie, nous avons choisi pour α la constante de James-Stein, $\dfrac{(k-3)(n-k+1)}{n-k+3}$.

TABLEAU 2

Risque de l'estimateur des moindres carrés, $\hat{\beta}^0$, et risques et gains de l'estimateur (17) pour diverses valeurs du paramètre β, en fonction de t

		0		$\beta^{(1)}$		$\beta^{(2)}$		$\beta^{(3)}$	
t	$\hat{\beta}^0$	R	G	R	G	R	G	R	G
1....	2,40	2,32	0,08	2,32	0,08	2,39	0,01	2,40	0,00
2....	3,12	2,91	0,21	2,93	0,19	3,09	0,03	3,12	0,00
3....	5,80	5,02	0,78	5,05	0,75	5,66	0,14	5,79	0,01
4....	9,32	6,67	2,63	6,78	2,54	8,86	0,52	9,25	0,07
5....	16,56	8,78	7,78	9,25	7,31	14,87	1,69	16,31	0,15
6....	34,79	12,57	22,22	13,26	21,33	28,26	6,53	33,67	1,12
7....	62,26	10,98	51,28	11,64	50,62	38,81	23,45	58,25	4,01
8....	144,12	6,87	137,25	7,56	136,56	54,99	89,13	125,45	18,67
9....	254,00	5,26	248,74	5,77	248,23	50,67	203,33	187,06	66,94
10....	511,68	,16	509,52	5,84	505,84	44,21	467,47	297,34	214,34
11....	1 076,70	1,30	1 075,40	2,81	1 073,89	45,39	1 031,31	388,90	687,80
12....	1 966,47	1,13	1 965,34	1,36	1 963,11	8,65	1 957,82	235,47	1 731,00
13....	4 358,50	1,28	4 357,22	1,20	4 357,30	14,24	4 344,26	152,90	4 205,60
14....	8 405,07	1,14	8 403,93	1,34	8 403,73	2,06	8 403,01	211,38	8 193,69
15....	15 613,50	1,23	15 612,27	1,24	15 612,26	2,47	15 611,03	54,97	15 527,53

Lorsqu'on observe les résultats obtenus, on constate le phénomène attendu, à savoir une certaine croissance du risque de φ_t^+ puis, pour un degré de multicolinéarité suffisant (qui dépend de la valeur des paramètres), un «décrochage» vers un risque très faible. Bien que φ_t^+ ne soit pas minimax, on peut noter que, pour toutes les valeurs de β retenues, φ_t^+ domine l'estimateur des moindres carrés pour tout t.

Comme on peut le remarquer sur les graphes 2 et 3, le comportement de φ_t^+ vis-à-vis de l'estimateur en composantes principales d'ordre 1 est conforme au résultat de la proposition 1 : quel que soit $\eta^{(t)}$, pour t suffisamment grand, φ_t^+ domine cet estimateur (dont le risque est indépendant de t d'après la proposition 5).

Remarque: La représentation graphique établie au vu de l'échantillon ne prétend pas donner une vision exacte du risque, en particulier au niveau de la «zone médiane» où la troncature en 0 ne s'applique pas totalement.

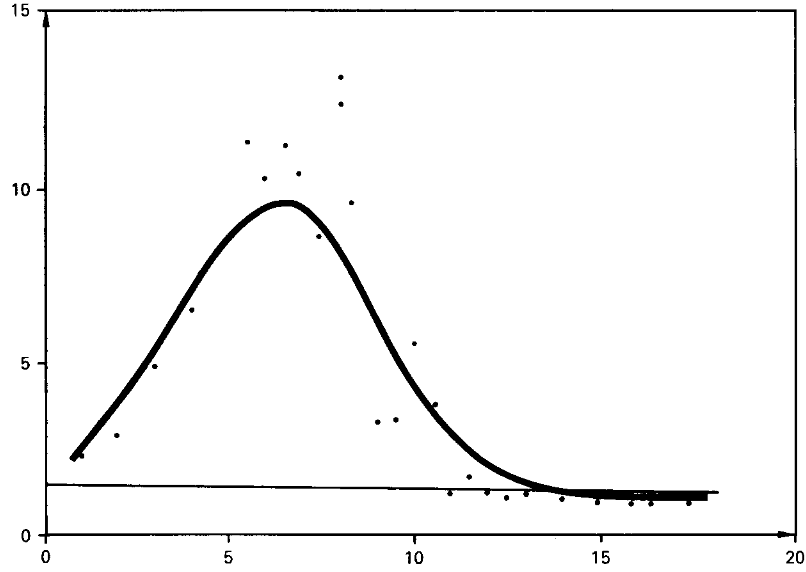
5.4. Conclusion

Lorsque la condition de multicolinéarité partielle n'est pas satisfaite, les deux objectifs exposés dans l'introduction ne sont pas compatibles. En pratique, l'expérimentateur devra choisir entre un estimateur assurant une domination uniforme de l'estimateur des moindres carrés comme ceux du type (8) et un estimateur intuitif, comme ceux de la classe (14), améliorant l'estimation des composantes touchées par la multicolinéarité. Comme nous l'avons vu en paragraphe 4, le critère de domination uniforme est très contraignant en situation de multicolinéarité et, pour un degré de multicolinéarité suffisant, le gain devient illusoire. A l'opposé, les simulations ont montré que l'estimateur (17), tout en assurant une meilleure estimation des composantes touchées par la multicolinéarité, permettait également un gain du point de vue du risque pour les valeurs des paramètres envisagées.

Le fait que les estimateurs de la classe (14) aient un risque fini, dominent l'estimateur des moindres carrés en situation de multicolinéarité partielle et dominent asymptotiquement les estimateurs en composantes principales nous incite à rejeter ces derniers au profit des estimateurs (14), plus souples face aux observations.

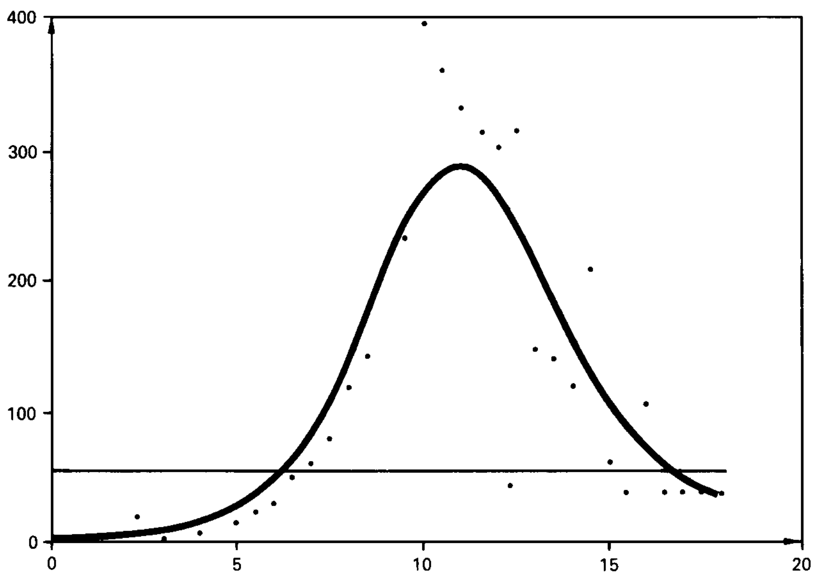
Au vu des résultats précédents, il nous semble justifié de reprendre les conclusions auxquelles CASELLA [1985] aboutit pour les estimateurs ridges : il est préférable, en situation de multicolinéarité, de sacrifier l'impératif d'uniforme domination au profit de la stabilisation numérique en utilisant un estimateur intuitif («. . . *the wisest course seems to be the forfeiture of minimaxity to achieve numerical stability*»).

Je tiens à remercier Jean-Pierre Raoult pour son aide lors de la rédaction de ce papier. Les commentaires d'Alain Monfort et de deux referees ont permis d'améliorer grandement sa présentation.



GRAPHIQUE 2

Risques de l'estimateur (17) et de l'estimateur en composantes principales pour $\beta = 0$ en fonction de t.



GRAPHIQUE 3

Risques de l'estimateur (17) et de l'estimateur en composantes principales pour $\beta = \beta^{(3)}$ en fonction de t.

• Références bibliographiques

BROWN, L. (1975). – « Estimation with Incompletely Specified Loss Functions », *JASA*, 70, p. 417-427.

CASELLA, G. (1980). – « Minimax Ridge Regression Estimation », *Ann. Stat.*, 8, p. 1036-1056.

CASELLA G. (1985). – « Condition Number and Minimax Ridge Regression Estimators », *JASA*, 80, p. 753-758.

CELLIER, D., FOURDRINIER, D. et ROBERT, (1986). – « Les estimateurs à rétrécisseur contrôlé », *Doc. Travail*, Université de Rouen, 1986-01.

FLURY, B. (1988). – *Common Principal Components*, à paraître chez Wiley.

FOMBY, T. et HILL, R. (1979). – « Multicolinearity and the Minimax Conditions of the Bock-Stein-Like Estimator », *Econometrica*, 47, p. 211-212.

FRAISSE, A. M., ROBERT, C. et ROY, M. (1987). – « Estimateurs à rétrécisseurs différentiable, pour un coût quadratique général », *Annales d'Économie et de Statistique*, 8.

HILL, R. et ZIEMER, R. (1982). – « The Application of Generalized Ridge and Stein-Like General Minimax Rules to Multicolinear Data », *Comm. Stat.*, A 11 (6), p. 623-638.

HILL, R. et ZIEMER, R. (1984). – « The Risk of General Stein-Like Estimators in the Presence of Multicolinearity », *J. Econometrics*, 25, p. 205-216.

HOCKING, R. (1976). – « The Analysis and Selection of Variables in Linear Regression », *Biometrics*, 32, p. 1-49.

HOERL, A. et KENNARD, A. (1970). – « Ridge-Regression: Applications to Non-Orthogonal Problems », *Technometrics*, 12, p. 62-82.

JAMES, W. et STEIN, C. (1961). – « Estimation with Quadratic Loss », *Proc. 4th Berkeley Symposium*, 1, p. 361-373.

JOHNSTON, S., REIMER, S. et ROTHROCK, T. (1973). – « Principal Components and the Problem of Multicolinearity », *Metroeconometrica*, 25 (3), p. 306-317.

JUDGE, G. et BOCK, M. (1978). – *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland.

STEIN, C. (1956). "Inadmissibility of the Usual Estimator of the Mean of a Multivariate Normal Distribution", *Proc. 3rd Berkeley Symposium*, 1, p. 197-206.

STEWARD, G. (1987). – « Collinearity and Least Squares Regression », *Stat. Sciences*, 2 (1), p. 68-100.

STRAWDERMAN, W. (1978). – « Minimax Adaptive Generalized Ridge Regression Estimators, *JASA* 73, p. 623-627.

THISTED, R. (1982). – « Decision Theoretic Regression Diagnostics », *Stat. Dec. Theo. and Rel. Topics*, III, (2), p. 363-382, SS. GUPTA et J. BERGER, éd., Academic Press.

THISTED, R. (1987). – « Comments on "Collinearity and Least Squares Regression" », *Stat. Sciences*, 2, (1), p. 91-93.

THISTED, R. et MORRIS, C. (1980). – « Theoretical Results for Adaptive Ordinary Ridge Regression Estimators », *Tech. Report*, 94, Univ. Chicago.