REVUE DE STATISTIQUE APPLIQUÉE

C. DENIAU
G. OPPENHEIM

Un critère simple auquel satisfont des estimateurs concurrents des moindres carrés

Revue de statistique appliquée, tome 24, n° 4 (1976), p. 35-42
<http://www.numdam.org/item?id=RSA_1976__24_4_35_0>

UN CRITÈRE SIMPLE AUQUEL SATISFONT DES ESTIMATEURS CONCURRENTS DES MOINDRES CARRÉS *

C. DENIAU et G. OPPENHEIM

U.E.R. de Mathématiques, Logique Formelle et Informatique

Université de Paris V, 12, rue Cujas - 75005 Paris

RESUME

L'étude de voisinages-boule de l'estimateur des moindres carrés (ou de l'estimateur identiquement nul) permet de montrer que des estimateurs usuels concurrents des moindres carrés satisfont à un critère quadratique très simple. Les métriques intervenant dans la définition des boules et du critère, sont les paramètres du problème.

INTRODUCTION

Dans la théorie de l'estimation, l'estimateur des moindres carrés $\hat{\beta}$ (L.S.) du paramètre β (ainsi que le B.L.U.E. d'ailleurs) a été souvent critiqué :

1) Il est peu robuste vis à vis d'éventuels outliers provenant par exemple de la contamination de la loi de ϵ (HUBER [3]).

2) La variance de ses coordonnées peut être très grande sous l'effet des faibles valeurs propres de ${}^tXX$; pour $m = 1$ et $\Sigma_\epsilon = \sigma^2 I$, $\operatorname{tr} \Sigma_\beta = \sigma^2 \sum_{j=1}^{p} (1/\lambda_j ({}^tXX)$ où $(\lambda_j ({}^tXX))_{1 \leqslant j \leqslant p}$ est la suite des valeurs propres de ${}^tXX$ (HOERL-KENNARD [2a , 2b]).

3) Sous les conditions de STEIN [8] ou de SCLOVE [7], il est inadmissible pour $p \geqslant 3$ pour les fonctions de pertes quadratiques.

Les recherches d'estimateurs pouvant remplacer le L.S. jugé défaillant, sont basées sur divers principes ; citons :

1) L'utilisation d'estimateurs légèrement biaisés (RIDGE REGRESSION [2a , 2b]) ;

2) L'utilisation d'estimateurs $\hat{b}$ plus courts, c'est-à-dire tels que pour tout $\omega \in \Omega$ la norme de $\hat{b}(\omega)$ soit inférieure à celle de $\hat{\beta}(\omega)$ (JAMES-STEIN [4]).

3) L'élaboration d'estimateurs solution de problème MINIMAX ; cette procédure est adaptée, en particulier, lorsqu'on dispose d'informations complémentaires sur le domaine $\mathcal{B}$ de variation possible de β. On cherche alors $\hat{b}$ solution de $\operatorname{Inf}_{\hat{b} \in \mathcal{A}} \operatorname{Sup}_{\beta \in \mathcal{B}} E| \, |\beta - \hat{b}| \, |$ où $\mathcal{A}$ est un ensemble à préciser (BUNKE [1], LAUTER [5], ROZANOV [6]).

(*) Ce texte a fait l'objet d'une communication au "Congrès Européen de la Statistique" GRENOBLE 1976.

Mots Clefs : **Régression multidimensionnelle. Estimateur des moindres carrés. Estimateurs Ridge et de Stein.**

Dans le présent texte nous montrons que la majeure partie des estimateurs concurrents de L.S. et rassemblés par BUNKE [1], peuvent être introduits par une problématique tout à fait simple associée à des critères quadratiques (partie 2) : il s'agit de choisir un estimateur dans le voisinage d'estimateurs privilégiés. Comme estimateurs privilégiés, nous étudierons le L.S. et l'estimateur identiquement nul (partie 1).

Le modèle linéaire multidimensionnel ($m \geqslant 1$) peut être précisé comme suit :

$$\underset{nm}{Z} = \underset{nm}{EZ} + \underset{nm}{\epsilon}$$

où Z et ϵ sont des variables aléatoires, ayant des seconds moments, définies sur $(\Omega, \mathfrak{A}, P)$, à valeurs dans $\mathbf{R}^{nm}$, d'espérance respective $EZ = \underset{np}{X}\ \underset{pm}{\beta}$ et et $E\epsilon = 0$; la matrice X est non aléatoire, de rang $p \leqslant n$.

1. PROPRIETE DES VOISINAGES-BOULES

Soit $\underset{pp}{\theta}$ une matrice symétrique définie positive. L'espace $\mathbf{R}^{pm}$ des matrices ($pm > 0$) est muni de la norme notée $\| \ \|_\Theta$, associée au produit scalaire noté $< | >_\Theta$ définit par :

$$Y \in \mathbf{R}^{pm}\ ,\ T \in \mathbf{R}^{pm}, < Y \mid T >_\Theta = \text{trace } {}^t Y\, \Theta\, T.$$

On note $y_k \in \mathbf{R}^p$ ($1 \leqslant k \leqslant m$) les colonnes de Y et I_p la matrice identité de $\mathbf{R}^{pp}$.

Définition 1.

On appelle voisinage-boule de rayon ϵ ($\geqslant 0$) de $a \in \mathbf{R}^{pm}$ l'ensemble

$$\{ d \mid d \in \mathbf{R}^{pm}\ ,\ \|d - a\|_\Theta^2 \leqslant \epsilon^2 \}$$

La variable aléatoire $\hat{b}$ appartient à un voisinage boule de $\hat{c}$ si et seulement si :

$$\forall\, \omega \in \Omega : \qquad \|\hat{b}(\omega) - \hat{c}(\omega)\|_\Theta^2 \leqslant \epsilon^2$$

On écrira alors

$$\|\hat{b} - \hat{c}\|_\Theta^2 \leqslant \epsilon^2$$

Propriété 1

Soit $\hat{b}$ un estimateur sans biais de β. Dans un voisinage-boule de $\hat{b}$ la norme du biais est borné :

$$\|\hat{d} - \hat{b}\|_\Theta^2 \leqslant \epsilon^2 \Rightarrow \sup_{\hat{d}} \|E\hat{d} - \beta\|_\Theta^2 \leqslant \epsilon^2.$$

Propriété 2

Dans un voisinage-boule de l'estimateur identiquement nul $\hat{O}$, on a

$$||\hat{d} - \hat{O}||_\Theta^2 \Rightarrow \sum_{k=1}^{m} \text{trace } (\Sigma_{d_k}\, \theta) \leqslant \epsilon^2$$

Corollaire de la propriété 2

$||\hat{d} - \hat{O}||_\Theta \leqslant \epsilon^2$ implique qu'il existe un scalaire $\eta = \eta(\epsilon)$ tel que Var $\hat{d}_{k_j} \leqslant \eta^2$, $\hat{d}_{j_k}$ étant les coordonnées de $\hat{d}$ $(1 \leqslant j \leqslant p,\ 1 \leqslant k \leqslant m)$.

DEMONSTRATIONS

Pour tout estimateur $\hat{u}$ on a

$$(1) \qquad E||\hat{u}||_\Theta^2 = \sum_{k=1}^{m} \text{trace } (\Sigma_{\hat{u}_k}\, \Theta) + ||E\hat{u}||_\Theta^2.$$

Propriété 1 : D'après (1)

$$(2) \qquad E||\hat{d} - \hat{b}||_\Theta^2 = \sum_{k=1}^{m} \text{trace } (\Sigma_{\hat{d}_k - \hat{b}_k}\, \Theta) + ||E\hat{d} - E\hat{b}||_\Theta^2$$

Cette quantité est majorée par ϵ^2 puisque :

$$\forall\, \omega \in \Omega: \qquad ||\hat{d}(\omega) - \hat{b}(\omega)||_\Theta^2 \leqslant \epsilon^2 \Rightarrow E||\hat{d} - \hat{b}||_\Theta^2 \leqslant \epsilon^2$$

Chacun des termes du membre de droite de (2) étant positif ou nul, est majoré par ϵ^2 : $||E\hat{d} - E\hat{b}||_\Theta^2 \leqslant \epsilon^2$, c'est-à-dire $||E\hat{d} - \beta|| \leqslant \epsilon^2$ puisque $\hat{b}$ est sans biais.

Propriété 2 : En écrivant (1) pour $\hat{u} = \hat{d}$, $\hat{b} = \hat{O}$ et tenant compte de l'hypothèse $||\hat{d}||_\Theta^2 \leqslant \epsilon^2$ qui implique

$$E||\hat{d}||_\Theta^2 \leqslant \epsilon^2\ ,\ E||\hat{d}||_\Theta^2 = \sum_{k=1}^{m} \text{trace } (\Sigma_{\hat{d}_k}\, \Theta) + ||E\hat{d}||_\Theta^2 \leqslant \epsilon^2$$

entraîne le résultat souhaité.

Corollaire : Montrons tout d'abord que

$$\text{trace } (\Sigma\Theta) \leqslant \epsilon^2 \Rightarrow \text{trace } \Sigma \leqslant \epsilon^2 \lambda_{\min}^{-1} = \eta^2$$

Le résultat cherché s'en déduira trivialement.

Σ étant symétrique et positive, il existe Σ_1 tel que $\Sigma = \Sigma_1\, {}^t\Sigma_1$; de plus, Θ admet des valeurs propres strictement positives, la plus petite étant notée $\lambda_{\min}$.

$$\text{trace } (\Sigma\Theta) = \text{trace } ({}^t\Sigma_1\, \Theta\, \Sigma_1) = ||\Sigma_1||_\Theta^2$$

La matrice $\Theta_1 = \lambda_{\min}\, I_p$ définit un produit scalaire sur R^{pm} qui est tel que

$$(3) \qquad \forall\ T \in R^{pm}\ :\ ||T||_{\Theta_1}^2 = \lambda_{\min}\, ||T||_I^2 \leqslant ||T||_\Theta^2$$

Appliquant (3) à Σ_1

$$||\Sigma_1||^2_\Theta = \lambda_{min}\ ||\Sigma_1||^2_I = \lambda_{min}\ \text{trace}\ \hat{\Sigma} \leqslant ||\hat{\Sigma}_1||^2 = \text{trace}\ (\hat{\Sigma}\ \Theta)$$

donc :

$$\text{trace}\ \hat{\Sigma} \leqslant \lambda^{-1}_{min}\ \text{trace}\ (\hat{\Sigma}\ \Theta) \leqslant \lambda^{-1}_{min}\ \epsilon^2$$

En conclusion de cette partie remarquons que ces propriétés élémentaires montrent qu'au voisinage des L.S., estimateurs sans biais, la norme du biais des estimateurs est majorée, la variance des coordonnées restant voisine de celle des L.S. Au voisinage de l'origine, la majoration concerne la variance, la norme du biais peut, bien sûr, être grande. Une idée consiste, à rechercher

– au voisinage L.S. : un estimateur proche de $\hat{O}$.
– au voisinage de $\hat{O}$: un estimateur proche des L.S.

2. PROBLEMES QUADRATIQUES

Dans ce qui suit, les matrices symétriques définies positives, ϕ , ϕ_1 (de forme n x n), ψ , ψ_1 (de forme p x p) doivent être considérés comme des paramètres de liberté laissés au choix de praticien.

On note $\beta = ({}^tX\ \phi\ X)^{-1}\ {}^tX\ \phi\ z$.

Soient les 4 fonctions suivantes de $b \in R^{mp}$:

$$(4)\quad \begin{cases} Q_1\ (b\ ,\ \psi) = ||b||^2_\psi & Q_2\ (b\ ,\ \phi) = ||Xb||^2_\phi = ||b||^2_{{}^tX\phi X} \\ & \qquad = Q_1\ (b,\ {}^tX\ \phi\ X) \\ Q_3\ (b\ ,\ \psi_1) = ||b - \beta||^2_{\psi_1} = Q_1\ (b-\beta, \psi_1) & Q_4\ (b,\ \phi_1) = ||Xb - X\beta||^2_{\phi_1} \\ & \qquad = Q_2\ (b\ -\ \beta,\ \phi_1) = Q_1\ (b - \beta\ ,\ {}^tX\phi_1 X) \end{cases}$$

Soit à déterminer, pour $1 \leqslant k \neq k' \leqslant 4$ les solutions b_o des problèmes notés $P_{kk'}$:

$Q_k\ (b_o) \leqslant Q_k\ (b)$ pour tout b tel que $Q'_k\ (b) \leqslant \epsilon^2$ (5) ce que l'on peut noter : $Q_1\ (b_o - a,\ A) \leqslant Q_1\ (b - a,\ A)$ pour tout b tel que $Q_1\ (b - c,\ C) \leqslant \epsilon^2$; les valeurs de a et c, de A et B sont déterminées à partir des expressions (4) ; a et c, égaux ou non, sont pris dans l'ensemble $\{\hat{o},\ \hat{\beta}\}$, A et C, égaux ou non, dans l'ensemble $\{\psi,\ \psi_1,\ {}^tX\ \phi\ X,\ {}^tX\ \phi_1\ X\}$.

Proposition 1

1. La solution de chacun des problèmes P_{kk}, définis par (5) existe et est unique.

2. Cette solution est formellement exprimée par :

(6) si $||a - c||^2_C \leqslant \epsilon^2$: $b_o = a$

(7) si $||a - c||^2_C > \epsilon^2$: $b_o = (A + \lambda_o C)^{-1}\ (Aa + \lambda_o Cc)$

λ_o étant déterminé par la propriété $||b_o - c||^2_C = \epsilon^2$

DEMONSTRATION

1. $Q_k(b)$ est la fonction continue de b qu'il s'agit de minimiser sur l'ensemble

$$E_c = \{b \mid Q_{k'}(b) \leqslant \epsilon^2\} = \{b \mid ||b - c||_C^2 \leqslant \epsilon^2\}$$

qui est ellipsoïde convexe fermé de R_{pm} de centre c. Le minimum existe et est atteint dans le domaine. La solution b_o est ou bien a si $a \in E_c$ ou bien la projection $P_{E_c}(a)$ (orthogonale pour le produit scalaire A) de a sur la frontière de E_c.

2. L'expression explicite s'obtient sans difficulté par la technique des multiplicateurs de Lagrange en annulant les dérivées partielles en x et λ de

$$U(x, \lambda) = ||x - a||_A^2 + \lambda (||x - c||_C^2 - \epsilon^2).$$

$$(8) \begin{cases} (A + \lambda_o C)\, b_o = Aa + \lambda_o\, Cc \\ \\ ||b_o - c||_C^2 = \epsilon^2 \end{cases} \Leftrightarrow (9) \begin{cases} A\,(bo - a) = -\lambda_o\, C(b_o - c) \\ \\ ||bo - c||_C^2 = \epsilon^2 \end{cases}$$

Remarquons que λ_o est positif ; $A + \lambda C$ somme de 2 matrices symétriques définies positives l'est aussi, et (8) devient :

$b_o = (A + \lambda_o\, C)^{-1}\, (Aa + \lambda_o\, Cc)$, λ_o étant déterminé par $||b_o - c||_C^2 = \epsilon^2$.

On a aussi

$$b_o = c + (A + \lambda_o\, C)^{-1}\ A\,(a - c)$$

Propriétés des solutions

i) Les solutions précédentes $b_o = P_{E_c}(a)$ sont obtenues dans un contexte géométrique. Dans le contexte probabiliste, les applications associées sont des variables aléatoires. En effet :

a) soit $\hat{b}_o$ l'application qui a tout $\omega \in \Omega$ associe $\hat{b}_o(\omega) = P_{E_o}\hat{\beta}(\omega)$, $\hat{b}_o$ est une variable aléatoire puisque P_{E_c} est continue ;

b) soit $\hat{b}_o$ l'application qui a tout $\omega \in \Omega$ associe

$$\hat{b}_o(\omega) = P_{E_{\hat{\beta}(\omega)}}\ (o) = P_{E_o}\ (-\hat{\beta}(\omega) + \hat{\beta}(\omega)),$$

c'est une variable aléatoire.

ii) Dans tous les cas étudiés, les estimateurs obtenus sont plus courts que β, c'est-à-dire que pout tout $\omega \in \Omega$:

$$||\hat{b}_o(\omega)||_A^2 \leqslant ||\hat{\beta}(\omega)||_A^2$$

Classement des familles de solutions

i) Les problèmes $P_{1,2}$ et $P_{2,1}$ admettent $\hat{O}$ pour solution ;
les problèmes $P_{3,4}$ et $P_{4,3}$ admettent $\hat{\beta}$ pour solution.

ii) Les solutions constituent une famille que l'on peut appeler (Φ, Ψ) – RIDGE Ainsi,

$$\begin{cases} P_{32}: a = \hat{\zeta},\ c = \hat{0},\ A = \Psi_1,\ C = t_{X\Phi X},\ b_o = \hat{\beta} \text{ si } \|\hat{\beta}\|^2_{t_{X\phi X}} < \epsilon^2 \\ \quad \text{sinon } b_o = (\Psi_1 + \lambda_o {}^t x\Phi x)^{-1}\ \Psi_1\beta \text{ avec } \|\hat{b_o}\|^2_{t_{X\phi X}} = \epsilon^2 \\ P_{23}: a = \hat{0},\ c = \hat{\beta},\ A = {}^t x\Phi x, C = \Psi_1,\ b_o = 0 \text{ si } \|\hat{\beta}\|^2{}_{\Psi 1} \leqslant \epsilon^2\ ; \\ \quad \text{sinon } b_o = \lambda_o\ ({}^t x\Phi x) + \lambda_o\ \Psi_1)^{-1}\ {}^t X\Phi X\hat{\beta} \text{ avec } \|b_o - \hat{\beta}\|^2 = \epsilon^2 \end{cases}$$

$$(10) \quad \begin{cases} P_{31}: a = \hat{\beta},\ c = \hat{0},\ A = \Psi_1,\ C = \Psi,\ b_o = \hat{\beta} \text{ si } \|\hat{\beta}\|^2_4 \leqslant \epsilon^2\ ; \\ \quad \text{sinon } b_o = (\Psi_1 + \lambda_o\Psi)^{-1}\ \Psi_1\hat{\beta} \text{ avec } \|\hat{b_o}\|^2_{\Psi} = \epsilon^2 \\ P_{13}: a = \hat{0},\ c = \hat{\beta}\ A = \Psi,\ C = \Psi_1,\ b_o = \hat{0} \text{ si } \|\hat{\beta}\|^2_{\Psi_1} \leqslant \epsilon^2\ ; \\ \quad \text{sinon } b_o = \lambda_o\ (\Psi + \lambda_o\Psi_1)^{-1}\Psi\hat{\beta} \text{ avec } \|b_o - \hat{\beta}\|^2 = \epsilon^2 \end{cases}$$

$$\begin{cases} P_{41}: a = \hat{\beta},\ c = \hat{0},\ A = {}^t X\Phi_1 X,\ C = \Psi,\ b_o = \hat{\beta} \text{ si } \|\hat{\beta}\|^2_{\Psi} \leqslant \epsilon^2 \\ \quad \text{sinon } b_o = ({}^t X\Phi_1 X + \lambda_o\Psi)^{-1}\ {}^t X\phi_1 X\hat{\beta} \text{ avec } \|b_o\|^2_{\Psi} = \epsilon^2 \\ P_{42}: a = \hat{\beta},\ c = \hat{0},\ A = {}^t X\Phi_1 X,\ C = {}^t X\Phi X,\ b_o = \hat{\beta} \text{ si } \|\hat{\beta}\|^2_{t_{X\Phi X}} \leqslant \epsilon^2 \\ \text{sinon } b_o = ({}^t X\ \Phi_1 X + \lambda_o\ {}^t X\ \Phi\ X)^{-1}\ {}^t X\Phi_1 X\hat{\beta} \text{ avec } \|b_o\|^2_{t_{X\Phi_1 X}} = \epsilon^2 \end{cases}$$

Les estimateurs RIDGE usuels [2a , 2b] son obtenus dans le problème P_{41} en posant $\phi_1 = I_n$ et $\psi = I_p$.

iii) La famille des estimateurs colinéaires et plus courts que β est communément appelée SHRUNKEN-FAMILY.

On peut se demander à quelle condition les solutions construites sont SHRUNKEN, c'est-à-dire telles que :

pour tout c il existe $\alpha_c, 0 \leqslant \alpha_c < 1$, tel que $b_o = \alpha_c \cdot C$.

La réponse est donnée dans la :

Proposition 2

Dans le cas où $a = \hat{0}$, $c = \hat{\beta}(\omega)$ et sous l'hypothèse que

$$\{\hat{\beta}(\omega),\ \omega \in \Omega\} = \mathbf{R}^{pm}$$

Une condition nécessaire suffisante pour que les solutions définies en (6) et (7) soient SHRUNKEN et qu'il existe $\eta > 0$ tel que $C = \eta A$.

DEMONSTRATION

Si les solutions sont SHRUNKEN, alors on a :

$$(11) \qquad (c \neq O\ ,\ \lambda_o = O) \Leftrightarrow (c \neq O\ ,\ \alpha_c = O)$$

– CONDITION SUFFISANTE

Si $||c||_C^2 \leqslant \epsilon^2$, $b_o = O = Oc$, b_o est colinéaire à c et plus court donc SHRUNKEN.

Si $||c||_C > \epsilon^2$ d'après (7) $Ab_o = \lambda_o C (c - b_o)$. Ceci et $C = \eta A$ entraînent $b_o = \lambda_o \eta (c - b_o)$, c'est-à-dire

$$b_o = (1 - \frac{1}{1 + \lambda_o \eta}) c = (1 - \frac{\epsilon}{||c||_C}) c$$

puisque

$$||b_o - c||_C^2 = \epsilon^2$$

– CONDITION NECESSAIRE – Rappelons que $\lambda_o = \lambda(c)$ dépend de c ce qui complique légèrement la démonstration.

. Le résultat cherché $C = \eta A$ n'est pas déduit de l'étude du cas $||c||_c^2 \leqslant \epsilon^2$ et des solutions SHRUNKEN.

. Si $||c||_C^2 > \epsilon^2$ et les solutions SHRUNKEN, d'après (11) $\alpha_c > 0$,

$$Ab_o = \lambda_c C (c - b_o) \text{ et } b_o = \alpha_c$$

entraînent $\alpha_c Ac = \lambda_c (1 - \alpha_c) Cc$ que l'on note $Ac = \beta_c Cc$ avec $\beta_c = \lambda_c \frac{1 - \alpha_c}{\alpha_c}$ scalaire positif.

Montrons que β_c est indépendant de c. On note $E_u = \{v \mid ||v - u||_C^2 \leqslant \epsilon^2\}$.

Soient c et d des vecteurs tels que $c \notin E_0$, $d \notin E_0$ $(c - d) \notin E_0$). On vérifie aisément que $\beta_c = \beta_d$.

Soit e tel que $e \notin E_0$ et $||e - c||_C > 2\epsilon$.

On a $\beta_e = \beta_c$ pour tout $f \in E_c \cap E_0$ alors $\beta_f = \beta_e = \beta_c$, donc β_c est indépendant de c dans E_0.

On conclut à $C = \eta A$ en notant que si $||c||_C > \epsilon$, $Ac = \frac{1}{\eta} Cc$. Cette relation entre les opérateurs linéaires A et C est en fait vraie pour tout c.

CONCLUSION

1) Les estimateurs RIDGE, SHRUNKEN usuels ou généralisés par le choix des métriques ϕ et ψ sur les espaces de matrices, sont obtenus comme solutions de problèmes quadratiques. Le choix des métriques, qui peuvent ou non dépendre du plan X, permet de construire de nouveaux estimateurs.

2) Les solutions exprimées sous la forme (6) – (7) dépendent du couple de matrices (A, C) ; si b_o est la solution associée à (A, C).

$$b_o = C^{-1/2} b_o'$$

où $C = {}^tC^{1/2} C^{1/2}$ et b_o' la solution associée au couple $({}^tC^{-1/2} AC^{-1/2}, I)$. Les procédures algorithmiques de calcul établies dans la RIDGE REGRESSION pour les couples (D, I) permettent de calculer b_o.

BIBLIOGRAPHIE

[1] BUNKE O. – Improved Inference in Models with Additional Information Math. *Operationforsh. Statist.* (1975).6. Pp 817-829.

[2a] HOERL A.E., KENNARD R.W. – Ridge Regression : biased information for non orthogonal problems. *Technometrics* 1970.12.Pp. 55-67.

[2b] HOERL A.E., KENNARD R.W. – Ridge Regression : Application to non orthogonal problems. *Technometrics* 1970, 12, Pp. 69-82.

[3] HUBER P.J – Robust Regression : Asymptotic, conjectures and Monte Carlo. *Ann. Stat.* 1973-1- Pp. 789-821.

[4] JAMES-STEIN C. – Estimation with quadratic loss. *Proceedings of the fourth Symposium of Math. Stat. and Proba.* 1961-1. pp. 361-379 – Univ. of California Press.

[5] LAUTER H.A. – Minimax linear estimator for linear parameters under restrictions in form of inequalities. *Math Operationforsh. Statist.* (1975) 6 Pp. 689-695.

[6] ROZANOV Y.A. – On the minimax estimator of an Unknown Mean Value *Journal of Multivariante Analysis* – vol. 1, Num. 2, June 1971.

[7] SCLOVE S.L. – Improved estimators for coefficients in linear regression. *JASA* 1968-63. Pp 597-606.

[8] STEIN C. – Multiple Regression in Contribution to Prob. and Stat. Essays in honnor to Harold Hotelling Ed. Olkin Stanford University Press 1960 Pp. 424-443.